EPIGENETIC BIOMARKER AND PERSONALIZED PRECISION MEDICINE

EDITED BY: Jiucun Wang, Dongyi He, Momiao Xiong and Yun Liu
PUBLISHED IN: Frontiers in Genetics and Frontiers in Cell and Developmental Biology



frontiers Research Topics



frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-88966-184-8
DOI 10.3389/978-2-88966-184-8

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

EPIGENETIC BIOMARKER AND PERSONALIZED PRECISION MEDICINE

Topic Editors:
Jiucun Wang, Fudan University, China
Dongyi He, Shanghai Guanghua Rheumatology Hospital, China
Momiao Xiong, University of Texas Health Science Center, United States
Yun Liu, Fudan University, China

Citation: Wang, J., He, D., Xiong, M., Liu, Y., eds. (2020). Epigenetic Biomarker and Personalized Precision Medicine. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88966-184-8

Table of Contents

06 ***Prognostic and Predictive Value of Three DNA Methylation Signatures in Lung Adenocarcinoma***
Yanfang Wang, Haowen Deng, Shan Xin, Kai Zhang, Run Shi and Xuanwen Bao

19 ***Genome-Wide Analysis of Lung Adenocarcinoma Identifies Novel Prognostic Factors and a Prognostic Score***
Donglai Chen, Yueqiang Song, Fuquan Zhang, Xiaofan Wang, Erjia Zhu, Xi Zhang, Gening Jiang, Siguang Li, Chang Chen and Yongbing Chen

30 ***F8 Inversions at Xq28 Causing Hemophilia A are Associated With Specific Methylation Changes: Implication for Molecular Epigenetic Diagnosis***
Muhammad Ahmer Jamil, Amit Sharma, Nicole Nuesgen, Behnaz Pezeshkpoor, André Heimbach, Anne Pavlova, Johannes Oldenburg and Osman El-Maarri

42 ***Epigenetic Changes in the Pathogenesis of Rheumatoid Arthritis***
Marina V. Nemtsova, Dmitry V. Zaletaev, Irina V. Bure, Dmitry S. Mikhaylenko, Ekaterina B. Kuznetsova, Ekaterina A. Alekseeva, Marina I. Beloukhova, Andrei A. Deviatkin, Alexander N. Lukashev and Andrey A. Zamyatnin Jr.

55 ***Transposable Elements and Their Epigenetic Regulation in Mental Disorders: Current Evidence in the Field***
Błażej Misiak, Laura Ricceri and Maria M. Sąsiadek

68 ***Epigenetic IVD Tests for Personalized Precision Medicine in Cancer***
Jesús Beltrán-García, Rebeca Osca-Verdegal, Salvador Mena-Mollá and José Luis García-Giménez

82 ***Gut Microbiota Composition is Associated With the Global DNA Methylation Pattern in Obesity***
Bruno Ramos-Molina, Lidia Sánchez-Alcoholado, Amanda Cabrera-Mulero, Raul Lopez-Dominguez, Pedro Carmona-Saez, Eduardo Garcia-Fuentes, Isabel Moreno-Indias and Francisco J. Tinahones

91 ***DNA Methylation Biomarkers Predict Objective Responses to PD-1/PD-L1 Inhibition Blockade***
Gang Xue, Ze-Jia Cui, Xiong-Hui Zhou, Yue-Xing Zhu, Ying Chen, Feng-Ji Liang, Da-Nian Tang, Bing-Yang Huang, Hong-Yu Zhang, Zhi-Huang Hu, Xi-Yu Yuan and Jianghui Xiong

103 ***A Hybrid Ensemble Approach for Identifying Robust Differentially Methylated Loci in Pan-Cancers***
Qi Tian, Jianxiao Zou, Yuan Fang, Zhongli Yu, Jianxiong Tang, Ying Song and Shicai Fan

115 ***Maternal Smoking During Pregnancy Induces Persistent Epigenetic Changes Into Adolescence, Independent of Postnatal Smoke Exposure and is Associated With Cardiometabolic Risk***
Sebastian Rauschert, Phillip E. Melton, Graham Burdge, Jeffrey M. Craig, Keith M. Godfrey, Joanna D. Holbrook, Karen Lillycrop, Trevor A. Mori, Lawrence J. Beilin, Wendy H. Oddy, Craig Pennell and Rae-Chi Huang

130 ***Predictive and Prognostic Value of Selected MicroRNAs in Luminal Breast Cancer***
Maria Amorim, João Lobo, Mário Fontes-Sousa, Helena Estevão-Pereira, Sofia Salta, Paula Lopes, Nuno Coimbra, Luís Antunes, Susana Palma de Sousa, Rui Henrique and Carmen Jerónimo

146 ***Epigenetic Biomarkers in the Management of Ovarian Cancer: Current Prospectives***
Alka Singh, Sameer Gupta and Manisha Sachan

181 ***A Novel DNA Methylation-Based Signature Can Predict the Responses of MGMT Promoter Unmethylated Glioblastomas to Temozolomide***
Rui-Chao Chai, Yu-Zhou Chang, Qiang-Wei Wang, Ke-Nan Zhang, Jing-Jun Li, Hua Huang, Fan Wu, Yu-Qing Liu and Yong-Zhi Wang

192 ***Comprehensive RNA-Seq Data Analysis Identifies Key mRNAs and lncRNAs in Atrial Fibrillation***
Dong-Mei Wu, Zheng-Kun Zhou, Shao-Hua Fan, Zi-Hui Zheng, Xin Wen, Xin-Rui Han, Shan Wang, Yong-Jian Wang, Zi-Feng Zhang, Qun Shan, Meng-Qiu Li, Bin Hu, Jun Lu, Gui-Quan Chen, Xiao-Wu Hong and Yuan-Lin Zheng

202 ***Essential Role of Histone Replacement and Modifications in Male Fertility***
Tong Wang, Hui Gao, Wei Li and Chao Liu

217 ***Epigenetic Biomarkers in Cardiovascular Diseases***
Carolina Soler-Botija, Carolina Gálvez-Montón and Antoni Bayés-Genís

248 ***Identification of Biomarkers in Neuropsychiatric Disorders Based on Systems Biology and Epigenetics***
Jacob Peedicayil

254 ***Recent Advances in the Targeting of Epigenetic Regulators in B-Cell Non-Hodgkin Lymphoma***
Marcelo L. Ribeiro, Diana Reyes-Garau, Marc Armengol, Miranda Fernández-Serrano and Gaël Roué

278 ***Association of Sperm Methylation at* LINE-1, *Four Candidate Genes, and Nicotine/Alcohol Exposure With the Risk of Infertility***
Wenjing Zhang, Min Li, Feng Sun, Xuting Xu, Zhaofeng Zhang, Junwei Liu, Xiaowei Sun, Aiping Zhang, Yupei Shen, Jianhua Xu, Maohua Miao, Bin Wu, Yao Yuan, Xianliang Huang, Huijuan Shi and Jing Du

287 ***Perspectives on miRNAs as Epigenetic Markers in Osteoporosis and Bone Fracture Risk: A Step Forward in Personalized Diagnosis***
Michela Bottani, Giuseppe Banfi and Giovanni Lombardi

312 ***From Genetics to Epigenetics, Roles of Epigenetics in Inflammatory Bowel Disease***
Zhen Zeng, Arjudeb Mukherjee and Hu Zhang

329 ***Mutations in* KIAA1109, CACNA1C, BSN, AKAP13, CELSR2, *and* HELZ2 *are Associated With the Prognosis in Endometrial Cancer***
Zhiwei Qiao, Ying Jiang, Ling Wang, Lei Wang, Jing Jiang and Jingru Zhang

337 ***Alteration of the DNA Methylation Signature of Renal Erythropoietin-Producing Cells Governs the Sensitivity to Drugs Targeting the Hypoxia-Response Pathway in Kidney Disease Progression***
Koji Sato, Naonori Kumagai and Norio Suzuki

345 ***DNA Methylation Cancer Biomarkers: Translation to the Clinic***
Warwick J. Locke, Dominic Guanzon, Chenkai Ma, Yi Jin Liew, Konsta R. Duesing, Kim Y.C. Fung and Jason P. Ross

367 ***Epigenetics of Bladder Cancer: Where Biomarkers and Therapeutic Targets Meet***
Victor G. Martinez, Ester Munera-Maravilla, Alejandra Bernardini, Carolina Rubio, Cristian Suarez-Cabrera, Cristina Segovia, Iris Lodewijk, Marta Dueñas, Mónica Martínez-Fernández and Jesus Maria Paramio

394 ***Comprehensive Analysis of Therapy-Related Messenger RNAs and Long Noncoding RNAs as Novel Biomarkers for Advanced Colorectal Cancer***
Jibin Li, Siping Ma, Tao Lin, Yanxi Li, Shihua Yang, Wanchuan Zhang, Rui Zhang and Yongpeng Wang

404 ***Testing Mediation Effects in High-Dimensional Epigenetic Studies***
Yuzhao Gao, Haitao Yang, Ruiling Fang, Yanbo Zhang, Ellen L. Goode and Yuehua Cui

417 ***Circulating Serum MicroRNAs as Potential Diagnostic Biomarkers of Posttraumatic Stress Disorder: A Pilot Study***
Clara Snijders, Julian Krauskopf, Ehsan Pishva, Lars Eijssen, Barbie Machiels, Jos Kleinjans, Gunter Kenis, Daniel van den Hove, Myeong Ok Kim, Marco P. M. Boks, Christiaan H. Vinkers, Eric Vermetten, Elbert Geuze, Bart P. F. Rutten and Laurence de Nijs

427 ***E-Cadherin Downregulation is Mediated by Promoter Methylation in Canine Prostate Cancer***
Carlos Eduardo Fonseca-Alves, Priscila Emiko Kobayashi, Antonio Fernando Leis-Filho, Patricia de Faria Lainetti, Valeria Grieco, Hellen Kuasne, Silvia Regina Rogatto and Renee Laufer-Amorim

437 ***The Current State of MicroRNAs as Restenosis Biomarkers***
Nelson Varela, Fernando Lanas, Luis A. Salazar and Tomás Zambrano

447 ***DNA Methylation Biomarkers in Aging and Age-Related Diseases***
Yasmeen Salameh, Yosra Bejaoui and Nady El Hajj

458 ***Detection and Comparative Analysis of Methylomic Biomarkers of Rheumatoid Arthritis***
Xin Feng, Xubing Hao, Ruoyao Shi, Zhiqiang Xia, Lan Huang, Qiong Yu and Fengfeng Zhou

471 ***New Analysis Framework Incorporating Mixed Mutual Information and Scalable Bayesian Networks for Multimodal High Dimensional Genomic and Epigenomic Cancer Data***
Xichun Wang, Sergio Branciamore, Grigoriy Gogoshin, Shuyu Ding and Andrei S. Rodin

frontiers in Genetics

ORIGINAL RESEARCH
published: 24 April 2019
doi: 10.3389/fgene.2019.00349

Prognostic and Predictive Value of Three DNA Methylation Signatures in Lung Adenocarcinoma

Yanfang Wang[1†], *Haowen Deng*[2†], *Shan Xin*[1,3], *Kai Zhang*[4], *Run Shi*[1*] and *Xuanwen Bao*[5,6*]

[1] Ludwig-Maximilians-Universität München, Munich, Germany, [2] Chair for Computer Aided Medical Procedures and Augmented Reality, Technical University Munich, Munich, Germany, [3] Institute of Molecular Toxicology and Pharmacology, Helmholtz Center Munich, German Research Center for Environmental Health, Neuherberg, Germany, [4] Department of Cardiology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China, [5] Institute of Radiation Biology, Helmholtz Center Munich, German Research Center for Environmental Health, Neuherberg, Germany, [6] Technical University Munich (TUM), Munich, Germany

OPEN ACCESS

Edited by:
Jiucun Wang,
Fudan University, China

Reviewed by:
Mariana Brait,
Johns Hopkins University,
United States
David D. Eisenstat,
University of Alberta, Canada
Nejat Dalay,
Istanbul University, Turkey
Yi Wang,
Fudan University, China

***Correspondence:**
Run Shi
shirun@outlook.com
Xuanwen Bao
xuanwen.bao@helmholtz-muenchen.de

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: 20 November 2018
Accepted: 01 April 2019
Published: 24 April 2019

Citation:
Wang Y, Deng H, Xin S, Zhang K, Shi R and Bao X (2019) Prognostic and Predictive Value of Three DNA Methylation Signatures in Lung Adenocarcinoma. Front. Genet. 10:349. doi: 10.3389/fgene.2019.00349

Background: Lung adenocarcinoma (LUAD) is the leading cause of cancer-related mortality worldwide. Molecular characterization-based methods hold great promise for improving the diagnostic accuracy and for predicting treatment response. The DNA methylation patterns of LUAD display a great potential as a specific biomarker that will complement invasive biopsy, thus improving early detection.

Method: In this study, based on the whole-genome methylation datasets from The Cancer Genome Atlas (TCGA) and several machine learning methods, we evaluated the possibility of DNA methylation signatures for identifying lymph node metastasis of LUAD, differentiating between tumor tissue and normal tissue, and predicting the overall survival (OS) of LUAD patients. Using the regularized logistic regression, we built a classifier based on the 3616 CpG sites to identify the lymph node metastasis of LUAD. Furthermore, a classifier based on 14 CpG sites was established to differentiate between tumor and normal tissues. Using the Least Absolute Shrinkage and Selection Operator (LASSO) Cox regression, we built a 16-CpG-based model to predict the OS of LUAD patients.

Results: With the aid of 3616-CpG-based classifier, we were able to identify the lymph node metastatic status of patients directly by the methylation signature from the primary tumor tissues. The 14-CpG-based classifier could differentiate between tumor and normal tissues. The area under the receiver operating characteristic (ROC) curve (AUC) for both classifiers achieved values close to 1, demonstrating the robust classifier effect. The 16-CpG-based model showed independent prognostic value in LUAD patients.

Interpretation: These findings will not only facilitate future treatment decisions based on the DNA methylation signatures but also enable additional investigations into the utilization of LUAD DNA methylation pattern by different machine learning methods.

Keywords: LUAD, DNA methylation, regularized logistic regression, recursive feature elimination, LASSO Cox regression, metastasis

INTRODUCTION

Lung cancer is the leading cause of cancer-related mortality globally, causing over a million deaths a year (Genome Atlas Research Network., 2014; Jemal et al., 2018). There are two clinical types, one is the aggressive subtype small cell lung cancer and the other is non-small cell lung cancer (Hankey et al., 1999). Non-small cell lung cancer is histologically classified into four major subtypes by pathological and molecular characteristics: adenocarcinoma, large cell lung cancer, squamous cell lung cancer, and other types (Ettinger et al., 2010). Adenocarcinoma is the most common histological subtype of non-small cell lung cancer. Tobacco smoking is the major cause of lung adenocarcinoma (Toh et al., 2006). However, with the decrease in the number of smokers in many countries, the occurrence of LUAD in non-smokers has increased (Genome Atlas Research Network., 2014).

An accurate diagnosis of LUAD is one precondition to achieve a better treatment effect. Although the Mayo Clinic stage, size, grade, and necrosis (SSIGN) score, as well as the University of California Integrated Staging System can help improve the accuracy of the prognosis (Travis et al., 2011), the outcomes of patients with similar clinical characteristics or integrated systems scores still differ. Molecular characteristics may provide an indication for predicting the LUAD prognosis and response to therapy, thus offering great potential for improving individual treatment. Moreover, molecular characterization-based methods do not generally require bulk tissue samples, which can improve the patients' tolerance and reduce unnecessary operation steps. Among all the molecular characteristics, DNA methylation of CpG sites plays a crucial role in epigenetic regulation by reducing the activity of a DNA segment and repressing gene transcription (Jones, 2012; Du et al., 2015; Schübeler, 2015). DNA methylation is associated with carcinogenesis by repressing the expression of the tumor suppressor gene and promoting the expression of oncogenes (Herman et al., 1995; Schübeler, 2015; Vizoso et al., 2015; Klutstein et al., 2016). Hence, the cancer tissues have a distinct DNA methylation pattern compared to normal tissues. More importantly, unlike somatic genetic mutations in tumor tissues, DNA methylation patterns are inherently reversible changes and can therefore be promising targets for drug treatments (Ramchandani et al., 1999). Using DNA methylation signatures can help us make a better prognosis and predict the treatment response, thus prolonging the patients' survival.

Machine learning is a novel method to learn concept from data, which will help researchers discover the hidden insights. Based on DNA methylation patterns, machine learning techniques are developed and used to design models for precise classification and accurate prediction in medicine. In this study, we evaluated the possibility of DNA methylation signatures in identifying LUAD lymph node metastasis, differentiating between tumor tissue and normal tissue and predicting the OS of LUAD patients by applying TCGA whole-genome methylation datasets to several machine learning methods. Our results showed robust classifier effects with the AUC of both classifiers achieving values close to one for identifying lymph node metastasis and differentiating between tumor tissue and normal tissue. Cross-validation was applied to prevent overfitting. The LASSO Cox regression model was used to evaluate the patients' OS. Risk scores from the LASSO Cox model were combined with other clinicopathological risk factors to generate a nomogram to predict the prognosis and help the doctors to manage LUAD patients.

METHODS

Data Source

The DNA methylation files and patients' information were obtained from Xena (https://xenabrowser.net/). Complete clinical, molecular, and histopathological data-sets are available at the TCGA website (https://portal.gdc.cancer.gov/).

Feature Selection for DNA CpG Sites

We formulate critical methylation identification as a feature selection problem. Each CpG site is treated as a feature here and our goal is to find out which features are important for different tasks.

Variance Based Filtering

Variance is the squared deviation of the data from its mean, showing the spread of numbers. It is an important characteristic that reflects the distribution and discriminability of a feature. The variance σ of an observed sample sequence of a given feature $\{x_1, x_2, \ldots, x_i, \ldots, x_N\}$ is computed by averaging across the squared difference of each value to the mean μ.

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$
$$\mu = \frac{\sum x_i}{N}$$

In general, a larger variance σ means a wider distributed and more separable feature space, which facilitates training a classifier to find class boundaries. On the other hand, variance σ is positively correlated with information entropy E, meaning that more information could be obtained with a larger variance σ. When σ is small, all the data is compressed and provides insufficient information for a classifier, so that we would avoid features with a small variance by setting a minimum threshold to filter out the indiscriminate features.

$$E = -\int P(x) \log P(x) dx$$

Regularized Logistic Regression Model

Logistic regression is a widely applied and useful statistical, non-linear model for predicting a binarized outcome based on a sequence of independent features. Assuming we have a general linear regression model y, which satisfies

$$y = \sum_{i=0}^{N} \beta_i x_i$$

where x_i stands for the i-th feature and β_i is the correspondent coefficient. Since there is no constraint on the range of β_i and x_i,

there is no maximum or minimum limit for y, i.e., $y \in [-\infty, \infty]$. Consider a standard logistic function

$$f(x) = \frac{1}{1 + e^{-x}}$$

This could map an input space from an infinite $[-\infty, \infty]$ to a finite $[0, 1]$. By combining the linear regression model with the logistic function, we obtain the logistic regression model

$$y = \frac{1}{1 + e^{-\sum_{i=1}^{N} \beta_i x_i}}$$

By thresholding y with threshold t, we obtain the binarized output.

$$o = \begin{cases} 1, y \geq t \\ 0, y < t \end{cases}$$

Note that, regularization term can be compounded with a logistic regression model, to force the learned coefficients to be sparser and more resistant to overfitting, which is highly beneficial for feature selection as well. We term the logistic regression model with regularization term as "Regularized logistic regression model."

Recursive Feature Elimination

Recursive feature elimination (RFE) adopts a brute-force and recursive way of undermining important features. Given a predefined model, which weighs all the features internally, RFE recursively uses the set of features to train the model and discard features that are the least important for the model (e.g., small weights) and repeats the training with the remaining features. This operation keeps recycling until certain expectations are reached, such as the maximum number of expected features N_{exp}. The process is described in **Algorithm 1**.

Algorithm 1: Recursive Feature Elimination.

INPUT: a set of features $S = \{f_1, f_2, \ldots, f_i, \ldots, f_n\}$, expected feature number N_{exp}

OUTPUT: a set of kept features $S_{best} = \{f_{s_1}, f_{s_2}, \ldots, f_{s_i}, \ldots, f_{s_n}\}$

WHILE size $(S) > N_{\text{exp}}$ **DO**

1 Train a model with the set of features in S

2 Get the coefficients for the features learned from the classification model

3 Prune features with small coefficients $S_{non-important}$

4 Update feature set $S = S - S_{non-important}$

END WHILE

Keep the final set of features as the set of most important features $S_{best} = S$

Cox Regression

Cox regression, also called Proportional Hazards Regression, is a survival analysis model. It can be used to analyze relationships between different features and the survival time. The Cox model is based on the proportional hazards condition, which assumes that features have a proportional relationship to the exponential change of hazard. Thus, the model is formulated as a multiplication of a baseline hazard function with a sole time variable t, and an exponential function of the linear combination of all of the features as an input. Given a set of n samples $\{(\boldsymbol{X}_i, Y_i, s_i) \mid 0 \leq i \leq n,\ i \in \boldsymbol{R}\}$, where $X_i = (x_{i0}, x_{i1}, \ldots, x_{ik})$ and stands for the i-th sample of all the k features, Y_i is the observation time and s_i is the survival status, the hazard function is

$$H_i(t) = H_0(t)\, e^{\boldsymbol{X}_i^T \beta}$$

$\boldsymbol{\beta} = (\beta_{\boldsymbol{0}}, \beta_1, \ldots, \beta_k)$ is the coefficient vector weighing the contribution of the features. The partial likelihood of all the samples is

$$\begin{aligned} L(\boldsymbol{\beta}) &= \prod_{i=1}^{n} L_i(\boldsymbol{\beta}) \\ &= \prod_{i=1}^{n} \frac{H_i(Y_i \mid \boldsymbol{X}_i)}{\sum_{j:\, Y_j \geq Y_i} H_i\left(Y_i \mid \boldsymbol{X}_j\right)} \\ &= \prod_{i=1}^{n} \frac{e^{\boldsymbol{X}_i^T \boldsymbol{\beta}}}{\sum_{j:\, Y_j \geq Y_i} e^{\boldsymbol{X}_j^T \boldsymbol{\beta}}} \end{aligned}$$

By penalizing -log $(L(\boldsymbol{\beta}))$, the optimal $\boldsymbol{\beta}$ could be uncovered.

LASSO Regularization

LASSO (Least Absolute Shrinkage and Selection Operator) is an important regularization in many regression analysis methods. The concept behind LASSO is that an L1-norm is used to penalize the weight of the model parameters. Assuming a model has a set of parameters $\{w_0, w_1, \ldots, w_n\}$, the LASSO regularization can be written as

$$\lambda \cdot \sum_{i=0}^{n} \|w_i\|_1$$

It can be also expressed as a constraint to the targeted objective function

$$\sum \left\| \boldsymbol{Y} - \boldsymbol{Y}^* \right\|_2 ,\ s.t.\ \ \|w_i\|_1 < t$$

An important property of the LASSO regularization term is that it can force the parameter values to be 0, thus generating a sparse parameter space, which is a desirable character for feature selection.

Workflow of the Coding Process

When it came to selecting the methylation features for the metastasis and tumor identification problems, we first used

variance-based filtering to eliminate some of the least important CpG sites, and to decrease the computation for the following Regularized Logistic Regression Model and RFE. To avoid model overfitting and bias in the feature selection, cross validation was used in the following stages. The dataset was evenly divided into 5-folds, and further feature selections were conducted by applying Logistic Regression Model and RFE, following the standard pipeline of cross validation.

When predicting the OS of LUAD patients, we built the Cox proportional hazard regression model with LASSO regularization. 5-fold cross-validation was applied to avoid the overfitting. We plotted the plots in R software (R Foundation for Statistical Computing, Vienna, Austria. Version 3.4.3) and Python (Python Software Foundation. Python Language Reference, version 3.7).

RESULTS

Preparation of LUAD DNA Methylation Datasets

LUAD DNA methylation data and corresponding clinical data were downloaded from Xena (https://xenabrowser.net/) (Cline et al., 2013). After removing samples without a survival status and normalization, a total of 478 samples were analyzed in the present study (**Supplementary File 1**). The datasets included 409 samples for the recognition of metastasis, 428 samples for the recognition of tumor from normal tissue, and 446 samples for the prediction of OS [(**Supplementary Files 2**–(**4**].

Identification of 3616-CpG-Based Signature for the Recognition of Metastasis

Variance-based selection was applied to filter features (methylation CpG sites). Features with small variances tend to be less discriminative, so we filtered out features with a standard variance smaller than 0.01 and 135,094 methylation signatures were selected. Regularized logistic regression and cross-validation were then applied to weigh the importance of each feature. The 428 LUAD samples were randomly assigned to a test set or a validation set by the cross-validation method. In short, five rounds of cross-validation were performed using different partitions and the validation results were combined over five rounds to overcome overfitting. By varying the value of the coefficient threshold, we obtained a different number of features that could be kept. When we used those kept features to regress the linear Logistic model by 5-fold cross-validation, the mean accuracy trend was as follows (**Figure 1A**). The number of kept features with regard to the different values of coefficient thresholds was shown in **Figure 1B**. The best performance was achieved at the threshold value 0.05 with 6,198 features kept with a 5-fold cross-validation. Recursive feature elimination with the same cross-validation configuration was tested and the result indicated that the kept features were the optimum minimal set of all the features (**Figure 1C**). The value of kept methylation CpG sites was shown in **Figure 1D**. We assessed the accuracy of the 3616-CpG-based classifier for detecting metastasis with a ROC analysis (**Figure 1E**) with the same cross-validation configuration, and averaged the weights of selected features across different set as the final coefficients. Furthermore, the metastatic probability of each sample were calculated by the coefficients of kept methylation CpG sites (**Figure 1F** and **Supplementary File 5**). The AUC for the classifier achieved values close to 1in all of the 5-fold cross-validation, indicating the robust classifier effect. The tumor tissues in total dataset were divided into high metastatic risk score and low metastatic risk score groups, respectively, using 0.5 as the cutoff. The patients in the low metastatic risk score group have a longer OS than those in the high metastatic risk score group in the total datasets ($p < 0.0001$, **Figure 1G**) as well as in the separated 5-fold training and validation sets ($p < 0.05$, **Supplementary Figure 1**). We assessed the prognostic accuracy of the 3616-CpG-based classifier metastatic classifier with a time-dependent ROC analysis at varying follow-up times (500, 1,000, 1,500, 2,000, 2,500, 3,000 days) (**Supplementary Figure 2**). The accuracy was all around 66%, indicating that the 3616-CpG-based classifier for identifying metastasis could also work well for predicting the OS of LUAD patients.

Identification of 14-CpG-Based Signature to Recognize Tumor and Normal Tissues

134,015 features were kept by variance thresholding (0.01). Regularized Logistic regression and cross-validation were applied to weigh the importance of each feature as mentioned above. An accuracy of 100% can easily be achieved for the number of features range from 14 to 43,246 (**Figure 2A**). The number of kept features with regard to the different thresholding values was shown in **Figure 2B**. Recursive feature elimination with cross-validation was tested and the result indicated that an accuracy of 100% can be achieved when the kept feature numbers reached 14 (**Figure 2C**). 14 CpG sites were kept: cg25774643, cg03502002, cg14789818, cg23479922, cg04864807, cg07915921, cg20146541, cg08862830, cg01016533, cg19191888, cg08094098, cg01912692, cg10707110, cg24103195. The value of kept methylation CpG sites was shown in **Figure 2D**. We then calculated the probability of being tumor for each sample by the coefficients of kept methylation CpG sites (**Supplementary File 6** and **Figure 2E**) in the same tradition as of the recognition of metastasis. The accuracy of the 14-CpG-based classifier was assessed by means of ROC analysis (**Figure 2F**). The results showed that the accuracy reached 100% in all 5-fold cross-validation, indicating the high sensitivity and specificity of the 14-CpG-based classifier in differentiating between LUAD tumor tissues and corresponding normal tissues. Furthermore, we applied the 14-CpG-based classifier on an external dataset to confirm the accuracy of the 14-CgG-based classifier (**Figure 2G**). The AUC value was 98.4% for differentiating the tumor and normal tissues (**Figure 2H**). The analysis before showed the regularized logistic model we applied worked well in different datasets.

Identification of 16-CpG-Based Signature to Predict the OS of LUAD Patients

We used a LASSO Cox regression to build a prognostic model, which selected 16 methylation CpG sites from the CpG sites identified by the DNA methylation 450 k chip: cg00161124, cg01105229, cg03923535, cg10976778, cg12141052, cg12240358,

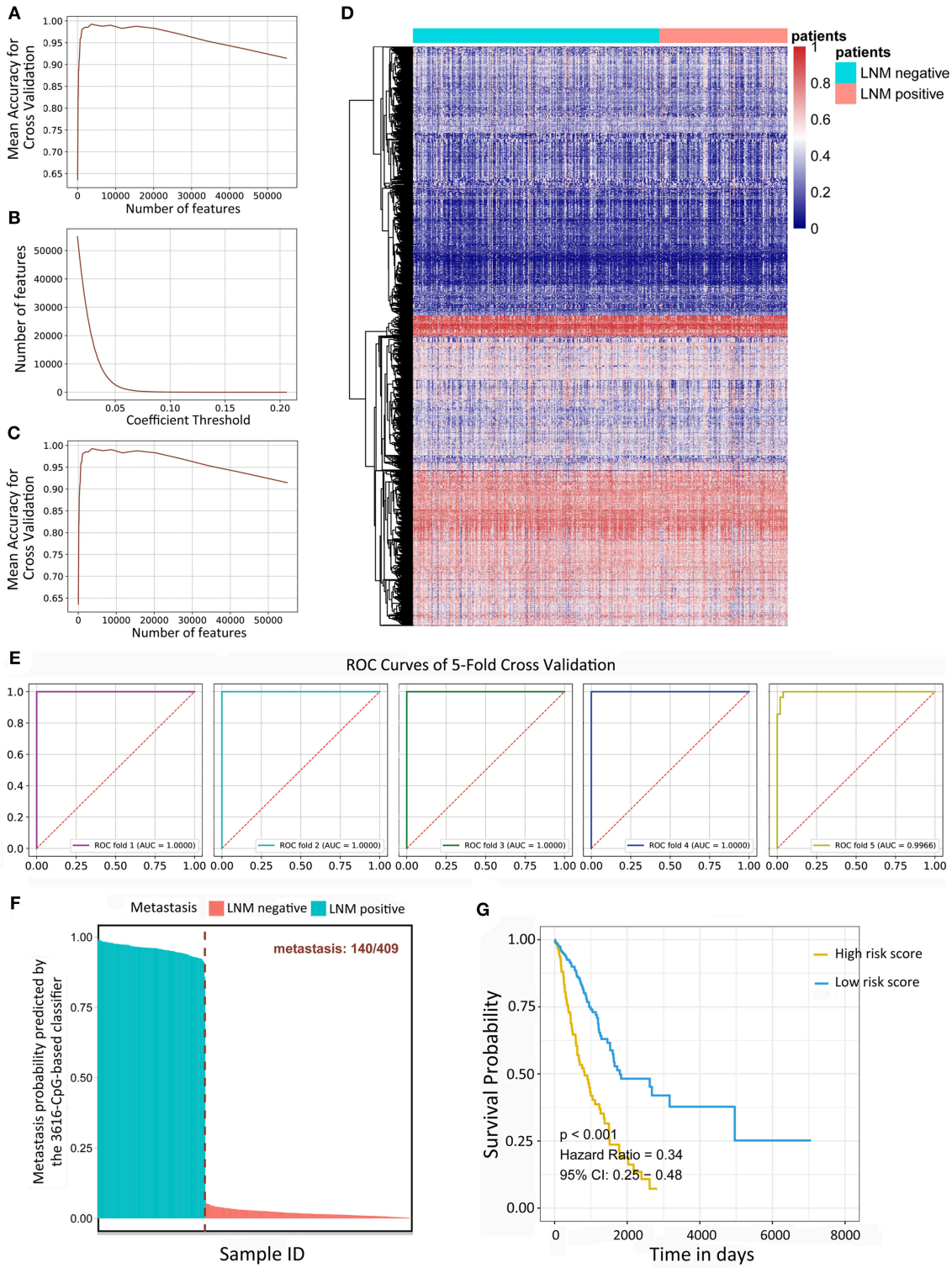


FIGURE 1 | Linear logistic classifier using a 5-fold cross-validation for the metastatic analysis and the efficiency for the 3616-CpG-based classifier. **(A)** The mean accuracy trend for the number of features and mean accuracy for the cross-validation. **(B)** The methylation features with regard to the different values of coefficient thresholds. **(C)** Recursive feature elimination with a cross-validation test. **(D)** Unsupervised hierarchical clustering and heat map associated with the methylation profile (according to the color scale shown) to recognize metastasis in LUAD. **(E)** ROC curves for 5-fold cross-validation showing the high sensitivity and specificity of the classifier in predicting metastasis. **(F)** The metastatic probability of each sample calculated by coefficients of the methylation signature. **(G)** Kaplan-Meier survival analysis for LUAD patients, which are divided into low-risk and high-risk groups using a cutoff value of 0.5. LNM: lymph node metastasis.

cg13297560, cg14139311, cg14184729, cg18140857, cg19410791, cg20268054, cg23146197, cg25229048, cg26709300, cg27018309 (**Figures 3A,B**). The values of the 16 methylation CpG sites for each patient were shown in **Figure 3B**. A formula was derived to calculate the risk score for every patient based on their individual 16 methylation β values (**Supplementary File 7**). The risk scores of tumor samples were calculated by the coefficients of the kept methylation CpG sites (**Figure 3C**). The patients were divided into high risk score and low risk score groups, respectively, with a cutoff of −0·54. Kaplan-Meier survival analysis (**Figure 3D**) showed that the survival probability of patients in lower risk score was significantly better than in high risk score group (log-rank test, all $p < 0.0001$). We assessed the prediction accuracy of the 16-CpG-based model by means of time-dependent ROC analysis at varying follow-up times. The AUC values for 500, 1,000, 1,500, 2,000, 2,500, and 3,000 days were 0.688, 0.681, 0.697, 0.685, 0.738, and 0.758, respectively, which confirmed the effectiveness of the 16-CpG-based model to predict the OS of LUAD patients (**Figure 3E**).

According to their clinicopathological conditions, like epidermal growth factor receptor (EGFR) mutation, K-ras or Ki-ras (KRAS) mutation, lymph node metastatic (LNM) condition, and AJCC stage, LUAD patients were divided up into several subgroups to validate the independent diagnostic value of the methylation signature. EGFR mutation showed a striking correlation with LUAD patient characteristics, which were correlated with the clinical treatment response and then affected the OS of LUAD patients. The Kaplan-Meier curves regarding EGFR mutation and wildtype groups were shown in **Figures 4A,B**. Patients with low risk scores generally had significantly better survival than those with high risk scores in both groups ($p < 0.0001$). Similarly, patients with low risk scores had a significantly longer OS than those with high risk scores in both KRAS mutation and wildtype subgroup and both LNM positive and negative groups (**Figures 4C–F**, $p < 0.0001$). For the patients in AJCC stage I and AJCC stage II-IV, the survival probability of patients with low risk scores was higher than those with low risk scores (**Figures 4G,H**). The stratification analysis above revealed that the 16-CpG-based model could effectively predict the OS of patients regardless of the patients' clinicopathological properties, and provide prognostic power to complement the clinical stage and SSIGN scores.

Lastly, the risk scores were applied to the Cox regression model with the clinicopathological risk factors to perform multivariable survival analysis, thereby generating a nomogram to predict patients' survival probability for 3 and 5 years (**Figure 5A**). In the multivariable survival analysis, we included age, gender, EGFR status, AJCC stage, and risk scores from 16-CpG-based model. The nomogram was further verified with calibration plots (**Figure 5B**). The results showed that the nomogram fared well with the ideal mode for 3 and 5 years, indicating the nomogram worked well in predicting the OS of LUAD patients. According to the risk scores from the nomogram, patients were divided into high risk and low risk group. Kaplan-Meier survival analysis showed that the survival probability of patients with low risk score was significantly higher than those with high risk score (**Figure 5C**, log-rank test, $p < 0.0001$). The prognostic accuracy of the nomogram was further accessed by time-dependent ROC curves (**Figure 5D**). The results showed that the AUC values were all around 0.7 at varying follow-up times (500, 1,000, 1,500, 2,000, 2,500, 3,000 days), indicating the high effectiveness of the nomogram in predicting the prognostic OS of LUAD patients.

DISCUSSION

The present study demonstrates the potential of using DNA methylation signatures to identify the lymph node metastasis of primary LUAD tissues, to differentiate between the LUAD tumor and normal tissues, and to predict the OS of LUAD patients. Invasive biopsy is the gold standard for the validation of tumor tissues and identification of histological subtypes. However, the collection of bulk tissue samples for immunohistochemical (IHC) staining may cause secondary damage to patients. An inadequate tissue yield or quality also creates barriers for the histological diagnosis. Besides, it may be difficult to identify lymph node metastasis during operation. Nowadays, molecular characterization methods provide new insights in pathological diagnosis (Tsou et al., 2007; Selamat et al., 2012; Zhang et al., 2013; Ogino et al., 2016). Since the global change of DNA methylation takes place at the beginning of carcinogenesis, DNA methylation has been considered a promising biomarker for the early detection and diagnosis of cancers (Franco et al., 2008; Hatano et al., 2015; Wu and Ni, 2015), which can complement the pathological IHC staining. Moreover, DNA methylation analysis does not require bulk tissue samples. Small amounts of tissue are enough for DNA extraction and methylation-chip or methylation-seq analysis, which will reduce the patients suffering. Hundreds of thousands of the DNA methylation CpG sites can be identified through genome-wide DNA methylation detection by DNA methylation chips or methylation-seq. Discovering a potential panel of DNA methylation-based biomarkers from the large DNA methylation files can be beneficial for the early diagnosis of cancer initiation and metastasis. Several research studies have shown the potential of utilizing DNA methylation profiles to help the diagnosis of different cancers (Diaz-Lagares et al., 2016; Zhang et al., 2017; Sandanger et al., 2018). One study applied an unsupervised clustering method on DNA methylation profiles to find potential subtypes of childhood B-cell acute lymphoblastic leukemia. The patients were allocated into two subgroups by the unsupervised hierarchical clustering of DNA methylation profiles, which showed a significant association between DNA methylation and disease-free survival (Sandoval et al., 2013a). Another study also utilized a similar strategy to find the association between DNA methylation signatures and the recurrence-free survival in non-small-cell lung cancer samples (Sandoval et al., 2013b). In our study, we applied supervised learning strategy (regularized logistic regression) to find the prognostic CpG cites in LUAD primary tissues. The RFE helped to eliminate the unnecessary features in regression, which constrained the numbers of key CpG sites for prognosis. Besides, LASSO Cox regression was useful to reduce the feature numbers in the COX survival analysis. One study built a

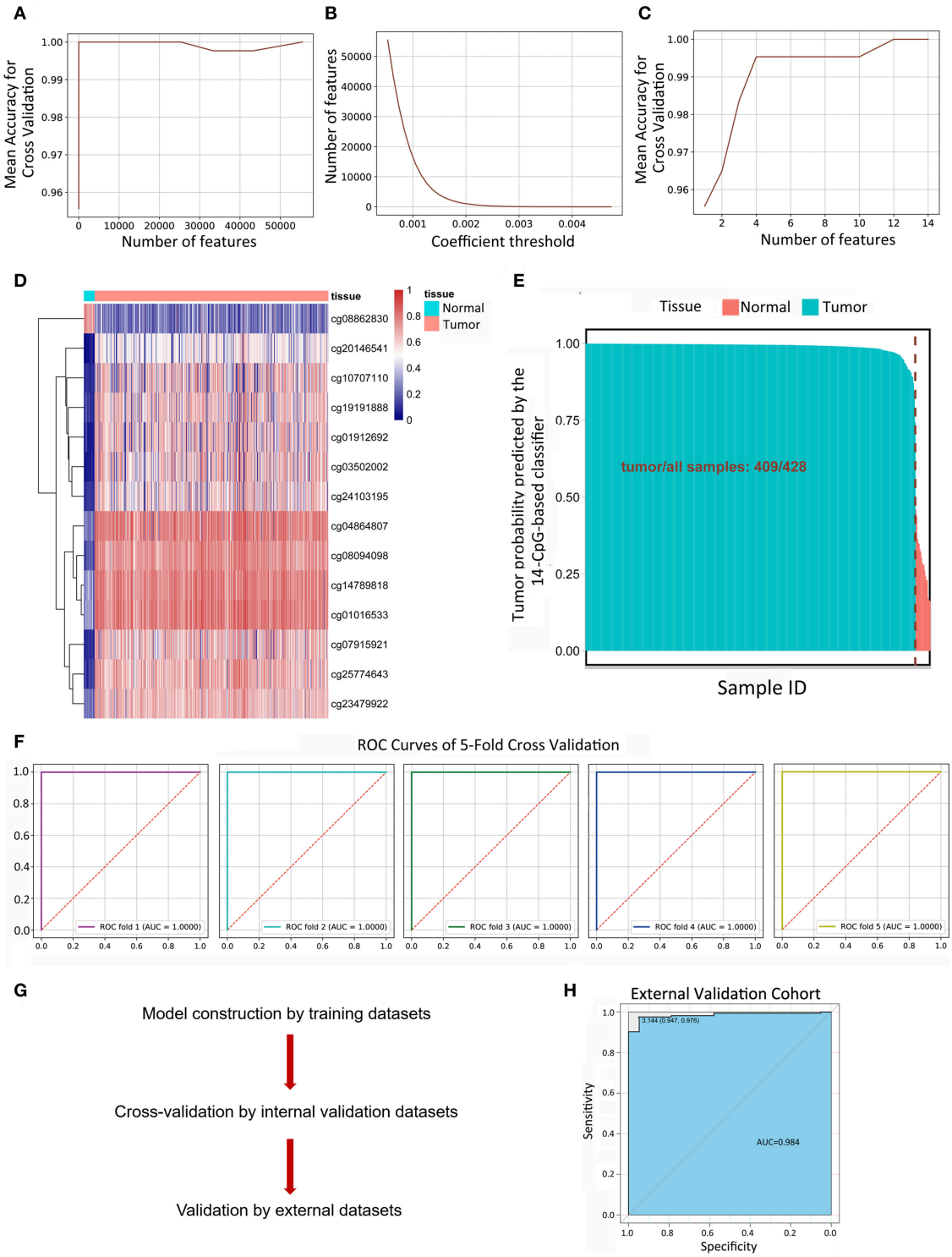


FIGURE 2 | Linear logistic classifier using a 5-fold cross-validation to differentiate between LUAD tumor tissues and corresponding normal tissues, and the efficiency for the 14-CpG-based classifier. **(A)** The mean accuracy trend for the number of features and mean accuracy for the cross-validation. **(B)** The methylation features with regard to the different values of coefficient thresholds. **(C)** Recursive feature elimination with a cross-validation test. **(D)** Unsupervised hierarchical clustering and heat map associated with the methylation profile (according to the color scale shown) to differentiate between LUAD tumor tissues and corresponding normal tissues. **(E)** The probability of being tumor for each sample calculated by the coefficients of methylation signatures. **(F)** ROC curves showing the high sensitivity and specificity in differentiating between LUAD tumor tissues and normal tissues. **(G)** The workflow of model construction, internal validation and external validation. **(H)** ROC curve showing the high sensitivity and specificity in differentiating between LUAD tumor tissues and corresponding normal tissues on an external dataset.

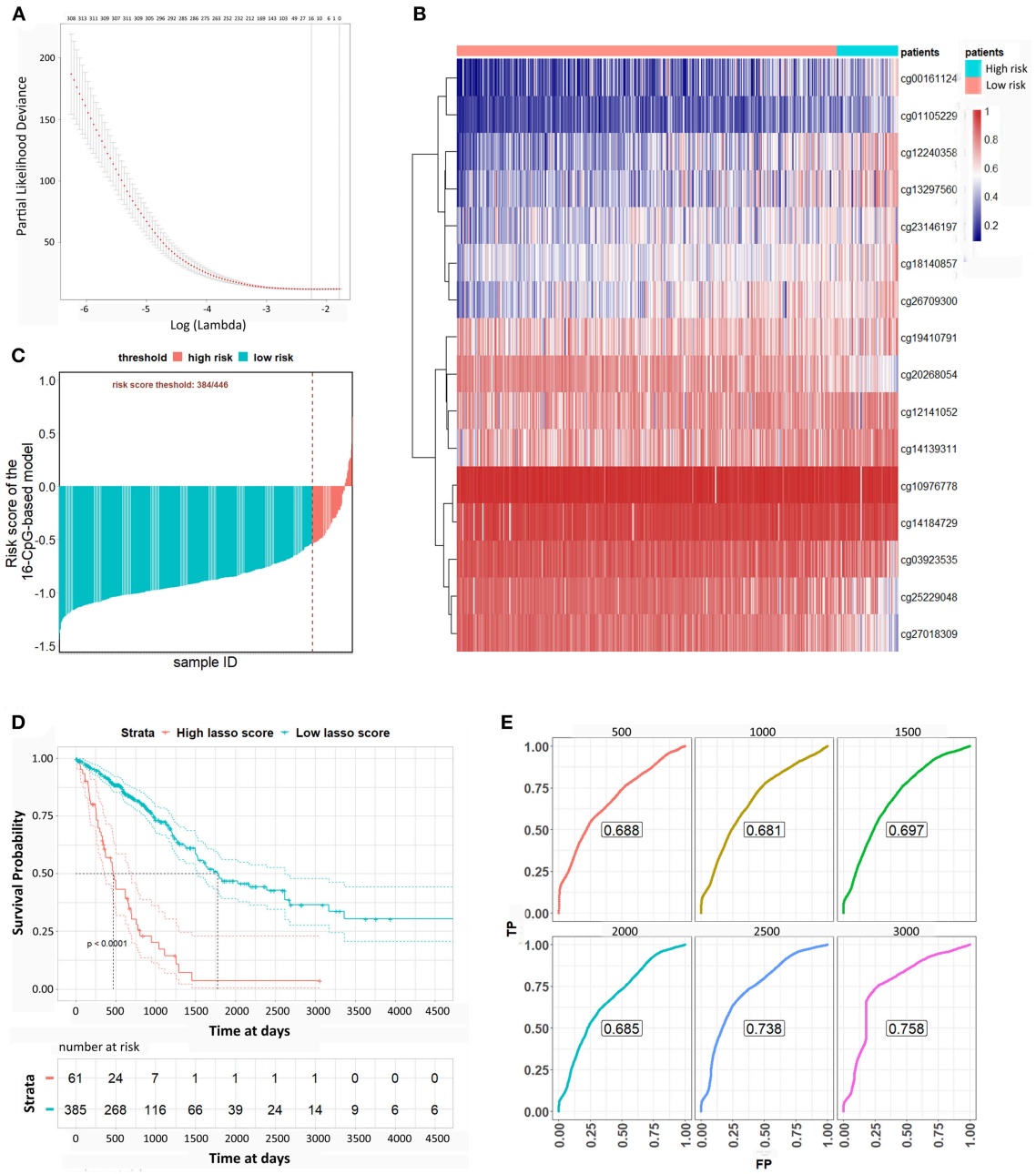


FIGURE 3 | Lasso Cox analysis to predict the OS of LUAD patients. **(A)** The selection of λ value for Lasso Cox regression. **(B)** Unsupervised hierarchical clustering and heat map associated with the methylation profile (according to the color scale shown) to predict the OS of LUAD patients. **(C)** The risk scores of each sample calculated by the coefficients of methylation signatures from Lasso Cox analysis. **(D)** Kaplan-Meier curves of LUAD patients with a low or high risk of death, according to risk scores from the 16-CpG-based classifier. **(E)** Time-dependent ROC analysis at varying follow-up times (500, 1,000, 1,500, 2,000, 2,500, 3,000 days). We used AUC values at 500, 1,000, 1,500, 2,000, 2,500, 3000 days to assess the prognostic accuracy.

prognostic signature by LASSO Cox regression to predict the progression-free survival of LUAD patients and demonstrated the potential biological significance of DNA methylation in the etiology of LUAD (Bjaanæs et al., 2016). Another study built a mortality risk score by LASSO Cox regression (Zhang et al., 2017). The signature based on ten selected CpG sites exhibited strong association with all-cause mortality. Moreover, one recent study used blood-derived DNA methylation and gene expression profiles to identify CpG lung cancer markers prior to diagnosis. They emphasized the difference of prognostic CpG sites in smoking and non-smoking lung cancer patients (Sandanger et al., 2018). In this study, based on the methylation

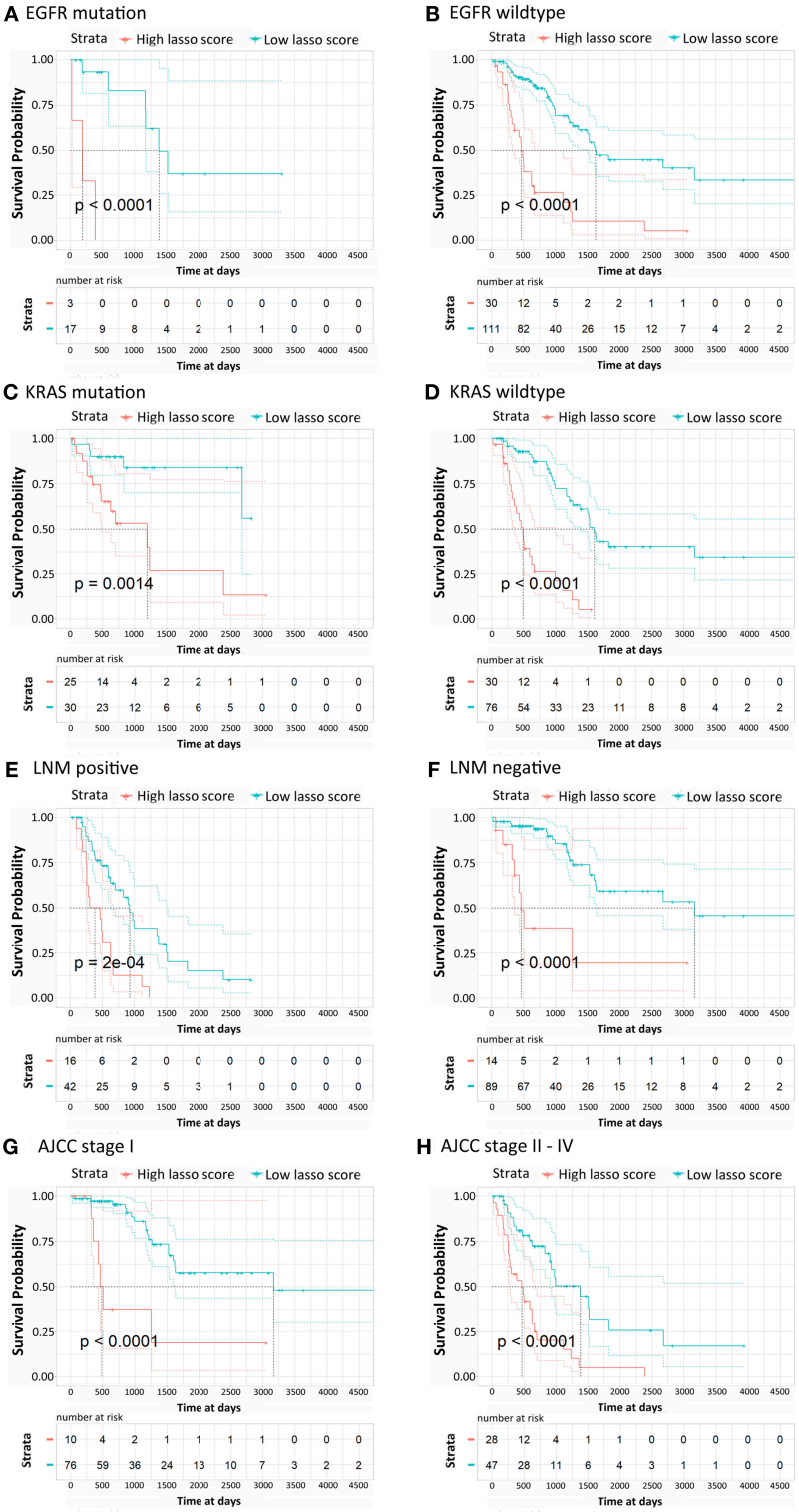


FIGURE 4 | Kaplan-Meier survival analysis for LUAD patients according to the 16-CpG-based classifier. Patients were classified according to clinicopathological risk factors. **(A,B)** EGFR status; **(C,D)** KRAS status; **(E,F)** lymph node metastatic (LNM) status; **(G,H)** AJCC stage I and II-IV. The patients were divided into low-risk and high-risk groups. *P*-values were calculated using the log-rank test.

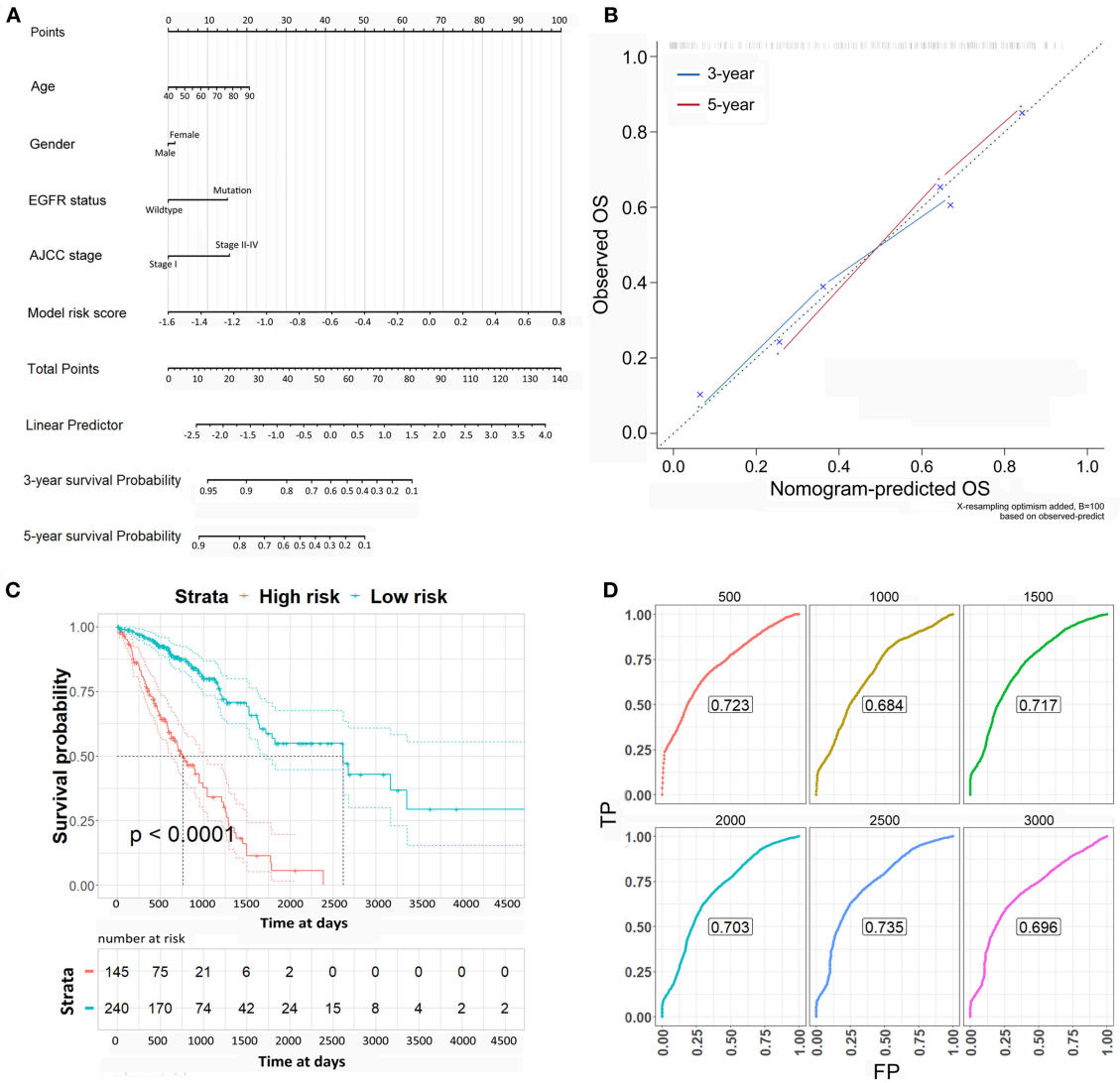


FIGURE 5 | The nomogram to predict the risk of death in 3 and 5 years of LUAD patients. **(A)** The nomogram to predict the risk of death in 3 and 5 years of LUAD patients. **(B)** Plots depict the calibration of each model in terms of agreement between predicted and observed 3- and 5-year outcomes. The dashed line represents the ideal line of a perfect match between nomogram-predicted and observed OS. The blue and red line indicate the performance of the proposed nomogram for 3 and 5 years, respectively. **(C)** Kaplan-Meier survival analysis for the OS of LUAD patients according to the risk scores from the nomogram. **(D)** Time-dependent ROC curves from the nomogram for overall survival in 3 and 5 years.

profiles of LUAD patients, we performed regularized logistic regression and LASSO Cox regression to identify the lymph node metastasis, to differentiate between tumor and normal tissues and to predict the OS of LUAD patients. From the primary LUAD tumor tissues, 3616 methylation CpG sites were kept to build a classifier to identify LUAD lymph node metastasis. ROC curves showed the high sensitivity and specificity of the 3616-CpG-based classifier in identifying lymph node metastasis from CpG sites of primary tumor tissues. All the samples came from the primary tumor tissues, which means that the metastatic behavior can be identified even without extracting tissues from lymph nodes. Therefore, it would work as a biomarker to predict the diagnosis of lymph node metastasis. Since the metastatic behavior of LUAD affects the OS of LUAD patients dramatically, we applied the metastatic classifier to check whether the model can be used to predict the OS of LUAD patients. The time-dependent ROC curves showed the effectiveness of the metastatic classifier in predicting the OS of LUAD patients at varying follow-up times. As expected, the patients in the high metastatic risk score group have a significantly worse OS than those in the low metastatic risk score group.

Tumor tissues are heterogeneous tissues that include cancer cells (epithelial cells), cancer stem cells, vascular epithelial cells and so on (Reya et al., 2001; Marusyk et al., 2012). More than 70% of the tumor tissues are cancer cells. The heterogeneity of tumor tissues may influence the accuracy of the diagnosis. We compared

the heterogeneous tumor tissues with the normal tissues, which also include the vascular epithelial cells and other cell types. Considering the heterogeneous tumor tissues and heterogeneous normal tissues as a whole for each, we tried to eliminate the influence brought about by heterogeneity (Li et al., 2014). In this study, to differentiate between tumor and normal tissues, we concluded that 14 CpG methylation sites were enough for the diagnosis. To check the overfitting potential, we applied 5-fold cross-validation. The efficiency of the model above was tested by a ROC curves in five different training and validation datasets, which showed the high efficiency and specificity of the 14-CpG-based classifier in differentiating between LUAD tumor tissues and the normal tissues. Furthermore, we also validated our regression model on the external dataset from one study (Bjaanæs et al., 2016). Results showed an AUC value of 98.4% to differentiate the tumor and normal tissues by the ROC analysis. The external dataset further confirmed the accuracy of the regularized logistic model which we applied to build the both classifiers above. From the two classifiers, we obtained an overlap cluster of CpG sites: cg03502002 and cg07915921. The information for cg07915921 is not clear. cg03502002 is on the CpG island of the promotor region of the GALR1 gene. The methylation status of the GALR1 promoter and the level of GALR1 gene expression have been correlated in a large number of head and neck squamous tumor specimens (Misawa et al., 2008). Ectopic expression of GALR1 suppresses tumor cell proliferation through Erk1/2-mediated regulation of cyclin-dependent kinase inhibitors and cyclin D1 (Kanazawa et al., 2009). One study revealed that hypermethylated GALR1 plays important roles in smoking-associated LUAD (Tan et al., 2013).

We also built a model to predict the OS of LUAD patients by means of methylation CpG sites. The LASSO Cox regression model generated risk score for each patient. When we assessed the survival status and distribution of risk scores, patients with low risk scores generally had a better OS than those with high risk scores. The model will help guide individualized follow-up schedules for LUAD patients. The high-risk patients have poor OS prediction. This could be the basis of a future clinical trial. The LASSO Cox regression results were further confirmed by the time-dependent ROC analysis. When we compared the time-dependent ROC from the OS-prediction model and metastasis-prediction classifier, the OS-prediction model turned out to be more precise in the long-term survival prediction while the metastasis-prediction classifier worked better in the short-term survival prediction. One explanation could be that when the LUAD patients were accompanied by lymph node metastasis, the tumor progressed and the patients had a poorer prognosis. The OS expectation of patients with lymph node metastasis was shorter than those without lymph node metastasis. Hence, the metastasis-prediction classifier would work better for the short-term prediction.

To further utilize the risk scores from the Cox regression model, we classified patients into several subgroups according to the clinicopathological risk factors (EGFR mutation, KRAS mutation, LNM status and AJCC stages). The 16-CpG-based classifier still showed clinical and statistical significance regardless of the clinicopathological status of LUAD patients.

The independent prognostic values of the 16-CpG-based model were validated by multivariable survival analysis, which integrated other clinicopathological risk factors for the OS of LUAD patients. The Cox regression risk scores were applied together with age, gender, EGFR status, AJCC stages as indicators to generate a nomogram to predict the 3- and 5-year survival probability. We verified the performance of the nomogram by calibration plots. The predicted OS of LUAD patients by the nomogram was highly consistent with the observed 3- and 5-year OS of LUAD patients. Log-rank test and time-dependent ROC curves at vary follow-up times further confirmed the nomogram. Thus, the nomogram could provide an accurate and simple prognostic prediction for LUAD patients.

In previous studies, mRNA expression files (Beer et al., 2002), the mutation of key genes (Takano et al., 2008; Kosaka et al., 2009), long no-coding RNA expressions (Kosaka et al., 2009; Huarte, 2015; Zhou et al., 2016), and histone modifications (Seligson et al., 2009; Zhou et al., 2016) showed the prognostic potential for different types of cancer. Here, we emphasized that the methylation patterns could also be a meaningful tool for the prognosis of LUAD patients. Some studies have identified that multiple CpG sites are differentially methylated in lung cancer compared to normal tissues (Genome Atlas Research Network., 2014; Poirier et al., 2015; Hao et al., 2017). The key for methylation pattern-based early diagnosis is the identification of crucial CpG sites in LUAD. The use of supervised machine learning methods allowed us to integrate all methylation CpG sites identified by the methylation chip into one model, which improved the prognostic accuracy over that of a single CpG site alone. Our findings show that three DNA CpG signature-based models can effectively identify lymph node metastasis by the CpG sites from primary tumor tissues, differentiate between tumor and normal tissues, and predict the OS of LUAD patients. The tissues would be collected by preoperative biopsy or at surgery. The classifiers for identifying lymph node metastasis and differentiation between tumor and normal tissues would help the preoperative diagnosis. The Lasso Cox model would be helpful for adjuvant treatment and prognostic planning. Therefore, the 3 methylation signatures could be of great value in assessing the status, predicting prognosis and achieving individualized treatments of LUAD patients.

The limitations of our study should be mentioned. The methylation 450 k chip did not identify as many CpG sites as the methylation 850 k chip or methylation sequencing. The methylation CpG site candidates identified here did not represent the complete CpG sites in the genome of LUAD patients.

In conclusion, we built three DNA CpG signature-based models to identify LUAD lymph node metastasis by the CpG sites from primary tumor tissues, differentiate between tumor tissue and normal tissue, and predict the OS of LUAD patients, which highlight the relationship between clinical results (metastasis, survival) and methylation biomarkers in LUAD patients. The nomogram comprising LASSO Cox risk scores and clinicopathological factors may help predict the OS of LUAD patients and help individualized treatment of LUAD patients.

AUTHOR CONTRIBUTIONS

XB and RS conceived and designed the experiments. XB and HD wrote the code (The Python code is available due to request). XB, YW, and SX wrote the paper. RS and KZ reviewed the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We greatly thank Dr. Michael Rosemann and Prof. Dr. Michael J. Atkinson for helpful discussions and suggestions. We sincerely thank Sino-German (CSC-DAAD) Postdoc Scholarship Program for supporting the research of YW, and greatly thank the China Scholarship Council (CSC) for supporting the research and work of SX, RS, and XB.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00349/full#supplementary-material

Supplementary Figure 1 | Kaplan-Meier survival according to risk scores from the 3616-CpG-based classifier in the training, validation sets for 5-fold cross-validation. **(A, C, E, G, I)** The training sets 1-5. **(B, D, F, H, J)** The validation sets 1-5.

Supplementary Figure 2 | Time-dependent ROC analysis at varying follow-up times (500, 1,000, 1,500, 2,000, 2,500, 3,000 days) according to risk scores from the 3616-CpG-based classifier.

REFERENCES

Beer, D. G., Kardia, S. L., Huang, C.-C., Giordano, T. J., Levin, A. M., Misek, D. E., et al. (2002). Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat. Med.* 8, 816–824. doi: 10.1038/nm733

Bjaanæs, M. M., Fleischer, T., Halvorsen, A. R., Daunay, A., Busato, F., Solberg, S., et al. (2016). Genome-wide DNA methylation analyses in lung adenocarcinomas: association with EGFR, KRAS and TP53 mutation status, gene expression and prognosis. *Mol. Oncol.* 10, 330–343. doi: 10.1016/j.molonc.2015.10.021

Cline, M. S., Craft, B., Swatloski, T., Goldman, M., Ma, S., Haussler, D., et al. (2013). Exploring TCGA pan-cancer data at the UCSC cancer genomics browser. *Sci. Rep.* 3:2652. doi: 10.1038/srep02652

Diaz-Lagares, A., Mendez-Gonzalez, J., Hervas, D., Saigi, M., Pajares, M. J., Garcia, D., et al. (2016). A novel epigenetic signature for early diagnosis in lung cancer. *Clin. Cancer Res.* 22, 3361–3371. doi: 10.1158/1078-0432.CCR-15-2346

Du, J., Johnson, L. M., Jacobsen, S. E., and Patel, D. J. (2015). DNA methylation pathways and their crosstalk with histone methylation. *Nat. Rev. Mol. Cell Biol.* 16, 519–532. doi: 10.1038/nrm4043

Ettinger, D. S., Akerley, W., Bepler, G., Blum, M. G., Chang, A., Cheney, R. T., et al. (2010). Non–small cell lung cancer. *J. Natl. Compreh. Cancer Net.* 8, 740–801. doi: 10.6004/jnccn.2010.0056

Franco, R., Schoneveld, O., Georgakilas, A. G., and Panayiotidis, M. I. (2008). Oxidative stress, DNA methylation and carcinogenesis. *Cancer Lett.* 266, 6–11. doi: 10.1016/j.canlet.2008.02.026

Genome Atlas Research Network. (2014). Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511, 543–550. doi: 10.1038/nature13385

Hankey, B. F., Ries, L. A., and Edwards, B. K. (1999). The surveillance, epidemiology, and end results program: a national resource. *Cancer Epidemiol. Prev. Biomarkers* 8, 1117–1121.

Hao, X., Luo, H., Krawczyk, M., Wei, W., Wang, W., Wang, J., et al. (2017). DNA methylation markers for diagnosis and prognosis of common cancers. *Proc. Natl. Acad. Sci. U.S.A.* 114, 7414–7419. doi: 10.1073/pnas.1703577114

Hatano, Y., Semi, K., Hashimoto, K., Lee, M. S., Hirata, A., Tomita, H., et al. (2015). Reducing DNA methylation suppresses colon carcinogenesis by inducing tumor cell differentiation. *Carcinogenesis* 36, 719–729. doi: 10.1093/carcin/bgv060

Herman, J. G., Merlo, A., Mao, L., Lapidus, R. G., Issa, J.-P. J., Davidson, N. E., et al. (1995). Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. *Cancer Res.* 55, 4525–4530.

Huarte, M. (2015). The emerging role of lncRNAs in cancer. *Nat. Med.* 21, 1253–1261. doi: 10.1038/nm.3981

Jemal, A., Miller, K. D., Ma, J., Siegel, R. L., Fedewa, S. A., Islami, F., et al. (2018). Higher lung cancer incidence in young women than young men in the United States. *N. Engl. J. Med.* 378, 1999–2009. doi: 10.1056/NEJMoa1715907

Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* 13, 484–492. doi: 10.1038/nrg3230

Kanazawa, T., Kommareddi, P. K., Iwashita, T., Kumar, B., Misawa, K., Misawa, Y., et al. (2009). Galanin receptor subtype 2 suppresses cell proliferation and induces apoptosis in p53 mutant head and neck cancer cells. *Clin. Cancer Res.* 15, 2222–2230. doi: 10.1158/1078-0432.CCR-08-2443

Klutstein, M., Nejman, D., Greenfield, R., and Cedar, H. (2016). DNA methylation in cancer and aging. *Cancer Res.* 76, 3446–3450. doi: 10.1158/0008-5472.CAN-15-3278

Kosaka, T., Yatabe, Y., Onozato, R., Kuwano, H., and Mitsudomi, T. (2009). Prognostic implication of EGFR, KRAS, and TP53 gene mutations in a large cohort of Japanese patients with surgically treated lung adenocarcinoma. *J. Thorac. Oncol.* 4, 22–29. doi: 10.1097/JTO.0b013e3181914111

Li, X., Shi, Y., Yin, Z., Xue, X., and Zhou, B. (2014). An eight-miRNA signature as a potential biomarker for predicting survival in lung adenocarcinoma. *J. Transl. Med.* 12:159. doi: 10.1186/1479-5876-12-159

Marusyk, A., Almendro, V., and Polyak, K. (2012). Intra-tumour heterogeneity: a looking glass for cancer? *Nat. Rev. Cancer* 12, 323–334. doi: 10.1038/nrc3261

Misawa, K., Ueda, Y., Kanazawa, T., Misawa, Y., Jang, I., Brenner, J. C., et al. (2008). Epigenetic inactivation of galanin receptor 1 in head and neck cancer. *Clin. Cancer Res.* 14, 7604–7613. doi: 10.1158/1078-0432.CCR-07-4673

Ogino, S., Nishihara, R., VanderWeele, T. J., Wang, M., Nishi, A., Lochhead, P., et al. (2016). The role of molecular pathological epidemiology in the study of neoplastic and non-neoplastic diseases in the era of precision medicine. *Epidemiology* 27, 602–611. doi: 10.1097/EDE.0000000000000471

Poirier, J. T., Gardner, E. E., Connis, N., Moreira, A. L., De Stanchina, E., Hann, C. L., et al. (2015). DNA methylation in small cell lung cancer defines distinct disease subtypes and correlates with high expression of EZH2. *Oncogene* 34, 5869–5878. doi: 10.1038/onc.2015.38

Ramchandani, S., Bhattacharya, S. K., Cervoni, N., and Szyf, M. (1999). DNA methylation is a reversible biological signal. *Proc. Natl. Acad. Sci. U.S.A.* 96, 6107–6112. doi: 10.1073/pnas.96.11.6107

Reya, T., Morrison, S. J., Clarke, M. F., and Weissman, I. L. (2001). Stem cells, cancer, and cancer stem cells. *Nature* 414, 105–111. doi: 10.1038/35102167

Sandanger, T. M., Nøst, T. H., Guida, F., Rylander, C., Campanella, G., Muller, D. C., et al. (2018). DNA methylation and associated gene expression in blood prior to lung cancer diagnosis in the Norwegian Women and Cancer cohort. *Sci. Rep.* 8:16714. doi: 10.1038/s41598-018-34334-6

Sandoval, J., Heyn, H., Méndez-González, J., Gomez, A., Moran, S., Baiget, M., et al. (2013a). Genome-wide DNA methylation profiling predicts relapse in childhood B-cell acute lymphoblastic leukaemia. *Br. J. Haematol.* 160, 406–409. doi: 10.1111/bjh.12113

Sandoval, J., Mendez Gonzalez, J., Nadal, E., Chen, G., Carmona, F. J., Sayols, S., et al. (2013b). A prognostic DNA methylation signature for stage I non-small-cell lung cancer. *J. Clin. Oncol.* 31, 4140–4147. doi: 10.1200/JCO.2012.48.5516

Schübeler, D. (2015). Function and information content of DNA methylation. *Nature* 517, 321–326. doi: 10.1038/nature14192

Selamat, S. A., Chung, B. S., Girard, L., Zhang, W., Zhang, Y., Campan, M., et al. (2012). Genome-scale analysis of DNA methylation in lung

adenocarcinoma and integration with mRNA expression. *Genome Res.* 22, 1197–1211. doi: 10.1101/gr.132662.111

Seligson, D. B., Horvath, S., McBrian, M. A., Mah, V., Yu, H., Tze, S., et al. (2009). Global levels of histone modifications predict prognosis in different cancers. *Am. J. Pathol.* 174, 1619–1628. doi: 10.2353/ajpath.2009.080874

Takano, T., Fukui, T., Ohe, Y., Tsuta, K., Yamamoto, S., Nokihara, H., et al. (2008). EGFR mutations predict survival benefit from gefitinib in patients with advanced lung adenocarcinoma: a historical comparison of patients treated before and after gefitinib approval in Japan. *J. Clin. Oncol.* 26, 5589–5595. doi: 10.1200/JCO.2008.16.7254

Tan, Q., Wang, G., Huang, J., Ding, Z., Luo, Q., Mok, T., et al. (2013). Epigenomic analysis of lung adenocarcinoma reveals novel DNA methylation patterns associated with smoking. *Onco. Targets. Ther.* 6:1471–1479. doi: 10.2147/OTT.S51041

Toh, C.-K., Gao, F., Lim, W.-T., Leong, S.-S., Fong, K.-W., Yap, S.-P., et al. (2006). Never-smokers with lung cancer: epidemiologic evidence of a distinct disease entity. *J. Clin. Oncol.* 24, 2245–2251. doi: 10.1200/JCO.2005.04.8033

Travis, W. D., Brambilla, E., Noguchi, M., Nicholson, A. G., Geisinger, K. R., Yatabe, Y., et al. (2011). International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. *J. Thorac. Oncol.* 6, 244–285. doi: 10.1097/JTO.0b013e318206a221

Tsou, J. A., Galler, J. S., Siegmund, K. D., Laird, P. W., Turla, S., Cozen, W., et al. (2007). Identification of a panel of sensitive and specific DNA methylation markers for lung adenocarcinoma. *Mol. Cancer* 6:70. doi: 10.1186/1476-4598-6-70

Vizoso, M., Puig, M., Carmona, F. J., Maqueda, M., Velásquez, A., Gómez, A., et al. (2015). Aberrant DNA methylation in non-small cell lung cancer-associated fibroblasts. *Carcinogenesis* 36, 1453–1463. doi: 10.1093/carcin/bgv146

Wu, Q., and Ni, X. (2015). ROS-mediated DNA methylation pattern alterations in carcinogenesis. *Curr. Drug Targets* 16, 13–19. doi: 10.2174/1389450116666150113121054

Zhang, J.-X., Song, W., Chen, Z.-H., Wei, J.-H., Liao, Y.-J., Lei, J., et al. (2013). Prognostic and predictive value of a microRNA signature in stage II colon cancer: a microRNA expression analysis. *Lancet Oncol.* 14, 1295–1306. doi: 10.1016/S1470-2045(13)70491-1

Zhang, Y., Wilson, R., Heiss, J., Breitling, L. P., Saum, K.-U., Schöttker, B., et al. (2017). DNA methylation signatures in peripheral blood strongly predict all-cause mortality. *Nat. Commun.* 8:14617. doi: 10.1038/ncomms14617

Zhou, M., Xu, W., Yue, X., Zhao, H., Wang, Z., Shi, H., et al. (2016). Relapse-related long non-coding RNA signature to improve prognosis prediction of lung adenocarcinoma. *Oncotarget* 7, 29720–29738. doi: 10.18632/oncotarget.8825

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Wang, Deng, Xin, Zhang, Shi and Bao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 22 May 2019
doi: 10.3389/fgene.2019.00493

Genome-Wide Analysis of Lung Adenocarcinoma Identifies Novel Prognostic Factors and a Prognostic Score

Donglai Chen[1†], Yueqiang Song[2†], Fuquan Zhang[3†], Xiaofan Wang[3], Erjia Zhu[1], Xi Zhang[2], Gening Jiang[1], Siguang Li[2], Chang Chen[1]* and Yongbing Chen[3]**

[1] Department of Thoracic Surgery, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai, China,
[2] Department of Regenerative Medicine, Stem Cell Center, Tongji University School of Medicine, Shanghai, China,
[3] Department of Thoracic Surgery, The Second Affiliated Hospital of Soochow University, Medical College of Soochow University, Suzhou, China

OPEN ACCESS

Edited by:
Dongyi He,
Shanghai Guanghua Rheumatology Hospital, China

Reviewed by:
Steven G. Gray,
St. James's Hospital, Ireland
David D. Eisenstat,
University of Alberta, Canada

***Correspondence:**
Siguang Li
siguangli@163.com
Chang Chen
chenthoracic@163.com
Yongbing Chen
chentongt@sina.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: 22 January 2019
Accepted: 06 May 2019
Published: 22 May 2019

Citation:
Chen D, Song Y, Zhang F, Wang X, Zhu E, Zhang X, Jiang G, Li S, Chen C and Chen Y (2019) Genome-Wide Analysis of Lung Adenocarcinoma Identifies Novel Prognostic Factors and a Prognostic Score. Front. Genet. 10:493. doi: 10.3389/fgene.2019.00493

Background and Objective: Lung adenocarcinoma (LUAD) is the most common histological type of all lung cancers and is associated with genetic and epigenetic aberrations. The tumor, node, and metastasis (TNM) stage is the most authoritative indicator of the clinical outcome in LUAD patients in current clinical practice. In this study, we attempted to identify novel genetic and epigenetic modifications and integrate them as a predictor of the prognosis for LUAD, to supplement the TNM stage with additional information.

Methods: A dataset of 445 patients with LUAD was obtained from The Cancer Genome Atlas database. Both genetic and epigenetic aberrations were screened for their prognostic impact on overall survival (OS). A prognostic score (PS) integrating all the candidate prognostic factors was then developed and its prognostic value validated.

Results: A total of two micro-RNAs, two mRNAs and two DNA methylation sites were identified as prognostic factors associated with OS. The low- and high-risk patient groups, divided by their PS level, showed significantly different OS ($p < 0.001$) and recurrence-free survival (RFS; $p = 0.005$). Patients in the early stages (stages I/II) and advanced stages (stages III/IV) of LUAD could be further subdivided by PS into four subgroups. PS remained efficient in stratifying patients into different OS ($p < 0.001$) and RFS ($p = 0.005$) when the low- and high-risk subgroups were in the early stages of the disease. However, there was only a significant difference in OS ($p = 0.04$) but not RFS ($p = 0.2$), between the low-risk and high-risk subgroups when both were in advanced stages.

Conclusion: PS, in combination with the TNM stage, provides additional precision in stratifying patients with significantly different OS and RFS prognoses. Further studies are warranted to assess the efficiency of PS and to explain the effects of the genetic and epigenetic aberrations observed in LUAD.

Keywords: lung adenocarcinoma, genome-wide, prognostic factor, survival prediction, TCGA

INTRODUCTION

Lung cancer is the leading cause of global cancer-related mortality, and ranks second in the estimated new cases of cancer in both sexes in the United States (Siegel et al., 2017). Lung adenocarcinoma (LUAD) is the most common histological type of lung cancer, accounting for approximately 50% (Shedden et al., 2008; Warth et al., 2012; Cancer Genome Atlas Research Network, 2014). Currently, the tumor, node, and metastasis (TNM) stage is the most accepted system for estimating the prognosis of patients with LUAD in clinical practice (Warth et al., 2012). However, prognoses of LUAD patients who share the same pathological stage vary considerably (Tsao et al., 2015; Zhang et al., 2015; Liang et al., 2017; Dalwadi et al., 2018). Therefore, a more accurate system is in demand to predict the outcomes of patients with LUAD that can add further valuable information to the TNM stage.

Aberrant genetic and epigenetic modifications of oncogenes and tumor suppressors contribute to the tumorigenesis and progression of LUAD (Khalil et al., 2018; Rowbotham et al., 2018; Tessema et al., 2018; Toyokawa et al., 2018; Wang et al., 2018). Genetic and epigenetic abnormalities have been associated with LUAD patient survival, especially the aberrant expression of cancer-related genes and DNA methylation at specific sites (Uruga et al., 2017; Zhang et al., 2017; Gonzalez-Vallinas et al., 2018; Wang et al., 2018). For instance, using genome-scale DNA methylation profiling, a study identified 164 hypermethylated genes and 57 hypomethylated genes involved in cell differentiation and the epithelial-to-mesenchymal transition in LUAD (Selamat et al., 2012). Notably, DNA methylation also accounts for the alteration of gene expression in LUAD (Zhang et al., 2017; Gao et al., 2018; He et al., 2018), and may thus indirectly affect the biological behaviors and processes of LUAD. Specifically, He et al. (2018) identified an association between aberrant CpG-methylation and the prognostic value of the corresponding gene expression based on 1095 LUAD samples, and identified 10 aberrantly methylated and dysregulated genes with independent prognostic value.

In recent years, a class of small non-coding RNA molecules, called microRNA (miRNA), has been increasingly investigated (Fu et al., 2017; Greenawalt et al., 2018; Othman and Nagoor, 2019). miRNAs can regulate the expression of protein-coding genes by base pairing with the target mRNAs, inducing the degradation or translational repression of the bound mRNAs (Ha and Kim, 2014; Hou et al., 2018). The prognostic significance of miRNAs has also been investigated and confirmed in many studies (Zhang et al., 2017; Gonzalez-Vallinas et al., 2018; Xu et al., 2018). For instance, mir-486 was shown to be a miRNA that is differentially expressed in LUAD and potentially interacts with ITGA11, a cancer-promoting gene (Zhang et al., 2017). Gonzalez-Vallinas et al also reported a significant association between mir-539, mir-323b, and mir-487a upregulation and worse disease-free survival in non-smoker patients with LUAD (Gonzalez-Vallinas et al., 2018).

So far, many studies have established panels of prognostic factors that predict the outcomes of patients with LUAD, based on multiple lines of evidence. However, most studies were conducted without integration of the network constituted by dysregulations at different levels. Because LUAD represents a set of heterogeneous diseases in which aberrations can exist at genome and epigenome levels, we performed a genome-wide analysis, which should provide more comprehensive insight into survival prediction. Using the data of 445 patients from The Cancer Genome Atlas (TCGA) database, we identified prognostic value of two miRNAs, two mRNAs and two methylation sites. A prognostic score (PS) was developed by integrating these factors to stratify LUAD patients with different lengths of survival into subgroups. From our data, combining PS and the TNM stage achieved greater accuracy in predicting the prognoses of patients with LUAD, indicating that PS is a promising system for personalized and precise medicine.

MATERIALS AND METHODS

Data Extraction and Prepossessing

The genome-wide data for 706 LUAD patients was downloaded from TCGA database[1], including the expression levels of 20530 mRNAs, 2228 miRNAs and 485577 DNA methylation sites, together with the outcomes of 630 patients. The exclusion criteria were listed as follows: (1) Patients whose genomic or epigenomic information was absent; (2) Genes lacking information on either their transcript (mRNA or miRNA) or DNA methylation levels in more than half the LUAD samples; (3) Patients whose survival information was unavailable. Ultimately, a total of 445 LUAD patients were included in the study, together with 16928 mRNAs, 453 miRNAs and 395963 DNA methylation sites.

Identification of Survival-Associated Transcripts and DNA Methylation Sites

A Cox regression model was used to evaluate the association of gene transcripts (mRNA or miRNA) and DNA methylation sites with lengths of overall survival (OS). A univariate Cox regression analysis was initially used, followed by the screening of included potential factors with a $p \leq 0.1$ for further analysis.

Afterward, considering the remaining large numbers of gene transcripts and methylated sites, we performed a Lasso-Cox analysis to screen and shrink the data. We then used multivariate Cox regression to further analyze the association between the gene transcripts or DNA methylation sites with OS, while adjusting for other clinicopathological factors.

Identifying and Screening Potential miRNA Targets

We retrieved the potential target genes of miRNAs that had already been shown to be significantly associated with OS from miRTarBase (the experimentally validated microRNA-target interactions database, release 7.0) (Hou et al., 2018). Lasso-Cox regression was then used to screen the mRNAs of

[1]https://xenabrowser.net

genes which were identified as targets of miRNAs with high-confidence ($p \leq 0.05$).

Calculation of Spearman's Correlation Coefficients

The direction of association among the transcripts and methylation sites were calculated with Spearman's correlation in the 445 LUAD tissues. If an mRNA tended to increase when miRNA or methylation increased, the Spearman's correlation coefficient was positive. If an mRNA tended to decrease when miRNA or methylation increased, the Spearman's correlation coefficient was negative. We set a threshold of 0 with which to assess the candidate miRNAs, mRNAs and methylation sites. Any pair with a correlation coefficient value < 0 was considered to be negatively correlated, whereas any pair with a correlation coefficient value > 0 as positively correlated.

External Validation of Identified Transcripts and Methylation Sites

We validated the prognostic value of our candidate transcripts in KM-Plot[2]. The impact of the candidate methylation sites on survival was confirmed in MethSurv[3] (Modhukur et al., 2018). We used Jetset to select the corresponding probe sets for the candidate mRNAs and miRNAs because a given gene may be detected by multiple probe sets, which may lead to inconsistent or even contradictory measurements (Li et al., 2011).

Construction and Validation of PS

To further assess the predictive ability of the significant factors identified, we constructed a PS as an integrated predictor. PS was calculated as a weighted sum of the expression levels of the transcripts and DNA methylation sites present in a given sample (Hou et al., 2018). For specimen i the calculation formula for PS was shown as follows:

$$\mathrm{PS} = \sum_{i=1}^{n} \beta i x i$$

The weight of each variable is represented by the Cox regression coefficient β, and the expression level is denoted by x. A greater value of PS indicates a worse prognosis.

We divided the patients into either the high-risk or low-risk group according to the median value of PS. Each group was subdivided into the early-stage (stages I–II) and advanced-stage (stages III–IV) subgroups based on the pathological stage. The Kaplan–Meier method and log-rank tests were used to assess the differences in OS and RFS between in the subgroups.

Statistical Analysis

All statistical analyses were performed with R version 3.4.4 (packages glmnet_2.0-16, survival_2.4-3; Institute for Statistics and Mathematics, Vienna, Austria). A two-tailed $p < 0.05$ was considered statistically significant.

[2]http://kmplot.com/analysis/index.php?p=background

[3]https://biit.cs.ut.ee/methsurv/

RESULTS

General Information on Patients With LUAD

The clinicopathological characteristics of the LUAD patients in our study are shown in **Table 1**. Of the 445 patients, 210 (47.2%)

TABLE 1 | Distributions of the demographic and clinical variables of 445 patients with lung adenocarcinoma patients.

Characteristic	Number (range)
Age at first diagnosis (median, range)	66 (39–88)
Gender	
Male	210
Female	235
Pathology (Histologic subtypes)	
Lepidic-predominant Adc	10
Acinar-predominant Adc	59
Papillary-predominant Adc	20
Micropapillary-predominant Adc	21
Solid-predominant Adc	34
Invasive mucinous Adc	7
Pathological stages	
Stage I	239
Stage II	109
Stage III	72
Stage IV	20
Unknown	5
Smoking history	
Lifelong non-smoker	65
Smoker	359
Current smoker	103
Current reformed smoker for < or = 15 years	145
Current reformed smoker for > 15 years	111
Number of packs smoked (N) per year	
N = 0	65
0 < N ≤ 20	112
20 < N ≤ 40	140
40 < N ≤ 60	113
60 < N ≤ 80	29
N > 80	38
Additional pharmaceutical therapy	
Yes	54
No	72
Additional radiation therapy	
Yes	65
No	63
Targeted molecular therapy	
Yes	138
No	247
Location	
RUL	159
RML	19
RLL	83
LUL	103
LLL	69

Adc, adenocarcinoma; RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe; LUL, left upper lobe; LLL, left lower lobe.

TABLE 2 | Risk factors in the patient cohort in our study.

	Univariate analysis			Multivariate analysis		
Characteristics	**HR**	**95%CI**	***p*-value**	**HR**	**95%CI**	***p*-value**
Age						
≤ 65						
> 66	1.233	0.900–1.688	0.192	1.153	0.652–2.041	0.624
Gender						
Female						
Male	1.016	0.744–1.386	0.922	1.259	0.712–2.226	0.428
Histologic subtypes						
Acinar predominant Adc						
Lepidic predominant Adc	1.067	0.356–3.197	0.907	0.782	0.233–2.625	0.69
Micropapillary predominant Adc	2.015	0.890–4.566	0.093	1.365	0.594–3.135	0.464
Papillary predominant Adc	1.48	0.685–3.197	0.318	0.899	0.397–2.036	0.799
Solid predominant Adc	2.66	1.375–5.148	0.004	1.752	0.855–3.593	0.126
Invasive mucinous Adc	0.81	0.186–3.527	0.779	0.709	0.156–3.223	0.656
Pathological stage						
Stage I						
Stage II	2.371	1.609–3.493	0	2.569	1.332–4.956	0.005
Stage III	3.419	2.280–5.128	0	2.16	1.060–4.400	0.034
Stage IV	3.863	2.146–6.951	0	2.139	0.715–6.403	0.174
T stage						
T1						
T2	1.42	0.978–2.060	0.065	NA	NA	NA
T3	2.411	1.341–4.333	0.003	NA	NA	NA
T4	2.727	1.396–5.330	0.003	NA	NA	NA
N stage						
N0						
N1	2.184	1.502–3.176	0	NA	NA	NA
N2	3.111	2.096–4.619	0	NA	NA	NA
M stage						
M0						
M1	2.155	1.208–3.845	0.009	NA	NA	NA

Adc, adenocarcinoma; HR, hazard ratio; CI, confidence interval; NA, not available.

were male and 235 (52.8%) were female. The median age was 66 years (ranging from 39 to 88). Patients with early-stage LUAD constituted the majority of our cohort. The primary tumor mainly was mainly located in the upper lobe on either side.

As shown in **Table 2**, we examined the association between each clinicopathological characteristic and OS. A univariate Cox regression analysis indicated that a higher TNM stage was significantly associated with poorer OS (**Table 2**). Meanwhile, a Kaplan-Meier survival analysis showed significantly different OS among the patients with different TNM stages (**Figure 1C**), but not among those who differed in age or sex (**Figures 1A,B**). Interestingly, a trend toward different OS among patients with different histologic subtypes was observed, but the *p* value was only marginally significant (**Figure 1D**). In the multivariate regression analysis, only a higher TNM stage remained a significant risk factor for OS (**Table 2**).

Identification of Transcripts and DNA Methylation Sites as Prognostic Factors

From 16928 mRNAs, 453 miRNA and 395963 DNA methylation sites, a total of 26 miRNAs, 15 mRNAs and 11 DNA methylation sites were initially identified as factors associated with OS using univariate Cox and Lasso-Cox analyses (**Supplementary Table S1**). Next, a list of 2882 genes was then retrieved from the miRTarBase database, of which 21 were identified as high-confidence ($p \leq 0.05$) targets of the 26 survival-associated miRNAs. A Lasso-Cox analysis was used to select the mRNAs of the 21 genes that interacted with the corresponding survival-related miRNAs. After a multivariate Cox regression analysis of the 73 potential prognostic factors, the overexpression of two miRNAs (MIMAT0002890 and MIMAT0000426) and the hypermethylation of two sites (cg12141052 and cg16404170) were confirmed as significant predictors of worse prognosis (**Table 3**), and the higher expression level of two mRNAs (CDADC1, FAHD2B) was significantly associated with a better prognosis (**Table 3**). Therefore, the final list of candidate prognostic factors for LUAD contained 6 biomarkers, including two miRNAs, two mRNAs and two methylation sites.

The Spearman's rank correlation coefficients for these candidate transcripts and methylation levels were then calculated for the LUAD cohort of 445 patients (**Table 4** and **Supplementary Figure S1**).

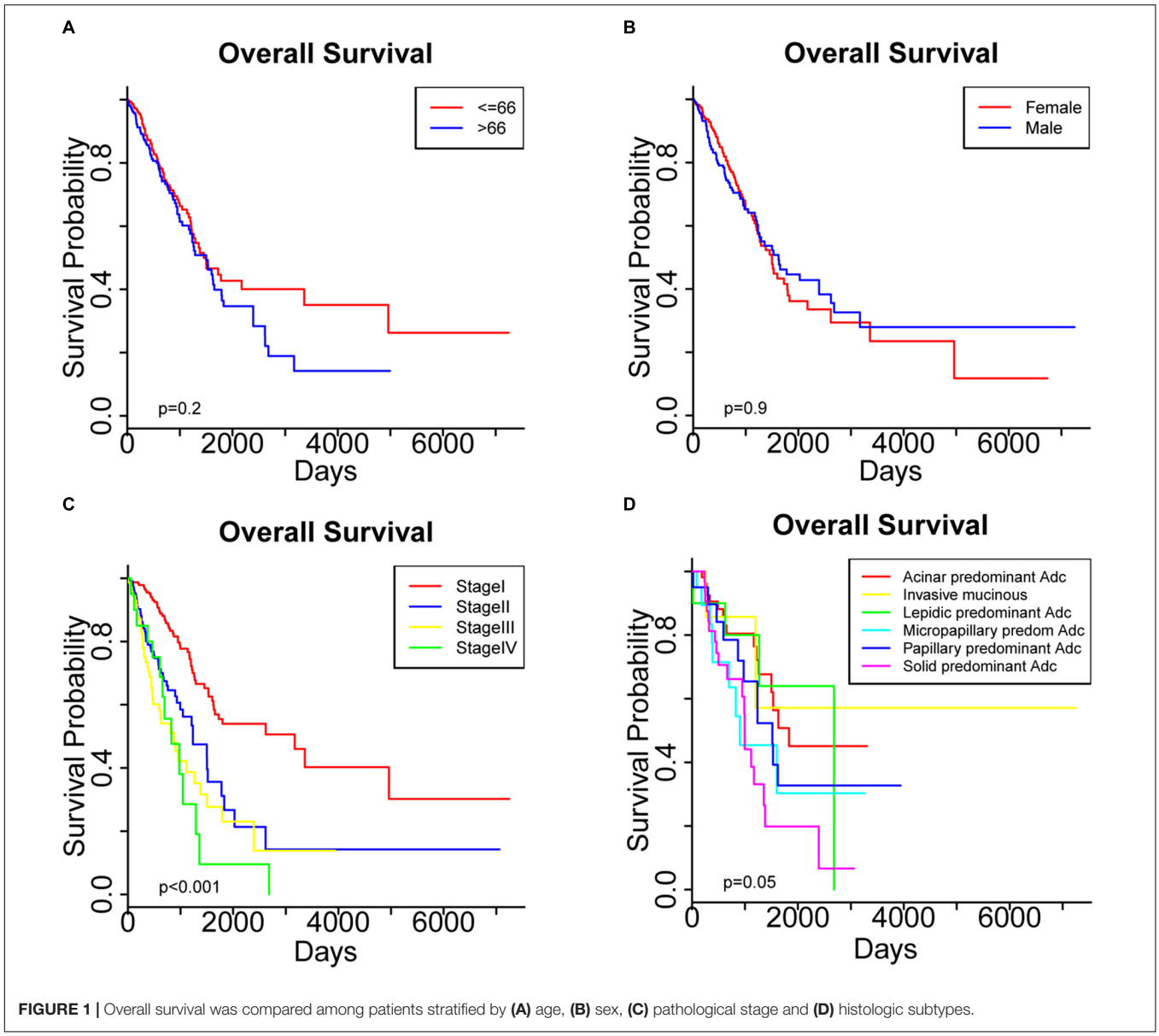


FIGURE 1 | Overall survival was compared among patients stratified by **(A)** age, **(B)** sex, **(C)** pathological stage and **(D)** histologic subtypes.

TABLE 3 | Genome-wide prognostic factors identified in our study.

Molecular	Name	Coefficient	HR	95% CI	SE	z-value	*p*-value
miRNA	MIMAT0002890	0.358	1.431	1.191, 1.718	0.094	3.829	0
	MIMAT0000426	0.246	1.278	0.999, 1.635	0.126	1.954	0.051
mRNA	CDADC1	−0.578	0.561	0.387, 0.814	0.19	−3.047	0.002
	FAHD2B	−0.276	0.759	0.643, 0.895	0.084	−3.281	0.001
Methylation site	cg12141052	4.045	57.117	8.653, 377.015	0.963	4.201	0
	cg16404170	3.495	32.959	6.593, 164.761	0.821	4.257	0

HR, hazard ratio; CI, confidence interval; SE, standard error; z value, Wald z-statistic value.

External Validation of Candidate Transcripts and Methylation Sites

As shown in **Supplementary Figures S2–S4**, a univariate Cox proportional hazards regression analysis showed that the six candidate factors identified from either genome or epigenome of LUAD were significantly associated with the survival of patient cohorts in other databases. Moreover, the relationships between their expression levels and the survival rate of LUAD patients were consistent with our findings.

TABLE 4 | Spearman's correlation coefficients among the prognostic factors identified in the study.

Factors	MIMAT0002890	MIMAT0000426	CDADC1	FAHD2B	cg12141052	cg16404170
MIMAT0002890	1.000	0.245	−0.004	−0.028	−0.119	0.136
MIMAT0000426	0.245	1.000	−0.042	0.000	−0.029	0.145
CDADC1	−0.004	−0.042	1.000	0.174	−0.160	−0.165
FAHD2B	−0.028	0.000	0.174	1.000	−0.053	−0.143
cg12141052	−0.119	−0.029	−0.160	−0.053	1.000	0.021
cg16404170	0.136	0.145	−0.165	−0.143	0.021	1.000

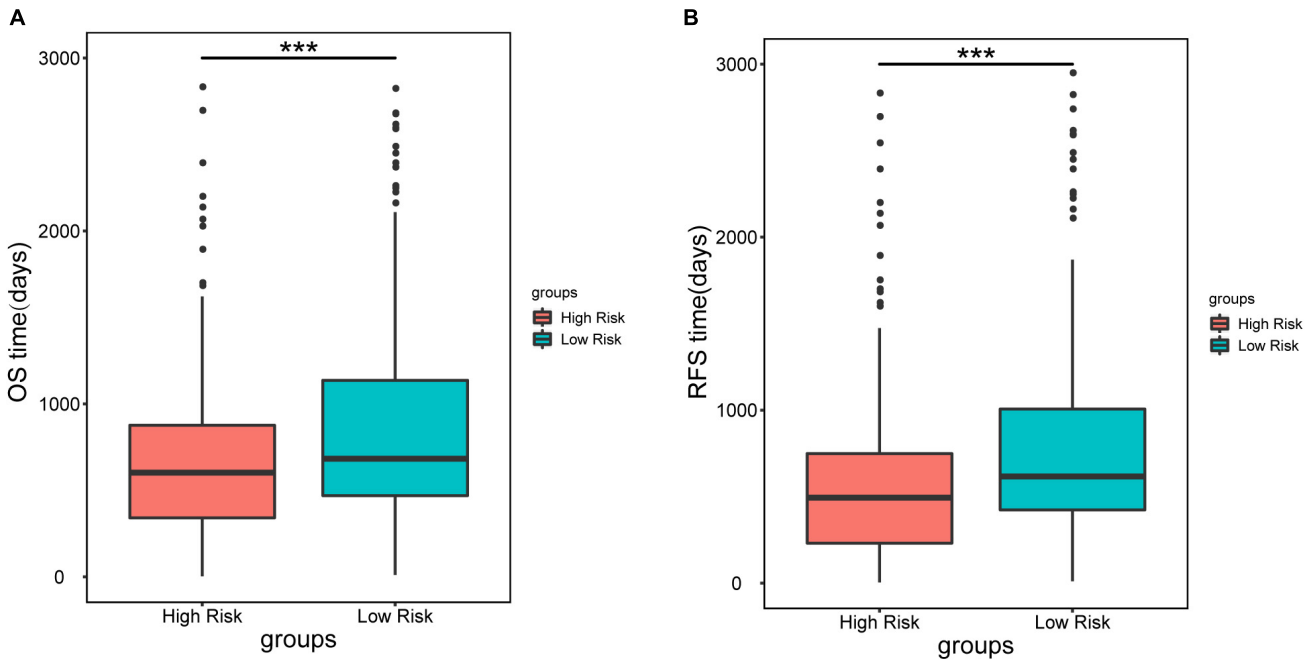


FIGURE 2 | Median overall survival **(A)** and recurrence-free survival **(B)** in patients in the high-risk (PS > 1.88) and low-risk groups (PS < 1.88). Patients in the high-risk group showed a significantly shorter survival than those in the low-risk group (***$p < 0.001$).

Validation of the Integrated Prognostic Factors

To further assess the predictive capacity of all the candidate prognostic factors, PS was established as an integrated prognostic predictor. To verify the efficiency of PS, the 445 LUAD patients were divided into two groups stratified by the median PS. The high-risk group (PS > 1.88) included 223 patients and the low-risk group (PS < 1.88) included 222 patients. As shown in **Figure 2**, the patients in the low- and high-risk groups displayed significantly different median OS (1070.8 vs. 753.9 days, $p < 0.001$) and RFS (900.8 vs. 668.2, $p = 0.005$). As shown in **Figures 3A**, **4A**, the Kaplan-Meier curves and log-rank tests indicated significant differences in the OS [hazard ratio (HR): 2.861, 95% confidence interval (CI): 2.052–3.988, $p < 0.001$] and RFS (HR: 1.77, 95% CI: 1.255–2.497, $p = 0.001$) between two groups.

Significantly different OS and RFS were also observed among the subgroups in further analyses (**Figures 3B**, **4B**). On the one hand, PS remained efficient in stratifying the patients into different OS (HR: 3.177, 95%CI: 2.110–4.783, $p < 0.001$) and RFS (HR: 1.752, 95% CI: 1.184–2.595, $p = 0.005$) when the low-risk and high-risk subgroups were in the early stages of the disease (**Figures 3C**, **4C**). However, there was only a significant difference in OS (HR: 1.806, 95% CI: 1.019–3.200, $p = 0.04$) but not RFS (HR: 1.594, 95% CI: 0.763–3.333, $p = 0.2$) between the two subgroups when both were in the advanced stages of the disease (**Figures 3D**, **4D**). On the other hand, the pathological stage could distinguish significantly different OS (low-risk group: HR: 3.341, 95% CI: 1.888–5.912, $p < 0.001$; high-risk group: HR: 1.955, 95% CI: 1.305–2.929, $p < 0.001$) but not RFS (low-risk group: HR: 1.472, 95% CI: 0.878–2.467, $p = 0.1$; high-risk group: HR: 1.604, 95% CI: 0.829–3.104, $p = 0.2$) in the low-risk and high-risk groups. Thus, PS was proved to be a useful prognostic indicator that can supplement additional information to the TNM stage, especially for LUAD patients in the early stages of the disease. Our study suggests that the combination of the TNM stage and PS increases the accuracy in predicting the outcomes of patients with LUAD.

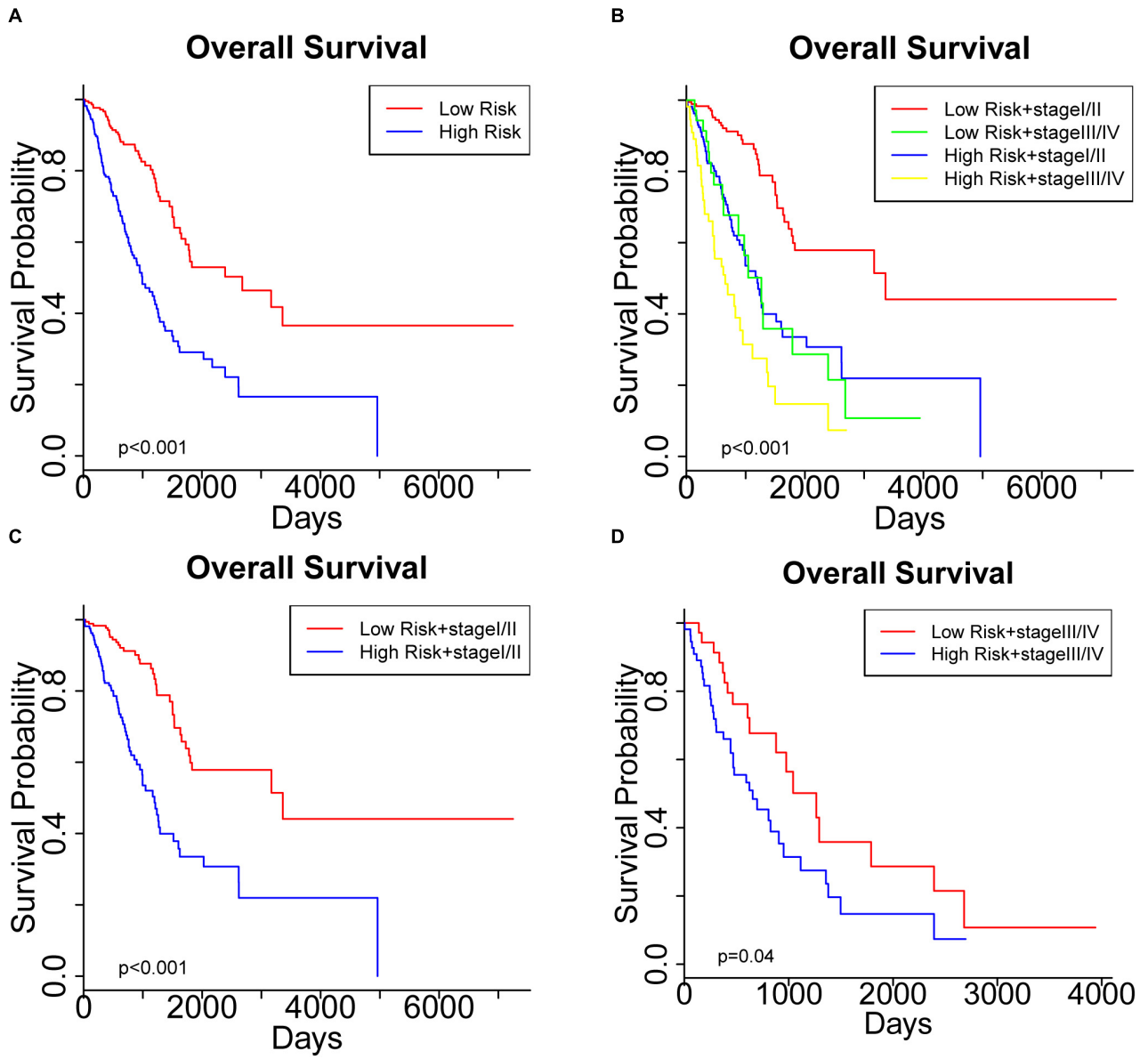


FIGURE 3 | Kaplan Meier curve showing the overall survival (OS) of the patient cohort grouped by **(A)** prognostic score (PS), and **(B)** PS plus pathological stage. OS of patients stratified by PS in subgroups with **(C)** early-stage tumors and **(D)** advanced- stage tumors.

DISCUSSION

The past decade has witnessed rapid progress in next-generation sequencing and its increasing application in preclinical practice. In recent years, several studies have attempted to associate the transcriptome or epigenome with the clinical outcomes of patients with LUAD (Selamat et al., 2012; Zhang et al., 2017; Gao et al., 2018; He et al., 2018). Zhang et al. analyzed and validated the expression profiles and prognostic values of the mRNAs of five differentially expressed genes associated with DNA methylation in LUAD (Zhang et al., 2017), increasing the likelihood that altered signature genes will become useful biomarkers. Using a TCGA dataset, He et al. (2018) disentangled the relationships between aberrant CpG-methylation and gene expression to identify 10 aberrantly methylated and dysregulated genes. However, their study only focused on the ability of individual genes to predict OS. Another TCGA-based study examined the feasibility of integrating prognosis-related methylation-driven genes into a risk model to predict the OS of patients with LUAD, which also involved a joint survival analysis based on methylation sites and gene expression (Gao et al., 2018). Nevertheless, it remained unclear whether a risk model could improve the accuracy of the TNM stage for survival estimation. Furthermore, no information was given on the predictive value of their method in distinguishing RFS in LUAD patients. None of these studies included the histologic

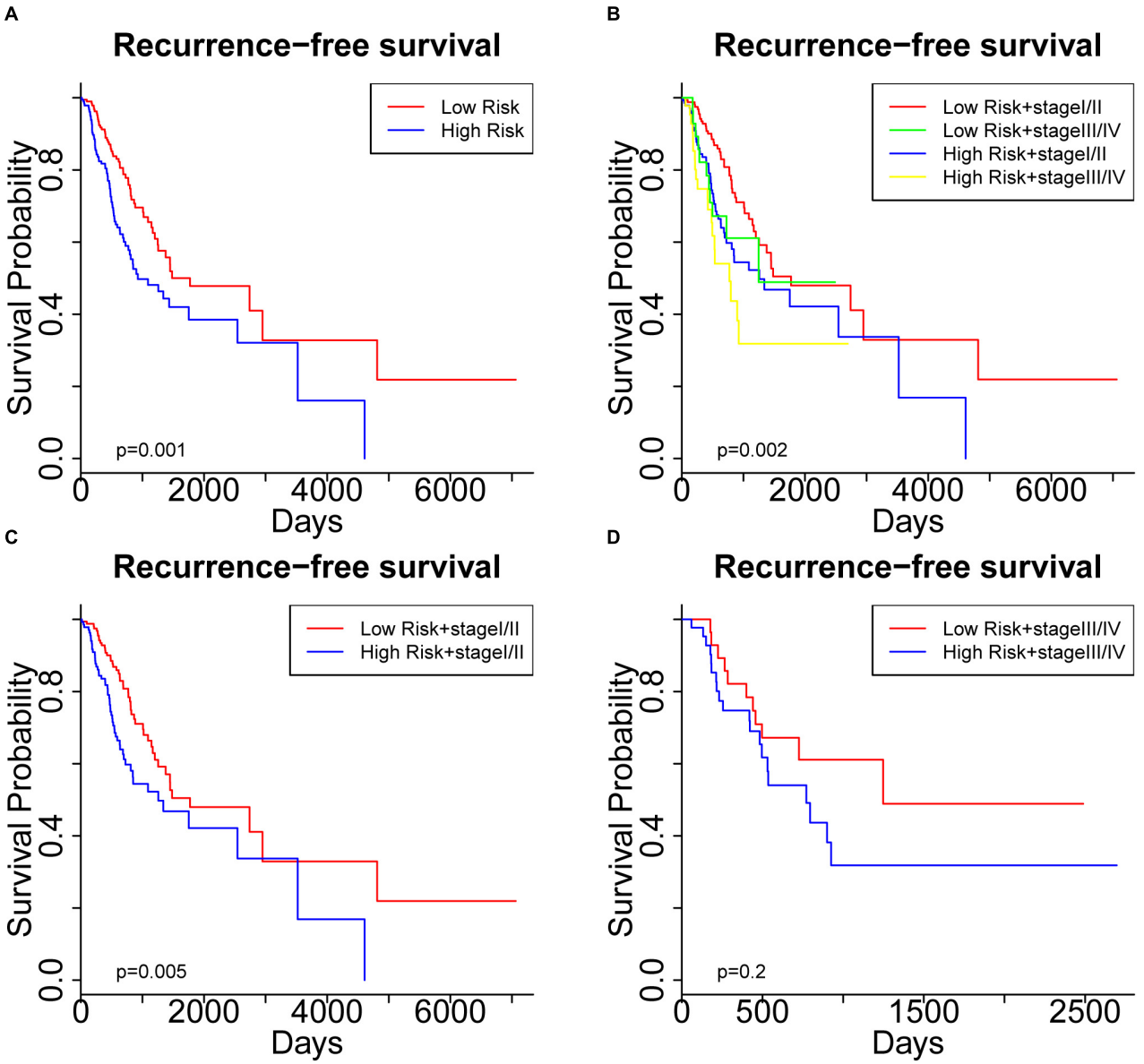


FIGURE 4 | Kaplan Meier curve showing recurrence-free survival (RFS) of the patient cohort grouped by **(A)** prognostic score (PS), and **(B)** PS plus pathological stage. RFS of the patients stratified by PS in subgroups with **(C)** early-stage tumors and **(D)** advanced-stage tumors.

subtypes proposed by the International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society (IASLC/ATS/ERS) (Travis et al., 2011; Warth et al., 2012; Hung et al., 2014) as an independent prognostic factor.

To the best of our knowledge, this is the first study to integrate genetic and epigenetic modifications for survival prediction in LUAD patients using TCGA samples. With a comprehensive analysis and screening of mRNA expression, miRNAs and DNA methylation sites based on samples from 445 patients, we identified a set of prognostic factors from both the transcriptome and epigenome. Notably, we included the histologic subtypes and the TNM stages in our initial survival analysis. In this way, we developed a novel subgrouping system that integrates PS and the TNM stage to predict the survival of patients with LUAD.

We started by identifying the clinicopathological characteristics associated with the OS of patients with LUAD. Both Cox regression and Kaplan-Meier survival analyses confirmed the significant prognostic impact of the TNM stage (**Table 2** and **Figure 1**). Further screening of genetic and epigenetic aberrations identified a collection of 26 miRNAs, 15 mRNAs and 11 DNA methylation sites whose expression or methylation levels were significantly associated with OS (**Supplementary Table S1**). Since miRNAs exert their function by regulating the expression of their target mRNAs, we retrieved the potential targets of these 26 miRNAs and performed a

LASSO-Cox analysis to select 21 mRNAs as high-confidence miRNA targets. This provided clues to the potential molecular interactions by which these miRNAs affect the clinical outcomes. Considering the interactions between these candidate prognostic factors into consideration, we performed a multivariate Cox regression to finally identify a list of six survival-related biomarkers. From our data, the expression levels of two mRNAs (CDADC1 and FAHD2B), two miRNAs (MIMAT0002890 and MIMAT0000426) and methylation of two sites (cg12141052 and cg16404170) were strongly associated with the clinical outcomes. PS was then computed as a predictor that integrated these candidate biomarkers and stratified the patients into low-risk (PS $<$ 1.88) and high-risk groups (PS $>$ 1.88). The efficiency of PS was confirmed by our success in distinguishing the OS and RFS of LUAD patients (**Figures 3**, **4**). A subgroup analysis further demonstrated that a more precise prediction of survival could be achieved for patients with LUAD by combining PS with the TNM stage, which should allow more timely therapeutic interventions.

To be noted, Targetscan[4] was preferentially considered for the validation of our candidate miRNAs, however, the small number of miRNA targets shared between miRTarBase and Targetscan limited its use (**Supplementary Figure S5**).

Ten survival-associated genes, whose aberrant expression was affected by methylation, have been identified previously by He et al. (2018) from the TCGA data portal[5]. Therefore we attempted to include the mRNAs of these 10 genes in our transcripts for further screening. However, as shown in **Supplementary Table S2** and **Supplementary Figures S6–S8**, integrating the mRNA of BLK which was identified in the Cox regression analysis into PS did not improve its predictive ability.

In terms of the disproportionate number of non-smokers in the selected patient cohort, secondary analyses were therefore performed to assess the potential value of PS in predicting the survival of the non-smokers and smokers in our cohort. As shown in **Supplementary Figure S9A**, Kaplan-Meier curves and the log-rank test indicated a significant difference in OS between two groups (HR: 2.785, 95% CI: 1.071–7.24, $p = 0.03$). Significantly different OS was also observed among subgroups stratified by PS plus the TNM stage (**Supplementary Figure S9B**). However, the performance of PS was not satisfactory for the non-smokers (**Supplementary Figures S9C,D**), especially in stratifying patients with advanced-stage LUAD into subgroups with different OS, which might be attributed to the limited number of non-smokers ($n = 65$). On the contrary, PS remained consistently efficient in stratifying OS in the smokers ($n = 359$) (**Supplementary Figure S10**).

There were several limitations to our study. For instance, risk factors such as packages of cigarettes and adjuvant therapy were not included in our analysis because of their interpatient heterogeneity. Moreover, the histologic subtypes was unsatisfactory in distinguishing prognoses in the multivariate analysis which could be explained by the missing histologic information for almost half the patients. It is noteworthy that the failure of PS to distinguish RFS in the advanced-stage subgroups ($p = 0.2$) should be possibly attributed to the limited number of LUAD patients with advanced-stage disease. Last but not least, the clinical utility of PS identified here may be limited in patients with small-sized lesions because of the difficulty in extracting sufficient RNA and protein. More studies are warranted to assess the roles of these candidate prognostic factors in LUAD.

[4]http://www.targetscan.org/vert_72/

[5]https://cancergenome.nih.gov

CONCLUSION

In conclusion, using a TCGA dataset of 445 LUAD patients, we identified six prognostic factors (two mRNAs, two miRNAs and two DNA methylation sites) for LUAD from the genome and epigenome, and developed PS from them. Combining the TNM stage and PS provided additional precision in stratifying patients into significantly different OS and RFS subgroups. Further studies are warranted to assess the efficiency of PS and to explain the effects of these observed genetic and epigenetic aberrations in LUAD.

AUTHOR CONTRIBUTIONS

DC, YS, SL, CC, and YC designed the study. DC, YS, FZ, and EZ performed the data analysis. DC, YS, FZ, XW, and XZ wrote the manuscript. GJ, SL, CC, and YC reviewed and edited the manuscript. YC and FZ afforded to conduct our study.

FUNDING

The work was supported from the National Natural Science Foundation pre Research Fund Project (SDFEYGJ1704), Jiangsu Provincial Commission of Health and Family Planning (Grant H201521), the Natural Science Foundation of Jiangsu Province (Grant BK20161224), Suzhou Key Discipline for Medicine (SZXK201803), and the Science and Technology Research Foundation of Suzhou Municipality (SYS2018063).

ACKNOWLEDGMENTS

We thank International Science Editing (http://www.internationalscienceediting.com) for editing this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00493/full#supplementary-material

FIGURE S1 | Plot of Spearman's rank correlation coefficients among candidate transcripts and methylation levels in LUAD ($n = 445$).

FIGURE S2 | Differential expression and prognostic impact of **(A)** FADH2B and **(B)** CDADC1 (two candidate mRNAs) in LUAD patients. Kaplan-Meier curves of 720 LUAD patients, who were separated into high-expression and low-expression groups, using as cutoffs the best-performing thresholds of the different genes. All values were significant ($p < 0.05$).

FIGURE S3 | Differential expression and prognostic impact of **(A)** MIMAT0002890 and **(B)** MIMAT0000426 (two candidate miRNAs) in LUAD patients. Kaplan-Meier curves of 720 LUAD patients, who were separated into high-expression and low-expression groups, using as cutoffs the best-performing thresholds of the different miRNAs. All values were significant ($p < 0.05$).

FIGURE S4 | Differential expression and prognostic impact of **(A)** cg12141052 and **(B)** 16404170 (two candidate methylation sites) in LUAD patients. Kaplan-Meier curves of 720 LUAD patients, who were separated into high-expression and low-expression groups, using as cutoffs the best-performing thresholds. All values were significant ($p < 0.05$).

FIGURE S5 | Venn diagrams representing 1161 miRNA targets that overlapped between miRTarBase and Targetscan.

FIGURE S6 | Median overall survival **(A)** and recurrence-free survival **(B)** of patients in the high-risk (PS > 2.78) and low-risk groups (PS < 2.78). Patients in the high-risk group showed a significantly shorter survival than those in the low-risk group (***$p < 0.001$).

FIGURE S7 | Kaplan Meier curve showing the overall survival (OS) of the patient cohort grouped by **(A)** recombinant prognostic score (PS), and **(B)** recombinant PS plus pathological stage. OS of the patients stratified by PS in subgroups with **(C)** early-stage disease and **(D)** advanced-stage disease.

FIGURE S8 | Kaplan Meier curve showing recurrence-free survival (RFS) of the patient cohort grouped by **(A)** recombinant prognostic score (PS), and **(B)** recombinant PS plus pathological stage. RFS of the patients stratified by PS in subgroups with **(C)** early-stage disease and **(D)** advanced-stage disease.

FIGURE S9 | Kaplan Meier curve showing the overall survival (OS) of the non-smokers in our cohort grouped by **(A)** prognostic score (PS), and **(B)** PS plus pathological stage. OS of the patients stratified by PS in subgroups with **(C)** early-stage disease and **(D)** advanced-stage disease.

FIGURE S10 | Kaplan Meier curve showing the overall survival (OS) of the smokers in our cohort grouped by **(A)** prognostic score (PS), and **(B)** PS plus pathological stage. OS of the patients stratified by PS in subgroups with **(C)** early-stage disease and **(D)** advanced-stage disease.

TABLE S1 | Transcripts and DNA methylation sites whose expression levels showed significant association with overall survival.

TABLE S2 | Integrated genome-wide prognostic factors in our study.

REFERENCES

Cancer Genome Atlas Research Network (2014). Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511, 543–550. doi: 10.1038/nature13385

Dalwadi, S. M., Szeja, S. S., Bernicker, E. H., Butler, E. B., Teh, B. S., and Farach, A. M. (2018). Practice patterns and outcomes in elderly stage i non-small-cell lung cancer: A 2004 to 2012 SEER analysis. *Clin. Lung Cancer* 19, e269–e276. doi: 10.1016/j.cllc.2017.11.004

Fu, Q., Song, X., Liu, Z., Deng, X., Luo, R., Ge, C., et al. (2017). miRomics and proteomics reveal a miR-296-3p/PRKCA/FAK/Ras/c-myc feedback loop modulated by hdgf/ddx5/beta-catenin complex in lung adenocarcinoma. *Clin. Cancer Res.* 23, 6336–6350. doi: 10.1158/1078-0432.ccr-16-2813

Gao, C., Zhuang, J., Li, H., Liu, C., Zhou, C., Liu, L., et al. (2018). Exploration of methylation-driven genes for monitoring and prognosis of patients with lung adenocarcinoma. *Cancer Cell Int.* 18:194. doi: 10.1186/s12935-018-0691-z

Gonzalez-Vallinas, M., Rodriguez-Paredes, M., Albrecht, M., Sticht, C., Stichel, D., Gutekunst, J., et al. (2018). Epigenetically regulated chromosome 14q32 mirna cluster induces metastasis and predicts poor prognosis in lung adenocarcinoma patients. *Mol. Cancer Res.* 16, 390–402. doi: 10.1158/1541-7786.mcr-17-0334

Greenawalt, E. J., Edmonds, M. D., Jain, N., Adams, C. M., Mitra, R., and Eischen, C. M. (2018). Targeting of SGK1 by miR-576-3p inhibits lung adenocarcinoma migration and invasion. *Mol. Cancer Res.* 17, 289–298. doi: 10.1158/1541-7786.mcr-18-0364

Ha, M., and Kim, V. N. (2014). Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* 15, 509–524. doi: 10.1038/nrm3838

He, W., Ju, D., Jie, Z., Zhang, A., Xing, X., and Yang, Q. (2018). Aberrant CpG-methylation affects genes expression predicting survival in lung adenocarcinoma. *Cancer Med.* 7, 5716–5726. doi: 10.1002/cam4.1834

Hou, X., He, X., Wang, K., Hou, N., Fu, J., Jia, G., et al. (2018). Genome-wide network-based analysis of colorectal cancer identifies novel prognostic factors and an integrative prognostic index. *Cell Physiol. Biochem.* 49, 1703–1716. doi: 10.1159/000493614

Hung, J. J., Yeh, Y. C., Jeng, W. J., Wu, K. J., Huang, B. S., Wu, Y. C., et al. (2014). Predictive value of the international association for the study of lung cancer/American Thoracic Society/European Respiratory Society classification of lung adenocarcinoma in tumor recurrence and patient survival. *J. Clin. Oncol.* 32, 2357–2364. doi: 10.1200/jco.2013.50.1049

Khalil, A., Dekmak, B., Boulos, F., Kantrowitz, J., Spira, A., Fujimoto, J., et al. (2018). Transcriptomic alterations in lung adenocarcinoma unveil new mechanisms targeted by the TBX2 subfamily of tumor suppressor genes. *Front. Oncol.* 8:482. doi: 10.3389/fonc.2018.00482

Li, Q., Birkbak, N. J., Gyorffy, B., Szallasi, Z., and Eklund, A. C. (2011). Jetset: selecting the optimal microarray probe set to represent a gene. *BMC Bioinform.* 12:474. doi: 10.1186/1471-2105-12-474

Liang, W., He, J., Shen, Y., Shen, J., He, Q., Zhang, J., et al. (2017). Impact of examined lymph node count on precise staging and long-term survival of resected non-small-cell lung cancer: a population study of the US SEER database and a chinese multi-institutional registry. *J. Clin. Oncol.* 35, 1162–1170. doi: 10.1200/jco.2016.67.5140

Modhukur, V., Iljasenko, T., Metsalu, T., Lokk, K., Laisk-Podar, T., and Vilo, J. (2018). MethSurv: a web tool to perform multivariable survival analysis using DNA methylation data. *Epigenomics* 10, 277–288. doi: 10.2217/epi-2017-0118

Othman, N., and Nagoor, N. H. (2019). Overexpression of miR3615p plays an oncogenic role in human lung adenocarcinoma through the regulation of SMAD2. *Int. J. Oncol.* 54, 306–314. doi: 10.3892/ijo.2018.4602

Rowbotham, S. P., Li, F., Dost, A. F. M., Louie, S. M., Marsh, B. P., Pessina, P., et al. (2018). H3K9 methyltransferases and demethylases control lung tumor-propagating cells and lung cancer progression. *Nat. Commun.* 9:4559. doi: 10.1038/s41467-018-07077-1

Selamat, S. A., Chung, B. S., Girard, L., Zhang, W., Zhang, Y., Campan, M., et al. (2012). Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. *Genome Res.* 22, 1197–1211. doi: 10.1101/gr.132662.111

Shedden, K., Taylor, J. M., Enkemann, S. A., Tsao, M. S., Yeatman, T. J., Gerald, W. L., et al. (2008). Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat. Med.* 14, 822–827. doi: 10.1038/nm.1790

Siegel, R. L., Miller, K. D., and Jemal, A. (2017). Cancer statistics, 2017. *CA Cancer J. Clin.* 67, 7–30. doi: 10.3322/caac.21387

Tessema, M., Rossi, M. R., Picchi, M. A., Yingling, C. M., Lin, Y., Ramalingam, S. S., et al. (2018). Common cancer-driver mutations and their association with abnormally methylated genes in lung adenocarcinoma from never-smokers. *Lung Cancer* 123, 99–106. doi: 10.1016/j.lungcan.2018.07.011

Toyokawa, G., Takada, K., Tagawa, T., Hamamoto, R., Yamada, Y., Shimokawa, M., et al. (2018). A positive correlation between the EZH2 and PD-L1 expression in resected lung adenocarcinomas. *Ann. Thorac. Surg.* 107, 393–400. doi: 10.1016/j.athoracsur.2018.08.056

Travis, W. D., Brambilla, E., Noguchi, M., Nicholson, A. G., Geisinger, K. R., Yatabe, Y., et al. (2011). International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. *J. Thorac. Oncol.* 6, 244–285. doi: 10.1097/JTO.0b013e318206a221

Tsao, M. S., Marguet, S., Le Teuff, G., Lantuejoul, S., Shepherd, F. A., Seymour, L., et al. (2015). Subtype classification of lung adenocarcinoma predicts benefit from adjuvant chemotherapy in patients undergoing complete resection. *J. Clin. Oncol.* 33, 3439–3446. doi: 10.1200/jco.2014.58.8335

Uruga, H., Bozkurtlar, E., Huynh, T. G., Muzikansky, A., Goto, Y., Gomez-Caraballo, M., et al. (2017). Programmed cell death ligand (PD-L1) expression

in stage II and III lung adenocarcinomas and nodal metastases. *J. Thorac. Oncol.* 12, 458–466. doi: 10.1016/j.jtho.2016.10.015

Wang, L., Zhao, H., Xu, Y., Li, J., Deng, C., Deng, Y., et al. (2018). Systematic identification of lincRNA-based prognostic biomarkers by integrating lincRNA expression and copy number variation in lung adenocarcinoma. *Int. J. Cancer* 144, 1723–1734. doi: 10.1002/ijc.31865

Warth, A., Muley, T., Meister, M., Stenzinger, A., Thomas, M., Schirmacher, P., et al. (2012). The novel histologic international association for the study of lung cancer/American Thoracic Society/European Respiratory Society classification system of lung adenocarcinoma is a stage-independent predictor of survival. *J. Clin. Oncol.* 30, 1438–1446. doi: 10.1200/jco.2011.37.2185

Xu, X., Yu, S., Sun, W., Qin, X., Chen, Y., Zhou, L., et al. (2018). MiRNA signature predicts the response of patients with advanced lung adenocarcinoma to platinum-based treatment. *J. Cancer Res. Clin. Oncol.* 144, 431–438. doi: 10.1007/s00432-017-2562-8

Zhang, J., Gold, K. A., Lin, H. Y., Swisher, S. G., Xing, Y., Lee, J. J., et al. (2015). Relationship between tumor size and survival in non-small-cell lung cancer (NSCLC): an analysis of the surveillance, epidemiology, and end results (SEER) registry. *J. Thorac. Oncol.* 10, 682–690. doi: 10.1097/jto.0000000000000456

Zhang, Y., Zhao, W., and Zhang, J. (2017). Comprehensive epigenetic analysis of the signature genes in lung adenocarcinoma. *Epigenomics* 9, 1161–1173. doi: 10.2217/epi-2017-0023

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Chen, Song, Zhang, Wang, Zhu, Zhang, Jiang, Li, Chen and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 29 May 2019
doi: 10.3389/fgene.2019.00508

F8 Inversions at Xq28 Causing Hemophilia A Are Associated With Specific Methylation Changes: Implication for Molecular Epigenetic Diagnosis

Muhammad Ahmer Jamil[1†], Amit Sharma[1†], Nicole Nuesgen[1], Behnaz Pezeshkpoor[1], André Heimbach[2], Anne Pavlova[1], Johannes Oldenburg[1] and Osman El-Maarri[1]*

[1] Institute of Experimental Hematology and Transfusion Medicine, University of Bonn, Bonn, Germany, [2] Institute of Human Genetics, School of Medicine, University of Bonn – University Hospital Bonn, Bonn, Germany

OPEN ACCESS

Edited by:
Dongyi He,
Shanghai Guanghua Rheumatology Hospital, China

Reviewed by:
Apiwat Mutirangura,
Chulalongkorn University, Thailand
Hehuang Xie,
Virginia Tech, United States

***Correspondence:**
Osman El-Maarri
osman.elmaarri@ukbonn.de

† These authors have contributed equally to this work as first authors

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *09 February 2019*
Accepted: *09 May 2019*
Published: *29 May 2019*

Citation:
Jamil MA, Sharma A, Nuesgen N, Pezeshkpoor B, Heimbach A, Pavlova A, Oldenburg J and El-Maarri O (2019) F8 Inversions at Xq28 Causing Hemophilia A Are Associated With Specific Methylation Changes: Implication for Molecular Epigenetic Diagnosis. Front. Genet. 10:508. doi: 10.3389/fgene.2019.00508

Diverse DNA structural variations (SVs) in human cancers and several other diseases are well documented. For genomic inversions in particular, the disease causing mechanism may not be clear, especially if the inversion border does not cross a coding sequence. Understanding about the molecular processes of these inverted genomic sequences, in a mainly epigenetic context, may provide additional information regarding sequence-specific regulation of gene expression in human diseases. Herein, we study one such inversion hotspot at Xq28, which leads to the disruption of F8 gene and results in hemophilia A phenotype. To determine the epigenetic consequence of this rearrangement, we evaluated DNA methylation levels of 12 CpG rich regions with the coverage of 550 kb by using bisulfite-pyrosequencing and next-generation sequencing (NGS)-based bisulfite re-sequencing enrichment assay. Our results show that this inversion prone area harbors widespread methylation changes at the studied regions. However, only 5/12 regions showed significant methylation changes, specifically in case of intron 1 inversion (two regions), intron 22 inversion (two regions) and one common region in both inversions. Interestingly, these aberrant methylated regions were found to be overlapping with the inversion proximities. In addition, two CpG sites reached 100% sensitivity and specificity to discriminate wild type from intron 22 and intron 1 inversion samples. While we found age to be an influencing factor on methylation levels at some regions, covariate analysis still confirms the differential methylation induced by inversion, regardless of age. The hemophilia A methylation inversion "HAMI" assay provides an advantage over conventional PCR-based methods, which may not detect novel rare genomic rearrangements. Taken together, we showed that genomic inversions in the F8 (Xq28) region are associated with detectable changes in methylation levels and can be used as an epigenetic diagnostic marker.

Keywords: structural rearrangement, epigenetic, DNA methylation, inversion, hemophilia, molecular diagnosis

INTRODUCTION

Implications of human DNA sequence variations have received considerable attention in recent years and structural variants (SVs) are considered an important contributor among them. SVs, particularly inversions, can vary in size from few nucleotides to large-scale chromosomal rearrangements. The inversions can have functional consequences by truncating a given gene (or genes) or by rearranging the regulatory element in the local proximity, both having a disproportionate impact on gene expression and transcriptional variability (Puig et al., 2015).

Hemophilia A (OMIM #306700), an inherited bleeding disorder, harbors two such rearrangements at chromosome X (Xq28) involving the coagulation factor VIII (F8) gene. *F8* (∼186 kb; 26 exons) is located at the telomeric end of the X- chromosome and contains regions with high GC content, which makes it more susceptible toward the methylated cytosine deamination mutations [2,537 mutations, reported by CDC Hemophilia A Mutation Project (CHAMP) (Payne et al., 2013)]. In addition, two hotspot inversions (known as intron 1 and intron 22 inversions) are reported accounting for 40–50% of patients with severe hemophilia A (Andrikovics et al., 2003; Oldenburg and El-Maarri, 2006; Zimmermann et al., 2011). These hotspot recurring inversions are caused by intra-chromosomal homologous recombination between identical inverted repeats: two long repeats located within the *F8* locus: the Int22h-1 in intron 22 and the Int1h-1 in intron 1 (Lakich et al., 1993; Naylor et al., 1995). The former is 9.1 kb in length and has two additional homologs (Int22h-2 and 3) at about 500–580 kb distance toward the telomere, while the latter is about 1 kb and has one homolog located 141 kb toward the telomere (**Figure 1**) (UCSC genome browser). Both repeats are prone to intra-chromosomal homologous recombination leading to an inversion of the intervening sequence, thus leaving the *F8* split into two parts of opposite transcriptional direction (Bagnall et al., 2002, 2006). The clinical result of such inversions is a severe hemophilia A (HA) phenotype with no functional FVIII protein. Inversion events leading to human diseases are not limited to *F8* gene, other genes, such as *IDS* gene (Hunter syndrome), *MSH2* gene (Lynch syndrome), *EML4-ALK* rearrangement in non-small cell lung cancer (NSCLC), *AP3B1* (Hermansky-Pudlak syndrome type 2), have been previously implicated (Bondeson et al., 1995; Soda et al., 2007; Jones et al., 2013; Rhees et al., 2014).

However, the effect of a given DNA inversion may not be as clear as the above examples. It is not known whether SVs without a clear gene-destruction effect are still benign in nature. For instance, it is likely that an inversion can disturb normal chromatin architecture and this could be translated into interchanging of hetero- and euchromatic states. This would lead to abnormal methylation patterns, and ultimately to alterations in gene expression that may have a phenotypic impact. Thus, a given gene could cross the borders between an actively transcribed and a non-active region as a result of the inversion. A clear example is what has been observed in *Drosophila* position-effect variegation, where an inversion of DNA shifts the w+ and rst+ genes from an euchromatin to a heterochromatin domain, thus resulting in white color eyes (Schotta et al., 2003).

In humans, indications for the none gene breaking effects of inversions came from Gonzalez et al. (2014) who reported on the effect of a common 0.45 Mb inversion at 16p11.2 on local gene expression and found that inverted alleles strongly correlated to neighboring gene expression. Expression effects were seen on single copy genes within the inverted regions as well as on genes flanking the duplicated regions (where the inversion breakpoints occur). Additionally, the multiple copy genes located in the duplications were also affected. Some genes are over-expressed, while others are under-expressed in the inverted allele. However, a large proportion remained unaffected. Although the molecular mechanism behind this set of observations goes beyond the scope of this particular study, it provides an indication for a cause-effect relationship between common human inversions and gene expression and its link to a disease phenotype: the joint susceptibility to asthma and obesity.

To date, there is no molecular mechanism that explains the biogenesis for the inversion effect on expression. Furthermore, it is still not possible to predict the effects of a given inversion. It has been hypothesized that changes in the chromatin structure comprise a possible underlying reason, but a clear model detailing the interplay between a given methylation and the different parameters that affect the gene expression, such as histones modifications, DNA methylation, nucleosome occupancy and three-dimensional chromatin structure, remains elusive. The above-described *F8* inversions are well characterized and their breakpoints are within defined unique repeats regions. Therefore, these two inversions are a suitable model for investigating the effect of inversions on gene expression as well as chromatin structure and epigenetic modifications.

In this study, we took advantage of the *F8* gene inversions model to analyze DNA methylation levels of CpG rich regions within and flanking the inverted DNA regions in wild type and inverted DNA (with intron 1 or intron 22 inversions). In summary, our results show clear detectable DNA methylation changes associated with inversions that are flanking the inverted regions. Therefore, methylation aberrations are a useful diagnostic tool to identify inversion structural variations.

MATERIALS AND METHODS

DNA Samples

DNA samples corresponding to healthy controls (21 non-hemophilic males) and to male hemophilia patients with known intron 1 (16 samples) or intron 22 inversions (19 samples) were obtained from the hemophilia center at the Institute of Experimental Hematology and Transfusion Medicine (University Clinic Bonn, Germany) and from the Institute for Human Genetics (University of Wuerzburg, Germany). The samples used are derived from DNA collected for molecular diagnostic purposes. All blood samples from patients and healthy controls were obtained upon written informed consent. The Ethics Committee of the University Clinic Bonn authorized the use of pre-collected DNA samples for research purposes (approval number 091/09 date 05/06/2009).

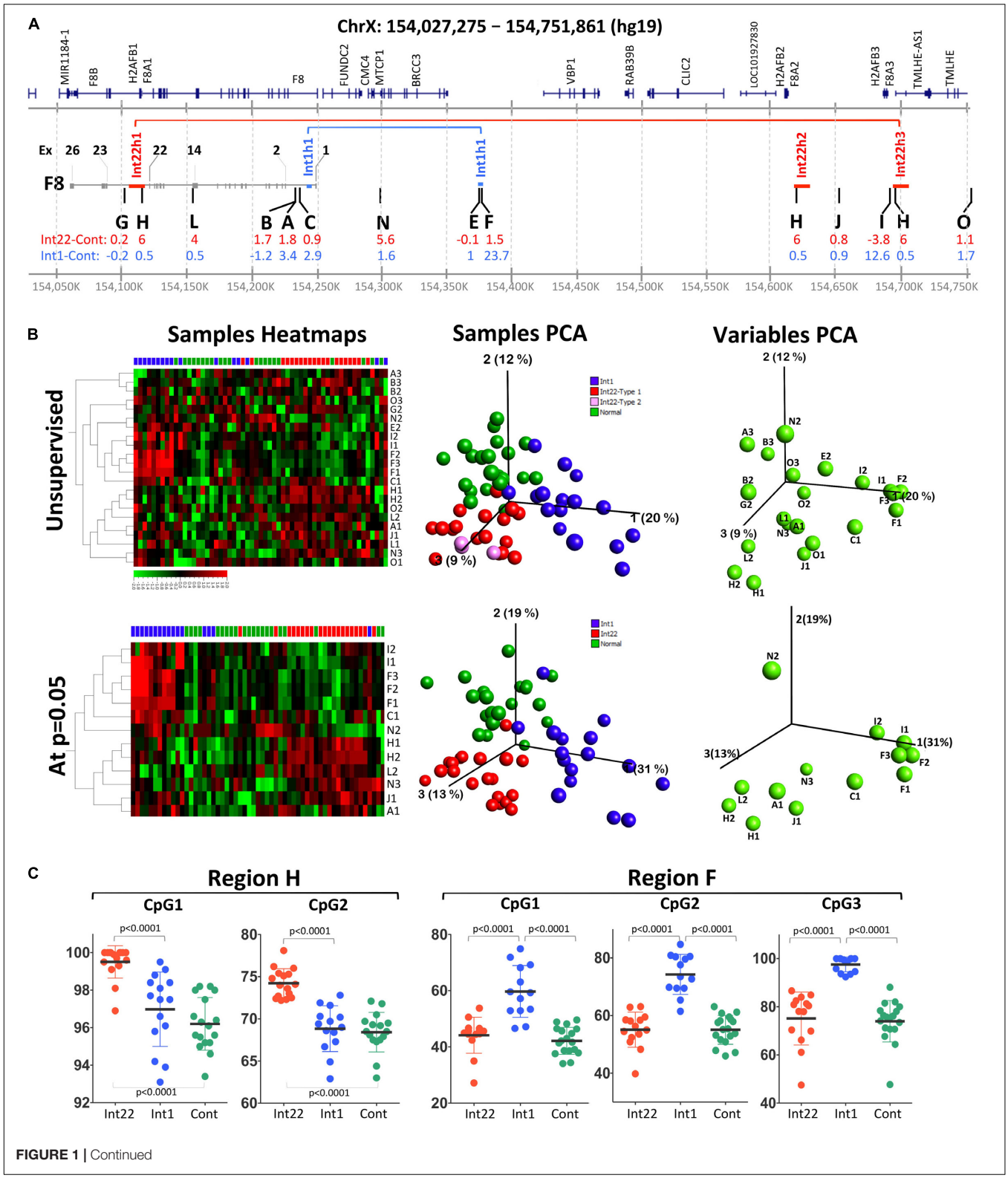


FIGURE 1 | Continued

FIGURE 1 | Pyrosequencing methylation data on 12 selected regions from intron 22 and intron 1-inversion samples as well as healthy male controls. **(A)** Detailed map on X chromosome (Chr X: 154,027,275-154,751,861:hg19) showing *F8*, the three Int22h and the two Int1h repeats involved in the inversion mutations. The positions of the studied regions are indicated in the middle region by capital letters. The inversion prone regions are labeled with red and blue horizontal lines for intron-22 and intron-1 inversions, respectively. **(B)** Methylation data represented by heatmaps, sample PCA and variable PCA plots. **(C)** Detailed data of the methylation values for individual samples at the two best regions (H and F) that clearly distinguish between inverted and non-inverted control samples.

Methylation Analysis and Pyrosequencing Assays

CpG rich regions within and around the inverted sequences were identified using the UCSC website. Feasible regions for primer designs were selected for methylation analysis. Primers for PCR amplifications as well as pyrosequencing primers were designed using the PyroMark assay-design Q24 software (Qiagen, Germany). Bisulfite treatment was done using the EZ 96-DNA methylation kit (Zymo Research, Irvine, CA, United States) following manufacturer's protocol. Bisulfite PCRs were done using HOT FIREPol (Solis Biodyne, Tartu, Estonia). Pyrosequencing was done on a PyroMark Q24 or Q96 machine (Qiagen, Germany). Primers used for amplification are listed in **Supplementary Table S1**. In total, 12 different regions were studied and designated as regions A to O.

Verification of Results by Bis-Seq NGS Based Assay

Since the pyrosequencing assay is restricted to check only few individual CpGs and provides an estimated average of methylation the results had to be verified by covering relatively larger regions and the spatial relationships (phase) between different CpGs in the same region had to be revealed. Such data could be provided by NGS-based resequencing assays. For this purpose, we chose the SeqCap Epi Enrichment system from NimbleGen (Roche, Switzerland). Using this system, we targeted the *F8* region: chrX: 154,027,275-154,751,861. Samples included four intron 1, six intron 22 inversions and four wild type controls. After obtaining the data we filtered for the overlapping reads with our pyrosequencing assays. All data are submitted to EBI as a mapped "BAM" file under study accession number "ERP113762."

Next Generation Sequencing Analysis

Sequencing data was generated using Illumina HiSeq 2500 v4 with read length of 2 × 125 bp. Reads were generated in fastq file format. Reads were pre-filtered for any adapters' sequence. Reads quality was tested using fastqc[1] and all reads were passed for the quality cut-off of 10. Reads were mapped using BSMAP (Xi and Li, 2009) program to HG38 genome downloaded from UCSC with parameter settings to WGBS mode ($-s = 16$), all four strands mapping ($-n = 1$) and with parallel computing of four processor cores ($-p = 4$). Mapped reads were split into top and bottom strand using bamtools (Barnett et al., 2011) to separately remove duplicates for both strands. Duplicates were removed using the "MarkDuplicates" function in picard tool[2]. Removed duplicate removed strands were merged together into single mapped file using bamtools. Filtered reads were filtered again for the properly paired reads using bamtools filter with parameters of "-isMapped true," "isPaired true" and "isProperPair true." Properly paired reads were further processed using "clipOverlap" function in "bamUtil" (Jun et al., 2015) to clip overlapping paired-end reads to correct bias for methylation calculation. Methylation percentages were determined using the "methRatio.py" function in BSMAP with the parameters of minimum number of reads per CpGs set to 1 ($m = 1$) and report to zero methylation ($-z$). A final methylation table with number of Cs, Ts and coverage for every CpG was created by removing the uncovered region via NimbleGen. Methylation analyses were further carried out in R using the "methylkit" (Akalin et al., 2012) package. Fisher's exact test was performed to calculate the *p*-value between samples for every CpG site.

Statistical Analysis and Data Visualization

Statistical analysis was done using R or Prism (GraphPad software). Additional data analysis and visualization were done using Qlucore Omics Explorer (Sweden) and ProFit software from Quatum Soft (Switzerland). Regression analyses using R were performed to understand the effect of covariates (age). Formula for regression analysis used were "aov(lm(MethDiff∼CaseControl+Age+CaseControl*Age)."

RESULTS

CpGs Regions at the Border of the Inverted DNA Are Prone to Significant Differential Methylation

The main aim of this study was to detect differential methylation region(s) that could serve as markers for identification of *F8* inversions rearrangements. We initially designed and selected the regions based on (1) feasibility of reading methylation of at least three CpGs, (2) their presence in a region between the three prime regions of F8 and the Int22h3 repeat regions, and (3) their presence in non-repetitive regions (like L1 and Alu). Next, we could retain 12 regions whose methylation was neither constantly 0 nor 100% for all samples: i.e., variable methylation. We then studied three groups of samples: int22 and int1 inversions and healthy controls. Three of the regions failed to show significant statistical differences when applying statistical test to compare between the groups, namely regions G, E, and O (**Figure 1**).

The rest of the nine regions showed statistical significance for at least one CpG at one of three comparisons (**Figure 1** and **Supplementary Table S2**). For intron 22 inversion samples, eight individual CpGs in six regions (regions H, L, A, N, J, and I) were

[1]http://www.bioinformatics.babraham.ac.uk/projects/fastqc

[2]http://broadinstitute.github.io/picard/

statistically different compared to healthy controls. The most significant region was region H (average meth. diff. = 6% at CpG2; *t*-test $p < 0.0001$) embedded within the Int22h repeats, followed by region L in exon 14 of factor 8 (average meth. diff. = 4% at CpG2; *t*-test $p = 0.0005$).

For intron 1 inversion samples, eight individual CpGs were statistically also significantly differentially methylated in comparison to healthy controls (**Figure 1** and **Supplementary Table S2**), covering five different regions: A, C, F, J, and I. The most significant region was region F embedded in the Int1h repeat (average meth. diff. = 23.7% at CpG3; *t*-test $p < 0.0001$), followed by region C (average meth. diff. = 2.9% at CpG1; *t*-test $p = 0.0004$). However, region I showed higher average differential methylation reaching 12.6%, but a *p*-value of 0.0011.

Of significance, the regions that showed the highest differential methylation were situated either within the repeats involved in the homologous recombination leading to the inversion (region H: intron 22 inversion and region F: intron 1 inversion) or close to that border (region C, L, and I).

Two Regions Show Promising Biomarkers Properties: High Sensitivity and Specificity, Making Them Eligible as Diagnostic Markers

In order to use methylation at a given CpG as a diagnostic marker to detect inverted DNA we calculated sensitivity and specificity for each CpG that showed a statistical significance difference between inversions and healthy controls. For this purpose, we defined sensitivity as the fraction of the inverted DNA sample that is identified as differentially methylated in comparison to the wild type controls. Whereas specificity is defined as the fraction of healthy samples within the normal range of methylation and not overlapping with inverted DNA. For intron 1 and intron 22 inversions, a sensitivity and specificity of 1 were reached for region F CpG3 and region H CpG2 (**Supplementary Table S2**).

Investigation of Factors Influencing DNA Methylation: Age and DNA Polymorphism

Age Effect: Healthy Group Shows Statistically Significant Linear Correlation Between Age and Methylation at Regions F and I and a Clear Tendency at Region L

A significant correlation between age and methylation difference was observed for some CpG sites. In order to understand whether the difference is due to age or rearrangements of the inversion region, we performed rigorous regression analysis between inversion samples and controls with age as a covariate (**Supplementary Figure S2**). Regression analysis revealed that some CpGs sites, i.e., F-CpG1, F-CpG2, F-CpG3, I-CpG1, and I-CpG2, showed statistical significance between intron-1 inversion and control in the difference between age and methylation and the difference between phenotype and methylation, while the difference between age and phenotype was not found to be significant. Thus, the difference in methylation due to intron-1 inversion will be statistically significant at any age range (**Supplementary Figure S2A**). Regarding intron-22 inversion, we found no statistically significant difference between age to methylation or phenotype to methylation (**Supplementary Figure S2B**).

In order to re-emphasize the age effect on region F and to exclude an effect on the ability to discriminate wild type from inversions at any age group, we calculated observed – expected – methylation levels for all samples in the healthy and the intron 1 inversion groups. For this purpose, we calculated the expected methylation values according to an equation of best fit linear regression model of healthy samples for each of the three CpGs and the average of the three CpGs in region F (**Supplementary Figure S3A**). Comparison of observed and expected levels showed a highly significant difference only at intron1 samples (at all three CpGs), which indicates that the observed differences between intron 1 inversions and healthy controls are not solely due to an age effect (**Supplementary Figure S3B**). Moreover, observed methylation values minus calculated expected methylation values (according to age using the linear regression fitting equation of the healthy samples) revealed high significance between healthy controls and intron 1 inversion compared to intron 22 inversions (**Supplementary Figure S3C**). This once more indicates that the differences in comparison to healthy controls are largely due to the intron 1 inversion of DNA.

DNA Polymorphism Effect

As the DNA polymorphism may affect the level of methylation at neighboring CpGs, we searched the UCSC databases for occurrences of polymorphisms in a window of 1 Kb surrounding each investigated CpG. The results are displayed in **Supplementary Table S3**. While we found some SNPs with minimum allele frequency up to 0.21 in European populations, especially in the two regions with high discrimination power to detect intron 1 (region F) and intron 22 (region H), no reported SNPs with MAF > 0.05 have been reported. Therefore, we could largely exclude a broader effect of polymorphism on the level of methylation at the two relevant regions H and F.

Methylation Correlation Between Different Regions Suggests Stochastic Random Effect, While Top Differentially Methylated Regions Are Indeed Correlated

In intron 1 and intron 22 inversions, we observed abnormal methylation at several CpG sites. Therefore, we queried whether these changes are coordinated and if they are correlated. In other words, are these changes in methylation in parallel at two or more altered regions for a given inversion type. If this is the case, a statistically significant correlation should be observed. Indeed, we calculated all correlations pairwise for all 22 CpG sites for every group (intron 22, intron 1 and healthy controls as separate groups). While we observed 15, 15, and 16 inter CpGs correlations in intron 22, intron 1 and healthy controls, we found little overlaps between all three groups. This was mainly observed at the intra-CpG correlations within the regions N and F

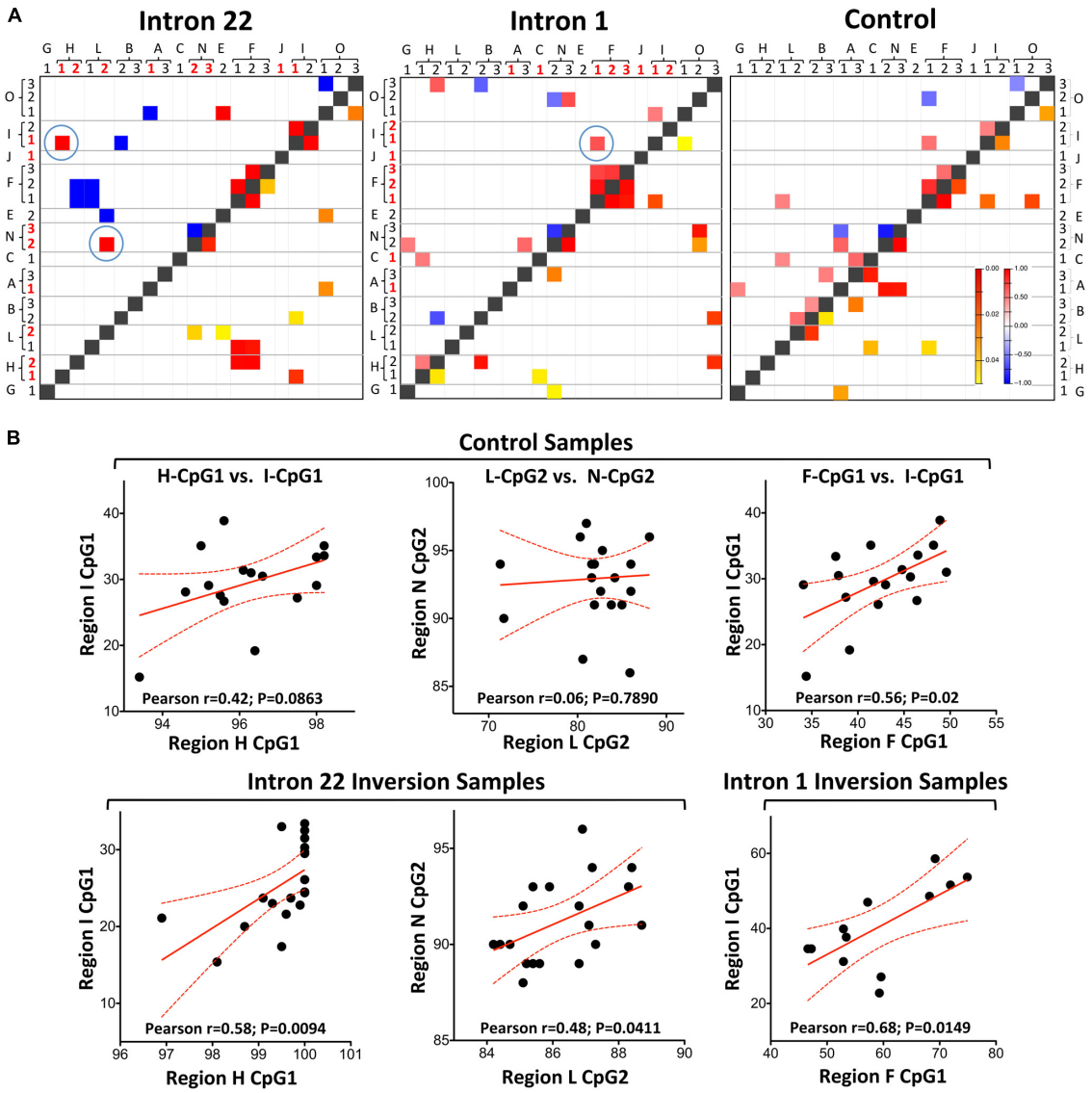


FIGURE 2 | Correlation analysis between the studied CpG sites. **(A)** Heatmaps representing Pearson correlation (left upper triangle) and *p*-values (right lower triangle). The CpG rich region names are labeled with capital letters, while the individual CpGs are labeled with numbers, whereby red ones represent statistically significant ones. When two significant CpGs from two regions are correlated they are highlighted with a blue circle. **(B)** Correlation graphs of the circled ones of part A. The best fit linear curves as well as the 95% confidence intervals are shown in red.

(**Figure 2A**). The absence of overlap suggests a change in the nature of epigenetic marks from the normal non-inverted to an inverted DNA.

In this context, three inter region correlations were observed in inversion groups that involve regions that are differentially methylated between inversions and controls. Two of these are observed in intron 22 only and are not present in controls, namely H-CpG1 vs. I-CpG1 and L-CpG2 vs. N-CpG2 (**Figure 2**). Possibly, this is specific for the inversion samples and is induced by the rearrangement. This hypothesis is supported by two arguments: (1) all four involved CpGs are at the top differentially methylated between intron 22 inversions and controls and (2) such correlation is absent in normal samples.

Yet, the third correlation was observed in intron 1 samples between F-CpG1 and I-CpG1. These CpGs also showed a significant methylation difference between intron 1 inversion samples and controls. In fact, these two CpGs are the highest two differentially methylated CpGs with an average difference between intron 1 samples and controls of 23.7 and 12.6% for F-CpG1 and I-CpG1, respectively. However, this correlation is induced by the age effect on methylation as this involves two CpGs that show high correlation between age and methylation (**Supplementary Figure S2**). In addition, this correlation is observed in healthy controls, re-emphasizing that the correlation between age and methylation is the driving force behind this correlation between the CpG1 at region F and CpG1 at region I.

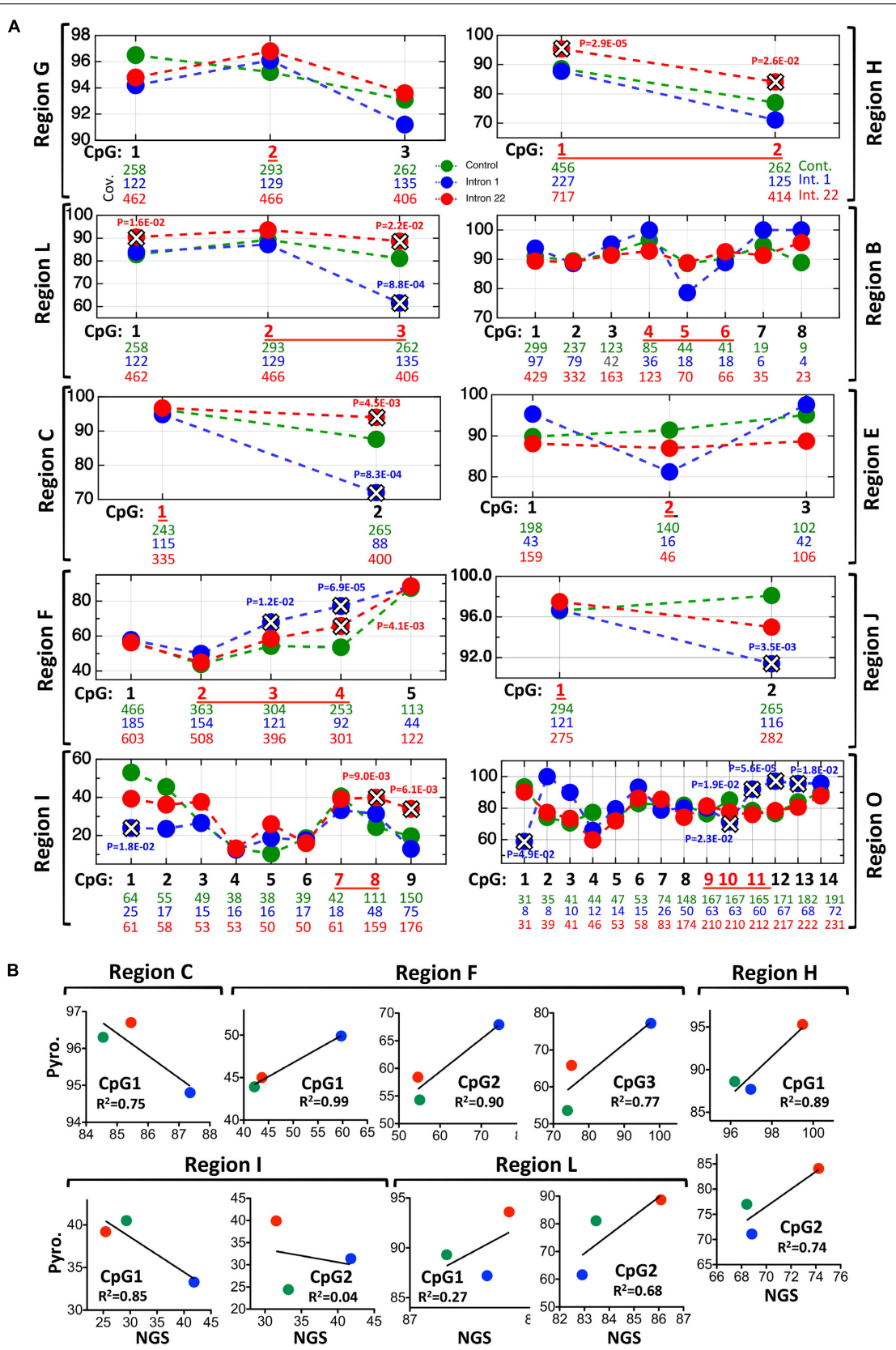


FIGURE 3 | Continued

FIGURE 3 | NGS results of the studied regions shown in **Figure 1A**. **(A)** Each graph represents one region; regions A and N have no enough coverage and are absent. The number of reads for each CpG is shown below the corresponding CpG, the *p*-value of Fisher's exact test is shown when significant (marked by X) between healthy samples (green) and intron 1-inversion samples (blue) or intron 22-inversion samples (red). The corresponding pyrosequencing CpGs are in red and underlined. **(B)** Correlation graphs between the pyrosequencing and the NGS methylation levels results.

Targeted Bisulfite Re-sequencing Largely Confirms the Pyrosequencing Results

Confirmation of Bisulfite Pyrosequencing

In order to further confirm the above results via an alternative method, we performed targeted bisulfite re-sequencing with the SeqCap Epi Enrichment system from NimbleGen (Roche, Switzerland) to capture a region containing the *F8* and extending up to the extragenic Int22h repeats (hg19, ChrX: 154,027,275–154,751,861; **Figures 3**, **4**). Ten of the studied pyrosequencing regions could be covered, while two were insufficiently covered with low read counts (regions A and N). In order to increase coverage and to decrease the effect of inter-individual differences, we merged the reads that belonged to the same group of samples. This resulted in a pool of reads of three groups including six, two and four individual DNA samples for intron 22 inversion, intron 1 inversion and healthy controls, respectively. Moreover, this merging approach increased the read numbers at each CpG site with ranges 23-717, 4-227, and 9-466 for intron 22 inversion, intron-1 inversion and healthy controls, respectively.

Nineteen individual CpGs were overlapping between the pyrosequencing and the NGS enrichment approach, of which 12 were showing complete concordance in results of significance (**Figure 3** and **Supplementary Figure S1**). This is also reflected by the correlation of average methylation between both approaches across the three samples cohorts (**Figure 3B**). Of particular interest are the two highly differentially methylated regions that showed high sensitivity and high specificity for distinguishing inversions from non-inversions, namely regions F and H for detecting intron 1 and intron 22, respectively. Both showed high significance in correlation values and overlapping, confirming results in both methods. The bisulfite targeted enrichment analysis provided additional confidence in the inversion-induced methylation aberrations and in the ability of such methylation assays to detect the inversions.

Trend Line of Methylation Changes Over the *F8* Till Intr22h3 Covered Region

Next, a global trend line was drawn of the methylation differences including all CpGs captured by the enrichment protocol (i.e., not only overlaps with pyrosequencing results as presented in the previous section). For this approach, we filtered the data to exclude any CpG overlapping with a repeat or a known SNP. Additionally, we excluded data for any CpG that had less than 30 reads in one of the two compared categories. In a next step, a trend line of difference in methylation to the healthy male controls was plotted in a map showing relative position to the studied pyrosequencing regions (**Figure 4A**). As expected, this approach indicated a major hypermethylated domain overlapping with the regions F and H for intron 1 and intron 22, respectively. However, we also noticed that the inversion breakpoints (shown as blue and red stars in **Figure 4A**) are lying in "methylation-disturbed" domains. The largest methylation disturbance in both magnitude and length of the domain appear to be overlapping with the inversion junctions. All of the above suggest that the observed methylation alterations are indeed reflection of new genomic architecture caused by the DNA inversion.

Characteristics of CpGs Showing Differential Methylation

We investigated the relationship between the degree of CpG methylation difference and the density of CpGs in a window of 50 bp where the CpG in question is in the center (**Figure 4B**). Using all data for all CpGs (regardless of statistical significance) we found that a clear and highly significant relationship between the methylation differences and CpG density exists where relatively CpG dense regions are more stable and show smaller methylation differences. This applies for the comparisons healthy vs. intron 1 ($r = -0.909$, Fisher's exact test $p < 0.0001$) and healthy vs. intron 22 ($r = -0.904$, Fisher's exact test $p < 0.0001$) (**Figure 4B**). It is our opinion that this is a general phenomenon of variability of methylation at "stand-alone" CpGs where they are more prone to uncontrolled "natural" fluctuations. However, the CpG methylation at significantly differentially methylated CpGs failed to show this correlation indicating that the latter are the result of induced aberrant methylation due to DNA rearrangement. From this analysis, we conclude that statistically significant methylation changes are more likely to occur at CpG rich regions or at clusters rather than at sole dispersed (non-clustered) CpGs.

DISCUSSION

The human genome shows significant variability between individuals (Auton et al., 2015). This variability is caused by single nucleotide polymorphisms, deletions, duplications, translocations and inversions. The effect of which may either be detectable as a change in the phenotype (which include disease manifestation) or be benign without observable phenotype. The molecular mechanism for the former can be explained for SNPs, deletions or duplications by virtue of possible changes in the DNA sequences leading to altered gene expression or protein structure. However, in the cases of translocations or inversions, there is no net gain or loss in DNA. Therefore, the association to a phenotype is difficult to explain by DNA changes unless the breakpoints disrupt a coding sequence or an expression-regulatory element (like a promoter or enhancer) (Harewood and Fraser, 2014). However, an additional scenario could be responsible to cause a

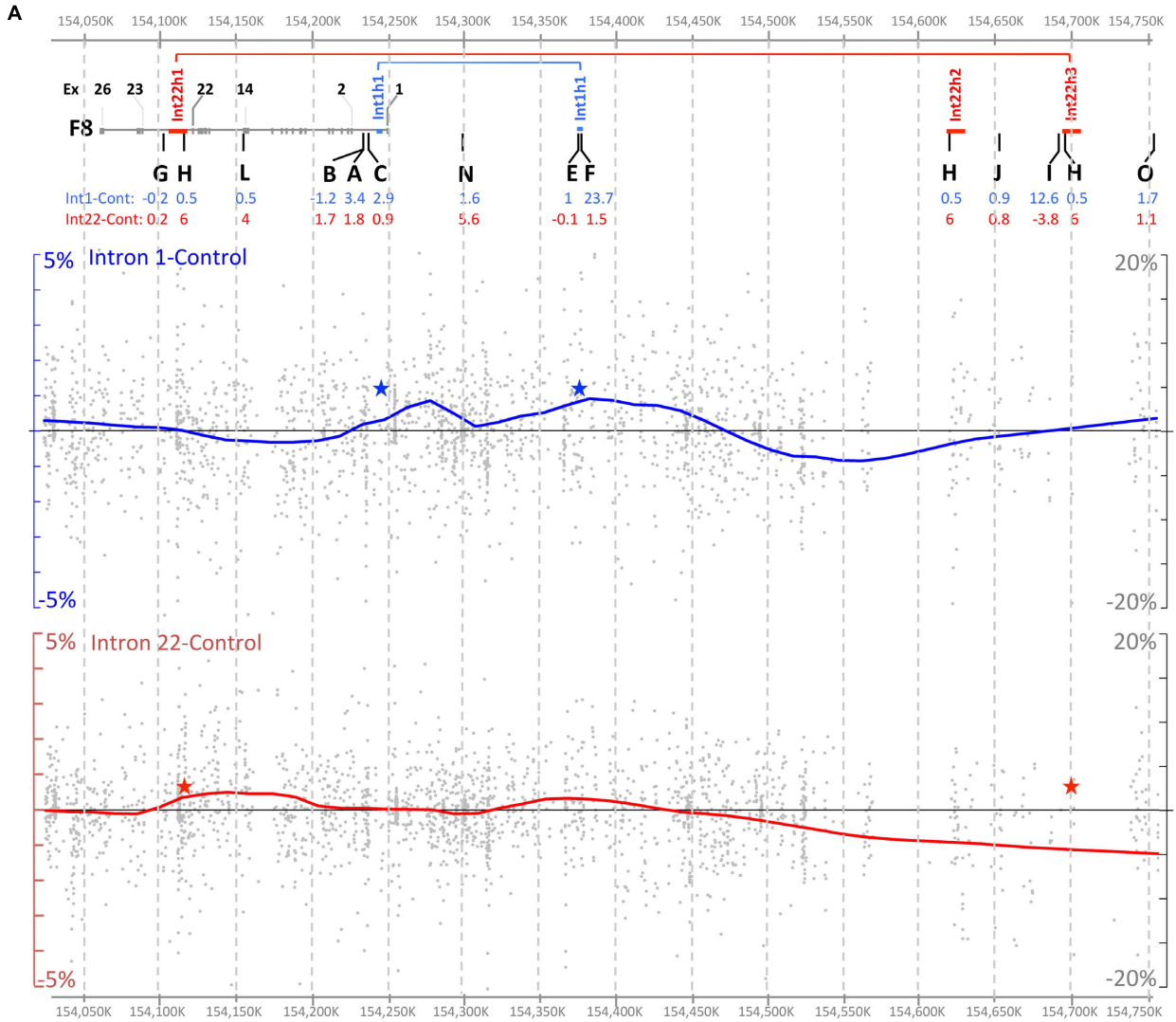


FIGURE 4 | Continued

FIGURE 4 | Global visualization of NGS data in the *F8* region (hg19: Chr X 154,027,275–154,751,861). **(A)** Upper panel shows the relative positions of the studied pyrosequencing regions, the middle panel shows the NGS data for intron 1 inversion samples and the lower panel the intron 22 inversion samples. The covered individual CpG methylation data are represented by a gray dot, while additionally, the data is represented by a smooth curve representing the trend of changes between the inverted and the control samples. CpG sites with less than 30 coverage or overlapping with known SNPs or repeats were excluded. Red and blue stars indicate the DNA inversion junctions. **(B)** Correlation between the methylation differences at a given CpG and the density of CpG within 50 bp flanking region. Left and right side include all CpG data and only significant data (Fisher's exact test), respectively.

phenotype: a shift in chromatin structure or – as it is also known – a position effect variegation (PEV).

Position effect variegation is one of similar phenomena which occur due to relocation of a genomic segment from one region to another and it has been extensively studied in *Drosophila*, yeast, mice and cultured human cells (Tham and Zakian, 2002; Pedram et al., 2006; Elgin and Reuter, 2013; Tchasovnikarova et al., 2015). Inversion prone position effects are not only limited to other species, it has also been reported in some human disease conditions, such as aniridia (*PAX6*), campomelic dysplasia (*SOX9*), familial adenomatous polyposis (*APC*) and Saethre-Chotzen syndrome (*TWIST1*) (Fantes et al., 1995; de Chadarevian et al., 2002; Cai et al., 2003; Velagaleti et al., 2005). Of note, some inversion variants can also act as risk factor for the offspring in microdeletion syndromes, such as Williams–Beuren syndrome, Angelman syndrome and Sotos syndrome (Osborne et al., 2001; Gimelli et al., 2003; Visser et al., 2005).

The above would lead to recreation of chromatin domains that result in local and regional epigenetic changes like DNA methylation aberrations. In this study, we used the two inversion hotspots in the *F8* gene at Xq28 as a model to investigate the global methylation aberration. Indeed, we found specific changes associated with each of the two inversions. With one specific region for each of the inversions showing high sensitivity and high specificity, our results pave the way for the use of methylation-based assay to detect the inversion. The hemophilia A methylation inversion "HAMI" assay will have several advantages over traditional assays. It is noteworthy to mention that repetitive elements also play an important role in generating structural variants (SVs) in humans (Xing et al., 2009). Among all mobile element types, long interspersed element-1 (LINE-1, or L1) has been previously investigated for DNA methylation-related changes in diseased conditions (Nusgen et al., 2015; Sharma et al., 2019). In this particular study, we took advantage of one such full length L1 repeat (region O) located in the vicinity of the *F8* gene and evaluated the methylation status of this repeat in both inversion type patients. However, no differences were found between inversion and wild type.

Currently, the gold standard molecular diagnostic assay to detect the inversion is the inverse based PCR assay (Rossetti et al., 2005), a procedure that needs up to 2–3 working days to complete and requires a skilled technician to perform a critical ligation step. In comparison, the HAMI assay includes three fail-free steps: (1) bisulfite conversion, (2) PCR and (3) quantitative pyrosequencing, all of which could be performed in 1 day. An additional advantage for HAMI is that it does not detect a specific DNA junction. Therefore, no specific amplification primers to detect only known inversions are required, while any rearrangement that could still be missed by specific amplification across known rearrangement junctions will be detected. However, disadvantages and limitations of such an assay include establishment of controls to define the relative borders (cut-off) of normal levels, as this could be population- or ethnicity-specific.

Overall, we could determine the methylation levels at multiple regions surrounding/overlapping *F8* associated genomic inversions at Xq28 region. Further evaluations are required to establish whether these epigenetic changes are cause or consequence of these inversion events.

DATA AVAILABILITY

The datasets generated for this study can be found in the ENA, https://www.ebi.ac.uk/ena/data/view/PRJEB31235.

ETHICS STATEMENT

The Ethic committee of the University Clinic in Bonn authorized the use of pre-collected DNA samples for research purposes, approval number 091/09 date 05/06/2009.

AUTHOR CONTRIBUTIONS

MJ, AS, and OE-M wrote the manuscript. MJ, AS, BP, and OE-M analyzed the results. MJ and OE-M performed the bioinformatics analysis. AS and NN performed the experiments. BP, AH, AP, and JO provided the samples and infrastructure and commented on the manuscript. OE-M designed the study.

FUNDING

This study was supported by an ASPIRE hemophilia research award (2014) from Pfizer (grant reference WI193463) and institutional fund of Institute of Experimental Hematology and Transfusion Medicine, University of Bonn, Germany.

ACKNOWLEDGMENTS

We thank the patients for donating DNA material for research purposes.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00508/full#supplementary-material

FIGURE S1 | Detailed pyrosequencing methylation data on 12 selected regions from intron 22 and intron 1-inversion samples as well as healthy male controls. **(A)** Detailed map of X chromosome (154,027,275–154,751,861:hg19) showing the F8, the three Int22h and the two Int1h repeats involved in the inversion mutations. The positions of the studied regions are represented in the middle by capital letters, below which is the max-observed difference in methylation average. The inversion prone regions are labeled with red and blue horizontal lines for the intron-22 and intron-1 inversions, respectively. (B) Vertical scatter plots represent the detail methylation levels of individual studied samples for all individual CpG sites. The significant *t*-test is shown in the figure.

FIGURE S2 | Age covariate regression analysis showing correlation between age and methylation levels for healthy controls in comparison to intron 1-inversion samples **(A)** and to intron 22-inversion samples. **(B)** Every plot shows methylation data vs. age. Above the individual plots are Pearson correlation p and rho values, while below the *p*-values of age-covariate analysis are shown. All significant *p*-values are written in bold. In case of significant Pearson correlation, the values are labeled with solid transparent red, green or blue rectangles. The plots corresponding to significant differences between cases and controls, even after considering age as covariates (*P*-value. MethDiff~CaseControls), are indicated by a red cadre.

FIGURE S3 | Calculation of observed vs. expected methylations values according to predicted linear regression formula of methylation vs. age of healthy group. **(A)** Linear regression curves of methylation vs. age for the three groups of samples (intron 1-inversion samples, intron 22-inversion samples and healthy controls). Also, equations are shown for every CpG (red, green, and blue for CpGs 1, 2, and 3, respectively) and for the average methylation of three CpGs (in black). **(B)** Comparison between observed and calculated expected values according to the linear regression equation of healthy controls. *T*-test *p*-values showed significance of all CpGs and their average only in the intron 1 inversion group. **(C)** Comparisons of observed-expected methylation values between inversion groups and healthy controls.

TABLE S1 | Primers list used in this study.

TABLE S2 | Summary values of data presented in **Figure 1** together with calculated sensitivity and specificity.

TABLE S3 | Common SNPs around the studied CpG site.

REFERENCES

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F. E., Figueroa, M. E., Melnick, A., et al. (2012). methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* 13:R87. doi: 10.1186/gb-2012-13-10-r87

Andrikovics, H., Klein, I., Bors, A., Nemes, L., Marosi, A., Varadi, A., et al. (2003). Analysis of large structural changes of the factor VIII gene, involving intron 1 and 22, in severe hemophilia A. *Haematologica* 88, 778–784.

Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. doi: 10.1038/nature15393

Bagnall, R. D., Giannelli, F., and Green, P. M. (2006). Int22h-related inversions causing hemophilia A: a novel insight into their origin and a new more discriminant PCR test for their detection. *J. Thromb. Haemost.* 4, 591–598. doi: 10.1111/j.1538-7836.2006.01840.x

Bagnall, R. D., Waseem, N., Green, P. M., and Giannelli, F. (2002). Recurrent inversion breaking intron 1 of the factor VIII gene is a frequent cause of severe hemophilia A. *Blood* 99, 168–174. doi: 10.1182/blood.v99.1.168

Barnett, D. W., Garrison, E. K., Quinlan, A. R., Stromberg, M. P., and Marth, G. T. (2011). BamTools: a C++ API and toolkit for analyzing and managing Bam files. *Bioinformatics* 27, 1691–1692. doi: 10.1093/bioinformatics/btr174

Bondeson, M. L., Dahl, N., Malmgren, H., Kleijer, W. J., Tonnesen, T., Carlberg, B. M., et al. (1995). Inversion of the IDS gene resulting from recombination with IDS-related sequences is a common cause of the hunter syndrome. *Hum. Mol. Genet.* 4, 615–621. doi: 10.1093/hmg/4.4.615

Cai, J., Goodman, B. K., Patel, A. S., Mulliken, J. B., Van Maldergem, L., Hoganson, G. E., et al. (2003). Increased risk for developmental delay in saethre-chotzen syndrome is associated with TWIST deletions: an improved strategy for TWIST mutation screening. *Hum. Genet.* 114, 68–76. doi: 10.1007/s00439-003-1012-7

de Chadarevian, J. P., Dunn, S., Malatack, J. J., Ganguly, A., Blecker, U., and Punnett, H. H. (2002). Chromosome rearrangement with no apparent gene mutation in familial adenomatous polyposis and hepatocellular neoplasia. *Pediatr. Dev. Pathol.* 5, 69–75. doi: 10.1007/s10024-001-0121-3

Elgin, S. C., and Reuter, G. (2013). Position-effect variegation, heterochromatin formation, and gene silencing in drosophila. *Cold Spring Harb. Perspect. Biol.* 5:a017780. doi: 10.1101/cshperspect.a017780

Fantes, J., Redeker, B., Breen, M., Boyle, S., Brown, J., Fletcher, J., et al. (1995). Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. *Hum. Mol. Genet.* 4, 415–422. doi: 10.1093/hmg/4.3.415

Gimelli, G., Pujana, M. A., Patricelli, M. G., Russo, S., Giardino, D., Larizza, L., et al. (2003). Genomic inversions of human chromosome 15q11-q13 in mothers of angelman syndrome patients with class II (BP2/3) deletions. *Hum. Mol. Genet.* 12, 849–858. doi: 10.1093/hmg/ddg101

Gonzalez, J. R., Caceres, A., Esko, T., Cusco, I., Puig, M., Esnaola, M., et al. (2014). A common 16p11.2 inversion underlies the joint susceptibility to asthma and obesity. *Am. J. Hum. Genet.* 94, 361–372. doi: 10.1016/j.ajhg.2014.01.015

Harewood, L., and Fraser, P. (2014). The impact of chromosomal rearrangements on regulation of gene expression. *Hum. Mol. Genet.* 23, R76–R82. doi: 10.1093/hmg/ddu278

Jones, M. L., Murden, S. L., Brooks, C., Maloney, V., Manning, R. A., Gilmour, K. C., et al. (2013). Disruption of AP3B1 by a chromosome 5 inversion: a new disease mechanism in hermansky-pudlak syndrome type 2. *BMC Med. Genet.* 14:42. doi: 10.1186/1471-2350-14-42

Jun, G., Wing, M. K., Abecasis, G. R., and Kang, H. M. (2015). An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res.* 25, 918–925. doi: 10.1101/gr.176552.114

Lakich, D., Kazazian, H. H. Jr., Antonarakis, S. E., and Gitschier, J. (1993). Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. *Nat. Genet.* 5, 236–241. doi: 10.1038/ng1193-236

Naylor, J. A., Buck, D., Green, P., Williamson, H., Bentley, D., and Giannelli, F. (1995). Investigation of the factor VIII intron 22 repeated region (int22h) and the associated inversion junctions. *Hum. Mol. Genet.* 4, 1217–1224. doi: 10.1093/hmg/4.7.1217

Nusgen, N., Goering, W., Dauksa, A., Biswas, A., Jamil, M. A., Dimitriou, I., et al. (2015). Inter-locus as well as intra-locus heterogeneity in LINE-1 promoter methylation in common human cancers suggests selective demethylation pressure at specific CpGs. *Clin. Epigenetics* 7:17. doi: 10.1186/s13148-015-0051-y

Oldenburg, J., and El-Maarri, O. (2006). New insight into the molecular basis of hemophilia A. *Int. J. Hematol.* 83, 96–102. doi: 10.1532/ijh97.06012

Osborne, L. R., Li, M., Pober, B., Chitayat, D., Bodurtha, J., Mandel, A., et al. (2001). A 1.5 million-base pair inversion polymorphism in families with williams-beuren syndrome. *Nat. Genet.* 29, 321–325. doi: 10.1038/ng753

Payne, A. B., Miller, C. H., Kelly, F. M., Michael Soucie, J., and Craig Hooper, W. (2013). The CDC hemophilia a mutation project (CHAMP) mutation list: a new online resource. *Hum. Mutat.* 34, E2382–E2391. doi: 10.1002/humu.22247

Pedram, M., Sprung, C. N., Gao, Q., Lo, A. W., Reynolds, G. E., and Murnane, J. P. (2006). Telomere position effect and silencing of transgenes near telomeres in the mouse. *Mol. Cell. Biol.* 26, 1865–1878. doi: 10.1128/mcb.26.5.1865-1878.2006

Puig, M., Casillas, S., Villatoro, S., and Caceres, M. (2015). Human inversions and their functional consequences. *Brief. Funct. Genomics* 14, 369–379. doi: 10.1093/bfgp/elv020

Rhees, J., Arnold, M., and Boland, C. R. (2014). Inversion of exons 1-7 of the MSH2 gene is a frequent cause of unexplained lynch syndrome in one local population. *Fam. Cancer* 13, 219–225. doi: 10.1007/s10689-013-9688-x

Rossetti, L. C., Radic, C. P., Larripa, I. B., and De Brasi, C. D. (2005). Genotyping the hemophilia inversion hotspot by use of inverse PCR. *Clin. Chem.* 51, 1154–1158. doi: 10.1373/clinchem.2004.046490

Schotta, G., Ebert, A., Dorn, R., and Reuter, G. (2003). Position-effect variegation and the genetic dissection of chromatin regulation in drosophila. *Semin. Cell Dev. Biol.* 14, 67–75. doi: 10.1016/s1084-9521(02)00138-6

Sharma, A., Jamil, M. A., Nuesgen, N., Dauksa, A., Gulbinas, A., Schulz, W. A., et al. (2019). Detailed methylation map of LINE-1 5' promoter region reveals hypomethylated CpG hotspots associated with tumor tissue specificity. *Mol. Genet. Genomic Med.* 2019:e601. doi: 10.1002/mgg3.601

Soda, M., Choi, Y. L., Enomoto, M., Takada, S., Yamashita, Y., Ishikawa, S., et al. (2007). Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* 448, 561–566. doi: 10.1038/nature05945

Tchasovnikarova, I. A., Timms, R. T., Matheson, N. J., Wals, K., Antrobus, R., Gottgens, B., et al. (2015). Gene silencing. epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science* 348, 1481–1485. doi: 10.1126/science.aaa7227

Tham, W. H., and Zakian, V. A. (2002). Transcriptional silencing at saccharomyces telomeres: implications for other organisms. *Oncogene* 21, 512–521. doi: 10.1038/sj.onc.1205078

Velagaleti, G. V., Bien-Willner, G. A., Northup, J. K., Lockhart, L. H., Hawkins, J. C., Jalal, S. M., et al. (2005). Position effects due to chromosome breakpoints that map approximately 900 Kb upstream and approximately 1.3 Mb downstream of SOX9 in two patients with campomelic dysplasia. *Am. J. Hum. Genet.* 76, 652–662. doi: 10.1086/429252

Visser, R., Shimokawa, O., Harada, N., Kinoshita, A., Ohta, T., Niikawa, N., et al. (2005). Identification of a 3.0-kb major recombination hotspot in patients with Sotos syndrome who carry a common 1.9-Mb microdeletion. *Am. J. Hum. Genet.* 76, 52–67. doi: 10.1086/426950

Xi, Y., and Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* 10:232. doi: 10.1186/1471-2105-10-232

Xing, J., Zhang, Y., Han, K., Salem, A. H., Sen, S. K., Huff, C. D., et al. (2009). Mobile elements create structural variation: analysis of a complete human genome. *Genome Res.* 19, 1516–1526. doi: 10.1101/gr.091827.109

Zimmermann, M. A., Oldenburg, J., Muller, C. R., and Rost, S. (2011). Unusual genomic rearrangements in introns 1 and 22 of the F8 gene. *Hamostaseologie* 31(Suppl. 1), S69–S73.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Jamil, Sharma, Nuesgen, Pezeshkpoor, Heimbach, Pavlova, Oldenburg and El-Maarri. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

REVIEW
published: 14 June 2019
doi: 10.3389/fgene.2019.00570

Epigenetic Changes in the Pathogenesis of Rheumatoid Arthritis

Marina V. Nemtsova[1,2], Dmitry V. Zaletaev[1,2], Irina V. Bure[1], Dmitry S. Mikhaylenko[1,2], Ekaterina B. Kuznetsova[1,2], Ekaterina A. Alekseeva[1,2], Marina I. Beloukhova[1], Andrei A. Deviatkin[1], Alexander N. Lukashev[1,3] and Andrey A. Zamyatnin Jr.[1,4]**

[1] Institute of Molecular Medicine, I.M. Sechenov First Moscow State Medical University (Sechenov University), Moscow, Russia, [2] Laboratory of Epigenetics, Research Centre for Medical Genetics, Moscow, Russia, [3] Martsinovsky Institute of Medical Parasitology, Tropical and Vector Borne Diseases, I.M. Sechenov First Moscow State Medical University (Sechenov University), Moscow, Russia, [4] A.N. Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia

OPEN ACCESS

Edited by:
Yun Liu,
Fudan University, China

Reviewed by:
Marco Magistri,
University of Miami, United States
Rowan Hardy,
University of Birmingham,
United Kingdom

***Correspondence:**
Marina V. Nemtsova
nemtsova_m_v@mail.ru
Andrey A. Zamyatnin Jr.
zamyat@belozersky.msu.ru

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: 26 March 2019
Accepted: 31 May 2019
Published: 14 June 2019

Citation:
Nemtsova MV, Zaletaev DV, Bure IV, Mikhaylenko DS, Kuznetsova EB, Alekseeva EA, Beloukhova MI, Deviatkin AA, Lukashev AN and Zamyatnin AA Jr (2019) Epigenetic Changes in the Pathogenesis of Rheumatoid Arthritis. Front. Genet. 10:570. doi: 10.3389/fgene.2019.00570

Rheumatoid arthritis (RA) is a systemic autoimmune disease that affects about 1% of the world's population. The etiology of RA remains unknown. It is considered to occur in the presence of genetic and environmental factors. An increasing body of evidence pinpoints that epigenetic modifications play an important role in the regulation of RA pathogenesis. Epigenetics causes heritable phenotype changes that are not determined by changes in the DNA sequence. The major epigenetic mechanisms include DNA methylation, histone proteins modifications and changes in gene expression caused by microRNAs and other non-coding RNAs. These modifications are reversible and could be modulated by diet, drugs, and other environmental factors. Specific changes in DNA methylation, histone modifications and abnormal expression of non-coding RNAs associated with RA have already been identified. This review focuses on the role of these multiple epigenetic factors in the pathogenesis and progression of the disease, not only in synovial fibroblasts, immune cells, but also in the peripheral blood of patients with RA, which clearly shows their high diagnostic potential and promising targets for therapy in the future.

Keywords: rheumatoid arthritis, epigenetics, DNA methylation, miRNA, histone modifications, circRNA

INTRODUCTION

Rheumatoid arthritis (RA) is a chronic auto-inflammatory disease of connective tissue with progressive joint damage and systemic disorders that affects around 1% of the world's population (Cribbs A. et al., 2015). RA can cause various symptoms, clinical forms and prognoses. The incidence of RA begins to increase at the age of 25 years; at the age of 55, the incidence of RA is reaching a plateau (Gabriel, 2001). For example, RA is more than six times greater in 60- to 64-year-old women compared to 18- to 29-year-old women (Melorose et al., 2015). The prevalence of RA varies in different ethnic groups. For example, the incidence of RA among American Indians is 7%, while for some other nations it is 0.2–0.4% (Ferucci et al., 2005). As in most other autoimmune diseases, RA is more common among women than among men in a ratio of 2–3 to 1

(van Vollenhoven, 2009). Based on this fact, there are assumptions that estrogens are actively involved in the pathogenesis of the disease (Wluka et al., 2000).

The etiology of RA remains unknown. It is considered to occur in the presence of genetic predispositions and provoking environmental factors. The heritability of RA has been shown from twin studies to be 60% (Yarwood et al., 2016). Earlier genealogical studies and modern molecular-genetic investigations confirm the involvement of genetic factors in RA development. Accumulation of the disease cases was revealed within families along with an increased risk of RA among first-degree relatives of the patients (Sparks and Costenbader, 2014).

The pathological process in RA represents an autoimmune inflammation of the synovial membrane of joints with synovial cells proliferation and pannus formation. This tumor-like aggressive granulation tissue promotes articular cartilage erosion and bones destruction. Synovial tissue dysfunction allows macrophages, fibroblasts and activated lymphocytes to penetrate into it. T-lymphocytes produce a variety of proinflammatory cytokines, predominantly belonging to tumor necrosis factor (TNF) and interleukin (IL) superfamilies as well as growth factors (Firenstein et al., 2013). B-lymphocytes are involved in the production of autoantibodies such as rheumatoid factor (RF) and antibodies against cyclic citrullinated peptide (anti−CCP). Differencies in expression of anti−CCP and RF, rate of disease manifestation and variability of response to therapy cause heterogeneity of RA patients indicating different pathophysiological mechanisms implication in the disease development and progression.

Genetic heterogeneity does not explain all the features of RA (Viatte et al., 2013). Thus, investigation of epigenetic factors and mechanisms associated with the progression of the disease and response to treatment is increasingly important. Investigation of the epigenetic landscape can provide novel therapeutic targets (Glant et al., 2014).

Different levels of DNA organization and chromatin packing in the eukaryotic cell's nucleus are points of application for the epigenetic regulation. Epigenetic mechanisms regulate chromatin structure and create clear patterns of gene expression during cell differentiation.

The chromatin structure regulates gene transcription by altering DNA regulatory regions (such as promoters and enhancers) availability for transcription factors (TF). An open chromatin structure, euchromatin, enables DNA-binding proteins and TF to interact with regulatory DNA sequences, leading to active gene transcription. Conversely, heterochromatin is a closed condensed chromatin state, where DNA is tightly bound to protein complexes forming a superspiralized structure. It prevents TF interaction with regulatory sequences, thus inactivating gene expression and leading to its silencing. Transcription factors, non-coding RNAs (ncRNAs), DNA methylation, histone modification and microRNAs (miRNAs) affect gene transcription without changing the DNA sequence itself (reviewed by Golbabapour et al., 2011). Specific epigenetic landscape of chromatin determines a differential gene expression and regulates various cellular processes in physiological and pathological conditions.

Epigenetic changes in RA have been studied both in mononuclear cells of peripheral blood and in different types of immune cells such as monocytes, T-cells and B-cells (Ospelt, 2016). At the same time, the epigenetic modifications in the rheumatoid arthritis synovial fibroblasts (RASFs) are of particular interest because of their aggressive phenotype, which is stable for several passages in cell culture (Hardy et al., 2013). RASFs are clue cells of joint damage and inflammation development due to pro-inflammatory and catabolic molecules synthesis, promoting abnormal proliferation and invasiveness. Implantation of RASFs together with normal human cartilage to immunodeficient mice revealed cell attachment and cartilage destruction without any proinflammatory stimuli. Such behavior was not observed in osteoarthritis (OA) synovial fibroblasts and is presumably related to epigenetic changes in these cells due to the specific pathology only (Lefèvre et al., 2009).

ABERRANT DNA METHYLATION IN IMMUNE CELLS AND PERIPHERAL BLOOD CELLS IN RA

DNA methylation is a biochemical process of methyl group binding with the cytosine ring carbon at position 5 to form 5-methylcytosine (5-mC). In mammals, DNA methylation occurs preferentially in CpG dinucleotides located throughout the whole gene either as single dinucleotide or concentrated into CpG-islands in vicinity of gene promoters. Hypermethylation of the promoters is an indicator of dense heterochromatin conformation, which blocks the binding of TF to DNA and leads to inactivation of gene transcription. The low-level methylation of promoters (hypomethylation) is associated with open chromatin conformation and active transcription of the gene (Eden and Cedar, 1994). DNA methylation is a reversible process, which could therefore be considered as a therapeutic target.

Recent studies confirmed a global DNA hypomethylation in T-cells and monocytes of RA patients compared to healthy individuals (de Andres et al., 2015). Genome-wide analysis of DNA methylation by microarrays revealed its alterations in B-cells on the early stages of RA in patients who have not yet received treatment compared to healthy donors (Glossop et al., 2016).

Cribbs et al. (2014) analyzed an aberrant function of regulatory T cells (Treg) in RA patients and found a specific region in the promoter of the *CTLA-4* (−658 CpG), which was hypermethylated in comparison with healthy controls. DNA hypermethylation prevents binding of the nuclear factor of activated T cells (NF-AT) with cytoplasmic one, called NF-ATc2, which leads to decrease of *CTLA-4* expression. As a consequence, Treg cells were unable to induce expression and activation of the tryptophan-degrading enzyme indoleamine 2,3-dioxygenase (IDO), which in turn resulted in a failure to activate the immunomodulatory kynurenine pathway (Cribbs et al., 2014). Furthermore, treatment with methotrexate induced DNA hypomethylation of *FoxP3* locus in Treg. This results in the gene

upregulation with consequent increase of CTLA-4 concentration and normalization of Treg function in RA. These studies clearly illustrate how aberrant DNA methylation can affect cell functions and how epigenetic mechanisms can be used in therapy (Cribbs A.P. et al., 2015). To determine differentially methylated regions as potential epigenetic risk factors and markers of RA predispositions, Liu et al. (2013) performed epigenome-wide association study. Using Illumina 450k microarrays they examined more than 485,000 CpG sites in peripheral blood of 354 RA patients and 337 healthy donors. As a result, 10 differentially methylated CpG sites were identified. All of them are localized on 6p12.1 and form two separate clusters within the locus also containing the genes of the major histocompatibility complex (MHC) that is known as the risk locus of RA (Raychaudhuri et al., 2012). This confirms the role of DNA methylation as an additional mechanism determining susceptibility to RA. Importantly, the heterogeneity of cell population isolated from a whole blood may cause diverse methylation profile. Thus, this factor should be taken into account in bioinformatic analysis to reduce possible biases.

Some of these results were confirmed by other studies. Aberrant DNA methylation was detected in peripheral blood mononuclear cells (PBMCs) of RA patients. For example, van Steenbergen et al. (2014) demonstrated that cg23325723 site was significantly associated with RA (p = 0.026) in PBMCs. Four other CpG sites (cg16609995, cg19555708, cg19321684, and cg25949002) demonstrated similar different methylation in PBMCs comparing to control samples, which was not, however, statistically significant.

Other studies have shown abnormal methylation of one cytosine in the *IL-6* promoter in RA PBMCs associated with reduction of its transcription (Nile et al., 2008). At the same time the loss of cytosine methylation in the *IL-10* promoter correlates with higher expression of IL-10 in such cells (Chen et al., 2011).

LRPAP1 gene is expressed in PBMCs and encodes the chaperone of low density lipoprotein receptor-related protein 1, that affects the activity of transforming growth factor beta (TGF-β) (Kolker et al., 2012). It was found that 4 CpG-dinucleotides in exon 7 of *LRPAP1* were hypermethylated in patients who demonstrated no response to the therapy by TNF inhibitors (etanercept) compared to responders. The locus of cg04857395 overlaps structures involved in alternative splicing: the region associated with trimethylation of histone H3 at lysine 36 (H3K36me3) and the binding site of CCCTC-binding factor, which is a methyl-sensitive transcriptional repressor (Lev Maor et al., 2015).

An important point to consider in epigenetic studies of PBMCs is the effect of cell heterogeneity. If the experimental data are not normalized according to the proportion of the cells of different types in the fraction of PBMCs, the differentially methylated regions (DMRs) in certain cell types could be missed.

DNA methylation in peripheral blood mononuclear cells was recently described by Zhu et al. (2019). DNA methylation profiling and gene expression profiling were measured in patients with RA and in healthy controls. Differentially methylated sites and genes identified an interferon inducible gene interaction network. The significance of PARP9 gene methylation and its associated change in the expression in the pathogenesis of RA was demonstrated. In addition, its ability to positively regulate IL2, which stimulates various cells of the immune response, has been revealed (Zhu et al., 2019).

Epigenetic regulation of immune cells can be crucial for the development and maintenance of autoimmune diseases, such as RA. Julià et al. (2017) investigated the methylation patterns of B lymphocytes in patients with RA and systemic lupus erythematosus. Differentially methylated in patients and in the control group CpG sites were located in the CD1C, TNFSF10, PARVG, NID1, DHRS12, ITPK1, ACSF3, and TNFRSF13C genes and two intergenic regions (10p12.31). Differential methylation of these genes was also reproduced in the cohort of patients with SLE. This indicates similar patterns of epigenetic changes in B-lymphocytes in these two autoimmune diseases (Julià et al., 2017).

ABERRATIONS OF DNA METHYLATION IN SYNOVIAL FIBROBLASTS IN RA

Rheumatoid arthritis synovial fibroblasts (RASFs) have a unique, non-random methylation pattern - methylome, which is specifically reorganized during the disease progression and varies depending on the joint localization. A precise mechanism of the methylome changing remains still unclear, but the overall pattern of differential methylation corresponds to the aggressive phenotype acquisition in SF results in the development of the disease. Earlier studies of epigenetic changes in RASFs demonstrated an abnormal expression of retroviral sequences LINE-1, associated with loss of silencing of these mobile elements as a result of hypomethylation (Neidhart et al., 2000). Global DNA hypomethylation is observed in many hyperproliferating tissues and is associated with a relative lack of methyl groups' donor *S*-adenosylmethionine (SAM). SAM is required to restore the DNA methylation after cell division, as well as for polyamines recycling. Increased cell proliferation leads to increased polyamines processing, competing with DNA methylation (Brooks, 2012). Interestingly, the key enzymes that are involved in the polyamine synthesis are encoded in the X chromosome and an elevated level of polyamines is found in many autoimmune diseases including RA (Furumitsu et al., 1993). Thus, high level of polyamines is associated not only with DNA hypomethylation, but also with an increased risk of RA development in women.

Nakano et al. (2013a) have shown that DNA hypomethylation increases expression of numerous genes: growth factors/receptors, extracellular matrix proteins, adhesion molecules, and matrix degrading enzymes, etc. Expression of the DNA methyltransferase-1 (DNMT1) is reduced on protein level in RASFs comparing to the osteoarthritis synovial fibroblasts (OASF), particularly when stimulated with cytokines or growth factors. However, DNMT1A transcripts levels are similar in both of these cell types. Additionally, transcription of DNMT1 could be reduced by stimulation of IL-1 (Nakano et al., 2013a).

A number of differentially methylated loci were reported in RASFs. For example, hypomethylation of *CXCL12*

(chemokine C-X-C motif ligand 12) promoter is associated with the gene's upregulation and accumulation of the protein in the joints of RA patients that contributes to chronic inflammation. TBX5 regulates expression of proinflammatory cytokines and chemokines in SF including CXCL12 chemokine, which is the downstream effector of the same pathway. Hypomethylation of *TBX5* (T-box transcription factor 5) increases its own expression as well as *CXCL12* expression in RASFs. Treatment of SF cell culture with 5-aza-2′-deoxycitidine (5-aza-dC) to achieve DNA demethylation induces hypomethylation of promoter region and subsequent re-activation of CXCL12 expression (Karouzakis et al., 2014).

Comprehensive analysis of DNA methylation in RASFs by Nakano et al. (2013b) identified 1 859 differently methylated loci. Some of the hypomethylated loci that were critically important for RA pathogenesis were located in the genes *CHI3L1*, *CASP1*, *STAT3*, *MAP3K5*, *MEFV*, and *WISP3*. Conversely, genes *TGFBR2* and *FOXO1* were hypermethylated. As shown by analyzing regulatory pathways, the aberrantly methylated genes were involved in cell migration, adhesion, transendothelial penetration and interactions in the extracellular matrix (Nakano et al., 2013b).

Analyzed patterns of DNA methylation suggest its aberrations do not occur randomly, but are specifically related to regulatory pathways involved in RA pathogenesis. Interestingly, RASFs DNA methylation patterns are altering during disease development and progression from early to chronic stage (Whitaker et al., 2013).

A recent study reported genome differently methylated sites that are localized in CpG-islands regions in promoters of RA patients with different clinical symptoms and age of manifestation (Karouzakis et al., 2018). Specifically methylated CpG-islands were found on every stage of the disease. Significant hypermethylation of CpG-islands was revealed in the promoters of peptidase M20 containing domain-1 gene (*PM20D1*), *SHROOM1* and engrailed-1 homeobox protein (*EN1*) at very early RASFs compared to normal SFs. *SHROOM1* gene is involved in the development of nervous tissue and the rearrangement of microtubules during cell division. The chondrocytes of knee joints in early RA and transient arthritis patients significantly differ in *SHROOM1* methylation making it a valuable biomarker for early diagnostics of the disease (Bonin et al., 2016).

Another set of genes with hypermethylated promoters has been identified in patients with chronic disease: microfibrillar-associated protein 2 (*MFAP2*), discoidin domain receptor (*DDR1*) tyrosine kinase and the major histocompatibility complex *HLA-C*. Several identified CpG-islands were specifically hypermethylated in the SFs of very early and/or chronic RA. The MFAP2 binds TGFβ and the members of bone morphogenetic protein (BMP) family, thus regulates release and activation of these factors involved in the development of arthritis (Weinbaum et al., 2008). Tyrosine kinase DDR1 binds collagens and can regulate various cellular processes such as cell migration, invasion, and proliferation (Juskaite et al., 2017). This indicates promising targets for further functional experiments that could explain the phenotypic changes in RASFs and their invasive behavior.

To identify novel therapeutic targets for RA treatment, the analysis of the methylome was recently suggested along with other data on RASFs. Whitaker et al. (2015) have combined findings from genome-wide association studies and analysis of differential gene expression and DNA methylation analysis in RASFs and OASFs. As a result, a number of genes were chosen as prominent candidates for further investigation relevant for RA pathogenesis: *ELMO1*, *LBH*, and *PTPN11*, which are directly involved in the pathogenesis of RA and may be used as therapeutic targets.

ELMO1 encodes a protein involved in cytoskeleton reorganization, which is crucial for phagocytosis of apoptotic cells and cell motility. *ELMO1* promoter is hypermethylated in RASFs, and its knockdown suppresses the RASFs migration and invasion by reducing the activation of RAC1 GTPase (Whitaker et al., 2015). These results demonstrate how the integration of datasets from genome-wide methylation and gene expression analyses allows identifying proteins with previously unknown critical role in RA development. Such a complex "omics" approach can be extended from studying only promoter regions of the genes to enhancers, silencers and other regulatory sequences with almost unknown effect of methylation. The application of such approaches led to discovery of novel differentially methylated loci in the RASFs.

Previously it was thought a protein encoded by *LBH* contributes only in embryonic development. However, *LBH* promotor was found to be hypomethylated in RASFs as well as its enhancer. The gene knockdown affected the transcriptome including pathways that control cell growth and proliferation in RASFs (Ekwall et al., 2015). Interestingly, its enhancer domain contains single nucleotide polymorphism (SNP) rs906868 associated with RA. The combination of the SNP genotype and methylation affects the activity of enhancer and, consequently, expression of *LBH* (Hammaker et al., 2016).

PTPN11 encodes the tyrosine phosphatase SHP2 and is upregulated in RASFs (Stanford et al., 2013). Analysis of the *PTPN11* enhancer in RASFs revealed hypermethylation, which increased the sensitivity of cells to glucocorticoids and their aggressiveness. This not only explained the mechanism of action of PTPN11 in RASFs but also demonstrated SHP2 as a potential therapeutic target in RA. The results were confirmed on mouse models of arthritis (Whitaker et al., 2016).

Summing up, analysis of the methylome in RASFs contributes to understanding of RA pathogenesis. Not only local cytokine environment but other factors can potentially affect DNA methylation pattern and participate in establishing a stable phenotype of RASFs. Identification of these environmental factors could shed light on the predisposition to RA development and progression. In addition, studies of RASFs used the novel "omics" technologies – include not only methylome but also the other types of epigenetic markers – will help discover novel molecular factors in the RA pathogenesis and determine potential therapeutic targets. Considering the role of differentially methylated genes from the pathway perspective, several cascades were reported to be usually disturbed in very

early RASFs and chronic RASFs in comparison with normal SFs but exhibit no changes in the SFs in transient arthritis. These regulatory pathways include cadherin, integrin, WNT signaling of cell adhesion, components of the actin cytoskeleton and antigen presentation.

HISTONE MODIFICATIONS IN IMMUNE CELLS IN RA

Histone modifications are important epigenetic marks that affect gene expression and determine phenotype of cells. Histone proteins are bound to DNA and regulate the accessibility of gene promoters for transcription factors. The basic functional unit of chromatin is the nucleosome. It contains 147 base pairs of DNA, which are wrapped around a histone octamer that consists of two copies each of histones H2A, H2B, H3, and H4. The epigenetic landscape of histones can be modified by numerous mechanisms including acetylation, methylation, citrullination, phosphorylation, ubiquitinylation, and sumoylation (Tessarz and Kouzarides, 2014). Some histone marks are associated with an open chromatin structure. This makes chromatin accessible to transcription factors and can significantly increase gene expression. These include histones lysine residues acetylation (H3K9, H3K14, H4K5, and H4K16) and methylation (H2BK5, H3K4, H3K36, and H3K79); phosphorylation of histone H3 threonine 3 (H3T3) and serines (H3S10 and H3S28), and as well as histone 4 serine 1 (H4S1) and H2BK120 ubiquitinylation. On the other hand, repressive histone marks that correlate with heterochromatin state and gene repression include methylation of H3K9, H3K27, and H4K20; ubiquitination of H2AK119; and sumoylation of H2AK126, H2BK6, and H2BK7 (Araki and Mimura, 2016). Most studies have focused on the acetylation and methylation of histones, although citrullinated histones in RASFs has also been reported (Wang et al., 2016).

Different types of histone modifications are colocalized throughout the genome in order to stabilize active or repressed chromatin states. This complicates analysis of RA-specific alterations of histone marks. The expression of histone-modifying enzymes (e.g., histone deacetylases and HDAC) could be studied instead. The interest to the expression of HDAC in RA arose after reports on its inhibitors, which were considered as novel and promising therapeutic strategy of inflammatory diseases. One of the studies revealed activation of HDAC in PBMCs of RA patients compared to healthy controls (Toussirot et al., 2013). Controversial data were also obtained for HDAC activity in the synovial tissues of RA patients. Presumably, HDAC activity depends strongly to the disease progression and therapy. This may explain discrepancies in the measurements among different cohorts of patients whereas the changes that were identified in larger and clearly defined groups got lost.

HISTONE MODIFICATIONS IN RASFs

Similarly as in immune cells, direct studies of histone modifications in synovial fibroblasts are quite rare. Huber et al. (2007) demonstrated an overall increase of acetylation associated with reduced HDAC activity as a result of decreased *HDAC1* and *HDAC2* gene expression in the synovial tissue of RA patients. Suppression of HDAC1 and HDAC2 suggests the balance between histone acetylases (HAT) and HDAC activity shifts toward hyperacetylation in RA synovial tissues. However, another study demonstrated that HDAC activity and *HDAC1* expression were upregulated in RA synovial tissue (Kawabata et al., 2010). Not all members of the HDAC family have pro-inflammatory effect. For example, HDAC5 demonstrates anti-inflammatory functions in SFs. That indicates the applicability of specific rather than general HDAC inhibitors for the RA treatment (Angiolilli et al., 2016). Recently, suppression of HDAC3 expression was found to be as effective in suppressing pro-inflammatory factors in the RASFs as general inhibition of HDAC, which makes HDAC3 a promising candidate for targeted therapy (Angiolilli et al., 2017).

Abnormal Modifications of Histones Are Involved in the Activation of RASFs

H3K27 specific histone methyltransferase (HMT) – the enhancer of zeste homolog 2 (EZH2) – is highly expressed in RASFs as a result of TNFα induction of nuclear factor kappa B (NF-KB) and mitogen-activated protein kinase (MAPK) pathways. The expression of the secreted frizzled-related protein 1 (SFRP1) – a EZH2 target gene – is increased under the active histone marks (H3K4me3 and H3K27me3) in its promoter that leads to the Wnt (wingless-type MMTV integration site signaling)-pathway inhibition in RASFs (Trenkmann et al., 2011).

Transcription factor T-box transcription factor 5 (TBX5) is overexpressed in RASFs and active histone marks – including H4K4me3 and histone acetylation – are widely represented in its promoter. The overexpression of TBX5 affects expression of 790 genes including *IL-8*, chemokine C-X-C motif ligand 12 (*CXCL12*) and chemokine C-C motif ligand 20 (*CCL20*) confirming its role as an inductor and regulator of chemokines important in RA development (Karouzakis et al., 2014).

Ai et al. (2018) performed the comprehensive study, describing histone modification, open chromatin, RNA expression, and genome-wide DNA methylation in synovial fibroblasts in RA patients. To determine complex multidimensional interactions in the epigenetic regulation of RA, an integrative analysis was performed using a new method for detecting genomic regions with similar profiles. In addition to the known pathological pathways that are activated in RA, the authors found novel pathway activated in RA that was previously known to be associated with Huntington's disease (Ai et al., 2018).

In the other paper (Webster et al., 2018) epigenetic changes in RASF in 79 pairs of discordant on RA monozygotic twins were revealed. An epigenetic signature has been shown to indicate the association of stress response pathways and RA pathogenesis. It is noteworthy that potential epigenetic disruption of multiple RUNX3 transcription factor binding sites was proposed to be associated with disease development.

Histones Modifications Affect Matrix Metalloproteinase Genes Regulation

Araki et al. (2015) reported significantly higher levels of the activating trimethylation mark H3K4me3 in promoters of *MMP-1*, *MMP-3*, *MMP-9*, and *MMP-13* along with reduced of the repressive modification H3K27me3 in promoters of *MMP-1* and *MMP-9* in RASFs. Furthermore, the elevated level of histone H4 acetylation was associated with upregulation of *MMP-1* (Maciejewska-Rodrigues et al., 2010).

Tryptophan-aspartate (WD) repeat-containing protein 5 (*WDR5*) is a major subunit of the proteins bound with SET1 (COMPASS) or COMPASS-like complexes that catalyze H3K4 methylation. *WDR5* knockdown reduces the level of H3K4me3 marks as well as the abundance of MMP-1, MMP-3, MMP-9, and MMP-13 in RASFs. IL-6 and soluble IL-6 receptor α (sIL-6Rα) induce the expression of *MMP-1*, *MMP-3*, and *MMP-13* but not of *MMP-9*: It has been shown that IL-6-induced signal transducer and activator of transcription 3 (STAT3) binds to *MMP-1*, *MMP-3*, and *MMP-13* promoters but not with that of *MMP-9*. High expression of IL-6 was associated with high level of histone H3 acetylation (H3ac) of the IL-6 promoter in RASFs (Wada et al., 2014).

microRNAs AS AN EPIGENETIC FACTOR ASSOCIATED WITH THE DEVELOPMENT OF RA

MicroRNAs (miRNAs) are small non-coding RNAs of 17–25 nt that regulate gene expression by either repressing the translation or causing degradation of multiple target mRNAs (Fabian et al., 2010). miRNAs play an important role in many biological processes including the development of the immune system and the subsequent regulation of immunity both innate and acquired (Chen et al., 2016). Currently more than 100 miRNAs have been identified that could potentially affect the molecular pathways in immune cells development and their functions regulation (Baulina et al., 2016).

ABERRATIONS OF microRNA EXPRESSION IN RA

Aberrant miRNA regulation occurs in various cells and tissues in RA (Tavasolian et al., 2018). The role of miRNA in the inflammatory process includes both control of cytokine production and protection of cartilage tissue by regulating catabolic activity, proliferation and resistance to apoptosis.

The increased production of IL-17 by T helper cells (Th17) in the synovial fluid may suppress miRNA-23b and so enhance expression of (TGF)-β-activated kinase 1/MAP3K7-binding protein and Iκβ kinase α contributing to inflammation. miRNA-21 promotes the differentiation of Th2 and Treg cells and is also associated with the regulation of Treg apoptosis (Salehi et al., 2015). Pro-inflammatory phenotype of Treg is determined by an aberrant miRNA-146a expression; thus its reduction in RA patients inversely correlated with disease activity and expression of its direct target gene *STAT1* (Zhou et al., 2015).

A number of studies have described miRNAs that modulate inflammatory or catabolic functions of RASFs thereby contributing to the development of the aggressive phenotype in RA (Vicente et al., 2016).

miRNA-146 and miRNA-155 were first described as abnormally expressed in synovial tissue, RASFs and synovial fluid of patients with RA and still remain the best characterized candidates. miRNA-155 is upregulated in RASFs, where – in addition to the pro-inflammatory activity – it regulates destructive processes due to the matrix metalloproteinases MMP-1 and MMP-3 expression repressing (Long et al., 2013). Similarly, miRNA-146a express increasingly in RASFs, despite a known role as a negative regulator of inflammation in immune cells (Vicente et al., 2016).

Increased expression in RASFs has also been determined for a number of other transcripts: miRNA-203 (Stanczyk et al., 2011), miRNA-221 (Yang and Yang, 2015), miRNA-663 (Miao et al., 2015a), miRNA-222 and miRNA-323-3p (Pandis et al., 2012). Besides their role in immune processes, these microRNAs are involved in oncogenesis by regulating cell invasiveness and migration in different types of tumors.

Earlier reports show miRNA-203 overexpression associated with hypomethylation of MMP1 and IL-6 gene promoters induces these proteins production. Regulation by DNA methylation was also observed in mir-203. Interestingly, this miRNA high-level expression in RASFs is observed with its own promoter hypomethylation (Stanczyk et al., 2011).

Suppression of miRNA-221 inhibits the production of pro-inflammatory cytokines by fibroblasts, causes cell apoptosis and reduces their migration and invasion of RASFs (Yang and Yang, 2015).

miRNA-663 regulates proliferation of RASFs and production of IL-6 via inhibition of the tumor suppressor *APC* thereby affecting the Wnt signaling pathway (Miao et al., 2015a).

Overexpression of miRNA-124a in RASFs may disrupt the cell cycle and lead to inhibition of cell proliferation through repression of its target genes *CDK-2* and *MCP-1*. Additionally, it was shown that miRNA-124a expression is regulated by methylation of the gene from which it is transcribed. Demethylation of the miRNA-124a gene by 5-aza-dC reduces RASF proliferation and expression of TNF-α (Zhou et al., 2016). Regulation by DNA methylation was also observed in mir-203. Its high expression in RASFs is associated with the promoter hypomethylation (Stanczyk et al., 2011).

The similar effect on RASFs proliferation has miRNA-34a* (Niederer et al., 2012). It regulates genes of the apoptosis inhibitor XIAP. Furthermore, direct correlation of miRNA-21, miRNA-25, and miRNA-124a from peripheral blood cells with estradiol level in plasma was described in women with RA. The effect of estradiol on miRNA-124a is particularly interesting, as this miRNA has an effect on synovial proliferation (Singh et al., 2013). Decreased expression of both miRNA-124a and miRNA-34a* as well as the others: miRNA-152, miRNA-375, and miRNA-22 in RASFs contributes to RA development (Vicente et al., 2016).

miRNA-152 and miRNA-375 directly target DNMT1. Thus, the increase of their expression leads to activation of Wnt-signaling pathway (Miao et al., 2015b).

Cyr61 expression is increased by miRNA-22 suppression, affecting various genes involved in different processes: angiogenesis, inflammation, matrix structure reorganization, IL-6 production with subsequent differentiation of Th17 and synovial hyperplasia. Earlier data show that the reduced expression of miRNA-22 is caused by p53 mutation that is frequent in RASFs (Lin et al., 2014).

The loss of miRNA-10a-5p expression in RASFs upregulates target gene *TBX5* thereby promoting the production of TLR3, MMP-13 and a number of pro-inflammatory cytokines (Hussain et al., 2018). Suppressed expression of miRNA-10a may cause the activation of NF-κB and enhancing the release of pro-inflammatory cytokines TNF-α, IL-1β, IL-6, and IL-8, chemokine MCP-1 and matrix metalloproteinases MMP-1 and MMP-13 (Mu et al., 2016).

Deregulations of expression affect entire miRNA clusters. However, the expression of individual miRNAs of a cluster may be altered differentially. For example, miRNA-18a is upregulated in the miRNA-17-92 cluster, which plays an important role in the regulation of apoptosis in the RASFs while miRNA-19a/b, miRNA-20a and miRNA-30a-3p are downregulated. Such an effect may be caused by the different target genes affection and different signal pathways regulation (Vicente et al., 2016).

The RA associated changes of miRNA expression were also confirmed in other cells of the joint capsule. For example, miRNA-323-3p is involved in the regulation of Wnt and cadherin signaling pathways, and its overexpression in chondrocytes causes degradation of the cartilaginous matrix and promotes bone erosion. Similar effect was demonstrated for miRNA-140, which contributes to reduction of joint destruction via suppression of *ADAMSTS5*. Upregulation of some miRNAs (miRNA-30a, miRNA-204, miRNA-211, miRNA-320, and miRNA-335) is associated with suppression of osteoblast differentiation via RUNX2 regulation (Moran-Moguel et al., 2018).

CIRCULATING miRNAs AS POTENTIAL MARKERS OF RA DEVELOPMENT

In addition to aberrant miRNAs expression in the joint area, they were also detected in blood: plasma, serum, and various blood cells (**Table 1**). In one of the first studies of miRNA expression in RA a significant difference between the 26 microRNAs expression patterns was demonstrated in patients compared to healthy donors. Three of those, namely miRNA-24, miRNA-26a, and miRNA-125a-5p have been proposed as a potential diagnostic panel with sensitivity and specificity of 78.4 and 92.3%, respectively (Murata et al., 2013).

Nevertheless, estimation of two other miRNAs expression level is considered more prominent at the moment. miRNA-146a and miRNA-155 are overexpressed both in whole blood and PBMCs of RA patients (Mookherjee and El-Gabalawy, 2013).

Many miRNAs that are aberrantly expressed in synovial tissues in RA are also deregulated in peripheral blood. However, these changes are not always similar. The concentrations of miRNA-16, miRNA-132, and miRNA-223 were significantly lower in the synovial fluid than in plasma of patients with RA, and no correlation between them was found (Murata et al., 2010). Elevated level of miRNA-125b was observed in both serum and synovial tissues of RA patients (Zhang B. et al., 2017).

While miRNA-146a is overexpressed in synovial tissue of RA patients comparing with healthy donors, a reduction of circulating miR-146a was reported in the peripheral blood (Wang et al., 2012). Besides, miRNA-146a from PBMCs does not demonstrate any correlation with disease activity. That is in contrast with the same miRNA quantified in serum. Thus, the miRNAs obtained from different types of samples may have different prognostic value or not have it at all (Ayeldeen et al., 2018).

Changes in the expression of miRNAs may be associated with the RA therapy (**Table 1**). Specifically, miRNA-146a, miRNA-155, and miRNA-16 levels were decreased in serum at the early RA stages after treatment with disease-modifying antirheumatic drugs (DMARDs) (Filková et al., 2014). In contrast the quantity of miRNA-16-5p, miRNA-23-3p, miRNA125b-5p, miRNA-126-3p, miRNA-146a-5p, and miRNA-223-3p in plasma was significantly elevated after combined anti-TNF-α/DMARD therapy (Castro-Villegas et al., 2015).

miRNAs Genes Polymorphism Associated With RA

miRNAs not only regulate gene expression in RA but also are themselves subject to regulation by various factors. As in protein-coding genes, polymorphism of miRNA genes or their target genes may be associated with predisposition to RA. The number of studies has reported a significant role of some nucleotide polymorphisms located in the genes of miRNA-146a and miRNA-499 precursors (Ayeldeen et al., 2018). For example, the importance of polymorphism rs3746444 in miRNA-499 was confirmed for the RA development among the patients of Caucasian race. Genotypes TC and CC and allele C at this SNP in miRNA-499 were characterized as independent risk factors of joint erosion in RA patients. The frequency of GG genotype of rs2910164 in miRNA-146a is significantly higher in patients with RA compared to healthy donors (Ayeldeen et al., 2018). Another study that included 200 RA patients and 120 healthy donors has demonstrated the correlation of SNP rs22928323 in miRNA-149 with RA development but no association with further clinical characteristics (Xiao et al., 2015).

DNA Methylation of miRNA Genes

Similarly to protein-coding genes, DNA methylation is an important mechanism of miRNA regulation. It was demonstrated

TABLE 1 | Potential diagnostic and prognostic markers of RA among miRNAs.

Clinical potential	miRNA	Type of sample	Expression	References
Diagnostic marker	miRNA-16	Synovial fluid	Increased, comparing to another autoimmune diseases	Mi et al., 2013
	miRNA -223	Synovial fluid	Increased, comparing to another autoimmune diseases	Shibuya et al., 2013
	miRNA-24, miRNA-30a-5p, miRNA-125a-5p	Plasma	Increased in ACPA-positive and ACPA-negative donors before clinical symptoms appear	Murata et al., 2013
	miRNA-22, miRNA-382, miRNA-486-3p	Whole blood	Increased in ACPA-positive donors before clinical symptoms appear	Ouboussad et al., 2017
	miRNA-10a	Whole blood	Decreased in RA patients	Hong et al., 2018
	miRNA-155	Serum	Increased in RA patients	Abdul-Maksoud et al., 2017
	miRNA-210	Serum	Decreased in RA patients	Abdul-Maksoud et al., 2017
	miRNA-26à	Serum/plasma	Increased in RA patients	Murata et al., 2013
Disease progression	miRNA-146a, miRNA-16, miRNA-150	PBMCs	Correlates with disease progression	Filkova et al., 2012
	miRNA-16, miRNA-146a, miRNA-155	Whole blood	Lower level on early stages, than in advanced disease	Filková et al., 2014
	miRNA-146a	Synovial fluid, blood	Correlates with disease progression	Chung et al., 2016
	miRNA-16	Plasma	Correlates with disease progression	Murata et al., 2010
Response to treatment:				
Adalimumab	miR-22	Whole blood	Better response to therapy when decreased	Krintel et al., 2016
	miR-886-3p	Whole blood	Better response to therapy when increased	Krintel et al., 2016
Rituximab	miRNA-125b	Serum	Better response to therapy when increased	Duroux-Richard et al., 2014
Methotrexate	miR-10a	Whole blood	Better response to therapy when decreased	Hong et al., 2018
Anti-TNF-α therapy	miR-5196	Serum	Better response to therapy when increased	Ciechomska et al., 2017
	miR−146a−5p	Serum	Better response to therapy when increased	Bogunia-Kubik et al., 2016
Anti-TNF/DMARD therapy	miR-16-5p, miR-23-3p, miR125b-5p, miR-126-3p, miR-146a-5p, miR-223-3p	Serum	Prognostic panel of therapy efficiency	Castro-Villegas et al., 2015
DMARD therapy and glucocorticoids	miR-16, miR-223	Serum	Better response to therapy when increased	Filková et al., 2014

that miRNA-124a and miRNA-203 are controlled by DNA methylation of respective genes in RASFs. *In vitro* 5-azacitidine treatment leads to DNA demethylation and transcriptional re-activation (Stanczyk et al., 2011; Zhou et al., 2016). de la Rica et al. (2013) provided additional evidence of such epigenetic regulation in the whole genome by profiling of the methylome and analysis of miRNA and mRNA expression in RASFs in parallel. Expression of 11 miRNAs was reduced in RA samples comparing to control osteoarthritis samples and was associated with hypermethylated genes thereof. In contrast, four other miRNAs were upregulated upon hypomethylation of CpG sites in vicinity of their genes (de la Rica et al., 2013).

Long Non-coding and Circular RNAs in RA

Several long non-coding RNAs (lncRNAs) were deregulated in RA by miRNAs. The expression of the most characterized lncRNA *HOTAIR* (HOX transcript antisense RNA) was significantly reduced in chondrocytes that were pretreated with lipopolysaccharide in order to suppress the inflammatory process. This transcript is able to suppress mir-138-mediated synthesis of NF-κB, since miRNA-138 is a *HOTAIR* direct target (Zhang H. et al., 2017).

Expression of the other lncRNA *ZFAS1* (zinc finger antisense 1) is increased in RA synovial tissue compared to healthy donors and can enhance migration and invasion of RASFs by directly affecting miRNA-27a (Ye et al., 2018).

lncRNA *GAPLINC* (Gastric Adenocarcinoma Predictive Long Intergenic Non-coding RNA) is overexpressed in RASFs and regulates their proliferation, migration and pro-inflammatory cytokine production by operating as a sponge for miRNA-382-5p, and miRNA-575 (Mo et al., 2018).

A possible effect of circular RNAs (circRNAs) on miRNAs in RA was revealed, too. Presumably, these transcripts

complementarily bind to miRNAs thereby preventing their interaction with target genes (Zheng et al., 2017).

SUMMARY AND FUTURE DIRECTIONS

In the last few decades, many studies have shown that epigenetic mechanisms are involved in the regulation of all biological processes in the body from impregnation to death. These functional mechanisms are involved in genome reorganization, control of gametogenesis and early embryogenesis, and play an important role in cell differentiation. Changes in DNA methylation and posttranslational modifications of histones are key epigenetic events contributing to the reorganization of chromatin into euchromatin, heterochromatin, and regions of nuclear compartmentalization, which allow to regulate gene expression by forcing them to be consistently switched on and off for the normal development of a multicellular organism. Epigenetic changes may form and appear over a long period of time, for example, during the training and organization of memory (Moosavi and Motevalizadeh Ardekani, 2016). Aberrations of epigenetic modifications can cause the development of congenital defects, hereditary diseases and multifactorial diseases including malignant tumors in different periods of life. DNA methylation, histone modifications, expression of proteins that generate or remove epigenetic marks, and ncRNAs affect inflammatory and matrix-degrading pathways and could be changed in RA. Epigenetic mechanisms play an important role in the pathogenesis of the disease.

Unlike genetic lesions, epigenetic alterations are reversible and could be modulated by diet, drugs and other environmental factors. This epigenetic flexibility suggests strategies for prevention and therapy of diseases with confirmed pathogenic role of epigenetic factors (Pashayan et al., 2016).

Therapeutic targeting of epigenetic mechanisms can be a successful approach in the treatment of chronic inflammatory diseases. Significant efforts have already been made to develop drugs that able to restore or alter the epigenetic mechanisms. DNA methyltransferase inhibitors (DNMT), 5-azacitidine (Vidaza) and 5-Aza-2′deoxy-5-azacitidine (Decitabine) are already being used to treat inflammatory conditions in pancreatitis therapy. Two types of HDAC inhibitors (HDACi) are used for treatment: pan-inhibitors with broad spectrum of action and specific inhibitors that target a specific class of HDAC enzyme. To date, the Food and Drug Administration (FDA) has approved four HDACi: Vorinostat, Romidepsin, Panobinostat, and Belinostat. These products have the minimal side effects and are mainly used for the hematological tumors treatment (Samanta et al., 2017).

Moreover epigenetic alterations are a source of diagnostic and prognostic markers. The gradual changes of epigenetic marks that may be caused by environmental factors can result in both the development and progression of pathological conditions. DNA methylation marks are the best characterized, its role is comprehensively studied in cancer development. The existence of loci and genes with differential methylation patterns, varying, respectively, the stage of the disease, was also revealed in patients during RA progression (Ai et al., 2015). However, the plasticity of the epigenome is a complication for the researches. The prognostic markers panels consisted of epigenetically modified genes could differs individually due to the variability of environmental factors affecting patients. Smoking is the most common environmental risk factor in RA development and its severity determination. In addition to directly affecting lung tissues, smoking alters the expression of sirtuins (SIRT), the proteins of the deacetylase families that are involved in modification of histone and non-histone proteins. SIRT maintain the integrity of the genome during the cellular response to stress by using epigenetic mechanisms, and are therefore key molecules in the body's adaptation process. The change of expression of SIRT1 and SIRT6 in RASFs was demonstrated (Engler et al., 2016).

In contrast to genetic aberrations that are persist throughout life, epigenetic changes may vary in different cell populations as well as in the same cell depending on conditions and developmental stage. Since the whole blood cells, T- and B-lymphocyte populations, and SFs exhibit different DNA methylation patterns, the analysis of epigenetic markers in a mixed cell population may hamper in the correct evaluation of the epigenetic panel (Liu et al., 2013; Glossop et al., 2016).

Similarly, miRNA profiles differ across individuals, cell populations and may be affected by concomitant diseases (Huang et al., 2011). Thus, it is important to consider the extracted cell type during the development of RA epigenetic markers panel. Individual patient therapeutic response could also be predicted with the epigenetic panel use. Currently, a large number of drugs for RA treatment is undergoing preclinical and clinical trials, however, the existing data linked the epigenetic changes and response to therapy is limited. Thus, intensive investigations in this field are critically needed. International cooperation will enable the access to larger patient cohorts thereby improving the quality of studies. However, when planning international consortia, it is important to consider the differences in ethnic composition that may be reflected in epigenetic variations (Rawlings-Goss et al., 2014). Some differences that were related to race and nationality of patients were demonstrated even for genetic markers.

Nevertheless, despite of the complications, investigation of epigenetic markers is undoubtedly a great achievement of molecular biology and molecular medicine. Epigenetic changes are the earliest factors that are associated with the development of the disease before its clinical manifestation. They can be used for prevention and monitoring of patient condition and also are the earliest to reflect the effect of the drug at the cellular level, before any systemic response of organism manifested as certain symptoms. Moreover, some of epigenetic markers, such

as changes in circulating miRNAs level in plasma, may be more accessible for evaluation than the other molecular markers.

AUTHOR CONTRIBUTIONS

MN, DZ, AL, and AZ contributed conception the review. MN, IB, DM, EK, EA, MB, and AD wrote sections of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

This research was supported by Ministry of Education and Science of the Russian Federation (Agreement 14.605.21.0003, unique project ID RFMEFI60518X0003).

REFERENCES

Abdul-Maksoud, R., Sediq, A., Kattaia, A., Elsayed, W., Ezzeldin, N., Abdel Galil, S., et al. (2017). Serum miR-210 and miR-155 expression levels as novel biomarkers for rheumatoid arthritis diagnosis. *Br. J. Biomed. Sci.* 74, 209–213. doi: 10.1080/09674845.2017.1343545

Ai, R., Laragione, T., Hammaker, D., Boyle, D. L., Wildberg, A., Maeshima, K., et al. (2018). Comprehensive epigenetic landscape of rheumatoid arthritis fibroblast-like synoviocytes. *Nat. Commun.* 9:1921. doi: 10.1038/s41467-018-04310-9

Ai, R., Whitaker, J. W., Boyle, D. L., Tak, P. P., Gerlag, D. M., Wang, W., et al. (2015). DNA methylome signature in synoviocytes from patients with early rheumatoid arthritis compared to synoviocytes from patients with longstanding rheumatoid arthritis. *Arthritis Rheumatol.* 67, 1978–1980. doi: 10.1002/art.39123

Angiolilli, C., Grabiec, A. M., Ferguson, B. S., Ospelt, C., Malvar Fernandez, B., van Es, I. E., et al. (2016). Inflammatory cytokines epigenetically regulate rheumatoid arthritis fibroblast-like synoviocyte activation by suppressing HDAC5 expression. *Ann. Rheum. Dis.* 75, 430–438. doi: 10.1136/annrheumdis-2014-205635

Angiolilli, C., Kabala, P. A., Grabiec, A. M., Van Baarsen, I. M., Ferguson, B. S., García, S., et al. (2017). Histone deacetylase 3 regulates the inflammatory gene expression programme of rheumatoid arthritis fibroblast-like synoviocytes. *Ann. Rheum. Dis.* 76, 277–285. doi: 10.1136/annrheumdis-2015-209064

Araki, Y., and Mimura, T. (2016). The mechanisms underlying chronic inflammation in rheumatoid arthritis from the perspective of the epigenetic landscape. *J. Immunol. Res.* 2016, 1–10. doi: 10.1155/2016/6290682

Araki, Y., Wada, T. T., Aizaki, Y., Sato, K., Yokota, K., Fujimoto, K., et al. (2015). Histone methylation and STAT3 differentially regulate IL-6-induced MMP gene activation in rheumatoid arthritis synovial fibroblasts. *Arthritis Rheumatol.* 68, 1111–1123. doi: 10.1002/art.39563

Ayeldeen, G., Nassar, Y., Ahmed, H., Shaker, O., and Gheita, T. (2018). Possible use of miRNAs-146a and -499 expression and their polymorphisms as diagnostic markers for rheumatoid arthritis. *Mol. Cell. Biochem.* 449, 145–156. doi: 10.1007/s11010-018-3351-7

Baulina, N. M., Kulakova, O. G., and Favorova, O. O. (2016). MicroRNAs: the role in autoimmune inflammation. *Acta Naturae* 8, 21–33.

Bogunia-Kubik, K., Wysoczańska, B., Piątek, D., Iwaszko, M., Ciechomska, M., and Świerkot, J. (2016). Significance of polymorphism and expression of miR-146a and NFkB1 genetic variants in patients with rheumatoid arthritis. *Arch. Immunol. Ther. Exp.* 64, 131–136. doi: 10.1007/s00005-016-0443-5

Bonin, C. A., Lewallen, E. A., Baheti, S., Bradley, E. W., Stuart, M. J., Berry, D. J., et al. (2016). Identification of differentially methylated regions in new genes associated with knee osteoarthritis. *Gene* 576, 312–318. doi: 10.1016/j.gene.2015.10.037

Brooks, W. H. (2012). Autoimmune diseases and polyamines. *Clin. Rev. Allergy Immunol.* 42, 58–70. doi: 10.1007/s12016-011-8290-y

Castro-Villegas, C., Pérez-Sánchez, C., Escudero, A., Filipescu, I., Verdu, M., Ruiz-Limón, P., et al. (2015). Circulating miRNAs as potential biomarkers of therapy effectiveness in rheumatoid arthritis patients treated with anti-TNFα. *Arthritis Res. Ther.* 17:49. doi: 10.1186/s13075-015-0555-z

Chen, H., Fu, L., Li, S., Ma, C., Zhang, J., and Cong, B. (2011). Hypomethylation of proximal CpG motif of interleukin-10 promoter regulates its expression in human rheumatoid arthritis. *Acta Pharmacol. Sin.* 32, 1373–1380. doi: 10.1038/aps.2011.98

Chen, J.-Q., Papp, G., Szodoray, P., and Zeher, M. (2016). The role of microRNAs in the pathogenesis of autoimmune diseases. *Autoimmun. Rev.* 15, 1171–1180. doi: 10.1016/j.autrev.2016.09.003

Chung, I.-M., Ketharnathan, S., Thiruvengadam, M., and Rajakumar, G. (2016). Rheumatoid arthritis: the stride from research to clinical practice. *Int. J. Mol. Sci.* 17:900. doi: 10.3390/ijms17060900

Ciechomska, M., Zarecki, P., Merdas, M., Swierkot, J., Morgiel, E., Wiland, P., et al. (2017). The role of microRNA-5196 in the pathogenesis of systemic sclerosis. *Eur. J. Clin. Invest.* 47, 555–564. doi: 10.1111/eci.12776

Cribbs, A., Feldmann, M., and Oppermann, U. (2015). Towards an understanding of the role of DNA methylation in rheumatoid arthritis: therapeutic and diagnostic implications. *Ther. Adv. Musculoskelet. Dis.* 7, 206–219. doi: 10.1177/1759720X15598307

Cribbs, A. P., Kennedy, A., Penn, H., Amjadi, P., Green, P., Read, J. E., et al. (2015). Methotrexate restores regulatory T cell function through demethylation of the FoxP3 upstream enhancer in patients with rheumatoid arthritis. *Arthritis Rheumatol.* 67, 1182–1192. doi: 10.1002/art.39031

Cribbs, A. P., Kennedy, A., Penn, H., Read, J. E., Amjadi, P., Green, P., et al. (2014). Treg cell function in rheumatoid arthritis is compromised by CTLA-4 promoter methylation resulting in a failure to activate the indoleamine 2,3-dioxygenase pathway. *Arthritis Rheumatol.* 66, 2344–2354. doi: 10.1002/art.38715

de Andres, M. C., Perez-Pampin, E., Calaza, M., Santaclara, F. J., Ortea, I., Gomez-Reino, J. J., et al. (2015). Assessment of global DNA methylation in peripheral blood cell subpopulations of early rheumatoid arthritis before and after methotrexate. *Arthritis Res. Ther.* 17:233. doi: 10.1186/s13075-015-0748-5

de la Rica, L., Urquiza, J. M., Gómez-Cabrero, D., Islam, A. B., López-Bigas, N., Tegnér, J., et al. (2013). Identification of novel markers in rheumatoid arthritis through integrated analysis of DNA methylation and microRNA expression. *J. Autoimmun.* 41, 6–16. doi: 10.1016/j.jaut.2012.12.005

Duroux-Richard, I., Pers, Y. M., Fabre, S., Ammari, M., Baeten, D., Cartron, G., et al. (2014). Circulating miRNA-125b is a potential biomarker predicting response to rituximab in rheumatoid arthritis. *Mediat. Inflamm.* 2014:342524. doi: 10.1155/2014/342524

Eden, S., and Cedar, H. (1994). Role of DNA methylation in the regulation of transcription. *Curr. Opin. Genet. Dev.* 4, 255–259. doi: 10.1016/S0959-437X(05)80052-8

Ekwall, A.-K. H., Whitaker, J. W., Hammaker, D., Bugbee, W. D., Wang, W., and Firestein, G. S. (2015). The rheumatoid arthritis risk gene LBH regulates growth in fibroblast-like synoviocytes. *Arthritis Rheumatol.* 67, 1193–1202. doi: 10.1002/art.39060

Engler, A., Tange, C., Frank-Bertoncelj, M., Gay, R. E., Gay, S., and Ospelt, C. (2016). Regulation and function of SIRT1 in rheumatoid arthritis synovial fibroblasts. *J. Mol. Med.* 94, 173–182. doi: 10.1007/s00109-015-1332-9

Fabian, M. R., Sonenberg, N., and Filipowicz, W. (2010). Regulation of mRNA translation and stability by microRNAs. *Annu. Rev. Biochem.* 79, 351–379. doi: 10.1146/annurev-biochem-060308-103103

Ferucci, E. D., Templin, D. W., and Lanier, A. P. (2005). Rheumatoid arthritis in american indians and alaska natives: a review of the literature. *Semin. Arthritis Rheum.* 34, 662–667. doi: 10.1016/j.semarthrit.2004.08.003

Filková, M., Aradi, B., Šenolt, L., Ospelt, C., Vettori, S., Mann, H., et al. (2014). Association of circulating miR-223 and miR-16 with disease activity in patients with early rheumatoid arthritis. *Ann. Rheum. Dis.* 73, 1898–1904. doi: 10.1136/annrheumdis-2012-202815

Filkova, M., Ospelt, C., Stanczyk, J., Vettori, S., Senolt, L., Frank, M., et al. (2012). MiRs in RA: possible biomarkers and therapeutic targets. *Arthritis Res. Ther.* 14:14. doi: 10.1186/ar3615

Firenstein, G. S., Budd, R. C., and Gabriel, S. E. (2013). *Kelley's Textbook of Rheumatology*. Philadelphia, PA: Elsevier, 117–431. doi: 10.1016/B978-1-4377-1738-9.00124-9

Furumitsu, Y., Yukioka, K., Kojima, A., Yukioka, M., Shichikawa, K., Ochi, T., et al. (1993). Levels of urinary polyamines in patients with rheumatoid arthritis. *J. Rheumatol.* 20, 1661–1665.

Gabriel, S. E. (2001). The epidemiology of rheumatoid arthritis. *Rheum. Dis. Clin. N. Am.* 27, 269–281. doi: 10.1016/S0889-857X(05)70201-5

Glant, T. T., Mikecz, K., and Rauch, T. A. (2014). Epigenetics in the pathogenesis of rheumatoid arthritis. *BMC Med.* 12:35. doi: 10.1186/1741-7015-12-35

Glossop, J. R., Emes, R. D., Nixon, N. B., Packham, J. C., Fryer, A. A., Mattey, D. L., et al. (2016). Genome-wide profiling in treatment-naive early rheumatoid arthritis reveals DNA methylome changes in T and B lymphocytes. *Epigenomics* 8, 209–224. doi: 10.2217/epi.15.103

Golbabapour, S., Abdulla, M. A., and Hajrezaei, M. (2011). A concise review on epigenetic regulation: insight into molecular mechanisms. *Int. J. Mol. Sci.* 12, 8661–8694. doi: 10.3390/ijms12128661

Hammaker, D., Whitaker, J. W., Maeshima, K., Boyle, D. L., Ekwall, A.-K. H., Wang, W., et al. (2016). LBH gene transcription regulation by the interplay of an enhancer risk allele and DNA methylation in rheumatoid arthritis. *Arthritis Rheumatol.* 68, 2637–2645. doi: 10.1002/art.39746

Hardy, R. S., Hülso, C., Liu, Y., Gasparini, S. J., Fong-Yee, C., Tu, J., et al. (2013). Characterisation of fibroblast-like synoviocytes from a murine model of joint inflammation. *Arthritis Res. Ther.* 15:R24. doi: 10.1186/ar4158

Hong, H., Yang, H., and Xia, Y. (2018). Circulating miR-10a as predictor of therapy response in rheumatoid arthritis patients treated with methotrexate. *Curr. Pharm. Biotechnol.* 19, 79–86. doi: 10.2174/1389201019666180417155140

Huang, R. S., Gamazon, E. R., Ziliak, D., Wen, Y., Im, H. K., Zhang, W., et al. (2011). Population differences in microRNA expression and biological implications. *RNA Biol.* 8, 692–701. doi: 10.4161/rna.8.4.16029

Huber, L. C., Brock, M., Hemmatazad, H., Giger, O. T., Moritz, F., Trenkmann, M., et al. (2007). Histone deacetylase/acetylase activity in total synovial tissue derived from rheumatoid arthritis and osteoarthritis patients. *Arthritis Rheum.* 56, 1087–1093. doi: 10.1002/art.22512

Hussain, N., Zhu, W., Jiang, C., Xu, J., Wu, X., Geng, M., et al. (2018). Down-regulation of miR-10a-5p in synoviocytes contributes to TBX5-controlled joint inflammation. *J. Cell. Mol. Med.* 22, 241–250. doi: 10.1111/jcmm.13312

Julià, A., Absher, D., López-Lasanta, M., Palau, N., Pluma, A., Waite Jones, L., et al. (2017). Epigenome-wide association study of rheumatoid arthritis identifies differentially methylated loci in B cells. *Hum. Mol. Genet.* 26, 2803–2811. doi: 10.1093/hmg/ddx177

Juskaite, V., Corcoran, D. S., and Leitinger, B. (2017). Collagen induces activation of DDR1 through lateral dimer association and phosphorylation between dimers. *eLife* 6, 1–27. doi: 10.7554/eLife.25716

Karouzakis, E., Raza, K., Kolling, C., Buckley, C. D., Gay, S., Filer, A., et al. (2018). Analysis of early changes in DNA methylation in synovial fibroblasts of RA patients before diagnosis. *Sci. Rep.* 8:7370. doi: 10.1038/s41598-018-24240-2

Karouzakis, E., Trenkmann, M., Gay, R. E., Michel, B. A., Gay, S., and Neidhart, M. (2014). Epigenome analysis reveals TBX5 as a novel transcription factor involved in the activation of rheumatoid arthritis synovial fibroblasts. *J. Immunol.* 193, 4945–4951. doi: 10.4049/jimmunol.1400066

Kawabata, T., Nishida, K., Takasugi, K., Ogawa, H., Sada, K., Kadota, Y., et al. (2010). Increased activity and expression of histone deacetylase 1 in relation to tumor necrosis factor-alpha in synovial tissue of rheumatoid arthritis. *Arthritis Res. Ther.* 12:R133. doi: 10.1186/ar3071

Kolker, E., Higdon, R., Haynes, W., Welch, D., Broomall, W., Lancet, D., et al. (2012). MOPED: model organism protein expression database. *Nucleic Acids Res.* 40, D1093–D1099. doi: 10.1093/nar/gkr1177

Krintel, S. B., Dehlendorff, C., Hetland, M. L., Hørslev-Petersen, K., Andersen, K. K., Junker, P., et al. (2016). Prediction of treatment response to adalimumab: a double-blind placebo-controlled study of circulating microRNA in patients with early rheumatoid arthritis. *Pharmacogenomics J.* 16, 141–146. doi: 10.1038/tpj.2015.30

Lefèvre, S., Knedla, A., Tennie, C., Kampmann, A., Wunrau, C., Dinser, R., et al. (2009). Synovial fibroblasts spread rheumatoid arthritis to unaffected joints. *Nat. Med.* 15, 1414–1420. doi: 10.1038/nm.2050

Lev Maor, G., Yearim, A., and Ast, G. (2015). The alternative role of DNA methylation in splicing regulation. *Trends Genet.* 31, 274–280. doi: 10.1016/j.tig.2015.03.002

Lin, J., Huo, R., Xiao, L., Zhu, X., Xie, J., Sun, S., et al. (2014). A Novel p53/microRNA-22/Cyr61 axis in synovial cells regulates inflammation in rheumatoid arthritis. *Arthritis Rheumatol.* 66, 49–59. doi: 10.1002/art.38142

Liu, Y., Aryee, M. J., Padyukov, L., Fallin, M. D., Hesselberg, E., Runarsson, A., et al. (2013). Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat. Biotechnol.* 31, 142–147. doi: 10.1038/nbt.2487

Long, L., Yu, P., Liu, Y., Wang, S., Li, R., Shi, J., et al. (2013). Upregulated MicroRNA-155 expression in peripheral blood mononuclear cells and fibroblast-like synoviocytes in rheumatoid arthritis. *Clin. Dev. Immunol.* 2013, 1–10. doi: 10.5167/uzh-62509

Maciejewska-Rodrigues, H., Karouzakis, E., Strietholt, S., Hemmatazad, H., Neidhart, M., Ospelt, C., et al. (2010). Epigenetics and rheumatoid arthritis: the role of SENP1 in the regulation of MMP-1 expression. *J. Autoimmun.* 35, 15–22. doi: 10.1016/j.jaut.2009.12.010

Melorose, J., Perroy, R., and Careas, S. (2015). *Harrison's Principles of Internal Medicine*, 17th Edn. Cambridge: Cambridge University Press, doi: 10.1017/CBO9781107415324.004

Mi, S., Zhang, J., Zhang, W., and Huang, R. S. (2013). Circulating MicroRNAs as biomarkers for inflammatory diseases. *MicroRNA* 2, 64–72. doi: 10.2174/2211536611302010007

Miao, C., Shi, W., Xiong, Y., Yu, H., Zhang, X., Qin, M., et al. (2015a). MicroRNA-663 activates the canonical Wnt signaling through the adenomatous polyposis coli suppression. *Immunol. Lett.* 166, 45–54. doi: 10.1016/j.imlet.2015.05.011

Miao, C., Shi, W., Xiong, Y., Yu, H., Zhang, X., Qin, M., et al. (2015b). miR-375 regulates the canonical Wnt pathway through FZD8 silencing in arthritis synovial fibroblasts. *Immunol. Lett.* 164, 1–10. doi: 10.1016/j.imlet.2015.01.003

Mo, B. Y., Guo, X. H., Yang, M. R., Liu, F., Bi, X., Liu, Y., et al. (2018). Long non-coding RNA gaplinc promotes tumor-like biologic behaviors of fibroblast-like synoviocytes as MicroRNA sponging in rheumatoid arthritis patients. *Front. Immunol.* 9:702. doi: 10.3389/fimmu.2018.00702

Mookherjee, N., and El-Gabalawy, H. S. (2013). High degree of correlation between whole blood and PBMC expression levels of miR-155 and miR-146a in healthy controls and rheumatoid arthritis patients. *J. Immunol. Methods* 400–401, 106–110. doi: 10.1016/j.jim.2013.10.001

Moosavi, A., and Motevalizadeh Ardekani, A. (2016). Role of epigenetics in biology and human diseases. *Iran. Biomed. J.* 20, 246–258. doi: 10.22045/ibj.2016.01

Moran-Moguel, M. C., Petarra-del Rio, S., Mayorquin-Galvan, E. E., and Zavala-Cerna, M. G. (2018). Rheumatoid arthritis and miRNAs: a critical review through a functional view. *J. Immunol. Res.* 2018, 1–16. doi: 10.1155/2018/2474529

Mu, N., Gu, J., Huang, T., Zhang, C., Shu, Z., Li, M., et al. (2016). A novel NF-κB/YY1/microRNA-10a regulatory circuit in fibroblast-like synoviocytes regulates inflammation in rheumatoid arthritis. *Sci. Rep.* 6:20059. doi: 10.1038/srep20059

Murata, K., Furu, M., Yoshitomi, H., Ishikawa, M., Shibuya, H., Hashimoto, M., et al. (2013). Comprehensive microRNA analysis identifies miR-24 and miR-125a-5p as plasma biomarkers for rheumatoid arthritis. *PLoS One* 8:e69118. doi: 10.1371/journal.pone.0069118

Murata, K., Yoshitomi, H., Tanida, S., Ishikawa, M., Nishitani, K., Ito, H., et al. (2010). Plasma and synovial fluid microRNAs as potential biomarkers of rheumatoid arthritis and osteoarthritis. *Arthritis Res. Ther.* 12:R86. doi: 10.1186/ar3013

Nakano, K., Boyle, D. L., and Firestein, G. S. (2013a). Regulation of DNA methylation in rheumatoid arthritis synoviocytes. *J. Immunol.* 190, 1297–1303. doi: 10.4049/jimmunol.1202572

Nakano, K., Whitaker, J. W., Boyle, D. L., Wang, W., and Firestein, G. S. (2013b). DNA methylome signature in rheumatoid arthritis. *Ann. Rheum. Dis.* 72, 110–117. doi: 10.1136/annrheumdis-2012-201526

Neidhart, M., Rethage, J., Kuchen, S., Künzler, P., Crowl, R. M., Billingham, M. E., et al. (2000). Retrotransposable L1 elements expressed in rheumatoid arthritis synovial tissue: association with genomic DNA hypomethylation and

influence on gene expression. *Arthritis Rheum.* 43, 2634–2647. doi: 10.1002/1529-0131(200012)43:12<2634::aid-anr3>3.0.co;2-1

Niederer, F., Trenkmann, M., Ospelt, C., Karouzakis, E., Neidhart, M., Stanczyk, J., et al. (2012). Down-regulation of microRNA-34a* in rheumatoid arthritis synovial fibroblasts promotes apoptosis resistance. *Arthritis Rheum.* 64, 1771–1779. doi: 10.1002/art.34334

Nile, C. J., Read, R. C., Akil, M., Duff, G. W., and Wilson, A. G. (2008). Methylation status of a single CpG site in the IL6 promoter is related to IL6 messenger RNA levels and rheumatoid arthritis. *Arthritis Rheum.* 58, 2686–2693. doi: 10.1002/art.23758

Ospelt, C. (2016). Epigenetic biomarkers in rheumatology – the future? *Swiss Med. Wkly.* 146:w14312. doi: 10.4414/smw.2016.14312

Ouboussad, L., Hunt, L., Hensor, E. M. A., Nam, J. L., Barnes, N. A., Emery, P., et al. (2017). Profiling microRNAs in individuals at risk of progression to rheumatoid arthritis. *Arthritis Res. Ther.* 19:288. doi: 10.1186/s13075-017-1492-9

Pandis, I., Ospelt, C., Karagianni, N., Denis, M. C., Reczko, M., Camps, C., et al. (2012). Identification of microRNA-221/222 and microRNA-323-3p association with rheumatoid arthritis via predictions using the human tumour necrosis factor transgenic mouse model. *Ann. Rheum. Dis.* 71, 1716–1723. doi: 10.1136/annrheumdis-2011-200803

Pashayan, N., Reisel, D., and Widschwendter, M. (2016). Integration of genetic and epigenetic markers for risk stratification: opportunities and challenges. *Per. Med.* 13, 93–95. doi: 10.2217/pme.15.53

Rawlings-Goss, R. A., Campbell, M. C., and Tishkoff, S. A. (2014). Global population-specific variation in miRNA associated with cancer risk and clinical biomarkers. *BMC Med. Genomics* 7:53. doi: 10.1186/1755-8794-7-53

Raychaudhuri, S., Sandor, C., Stahl, E. A., Freudenberg, J., Lee, H.-S., Jia, X., et al. (2012). Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat. Genet.* 44, 291–296. doi: 10.1038/ng.1076

Salehi, E., Eftekhari, R., Oraei, M., Gharib, A., and Bidad, K. (2015). MicroRNAs in rheumatoid arthritis. *Clin. Rheumatol.* 34, 615–628. doi: 10.1007/s10067-015-2898-x

Samanta, S., Rajasingh, S., Cao, T., Dawn, B., and Rajasingh, J. (2017). Epigenetic dysfunctional diseases and therapy for infection and inflammation. *Biochim. Biophys. Acta* 1863, 518–528. doi: 10.1016/j.bbadis.2016.11.030

Shibuya, H., Nakasa, T., Adachi, N., Nagata, Y., Ishikawa, M., Deie, M., et al. (2013). Overexpression of microRNA-223 in rheumatoid arthritis synovium controls osteoclast differentiation. *Mod. Rheumatol.* 23, 674–685. doi: 10.3109/s10165-012-0710-1

Singh, R. P., Massachi, I., Manickavel, S., Singh, S., Rao, N. P., Hasan, S., et al. (2013). The role of miRNA in inflammation and autoimmunity. *Autoimmun. Rev.* 12, 1160–1165. doi: 10.1016/j.autrev.2013.07.003

Sparks, J. A., and Costenbader, K. H. (2014). Genetics, environment, and gene-environment interactions in the development of systemic rheumatic diseases. *Rheum. Dis. Clin. N. Am.* 40, 637–657. doi: 10.1016/j.rdc.2014.07.005

Stanczyk, J., Ospelt, C., Karouzakis, E., Filer, A., Raza, K., Kolling, C., et al. (2011). Altered expression of microRNA-203 in rheumatoid arthritis synovial fibroblasts and its role in fibroblast activation. *Arthritis Rheum.* 63, 373–381. doi: 10.1002/art.30115

Stanford, S. M., Maestre, M. F., Campbell, A. M., Bartok, B., Kiosses, W. B., Boyle, D. L., et al. (2013). Protein tyrosine phosphatase expression profile of rheumatoid arthritis fibroblast-like synoviocytes: a novel role of SH2 domain-containing phosphatase 2 as a modulator of invasion and survival. *Arthritis Rheum.* 65, 1171–1180. doi: 10.1002/art.37872

Tavasolian, F., Abdollahi, E., Rezaei, R., Momtazi-borojeni, A. A., Henrotin, Y., and Sahebkar, A. (2018). Altered expression of MicroRNAs in rheumatoid arthritis. *J. Cell. Biochem.* 119, 478–487. doi: 10.1002/jcb.26205

Tessarz, P., and Kouzarides, T. (2014). Histone core modifications regulating nucleosome structure and dynamics. *Nat. Rev. Mol. Cell Biol.* 15, 703–708. doi: 10.1038/nrm3890

Toussirot, E., Abbas, W., Khan, K. A., Tissot, M., Jeudy, A., Baud, L., et al. (2013). Imbalance between HAT and HDAC activities in the PBMCs of patients with ankylosing spondylitis or rheumatoid arthritis and influence of HDAC inhibitors on TNF alpha production. *PLoS One* 8:e70939. doi: 10.1371/journal.pone.0070939

Trenkmann, M., Brock, M., Gay, R. E., Kolling, C., Speich, R., Michel, B. A., et al. (2011). Expression and function of EZH2 in synovial fibroblasts: epigenetic repression of the Wnt inhibitor SFRP1 in rheumatoid arthritis. *Ann. Rheum. Dis.* 70, 1482–1488. doi: 10.1136/ard.2010.143040

van Steenbergen, H. W., Luijk, R., Shoemaker, R., Heijmans, B. T., Huizinga, T. W. J., and van der Helm-van Mil, A. H. M. (2014). Differential methylation within the major histocompatibility complex region in rheumatoid arthritis: a replication study. *Rheumatology* 53, 2317–2318. doi: 10.1093/rheumatology/keu380

van Vollenhoven, R. F. (2009). Sex differences in rheumatoid arthritis: more than meets the eye. *BMC Med.* 7:12. doi: 10.1186/1741-7015-7-12

Viatte, S., Plant, D., and Raychaudhuri, S. (2013). Genetics and epigenetics of rheumatoid arthritis. *Nat. Rev. Rheumatol.* 9, 141–153. doi: 10.1038/nrrheum.2012.237

Vicente, R., Noël, D., Pers, Y.-M., Apparailly, F., and Jorgensen, C. (2016). Deregulation and therapeutic potential of microRNAs in arthritic diseases. *Nat. Rev. Rheumatol.* 12, 211–220. doi: 10.1038/nrrheum.2015.162

Wada, T. T., Araki, Y., Sato, K., Aizaki, Y., Yokota, K., Kim, Y. T., et al. (2014). Aberrant histone acetylation contributes to elevated interleukin-6 production in rheumatoid arthritis synovial fibroblasts. *Biochem. Biophys. Res. Commun.* 444, 682–686. doi: 10.1016/j.bbrc.2014.01.195

Wang, F., Chen, F.-F., Gao, W.-B., Wang, H.-Y., Zhao, N.-W., Xu, M., et al. (2016). Identification of citrullinated peptides in the synovial fluid of patients with rheumatoid arthritis using LC-MALDI-TOF/TOF. *Clin. Rheumatol.* 35, 2185–2194. doi: 10.1007/s10067-016-3247-4

Wang, H., Peng, W., Ouyang, X., Li, W., and Dai, Y. (2012). Circulating microRNAs as candidate biomarkers in patients with systemic lupus erythematosus. *Transl. Res.* 160, 198–206. doi: 10.1016/j.trsl.2012.04.002

Webster, A. P., Plant, D., Ecker, S., Zufferey, F., Bell, J. T., Feber, A., et al. (2018). Increased DNA methylation variability in rheumatoid arthritis-discordant monozygotic twins. *Genome Med.* 10:64. doi: 10.1186/s13073-018-0575-9

Weinbaum, J. S., Broekelmann, T. J., Pierce, R. A., Werneck, C. C., Segade, F., Craft, C. S., et al. (2008). Deficiency in microfibril-associated glycoprotein-1 leads to complex phenotypes in multiple organ systems. *J. Biol. Chem.* 283, 25533–25543. doi: 10.1074/jbc.M709962200

Whitaker, J. W., Boyle, D. L., Bartok, B., Ball, S. T., Gay, S., Wang, W., et al. (2015). Integrative omics analysis of rheumatoid arthritis identifies non-obvious therapeutic targets. *PLoS One* 10:e0124254. doi: 10.1371/journal.pone.0124254

Whitaker, J. W., Shoemaker, R., Boyle, D. L., Hillman, J., Anderson, D., Wang, W., et al. (2013). An imprinted rheumatoid arthritis methylome signature reflects pathogenic phenotype. *Genome Med.* 5:40. doi: 10.1186/gm444

Whitaker, J. W., Zhang, V., Stanford, S. M., Ai, R., Firestein, G. S., Aleman Muench, G. R., et al. (2016). Abnormal PTPN11 enhancer methylation promotes rheumatoid arthritis fibroblast-like synoviocyte aggressiveness and joint inflammation. *JCI Insight* 1, 1–13. doi: 10.1172/jci.insight.86580

Wluka, A. E., Cicuttini, F. M., and Spector, T. D. (2000). Menopause, oestrogens and arthritis. *Maturitas* 35, 183–199. doi: 10.1016/S0378-5122(00)00118-3

Xiao, M., Ma, Y., Chen, X., and Kuang, B. (2015). Single nucleotide polymorphism of miR-149 and susceptibility of rheumatoid arthritis. *Zhong Nan Da Xue Xue Bao. Yi Xue Ban* 40, 495–498. doi: 10.11817/j.issn.1672-7347.2015.05.006

Yang, S., and Yang, Y. (2015). Downregulation of microRNA-221 decreases migration and invasion in fibroblast-like synoviocytes in rheumatoid arthritis. *Mol. Med. Rep.* 12, 2395–2401. doi: 10.3892/mmr.2015.3642

Yarwood, A., Huizinga, T. W. J., and Worthington, J. (2016). The genetics of rheumatoid arthritis: risk and protection in different stages of the evolution of RA. *Rheumatology* 55, 199–209. doi: 10.1093/rheumatology/keu323

Ye, Y., Gao, X., and Yang, N. (2018). LncRNA ZFAS1 promotes cell migration and invasion of fibroblast-like synoviocytes by suppression of miR-27a in rheumatoid arthritis. *Hum. Cell* 31, 14–21. doi: 10.1007/s13577-017-0179-5

Zhang, B., Wang, L.-S., and Zhou, Y.-H. (2017). Elevated microRNA-125b promotes inflammation in rheumatoid arthritis by activation of NF-κB pathway. *Biomed. Pharmacother.* 93, 1151–1157. doi: 10.1016/j.biopha.2017.07.042

Zhang, H., Wei, Q., Wang, S., Zhang, H., Zhang, X., Geng, Q., et al. (2017). LncRNA HOTAIR alleviates rheumatoid arthritis by targeting miR-138 and inactivating NF-κB pathway. *Int. Immunopharmacol.* 50, 283–290. doi: 10.1016/j.intimp.2017.06.021

Zheng, F., Yu, X., Huang, J., and Dai, Y. (2017). Circular RNA expression profiles of peripheral blood mononuclear cells in rheumatoid arthritis patients, based on microarray chip technology. *Mol. Med. Rep.* 16, 8029–8036. doi: 10.3892/mmr.2017.7638

Zhou, Q., Haupt, S., Kreuzer, J. T., Hammitzsch, A., Proft, F., Neumann, C., et al. (2015). Decreased expression of miR-146a and miR-155 contributes to an abnormal Treg phenotype in patients with rheumatoid arthritis. *Ann. Rheum. Dis.* 74, 1265–1274. doi: 10.1136/annrheumdis-2013-204377

Zhou, Q., Long, L., Zhou, T., Tian, J., and Zhou, B. (2016). Demethylation of MicroRNA-124a genes attenuated proliferation of rheumatoid arthritis derived fibroblast-like synoviocytes and synthesis of tumor necrosis Factor-α. *PLoS One* 11:e0164207. doi: 10.1371/journal.pone.0164207

Zhu, H., Wu, L.-F., Mo, X.-B., Lu, X., Tang, H., Zhu, X.-W., et al. (2019). Rheumatoid arthritis-associated DNA methylation sites in peripheral blood mononuclear cells. *Ann. Rheum. Dis.* 78, 36–42. doi: 10.1136/annrheumdis-2018-213970

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Nemtsova, Zaletaev, Bure, Mikhaylenko, Kuznetsova, Alekseeva, Beloukhova, Deviatkin, Lukashev and Zamyatnin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

REVIEW
published: 25 June 2019
doi: 10.3389/fgene.2019.00580

Transposable Elements and Their Epigenetic Regulation in Mental Disorders: Current Evidence in the Field

Błażej Misiak[1], Laura Ricceri[2] and Maria M. Sąsiadek[1]*

[1] Department of Genetics, Wrocław Medical University, Wrocław, Poland, [2] Centre for Behavioural Sciences and Mental Health, Istituto Superiore di Sanità, Rome, Italy

Transposable elements (TEs) are highly repetitive DNA sequences in the human genome that are the relics of previous retrotransposition events. Although the majority of TEs are transcriptionally inactive due to acquired mutations or epigenetic processes, around 8% of TEs exert transcriptional activity. It has been found that TEs contribute to somatic mosaicism that accounts for functional specification of various brain cells. Indeed, autonomous retrotransposition of long interspersed element-1 (LINE-1) sequences has been reported in the neural rat progenitor cells from the hippocampus, the human fetal brain and the human embryonic stem cells. Moreover, expression of TEs has been found to regulate immune-inflammatory responses, conditioning immunity against exogenous infections. Therefore, aberrant epigenetic regulation and expression of TEs emerged as a potential mechanism underlying the development of various mental disorders, including autism spectrum disorders (ASD), schizophrenia, bipolar disorder, major depression, and Alzheimer's disease (AD). Consequently, some studies revealed that expression of some sequences of human endogenous retroviruses (HERVs) appears only in a certain group of patients with mental disorders (especially those with schizophrenia, bipolar disorder, and ASD) but not in healthy controls. In addition, it has been found that expression of HERVs might be related to subclinical inflammation observed in mental disorders. In this article, we provide an overview of detrimental effects of transposition on the brain development and immune mechanisms with relevance to mental disorders. We show that transposition is not the only mechanism, explaining the way TEs might shape the phenotype of mental disorders. Other mechanisms include the regulation of gene expression and the impact on genomic stability. Next, we review current evidence from studies investigating expression and epigenetic regulation of specific TEs in various mental disorders. Most consistently, these studies indicate altered expression of HERVs and methylation of LINE-1 sequences in patients with ASD, schizophrenia, and mood disorders. However, the contribution of TEs to the etiology of AD is poorly documented. Future studies should further investigate the mechanisms linking epigenetic processes, specific TEs and the phenotype of mental disorders to disentangle causal associations.

Keywords: retrotransposon, DNA methylation, LINE-1, Alu, SINE, SVA

OPEN ACCESS

Edited by:
Yun Liu,
Fudan University, China

Reviewed by:
Nicole Grandi,
University of Cagliari, Italy
Apiwat Mutirangura,
Chulalongkorn University, Thailand

****Correspondence:***
Błażej Misiak
blazej_misiak@interia.pl;
mblazej@interia.eu

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *10 April 2019*
Accepted: *04 June 2019*
Published: *25 June 2019*

Citation:
Misiak B, Ricceri L and Sąsiadek MM (2019) Transposable Elements and Their Epigenetic Regulation in Mental Disorders: Current Evidence in the Field. Front. Genet. 10:580. doi: 10.3389/fgene.2019.00580

INTRODUCTION

Mental disorders represent complex phenotypes and are the leading causes of global disease burden (Vigo et al., 2016). The phenotype complexity of mental disorders manifests in symptomatic and biological overlap, impeding a diagnostic process that is based on a clinical consensus without a crucial role of biological markers. Heritability rates of mental disorders are high, exceeding 80% in twin studies of schizophrenia and bipolar disorder (Cardno and Gottesman, 2000; McGuffin et al., 2003; Misiak et al., 2016). However, monogenic determinants with high penetrance rates have not been identified so far, and the concept of major mental disorders as the polygenic phenotypes prevails in the research approaches. Consequently, a paradigm shift toward investigating polygenic signatures, gene × environment (G × E) interactions and epigenetic mechanisms has been widely observed in the recent years.

The term 'epigenetics' refers to a number of reversible mechanisms that impacts gene expression without altering DNA sequence, and include DNA methylation and hydroxymethylation at the CpG islands, histone modifications as well as the regulation by microRNA species. It is now increasingly being recognized that the brain development is a complex process during which there is an increased sensitivity to the regulatory effects of epigenetic mechanisms (Nagy and Turecki, 2012). In light of existing evidence, major mental disorders, especially schizophrenia and autism spectrum disorders (ASD), are perceived as the neurodevelopmental disorders, occurring due to the effects of various genetic and environmental factors that affect critical periods of the brain development (Meredith, 2015).

Transposable elements (TEs) are the highly repetitive DNA sequences that constitute more than 50% of the human genome and contain about 52% of all CpG dinucleotides (Cordaux and Batzer, 2009; Su et al., 2012). Therefore, methylation at TEs is believed to serve as a proxy measure of global DNA methylation. Some TEs share similarity to exogenous viral agents and thus they are called endogenous retroviruses (Griffiths, 2001). Only about 7% of TEs are transcriptionally active (Oja et al., 2008). It has been estimated that approximately 0.27% of human genetic diseases are caused by retrotransposition (Callinan and Batzer, 2006).

Less is known about the contribution of TEs to the etiology of mental disorders. However, there is accumulating evidence that retrotransposition plays an important role in shaping somatic mosaicism that accounts for functional specification of brain cells (Baillie et al., 2011; Poduri et al., 2013). For instance, it has been reported that the transposition of long interspersed element (LINE)-1 sequences may play a role in differentiation of neurons during the brain development (Muotri et al., 2010). Moreover, this sequence exerts autonomous retrotransposition activity in the neural rat progenitor cells from the hippocampus, the human fetal brain and the human embryonic stem cells (Muotri et al., 2005; Coufal et al., 2009). Therefore, aberrant epigenetic regulation of TEs has been hypothesized to play an important role in the development of mental disorders. In this article, we provide an overview of transposition processes with relevance to major psychiatric disorders. Next, we review human and animal model studies investigating expression and epigenetic regulation of TEs in various mental disorders. Finally, we provide a summary of evidence with future directions and potential translation of findings to personalized precision medicine.

BRIEF OVERVIEW OF TEs IN THE HUMAN GENOME – CLASSIFICATION AND NOMENCLATURE

Classification of TEs in the human genome was shown in **Figure 1**. A detailed description of the structure and function of various TEs can be found elsewhere (Darby and Sabunciyan, 2014). All TEs can be divided into two subgroups – type I TEs (retrotransposons) and type II TEs (DNA transposons). Type I TEs can be divided into two subgroups – long terminal repeat (LTR) elements, represented by the human endogenous retroviruses (HERVs) and non-LTR sequences that include LINEs, short interspersed elements (SINEs) and processed pseudogenes (Dewannieux and Heidmann, 2005). Retrotransposons act via RNA intermediates that are converted to DNA sequences before transposition (reverse transcription) (Munoz-Lopez and Garcia-Perez, 2010). Type II TEs encode enzymes required for insertion and excision, enabling direct transposition processes without the use of RNA intermediates (Pray, 2008b). Some TEs are autonomous and encode all enzymes that are necessary for transposition, while the rest of them require a transcriptional activity of other transposons. Type II TEs have lost a transposition activity (Darby and Sabunciyan, 2014).

The HERV sequences have likely existed as exogenous infectious factors; however, they have lost this activity due to acquired mutations (Bannert and Kurth, 2006). These TEs constitute 8% of the human genome and contain genes that are conservative for all retroviruses, including the *gag*, *pro*, *pol*, and *env* genes (Lander et al., 2001; Vargiu et al., 2016). The *gag* gene encodes proteins that build up matrix, capsid and nucleocapsid. *Pro* and *pol* encode protease, reverse transcriptase and integrase. In turn, the *env* gene is expressed to surface and transmembrane proteins. The HERV sequences in the human genome represent three classes of retroviruses: class I (e.g., HERV-E, HERV-W, HERV-FRD, and HERV-H), class II (e.g., HERV-K), and class III (e.g., HERV-L). This classification is based on the similarity to exogenous retroviruses. The HERV-K sequences are the youngest and exert the highest transcriptional activity. The HERV sequences can provide promoters, enhancers, repressors, poly-A signals and alternative splicing sites for human transcripts (Vargiu et al., 2016).

The LINEs that represent non-LTR elements, possess an autonomous retrotransposition activity and include LINE-1 and LINE-2 sequences. These sequences make up approximately 21% of the human genome (Lander et al., 2001; Schumann et al., 2010). The LINE-1 sequences contain their own promoters and encode two open reading frame proteins – ORF1 that is an RNA-binding protein and ORF2 with endonuclease and reverse transcriptase activities. They are the most abundant sequences from the LINE family, making up 18% of the human genome (Lander et al., 2001). The majority of LINE-1 sequences are transcriptionally inactive. The LINE-2 sequences in the human

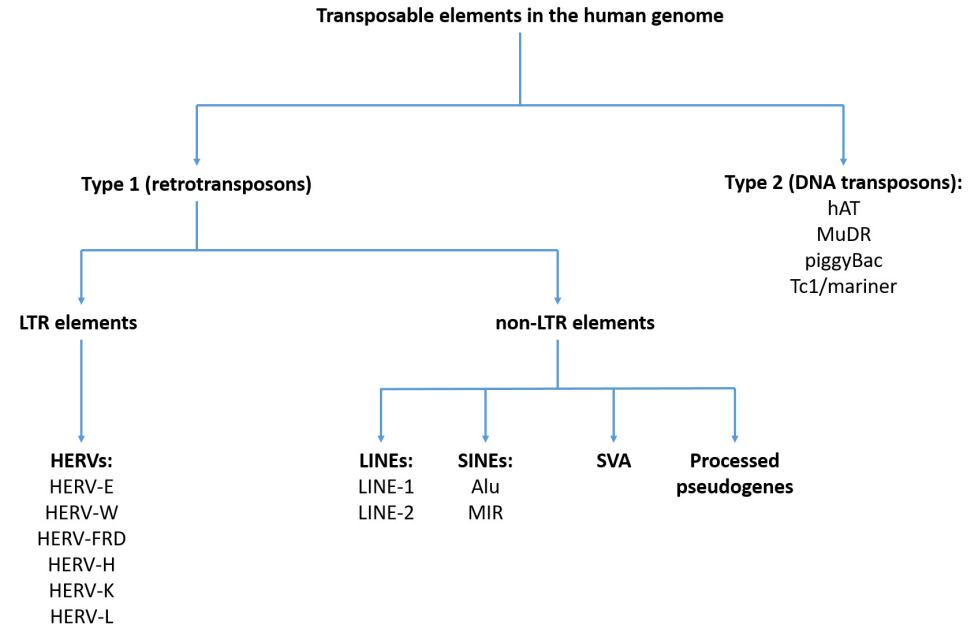


FIGURE 1 | Classification of transposable elements in the human genome.

genome are highly defective and can encode either one or two ORF proteins (Darby and Sabunciyan, 2014).

The SINEs are active and non-autonomous TEs, represented by the Alu and the Mammalian wide Interspersed Repeat (MIR) elements (11 and 3% of the human genome, respectively). The Alu sequences were named based on sharing a common cleavage site for the *Alu*I restriction enzyme (Houck et al., 1979). The Alu sequences are active but require the reverse transcriptase that is encoded by LINE-1 sequences (Pray, 2008a). In turn, the MIR elements are inactive. It has recently been shown that SINEs may form more complex sequences that are classified as the SVA retrotransposons. The SVA sequences have been formed by coupling the SINEs, a variable number of tandem repeats and the Alu retrotransposons. The SVAs also require the LINE-1 expression for mobilization. These sequences contribute to about 0.1% of the human genome and are the most active group of retrotransposons (Ostertag et al., 2003; Wang et al., 2005).

In turn, pseudogenes are DNA sequences that are related to real genes but they have lost at least some protein-coding abilities. It has been found that mRNA of pseudogenes can be reverse transcribed by the proteins encoded by LINE-1 sequences and transferred into other regions of the genome, creating processed pseudogenes. It has been estimated that the human genome consists of over 7,800 pseudogenes (Zhang et al., 2003). In case of integration close to active promoters, processed pseudogenes can be further transcribed. As listed by Kazazian (2014), they share the following characteristics: (1) their sequences are similar to the transcribed part of the parent gene; (2) they lack all or most introns; (3) they contain a poly-A tail attached to the 3′-most transcribed nucleotide; and (4) they are flanked at their 5′ and 3′ terminals by target site duplications of 5–20 nucleotides.

Finally, little is known about type II TEs (DNA transposons) that have never been active in the human genome. Type II TEs include the hAT, MuDR, piggyBac, and Tc1/mariner sequences (Munoz-Lopez and Garcia-Perez, 2010). These transposons do not act via RNA intermediates and encode enzymes that enable their mobilization. Due to their inactivity their causal role in the etiology of human diseases is less likely (Darby and Sabunciyan, 2014).

INSIGHTS INTO POTENTIAL MECHANISMS UNDERLYING THE ROLE OF TEs IN MENTAL DISORDERS

A recent review of human monogenic diseases that occur due to retrotransposition suggests that only the transposition of LINE-1, Alu, and SVA sequences might be deleterious, underlying the development of monogenic diseases (Kaer and Speek, 2013). Retrotransposition might affect various gene regions via altering their sequence or influencing expression activity. For instance, the Alu sequences contain several stop codons that may result in a truncated protein (Mighell et al., 1997). This mechanism has been discovered in patients with hemophilia B caused by transferring the Alu-Ya5 element into a protein coding region of the factor IX gene (Vidaud et al., 1993). In case of transposition into promoter regions, these sequences might impact gene expression. Another scenario originates from sequence homology that can promote homologous recombination, leading to insertions and deletions. Finally, the SVA tandems can mobilize exons, contributing to complex rearrangements.

However, the effects of alterations in DNA sequence triggered by retrotransposition have not been found to underlie the

development of common mental disorders. In the majority of studies of patients with mental disorders (reviewed in detail below), altered expression and/or epigenetic regulation of retrotransposons have been reported. There are several epigenetic processes that act as defense mechanism against retrotransposition, including DNA methylation, histone modifications, small RNA-mediated regulation and post-transcriptional silencing by DICER and siRNA (Lapp and Hunter, 2016). Indeed, the majority of TEs in the human genome are hypermethylated (Pray, 2008b). Although DNA methylation acts as a defense mechanism, it cannot be excluded that hypermethylation of newly inserted TEs can lead to further changes in chromatin conformation, triggering changes in the expression of adjacent genes. It is most likely that retrotransposition occurs during early development when epigenetic marks are removed (Darby and Sabunciyan, 2014). There are also some well characterized histone modifications, including trimethylation of lysine 9 and lysine 27 at histone H3 (H3K9me3 and H3K27me3, respectively), which lead to heterochromatin formation and transcriptional silencing of TEs (Day et al., 2010; Baker et al., 2012).

It should be noted that only a small subset of TEs has been reported to be involved in retrotransposition. For instance, only 30–60 LINE-1 sequences in diploid cells are capable of retrotransposition (Sassaman et al., 1997). In addition, the majority of LINE-1 sequences are methylated to a certain degree. It has been found that LINE-1 methylation might impact gene expression via specific mechanisms [for review see (Kitkumthorn and Mutirangura, 2011)]. Firstly, LINE-1 sequences may produce unique RNA transcripts that act beyond the LINE-1 location. Alternatively, the reverse LINE-1 promoter can transcribe unique DNA sequences beyond the 5′ end of LINE-1. The second scenario is that intragenic LINE-1 RNAs decrease the expression of host gene via the nuclear RNA-induced silencing complexes. Indeed, it has been found that the Argonaute-2 (AGO2) protein targets intronic LINE-1 pre-mRNA complexes leading to down-regulation of gene expression in cancer cells (Aporntewan et al., 2011).

Global DNA hypomethylation that progresses with aging has been associated with genomic instability (Jung and Pfeifer, 2015). Hypomethylated genome regions are prone to accumulate various types of DNA lesions that include oxidative damage, depurination, depyrimidation and pathologic endogenous double-strand breaks (Mutirangura, 2019). The latter ones are now believed to act as intermediate products that drive genomic instability (Mutirangura, 2019). Accumulating evidence indicates that methylation of TEs might protect against genomic instability processes. For instance, it has been demonstrated that Alu siRNA increases Alu methylation levels, lowers endogenous DNA damage and increases DNA resistance to DNA damaging agents (Patchsung et al., 2018). Similarly, LINE-1 hypomethylation may contribute to genomic instability via interactions with the *ATM* gene expression (Kitkumthorn and Mutirangura, 2011). However, the contribution of a reduction in the Alu methylation to genomic instability might be greater than that of LINE-1 or HERV sequences (Jintaridth and Mutirangura, 2010).

It remains largely unknown how changes in the expression of TEs might contribute to the development of mental disorders. It has been hypothesized that the presence of TEs in the human genome provides immunity against several infectious agents. Indeed, the mechanisms that contributed to HERV insertions are analogous to those used for replication by exogenous retroviruses (Grandi and Tramontano, 2018). Therefore, changes in the expression of TEs, e.g., via epigenetic processes, might impact immune responses and make the host more liable to exogenous infections. There is evidence that HERV-derived peptides may interact with innate immunity via various mechanisms. For instance, HERV proteins are able to interact with pattern recognition receptors (PRRs) that play a pivotal role in antiviral responses (Hurst and Magiorkinis, 2015). Emerging evidence indicates that exogenous viruses, including herpesviruses and influenza virus, might modulate the expression of HERV sequences. This mechanism might play a protective role and has been reviewed in detail by Grandi and Tramontano (2018). In brief, HERV transcripts might interact with homologous RNA from exogenous retroviruses, leading to the formation of molecules that are recognized by PRRs, acting as innate immunity sensors. The similarity of HERV proteins to those exogenous retroviruses allow them to compete with cellular receptors. This similarity might also trigger complementation events that impair formation of viral particles after cellular infection. On the other site, HERV proteins may suppress innate immunity. It has been reported that HERV-K proteins inhibit the activation of T cells (Morozov et al., 2013) as well as reduce the levels of interleukin-6 and Toll-like Receptor 7 (Laska et al., 2017).

TRANSPOSABLE ELEMENTS AND THEIR EPIGENETIC REGULATION IN MENTAL DISORDERS

As mentioned above, expression of TEs might play an important role in shaping immune responses against exogenous infections. Aberrant immune-inflammatory responses have been reported in several mental disorders. Also, a number of exogenous infections have been found to impact a risk of mental disorders. Below, we review studies investigating TEs and their epigenetic regulation in specific mental disorders, starting from the rationale of these studies that is based on the contribution of immune-inflammatory processes. A summary of human studies was provided in **Table 1**.

Autism Spectrum Disorders (ASD)

Overexpression of HERV-H has been observed in peripheral blood mononuclear cells (PBMC) of children with ASD (Balestrieri et al., 2012, 2016). Similar findings have also been observed in two different ASD mouse models – inbred BTBR T+tf/J mice and CD-1 outbred mice prenatally exposed to valproic acid. In both of these mouse models, the expression of several endogenous retrovirus (ERV) families (ETnI, ETnII-α, ETnII-β, ETnII-γ, MusD, and IAP) was significantly higher in comparison with corresponding controls (Cipriani et al., 2018). Interestingly, the studies in mouse models provide additional

TABLE 1 | Overview of human studies investigating the role of TEs in mental disorders.

Diagnosis	Study	*N*	Methods	Main findings
ASD	Balestrieri et al., 2012	28 ASD children 28 HCs	Expression of HERV sequences (E, H, K, and W) in PBMCs – RT-PCR	The percentage of HERV-H and HERV-W positive samples was significantly higher in ASD patients compared to HCs. HERV-H was more abundantly expressed and HERV-W had lower abundance in ASD patients than in HCs. PBMCs from ASD patients had an increased potential to up-regulate the HERV-H expression upon stimulation.
	Balestrieri et al., 2016	30 ASD children 30 HCs	Expression of HERV sequences (H, K, and W) in PBMCs – RT-PCR	There were significantly higher expression levels of HERV-H in PBMCs from ASD children compared to HCs. In turn, expression levels of HERV-K and HERV-W were significantly higher in PBMCs from healthy controls compared to ASD children.
	Saeliw et al., 2018	36 ASD patients 20 HCs Publicly available datasets of DNA expression profiling (465 ASD patients and 276 HCs)	Methylation (COBRA) and expression of Alu sequences (RT-PCR) in lymphoblastoid cell lines	A total of 320 Alu-inserted genes were differentially expressed. These genes are known to be associated with neurodevelopmental and neurological disorders. Pathway analysis revealed that these genes are involved in inflammation, estrogen receptor and androgen signaling.
	Shpyleva et al., 2018	13 ASD patients 13 HCs	LINE-1 methylation (the 5-methylcytosine MeDIP assay) and expression (RT-PCR) as well as H3K9me3 (ChIP) in post-mortem brain samples	LINE-1 expression was significantly higher in the cerebellum but not in the BA9, BA22, and BA24 brain regions from ASD patients. The binding of repressive MeCP2 protein and histone H3K9me3 to LINE-1 was significantly lower in the cerebellum of ASD patients. Higher LINE-1 expression was associated with elevated levels of oxidative stress. Overall, no significant differences in methylation levels between ASD patients and HCs were found. However, significantly altered Alu methylation patterns were observed in ASD cases sub-grouped based on clinical manifestation compared with HCs.
SCZ	Bundo et al., 2014	48 SCZ patients 13 BD patients 12 MDD patients 47 HCs	The number of LINE-1 copies in brain samples and induced pluripotent cells (RT-PCR)	The LINE-1 retrotransposition in neurons from the prefrontal cortex of patients with schizophrenia was increased, especially in the genes involved in synaptic functions. These findings were confirmed in induced pluripotent cells from patients with 22q11 deletion syndrome as well as in a mouse model of schizophrenia (maternal immune activation paradigm).
	Doyle et al., 2017	36 SCZ patients 26 HCs	LINE-1 insertions in the dorsolateral prefrontal cortex samples (qPCR)	A significant increase in the number of intragenic LINE-1 insertions has been observed in the dorsolateral prefrontal cortex of patients with schizophrenia compared to healthy controls.
	Fachim et al., 2019	35 FES patients 21 siblings of SCZ patients 35 HCs	LINE-1 methylation (pyrosequencing) in peripheral blood leukocytes	LINE-1 methylation was significantly higher in FES patients and siblings of schizophrenia patients compared to HCs.
	Frank et al., 2005	39 SCZ patients 39 HCs	Expression of HERV sequences in brain samples (HERV chip and RT-PCR)	Overrepresentation of the HERV-K(HML2) group in the brain samples of SCZ patients was found.
	Huang et al., 2006	58 recent-onset SCZ patients 38 HCs	Expression of the HERV *pol* gene in peripheral blood leukocytes (RT-PCR)	The HERV pol gene expression was detected in 34.5% of SCZ patients but not in HCs.
	Kalayasiri et al., 2019	31 paranoid SCZ patients 94 MIP patients 163 HCs	LINE-1 methylation in PBMCs (COBRA)	Methylation levels of LINE-1 sequences were significantly higher in paranoid SCZ patients and MIP patients compared to HCs.
	Karlsson et al., 2002	35 FES and FESaff patients 20 SCZ and SCZaff patients 22 patients with non-inflammatory neurological disorders 30 HCs	Detection of HERV sequences in CSF (cDNA synthesis, pan-retroviral PCR, cloning and sequencing)	Nucleotide sequences homologous to those of the HERV-W *pol* gene were found in CSF of 28.6% of FES patients and in 5% of patients with SCZ and SCZaff. These sequences were not detected in the CSF of individuals with non-inflammatory neurological diseases and healthy controls.

(Continued)

TABLE 1 | Continued

Diagnosis	Study	*N*	Methods	Main findings
	Karlsson et al., 2004	54 FES and FESaff patients 46 HCs	The presence of HERV-W sequences in plasma samples (RT-PCR)	The HERV-W gag sequences were detected in 16.7% of patients and 4.3% of HCs.
	Li et al., 2019	32 SCZ patients 51 HCs	Analysis of post-mortem brain samples (RNAseq data)	Increased transcription of HERV, especially HERV-W and HERV-H elements, was found in the anterior cingulate cortex, hippocampus, and orbitofrontal cortex of patients with SCZ.
	Mak et al., 2019	100 MES patients 49 FES patients 97 HCs	Methylation of HERV-K sequences in peripheral blood leukocytes (COBRA)	There were no significant differences in the level of HERV-K methylation between MES patients and HCs. In turn, FES patients had significantly lower HERV-K methylation than HCs. There was a significant positive correlation between a cumulative dosage of antipsychotic and the HERV-K methylation level in MES patients.
	Melbourne et al., 2018	17 SCZ patients 16 HCs	Expression of HERV-W *env* and *gag* in PBMCs (RT-PCR)	There were no significant differences in the level of expression. In all participants, higher expression of HERV-W *env* and *gag* was related to higher levels of interleukin-6. There was a negative correlation between the dosage of atypical antipsychotic and the level of HERV-W *env* and *gag* expression.
	Misiak et al., 2015	48 FES patients 48 HCs	Methylation of LINE-1 sequences in peripheral blood leukocytes (COBRA)	There were no significant differences in the level of LINE-1 methylation between FES patients and HCs. However, FES patients with a positive history of childhood trauma had significantly lower LINE-1 methylation than HCs. More specifically, a higher level of emotional abuse was related to lower LINE-1 methylation in FES patients.
	Perron et al., 2012	45 SCZ patients 91 BD patients 73 HCs	Expression of HERV-W env gene in PBMCs (RT-PCR)	There were significantly elevated transcription levels of the HERV-W *env* sequence in SCZ patients compared to HCs. However, the levels of HERV-W *env* expression were significantly lower compared to BD patients.
	Weis et al., 2007	15 SCZ patients 15 HCs	Expression of HERV-W gag protein in brain samples (immunohistochemistry)	Expression level of the HERV-W gag protein has been found to be decreased in the cingulate gyrus and the hippocampus of patients with SCZ.
Mood disorders	Frank et al., 2005	38 BD patients 39 HCs	Expression of HERV sequences in brain samples (HERV chip and RT-PCR)	Overrepresentation of the HERV-K(HML2) group in the brain samples of BD patients was found.
	Li et al., 2018	24 BD patients 51 HCs	Analysis of post-mortem brain samples (RNAseq data)	Increased transcription of HERV, especially HERV-W and HERV-H elements, was found in the anterior cingulate cortex, hippocampus, and orbitofrontal cortex of patients with BD.
	Perron et al., 2012	91 BD patients 45 SCZ patients 73 HCs	Expression of HERV-W env gene in PBMCs (RT-PCR)	There were elevated transcription levels of the HERV-W env sequence in SCZ patients compared to SCZ and HCs.
	Weis et al., 2007	15 BD patients 15 MDD patients 15 HCs	Expression of HERV-W gag protein in brain samples (immunohistochemistry)	Expression level of the HERV-W gag protein has been found to be decreased in the cingulate gyrus and the hippocampus of patients with BD and SCZ.
AD	Bollati et al., 2011	43 AD patients 38 HCs	Methylation of LINE-1 sequences in peripheral blood leukocytes (pyrosequencing)	Increased LINE-1 methylation level was found in AD patients, especially those with better cognitive performance.
	Hernández et al., 2014	28 AD patients 30 HCs	Methylation of LINE-‘ sequences in peripheral blood leukocytes (MS-HRM)	No significant differences in LINE-1 methylation levels between AD patients and HCs were found.
	Protasova et al., 2017	18 AD patients 20 HCs	LINE-1 – the number of copies and expression level in post-mortem samples of frontal cortex (RT-PCR)	No significant differences in LINE-1 expression or the number of LINE-1 copies between AD patients and HCs were found.

AD, Alzheimer's disease; ASD, autism-spectrum disorder; BD, bipolar disorder; COBRA, combined bisulfite restriction analysis; CSF, cerebrospinal fluid; FES, first-episode schizophrenia; FESaff, first-episode schizoaffective disorder; H3K9me3, histone H3 lysine 9 trimethylation; HCs, healthy controls; MeCP2, methyl CpG binding protein 2; MDD, major depressive disorder; MIP, methamphetamine-induced psychosis; MS-HRM, methylation-sensitive high-resolution melting; PBMCs, peripheral blood mononuclear cells; RT-PCR, real-time polymerase chain reaction; SCZ, schizophrenia; SCZaff, schizoaffective disorder; qPCR, quantitative polymerase chain reaction.

information on the potential use of ERV sequences as biomarkers: (i) a higher expression of ERV was observed both in the peripheral blood mononuclear cells and the brain, suggesting that altered profile of peripheral ERV sequences may reflect similar alterations at the brain level; (ii) ERV overexpression in ASD mouse models is detectable from prenatal stage till the adulthood and (iii) ERV overexpression in ASD mouse models is also accompanied by increased expression of pro-inflammatory

cytokines and Toll-like receptors. Furthermore, a subsequent study in one of the models (mice prenatally exposed to valproic acid) provided evidence that higher levels of ERVs are also detectable in the offspring (second and third generations) of those mice exposed prenatally to valproic acid (Tartaglione et al., 2018).

Also LINE-1 retrotransposons have been associated with ASD (Shpyleva et al., 2018; Suarez et al., 2018). The levels of LINE-1 ORF1 and ORF2 transcripts have been investigated in four brain regions of patients with idiopathic autism (the frontal cortex, anterior cingulate, auditory cortex, and cerebellum). Elevated LINE-1 expression together with lower binding affinity of repressive MeCP2 protein and histone H3K9me3 to LINE-1 sequences was observed only in the cerebellum, suggesting a lessening of epigenetic repression and consequently an increase in chromatin accessibility. Interestingly, the increase in LINE-1 expression was also inversely correlated with glutathione redox status, consistent with reports indicating that LINE-1 expression is increased under pro-oxidant conditions (Shpyleva et al., 2018). The overexpression of LINE-1 within a single brain region is suggestive of a mosaicism-like impact of retrotransposons and definitively needs further investigation. In partial agreement with the findings of increased LINE-1 expression in ASD, data concerning LINE-1 methylation status in lymphoblastoid peripheral cells have provided evidence of reduced methylation in a subgroup of patients with severe language impairment (Tangsuwansri et al., 2018).

It has also been shown that the Alu sequence, the most abundant of all TEs in the human genome, deserves further research in ASD (Saeliw et al., 2018). Indeed, this study investigated the Alu methylation and expression in lymphoblastoid peripheral cells from ASD patients. Although the global methylation of Alu subfamilies was not significantly different between ASD and control group, when ASD samples were divided according to phenotypic subgroups, methylation patterns of the AluS subfamily were different from those in relative controls in two of the ASD subgroups, and within one of the subgroup (mild phenotype), the Alu expression was correlated with methylation status. Despite the limited sample size (particularly of subgroups), these data suggest that classification of ASD patients in phenotypic subgroups may represent a useful tool in investigating associations of TEs with the highly heterogeneous ASD diagnostic construct.

Schizophrenia-Spectrum Disorders

It has been clearly demonstrated that winter-spring seasonality of birth as well as prenatal and postnatal infections increase a risk of developing schizophrenia (McGrath and Welham, 1999; Davies et al., 2003; Khandaker et al., 2013). Moreover, the largest genome-wide association study revealed that variation within the HLA genes is strongly associated with schizophrenia susceptibility (Ripke et al., 2014). Finally, schizophrenia patients present with several indices of subclinical inflammation in terms of pro-inflammatory cytokine profiles (Miller et al., 2011; Frydecka et al., 2018), alterations of lymphocyte counts (Miller et al., 2013; Karpiński et al., 2016, 2018) and elevated levels of C-reactive protein (CRP) (Fernandes et al., 2016). On the basis of a meta-analysis, Arias et al. (2012) found a higher prevalence of infections with several agents, including *Human Herpesvirus* (HHV) *2*, *Borna Disease Virus*, *Chlamydia pneumoniae*, *Chlamydia psittaci,* and *Toxoplasma gondii* in patients with schizophrenia compared to healthy controls.

Accumulating evidence indicates altered expression of HERV sequences in patients with schizophrenia. Karlsson et al. (2002) found nucleotide sequences homologous to those of the HERV-W *pol* gene in the cerebrospinal fluid (CSF) of 28.6% of first-episode schizophrenia patients and in 5% of patients with chronic schizophrenia. These sequences were not detected in the CSF of individuals with non-inflammatory neurological diseases and healthy controls. Increased levels of HERV-W-related *gag* and *pol* transcripts and a higher prevalence of the *gag* and *pol* antigenemia in peripheral blood from patients with schizophrenia compared to healthy controls have been reported by several studies (Karlsson et al., 2004; Huang et al., 2006; Perron et al., 2008; Yao et al., 2008). The study by Perron et al. (2008) also revealed significantly higher rates of positive HERV-W *env* antigenemia in patients with schizophrenia than in healthy controls. The HERV-W *gag* and *env* antigenemia has been also associated with subclinical inflammation in terms of elevated levels of CRP and pro-inflammatory cytokines (Perron et al., 2008; Melbourne et al., 2018). Interestingly, Huang et al. (2011) found that overexpression of the HERV-W *env* in the human U251 glioma cells up-regulated a number of schizophrenia-associated genes, including those that encode brain-derived neurotrophic factor, neurotrophic tyrosine kinase receptor type 2 and the dopamine D3 receptor as well as increased the phosphorylation of cyclic adenosine monophosphate response element-binding protein. In this study, mRNA of the HERV-W *env* gene was detected in plasma from 42 out of 118 recent-onset schizophrenia patients but not in healthy controls. There is also evidence that the expression of HERV-W *env* induces calcium influx and down-regulates the *DISC1* gene expression in the human neuroblastoma cells (Chen et al., 2018). Interestingly, expression level of the HERV-W *gag* protein has been found to be decreased in the cingulate gyrus and the hippocampus of patients with schizophrenia (Weis et al., 2007). However, a recent analysis of RNA-seq data in the human *post mortem* brain samples revealed increased transcription of HERV, especially HERV-W and HERV-H elements, in the anterior cingulate cortex, hippocampus and orbitofrontal cortex of patients with schizophrenia and bipolar disorder (Li et al., 2019). Interestingly, the HERV sequences within the *ERVWE1* gene (7q21.2) exhibited the highest levels of transcription across all brain regions examined in this analysis. The *env* gene in this locus encodes syncytin-1, expressed at high levels in the human placenta (Blond et al., 2000; Mi et al., 2000). However, altered expression of this gene has been reported in the areas of active demyelination in patients with multiple sclerosis (Mameli et al., 2007). At this point, it should be noted that myelin alterations are widely observed in patients with schizophrenia (Mighdoll et al., 2015). Although initial results regarding expression of the HERV-W sequences in schizophrenia patients are promising, caution should be taken on the way these results are being interpreted. Indeed, the majority of studies in this field analyzed the overall expression of HERV-W sequences

without investigating specific HERV-W loci. Moreover, no conclusive association between the HERV-W expression and other human pathologies has been documented so far [for review see (Grandi and Tramontano, 2017)].

Less is known about other families of HERVs in patients with schizophrenia. Frank et al. (2005) found overrepresentation of the HERV-K(HML2) group in brain samples of patients with schizophrenia and bipolar disorder. Our group also tested peripheral blood methylation levels of HERV-K sequences in first-episode and multi-episode schizophrenia patients (Mak et al., 2019). We found significantly lower levels of HERV-K methylation in first-episode schizophrenia patients compared to healthy controls. These alterations were not observed in multi-episode schizophrenia patients. Moreover, we did not find an association between HERV-K methylation levels and the deficit schizophrenia subtype that refers to a subgroup of patients with enduring and persistent negative symptoms. However, we found a significant positive correlation between the dosage of antipsychotics and HERV-K methylation levels in multi-episode schizophrenia patients. These findings imply that the HERV-K methylation might normalize in the course of schizophrenia. It is also likely that antipsychotic drugs might impact methylation and expression of HERV-K sequences. In contrast to our findings, Diem et al. (2012) found no significant effects of valproic acid, haloperidol, risperidone and clozapine on the HERV-K expression levels in the human brain cell lines. However, valproic acid was found to strongly up-regulate expression of HERV-W and ERV9 elements.

Some studies also investigated methylation status and expression levels of non-LTR sequences in patients with schizophrenia. Bundo et al. (2014) demonstrated increased LINE-1 retrotransposition in neurons from the prefrontal cortex of patients with schizophrenia, especially in the genes involved in synaptic functions. These findings were confirmed in induced pluripotent cells from patients with 22q11 deletion syndrome as well as in a mouse model of schizophrenia (maternal immune activation paradigm). In agreement with these results, a significant increase in the number of intragenic LINE-1 insertions has been observed in the dorsolateral prefrontal cortex of patients with schizophrenia compared to healthy controls (Doyle et al., 2017). Over-representation of these insertions appeared within the gene ontologies called "cell projection" and "postsynaptic membrane," suggesting their role in the brain development. In some studies, LINE-1 methylation was tested in peripheral blood leukocytes of patients with schizophrenia, providing mixed findings (Misiak et al., 2015; Li et al., 2018; Fachim et al., 2019; Kalayasiri et al., 2019). The study by our group revealed lower LINE-1 methylation only in patients with first-episode schizophrenia and a positive history of childhood trauma. Among various childhood adversities, emotional trauma was most strongly associated with the LINE-1 methylation status. These results are in agreement with a previous study, showing that the LINE-1 methylation might be involved in resilience and susceptibility to develop post-traumatic stress disorder (Rusiecki et al., 2012). Moreover, increased expression of LINE-1 in response to stress has been reported in various cell lines (Li and Schmid, 2001; Capomaccio et al., 2010). Lower LINE-1 methylation levels in patients with schizophrenia and bipolar disorder were also reported by Li et al. (2018). Other studies revealed hypermethylation of LINE-1 sequences in patients with first-episode psychosis, paranoid schizophrenia and methamphetamine-induced paranoia (Fachim et al., 2019; Kalayasiri et al., 2019).

Mood Disorders

A recent systematic review indicates that prenatal infections might impact the risk of bipolar disorder (Marangoni et al., 2016). However, this observation is based on a lower number of studies compared to studies addressing the impact of prenatal infections on schizophrenia risk. There is evidence that influenza infection during pregnancy is associated with a fourfold increase in the risk of bipolar disorder in the offspring (Parboosing et al., 2013). Another study demonstrated that prenatal flu exposure increases the risk of bipolar disorder with psychotic features (Canetta et al., 2014). However, no association was found between prenatal infections with HHV-1, HHV-2, *Cytomegalovirus* or *Toxoplasma gondii* and bipolar disorder risk (Mortensen et al., 2011). Maternal infections in the second trimester might also contribute to the development of depressive symptoms in the adolescent offspring (Murphy et al., 2017). However, the impact of specific infectious agents has not been tested so far.

Although all major mental disorders are characterized by co-existing subclinical inflammation, some differences, regarding specific pro-inflammatory markers can be indicated (Goldsmith et al., 2016; Misiak et al., 2019). Therefore, it might be hypothesized that the mechanisms leading to subclinical inflammation in bipolar disorder, major depression and schizophrenia-spectrum disorders are different. However, studies investigating expression of TEs do not support this hypothesis. For instance, over-expression of HERV-K sequences has been reported in brain samples of patients with bipolar disorder and schizophrenia (Frank et al., 2005). Similarly, decreased expression of the HERV-W *gag* protein has been reported in the cingulate gyrus and hippocampus of patients with schizophrenia, bipolar disorder, and major depression (Weis et al., 2007). Finally, hypomethylation of LINE-1 elements in peripheral blood has been observed in patients with bipolar disorder and schizophrenia (Li et al., 2018). Some differences have been detected with respect to the expression of HERV-W sequences. Indeed, Perron et al. (2012) found elevated transcription levels of the HERV-W *env* sequence in the peripheral blood of patients with bipolar disorder and schizophrenia compared to healthy controls. Expression levels of the HERV-W *env* sequence were also significantly higher in patients with bipolar disorder than in those with schizophrenia.

Alzheimer's Disease

There is a general consensus that aging processes are associated with progressive loss of global DNA methylation and site-specific DNA hypermethylation (Jung and Pfeifer, 2015). Similarly, TEs are subjected to profound epigenetic modifications during aging that appear in the context of organismal and cellular senescence (Cardelli, 2018). For instance, age-related loss of Alu and HERV-K methylation has been well-documented (Bollati et al., 2009;

Jintaridth and Mutirangura, 2010; Gentilini et al., 2013). Moreover, it has been found that the expression of HERV-H, HERV-K and HERV-W changes during the lifespan with distinct patterns (Balestrieri et al., 2015). Importantly, the study by Gentilini et al. (2013) demonstrated that age-related loss of Alu methylation was less apparent in the offspring of centenarians, suggesting the effects of genetic factors associated with longevity. In turn, studies investigating changes of LINE-1 methylation have provided mixed findings (Bollati et al., 2009; Talens et al., 2012; Cho et al., 2015). Finally, there is evidence that chromatin of Alu, SVA, and LINE-1 becomes relatively more open in senescent cells (De Cecco et al., 2013).

Age-related changes in epigenetic modifications of TEs have provided basis for investigating alterations of these processes in Alzheimer's disease. In a single study of early-onset Alzheimer's disease family, it has been reported that large genomic rearrangements might affect the presenilin-1 gene via the mechanisms involving recombination stimulated by the Alu sequence (Hiltunen et al., 2000). However, subsequent studies have not provided compelling evidence regarding the contribution of TEs to the etiology of Alzheimer's disease. There is only one study, showing increased LINE-1 methylation in peripheral blood leukocytes of patients with Alzheimer's disease, especially those with better cognitive performance, compared to healthy controls (Bollati et al., 2011). However, the authors did not find significant between-group differences in the levels of Alu and SAT-α methylation. Other studies did not confirm these findings regarding LINE-1 methylation (Hernández et al., 2014; Protasova et al., 2017). Alterations of other TEs in patients with Alzheimer's disease have not been tested so far.

SUMMARY OF EVIDENCE AND FUTURE DIRECTIONS

Although specific retrotransposition events that may account for mental disorders in the manner observed in case of Mendelian diseases have not been identified so far, accumulating evidence indicates the involvement of altered expression and epigenetic regulation of TEs in the pathophysiology of schizophrenia, mood disorders and ASD. Most consistently, previous studies indicate altered expression of HERVs and methylation of LINE-1 sequences. However, specific findings are similar in patients with various mental disorders and thus their use as biomarkers is largely limited. Moreover, the direction of causality is yet to be determined. For instance, it cannot be excluded that altered expression of HERV appears as a consequence of other epigenetic dysregulations that are widely observed in mental disorders. Additionally, severe mental disorders, including schizophrenia and mood disorders, are associated with high prevalence rates of somatic comorbidities, including autoimmune diseases, type 2 diabetes and cardiovascular diseases that have also been associated with altered epigenetic regulation of TEs (Cash et al., 2011; De Hert et al., 2011; Misiak et al., 2013; Nestler et al., 2016; Zhao et al., 2018). Interestingly, there are studies showing that the expression of various HERV sequences appears in a certain subgroup of patients with schizophrenia but not in healthy controls. These findings are consistent with previous studies, showing that immune alterations can be observed only in a subgroup of patients characterized by poor response to treatment and support the concept of psychosis subtypes (Frydecka et al., 2015; Mondelli et al., 2015; Fillman et al., 2016). Other clinical correlates of subclinical inflammation in schizophrenia include, i.e., more severe cognitive deficits (Misiak et al., 2017b), persistent negative symptoms (Goldsmith et al., 2018) and certain neurostructural abnormalities (Najjar and Pearlman, 2015). However, so far studies investigating expression and epigenetic regulation of TEs in schizophrenia have been based on relatively small samples without comprehensive clinical assessment. Similarly, studies investigating the expression of TEs in patients with bipolar disorder did not control for mood status and a severity of psychopathological symptoms.

Another important point is that causal inferences between TEs and mental disorders cannot be established. Firstly, it remains unknown what are the critical periods when alterations in epigenetic regulation and expression of TEs appear. Therefore, future studies should examine epigenetic processes that regulate expression of TEs in patients at early stages of mental disorders or individuals from clinical high risk groups. This is particularly important since several lifestyle characteristics that are highly prevalent among patients with mental disorders, e.g., cigarette smoking and poor dietary habits, might impact TEs *per se* (Miglino et al., 2014; Miousse et al., 2015). Secondly, the role of HERVs in shaping innate immunity also remains problematic with respect to understanding causal associations. On one side, expression of HERVs might condition resistance to exogenous infections; on the other, exogenous retroviruses have been found to impact the expression of HERVs. Therefore, it remains unknown whether altered expression profiles of HERVs in mental disorders represent cause or consequence of exogenous infections. Future studies should necessarily examine the biological nature and the extent of associations between immune alterations in mental disorders and expression of various TEs.

Finally, more global concordance patterns of different TEs expression in mental disorders are yet to be examined: this could provide further insight into specificity of methylation patterns across different TEs and provide additional information of their use as potential biomarkers. At this point, it is important to note that similar DNA methylation patterns have been described in brain samples and peripheral blood leukocytes of patients with schizophrenia (Van Den Oord et al., 2016).

Another direction for the field is to disentangle the effects of stressful life events on epigenetic regulation of TEs expression. Early-life stress is a known risk factor for mood and psychotic disorders as well as correlates with a number of biological dysregulations in adults (Misiak et al., 2017a; Bielawski et al., 2019; Jaworska-Andryszewska and Rybakowski, 2019). Acute stress has been found to increase the levels of H3K9me3 as well as decrease the levels of H3K9me1 and H3K27me3 in the dentate gyrus and the CA1 layer of the hippocampus in rats (Milne et al., 2009). In turn, chronic restraint stress for 21 days mildly increased the levels of H3Kme4 and reduced the levels of H3K9me3 in the dentate gyrus. Treatment with

fluoxetine reversed changes in the levels of H3K9me3 during chronic restraint stress. More specifically, the same group found that acute stress had increased H3K9me3 enrichment at SINEs (Baker et al., 2012). In turn, our group found lower methylation of LINE-1 sequences in peripheral blood leukocytes of patients with first-episode schizophrenia reporting a positive history of childhood trauma (Misiak et al., 2015). In light of these findings, future studies should further examine the effects of stress on the expression of TEs in patients from various clinical groups and preclinical studies could contribute to this aim.

AUTHOR CONTRIBUTIONS

BM and MS conceived the concept of this article. BM wrote the Sections "Introduction", "Insights Into Potential Mechanisms Underlying the Role of TEs in Mental Disorders", "Schizophrenia-Spectrum Disorders" and "Mood Disorders". LR wrote the Section "Autism Spectrum Disorders (ASD)". MS prepared the Sections "Brief Overview of TEs in the Human Genome – Classification and Nomenclature" and "Alzheimer's Disease". BM, LR, and MS prepared the Section "Summary of Evidence and Future Directions". All authors contributed to the manuscript revision, read, and approved the submitted version.

FUNDING

This work was supported by the statutory project funded by the Wrocław Medical University, Wrocław, Poland (Task Number: ST-A.290.17.032).

REFERENCES

Aporntewan, C., Phokaew, C., Piriyapongsa, J., Ngamphiw, C., Ittiwut, C., Tongsima, S., et al. (2011). Hypomethylation of intragenic LINE-1 represses transcription in cancer cells through AGO2. *PLoS One* 6:e17934. doi: 10.1371/journal.pone.0017934

Arias, I., Sorlozano, A., Villegas, E., Luna, J., de, D., McKenney, K., et al. (2012). Infectious agents associated with schizophrenia: a meta-analysis. *Schizophr. Res.* 136, 128–136. doi: 10.1016/j.schres.2011.10.026

Baillie, J. K., Barnett, M. W., Upton, K. R., Gerhardt, D. J., Richmond, T. A., De Sapio, F., et al. (2011). Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* 479, 534–537. doi: 10.1038/nature10531

Baker, M. E. R., Pfaff, D. W., McEwen, B. S., Seligsohn, M., Hunter, R. G., Datson, N. A., et al. (2012). Acute stress and hippocampal histone H3 lysine 9 trimethylation, a retrotransposon silencing response. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17657–17662. doi: 10.1073/pnas.1215810109

Balestrieri, E., Arpino, C., Matteucci, C., Sorrentino, R., Pica, F., Alessandrelli, R., et al. (2012). HERVs expression in autism spectrum disorders. *PLoS One* 7:e48831. doi: 10.1371/journal.pone.0048831

Balestrieri, E., Cipriani, C., Matteucci, C., Capodicasa, N., Pilika, A., Korea, I., et al. (2016). Transcriptional activity of human endogenous retrovirus in Albanian children with autism spectrum disorders. *New Microbiol.* 39, 228–231.

Balestrieri, E., Pica, F., Matteucci, C., Zenobi, R., Sorrentino, R., Argaw-Denboba, A., et al. (2015). Transcriptional activity of human endogenous retroviruses in human peripheral blood mononuclear cells. *Biomed Res. Int.* 2015:164529. doi: 10.1155/2015/164529

Bannert, N., and Kurth, R. (2006). The evolutionary dynamics of human endogenous retroviral families. *Annu. Rev. Genomics Hum. Genet.* 7, 149–173. doi: 10.1146/annurev.genom.7.080505.115700

Bielawski, T., Misiak, B., Moustafa, A., and Frydecka, D. (2019). Epigenetic mechanisms, trauma, and psychopathology: targeting chromatin remodeling complexes. *Rev. Neurosci.* doi: 10.1515/revneuro-2018-0055 [Epub ahead of print].

Blond, J. L., Lavillette, D., Cheynet, V., Bouton, O., Oriol, G., Chapel-Fernandes, S., et al. (2000). An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placenta and fuses cells expressing the type D mammalian retrovirus receptor. *J. Virol.* 74, 3321–3329. doi: 10.1128/jvi.74.7.3321-3329.2000

Bollati, V., Galimberti, D., Pergoli, L., Dalla Valle, E., Barretta, F., Cortini, F., et al. (2011). DNA methylation in repetitive elements and Alzheimer disease. *Brain Behav. Immun.* 25, 1078–1083. doi: 10.1016/j.bbi.2011.01.017

Bollati, V., Schwartz, J., Wright, R., Litonjua, A., Tarantini, L., Suh, H., et al. (2009). Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mech. Ageing Dev.* 130, 234–239. doi: 10.1016/j.mad.2008.12.003

Bundo, M., Toyoshima, M., Okada, Y., Akamatsu, W., Ueda, J., Nemoto-Miyauchi, T., et al. (2014). Increased L1 retrotransposition in the neuronal genome in schizophrenia. *Neuron* 81, 306–313. doi: 10.1016/j.neuron.2013.10.053

Callinan, P. A., and Batzer, M. A. (2006). Retrotransposable elements and human disease. *Genome Dis.* 1, 104–115. doi: 10.1159/000092503

Canetta, S. E., Bao, Y., Co, M. D. T., Ennis, F. A., Cruz, J., Terajima, M., et al. (2014). Serological documentation of maternal influenza exposure and bipolar disorder in adult offspring. *Am. J. Psychiatry* 171, 557–563. doi: 10.1176/appi.ajp.2013.13070943

Capomaccio, S., Verini-Supplizi, A., Galla, G., Vitulo, N., Barcaccia, G., Felicetti, M., et al. (2010). Transcription of LINE-derived sequences in exercise-induced stress in horses. *Anim. Genet.* 41(Suppl. 2), 23–27. doi: 10.1111/j.1365-2052.2010.02094.x

Cardelli, M. (2018). The epigenetic alterations of endogenous retroelements in aging. *Mech. Ageing Dev.* 174, 30–46. doi: 10.1016/j.mad.2018.02.002

Cardno, A. G., and Gottesman, I. I. (2000). Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. *Am. J. Med. Genet. Semin. Med. Genet.* 97, 12–17. doi: 10.1002/(sici)1096-8628(200021)97:1<12::aid-ajmg3>3.3.co;2-l

Cash, H. L., Mcgarvey, S. T., Houseman, E. A., Marsit, C. J., Hawley, N. L., Lambert-Messerlian, G. M., et al. (2011). Cardiovascular disease risk factors and DNA methylation at the LINE-1 repeat region in peripheral blood from Samoan Islanders. *Epigenetics* 6, 1257–1264. doi: 10.4161/epi.6.10.17728

Chen, Y., Yan, Q., Zhou, P., Li, S., and Zhu, F. (2018). HERV-W env regulates calcium influx via activating TRPC3 channel together with depressing DISC1 in human neuroblastoma cells. *J. Neurovirol.* 25, 101–113. doi: 10.1007/s13365-018-0692-7

Cho, Y. H., Woo, H. D., Jang, Y., Porter, V., Christensen, S., Hamilton, R. F., et al. (2015). The association of LINE-1 hypomethylation with age and centromere positive micronuclei in human lymphocytes. *PLoS One* 10:e0133909. doi: 10.1371/journal.pone.0133909

Cipriani, C., Ricceri, L., Matteucci, C., De Felice, A., Tartaglione, A. M., Argaw-Denboba, A., et al. (2018). High expression of Endogenous Retroviruses from intrauterine life to adulthood in two mouse models of Autism Spectrum Disorders. *Sci. Rep.* 8:629. doi: 10.1038/s41598-017-19035-w

Cordaux, R., and Batzer, M. A. (2009). The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.* 10, 691–703. doi: 10.1038/nrg2640

Coufal, N. G., Garcia-perez, J. L., Peng, G. E., Yeo, G. W., Mu, Y., Lovci, M. T., et al. (2009). L1 retrotransposition in human neural progenitor cells Nicole. *Nature* 460, 1127–1131. doi: 10.1038/nature08248.L1

Darby, M. M., and Sabunciyan, S. (2014). Repetitive elements and epigenetic marks in behavior and psychiatric disease. *Adv. Genet.* 86, 185–252. doi: 10.1016/B978-0-12-800222-3.00009-7

Davies, G., Welham, J., Chant, D., Torrey, E. F., and McGrath, J. (2003). A systematic review and meta-analysis of Northern Hemisphere season of birth studies in schizophrenia. *Schizophr. Bull.* 29, 587–593. doi: 10.1093/oxfordjournals.schbul.a007030

Day, D. S., Luquette, L. J., Park, P. J., and Kharchenko, P. V. (2010). Estimating enrichment of repetitive elements from high-throughput sequence data. *Genome Biol.* 11:R69. doi: 10.1186/gb-2010-11-6-r69

De Cecco, M., Criscione, S. W., Peckham, E. J., Hillenmeyer, S., Hamm, E. A., Manivannan, J., et al. (2013). Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. *Aging Cell* 12, 247–256. doi: 10.1111/acel.12047

De Hert, M., Correll, C. U., Bobes, J., Cetkovich-Bakmas, M., Cohen, D. A. N., Asai, I., et al. (2011). Physical illness in patients with severe mental disorders. I. Prevalence, impact of medications and disparities in health care. *World Psychiatry* 10, 52–77. doi: 10.1002/j.2051-5545.2011.tb00014.x

Dewannieux, M., and Heidmann, T. (2005). LINEs, SINEs and processed pseudogenes: parasitic strategies for genome modeling. *Cytogenet. Genome Res.* 110, 35–48. doi: 10.1159/000084936

Diem, O., Schäffner, M., Seifarth, W., and Leib-Mösch, C. (2012). Influence of antipsychotic drugs on human endogenous retrovirus (HERV) transcription in brain cells. *PLoS One* 7:e30054. doi: 10.1371/journal.pone.0030054

Doyle, G. A., Crist, R. C., Karatas, E. T., Hammond, M. J., Ewing, A. D., Ferraro, T. N., et al. (2017). Analysis of LINE-1 elements in DNA from postmortem brains of individuals with schizophrenia. *Neuropsychopharmacology* 42, 2602–2611. doi: 10.1038/npp.2017.115

Fachim, H., Loureiro, C., Corsi-Zuelli, F., Shuhama, R., Louzada-Junior, P., Menezes, P., et al. (2019). GRIN2B promoter methylation deficits in early-onset schizophrenia and its association with cognitive function. *Epigenomics* 11, 401–410. doi: 10.2217/epi-2018-0127

Fernandes, B. S., Steiner, J., Bernstein, H. G., Dodd, S., Pasco, J. A., Dean, O. M., et al. (2016). C-reactive protein is increased in schizophrenia but is not altered by antipsychotics: meta-analysis and implications. *Mol. Psychiatry* 21, 554–564. doi: 10.1038/mp.2015.87

Fillman, S. G., Weickert, T. W., Lenroot, R. K., Catts, S. V., Bruggemann, J. M., Catts, V. S., et al. (2016). Elevated peripheral cytokines characterize a subgroup of people with schizophrenia displaying poor verbal fluency and reduced Broca's area volume. *Mol. Psychiatry* 21, 1090–1098. doi: 10.1038/mp.2015.90

Frank, O., Giehl, M., Hehlmann, R., Zheng, C., Leib-Mosch, C., and Seifarth, W. (2005). Human endogenous retrovirus expression profiles in samples from brains of patients with schizophrenia and bipolar disorders. *J. Virol.* 79, 10890–10901. doi: 10.1128/jvi.79.17.10890-10901.2005

Frydecka, D., Krzystek-Korpacka, M., Lubeiro, A., Stramecki, F., Stańczykiewicz, B., Aleksander Beszłej, J., et al. (2018). Profiling inflammatory signatures of schizophrenia: a cross-sectional and meta-analysis study. *Brain Behav. Immun.* 71, 28–36. doi: 10.1016/j.bbi.2018.05.002

Frydecka, D., Misiak, B., Pawlak-Adamska, E., Karabon, L., Tomkiewicz, A., Sedlaczek, P., et al. (2015). Interleukin-6: the missing element of the neurocognitive deterioration in schizophrenia? The focus on genetic underpinnings, cognitive impairment and clinical manifestation. *Eur. Arch. Psychiatry Clin. Neurosci.* 265, 449–459. doi: 10.1007/s00406-014-0533-5

Gentilini, D., Mari, D., Castaldi, D., Remondini, D., Ogliari, G., Ostan, R., et al. (2013). Role of epigenetics in human aging and longevity: genome-wide DNA methylation profile in centenarians and centenarians' offspring. *Age* 35, 1961–1973. doi: 10.1007/s11357-012-9463-1

Goldsmith, D. R., Haroon, E., Miller, A. H., Strauss, G. P., Buckley, P. F., and Miller, B. J. (2018). TNF-α and IL-6 are associated with the deficit syndrome and negative symptoms in patients with chronic schizophrenia. *Schizophr. Res.* 199, 281–284. doi: 10.1016/j.schres.2018.02.048

Goldsmith, D. R., Rapaport, M. H., and Miller, B. J. (2016). A meta-analysis of blood cytokine network alterations in psychiatric patients: comparisons between schizophrenia, bipolar disorder and depression. *Mol. Psychiatry* 199, 281–284.

Grandi, N., and Tramontano, E. (2017). Type W human endogenous retrovirus (HERV-W) integrations and their mobilization by L1 machinery: contribution to the human transcriptome and impact on the host physiopathology. *Viruses* 9:E162. doi: 10.3390/v9070162

Grandi, N., and Tramontano, E. (2018). Human endogenous retroviruses are ancient acquired elements still shaping innate immune responses. *Front. Immunol.* 9:2039. doi: 10.3389/fimmu.2018.02039

Griffiths, D. J. (2001). Endogenous retroviruses in the human genome sequence. *Genome Biol.* 2, reviews1017.1–reviews1017.5.

Hernández, H. G., Mahecha, M. F., Mejía, A., Arboleda, H., and Forero, D. A. (2014). Global long interspersed nuclear element 1 DNA methylation in a Colombian sample of patients with late-onset Alzheimer's disease. *Am. J. Alzheimers Dis. Other Demen.* 29, 50–53. doi: 10.1177/1533317513505132

Hiltunen, M., Helisalmi, S., Mannermaa, A., Alafuzoff, I., Koivisto, A. M., Lehtovirta, M., et al. (2000). Identification of a novel 4.6-kb genomic deletion in presenilin-1 gene which results in exclusion of exon 9 in a Finnish early onset Alzheimer's disease family: an Alu core sequence-stimulated recombination? *Eur. J. Hum. Genet.* 8, 259–266. doi: 10.1038/sj.ejhg.5200423

Houck, C. M., Rinehart, F. P., and Schmid, C. W. (1979). A ubiquitous family of repeated DNA sequences in the human genome. *J. Mol. Biol.* 132, 289–306. doi: 10.1016/0022-2836(79)90261-4

Huang, W., Li, S., Hu, Y., Yu, H., Luo, F., Zhang, Q., et al. (2011). Implication of the env gene of the human endogenous retrovirus W family in the expression of BDNF and DRD3 and development of recent-onset schizophrenia. *Schizophr. Bull.* 37, 988–1000. doi: 10.1093/schbul/sbp166

Huang, W. J., Liu, Z. C., Wei, W., Wang, G. H., Wu, J. G., and Zhu, F. (2006). Human endogenous retroviral pol RNA and protein detected and identified in the blood of individuals with schizophrenia. *Schizophr. Res.* 83, 193–199. doi: 10.1016/j.schres.2006.01.007

Hurst, T. P., and Magiorkinis, G. (2015). Activation of the innate immune response by endogenous retroviruses. *J. Gen. Virol.* 96, 1207–1218. doi: 10.1099/jgv.0.000017

Jaworska-Andryszewska, P., and Rybakowski, J. K. (2019). Childhood trauma in mood disorders: neurobiological mechanisms and implications for treatment. *Pharmacol. Rep.* 71, 112–120. doi: 10.1016/j.pharep.2018.10.004

Jintaridth, P., and Mutirangura, A. (2010). Distinctive patterns of age-dependent hypomethylation in interspersed repetitive sequences. *Physiol. Genomics* 41, 194–200. doi: 10.1152/physiolgenomics.00146.2009

Jung, M., and Pfeifer, G. P. (2015). Aging and DNA methylation. *BMC Biol.* 13:7. doi: 10.1186/s12915-015-0118-4

Kaer, K., and Speek, M. (2013). Retroelements in human disease. *Gene* 518, 231–241. doi: 10.1016/j.gene.2013.01.008

Kalayasiri, R., Kraijak, K., Mutirangura, A., and Maes, M. (2019). Paranoid schizophrenia and methamphetamine-induced paranoia are both characterized by a similar LINE-1 partial methylation profile, which is more pronounced in paranoid schizophrenia. *Schizophr. Res.* 208, 221–227. doi: 10.1016/j.schres.2019.02.015

Karlsson, H., McArthur, J., Schroder, J., Torrey, E. F., Bachmann, S., and Yolken, R. H. (2002). Retroviral RNA identified in the cerebrospinal fluids and brains of individuals with schizophrenia. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4634–4639. doi: 10.1073/pnas.061021998

Karlsson, H., Schröder, J., Bachmann, S., Bottmer, C., and Yolken, R. H. (2004). HERV-W-related RNA detected in plasma from individuals with recent-onset schizophrenia or schizoaffective disorder. *Mol. Psychiatry* 9, 12–13. doi: 10.1038/sj.mp.4001439

Karpiński, P., Frydecka, D., Sasiadek, M. M., and Misiak, B. (2016). Reduced number of peripheral natural killer cells in schizophrenia but not in bipolar disorder. *Brain Behav. Immun.* 54, 194–200. doi: 10.1016/j.bbi.2016.02.005

Karpiński, P., Samochowiec, J., Frydecka, D., Sąsiadek, M. M., and Misiak, B. (2018). Further evidence for depletion of peripheral blood natural killer cells in patients with schizophrenia: a computational deconvolution study. *Schizophr. Res.* 54, 194–200. doi: 10.1016/j.schres.2018.04.026

Kazazian, H. H. (2014). Processed pseudogene insertions in somatic cells. *Mob. DNA* 5:20. doi: 10.1186/1759-8753-5-20

Khandaker, G. M., Zimbron, J., Lewis, G., and Jones, P. B. (2013). Prenatal maternal infection, neurodevelopment and adult schizophrenia: a systematic review of population-based studies. *Psychol. Med.* 43, 239–257. doi: 10.1017/S0033291712000736

Kitkumthorn, N., and Mutirangura, A. (2011). Long interspersed nuclear element-1 hypomethylation in cancer: biology and clinical applications. *Clin. Epigenetics* 2, 315–330. doi: 10.1007/s13148-011-0032-8

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi: 10.1038/35057062

Lapp, H. E., and Hunter, R. G. (2016). The dynamic genome: transposons and environmental adaptation in the nervous system. *Epigenomics* 8, 237–249. doi: 10.2217/epi.15.107

Laska, M. J., Troldborg, A., Hauge, E. M., Bahrami, S., and Stengaard-Pedersen, K. (2017). Human endogenous retroviral genetic element with immunosuppressive activity in both human autoimmune diseases and experimental arthritis. *Arthritis Rheumatol.* 69, 398–409. doi: 10.1002/art.39867

Li, F., Sabunciyan, S., Yolken, R. H., Lee, D., Kim, S., and Karlsson, H. (2019). Transcription of human endogenous retroviruses in human brain by RNA-seq analysis. *PLoS One* 14:e0207353. doi: 10.1371/journal.pone.0207353

Li, S., Yang, Q., Hou, Y., Jiang, T., Zong, L., Wang, Z., et al. (2018). Hypomethylation of LINE-1 elements in schizophrenia and bipolar disorder. *J. Psychiatr. Res.* 107, 68–72. doi: 10.1016/j.jpsychires.2018.10.009

Li, T. H., and Schmid, C. W. (2001). Differential stress induction of individual Alu loci: implications for transcription and retrotransposition. *Gene* 276, 135–141. doi: 10.1016/s0378-1119(01)00637-0

Mak, M., Samochowiec, J., Frydecka, D., Pełka-Wysiecka, J., Szmida, E., Karpiński, P., et al. (2019). First-episode schizophrenia is associated with a reduction of HERV-K methylation in peripheral blood. *Psychiatry Res.* 271, 459–463. doi: 10.1016/j.psychres.2018.12.012

Mameli, G., Astone, V., Arru, G., Marconi, S., Lovato, L., Serra, C., et al. (2007). Brains and peripheral blood mononuclear cells of multiple sclerosis (MS) patients hyperexpress MS-associated retrovirus/HERV-W endogenous retrovirus, but not human herpesvirus 6. *J. Gen. Virol.* 88, 264–274. doi: 10.1099/vir.0.81890-0

Marangoni, C., Hernandez, M., and Faedda, G. L. (2016). The role of environmental exposures as risk factors for bipolar disorder: a systematic review of longitudinal studies. *J. Affect. Disord.* 193, 165–174. doi: 10.1016/j.jad.2015.12.055

McGrath, J. J., and Welham, J. L. (1999). Season of birth and schizophrenia: a systematic review and meta-analysis of data from the Southern Hemisphere. *Schizophr. Res.* 35, 237–242. doi: 10.1016/s0920-9964(98)00139-x

McGuffin, P., Rijsdijk, F., Andrew, M., Sham, P., Katz, R., and Cardno, A. (2003). The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Arch. Gen. Psychiatry* 60, 497–502.

Melbourne, J. K., Chase, K. A., Feiner, B., Rosen, C., and Sharma, R. P. (2018). Long non-coding and endogenous retroviral RNA levels are associated with proinflammatory cytokine mRNA expression in peripheral blood cells: implications for schizophrenia. *Psychiatry Res.* 262, 465–468. doi: 10.1016/j.psychres.2017.09.025

Meredith, R. M. (2015). Sensitive and critical periods during neurotypical and aberrant neurodevelopment: a framework for neurodevelopmental disorders. *Neurosci. Biobehav. Rev.* 50, 180–188. doi: 10.1016/j.neubiorev.2014.12.001

Mi, S., Lee, X., Li, X., Veldman, G. M., Finnerty, H., Racie, L., et al. (2000). Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403, 785–789. doi: 10.1038/35001608

Mighdoll, M. I., Tao, R., Kleinman, J. E., and Hyde, T. M. (2015). Myelin, myelin-related disorders, and psychosis. *Schizophr. Res.* 161, 85–93. doi: 10.1016/j.schres.2014.09.040

Mighell, A. J., Markham, A. F., and Robinson, P. A. (1997). Alu sequences. *FEBS Lett.* 417, 1–5. doi: 10.1016/s0014-5793(97)01259-3

Miglino, N., Tamm, M., and Borger, P. (2014). Transposable element LINE1 is activated after exposure to cigarette smoke in primary human lung fibroblasts. *Eur. Respir. J.* 44:S2023.

Miller, B. J., Buckley, P., Seabolt, W., Mellor, A., and Kirkpatrick, B. (2011). Meta-analysis of cytokine alterations in schizophrenia: clinical status and antipsychotic effects. *Biol. Psychiatry* 70, 663–671. doi: 10.1016/j.biopsych.2011.04.013

Miller, B. J., Gassama, B., Sebastian, D., Buckley, P., and Mellor, A. (2013). Meta-analysis of lymphocytes in schizophrenia: clinical status and antipsychotic effects. *Biol. Psychiatry* 73, 993–999. doi: 10.1016/j.biopsych.2012.09.007

Milne, T. A., McEwen, B. S., Hunter, R. G., Pfaff, D. W., and McCarthy, K. J. (2009). Regulation of hippocampal H3 histone methylation by acute and chronic stress. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20912–20917. doi: 10.1073/pnas.0911143106

Miousse, I. R., Chalbot, M. C. G., Lumen, A., Ferguson, A., Kavouras, I. G., and Koturbash, I. (2015). Response of transposable elements to environmental stressors. *Mutat. Res. Rev. Mutat. Res.* 765, 19–39. doi: 10.1016/j.mrrev.2015.05.003

Misiak, B., Frydecka, D., Piotrowski, P., and Kiejna, A. (2013). The multidimensional nature of metabolic syndrome in schizophrenia: lessons from studies of one-carbon metabolism and DNA methylation. *Epigenomics* 5, 317–329. doi: 10.2217/epi.13.22

Misiak, B., Frydecka, D., and Rybakowski, J. K. (2016). Editorial: endophenotypes for schizophrenia and mood disorders: implications from genetic, biochemical, cognitive, behavioral, and neuroimaging studies. *Front. Psychiatry* 7:83. doi: 10.3389/fpsyt.2016.00083

Misiak, B., Frydecka, D., Stańczykiewicz, B., and Samochowiec, J. (2019). Editorial: peripheral markers of immune response in major psychiatric disorders: where are we now and where do we want to be? *Front. Psychiatry* 10:5. doi: 10.3389/fpsyt.2019.00005

Misiak, B., Krefft, M., Bielawski, T., Moustafa, A. A., Sąsiadek, M. M., and Frydecka, D. (2017a). Toward a unified theory of childhood trauma and psychosis: a comprehensive review of epidemiological, clinical, neuropsychological and biological findings. *Neurosci. Biobehav. Rev.* 75, 393–406. doi: 10.1016/j.neubiorev.2017.02.015

Misiak, B., Stańczykiewicz, B., Kotowicz, K., Rybakowski, J. K., Samochowiec, J., and Frydecka, D. (2017b). Cytokines and C-reactive protein alterations with respect to cognitive impairment in schizophrenia and bipolar disorder: a systematic review. *Schizophr. Res.* 192, 16–29. doi: 10.1016/j.schres.2017.04.015

Misiak, B., Szmida, E., Karpiński, P., Loska, O., Sąsiadek, M. M., and Frydecka, D. (2015). Lower LINE-1 methylation in first-episode schizophrenia patients with the history of childhood trauma. *Epigenomics* 7, 1275–1285. doi: 10.2217/epi.15.68

Mondelli, V., Ciufolini, S., Murri, M. B., Bonaccorso, S., Di Forti, M., Giordano, A., et al. (2015). Cortisol and inflammatory biomarkers predict poor treatment response in first episode psychosis. *Schizophr. Bull.* 41, 1162–1170. doi: 10.1093/schbul/sbv028

Morozov, V. A., Dao Thi, V. L., and Denner, J. (2013). The transmembrane protein of the human endogenous retrovirus - K (HERV-K) modulates cytokine release and gene expression. *PLoS One* 8:e70399. doi: 10.1371/journal.pone.0070399

Mortensen, P. B., Pedersen, C. B., Mcgrath, J. J., Hougaard, D. M., Nørgaard-Petersen, B., Mors, O., et al. (2011). Neonatal antibodies to infectious agents and risk of bipolar disorder: a population-based case-control study. *Bipolar Disord.* 13, 624–629. doi: 10.1111/j.1399-5618.2011.00962.x

Munoz-Lopez, M., and Garcia-Perez, J. (2010). DNA transposons: nature and applications in genomics. *Curr. Genomics* 11, 115–128. doi: 10.2174/138920210790886871

Muotri, A. R., Chu, V. T., Marchetto, M. C. N., Deng, W., Moran, J. V., and Gage, F. H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* 435, 903–910. doi: 10.1038/nature03663

Muotri, A. R., Marchetto, M. C. N., Coufal, N. G., Oefner, R., Yeo, G., Nakashima, K., et al. (2010). L1 retrotransposition in neurons is modulated by MeCP2. *Nature* 468, 443–446. doi: 10.1038/nature09544

Murphy, S. K., Fineberg, A. M., Maxwell, S. D., Alloy, L. B., Zimmermann, L., Krigbaum, N. Y., et al. (2017). Maternal infection and stress during pregnancy and depressive symptoms in adolescent offspring. *Psychiatry Res.* 257, 102–110. doi: 10.1016/j.psychres.2017.07.025

Mutirangura, A. (ed.). (2019). "A hypothesis to explain how the DNA of elderly people is prone to damage: genome-wide hypomethylation drives genomic instability in the elderly by reducing youth-associated gnome-stabilizing DNA gaps," in *Epigenetics [Working Title]*, (London: Intechopen). doi: 10.5772/intechopen.83372

Nagy, C., and Turecki, G. (2012). Sensitive periods in epigenetics: bringing us closer to complex behavioral phenotypes. *Epigenomics* 4, 445–457. doi: 10.2217/epi.12.37

Najjar, S., and Pearlman, D. M. (2015). Neuroinflammation and white matter pathology in schizophrenia: systematic review. *Schizophr. Res.* 161, 102–112. doi: 10.1016/j.schres.2014.04.041

Nestler, E. J., Peña, C. J., Kundakovic, M., Mitchell, A., and Akbarian, S. (2016). Epigenetic basis of mental illness. *Neuroscientist* 22, 447–463. doi: 10.1177/1073858415608147

Oja, M., Peltonen, J., Blomberg, J., and Kaski, S. (2008). Methods for estimating human endogenous retrovirus activities from EST databases. *BMC Bioinformatics* 8(Suppl. 2):S11. doi: 10.1186/1471-2105-8-S2-S11

Ostertag, E. M., Goodier, J. L., Zhang, Y., and Kazazian, H. H. (2003). SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am. J. Hum. Genet.* 73, 1444–1451. doi: 10.1086/380207

Parboosing, R., Bao, Y., Shen, L., Schaefer, C. A., and Brown, A. S. (2013). Gestational influenza and bipolar disorder in adult offspring. *JAMA Psychiatry* 70, 677–685. doi: 10.1001/jamapsychiatry.2013.896

Patchsung, M., Settayanon, S., Pongpanich, M., Mutirangura, D., Jintarith, P., and Mutirangura, A. (2018). Alu siRNA to increase Alu element methylation and prevent DNA damage. *Epigenomics* 10, 175–185. doi: 10.2217/epi-2017-0096

Perron, H., Hamdani, N., Faucard, R., Lajnef, M., Jamain, S., Daban-Huard, C., et al. (2012). Molecular characteristics of Human Endogenous Retrovirus type-W in schizophrenia and bipolar disorder. *Transl. Psychiatry* 2:e201. doi: 10.1038/tp.2012.125

Perron, H., Mekaoui, L., Bernard, C., Veas, F., Stefas, I., and Leboyer, M. (2008). Endogenous retrovirus type W GAG and envelope protein antigenemia in serum of schizophrenic patients. *Biol. Psychiatry* 64, 1019–1023. doi: 10.1016/j.biopsych.2008.06.028

Poduri, A., Evrony, G. D., Cai, X., and Walsh, C. A. (2013). Somatic mutation, genomic variation, and neurological disease. *Science* 341:1237758. doi: 10.1126/science.1237758

Pray, L. A. (2008a). Functions and utility of alu jumping genes. *Nature* 1:93.

Pray, L. A. (2008b). Transposons: the jumping genes. *Nat. Educ.* 1:204.

Protasova, M. S., Gusev, F. E., Grigorenko, A. P., Kuznetsova, I. L., Rogaev, E. I., and Andreeva, T. V. (2017). Quantitative analysis of L1-retrotransposons in Alzheimer's disease and aging. *Biochemistry* 82, 962–971. doi: 10.1134/S0006297917080120

Ripke, S., Neale, B. M., Corvin, A., Walters, J. T. R., Farh, K. H., Holmans, P. A., et al. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511, 421–427. doi: 10.1038/nature13595

Rusiecki, J. A., Chen, L., Srikantan, V., Zhang, L., Yan, L., Polin, M. L., et al. (2012). DNA methylation in repetitive elements and post-traumatic stress disorder: a case-control study of US military service members. *Epigenomics* 4, 29–40. doi: 10.2217/epi.11.116

Saeliw, T., Tangsuwansri, C., Thongkorn, S., Chonchaiya, W., Suphapeetiporn, K., Mutirangura, A., et al. (2018). Integrated genome-wide Alu methylation and transcriptome profiling analyses reveal novel epigenetic regulatory networks associated with autism spectrum disorder. *Mol. Autism* 9:27. doi: 10.1186/s13229-018-0213-9

Sassaman, D. M., Dombroski, B. A., Moran, J. V., Kimberland, M. L., Naas, T. P., DeBerardinis, R. J., et al. (1997). Many human L1 elements are capable of retrotransposition. *Nat. Genet.* 16, 37–43. doi: 10.1038/ng0597-37

Schumann, G. G., Gogvadze, E. V., Osanai-Futahashi, M., Kuroki, A., Münk, C., Fujiwara, H., et al. (2010). Unique functions of repetitive transcriptomes. *Int. Rev. Cell Mol. Biol.* 285, 115–188. doi: 10.1016/B978-0-12-381047-2.00003-7

Shpyleva, S., Melnyk, S., Pavliv, O., Pogribny, I., and Jill James, S. (2018). Overexpression of LINE-1 retrotransposons in autism brain. *Mol. Neurobiol.* 55, 1740–1749. doi: 10.1007/s12035-017-0421-x

Su, J., Shao, X., Liu, H., Liu, S., Wu, Q., and Zhang, Y. (2012). Genome-wide dynamic changes of DNA methylation of repetitive elements in human embryonic stem cells and fetal fibroblasts. *Genomics* 99, 10–17. doi: 10.1016/j.ygeno.2011.10.004

Suarez, N. A., Macia, A., and Muotri, A. R. (2018). LINE-1 retrotransposons in healthy and diseased human brain. *Dev. Neurobiol.* 78, 434–455. doi: 10.1002/dneu.22567

Talens, R. P., Christensen, K., Putter, H., Willemsen, G., Christiansen, L., Kremer, D., et al. (2012). Epigenetic variation during the adult lifespan: cross-sectional and longitudinal data on monozygotic twin pairs. *Aging Cell* 11, 694–703. doi: 10.1111/j.1474-9726.2012.00835.x

Tangsuwansri, C., Saeliw, T., Thongkorn, S., Chonchaiya, W., Suphapeetiporn, K., Mutirangura, A., et al. (2018). Investigation of epigenetic regulatory networks associated with autism spectrum disorder (ASD) by integrated global LINE-1 methylation and gene expression profiling analyses. *PLoS One* 13:e0201071. doi: 10.1371/journal.pone.0201071

Tartaglione, A. M., Cipriani, C., Chiarotti, F., Perrone, B., Balestrieri, E., Matteucci, C., et al. (2018). Early behavioral alterations and increased expression of endogenous retroviruses are inherited across generations in mice prenatally exposed to valproic acid. *Mol. Neurobiol.* 56, 3736–3750. doi: 10.1007/s12035-018-1328-x

Van Den Oord, E. J. C. G., Clark, S. L., Xie, L. Y., Shabalin, A. A., Dozmorov, M. G., Kumar, G., et al. (2016). A whole methylome CpG-SNP association study of psychosis in blood and brain tissue. *Schizophr. Bull.* 42, 1018–1026. doi: 10.1093/schbul/sbv182

Vargiu, L., Rodriguez-Tomé, P., Sperber, G. O., Cadeddu, M., Grandi, N., Blikstad, V., et al. (2016). Classification and characterization of human endogenous retroviruses mosaic forms are common. *Retrovirology* 13:7. doi: 10.1186/s12977-015-0232-y

Vidaud, D., Vidaud, M., Bahnak, B. R., Siguret, V., Gispert Sanchez, S., Laurian, Y., et al. (1993). Haemophilia B due to a de novo insertion of a human-specific Alu subfamily member within the coding region of the factor IX gene. *Eur. J. Hum. Genet.* 1, 30–36.

Vigo, D., Thornicroft, G., and Atun, R. (2016). Estimating the true global burden of mental illness. *Lancet Psychiatry* 3, 171–178. doi: 10.1016/S2215-0366(15)00505-2

Wang, H., Xing, J., Grover, D., Hedges Kyudong, Han, D. J., Walker, J. A., et al. (2005). SVA elements: a hominid-specific retroposon family. *J. Mol. Biol.* 354, 994–1007. doi: 10.1016/j.jmb.2005.09.085

Weis, S., Llenos, I. C., Sabunciyan, S., Dulay, J. R., Isler, L., Yolken, R., et al. (2007). Reduced expression of human endogenous retrovirus (HERV)-W GAG protein in the cingulate gyrus and hippocampus in schizophrenia, bipolar disorder, and depression. *J. Neural Transm.* 114, 645–655. doi: 10.1007/s00702-006-0599-y

Yao, Y., Schröder, J., Nellåker, C., Bottmer, C., Bachmann, S., Yolken, R. H., et al. (2008). Elevated levels of human endogenous retrovirus-W transcripts in blood cells from patients with first episode schizophrenia. *Genes Brain Behav.* 7, 103–112.

Zhang, Z., Harrison, P. M., Liu, Y., and Gerstein, M. (2003). Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res.* 13, 2541–2558. doi: 10.1101/gr.1429003

Zhao, K., Du, J., Peng, Y., Li, P., Wang, S., Wang, Y., et al. (2018). LINE1 contributes to autoimmunity through both RIG-I- and MDA5-mediated RNA sensing pathways. *J. Autoimmun.* 90, 105–115. doi: 10.1016/j.jaut.2018.02.007

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Misiak, Ricceri and Sąsiadek. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

REVIEW
published: 28 June 2019
doi: 10.3389/fgene.2019.00621

Epigenetic IVD Tests for Personalized Precision Medicine in Cancer

Jesús Beltrán-García[1,2,3], *Rebeca Osca-Verdegal*[2,3], *Salvador Mena-Mollá*[3,4]* and *José Luis García-Giménez*[1,2,3,4]*

[1] Center for Biomedical Network Research on Rare Diseases (CIBERER), Institute of Health Carlos III, Valencia, Spain, [2] INCLIVA Biomedical Research Institute, Valencia, Spain, [3] Department of Physiology, School of Medicine and Dentistry, Universitat de València (UV), Valencia, Spain, [4] EpiDisease S.L. Spin-Off of CIBERER (ISCIII), Valencia, Spain

Epigenetic alterations play a key role in the initiation and progression of cancer. Therefore, it is possible to use epigenetic marks as biomarkers for predictive and precision medicine in cancer. Precision medicine is poised to impact clinical practice, patients, and healthcare systems. The objective of this review is to provide an overview of the epigenetic testing landscape in cancer by examining commercially available epigenetic-based *in vitro* diagnostic tests for colon, breast, cervical, glioblastoma, lung cancers, and for cancers of unknown origin. We compile current commercial epigenetic tests based on epigenetic biomarkers (i.e., DNA methylation, miRNAs, and histones) that can actually be implemented into clinical practice.

Keywords: precision medicine, epigenetic biomarker, In Vitro Diagnostic (IVD), DNA methylation, miRNA, cfDNA, circulating nucleosomes

OPEN ACCESS

Edited by:
Yun Liu,
Fudan University, China

Reviewed by:
Jorg Tost,
Institut de Biologie François Jacob, Commissariat à l'Energie Atomique et aux Energies Alternatives, France
Beisi Xu,
St. Jude Children's Research Hospital, United States

***Correspondence:**
José Luis García-Giménez
j.luis.garcia@uv.es
Salvador Mena-Mollá
salvador.mena@uv.es

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *29 March 2019*
Accepted: *13 June 2019*
Published: *28 June 2019*

Citation:
Beltrán-García J, Osca-Verdegal R, Mena-Mollá S and García-Giménez JL (2019) Epigenetic IVD Tests for Personalized Precision Medicine in Cancer. Front. Genet. 10:621. doi: 10.3389/fgene.2019.00621

INTRODUCTION

Epigenetics, a breakthrough discipline in biomedicine, aims to improve precision medicine by discovering new epigenetic mechanisms and providing new epigenetic biomarkers, therapeutic targets, and epigenetic drugs with potential uses in clinical practice.

Most human diseases have complex multifactorial pathologies that result from a pathogenic polymorphism in human genes, besides epigenetic mechanisms, which can modulate the expression of functional genes. Currently, several IVD molecular-based tests contribute to the development of precision oncology, which already offers viable alternatives for cancer diagnostics and prognostics. The Food and Drug Administration (FDA) lists several IVD tests that have been cleared and approved for diagnostics, which can be consulted by searching *Nucleic Acid-Based Test* (Food and Drug Administration, 2019a) and *List of Cleared or Approved Companion Diagnostic Devices (In Vitro and Imaging Tools)* (Food and Drug Administration, 2019b).

For a given phenotype, there is a causal contribution of genetic mutations, copy number variations, epigenetic control, and altered transcription programs and altered complex metabolic inputs. The contribution of the aforementioned factors renders the use of different approaches necessary to understand the physiopathology of complex and multifactorial diseases. In line with this, epigenetic biomarkers can help early diagnosis, disease progression monitoring, disease outcome prediction, selection and stratification of patients by risk, prediction of future comorbidities, and even the evaluation of the positive or negative effects of therapeutic interventions in specific patient subsets. Among others, DNA methylation and microRNAs are markedly more stable than RNA and proteins, which renders the use of these biomarkers more practical and viable in clinical settings (Faruq and Vecchione, 2015; Hashimoto et al., 2016; García-Giménez et al., 2017b). In particular, DNA methylation, microRNAs, and post-translational modifications of histones offer high stability

in biofluids and in samples with a compromised quality, such as formalin-fixed paraffin embedded (FFPE). Other advantages of epigenetic biomarkers over genetic or protein-based biomarkers are as follows: 1) their dynamic nature; 2) they provide information about the gene function; 3) they inform about the specific genetic programs that alter during disease; and 4) most techniques to analyze epigenetic biomarkers (i.e., RT-qPCR) have already been introduced into clinical laboratories. Therefore, epigenetics has a tremendous potential to improve predictive and precision medicine.

Precision Medicine was defined by the National Research Council's Toward Precision Medicine in 2008 as: "The tailoring of medical treatment to the individual characteristics of each patient … to classify individuals into subpopulations that differ in their susceptibility to a particular disease or their response to a specific treatment. Preventative or therapeutic interventions can then be concentrated on those who will benefit, sparing expense and side effects for those who will not" (Ginsburg and Phillips, 2018). Therefore, precision medicine has started to use potential epigenetic biomarkers in clinical settings.

We recently defined an epigenetic biomarker as "any epigenetic mark or altered epigenetic mechanism which generally serves to evaluate health or disease status and is particularly stable and reproducible during sample processing." An ideal biomarker can be measured in body fluids (i.e., plasma, serum, saliva, semen, urine, etc.) or primary tissue samples (fresh tissue, cells, single cell isolated, fine-needle aspirates, FFPE, etc.). However, for clinical settings, minimal invasive procedures are preferable. In line with this, human plasma as a source of miRNAs and circulating cell-free DNA (cfDNA) is, therefore, the best option. An ideal epigenetic biomarker for precision medicine applications may cover at least one of the following properties: i) predicts the risk of future disease development (risk); ii) defines a disease (detection); iii) reveals information about the natural history of the disease; iv) predicts the outcome of disease (prognostic); v) responds to therapy (predictive); vi) monitors responses to therapy or medication (therapy monitoring); vii) allows to simultaneously make a diagnosis and perform targeted therapy (theragnosis) (García-Giménez et al., 2017b).

To achieve the precision medicine goals, the current challenge is knowing how to obtain a reliable useful biomarker for clinical routine because, for this purpose, the new biomarker requires high accuracy and robustness (Li et al., 2010; Diamandis, 2012) and cost-effectiveness. It is noteworthy that less than 1% of the biomarkers obtained in biomedical research is finally implemented into the clinical laboratory (Kern, 2012), with an even lower percentage for epigenetic biomarkers. This low percentage of commercialized IVD tests based on epigenetic biomarkers suggests that the precision medicine ecosystem formed by distinct stakeholders (i.e., patients, providers, payers, and regulators) may increase their knowledge about the impact of epigenetic biomarkers on precision medicine, and might also work together to successfully implement this breakthrough technology in clinical practice.

A number of precision medicine applications are contributing to health care improvements by allowing the precise diagnosis of diseases or by identifying specific disease subsets or stages, and by also improving personalized treatments. Specifically, for cancer, which remains the second leading cause of death worldwide, early detection, the identification of cancer subtypes, and the selection of appropriate therapies are crucial to increase the survival of cancer patients. However, the identification of new tumor biomarkers, especially those based on epigenetic biomarkers with the capability to identify tumor origin or cancer subsets, advances in assay technologies, and the development of sophisticated analytical software techniques (i.e., machine learning and artificial intelligence), will help to improve precision medicine in cancer (Ahlquist, 2018).

TECHNOLOGIES FOR EPIGENETIC BIOMARKER ANALYSES IN CLINICAL LABORATORIES

Given the prevalence of the DNA methylation alterations at specific genes under a variety of human disease conditions, a promising future is coming for the DNA methylation analysis as an epigenetic biomarker. In fact, DNA methylation is the best-studied epigenetic modification since it was discovered. In addition, miRNAs have attracted a great deal of interest in clinical research for their role in gene regulation, tissue signaling and cellular homeostasis, their high stability in practically all types of biospecimens, and the relatively easy way by which to measure miRNAs in a wide array of biospecimens. Histone variants and histone post-translational modifications are other potential markers that can be analyzed in a wide array of biospecimens for clinical settings.

Therefore, it is not surprising that most current commercial *in vitro* diagnostic tests are based on either the analysis of DNA methylation of specific genes or the measurement of the relative expression of microRNAs, which can be easily measured by RT-qPCR-based methods (i.e., methyLight, methyl-specific PCR, and methylation-sensitive high-resolution melting) and pyrosequencing technologies (García-Giménez et al., 2017a).

There are other assays based on high-throughput analyses to simultaneously measure several CpG sites. This is, for example, the case of the EPICUP® assay, which is based on using human methylation array Beadchip 450K (Illumina). In the following section, we provide details of a selection of current IVD tests based on epigenetic biomarkers that are currently being commercialized for *in vitro* diagnostic in cancer (**Table 1**).

IN VITRO DIAGNOSTIC TESTS BASED ON EPIGENETIC BIOMARKERS

Epigenetic-Based IVD Test for Colorectal Cancer

Colorectal cancer (CRC) (MIM 11 4500) is the third most frequent cancer in men and the second most frequent cancer in women worldwide, and accounts for nearly 10% of cancers (Ferlay et al., 2015). CRC is the second leading cause of death by cancer. Five-year survival rates range from more than 90% for stage I to less than 10% for stage IV CRC (Siegel et al., 2012). CRC is characterized by slow progression from detectable precancerous lesions and has a good prognosis when patients are diagnosed

TABLE 1 | Commercially available Epigenetic IVD tests with the potential of improving precision medicine in cancer.

Diseases	Epigenetic biomarkers	Commercial tests	Technology for the analysis	Biospecimen	Sn (%)	Sp (%)
Colorectal cancer	DNA methylation (*NDRG4* and *BMP3*)	*Cologuard® stool-DNA-based test*	Stool-based CRC test	Stool	92.3	86.6
	DNA methylation (*SEPT9*)	*Epi proColon® 2.0 test*	MethyLight	CfDNA from blood	75-81	96-91
	DNA methylation (*SDC2*)	*EarlyTect®* CRC *assay*	MethyLight	CfDNA from blood	87	95.2
	miR-31-3p	*miRPreDX-31-3p*	RT-qPCR	FFPE	NA	NA
Breast cancer	DNA methylation (*PITX2*)	*Therascreen PITX2 RGQ PCR kit.*	MethyLight	FFPE; DNA from blood	NA	NA
Cervical cancer	DNA methylation (*ZNF582*)	*Cervi-M® assay*	Methyl-specific PCR	Epithelial cells from cervical brush	73	80
Glioblastoma	DNA methylation (*MGMT*)	*Therascreen MGMT Pyro Kit*	Pyrosequencing	FFPE, DNA from blood	95-97	NA
Lung cancer	DNA methylation (*SHOX2* and *PTGER4*)	*Epi proLung BL Reflex Assay®*	Methyl-specific PCR	CfDNA from blood	78	96
Cancers of unknown origin	Analysis of 450K CpGs	*EPICUP™*	Human methylation Beadchip 450 K (Illumina)	FFPE	97.7	99.6

cfDNA, circulating cell-free DNA; FFPE, formalin-fixed, paraffin-embedded; Sn, Sensitivity; Sp, specificity; NA, data not available.

in early stages. Non-invasive fecal immunochemical test (FIT) for hemoglobin detection in stools is the most widely used test, but its sensitivity is relatively low in detecting early stage I CRC (53%) and advanced adenomas (≥ 1.0 cm) (27%) (Morikawa et al., 2005). Therefore, the potential for reducing the burden of CRC by early detection is significant, and efforts are currently being made to develop CRC screening tests and to improve the adherence rates of participation for screening because people scarcely comply with currently available methods (Issa and Noureddine, 2017). The selection of appropriate therapies for CRC patients is also a clinical need. Among the therapies proposed for CRC, anti-epidermal growth factor receptor (EGFR) mAb therapy is not indicated for carriers of RAS mutations [approximately 50% of patients with metastatic CRC because the mutations in the *RAS* gene (mainly in exons 2, 3, and 4 of *KRAS* and *NRAS*) make metastatic CRC patients non responders to anti-EGFRs mAB treatment] (Boleij et al., 2016). So, the identification of additional biomarkers to allow clinicians to select those patients who could benefit by the established therapies is needed.

The Cologuard® Stool DNA-Based Test

The first FDA-approved DNA methylation assay for general CRC screening for average-risk adults older than 50 years was Cologuard® (Exact Sciences Corp., Madison, WI). The Cologuard® IVD test is a multitarget stool deoxyribonucleic acid (MT-sDNA) screening test based on the analysis of the methylation levels of genes N-Myc downstream-regulated gene 4 (*NDRG4*) and bone morphogenetic protein 3 (*BMP3*), a mutation in the *KRAS* gene (exon 2, codons 12, 13, using ß-actin as the reference gene), and a non-DNA immunochemical assay for human hemoglobin that allows the precise detection of colon neoplasia (Imperiale et al., 2014). The methylation analysis of *NDRG4* and *BMP3* using *ACTB* (ß-actin) as the reference gene is performed according to the method described by Zou et al. (2012), while fecal hemoglobin biomarker values are obtained by the analytical method described by Lidgard et al. (2013). Cologuard® uses a composite score algorithm that is incorporated into the multitarget stool DNA analytic device software as described by Imperiale et al. (2014).

Cologuard® sensitivity and specificity for CRC detection in a study performed with 9,989 subjects was 92.3% and 86.6%, respectively (Imperiale et al., 2014). Although the Cologuard® test sensitivity was higher than FIT (which measures the presence of blood in the colon in fewer fecal samples) for detecting CRC (92% vs. 74%, $p = 0.015$), specificity was lower than that shown by the FIT ((87% vs. 95%) (Imperiale et al., 2014). Moreover, the Cologuard® test detected less than half largely advanced adenomas (precancerous lesions), but performs better than the FIT. In fact, the sensitivity for detecting advanced precancerous lesions was 42.4% with DNA testing and 23.8% with the FIT ($P < 0.001$). These results reinforce the potential of the Cologuard® test as an alternative for surveillance colonoscopy (van Lanschot et al., 2017). However, its high cost and difficult sample pretreatment and management for each analysis type are considered disadvantages for its rapid implementation into clinical routine. Accordingly, the results obtained with the Cologuard® test are delivered to the healthcare provider within 2 weeks from receiving the stool sample.

Despite these disadvantages, both the US Food and Drug Administration and the US Preventive Services Task Force (USPSTF) include the Cologuard® test in their screening exam recommendations (Lin et al., 2016).

The Epi proColon® 2.0 Test

The Epi proColon® test (Epigenomics AG, Berlin, Germany) was designed to minimize invasive tests and to increase the adherence rates of the participation of those people screened for CRC. The Epi proColon® test uses peripheral blood samples to analyze the methylation status of the *SEPT9* gene. Septins are essential proteins during cell division, and *SEPT9* hypermethylation has been proposed as a key factor in CRC (Song and Li, 2015). The original assay was designed to extract DNA from 5 ml of plasma samples, bisulfite conversion of DNA, and its purification by a particle-based bis-DNA purification method to improve the recovery of bisulfite-treated DNA, the quantification of converted DNA by real-time PCR, and the subsequent measurement of *SEPT9* methylation, and *ACTB* (ß-actin) as a reference gene, by real-time PCR in a Lightcycler LC480 system (Roche Applied

Science) and the Quantitect Multiplex PCR mastermix (Qiagen) (DeVos et al., 2009). Epi proColon® 2.0 (Epigenomics Inc., Germany) was approved by the FDA in 2016 as the first blood test intended for early CRC detection. In a large clinical trial using 1,544 plasma samples from the PRESEPT study cohort (ClinicalTrials.gov, Trial Registration ID: NCT00855348), Epi proColon® demonstrated high sensitivity, which ranged from 77.0% to 81.4%, and specificity from 77.9% to 92.1% (Potter et al., 2014). However, some studies have shown some flaws in the use of Epi proColon® to diagnose CRC, such as its lower sensitivity for stage I than for stages II, III, or IV (Jin et al., 2015). A large multicenter prospective study using blood samples from 53 CRC cases and from 1,457 subjects without CRC from the PRESEPT cohort (ClinicalTrials.gov, Trial Registration ID: NCT00855348) showed low sensitivity (48.2%) for detecting CRC and very low sensitivity (11.2%) for identifying advanced adenoma, with 91.5% specificity (Church et al., 2014). One noteworthy result was that the positive detection rate of the *SEPT9* methylation assay increased exponentially as colorectal lesions became more severe and with more advanced CRC stages (Song et al., 2018), although a negative result does not guarantee absence of cancer.

The results obtained by Song et al. (2018) and He et al. (2018) suggest that the methylation status of *SEPT9* could be applied to CRC stage, size, invasion depth, future risk assessment, metastasis, disease progression monitoring, and therapeutic effect evaluation. A possible flaw of this test is that Epi proColon® detected the methylated status of the same region of the *SEPT9* gene in some patients affected by other cancers (i.e., prostate, breast, lung or other diseases, hypertension, hyperlipidemia, diverticulitis, chronic gastritis, or cardiovascular) and according to their age (Ørntoft et al., 2015). Indeed, the Epi proColon® test was positive in 72 (42%) of 173 patients with other cancers and positive in 33 of 191 patients (17%) with other diseases. In addition, an active clinical trial was run to evaluate the potential of the Epi proColon® test for detecting hepatocellular carcinoma among cirrhotic patients (ClinicalTrials.gov, Trial Registration ID: NCT03311152). These scenarios suggest the potential of this test to diagnose other cancers, such as breast cancer, as demonstrated by Shen et al. (2018), but also the inconvenience of the positive results given for cancer patients who were negative for CRC.

It is worth mentioning that colonoscopy remains the universal gold standard method for CRC diagnostics. In Europe and Asia Pacific, only the use of fecal occult blood test (gFOBT) or quantitative FIT for non-invasive screening is still recommended. However, Chinese guidelines have recently recommended using the test as a complement to other diagnostic approaches, like the guaiac-based gFOBT. In the United States, Epi proColon® is not intended to replace the CRC screening tests recommended by clinical guidelines (i.e., colonoscopy, sigmoidoscopy, and gFOBT), but the Epi proColon® test was FDA-approved for CRC screening those patients unwilling or unable to be screened by recommended methods following guidelines.

The EarlyTect® Colorectal Cancer Assay

The EarlyTect® CRC test (Genomictree Inc. Daejeon, South Korea) has recently received CE-IVD certification for the diagnosis of CRC. The EarlyTect™-GI Syndecan2 Methylation Assay is an IVD assay that uses cfDNA isolated from 0.5 ml of serum to analyze the methylation status of *SDC2* (Syndecan-2).

Previous studies have demonstrated the potential of the analysis of the methylation status of the *SDC2* gene for the early diagnosis of CRC. For example, the studies performed by Mitchell et al. (2016) showed lower sensitivity (59%), but relatively good specificity (84%), of methylation-specific PCR assays (probe-based MethyLight assays) for *SDC2* in the early detection of CRC. At this point, it is worth mentioning that the amplicon selected to study the methylation status of this gene slightly differed (420 bp downstream to the CpG proposed by Oh et al., 2017).

More recent studies performed by Oh et al. (2017), which evaluated the methylation analysis of *SDC2* to detect CRC using isolated DNA from stool samples, demonstrated a good sensitivity of 90.0% for detecting CRC and 33.3% for small polyps, with a specificity of 90.9%. Furthermore, these authors demonstrated that the *SDC2* methylation level was linked to cancer severity in CRC patients in stages I to IV (n = 50). Similarly, Niu et al. (2017) evaluated the methylation levels of the *SDC2* gene in 497 stool samples and found sensitivities of 81.1% and 58.2% for detecting CRC (n = 196) and adenoma (≥1 cm) (n = 122), respectively, with 93.3% specificity. These results were comparable to that observed by Park et al. (2018), who found that the *SDC2* gene methylation analysis performed with methyl-specific PCR in bowel lavage fluid collected during colonoscopy could detect CRC and precancerous lesions. In this study, *SDC2* methylation was positive in 100% of villous adenoma, high-grade dysplasia, and hyperplastic polyp biopsies in 88.9% of tubular adenoma samples and in 0% of normal mucosal samples. These findings indicate the potential of *SDC2* methylation as a biomarker for early CRC detection with a sensitivity of 80% and a specificity of 88.9%.

The clinical validation of *SDC2* methylation in serum DNA from the CRC patients (n = 131) in stages I to IV (stage I, 26; II, 57; III, 36; IV, 12) and from healthy individuals (n = 125) by quantitative methylation-specific PCR using the methylation-specific TaqMan probe demonstrated 87% sensitivity [114/141; 95% confidence interval (CI), 80.0% to 92.3%] and 95.2% specificity (10/125; 95%CI, 89.8% to 98.2%) (Oh et al., 2013). The sensitivity of the patients in stage I was particularly high with 92%, which suggests the potential utility of this test for early CRC detection and identification of precancerous lesions, such as polyps.

A recent observational clinical trial conducted with the EarlyTect® CRC test (ClinicalTrials.gov, Trial Registration ID: NCT03146520) was designed to validate the clinical performance of the *EarlyTect® Colon Cancer test* in stool DNA to detect CRC in a case-control study with 634 participants (Dae Han et al., 2019). Of the 585 evaluated subjects, 245 had CRC, 44 had various sized adenomatous polyps, and 245 obtained negative colonoscopy results. The EarlyTect® CRC test gave an overall sensitivity of 90.2% with area under the curve (AUC) of 0.902 in detecting CRC (0–IV) not associated with tumor stage, and a specificity of 90.2%. The sensitivity for detecting early stages (0-II) was 89.1% (114/128). The EarlyTect® CRC test also detected 66.7% (2/3) and 24.4% (10/41) of advanced and non-advanced adenomas, respectively (Dae Han et al., 2019).

Genomictree Inc. has performed experiments to evaluate the cross-reactivity of the EarlyTect® CRC test in an interim clinical validation with stool DNA from 50 CRC patients (stage I, 10; II, 16; III, 14; IV, 10), 14 irritable bowel syndrome (no colonoscopy was performed), 4 with acute colitis, 11 Crohn's disease (colonoscopy was performed), 14 ulcerative enteritis (colonoscopy was performed), and 50 healthy subjects (endoscopy was not performed). In this study, the sensitivity was 90.0% (45/50) and specificity was 90.9% (5/55). The methylation positivity for *SDC2* was observed in 14.3% (2/14) of the irritable bowel syndrome patients, 25.0% (1/4) of the acute colitis patients, and 35.7% (5/14) of the ulcerative colitis patients, while no Crohn's disease case was positive for the EarlyTect® assay. Notably, sensitivity was 84.6% (22/26) in CRC in stages I and II, which suggests the potential applicability of this test for colorectal detection testing using stool DNA.

miRPreDX-31-3p

The miRpredX-31-3p kit (IntegraGen S.A., France) is a CE-IVD marked theranostic test intended to identify patients with metastatic CRC who can benefit from anti-EGFR (epidermal growth factor receptor) therapy. The miRpredX-31-3p kit quantifies relative miR-31-3p levels by RT-qPCR from the total RNA extracted from FFPE samples in primary tumors of patients with metastatic CRC, using a cutoff value of 1.36 for the miR-31-3p expression level to define patients as being low or high expressers of this miRNA (Ramon et al., 2018).

miRpredX-31-3p predicts the potential clinical benefits associated with first-line anti-EGFR (epidermal growth factor receptor) therapy compared with anti-vascular endothelial growth factor receptor (VEGF) therapy or when second or further lines of treatment with anti-EGFR mAB therapy is more beneficial *versus* chemotherapy alone for multiple patient outcomes (Laurent-Puig et al., 2015). Specifically, on one hand, a low miR-31-3p expression in affected tissue is associated with a 12-month survival advantage and a 40% reduced risk of death when using anti-EGFR (cetuximab) therapy *versus* anti-VEGF (bevacizumab) therapy in patients with metastatic CRC. On the other hand, those patients expressing high miR-31-3p levels displayed no differences in outcomes when treated with either anti-EGFR or anti-VEGF therapy (Laurent-Puig et al., 2015; Laurent-Puig et al., 2017). Furthermore, the miR-31-3p expression was evaluated for its potential as a predictive biomarker for anti-EGFR mAb therapy in the patients without mutations in *KRAS* with operable colorectal liver metastases (Pugh et al., 2017).

In an interventional clinical trial in 1,808 subjects (ClinicalTrials.gov, Trial Registration ID: NCT03362684), the predictive potential of the miR-31-3p expression level was studied for the prognostic of patient outcomes, as was the predictive value of the benefit of anti-EGFR therapy (cetuximab) in stage III CRC patients (the patients enrolled in the PETACC-8 Study) (Taieb et al., 2014). The results obtained from this clinical trial demonstrated that patients with the RAS/BRAF wild type who showed low miR-31-3p expression when tumors were treated with cetuximab plus FOLFOX-4 presented improved disease-free survival, overall survival, and survival after recurrence compared with the patients treated with FOLFOX-4 alone.

More recently with logistic regression models, including the miR-31-3p expression level adjusted for potential confounding factors, Laurent-Puig et al. (2019) validate the use of miR-31-3p to differentiate RAS wt metastatic CRC patient outcomes from patients treated with anti-EGFR mAb or anti-VEGF mAb therapy. Those patients with low miR-31-3p levels showed better outcomes when treated with cetuximab compared with bevacizumab.

The miRpredX-31-3p kit was developed on the basis of a standardized RT-qPCR assay for miRNA detection. Several extraction kits (miRNeasy FFPE kit (Qiagen), AllPrep DNA/RNA FFPE Kit (Qiagen), QIAsymphony RNA kit (Qiagen), and Maxwell 16 LEV RNA FFPE kit (Promega)) have been tested to evaluate the efficiency of miRNA extraction from five formalin-fixed, paraffin-embedded (FFPE) 5-mm-thick slides. In addition, the analytical sensitivity and specificity, assay robustness, reproducibility, and accuracy of miR-31-3p detection were also demonstrated in different quantitative PCR systems like ABI 7900HT®, ABI StepOne+®, and ABI QS5® (Applied Biosystems) and LightCycler® 480 (Roche) (Ramon et al., 2018). These results demonstrated the good versatility of the miRpredX-31-3p assay and its feasibility for being easily implemented into clinical diagnostic laboratories. The time to perform the assay was not as long after total RNA was isolated from FFPE tumor samples because the assay is based on a simple RT-qPCR reaction (reverse transcription and subsequent real-time PCR). Hence the miRpredX-31-3p assay can analyze up to 12 samples and provide the results in 1 day (see version 8 of the mirpredx instructions manual).

The Nu.Q™ Colorectal Cancer Screening Triage Test

NuQ® tests (Volition SA; Namur, Belgium) are intended for diagnosing CRC by analyzing different nucleosome characteristics, including the DNA methylation of DNA bound to nucleosomes, post-translational modifications in histones and histone variants, and the detection of cell-free nucleosomes, although the company is developing a new test based on these biomarkers. NuQ® tests are based on Enzyme-linked immunosorbent assay (ELISA) technology and require only one drop of blood from patients (a 10-µl sample).

The most advanced test is the Nu.Q™ Colorectal Cancer Screening Triage Test, which consists in combining different NuQ® previously CE-IVD marked tests. One of them is the NuQ®X test, which detects the 5-methylcytosine levels present in DNA bound to cell-free circulating nucleosomes.

In a validation study performed by Holdenrieder et al. (2014), serum samples were used in two independent cohorts of subjects: i) 90 subjects, including CRC patients (n = 24), benign colorectal diseases (BCD) (n = 10), and healthy controls (n = 56); ii) 113 subjects, including CRC patients (n = 49), BCD (n = 26), and healthy controls (n = 38). Holdenrieder et al. (2014) used the Nu.Q®X test to evaluate its differential diagnostic performance. Their study showed that the circulating methylated DNA levels significantly lowered in CRC and BCD compared with the healthy controls ($p < 0.05$), although no difference was found between BCD and CRC. The AUC on the receiver operating characteristic curve was 0.78, and sensitivity was 33% at 95% specificity for CRC and BCD compared to HC, with a sensitivity of 75% at 70% specificity for CRC compared to HC.

To improve both the sensitivity and specificity of the assays, two new tests were designed: the Nu.Q®T test and the NuQ®V test. Both obtained the CE-IVD mark. The Nu.Q®T test was designed for the diagnostic of CRC by detecting total free circulating nucleosomes (cell-free nucleosome). Nu.Q®V focused on detecting CRC by analyzing histone variants. Finally, they were included in a NuQ® test based on the same nucleosomics ELISA technology (Holdenrieder et al., 2014).

Rahier et al. (2017) used the Nu.Q® assay to evaluate the levels of 12 epitopes [including nucleosome-associated histone modifications: H4K20me3 (mAb), H4PanAc (mAb), pH2AX (mAb), H3K9Me3 (pAb), H2AK119Ub (mAb), H3K9Ac (mAb), and H3K27Ac (mAb); nucleosome-associated DNA modification: 5mC (mAb); nucleosome containing histone variants: H2AZ (mAb); nucleosome-protein adducts: HMGB1 (mAb) and EZH2 (mAb); and finally a conserved nucleosome epitope as reference of total nucleosome content] in the sera of 58 individuals referred for endoscopic CRC detection [patients with CRC (n = 23), patients with pre-cancerous lesions (polyps) (n = 16), and healthy controls (n = 19)]. The multivariate analysis defined a panel of four age-adjusted cf-nucleosomes that provided an AUC of 0.97 for the CRC discrimination of healthy controls with high sensitivity in initial stages (sensitivity of 75% and 86% and specificity of 90% for stages I and II, respectively). The second combination of four cf-nucleosome biomarkers provided an AUC of 0.72 for the identification of patients with pre-cancerous lesions (polyps) (n = 16) in healthy subjects (Rahier et al., 2017).

The Nu.Q™ Colorectal Cancer Screening Triage Test, which is based on the previous described Nu.Q® tests, was evaluated in blinded serum samples from 1,961 FIT-positive individuals. In a set of samples "training set" (n = 1,907), the Nu.Q™ Colorectal Cancer Screening Triage test had the potential to identify a subset of 477 subjects in which colonoscopy was applied and could be avoided. Moreover, the test detected 96.6% of CRCs and 88.5% of high-risk adenomas. The results were corroborated in the "validation set" of samples (n = 1,961), which gave a sensitivity of 91.2% for CRC and 83.0% for high-risk adenoma. Ii was noteworthy that the sensitivity for "screen relevant neoplasia" (considering patients with CRC and high-risk adenomas) was about 86% compared with the 80% obtained with positive FIT and a cutoff value of 200 ng/ml. The results of this large cohort evaluation were promising as the Nu.Q™ Colorectal Cancer Screening Test can reduce unnecessary colonoscopies by 20%, while maintaining sensitivity for CRC close to 90% (Marielle et al., 2017).

The Volition Company announced that the new Nu.Q™ assay would have the potential to detect 81% of CRCs with a specificity of 78% in a cohort of 4,800 CRC patients. Furthermore, the new Nu.Q™ assay detected up to 67% of high-risk adenomas with a specificity of 80% in a cohort of 530 symptomatic patients and initial stage I cancers with a sensitivity of 74% and a specificity of 90% in a pilot study of 58 asymptomatic patients. However, we were unable to find any published results or any registered clinical trial results of this study apart from the company's published interim results.

The Volition Company is developing new-generation Nu.Q assays for other intended uses, such as pancreatic cancer. In fact the Nu.Q assay was also evaluated for the diagnostic of pancreatic cancer. By using a combination of carbohydrate antigen 19-9 (CA 19-9) levels with a panel of four cf-nucleosome markers, Bauden et al. (2015) obtained an AUC of 0.98 with an overall sensitivity of 92% at a 90% specificity to detect pancreatic cancer in serum samples from a cohort of 59 subjects [including patients with resectable pancreatic cancer (n = 25), patients with benign pancreatic disease (n 0), and healthy individuals (n = 24)].

An Epigenetic-Based IVD Test for Breast Cancer

Breast cancer (MIM 114480) is the most commonly diagnosed cancer in women (Torre et al., 2017) and the leading cause of death from cancer in women worldwide (Torre et al., 2016). It is noteworthy that breast cancer can also affect men and, consequently, around 2,670 new cases of invasive breast cancer are expected to be diagnosed in men in 2019. About 20% of breast cancers worldwide are due to environmental or lifestyle risk factors, such as alcohol abuse, excess body weight and fat, and a sedentary lifestyle (Danaei et al., 2005). In addition, screening with the mammography technique has demonstrated its ability to detect breast cancer in early stages, which reduces the mortality risk and increases treatment success (Lauby-Secretan et al., 2015). As a result, new methods that contribute to early diagnosis, the identification of specific subtypes, and the selection of patients who can benefit from specific therapies will increase patient survival for this cancer.

Breast cancer mortality rates are higher than those for any other cancer and account for 25% of cancer cases and 15% of cancer-related deaths (Ferlay et al., 2015). Breast cancer mortality also depends on the cancer subtype. Breast cancer presents several classifications depending on different aspects. It can be classified according to their histological origin, cell differentiation degree, stage, the presence or absence of certain hormone receptors [i.e., hormonal receptors, like estrogen receptor (ER), and progesterone receptor (PR); and the ERBB2 receptor], and molecular subtype (i.e., luminal A, luminal B, HER2, basal-like subtype, normal-like subtype, and Claudin-low subtype).

Tumors classified as triple-negative breast cancer (TNBC) and HER2-positive breast cancer are classified as high-risk cancer with a poor prognosis (Harbeck and Gnant, 2017). Enhancing breast cancer survival and outcome by early detection remains one of the main breast cancer priorities according to the World Health Organization (WHO). Therefore, several efforts are being made by the research community to provide not only new drugs and therapies to treat breast cancer, but to also identify new biomarkers to help implement precision medicine into the clinical management of breast cancer patients (Low et al., 2018). Breast cancer treatment depends partially on the disease state and the breast cancer subtype. Generally speaking, the commonest treatments are targeted therapy, hormonal therapy, radiation therapy, surgery, and chemotherapy, although immunotherapy is being increasingly utilized. Fortunately, the therapeutic options for breast cancer patients are further improved thanks to the use of biomarkers and the implementation of precision medicine

(Meisel et al., 2018), in which epigenetic biomarkers can further improve the battery of *in vitro* assays to manage breast cancer.

The Therascreen PITX2 RGQ PCR Kit

The *Therascreen PITX2 RGQ PCR kit* (Qiagen, Germany) is a methylation-based CE-IVD marked assay that predicts the response of lymph node-positive, ER-positive, and HER2-negative high-risk breast cancer patients. The test differentiates between the patients more likely to respond to anthracyclines chemotherapy (Aubele et al., 2017), and it obtained the CE-IVD mark in 2018.

The methylation analysis of *PITX2* (a promoter of transcription factor 2 of the pituitary homeobox) demonstrates a high correlation with other diagnostic techniques, has the predictive and prognostic capability for patient identification, and supports clinicians by being the most effective therapy option. *PITX2* methylation has attracted the attention of clinicians for not only breast cancer (Widschwendter et al., 2004; Aubele et al., 2017), but also for other tumor types. Continuous scientific evidence indicates the potential of the *PITX2* methylation analysis to predict breast cancer outcomes in lymph node-positive, ER-positive, and HER2-negative breast cancer patients to adjuvant anthracycline-based chemotherapy. Therefore, these clinical observations reinforce the idea of using *PITX2* methylation status to support clinicians as the most effective therapy option (Hartmann et al., 2009; Absmaier et al., 2018).

Hartmann et al. (2009) showed that the *PITX2* DNA methylation improved the prediction by using only clinical factors like tumor stage, grade, or age in a cohort of >200 patients. *PITX2* plays an essential role in the disease pathogenesis. In fact, tumors with a hypermethylated *PITX2* status correlate with poorer survival (overall survival and reduced metastasis-free survival), and also with resistance to treatment. In addition, *PITX2* methylation has been associated with the response to adjuvant chemotherapy (Absmaier et al., 2018; Sheng et al., 2017).

Absmaier et al. (2018) explored the validity of this new predictive candidate biomarker in a retrospective exploratory study. To do so, these authors determined the *PITX2* DNA methylation status in non-metastatic TNBC patients treated with adjuvant chemotherapy with anthracycline by a molecular analysis of breast cancer tissues. Univariate and multivariate analyses demonstrated the statistically independent predictive value of *PITX2* DNA methylation. The authors concluded that for those patients with non-metastatic TNBC, the selective determination of the *PITX2* DNA methylation status can serve as a cancer biomarker to predict responses to anthracycline-based adjuvant chemotherapy (Absmaier et al., 2018).

Schriker et al. (2018) performed a clinical study to analyze the performance of the *PITX2* DNA methylation assay compared to microarray technology. These authors concluded that the performance of the *Therascreen PITX2 RGQ PCR* assay showed high reliability and robustness to predict the outcome of patients with high-risk breast cancer to anthracycline-based chemotherapy. In this study, three CpGs from the *PITX2* promoter 2 gene (*PITX2P2*; 4q25) contained in the methylation array (Maier et al., 2007) were selected, and the appropriate Taqman probes were designed to cover these three CpGs in the Therascreen PITX2 RGQ assay (Schricker et al., 2018).

The *Therascreen PITX2 RGQ PCR assay*, developed by Perkins et al. (2018) in conjunction with the Therawis Diagnostics Company, analyzes the methylation status of the *PITX2* gene obtained from the DNA of FFPE biospecimens. *PITX2* methylation is assessed by methylation-specific real-time PCR and exploits the quantitative PCR (qPCR) oligonucleotide hydrolysis principle of two TaqMan probes labeled with different fluorescent dies (FAM™ for fully methylated and HEX™ for fully unmethylated DNA) in combination with methylation nonspecific primers to measure the methylation status of the target sequences of PITX2 gene promoter 2 in bisulfite-treated DNA. The *Therascreen PITX2 RGQ PCR kit* (Qiagen, Catalog no. 873211) has been currently CE-IVD marked and is commercially available (Aubele et al., 2017; Schricker et al., 2018). It runs in the real-time Rotor-Gene Q MDx thermal cycler (Qiagen) or a Rotor-Gene Q MDx 5plex HRM instrument (Qiagen). The percentage of the methylation ratio (PMR = 100/(1 + 2exp(CtFAM(methylated) – CtHEX(unmethylated))]) is calculated by the Rotor-Gene AssayManager® software with a Gamma Plug-in plus a kit-specific *PITX2* Assay Profile for automated analyses and quality control, including all the validity criteria. Detailed information about the method is described by Schricker et al. (2018) and Maier et al. (2007). The *Therascreeen PITX2 RGQ PCR assay* can be easily adopted in clinical laboratories that already run other Therascreen assays commercialized by Qiagen. The complete workflow is streamlined throughput for a medium sample with highly reliable and robust readouts and can be performed in a time of 2 working days (Perkins et al., 2018).

The Epigenetic-Based IVD Test for Cervical Cancer

Cervical cancer (MIM 603956) is the fourth most frequent cancer in women, with an estimation of 570,000 new cases in 2018, which represents 6.6% of all female cancers. Cervical cancer is the fourth commonest cause of death from cancer in women (Vu et al., 2018), which is approximately 8% of the total deaths from cancer. Furthermore, as cervical cancer has no shown symptoms in its early stages, early identification of cervical precancerous lesions is of critical importance (Gradíssimo and Burk, 2017).

More than 90% of cases are due to infection with human papillomavirus (HPV) (Kumar et al., 2007; Crosbie et al., 2013), and despite people having had HPV infections and them not developing cervical cancer (Dunne and Park, 2013), organized vaccination and screening programs are essential to lower the cervical cancer incidence (Vu et al., 2018).

Thus, cytology-based screening is widespread and has proven to effectively lower cervical cancer incidence rates in many countries (Anttila and Nieminen, 2000). However, the relatively low sensitivity of a single Pap smear and the higher false-negative results, and sometimes the requirement of multiple Pap tests, make cytology-based screening costs prohibitive for the early identification of precancerous lesions. Therefore, preventive programs focus on HPV testing as a primary screening tool for the early detection of the causative agent of cervical cancer (Dillner et al., 2008). In fact, primary high-risk HPV (hrHPV) screening has recently become an accepted stand-alone or

co-test with conventional cytology in preventive cervical cancer programs.

Chang et al. (2015) found that several genes, such as *PAX1*, *ZNF582*, and *SOX1,* were hypermethylated in cervical cancer compared to normal cervical tissue. Shen-Gunter et al. (2016) evaluated the performance of analyzing the HPV genotype and measuring DNA methylation at promoters *ADCY8*, *CDH8,* and *ZNF582* correlated with the cytological grade, therefore demonstrating their potential to be useful biomarkers for the molecular classification of Pap smears. With their systematic literature review, Wentzensen et al. (2009) attempted to identify promising methylation-based biomarkers for the early detection of cervical cancer. These authors found that the elevated methylation of *DAPK1*, *CADM1,* and *RARB* in cervical cancer was consistently observed in several studies and thus became interesting candidates to be validated in large cohorts during standardized clinical trials (Wentzensen et al., 2009). However, no consensus has been reached about which promoter or gene methylation should be analyzed, and whether these will develop into molecule tests with sufficient predictive values or be useful for the early detection of precancerous lesions. One epigenetic test, based on the analysis of the methylation of genes *ZNF582* and *PAX1,* is being commercialized.

The Cervi-M® and Oral-M® DNA assays

The Cervi-M® and Oral-M® DNA assays (by Epigene, iStat Biomedical Co.; Taiwan) obtained CE-IVD approval for the diagnostic of cervical and oral cancers. iStat Biomedical Co. commercializes the Cervi-M®, *ZNF582* DNA, and the Oral-M® assay, which are based on the methylation analysis of *genes ZNF582* and *PAX1*. These genes are highly methylated in cervical and oral cancers, as described by Lin et al. (2014) and Chang et al. (2015). Gene *ZNF582* codifies for zinc finger protein 582, which plays a key role in transcriptional regulation. *ZNF582* methylation status has been demonstrated as a good biomarker for cervical cancer induced by HPV, with a sensitivity of 73% and a specificity of 80% (Lin et al., 2014). Furthermore, *ZNF582* methylation status shows high sensitivity for the detection of grade-3 cervical intraepithelial neoplasia or in a higher stage (CIN3+) (Liou et al., 2016), and demonstrates its utility to improve diagnostic accuracy more than single HPV DNA testing (Li et al., 2019). In addition, the *PAX1* DNA methylation assay allows the detection of cervical cancers graded as CIN3+, as described by Lai et al. (2008) and Lai et al. (2010). This assay generates clinical sensitivity and specificity above 80% when used with the DNA purified from Pap smears (data provided by the company).

The ZNF582/PAX1 assay consists of the bisulfite treatment of DNA obtained from human epithelial cells collected by cervical brush. Then 20 to 80 ng of bisulfite-converted DNA is analyzed by methyl-specific quantitative PCR in a LightCycler® 480 Instrument (Roche) or an Applied Biosystems® 7500 fast system following the protocol described by the manufacturer (see the instructions in the manual). As the analysis depends on the kit used for the bisulfite treatment of DNA, which can last up to 1 day, the complete workflow to perform the Cervi-M® and Oral-M® DNA assays takes about 2 working days.

It is interesting to note that although the Cervi-M® assay has been tested only in the DNA obtained from epithelial cells collected by cervical brush, as the female reproductive tract and regular uterine endometria shedding into the vagina may exfoliate cells, Bakkum-Gamez et al. (2015) proposed using vaginal tampons as a source of DNA to detect endometrial cancer by an assay of methylated DNA markers.

The Epigenetic-Based IVD Test for Glioblastoma

Glioblastoma (GBM, MIM 137800) is the most common primary malignant brain tumor in adults with an unfavorable prognosis and limited treatment options despite innovative diagnostic strategies and new therapies having been developed (Lombardi and Assem, 2017). GBM constitutes approximately 45% to 50% of all primary malignant brain tumors and is diagnosed more frequently in patients aged between 55 and 85 years, with a mean age of 64 years in the United States (Louis et al., 2016). Evidence in recent years has demonstrated that tumors are made of multiple populations of cancerous cells by harboring specific genetic alterations in addition to the classic founder genetic abnormalities and epigenetic alterations that drive intratumor heterogeneity with multiple different cell subpopulations (Gerlinger and Swanton, 2010; Lombardi and Assem, 2017).

EGFR amplification, *IDH1/2* mutations, and *MGMT* promoter methylation have been proposed as prognostic biomarkers for their molecular and clinical significance. *MGMT* promoter methylation is one of the most relevant prognostic markers and can be used to also predict the therapeutic response to one of the therapeutic strategies for GBM based on the use of alkylating agents like carmustine (BCNU, Gliadel®) and temozolomide (Temodar®). This is because *MGMT*, an O6-methylguanine-DNA-methyltransferase, is a DNA-repairing gene whose silencing may increase the susceptibility of cells to temozolomide concurrently with radiation therapy (Zawlik et al., 2009). Furthermore, increased methylation of the *MGMT* promoter measured by pyrosequencing has been related to increased GBM patient survival (Zhao et al., 2016).

The PyroMark Therascreen MGMT Kit and the PyroMark Q96 CpG MGMT Kit

The *MGMT* methylated status is a strong predictor of the response to temozolomide in patients with GBM during therapy with alkylating agents. Therefore, the DNA methylation of this gene has been postulated as a biomarker to classify gliomas and to guide treatment decision-making (Gusyatiner and Hegi, 2018).

Quillien et al. (2012) found that pyrosequencing led to the highest reproducibility and sensitivity in *MGMT* methylation status analyses, as was also confirmed by Hsu et al. (2017) after testing four different techniques (e.g., immunohistochemistry, MSP, qMSP, and pyrosequencing) to analyze the *MGMT* methylation status. Different commercialized kits are available for the pyrosequencing methodology, such as the PyroMark Q96 CpG MGMT kit (cat. number 972032; Qiagen), which uses the PyroMark Q96 MD system (Qiagen), and the test Therascreen MGMT PyroKit (cat. number 972032; Qiagen), which uses the pyrosequencing

PyroMark Q24 system (Qiagen) with specific sequencing primers. The PyroMark Q96 CpG MGMT kit detects five CpG sites located in exon 1 (CpG 74–78), whereas the CE-IVD commercialized kit, the PyroMark Therascreen MGMT kit, detects four CpG sites in the same location (CpG 76–79) of the human *MGMT* gene in DNA samples obtained from blood or FFPE biospecimens. Briefly, the assay consists of using bisulfite converted genomic DNA (with the EpiTect Bisulfite kit, cat. number 59104; Qiagen) for subsequent PCR amplification to sequencing it by pyrosequencing using the kits and systems described above to analyze the methylation status of exon 1 of the *MGMT* gene. The sequences surrounding the defined positions serve as normalization and reference peaks for the quantification and quality assessment of the analysis (see the manufacturer's instructions). The time it takes to obtain results relies on the bisulfite treatment of DNA, which needs 6 to 8 h to complete the workflow of the *MGMT* methylation status analysis and lasts about 2 working days.

After performing the PCR using primers by targeting the defined region of exon 1, amplicons are immobilized on Streptavidin Sepharose High Performance beads. Then single-stranded DNA is prepared, and the sequencing primers are annealed to DNA. Samples are then analyzed in the PyroMark Q24 system.

Both kits (PyroMark Q24 CpG MGMT and Therascreen MGMT PyroKit) have demonstrated their capability to stratify patients with GBM according to its prognostic after measuring *MGMT* promoter methylation (Johannessen et al., 2018). Quillien et al. (2017) evaluated the ability of the Therascreen MGMT kit in 102 glioblastoma patients and found that using a binary classification of methylated/unmethylated *MGMT* gene with cutoffs of 8% and 12%, 95% and 97% of GBM patients were well classified. Quillien et al. (2017) also found an excellent prognostic capability of the assay and indicated median overall survival of 15.9 and 34.9 months, respectively, for unmethylated and methylated patients. Moreover, the use of the *MGMT* methylated status as a predictor of meningioma has been recently tested by Panagopoulos et al., but these authors concluded that the methylation frequency of the *MGMT* promoter in meningioma is very low (6%) and, therefore, suggested that *Therascreen MGMT PyroKit* is not suitable for meningiomas.

As *MGMT* is methylated to 25% to 50% in numerous cancers, including brain, colon, lung, breast, gastric, and ovarian cancer (Gerson, 2004), it involves the risk of offering positive results for cancer patients who were found negative for GBM.

The Epigenetic-Based IVD Test for Lung Cancer

Lung cancer (MIM 211980) is the leading cause of death from cancer worldwide (Siegel et al., 2017), and 8 or 9 of 10 lung cancer cases occur in smokers. Thus, smoking is the biggest risk factor of this disease. The 5-year survival rate after diagnosis is 15.6%, which is lower than the survival rates for breast, colon, and prostate cancers. The WHO classifies lung cancer into two broad histological subtypes. The first one is non–small-cell lung cancer (NSCLC), which causes about 85% of cases, including lung squamous carcinoma (LUSC), lung adenocarcinoma (LUAD), and large cell carcinoma subtypes. The second subtype is small-cell lung cancer (SCLC), which accounts for the remaining 15% (Couraud et al., 2012).

The treatment that includes surgical, medical, and radiotherapeutic interventions did not much improve the long-term survival rate of those patients diagnosed with primary lung neoplasms. Moreover, classic cisplatin-based chemotherapy for NSCLC, which can be combined with anti-angiogenic bevacizumab, gives low to moderate satisfactory results. The use of specific tyrosine kinase inhibitors (TKi) in *EGFR*-mutated, ALK/ROS1-rearranged NLSC, and the development of new immunotherapy strategies based on anti-PD1/PD-L1 mAb are currently improving the clinical outcomes of lung cancer patients (Duruisseaux and Esteller, 2018). Yet despite new therapies having been designed and applied, tumor resistance to treatments mean that about 154,050 people died from lung cancer in 2018 only in the United States (https://www.cancer.org/). To increase survival rates in lung cancer, early diagnosis is a priority. However, one of the most widely used techniques is the computed tomography (CT) of the thorax and bronchoscopy. CT gives rise to false positives in lung-cancer free patients, delays lung cancer diagnosis, and also exposes these subjects unnecessarily to radiation. Bronchoscopy fails in about half those diagnosed with lung cancer. Therefore, a diagnostic test based on the biological material obtained from non-invasive or minimally invasive samples with high specificity may cut the need for more costly invasive diagnostic procedures.

The current hypothesis to explain lung carcinogenesis considers that tumor development occurs in a multistage stepwise manner that contributes to the accumulation of genetic and epigenetic alterations (Lantuéjoul et al., 2009). Therefore, epigenetic signatures based on dysregulated DNA methylation differentially express miRNA, and altered post-translational modified histones can reflect the driving force of lung carcinogenesis. Accordingly, given the pivotal role of epigenetic disruption during this process, the epigenomic marks detected in tissue or body fluids represent a feasible biomarker to identify disease in its early stages, establish a prognostic, and monitor treatment response (Bhargava et al., 2018). In a recent relevant work, Duruisseaux and Esteller (2018) describe several epigenetic mechanisms that underlie the acquisition of the cancerous phenotype and the aggressive behavior of lung cancer. They also propose circulating epigenetic biomarkers and the therapeutic potential of epigenetic drugs to implement precision medicine in lung cancer.

The Epi proLung BL Reflex Assay®

SHOX2, or short stature homeobox gene two, methylation has been identified as a biomarker capable of reliably differentiating between lung tumor tissue and normal tissues (Lewin et al., 2007; Schmidt et al., 2010).

SHOX2 methylation, as determined from bronchial aspirates, has demonstrated good sensitivity and a high specificity as a biomarker for lung cancer (Dietrich et al., 2012b). *Epigenomics AG* commercializes the Epi proLung BL Reflex Assay® (Epigenomics AG, Berlin, Germany), a CE-IVD test for quantifying *SHOX2* methylation using methyl-specific PCR with AUC [95% confidence intervals] = 0.94 [0.91–0.98], sensitivity 78% [69–86%], and specificity 96% [90–99%] in bronchial lavage

specimens (Dietrich et al., 2012b), albeit with lower sensitivity (about 40%) in malignant pleural effusions (Ilse et al).

The Epi proLung BL Reflex Assay® is composed of three individual kits: The Epi proLung BL DNA Preparation Kit to prepare bisulfite converted DNA by ammonium bisulfite chemistry, the Epi proLung BL real-time PCR Kit for the quantitative and sensitive analyses of the relative amount of methylated *SHOX2* gene, and the Epi proLung BL Work Flow Control Kit for monitoring and controlling the whole workflow. A detailed explanation of the different steps performed in the *SHOX2* gene methylation analysis using the Epi proLung BL Reflex assay® is described by Dietrich et al. (2012a). Like other methylation-based assays, the time to obtain the results relies on the bisulfite treatment of DNA, which requires about 8 h. Therefore, 2 working days are needed to complete the workflow of the Epi proLung BL Reflex Assay®.

In 2011, *SHOX2* methylation was assessed in circulating cell-free DNA obtained from blood plasma and showed a sensitivity of 60% and a specificity of 90% for lung cancer diagnosis in a case-control study with 343 subjects (Kneip et al., 2011). Since then, Epigenomics AG has been working on demonstrating the test's utility. In 2017, the Epi proLung® blood-based version for the lung cancer test received the CE-IVD mark, which is based on a combination of the methylation analyses of *SHOX2* and *PTGER4* (the prostaglandin E receptor 4 gene). In fact, Weiss et al. (2017) demonstrated significant discriminatory performance for distinguishing patients with lung cancer from subjects with no malignancy (AUC [95% confidence intervals] = 0.88, sensitivity 90%, and specificity 73%) in circulating DNA from plasma samples by the methylation analysis of genes *SHOX2* and *PTGER4*.

The current commercial Epi proLung® assay consists of the Epi proLung PCR Kit (M6-02-002) and the Epi proLung Control Kit (M6-02-003), and has been validated with bisulfited-treated DNA prepared with the Epigenomics Epi BiSKit (M7-01-001) from cell free-circulating DNA present and isolated from 3.5 ml of plasma. The methylation of the *ACTB* gene (ß-actin) is measured as an internal control to assess input adequacy. It also provides positive and negative controls for each run by starting with DNA extraction from plasma. Two methylated *SHOX2*- and *PTGER4*-specific fluorescent detection probes are used in this MethyLight-based assay to exclusively identify the methylated target sequences amplified during the PCR reaction. The assay, with an area under the ROC curve (AUC = 0.82), displays the observed likelihood of being diagnosed with lung cancer according to the EPLT score (ranging from threshold −0.43 to −1.85), together with the corresponding sensitivity (59% to 85%, respectively) and specificity (95% to 50%, respectively), which depends on a given specific threshold (see the Epi proLung® instruction manual for more details)

Epigenomics has performed experiments to evaluate the cross-reactivity of the Epi proLung® assay. Both BLAST alignment searches and PCR analyses against the human genome with the Epi proLung PCR assay (blockers, primers, and probes) have been performed. This analysis showed that the test is specific and only gives the amplification of the bisulfite-treated DNA sequence of methylated *SHOX2* and *PTGER4*, respectively, and not the other regions in the human genome. Epi proLung® was also checked to evaluate the methylated status of *SHOX2* and *PTGER4* in the patients affected by other lung-associated diseases. Fifty-seven (57) samples from patients with non-malignant lung diseases [Chronic obstructive pulmonary disease (COPD), pneumonia, lung emphysema, interstitial lung disease] were evaluated to determine cross-reactivity. The Epi proLung® test discriminated malignant disease from non-malignant disease with an AUC of 0.73.

The Epigenetic-Based IVD Test for Cancers of Unknown Origin

Cancer of an unknown primary site (CUP) is a heterogeneous group of cancers for which the anatomical site of origin remains hidden after detailed clinical and histological investigations (Briasoulis et al., 2005; Varadhachary and Raber, 2014). CUP is clinically characterized as an aggressive disease with early dissemination (Pentheroudakis et al., 2013) that contributes to their presentation (Varadhachary and Raber, 2014). CUP accounts for 3% to 5% of all cancer diagnoses and is the third commonest cause of death from cancer because, unfortunately, most patients (80–85%) do not respond appropriately to treatment (Pavlidis and Fizazi, 2009; Pavlidis and Pentheroudakis, 2012). Therefore, patient survival is very limited.

Tumors in CUP share biologic and molecular properties, but tumors in CUP are currently indicated to maintain the signature of the putative primary origin. The general characteristics of CUP are: 1) short natural history with symptoms and signs associated with metastatic sites; 2) early rapid dissemination in the absence of a primary tumor (three organs or more are involved upon diagnosis in one third of patients); 3) aggressive clinical progression; and 4) sometimes an unpredictable metastatic pattern that differs from those of known primary tumors (Pavlidis and Fizazi, 2009; Pavlidis and Pentheroudakis, 2012).

The heterogeneous CUP presentations mean that immunohistochemical testing, the characterization of tissue-of-origin molecular profiling, and the assignation of appropriate therapies present a challenge (Varadhachary and Raber, 2014). Classifying CUP patients into several clinicopathological subsets is necessary for oncologists to manage these patients and to decide about appropriate therapies. This classification is done according to socio-demographic criteria, such as age and gender, histopathology patterns, clinico-pathological data, laboratory tests, and image data (MRI, PET, CT scanning, mammography, etc.), and also to the affected organ or site. Despite several immunohistochemistry panels having been developed to diagnose CUP, the primary cancer site remains unknown in about 75% of patients (Varadhachary and Raber, 2014). Therefore, the need to find new diagnostic tools to discover the tissue of origin is substantial.

EPICUP™

EPICUP™ (Ferrer, Spain) is a CE-IVD test used to biologically define the tissue of origin in CUP. EPICUP™ was the first epigenetic test designed to identify tumors of unknown primary and claims that it can identify up to 87% of cases of cancer of unknown origin (Moran et al., 2016). The EPICUP™ test is based on the analysis of 485,577 CpG sites measured by the human methylation matrix Infinium HumanMethylation450 Beadchip

microarray (Illumina), and the test was designed to look for similarities in the methylation patterns of cancers of unknown primary and known primary tumors. Based on the results, the EPICUP™ test is able to perform an epigenetic identification and subsequent categorization of the primary site in CUP cancers from FFPE or frozen tissue samples (Moran et al., 2016). This is not a suitable assay for all clinical laboratories because the EPICUP™ test is based on Illumina methylation BeadChip. Therefore, the mean time to provide results takes about 2 weeks if it is to consider DNA purification from tissue, the bisulfite treatment of purified DNA, array hybridization, and, finally, bioinformatic data analyses and their interpretations.

EPICUP™ classifies the tumor type based on the study of DNA methylation profiles using the Infinium HumanMethylation450 Beadchip microarray DNA methylation signature. It offers a specificity of 99.6%, a sensitivity of 97.7%, a positive predictive value of 88.6%, and a negative predictive value of 99.9% in a validation set of 7,691 tumors. Thus, with the samples of 216 CUP patients (FFPE tissue), the DNA methylation profile was able to predict a cancer of primary origin in 188 patients (87%) (Moran et al., 2016).

EPICUP™ demonstrates its ability to provide the correct treatment to CUP patients. In fact, the patients who received tumor-specific therapy diagnosed with EPICUP showed better overall survival than those who received empirical therapy [hazard ratio (HR) 3.24, p = 0.0051 (95% CI, 1.42–7.38); log-rank p = 0.0029] (Moran et al., 2016). Likewise, EPICUP in a study of DNA methylation profiles was proven a cost-effective test in breast, pancreas, colon, lung (NSCLC), and prostate cancers and increased the overall survival adjusted for quality (Gracia et al., 2015).

CONCLUSIONS

Modern medicine moves toward more personalized practice and theragnosis, and epigenetic biomarkers can further contribute to all of this. This review describes the most advanced and commercially available tests based on epigenetic biomarkers that help to improve precision medicine. In some cancers, such as CRC, several options are available, based on stool DNA (i.e., Cologuard® and EarlyTect®), liquid biopsy (Epi ProColon®, EarlyTect® and NuQ™), and FFPE (miRPredX-31-3p). Other tests, such as the Therascreen MGMT Pyro kit for glioblastoma, can be used in the DNA obtained from blood and FFPE tissues. Obviously, for clinical settings and to avoid invasive procedures, tests based on a liquid biopsy are preferable.

Methodologically speaking, to implement these new epigenetic tests into clinical routine, most of these tests have adopted easy-to-use inexpensive analytical methods, like those based on RT-qPCR and microarrays for both DNA methylation and miRNA analyses. There is still a long way ahead before these epigenetic tests can be completely implemented into clinical routine. The companies developing epigenetic tests should focus their efforts on simplifying the technology used to analyze epigenetic biomarkers in a clinical laboratory environment by, for example, using qPCR-based technology, which is easy to use and cost-effective. Moreover, companies have to make efforts to identify biomarkers in non-invasive biospecimens, which will contribute to anticipate cancer diagnosis and to also increase patient compliance with screening campaigns.

We are witnessing a revolution by adapting machine learning procedures to epigenetic biomarkers analyses that will contribute to definitely implement new epigenetic biomarkers into clinical routine. In fact, advanced computational techniques have taken us closer to realize the application of epigenetics to personalized medicine (Holder et al., 2017). One important scenario is that the cost of specific treatments and the appropriate use of targeted therapies guided by epigenetic biomarkers are expected to streamline the immense cost required to receive personalized therapies.

AUTHOR CONTRIBUTIONS

JB-G, RO-V, SM-M, and JLG-G contributed to bibliography compilation and analysis. JB-G, RO-V, SM-M, and JLG-G contributed to manuscript drafting. JB-G, RO-V, SM-M, and JLG-G reviewed the manuscript content. All authors approved the final version of the manuscript.

FUNDING

This work was supported by a 2017 VLC-Bioclinics grant and the Generalitat Valenciana (GV/2014/132), AES2016 (ISCIII) with grant number PI16/01036 and Proyectos de Desarrollo Tecnológico en Salud with grant number DTS17/132 AES2017, co-financed by the European Regional Development Fund (ERDF), Instituto de Salud Carlos III through CIBERer (Biomedical Network Research Center for Rare Diseases and INGENIO2010). JB-G is supported by a grant Contratos i-PFIS (IFI18/00015) and co-financed by the European Social Fund. RO-V is supported by the grant APOTIP/2019/A/015 "Subvenció per a la Contractació de Personal de Suport vinculat a un projecte d'investigació o de Transferència Tecnològica" Conselleria de Educación, Investigación, Cultura y Deporte de la Generalitat Valenciana.

REFERENCES

Absmaier, M., Napieralski, R., Schuster, T., Aubele, M., Walch, A., Magdolen, V., et al. (2018). PITX2 DNA-methylation predicts response to anthracycline-based adjuvant chemotherapy in triple-negative breast cancer patients. *Int. J. Oncol.* 52 (3), 755–767. doi: 10.3892/ijo.2018.4241

Ahlquist, D. A. (2018). Universal cancer screening: revolutionary, rational, and realizable. *npj Precis. Oncol.* 2, 23. doi: 10.1038/s41698-018-0066-x

Anttila, A., and Nieminen, P. (2000). Cervical cancer screening programme in Finland. *Eur. J. Cancer* 36, 2209–2214. doi: 10.1016/S0959-8049(00)00311-7

Aubele, M., Schmitt, M., Napieralski, R., Paepke, S., Ettl, J., Absmaier, M., et al. (2017). The predictive value of PITX2 DNA methylation for high-risk breast cancer therapy: current guidelines, medical needs, and challenges. *Dis. Markers* 2017, 1–14. doi: 10.1155/2017/4934608

Bakkum-Gamez, J. N., Wentzensen, N., Maurer, M. J., Hawthorne, K. M., Voss, J. S., Kroneman, T. N., et al. (2015). Detection of endometrial cancer via molecular

analysis of DNA collected with vaginal tampons. *Gynecol. Oncol.* 137, 14–22. doi: 10.1016/j.ygyno.2015.01.552

Bauden, M., Pamart, D., Ansari, D., Herzog, M., Eccleston, M., Micallef, J., et al. (2015). Circulating nucleosomes as epigenetic biomarkers in pancreatic cancer. *Clin. Epigenetics*. 7 (1), 106. doi: 10.1186/s13148-015-0139-4

Bhargava, A., Bunkar, N., Aglawe, A., Pandey, K. C., Tiwari, R., Chaudhury, K., et al. (2018). Epigenetic biomarkers for risk assessment of particulate matter associated lung cancer. *Curr. Drug Targets* 19, 1127–1147. doi: 10.2174/1389450118666170911114342

Boleij, A., Tack, V., Taylor, A., Kafatos, G., Jenkins-Anderson, S., Tembuyser, L., et al. (2016). RAS testing practices and RAS mutation prevalence among patients with metastatic colorectal cancer: results from a Europe-wide survey of pathology centres. *BMC Cancer* 16, 825. doi: 10.1186/s12885-016-2810-3

Briasoulis, E., Tolis, C., Bergh, J., and Pavlidis, N. ESMO Guidelines Task Force (2005). ESMO Minimum Clinical Recommendations for diagnosis, treatment and follow-up of cancers of unknown primary site (CUP). *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* 16 (Suppl 1), i75–76. doi: 10.1093/annonc/mdi804

Chang, C.-C., Huang, R.-L., Liao, Y.-P., Su, P.-H., Hsu, Y.-W., Wang, H.-C., et al. (2015). Concordance analysis of methylation biomarkers detection in self-collected and physician-collected samples in cervical neoplasm. *BMC Cancer* 15, 418. doi: 10.1186/s12885-015-1411-x

Church, T. R., Wandell, M., Lofton-Day, C., Mongin, S. J., Burger, M., Payne, S. R., et al. (2014). Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. *Gut* 63, 317–325. doi: 10.1136/gutjnl-2012-304149

Couraud, S., Zalcman, G., Milleron, B., Morin, F., and Souquet, P.-J. (2012). Lung cancer in never smokers – A review. *Eur. J. Cancer* 48, 1299–1311. doi: 10.1016/j.ejca.2012.03.007

Crosbie, E. J., Einstein, M. H., Franceschi, S., and Kitchener, H. C. (2013). Human papillomavirus and cervical cancer. *Lancet (London, England)* 382, 889–899. doi: 10.1016/S0140-6736(13)60022-7

Dae Han, Y., Jeong Oh, T., Chung, T.-H., Won Jang, H., Nam Kim, Y., An, S., et al. (2019). Early detection of colorectal cancer based on presence of methylated syndecan-2 (SDC2) in stool DNA. *Clin. Epigenetics*. 11, 51. doi: 10.1186/s13148-019-0642-0

Danaei, G., Vander Hoorn, S., Lopez, A. D., Murray, C. J. L., and Ezzati, M. Comparative risk assessment collaborating group (Cancers) (2005). Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. *Lancet (London, England)* 366, 1784–1793. doi: 10.1016/S0140-6736(05)67725-2

DeVos, T., Tetzner, R., Model, F., Weiss, G., Schuster, M., Distler, J., et al. (2009). Circulating methylated SEPT9 DNA in plasma is a biomarker for colorectal cancer. *Clin. Chem.* 55, 1337–1346. doi: 10.1373/clinchem.2008.115808

Diamandis, E. P. (2012). The failure of protein cancer biomarkers to reach the clinic: why, and what can be done to address the problem? *BMC Med.* 10, 87. doi: 10.1186/1741-7015-10-87

Dietrich, D., Kneip, C., Raji, O., Liloglou, T., Seegebarth, A., Schlegel, T., et al. (2012a). Performance evaluation of the DNA methylation biomarker SHOX2 for the aid in diagnosis of lung cancer based on the analysis of bronchial aspirates. *Int. J. Oncol.* 40, 825–832. doi: 10.3892/ijo.2011.1264

Dietrich, D., Liebenberg, V., Schmidt, B., Fleischhacker, M., Kneip, C., Kristiansen, G., et al. (2012b). Development and performance evaluation of a CE-IVD for measuring SHOX2 DNA methylation in bronchial aspirates for the diagnosis of lung cancer. *Lung Cancer* 77, S22. doi: 10.1016/j.lungcan.2012.05.036

Dillner, J., Rebolj, M., Birembaut, P., Petry, K.-U., Szarewski, A., Munk, C., et al. (2008). Long term predictive values of cytology and human papillomavirus testing in cervical cancer screening: joint European cohort study. *BMJ* 337, a1754. doi: 10.1136/bmj.a1754

Dunne, E. F., and Park, I. U. (2013). HPV and HPV-associated diseases. *Infect. Dis. Clin. North Am.* 27, 765–778. doi: 10.1016/j.idc.2013.09.001

Duruisseaux, M., and Esteller, M. (2018). Lung cancer epigenetics: from knowledge to applications. *Semin. Cancer Biol.* 51, 116–128. doi: 10.1016/j.semcancer.2017.09.005

Faruq, O., and Vecchione, A. (2015). microRNA: diagnostic perspective. *Front. Med.* 2, 51. doi: 10.3389/fmed.2015.00051

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., et al. (2015). Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* 136, E359–E386. doi: 10.1002/ijc.29210

Food and Drug Administration (2019a). Nucleic Acid Based Test. https://www.fda.gov/medical-devices/vitro-diagnostics/nucleic-acid-based-tests

Food and Drug Administration (2019b). List of cleared or approved companion diagnostic devices (in vitro and imaging tools). https://www.fda.gov/medical-devices/vitro-diagnostics/list-cleared-or-approved-companion-diagnostic-devices-vitro-and-imaging-tools

García-Giménez, J. L., Mena-Mollá, S., Beltrán-García, J., and Sanchis-Gomar, F. (2017a). Challenges in the analysis of epigenetic biomarkers in clinical samples. *Clin. Chem. Lab. Med.*, 55 (10), 1–4. doi: 10.1515/cclm-2016-1162

García-Giménez, J. L., Seco-Cervera, M., Tollefsbol, T. O., Romá-Mateo, C., Peiró-Chova, L., Lapunzina, P., et al. (2017b). Epigenetic biomarkers: current strategies and future challenges for their use in the clinical laboratory. *Crit. Rev. Clin. Lab. Sci.* 54, 529–550. doi: 10.1080/10408363.2017.1410520

Gerlinger, M., and Swanton, C. (2010). How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. *Br. J. Cancer* 103, 1139–1143. doi: 10.1038/sj.bjc.6605912

Gerson, S. (2004). MGMT: its role in cancer aetiology and cancer therapeutics. *Nat. Rev. Cancer*. 4, 296–307. doi: 10.1038/nrc1319

Ginsburg, G. S., and Phillips, K. A. (2018). Precision medicine: from science to value. *Health Aff. (Millwood).* 37, 694–701. doi: 10.1377/hlthaff.2017.1624

Gracia, A., Balañá, C., Kaskens, L., Chiavenna, S., Matias-Guiu, X., Rubio-Rodríguez, D., et al. (2015). Economic analysis of epicup, an epigenetic test to predict the tissue of origin in cancer of unknown primary site, the USA payors perspective. *Value Heal.* 18, A356. doi: 10.1016/j.jval.2015.09.670

Gradíssimo, A., and Burk, R. D. (2017). Molecular tests potentially improving HPV screening and genotyping for cervical cancer prevention. *Expert Rev. Mol. Diagn.* 17, 379–391. doi: 10.1080/14737159.2017.1293525

Gusyatiner, O., and Hegi, M. E. (2018). Glioma epigenetics: from subclassification to novel treatment options. *Semin. Cancer Biol.* 51, 50–58. doi: 10.1016/j.semcancer.2017.11.010

Harbeck, N., and Gnant, M. (2017). Breast cancer. *Lancet* 389, 1134–1150. doi: 10.1016/S0140-6736(16)31891-8

Hartmann, O., Spyratos, F., Harbeck, N., Dietrich, D., Fassbender, A., Schmitt M., et al. (2009). DNA methylation markers predict outcome in node-positive, estrogen receptor-positive breast cancer with adjuvant anthracycline-based chemotherapy. *Clin. Cancer Res.* 15, 315–323. doi: 10.1158/1078-0432.CCR-08-0166

Hashimoto, Y., Zumwalt, T. J., and Goel, A. (2016). DNA methylation patterns as noninvasive biomarkers and targets of epigenetic therapies in colorectal cancer. *Epigenomics* 8, 685–703. doi: 10.2217/epi-2015-0013

He, N., Song, L., Kang, Q., Jin, P., Cai, G., Zhou, J., et al. (2018). The pathological features of colorectal cancer determine the detection performance on blood ctDNA. *Technol. Cancer Res. Treat.* 17, 153303381879179. doi: 10.1177/1533033818791794

Holdenrieder, S., Dharuman, Y., Standop, J., Trimpop, N., Herzog, M., Hettwer, K., et al. (2014). Novel serum nucleosomics biomarkers for the detection of colorectal cancer. *Anticancer Res.* 34 (5), 2357–2362.

Holder, L. B., Haque, M. M., and Skinner, M. K. (2017). Machine learning for epigenetics and future medical applications. *Epigenetics* 12, 505–514. doi: 10.1080/15592294.2017.1329068

Hsu, C.-Y., Ho, H.-L., Lin, S.-C., Chen, M.-H., Hsu, S. P.-C., Yen, Y.-S., et al. (2017). Comparative assessment of 4 methods to analyze MGMT status in a series of 121 glioblastoma patients. *Appl. Immunohistochem. Mol. Morphol. AIMM* 25, 497–504. doi: 10.1097/PAI.0000000000000331

Ilse, P., Biesterfeld, S., Pomjanski, N., Fink, C., and Schramm, M. (2013). SHOX2 DNA methylation is a tumour marker in pleural effusions. *Cancer Genomics Proteomics* 10 (5), 217–223.

Imperiale, T. F., Ransohoff, D. F., Itzkowitz, S. H., Levin, T. R., Lavin, P., Lidgard, G. P., et al. (2014). Multitarget stool DNA testing for colorectal-cancer screening. *N. Engl. J. Med.* 370, 1287–1297. doi: 10.1056/NEJMoa1311194

Issa, I. A., and Noureddine, M. (2017). Colorectal cancer screening: an updated review of the available options. *World J. Gastroenterol.* 23, 5086–5096. doi: 10.3748/wjg.v23.i28.5086

Jin, P., Kang, Q., Wang, X., Yang, L., Yu, Y., Li, N., et al. (2015). Performance of a second-generation methylated SEPT9 test in detecting colorectal neoplasm. *J. Gastroenterol. Hepatol.* 30, 830–833. doi: 10.1111/jgh.12855

Johannessen, L., Bradal, P., Myklebust, T. A., Heim, S., Micci, F., and Panagopoulos, I. (2018). MGMT gene promoter methylation status – assessment of two

pyrosequencing kits and three methylation-specific PCR methods for their predictive capacity in glioblastomas. *Cancer Genomics Proteomics* 15, 437–446. doi: 10.21873/cgp.20102

Kern, S. E. (2012). Why your new cancer biomarker may never work: recurrent patterns and remarkable diversity in biomarker failures. *Cancer Res.* 72, 6097–6101. doi: 10.1158/0008-5472.CAN-12-3232

Kneip, C., Schmidt, B., Seegebarth, A., Weickmann, S., Fleischhacker, M., Liebenberg, V., et al. (2011). SHOX2 DNA methylation is a biomarker for the diagnosis of lung cancer in plasma. *J. Thorac. Oncol.* 6, 1632–1638. doi: 10.1097/JTO.0b013e318220ef9a

Kumar, V., Abbas, A. K., Fausto, N., and Mitchell., R. (2007). *Robbins Basic Pathology (8th ed)*. 8th ed. Philadelphia: Saunders Elseviers.

Lai, H., Lin, Y., Huang, R., Chung, M., Wang, H., Liao, Y., et al. (2008). Identification of novel DNA methylation markers in cervical cancer. *Int. J. Cancer.* 123 (1), 161–167. doi: 10.1002/ijc.23519

Lai, H., Lin, Y., Huang, R., Chung, M., Wang, H., Liao, Y., et al. (2010). Quantitative DNA methylation analysis detects cervical intraepithelial neoplasms type 3 and worse. *Cancer*. 116 (8), 4266–4274. doi: 10.1002/cncr.25252

Lantuéjoul, S., Salameire, D., Salon, C., and Brambilla, E. (2009). Pulmonary preneoplasia–sequential molecular carcinogenetic events. *Histopathology* 54, 43–54. doi: 10.1111/j.1365-2559.2008.03182.x

Lauby-Secretan, B., Scoccianti, C., Loomis, D., Benbrahim-Tallaa, L., Bouvard, V., Bianchini, F., et al. (2015). Breast-cancer screening—viewpoint of the IARC Working Group. *N. Engl. J. Med.* 372, 2353–2358. doi: 10.1056/NEJMsr1504363

Laurent-Puig, P., Grisoni, M.-L., Heinemann, V., Bonnetain, F., Fontaine, K., Vazart, C., et al. (2017). MiR 31 3p as a predictive biomarker of cetuximab efficacy effect in metastatic colorectal cancer (mCRC) patients enrolled in FIRE-3 study. *J. Clin. Oncol.*, 34, 3516–3516. doi: 10.1200/JCO.2016.34.15_suppl.3516

Laurent-Puig, P., Grisoni, M.-L., Heinemann, V., Liebaert, F., Neureiter, D., Jung, A., et al. (2019). Validation of miR-31-3p expression to predict cetuximab efficacy when used as first-line treatment in RAS wild-type metastatic colorectal cancer. *Clin. Cancer Res.* 25, 134–141. doi: 10.1158/1078-0432.CCR-18-1324

Laurent-Puig, P., Paget-Bailly, S., Vernerey, D., Vazart, C., Decaulne, V., Fontaine, K., et al. (2015). Evaluation of miR 31 3p as a biomarker of prognosis and panitumumab benefit in RAS -wt advanced colorectal cancer (aCRC): analysis of patients (pts) from the PICCOLO trial. *J. Clin. Oncol.* 33, 3547–3547. doi: 10.1200/jco.2015.33.15_suppl.3547

Lewin, J., Plum, A., Hildmann, T., Rujan, T., Eckhardt, F., Liebenberg, V., et al. (2007). Comparative DNA methylation analysis in normal and tumour tissues and in cancer cell lines using differential methylation hybridisation. *Int. J. Biochem. Cell Biol.* 39, 1539–1550. doi: 10.1016/j.biocel.2007.03.006

Li, J., Lenferink, A. E. G., Deng, Y., Collins, C., Cui, Q., Purisima, E. O., et al. (2010). Identification of high-quality cancer prognostic markers and metastasis network modules. *Nat. Commun.* 1, 1–8. doi: 10.1038/ncomms1033

Li, N., He, Y., Mi, P., and Hu, Y. (2019). ZNF582 methylation as a potential biomarker to predict cervical intraepithelial neoplasia type III/worse: a meta-analysis of related studies in Chinese population. *Medicine (Baltimore)* 98, e14297. doi: 10.1097/MD.0000000000014297

Lidgard, G., Domanico, M., Bruinsma, J., Light, J., Gagrat, Z., Oldham-Haltom, R., et al. (2013). Clinical performance of an automated stool DNA assay for detection of colorectal neoplasia. *Clin. Gastroenterol. Hepatol.* 11, 1313–1318. doi: 10.1016/j.cgh.2013.04.023

Lin, H., Chen, T.-C., Chang, T.-C., Cheng, Y.-M., Chen, C.-H., Chu, T.-Y., et al. (2014). Methylated ZNF582 gene as a marker for triage of women with Pap smear reporting low-grade squamous intraepithelial lesions—a Taiwanese Gynecologic Oncology Group (TGOG) study. *Gynecol. Oncol.* 135, 64–68. doi: 10.1016/j.ygyno.2014.08.012

Lin, J., Piper, M., Perdue, L., Rutter, C., Webber, E., O'Connor, E., et al. (2016). Screening for colorectal cancer: updated evidence report and systematic review for the US Preventive Services Task Force. *JAMA* 315, 2576–2594. doi: 10.1001/jama.2016.3332

Liou, Y.-L., Zhang, T.-L., Yan, T., Yeh, C.-T., Kang, Y.-N., Cao, L., et al. (2016). Combined clinical and genetic testing algorithm for cervical cancer diagnosis. *Clin. Epigenetics* 8, 66. doi: 10.1186/s13148-016-0232-3

Lombardi, M. Y., and Assem, M., (2017). *Glioblastoma Genomics: A Very Complicated Story*. ed. S. De Vleeschouwer, Brisbane, AU: Codon Publications, p. 3–25. doi: 10.15586/codon.glioblastoma.2017.ch1

Louis, D. N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W. K., et al. (2016). The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol.* 131, 803–820. doi: 10.1007/s00401-016-1545-1

Low, S.-K., Zembutsu, H., and Nakamura, Y. (2018). Breast cancer: the translation of big genomic data to cancer precision medicine. *Cancer Sci.* 109, 497–506. doi: 10.1111/cas.13463

Maier, S., Nimmrich, I., Koenig, T., Eppenberger-Castori, S., Bohlmann, I., Paradiso, A., et al. (2007). DNA-methylation of the homeodomain transcription factor PITX2 reliably predicts risk of distant disease recurrence in tamoxifen-treated, node-negative breast cancer patients—technical and clinical validation in a multi-centre setting in collaboration wi. *Eur. J. Cancer* 43, 1679–1686. doi: 10.1016/j.ejca.2007.04.025

Marielle, H., Mark, E., Jake, M., Dorian, P., Brieuc, C., Jason, T., et al. (2017). Validation of Nu.QTM colorectal cancer screening triage test to identify FIT positive individuals at low risk of screen relevant neoplasia. *Ann. Oncol.* 28 (3), mdx262.021. doi: 10.1093/annonc/mdx262.021

Meisel, J. L., Venur, V. A., Gnant, M., and Carey, L. (2018). Evolution of targeted therapy in breast cancer: where precision medicine began. *Am. Soc. Clin. Oncol. Educ. Book*, 38, 78–86. doi: 10.1200/EDBK_201037

Mitchell, S. M., Ho, T., Brown, G. S., Baker, R. T., Thomas, M. L., McEvoy, A., et al. (2016). Evaluation of methylation biomarkers for detection of circulating tumor DNA and application to colorectal cancer. *Genes (Basel).* 7, 125, doi: 10.3390/genes7120125

Moran, S., Martínez-Cardús, A., Sayols, S., Musulén, E., Balañá, C., Estival-Gonzalez, A., et al. (2016). Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. *Lancet Oncol.* 17, 1386–1395. doi: 10.1016/S1470-2045(16)30297-2

Morikawa, T., Kato, J., Yamaji, Y., Wada, R., Mitsushima, T., and Shiratori, Y. (2005). A comparison of the immunochemical fecal occult blood test and total colonoscopy in the asymptomatic population. *Gastroenterology* 129, 422–428. doi: 10.1016/j.gastro.2005.05.056

Niu, F., Wen, J., Fu, X., Li, C., Zhao, R., Wu, S., et al. (2017). Stool DNA test of methylated syndecan-2 for the early detection of colorectal neoplasia. *Cancer Epidemiol. Biomarkers Prev.* 26, 1411–1419. doi: 10.1158/1055-9965.EPI-17-0153

Oh, T. J., Oh, H. Il, Seo, Y. Y., Jeong, D., Kim, C., Kang, H. W., et al. (2017). Feasibility of quantifying SDC2 methylation in stool DNA for early detection of colorectal cancer. *Clin. Epigenetics* 9, 126. doi: 10.1186/s13148-017-0426-3

Oh, T., Kim, N., Moon, Y., Kim, M. S., Hoehn, B. D., Park, C. H., et al. (2013). Genome-wide identification and validation of a novel methylation biomarker, SDC2, for blood-based detection of colorectal cancer. *J. Mol. Diagnostics* 15, 498–507. doi: 10.1016/j.jmoldx.2013.03.004

Ørntoft, M., Nielsen, H., Ørntoft, T., and Andersen, C. (2015). Performance of the colorectal cancer screening marker Sept9 is influenced by age, diabetes and arthritis: a nested case-control study. *BMC Cancer* 819. doi: 10.1186/s12885-015-1832-6

Panagopoulos, I., Gorunova, L., Leske, H., Niehusmann, P., Johannessen, L. E., Staurseth, J., et al. (2018). Pyrosequencing analysis of MGMT promoter methylation in meningioma. *Cancer Genomics Proteomics* 15, 379–385. doi: 10.21873/cgp.20096

Park, Y. S., Kim, D. S., Cho, S. W., Park, J. W., Jeon, S. J., Moon, T. J., et al. (2018). Analysis of syndecan-2 methylation in bowel lavage fluid for the detection of colorectal neoplasm. *Gut Liver* 12, 508–515. doi: 10.5009/gnl17357

Pavlidis, N., and Fizazi, K. (2009). Carcinoma of unknown primary (CUP). *Crit. Rev. Oncol. Hematol.* 69, 271–278. doi: 10.1016/j.critrevonc.2008.09.005

Pavlidis, N., and Pentheroudakis, G. (2012). Cancer of unknown primary site. *Lancet* 379, 1428–1435. doi: 10.1016/S0140-6736(11)61178-1

Pentheroudakis, G., Pavlidis, N., Fountzilas, G., Krikelis, D., Goussia, A., Stoyianni, A., et al. (2013). Novel microRNA-based assay demonstrates 92% agreement with diagnosis based on clinicopathologic and management data in a cohort of patients with carcinoma of unknown primary. *Mol. Cancer* 12, 57. doi: 10.1186/1476-4598-12-57

Perkins, J., Napieralski, R., Schricker, G., Piednoir, E., Manner, O., Bona, A., et al. (2018). TherascreenPITX2 RGQ PCR assay for the assessment of PITX2 DNA-methylation status to investigate the role of the transcription factor PITX2 and the regulation of the Wnt/ß-catenin pathway in pathophysiological processes. *Protoc. Exch.* doi: 10.1038/protex.2018.022

Potter, N., Hurban, P., White, M., Whitlock, K., Lofton-Day, C., Tetzner, R., et al. (2014). Validation of a real-time PCR-based qualitative assay for the detection of methylated SEPT9 DNA in human plasma. *Clin. Chem.* 60, 1183–1191. doi: 10.1373/clinchem.2013.221044

Pugh, S., Thiébaut, R., Bridgewater, J., Grisoni, M.-L., Moutasim, K., Rousseau, F., et al. (2017). Association between miR-31-3p expression and cetuximab efficacy in patients with KRAS wild-type metastatic colorectal cancer: a post-hoc analysis of the New EPOC trial. *Oncotarget*. 8, 93856–93866. doi: 10.18632/oncotarget.21291

Quillien, V., Lavenu, A., Ducray, F., Meyronet, D., Chinot, O., Fina, F., et al. (2017). Clinical validation of the CE-IVD marked Therascreen MGMT kit in a cohort of glioblastoma patients. *Cancer Biomarkers* 20, 435–441. doi: 10.3233/CBM-170191

Quillien, V., Lavenu, A., Karayan-Tapon, L., Carpentier, C., Labussiãre, M., Lesimple, T., et al. (2012). Comparative assessment of 5 methods (methylation-specific polymerase chain reaction, methylight, pyrosequencing, methylation-sensitive high-resolution melting, and immunohistochemistry) to analyze O6-methylguanine-DNA-methyltranferase in a series of 100. *Cancer* 118, 4201–4211. doi: 10.1002/cncr.27392

Rahier, J., Druez, A., Faugeras, L., Martinet, J., Géhénot, M., Josseaux, E., et al. (2017). Circulating nucleosomes as new blood-based biomarkers for detection of colorectal cancer. *Clin. Epigenetics*. 9, 53. doi: 10.1186/s13148-017-0351-5

Ramon, L., David, C., Fontaine, K., Lallet, E., Marcaillou, C., Martin-Lannerée, S., et al. (2018). Technical validation of a reverse-transcription quantitative polymerase chain reaction in vitro diagnostic test for the determination of MiR-31-3p expression levels in formalin-fixed paraffin-embedded metastatic colorectal cancer tumor specimens. *Biomark. Insights* 13, 1177271918763357. doi: 10.1177/1177271918763357

Schmidt, B., Liebenberg, V., Dietrich, D., Schlegel, T., Kneip, C., Seegebarth, A., et al. (2010). SHOX2 DNA methylation is a biomarker for the diagnosis of lung cancer based on bronchial aspirates. *BMC Cancer* 10, 600. doi: 10.1186/1471-2407-10-600

Schricker, G., Napieralski, R., Noske, A., Piednoir, E., Manner, O., Schüren, E., et al. (2018). Clinical performance of an analytically validated assay in comparison to microarray technology to assess PITX2 DNA-methylation in breast cancer. *Sci. Rep.* 8, 16861. doi: 10.1038/s41598-018-34919-1

Shen-Gunther, J., Wang, C.-M., Poage, G. M., Lin, C.-L., Perez, L., Banks, N. A., et al. (2016). Molecular Pap smear: HPV genotype and DNA methylation of ADCY8, CDH8, and ZNF582 as an integrated biomarker for high-grade cervical cytology. *Clin. Epigenetics* 8, 96. doi: 10.1186/s13148-016-0263-9

Shen, C., Changxin, Z., Wei, L., Shanping, S., Anqi, Z., Wenqiang, T., et al. (2018). Methylated septin 9 gene for noninvasive diagnosis and therapy monitoring of breast cancer. *Transl. Cancer Res.* 7, 587–599. doi: 10.21037/tcr.2018.05.24

Sheng, X., Guo, Y., and Lu, Y. (2017). Prognostic role of methylated GSTP1, p16, ESR1 and PITX2 in patients with breast cancer. *Medicine (Baltimore)* 96, e7476. doi: 10.1097/MD.0000000000007476

Siegel, R. L., Miller, K. D., and Jemal, A. (2017). Cancer statistics, 2017. *CA. Cancer J. Clin.* 67, 7–30. doi: 10.3322/caac.21387

Siegel, R., Naishadham, D., and Jemal, A. (2012). Cancer statistics, 2012. *CA. Cancer J. Clin.* 62, 10–29. doi: 10.3322/caac.20138

Song, L., and Li, Y. (2015). Methylated Sept9 gene is a sensitive biomarker for all stages of colorectal cancer. *iMedPub J.* 1 (1), 1–7. doi: 10.21767/2471-9943.100002

Song, L., and Wang, J., Wang, H., Chen, Y., Jia, J., Guo, S., et al. (2018). The quantitative profiling of blood mSEPT9 determines the detection performance on colorectal tumors. *Epigenomics* 10, 1569–1583. doi: 10.2217/epi-2017-0154

Taieb, J., Tabernero, J., Mini, E., Subtil, F., Folprecht, G., Van Laethem, J., et al. (2014). Oxaliplatin, fluorouracil, and leucovorin with or without cetuximab in patients with resected stage III colon cancer (PETACC-8): an open-label, randomised phase 3 trial. *Lancet Oncol.*, 15 (8), 862–873. doi: 10.1016/S1470-2045(14)70227-X

Torre, L. A., Islami, F., Siegel, R. L., Ward, E. M., and Jemal, A. (2017). Global cancer in women: burden and trends. *Cancer Epidemiol. Biomarkers Prev.* 26, 444–457. doi: 10.1158/1055-9965.EPI-16-0858

Torre, L. A., Siegel, R. L., Ward, E. M., and Jemal, A. (2016). Global cancer incidence and mortality rates and trends—an update. *Cancer Epidemiol. Biomarkers Prev.* 25, 16–27. doi: 10.1158/1055-9965.EPI-15-0578

van Lanschot, M. C. J., Carvalho, B., Coupé, V. M. H., van Engeland, M., Dekker, E., and Meijer, G. A. (2017). Molecular stool testing as an alternative for surveillance colonoscopy: a cross-sectional cohort study. *BMC Cancer* 17, 116. doi: 10.1186/s12885-017-3078-y

Varadhachary, G. R., and Raber, M. N. (2014). Cancer of unknown primary site. *N. Engl. J. Med.* 371, 757–765. doi: 10.1056/NEJMra1303917

Vu, M., Yu, J., Awolude, O. A., and Chuang, L. (2018). Cervical cancer worldwide. *Curr. Probl. Cancer* 42, 457–465. doi: 10.1016/j.currproblcancer.2018.06.003

Weiss, G., Schlegel, A., Kottwitz, D., König, T., and Tetzner, R. (2017). Validation of the SHOX2/PTGER4 DNA Methylation marker panel for plasma-based discrimination between patients with malignant and nonmalignant lung disease. *J. Thorac. Oncol.* 12, 77–84. doi: 10.1016/j.jtho.2016.08.123

Wentzensen, N., Sherman, M. E., Schiffman, M., and Wang, S. S. (2009). Utility of methylation markers in cervical cancer early detection: appraisal of the state-of-the-science. *Gynecol. Oncol.* 112, 293–299. doi: 10.1016/j.ygyno.2008.10.012

Widschwendter, M., Siegmund, K. D., Müller, H. M., Fiegl, H., Marth, C., Müller-Holzner, E., et al. (2004). Association of breast cancer DNA methylation profiles with hormone receptor status and response to tamoxifen. *Cancer Res.* 64, 3807–3813. doi: 10.1158/0008-5472.CAN-03-3852

Zawlik, I., Vaccarella, S., Kita, D., Mittelbronn, M., Franceschi, S., and Ohgaki, H. (2009). Promoter methylation and polymorphisms of the MGMT gene in glioblastomas: a population-based study. *Neuroepidemiology* 32, 21–29. doi: 10.1159/000170088

Zhao, H., Wang, S., Song, C., Zha, Y., and Li, L. (2016). The prognostic value of MGMT promoter status by pyrosequencing assay for glioblastoma patients' survival: a meta-analysis. *World J. Surg. Oncol.* 14, 261. doi: 10.1186/s12957-016-1012-4

Zou, H., Allawi, H., Cao, X., Domanico, M., Harrington, J., Taylor, W., et al. (2012). Quantification of methylated markers with a multiplex methylation-specific technology. *Clin. Chem.* 58, 375–383. doi: 10.1373/clinchem.2011.171264

Conflict of Interest Statement: JLG-G is the Chief Executive Officer and SM-M is the Chief Scientific Officer of EpiDisease S.L. Both own shares in EpiDisease SL., an epigenetics company that focuses on developing epigenetic biomarkers. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Beltrán-García, Osca-Verdegal, Mena-Mollá and García-Giménez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 03 July 2019
doi: 10.3389/fgene.2019.00613

Gut Microbiota Composition Is Associated With the Global DNA Methylation Pattern in Obesity

Bruno Ramos-Molina[1,2], Lidia Sánchez-Alcoholado[1,2], Amanda Cabrera-Mulero[1,2], Raul Lopez-Dominguez[3], Pedro Carmona-Saez[3], Eduardo Garcia-Fuentes[2,4], Isabel Moreno-Indias[1,2]* and Francisco J. Tinahones[1,2]*

[1] Deparment of Endocrinology and Nutrition, Virgen de la Victoria University Hospital, Institute of Biomedical Research in Malaga (IBIMA) and University of Malaga, Malaga, Spain, [2] CIBER Physiopathology of Obesity and Nutrition (CIBERobn), Institute of Health Carlos III, Madrid, Spain, [3] Bioinformatics Unit, Centre for Genomics and Oncological Research: Pfizer/University of Granada/Andalusian Regional Government, PTS, Granada, Spain, [4] Department of Gastroenterology, Virgen de la Victoria University Hospital, Institute of Biomedical Research in Malaga (IBIMA) and University of Malaga, Malaga, Spain

OPEN ACCESS

Edited by:
Yun Liu,
Fudan University, China

Reviewed by:
Apiwat Mutirangura,
Chulalongkorn University, Thailand
Neil Youngson,
University of New South Wales, Australia

***Correspondence:**
Bruno Ramos-Molina
bruno.ramos@ibima.eu
Isabel Moreno-Indias
isabel.moreno@ibima.eu

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *12 April 2019*
Accepted: *12 June 2019*
Published: *03 July 2019*

Citation:
Ramos-Molina B, Sánchez-Alcoholado L, Cabrera-Mulero A, Lopez-Dominguez R, Carmona-Saez P, Garcia-Fuentes E, Moreno-Indias I and Tinahones FJ (2019) Gut Microbiota Composition Is Associated With the Global DNA Methylation Pattern in Obesity. Front. Genet. 10:613. doi: 10.3389/fgene.2019.00613

Objective: Obesity and obesity-related metabolic diseases are characterized by gut microbiota and epigenetic alterations. Recent insight has suggested the existence of a crosstalk between the gut microbiome and the epigenome. However, the possible link between alterations in gut microbiome composition and epigenetic marks in obesity has been not explored yet. The aim of this work is to establish a link between the gut microbiota and the global DNA methylation profile in a group of obese subjects and to report potential candidate genes that could be epigenetically regulated by gut microbiota in adipose tissue.

Methods: Gut microbiota composition was analyzed in DNA stool samples from 45 obese subjects by 16S ribosomal RNA (rRNA) gene sequencing. Twenty patients were selected based on their Bacteroidetes-to-Firmicutes ratio (BFR): HighBFR group (BFR > 2.5, n = 10) and LowBFR group (BFR < 1.2, n = 10). Genome-wide analysis of DNA methylation pattern in both whole blood and visceral adipose tissue of these selected patients was performed with an Infinium EPIC BeadChip array-based platform. Gene expression analysis of candidate genes was done in adipose tissue by real-time quantitative PCR.

Results: Genome-wide analysis of DNA methylation revealed a completely different DNA methylome pattern in both blood and adipose tissue in the low BFR group vs. the high BFR group. Two hundred fifty-eight genes were differentially methylated in both blood and adipose tissue, of which several potential candidates were selected for gene expression analysis. We found that in adipose tissue, both *HDAC7* and *IGF2BP2* were hypomethylated and overexpressed in the low BFR group compared with the high BFR group. β values of both genes significantly correlated with the BFR ratio and the relative abundance of Bacteroidetes and/or Firmicutes.

Conclusions: In this study, we demonstrate that the DNA methylation status is associated with gut microbiota composition in obese subjects and that the expression levels of candidate genes implicated in glucose and energy homeostasis (e.g., *HDAC7* and *IGF2BP2*) could be epigenetically regulated by gut bacterial populations in adipose tissue.

Keywords: obesity, gut microbiota, methylation, epigenetics, adipose tissue

INTRODUCTION

Obesity has reached a pandemic scale worldwide, mainly caused by changes in lifestyles that include regular consumption of high-calorie food and a critical reduction of physical activity. Emerging evidence suggests that an altered composition and diversity of gut microbiota could play an important role in the development of obesity and related metabolic disorders such as type 2 diabetes (T2D) or non-alcoholic fatty liver disease (Cani, 2013; Han and Lin, 2014; Moreno-Indias et al., 2014; Leung et al., 2016; Cani, 2019). The relative amount of the two dominant phyla in gut microbiota, Firmicutes and Bacteroidetes, is altered in obesity conditions both in humans and in animal models (Ley et al., 2005; Ley et al., 2006; Turnbaugh et al., 2006; Verdam et al., 2013). Besides, the Bacteroidetes-to-Firmicutes ratio (BFR) has been widely associated with the inflammatory and metabolic state in obesity (Cani et al., 2009; de La Serre et al., 2010; Verdam et al., 2013). Several mechanisms have been proposed as a link between obesity and gut microbiota, for instance, the production of microbial metabolites that regulate energy metabolism, metabolic endotoxemia, or the modulation of the secretion of hormones by intestinal cells (Cani, 2019).

Epigenome captures environmental and lifestyle events. Recent insight has suggested a role of epigenetics in the development of obesity and related metabolic disorders (van Dijk et al., 2015; Davegardh et al., 2018). More recently, the existence of a crosstalk between the gut microbiome and the epigenome has been suggested (Qin and Wade, 2018). It has been proposed that certain metabolites generated by the gut microbiota such as short-chain fatty acids (SCFAs), folate, and polyamines can act as epigenetic modulators by affecting DNA methylation and inducing histone modifications (Crider et al., 2012; Paul et al., 2015; Bhat and Kapila, 2017; Soda, 2018; Cuevas-Sierra et al., 2019; Ramos-Molina et al., 2019). However, the possible link between alterations in gut microbiome composition and epigenetic marks in the context of obesity has been not explored yet.

In this work, we have established a link between the gut microbiota and the global DNA methylation profile in a group of obese subjects by integrating 16S rRNA gene sequence analysis and epigenome-wide association studies, and we have reported potential candidate genes that could be epigenetically regulated by gut microbiota in adipose tissue.

MATERIAL AND METHODS

Study Participants

This is a cross-sectional analysis of 45 morbidly obese subjects [body mass index (BMI) > 40 kg/m^2] who were consecutively recruited at the Virgen de la Victoria University Hospital for bariatric surgery (Malaga, Spain) from 2015 to 2017. All participants provided written informed consent, and the study protocol and procedures were approved according to the ethical standards of the Declaration of Helsinki by the Research Ethics Committees from all the participating institutions.

Laboratory Measurements

Blood samples were obtained from the antecubital vein and placed in vacutainer tubes after an overnight fast. The serum was separated by centrifugation for 15 min at 4,000 rpm at 4°C and frozen at −80°C until analysis. Enzymatic methods (Randox Laboratories Ltd). were employed to analyze the levels of serum cholesterol, triglycerides, HDL-cholesterol, glucose, and glycosylated hemoglobin (HbA1c) using a Dimension Vista autoanalyzer (Siemens Healthcare Diagnostics). Serum insulin levels were measured by immunoassay using an ADVIA Centaur autoanalyzer (Siemens Healthcare Diagnostics). Insulin resistance (IR) was calculated from the homeostasis model assessment of IR (HOMA-IR) with the following formula: HOMA-IR = [fasting serum insulin (μU/ml) × fasting blood glucose (mmol/L)]/22.5.

Gut Microbiota Analysis

Stool samples were collected and immediately frozen at −80°C until DNA extraction. DNA was extracted from fecal samples using the QIAamp DNA Stool Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. Ribosomal 16S rRNA gene sequences were amplified from cDNA using the 16S Metagenomics Kit (Thermo Fisher Scientific, Italy). The kit included two primer sets that selectively amplify the corresponding hypervariable regions of the 16S region in bacteria: primer set V2–4–8 and primer set V3–6, 7–9. Libraries were created using the Ion Plus Fragment Library Kit (Thermo Fisher Scientific). Barcodes were added to each sample using the Ion Xpress Barcode Adapters kit (Thermo Fisher Scientific). Emulsion PCR and sequencing of the amplicon libraries were performed on an Ion 520 chip (Ion 520™ Chip Kit) using the Ion Torrent S5™ system and the Ion 520™/530™ Kit-Chef (Thermo Fisher Scientific) according to the manufacturer's instructions.

Base calling and run demultiplexing were performed by using Torrent Suite™ Server software (Thermo Fisher), version 5.4.0, with default parameters for the 16S Target Sequencing (bead loading ≤ 30, key signal ≤ 30, and usable sequences ≤ 30). Quality sequences were analyzed using QIIME 1.9.1 software. Briefly, the workflow was the following: operational taxonomic units (OTUs) were calculated by clustering sequences at a similarity of 97% with a closed-reference OTU picking approach. The representative sequences were submitted to the UCLUST to obtain the taxonomy assignment and the relative abundance of each OTU using the Greengenes 16S rRNA gene database. OTUs were collapsed to phylum level in order to calculate the BFR. Raw data can be found in the SRA database public repository from NCBI within the BioProject accession number PRJNA539905.

DNA Methylation Profiling Using Universal Bead Array

Visceral adipose tissue (VAT) was obtained during bariatric surgery. Biopsy samples were washed in physiological saline and immediately frozen at −80°C until analysis. DNA was extracted from blood and VAT using Zymo ZR 96 Quick gDNA kit (Zymo

Research Corp., Irvine, CA, USA) following manufacturer's instructions. After quantification and purity assessment, a total of 500 ng of genomic DNA was bisulfite treated using the ZymoResearch Infinitum HD FFPE Methylation kit (Zymo Research Corp, Irvine, CA, USA) and was purified by DNA-Clean-Up kit (Zymo Research Corp, Irvine, CA, USA). Over 850,000 methylation sites were interrogated with the Infinium Methylation EPIC Bead Chip Kit (Illumina, San Diego, CA, USA) following the Infinium HD Assay Methylation protocol, and raw data (idat files) were obtained from iScan (Illumina) software.

Methylation Data Analysis

Raw data files (idat files) were processed to derive beta values after background correction and normalization by BMIQ (Teschendorff et al., 2013). The beta value is the ratio of the methylated probe intensity and the overall intensity, which resulted from the sum of methylated and unmethylated probe intensities. The beta value results in a number between 0 and 1, in which a value of zero indicates that all copies of the CpG site in the sample were completely unmethylated and a value of one indicates that every copy of the site was methylated (Du et al., 2010). Differential methylation, gene set enrichment, and pathway analyses were performed using Partek Genomics Suit with Pathway (version 7.0). To complete the analysis, we used EnrichR (https://amp.pharm.mssm.edu/Enrichr/) (Chen et al., 2013) and the analysis of some ontologies such as Online Mendelian Inheritance in Man (OMIM) diseases. Raw data can be found in the GEO database public repository from NCBI within the accession number GSE131461.

Gene Expression Analysis

Frozen VAT was homogenized with an Ultra-Turrax 8 (Ika, Staufen, Germany). Total RNA was extracted by RNeasy lipid tissue midi kit (QIAGEN Science, Hilden, Germany) and treated with 55 U of RNase-free deoxyribonuclease (QIAGEN Science, Hilden, Germany), following the manufacturer's instructions. RNA purity and concentration were determined by 260/280 absorbance ratios on a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA). Total purified RNA integrity was checked by denaturing agarose gel electrophoresis and SYBR Safe DNA gel staining (Invitrogen). Total RNA was reverse transcribed to cDNA by a high-capacity cDNA reverse transcription kit with RNase inhibitor (Applied Biosystems, Foster City, CA). Quantitative real-time PCR with duplicates was done with the cDNA. The amplifications were performed using a MicroAmpH Optical 96-well reaction plate (Applied Biosystems, Foster City, CA) on an ABI 7500 Fast Real-Time PCR System (Applied Biosystems, Foster City, CA). Commercially available and pre-validated TaqMan® primer/probe sets were used as follows: cyclophilin A (*PPIA*, 4333763), used as endogenous control for the target gene in each reaction; fibroblast growth factor 1 (*FGF1*, Hs01092738_m1); fibroblast growth factor 10 (*FGF10*, Hs00610298_m1); lysine demethylase 4B (*KDM4B*, Hs00392119_m1); interleukin-7 (*IL7*, Hs00174202_m1); insulin-like growth factor 2 mRNA-binding protein 2 (*IGF2BP2*, Hs01118009_m1); histone deacetylase 7 (*HDAC7*, Hs01045864_m1); ER degradation enhancing alpha-mannosidase-like protein 1 (*EDEM1*, Hs00976004_m1); activating transcription factor 6 (*ATF6*, Hs00232586_m1); and cyclin-dependent kinase 6 (*CDK6*, Hs01026371_m1). A threshold cycle (*Ct* value) was obtained for each amplification curve and normalized by subtracting the *Ct* value of the endogenous gene and expressed as ΔCt value and expressed in linear scale as $2^{-\Delta Ct}$.

Statistical Analysis

Continuous variables are summarized as means ± SD or SE. Discrete variables are presented as frequencies and percentages. Differences in clinical characteristics between two groups were analyzed using the Mann–Whitney *U* test. The Spearman correlation coefficients were calculated to estimate the correlations between variables. Statistical analyses were carried out with the statistical software package SPSS version 15.0 (SPSS Inc., Chicago, IL, United States). Values were considered to be statistically significant when the $p < 0.05$. Association analysis between phenotypes and probes was assessed with the R package CpGassoc in R 3.3.3 (Barfield et al., 2012). FDR-corrected $p < 0.01$ was considered statistically significant.

RESULTS

Analysis of the gut microbiota composition was performed in a group of 45 patients with morbid obesity (**Table S1**). Moreover, to determine the potential contribution of gut microbiota composition to the global DNA methylome, we extracted genomic DNA from whole blood and VAT of these 45 patients and performed EWAS on the Illumina platform using the Infinium HumanEPIC BeadChip array. From the 45 patients with morbid obesity, 20 subjects were selected based on the relative abundances of the predominant phyla, Bacteroidetes and Firmicutes: high BFR (HighBFR group; BFR > 2.5; $n = 10$) vs. low BFR (LowBFR group; BFR < 1.2; $n = 10$) (**Table S2**). As expected, the HighBFR group ($n = 10$) exhibited predominance of the Bacteroidetes phylum ($p < 0.0001$), whereas Firmicutes was predominant in the LowBFR group ($n = 10$) subjects ($p < 0.0001$) (**Figure 1**). No statistical differences between other phyla such as Proteobacteria, Actinobacteria, or Fusobacteria were found between groups (**Figure 1**). The general characteristics of both study groups are summarized in **Table 1**. There were no significant differences in age, sex, and BMI between the two study groups. Glucose and HbA1c levels were significantly lower in the HighBFR group when compared with the LowBFR group ($p < 0.05$). There were no significant differences in HOMA-IR, HDL-cholesterol, and triglycerides between the two study groups ($p > 0.05$).

To determine the potential contribution of gut microbiota composition to the global DNA methylome, we extracted genomic DNA from whole blood and VAT and performed EWAS on the Illumina platform using the Infinium HumanEPIC BeadChip array. As shown in **Figure 2**, the two

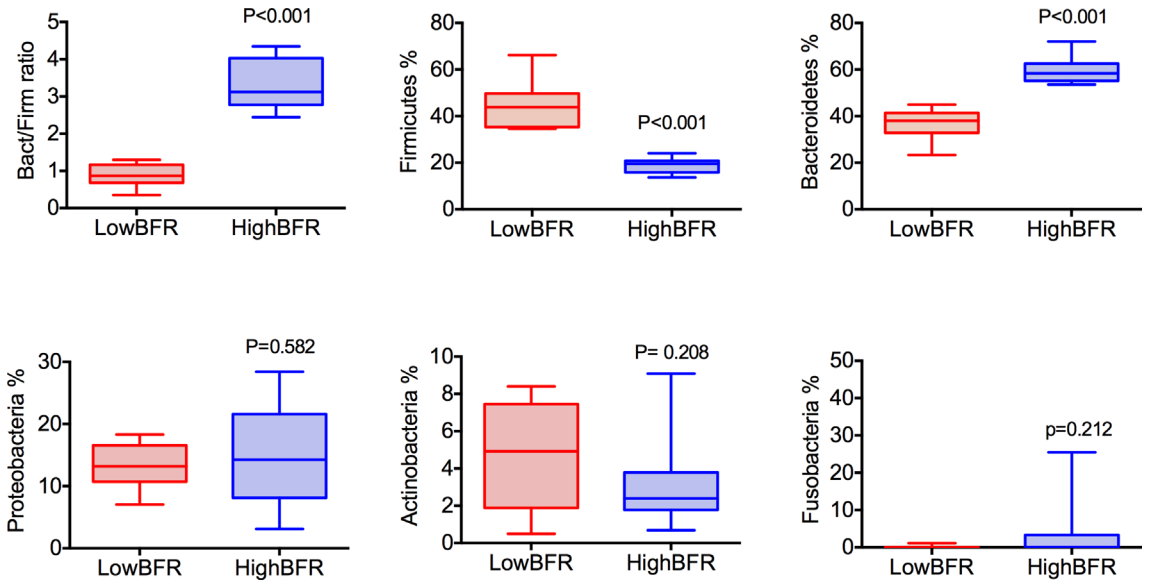


FIGURE 1 | Categorization of study patients into the HighBFR and LowBFR groups according to dominant bacterial phyla. Box plots (range min–max) show the Bacteroidetes-to-Firmicutes ratio and the relative abundance (%) of five more predominant bacterial phyla: Bacteroidetes, Firmicutes, Proteobacteria, Acidobacteria, and Fusobacteria. Comparisons between groups were performed using *t* test analysis (***$p < 0.0001$).

groups showed a different methylation profile in both whole blood and VAT (**Figure 2A and B**). We found 1,658 and 1,421 differentially methylated genes between study groups in whole blood and VAT, respectively (**Figure 2C**; **Tables S3 and S4**). We classified them as hypermethylated and hypomethylated, differentiating those that were significant in both whole blood and VAT from those that were only significant in whole blood or VAT (**Table S5**). Remarkably, 258 genes were differentially methylated both whole blood and VAT (**Figure 2C; Table S6**). Pathway enrichment analysis revealed that most of the genes differentially methylated in whole blood and VAT were involved in glycerophospholipid metabolism and cell adhesion molecules, respectively (**Table S7**). Moreover, a further enrichment analysis on an ontology basis such as OMIM diseases revealed that the top three categories enriched were related to diabetes. In order to better understand these results, a further association analysis between the probes and the phenotype characteristics of the patients was performed (**Table S8**). Many associations were found, but HOMA-IR, HbA1c, weight, and BMI were found to be the most relevant variables.

Following an exhaustive analysis of the list of genes, we focused on genes previously related to obesity, metabolic disease, and/or T2D. Thus, we tested the impact of changes in the methylation levels on the mRNA expression levels in VAT of the following genes: *FGF1*, *FGF10*, *KDM4B*, *HDCA7*, *IGF2BP2*, *IL7*, *EDEM1*, *ATF6*, and *CDK6* (Sharma et al., 2008; Makki et al., 2013; Ohta and Itoh, 2014; Dai et al., 2015; Cheng et al., 2018; Davegardh et al., 2018; Hou et al., 2018). As shown in **Figure S1**, *HDAC7* and *IGF2BP2* mRNA levels were significantly different between study groups; no differences in the expression levels were found for the rest

TABLE 1 | Baseline clinical and biochemical variables of the study subjects divided by Bact/Firm ratio.

	HighBFR (*n* = 10)	LowBFR (*n* = 10)	*p* value
Age, years	42.2 ± 10.6	46.1 ± 7.2	0.353
Gender, f (%)	8/10 (80)	8/10 (80)	—
Weight, kg	144.1 ± 18.9	133.8 ± 25.5	0.796
BMI, mg/kg^2	52.5 ± 5.4	49.4 ± 8.1	0.579
Glucose, mg/dl	93.0 ± 17.9	141.2 ± 75.3	0.015
HbA1c, %	5.7 ± 0.6	6.8 ± 2.0	0.011
HOMA-IR	6.3 ± 4.0	10.7 ± 6.6	0.089
HDL-cholesterol, mg/dl	42.5 ± 10.1	48.5 ± 9.4	0.247
Triglycerides, mg/dl	168.2 ± 111.9	147.0 ± 108.1	0.796

Data are means ± SD. p values were calculated for the difference between study groups using Mann–Whitney U test. $p < 0.05$ was considered significant. HOMA-IR, homeostasis model assessment of IR; HDL, high-density lipoprotein.

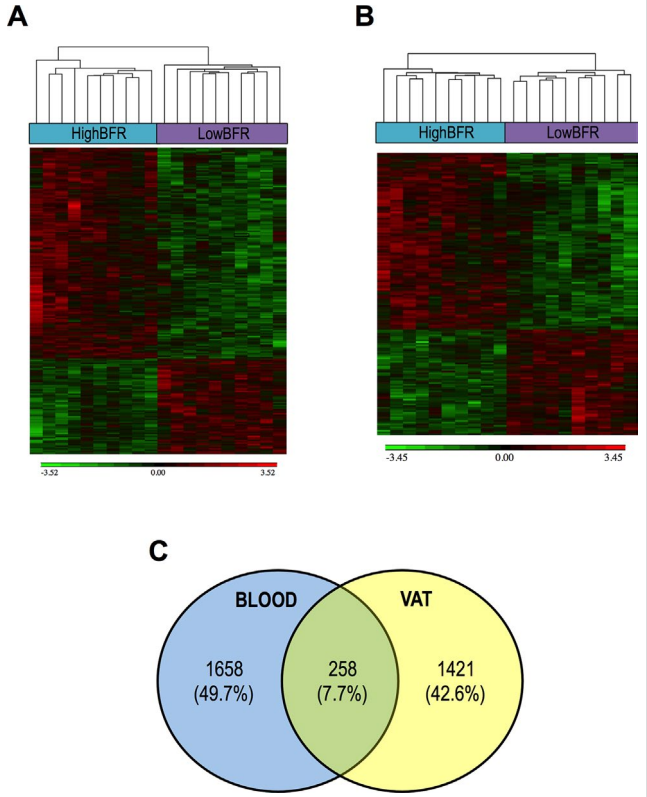


FIGURE 2 | Association between the gut microbiota composition and the DNA methylome in blood **(A)** and visceral adipose tissue (VAT) **(B)**. The analysis of methylation data yielded a set of genes that were significant differentially methylated among subjects with a predominance of either Bacteroidetes (HighBFR) or Firmicutes (LowBFR) in the gut microbial population. Green indicates decreased and red indicates increased methylation in the LowBFR group compared with the promoter methylation in the HighBFR group. **(C)** Venn diagram of genes differentially methylated in both whole blood and visceral adipose tissue.

of the analyzed genes. Further, we assessed the association between the gut microbiota composition and the methylation levels of both *HDAC7* and *IGF2BP2*. As shown in **Figure 3A**, the β values of *HDAC7* (indicatives of the DNA methylation status of the gene) were significantly higher in the HighBFR group in both whole blood and VAT. Correlation analysis in the whole cohort of obese subjects ($n = 45$) demonstrated that the β values of *HDAC7* in both whole blood and VAT were positively associated with the BFR (**Figure 3B**). Additionally, whereas β values of *HDAC7* in blood correlated negatively with the relative abundance of Firmicutes (**Figure 3C**), β values of HDAC7 in VAT were positively associated with the relative abundance of Bacteroidetes (**Figure 3D**). Like *HDAC7*, the β values of *IGF2BP2* were significantly higher in the HighBFR group in both whole blood and VAT (**Figure 4A**). Furthermore, *IGF2BP2* β values in VAT significantly correlated with the BFR and the relative abundance of Bacteroidetes (**Figure 4B**). No significant correlation was observed between gut microbiota composition and β values of *IGF2BP2* in whole blood. It is noteworthy that some of these correlations remain still significant when patients with the most extreme BFR values were excluded (validation cohort; $n = 25$). Thus, we found that the relative abundance of Bacteroidetes positively correlated with the methylation levels of *HDAC7* ($r = 0.500$, $p = 0.011$) and *IGF2BP2* ($r = 0.597$, $p = 0.002$) in adipose tissue, and the relative abundance of Firmicutes negatively correlated with the methylation levels of *HDAC7* ($r = -0.465$, $p = 0.019$) in whole blood. These results reinforce the relationship between gut microbiota and DNA methylation within these genes.

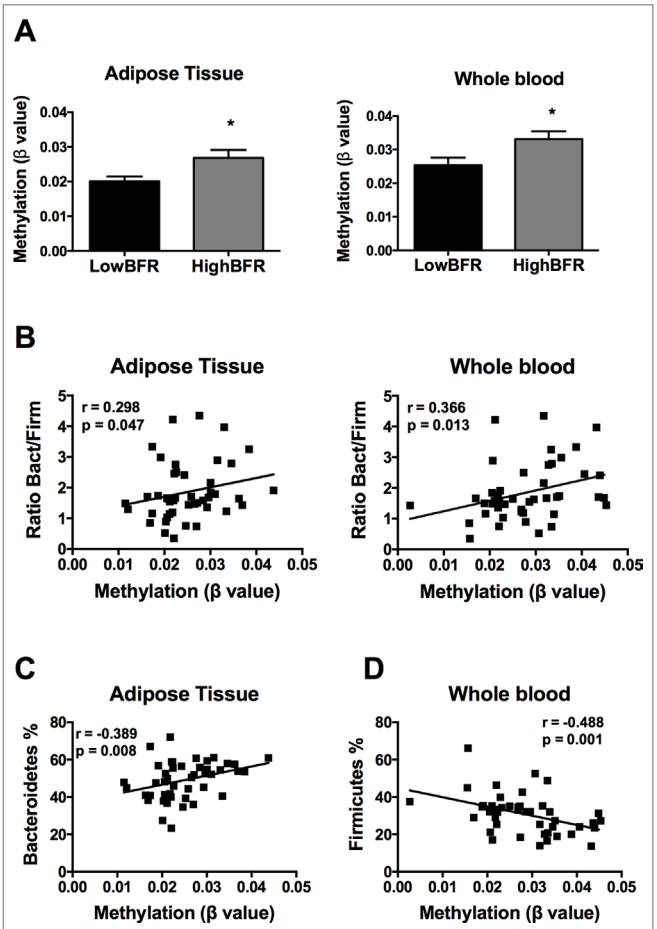


FIGURE 3 | Association between the DNA methylation status of histone deacetylase 7 (HDAC7) and the gut microbiota composition in obese subjects. **(A)** Methylation of *HDAC7* (β value) in the HighBFR vs. LowBFR groups in both visceral adipose tissue and whole blood. Data ($n = 10$ per group) are plotted as means ± SE. Significance was tested using Mann–Whitney *U* test and is indicated as $^*p < 0.05$. **(B)** Spearman correlations between *HDAC7* methylation and the ratio Bact/Firm in both visceral adipose tissue and whole blood. **(C)** Spearman correlation between *HDAC7* methylation and the relative abundance (%) of Firmicutes in whole blood. **(D)** Spearman correlation between *HDAC7* methylation and the relative abundance (%) of Bacteroidetes in visceral adipose tissue.

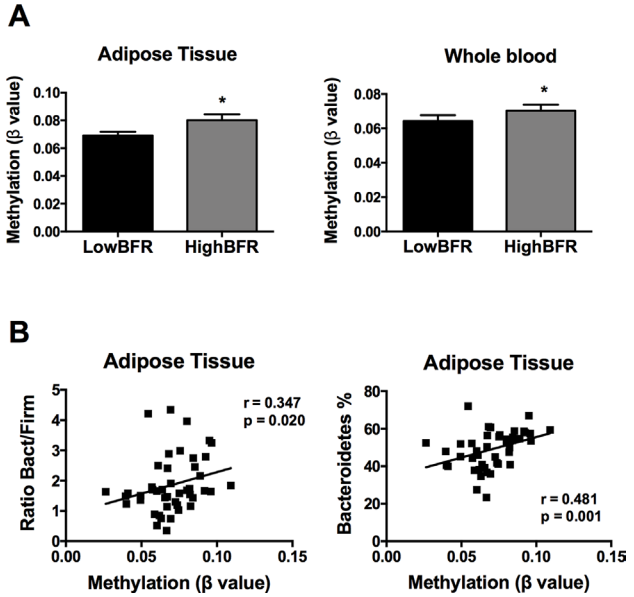


FIGURE 4 | Association between the DNA methylation status of insulin like growth factor 2 mRNA binding protein 2 (IGF2BP2) and the gut microbiota composition in obese subjects. **(A)** Methylation of *IGF2BP2* (β value) in the HighBFR vs. LowBFR groups in both visceral adipose tissue and whole blood. Data (n = 10 per group) are plotted as means ± SE. Significance was tested using Mann–Whitney U test and is indicated as $^*p < 0.05$. **(B)** Spearman correlations between *IGF2BP2* methylation in visceral adipose tissue and the ratio Bact/Firm or the relative abundance (%) of Bacteroidetes.

DISCUSSION

Obesity is a pathological condition highly associated with lifestyle. Epigenome and gut microbiota are two factors clearly impacted by lifestyle. Recent evidence has proposed that certain metabolites produced by microbial metabolism can influence the epigenetic profile in several conditions (Hullar and Fu, 2014; Cuevas-Sierra et al., 2019). Despite the possible role of gut microbiota as epigenetic regulator, the number of works associating gut microbiome and epigenetics is scarce. Moreover, most of these studies were focused on histone acetylation, with little attention paid to DNA methylation status. Here, we have demonstrated for the first time an association between the composition of certain bacterial populations in the gastrointestinal tract with specific DNA whole-genome methylation states in both blood samples and adipose tissue biopsies in the context of extreme obesity. Overall, the subjects included in the present study were characterized by a heterogeneous gut microbiota composition. However, we found that, independently on their clinical characteristics, classification of patients clustered into two groups according to their gut microbiota profile measured by the relative abundance of the predominant phyla Bacteroidetes and Firmicutes. These clusters of obese individuals presented similar BMI and clinical parameters related to lipid metabolism but significant differences in markers of glucose metabolism. In particular, individuals with low Bact/Firm ratio displayed higher levels of fasting glucose and HbA1c. Microbiota profile is influenced by the environmental conditions (Rothschild et al., 2018). Within the gut, microbiota is influenced by the host phenotype. Gut microbiota has been extensively related to glucose levels and metabolism, although a clear conclusion about the cause or consequence has not completely been achieved (Utzschneider et al., 2016). Thus, glucose levels could drive the clusters of these patients and could influence the gut microbiota profiles and consequently the Bact/Firm ratio used in the study. The classification of obese patients according to their Bact/Firm ratio showed a clear association between the relative abundance of these phyla with the DNA methylation profile in both blood and adipose tissue, supporting the idea that the gut microbiota could act as an epigenetic regulator in obesity, as previously indicated by others for other pathological conditions (Yang et al., 2013; Hullar and Fu, 2014; Sook Lee et al., 2017; Watson and Søreide, 2017; Qin and Wade, 2018). In fact, the furthest association analysis between the DNA methylation results and the phenotypes of the patients revealed that weight and BMI, as well as HOMA-IR and HbA1c levels, were the variables more related to DNA methylation status. Interestingly, the enrichment analysis based on OMIM diseases database showed that diabetes, and particularly type 2 diabetes, was the disease most related to the DNA methylation status, which mirrored the results showed through the clustering of the patients according to their Bact/Firm ratio.

Previous studies have suggested that gut microbiota may impact the epigenetic landscape of the host. In animal models, it has been previously shown that microbial metabolites such as SCFAs can influence epigenetic programming in various tissues, including proximal colon, liver, and white adipose tissue (Krautkramer et al., 2016). Because most of butyrate-producing bacteria belong to the Firmicutes phyla (Vital et al., 2014), differences in the Bact/Firm ratio within our cohort of obese individuals could result in different circulating levels of butyrate or other SCFAs, which would explain the observed differences in the DNA methylation status in both blood cells and adipose tissue. In addition to SCFAs, other metabolites produced by the bacteria from the gastrointestinal tract have been related to epigenetic modifications (Bhat and Kapila, 2017). In particular, gut bacteria can produce high levels of folic acid and polyamines, which are molecules highly related to carbon metabolism and therefore with potential impact in the DNA methylation status (Crider et al., 2012; Soda, 2018; Ramos-Molina et al., 2019). Nevertheless, whether the changes in the methylome associated with alterations in gut microbiome are related to changes in the levels of these or other bacterial metabolites requires further investigation.

As described above, in this study, we report for the first time a possible crosstalk between the gut microbiome and the DNA methylation state in obesity. Our results are supported by multiple studies performed in cohorts of non-obese individuals. For instance, a recent pilot study performed in pregnant women demonstrated an association between the relative abundance of dominant phyla (Bacteroidetes and Firmicutes) and the DNA methylation profile in blood samples (Kumar et al., 2014). In another interesting work, Kelly et al. reported an association between gut microbiota and histone methylation signature of intestinal epithelial cells in patients with inflammatory bowel syndrome (Kelly et al., 2018).

In a mouse model of diet-induced obesity, Qin et al. demonstrated that changes in the gut microbiome could result in epigenetic alterations associated with the development of colon cancer (Qin et al., 2018). There is no evidence, however, of a relationship between the composition of the gut microbiome and the methylation status in adipose tissue. Previous work from our lab demonstrated that the DNA methylation of certain genes related to adipogenesis and lipid metabolism is impaired in adipose tissue of subjects with metabolic syndrome (Castellano-Castillo et al., 2019). Nevertheless, whether these changes in the methylation pattern are related to differences in the gut microbiota composition remains unknown.

On the other hand, we have found that the promoters of both *HDAC7* and *IGF2BP2* genes were hypomethylated in whole blood and adipose tissue of the study patients with low Bact/Firm ratio. These genes were further studied based on their relationship with metabolism. However, it is worthy to mention that only two of the nine studied genes achieved a statistically significant difference between BFR groups, indicative of the complex machinery regulating gene expression and representing DNA methylation in only one of the mechanisms implicated. On the one hand, *HDAC7* gene encodes a histone deacetylase (HDAC). Histone deacetylase enzymes repress gene expression by removing an acyl group bound to chromatin. Although it is widely known that class I HDACs (mainly 1, 2, and 3) are inhibited by microbial products as SCFAs, mainly butyrate (Yuille et al., 2018), this is the first time that a class IIb HDAC is related with gut microbiota. In line with our results, a previous work demonstrated that the *HDAC7* gene was hypomethylated and overexpressed in islets from donors with T2D (Dayeh et al., 2014), which could have pathological implications given that *Hdac7* overexpression in rat islets and β-cell lines resulted in impaired insulin secretion (Daneshpajooh et al., 2017). Our results show that hypomethylation in the *HDAC7* promoter in both whole blood and adipose tissue is also associated with disturbances in glucose metabolism, as both study groups displayed marked differences in glucose and HbA1c levels. This suggests that the changes in the methylation profile in the *HDAC7* gene are related not only to the composition of the gut microbiota but also to the metabolic profile of the subjects, at least in blood and adipose tissue. However, further investigation is required to examine in detail the implication of the microbial population.

On the other hand, hypomethylation of *IGF2BP2* also resulted in higher mRNA levels in adipose tissue. In adipose tissue, *IGF2BP2* is able to downregulate the expression of *IGF2*, a growth factor that plays a pivotal role in controlling adipogenesis (Louveau and Gondret, 2004). Therefore, impaired *IGF2BP2* expression levels may contribute to the development of metabolic disorders such as obesity and T2D through alterations in the function of the adipose tissue. In this regard, inactivation of the *IGF2BP2* in mice induces resistance to diet-induced obesity and fatty liver due in part to increased energy expenditure, suggesting that *IGF2BP2* has an important role in the regulation of energy homeostasis (Dai et al., 2015). Thus, gut microbiota profile could be participating in the homeostasis of the host through the methylation of particular genes as *IGF2BP2*. Interestingly, these associations between gut microbiota and *HDAC7 and IGF2BP2* gene expression and methylation levels seem to be driven by the phylum Bacteroidetes. The major end products of Bacteroidetes are succinate, acetate, and, in some cases, propionate (Chakraborti, 2015). Methylation rates depend on the availability of one- and two-carbon substrates (Su et al., 2016). Acetate is a two-carbon substrate, while succinate is able to follow the tricarboxylic acid cycle. Thus, although classically Firmicutes has been the main phylum related to epigenetic modifications, Bacteroidetes could be more related to methylation and Firmicutes to acetylation modifications. However, phylum is a phylogenetic level that groups different microbial members with different SCFAs and other metabolites that should be carefully studied.

In conclusion, we demonstrate that the methylation status could be largely affected by the gut microbiota composition in obese subjects and that the expression levels of genes implicated in glucose and energy homeostasis (e.g., *HDAC7* and *IGF2BP2*) could be epigenetically regulated by gut bacterial populations in adipose tissue. In order to understand how gut microbiota can influence DNA methylation in adipose tissue and other target organs, further studies are needed.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the manuscript and the supplementary files.

ETHICS STATEMENT

All participants provided written informed consent, and the study protocol and procedures were approved according to the ethical standards of the Declaration of Helsinki by the Research Ethics Committees from the participating institution (Virgen de la Victoria University Hospital, Malaga, Spain)

AUTHOR CONTRIBUTIONS

BR-M, IM-I, and FT designed research. BR-M, LS-A, AC-M, RL-D, and PC-S conducted research. EG-F provided essential materials. BR-M and IM-I analyzed the data. BR-M, IM-I, and FT wrote the paper. BR-M, IM-I, and FT had primary responsibility for the final content. All authors read and approved the final manuscript.

FUNDING

This study was supported by the "Centros de Investigación Biomédica en Red" (CIBER) of the Institute of Health Carlos III (ISCIII) (CB06/03/0018) and research grants from the ISCIII (grant numbers PI15/01114 and PI18/01160) and co-financed by the European Regional Development Fund (ERDF). BR-M was a recipient of a Sara Borrell postdoctoral fellowship from the ISCIII (CD16/0003) and co-funded by the ERDF. IM-I was supported by the Miguel Servet Type I program (CP16/00163) from the ISCIII and co-funded by the ERDF.

ACKNOWLEDGMENTS

The authors thank the Metagenomic Platform of the CIBER Physiopathology of Obesity and Nutrition (CIBERobn), Institute of Health Carlos III (ISCIII), Madrid, Spain.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00613/full#supplementary-material

TABLE S1 | Baseline clinical characteristics of study subjects.

TABLE S2 | Relative abundance of bacterial phyla in fecal microbiota.

TABLE S3 | List of genes differentially methylated in whole blood.

TABLE S4 | List of genes differentially methylated in visceral adipose tissue.

TABLE S5 | Genes differentially hypomethylated and hypermethylated in both whole blood and visceral adipose tissue.

TABLE S6 | List of genes differentially methylated both in whole blood and visceral adipose tissue.

TABLE S7 | Pathway analysis.

TABLE S8 | Association analysis between methylation levels and the phenotype characteristics of the study patients.

REFERENCES

Barfield, R. T., Kilaru, V., Smith, A. K., and Conneely, K. N. (2012). CpGassoc: an R function for analysis of DNA methylation microarray data. *Bioinformatics* 28 (9), 1280–1281. doi: 10.1093/bioinformatics/bts124

Bhat, M. I., and Kapila, R. (2017). Dietary metabolites derived from gut microbiota: critical modulators of epigenetic changes in mammals. *Nutr. Rev.* 75 (5), 374–389. doi: 10.1093/nutrit/nux001

Cani, P. D. (2013). Gut microbiota and obesity: lessons from the microbiome. *Briefings Func. Genomics* 12 (4), 381–387. doi: 10.1093/bfgp/elt014

Cani, P. D. (2019). Microbiota and metabolites in metabolic diseases. *Nat. Rev. Endocrinol.* 15 (2), 69–70. doi: 10.1038/s41574-018-0143-9

Cani, P. D., Possemiers, S., Van de Wiele, T., Guiot, Y., Everard, A., Rottier, O., et al. (2009). Changes in gut microbiota control inflammation in obese mice through a mechanism involving GLP-2-driven improvement of gut permeability. *Gut* 58 (8), 1091–1103. doi: 10.1136/gut.2008.165886

Castellano-Castillo, D., Moreno-Indias, I., Sanchez-Alcoholado, L., Ramos-Molina, B., Alcaide-Torres, J., Morcillo, S., et al. (2019). Altered adipose tissue DNA methylation status in metabolic syndrome: relationships between global DNA methylation and specific methylation at adipogenic, lipid metabolism and inflammatory candidate genes and metabolic variables. *J. Clin. Med.* 8 (1). doi: 10.3390/jcm8010087

Chakraborti, C. K. (2015). New-found link between microbiota and obesity. *World J. Gastrointest. Pathophysiol.* 6 (4), 110–119. doi: 10.4291/wjgp.v6.i4.110

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., et al. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinf.* 14, 128. doi: 10.1186/1471-2105-14-128

Cheng, Y., Yuan, Q., Vergnes, L., Rong, X., Youn, J. Y., Li, J., et al. (2018). KDM4B protects against obesity and metabolic dysfunction. *Proc. Natl. Acad Sci. U. S. A.* 115 (24), E5566–E5575. doi: 10.1073/pnas.1721814115

Crider, K. S., Yang, T. P., Berry, R. J., and Bailey, L. B. (2012). Folate and DNA methylation: a review of molecular mechanisms and the evidence for folate's role. *Adv. Nutr.* 3 (1), 21–38. doi: 10.3945/an.111.000992

Cuevas-Sierra, A., Ramos-Lopez, O., Riezu-Boj, J. I., Milagro, F. I., and Martinez, J. A. (2019). Diet, gut microbiota, and obesity: links with host genetics and epigenetics and potential applications. *Adv. Nutr.* 10 (suppl_1), S17–S30. doi: 10.1093/advances/nmy078

Dai, N., Zhao, L., Wrighting, D., Kramer, D., Majithia, A., Wang, Y., et al. (2015). IGF2BP2/IMP2-Deficient mice resist obesity through enhanced translation of Ucp1 mRNA and Other mRNAs encoding mitochondrial proteins. *Cell Metab.* 21 (4), 609–621. doi: 10.1016/j.cmet.2015.03.006

Daneshpajooh, M., Bacos, K., Bysani, M., Bagge, A., Ottosson Laakso, E., Vikman, P., et al. (2017). HDAC7 is overexpressed in human diabetic islets and impairs insulin secretion in rat islets and clonal beta cells. *Diabetologia* 60 (1), 116–125. doi: 10.1007/s00125-016-4113-2

Davegardh, C., Garcia-Calzon, S., Bacos, K., and Ling, C. (2018). DNA methylation in the pathogenesis of type 2 diabetes in humans. *Mol. Metab.* 14, 12–25. doi: 10.1016/j.molmet.2018.01.022

Dayeh, T., Volkov, P., Salo, S., Hall, E., Nilsson, E., Olsson, A. H., et al. (2014). Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. *PLoS Genet.* 10 (3), e1004160. doi: 10.1371/journal.pgen.1004160

de La Serre, C. B., Ellis, C. L., Lee, J., Hartman, A. L., Rutledge, J. C., and Raybould, H. E. (2010). Propensity to high-fat diet-induced obesity in rats is associated with changes in the gut microbiota and gut inflammation. *Am. J. Physiol. Gastrointest. Liver Physiol.* 299 (2), G440–448. doi: 10.1152/ajpgi.00098.2010

Du, P., Zhang, X., Huang, C. C., Jafari, N., Kibbe, W. A., Hou, L., et al. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinf.* 11, 587. doi: 10.1186/1471-2105-11-587

Han, J.-L., and Lin, H.-L. (2014). Intestinal microbiota and type 2 diabetes: from mechanism insights to therapeutic perspective. *World J. Gastroenterol.* 20 (47), 17737–17745. doi: 10.3748/wjg.v20.i47.17737

Hou, X., Zhang, Y., Li, W., Hu, A. J., Luo, C., Zhou, W., et al. (2018). CDK6 inhibits white to beige fat transition by suppressing RUNX1. *Nat. Commun.* 9 (1), 1023. doi: 10.1038/s41467-018-03451-1

Hullar, M. A. J., and Fu, B. C. (2014). Diet, the gut microbiome, and epigenetics. *Cancer J.* 20 (3), 170–175. doi: 10.1097/PPO.0000000000000053

Kelly, D., Kotliar, M., Woo, V., Jagannathan, S., Whitt, J., Moncivaiz, J., et al. (2018). Microbiota-sensitive epigenetic signature predicts inflammation in Crohn's disease. *JCI Insight* 3 (18). doi: 10.1172/jci.insight.122104

Krautkramer, K. A., Kreznar, J. H., Romano, K. A., Vivas, E. I., Barrett-Wilt, G. A., Rabaglia, M. E., et al. (2016). Diet–microbiota interactions mediate global epigenetic programming in multiple host tissues. *Mol. Cell* 64 (5), 982–992. doi: 10.1016/j.molcel.2016.10.025

Kumar, H., Lund, R., Laiho, A., Lundelin, K., Ley, R. E., Isolauri, E., et al. (2014). Gut microbiota as an epigenetic regulator: pilot study based on whole-genome methylation analysis. *MBio* 5 (6). doi: 10.1128/mBio.02113-14

Leung, C., Rivera, L., Furness, J. B., and Angus, P. W. (2016). The role of the gut microbiota in NAFLD. *Nat. Rev. Gastroenterol. Hepatol.* 13 (7), 412–425. doi: 10.1038/nrgastro.2016.85

Ley, R. E., Turnbaugh, P. J., Klein, S., and Gordon, J. I. (2006). Microbial ecology: human gut microbes associated with obesity. *Nature* 444 (7122), 1022–1023. doi: 10.1038/4441022a

Ley, R. E., Backhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D., and Gordon, J. I. (2005). Obesity alters gut microbial ecology. *Proc. Natl. Acad Sc.i U. S. A.* 102 (31), 11070–11075. doi: 10.1073/pnas.0504978102

Louveau, I., and Gondret, F. (2004). Regulation of development and metabolism of adipose tissue by growth hormone and the insulin-like growth factor system. *Domest. Anim. Endocrinol.* 27 (3), 241–255. doi: 10.1016/j.domaniend.2004.06.004

Makki, K., Froguel, P., and Wolowczuk, I. (2013). Adipose tissue in obesity-related inflammation and insulin resistance: cells, cytokines, and chemokines. *ISRN Inflamm.* 2013, 139239. doi: 10.1155/2013/139239

Moreno-Indias, I., Cardona, F., Tinahones, F. J., and Queipo-Ortuno, M. I. (2014). Impact of the gut microbiota on the development of obesity and type 2 diabetes mellitus. *Front. Microbiol.* 5, 190. doi: 10.3389/fmicb.2014.00190

Ohta, H., and Itoh, N. (2014). Roles of FGFs as adipokines in adipose tissue development, remodeling, and metabolism. *Front. Endocrinol. (Lausanne)* 5, 18. doi: 10.3389/fendo.2014.00018

Paul, B., Barnes, S., Demark-Wahnefried, W., Morrow, C., Salvador, C., Skibola, C., et al. (2015). Influences of diet and the gut microbiome on epigenetic modulation in cancer and other diseases. *Clin. Epigenet.* 7, 112. doi: 10.1186/s13148-015-0144-7

Qin, Y., and Wade, P. A. (2018). Crosstalk between the microbiome and epigenome: messages from bugs. *J. Biochem.* 163 (2), 105–112. doi: 10.1093/jb/mvx080

Qin, Y., Roberts, J. D., Grimm, S. A., Lih, F. B., Deterding, L. J., Li, R., et al. (2018). An obesity-associated gut microbiome reprograms the intestinal epigenome and leads to altered colonic gene expression. *Genome Biol.* 19 (1), 7. doi: 10.1186/s13059-018-1389-1

Ramos-Molina, B., Queipo-Ortuño, M. I., Lambertos, A., Tinahones, F. J., and Peñafiel, R. (2019). Dietary and gut microbiota polyamines in obesity- and age-related diseases. *Front. Nutr.* 6. doi: 10.3389/fnut.2019.00024

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., et al. (2018). Environment dominates over host genetics in shaping human gut microbiota. *Nature* 555 (7695), 210–215. doi: 10.1038/nature25973

Sharma, N. K., Das, S. K., Mondal, A. K., Hackney, O. G., Chu, W. S., Kern, P. A., et al. (2008). Endoplasmic reticulum stress markers are associated with obesity in nondiabetic subjects. *J. Clin. Endocrinol. Metab.* 93 (11), 4532–4541. doi: 10.1210/jc.2008-1001

Soda, K. (2018). Polyamine metabolism and gene methylation in conjunction with one-carbon metabolism. *Int. J. Mol. Sci.* 19 (10). doi: 10.3390/ijms19103106

Sook Lee, E., Ji Song, E., and Do Nam, Y. (2017). Dysbiosis of gut microbiome and its impact on epigenetic regulation. *J. Clin. Epigenet.* 03 (02). doi: 10.21767/2472-1158.100048

Su, X., Wellen, K. E., and Rabinowitz, J. D. (2016). Metabolic control of methylation and acetylation. *Curr. Opin. Chem. Biol.* 30, 52–60. doi: 10.1016/j.cbpa.2015.10.030

Teschendorff, A. E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., et al. (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* 29 (2), 189–196. doi: 10.1093/bioinformatics/bts680

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., and Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444 (7122), 1027–1031. doi: 10.1038/nature05414

Utzschneider, K. M., Kratz, M., Damman, C. J., and Hullar, M. (2016). Mechanisms linking the gut microbiome and glucose metabolism. *J. Clin. Endocrinol. Metab.* 101 (4), 1445–1454. doi: 10.1210/jc.2015-4251

van Dijk, S. J., Tellam, R. L., Morrison, J. L., Muhlhausler, B. S., and Molloy, P. L. (2015). Recent developments on the role of epigenetics in obesity and metabolic disease. *Clin. Epigenet.* 7 (1). doi: 10.1186/s13148-015-0101-5

Verdam, F. J., Fuentes, S., de Jonge, C., Zoetendal, E. G., Erbil, R., Greve, J. W., et al. (2013). Human intestinal microbiota composition is associated with local and systemic inflammation in obesity. *Obesity (Silver Spring)* 21 (12), E607–615. doi: 10.1002/oby.20466

Vital, M., Howe, A. C., Tiedje, J. M., and Moran, M. A. (2014). Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. *MBio* 5 (2). doi: 10.1128/mBio.00889-14

Watson, M. M., and Søreide, K. (2017). "Chapter 32 the gut microbiota influence on human epigenetics, health, and disease" in *Handbook of Epigenetics (The New Molecular and Medical Genetics, 2nd Edn)*. 495–510. doi: 10.1016/B978-0-12-805388-1.00032-8

Yang, T., Owen, J. L., Lightfoot, Y. L., Kladde, M. P., and Mohamadzadeh, M. (2013). Microbiota impact on the epigenetic regulation of colorectal cancer. *Trends Mol. Med.* 19 (12), 714–725. doi: 10.1016/j.molmed.2013.08.005

Yuille, S., Reichardt, N., Panda, S., Dunbar, H., and Mulder, I. E. (2018). Human gut bacteria as potent class I histone deacetylase inhibitors in vitro through production of butyric acid and valeric acid. *PLoS One* 13 (7), e0201073. doi: 10.1371/journal.pone.0201073

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Ramos-Molina, Sánchez-Alcoholado, Cabrera-Mulero, Lopez-Dominguez, Carmona-Saez, Garcia-Fuentes, Moreno-Indias and Tinahones. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 16 August 2019
doi: 10.3389/fgene.2019.00724

DNA Methylation Biomarkers Predict Objective Responses to PD-1/PD-L1 Inhibition Blockade

Gang Xue[1,2†], Ze-Jia Cui[2,3†], Xiong-Hui Zhou[2], Yue-Xing Zhu[1], Ying Chen[1], Feng-Ji Liang[3], Da-Nian Tang[4], Bing-Yang Huang[5], Hong-Yu Zhang[2], Zhi-Huang Hu[6], Xi-Yu Yuan[7]* and Jianghui Xiong[1,3]**

[1] SPACEnter Space Science and Technology Institute, Shenzhen, China, [2] College of Informatics, Huazhong Agricultural University, Wuhan, China, [3] State Key Laboratory of Space Medicine Fundamentals and Application, China Astronaut Research and Training Center, Beijing, China, [4] Gastro-Intestinal Surgery Department, Beijing Hospital, Beijing, China, [5] Department of Cardiothoracic Surgery, Strategic Support Force Medical Center of PLA. No. 9, Beijing, China, [6] Department of Medical Oncology, Fudan University Shanghai Cancer Center, Shanghai, China, [7] Department of General Surgery, Dongguan People's Hospital affiliated to Southern Medical University, Dongguan, China

OPEN ACCESS

Edited by:
Yun Liu,
Fudan University, China

Reviewed by:
Erietta Stelekati,
University of Pennsylvania, United States
Chang Sun,
Shaanxi Normal University, China

***Correspondence:**
Zhihuang Hu
zhihuanghu@hotmail.com
Xiyu Yuan
15816818820@qq.com
Jianghui Xiong
xiongjh77@163.com

†These authors have contributed equally to this work.

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *05 March 2019*
Accepted: *10 July 2019*
Published: *16 August 2019*

Citation:
Xue G, Cui Z-J, Zhou X-H, Zhu Y-X, Chen Y, Liang F-J, Tang D-N, Huang B-Y, Zhang H-Y, Hu Z-H, Yuan X-Y and Xiong J (2019) DNA Methylation Biomarkers Predict Objective Responses to PD-1/PD-L1 Inhibition Blockade. Front. Genet. 10:724. doi: 10.3389/fgene.2019.00724

Immune checkpoint inhibitor (ICI) treatment could bring long-lasting clinical benefits to patients with metastatic cancer. However, only a small proportion of patients respond to PD-1/PD-L1 blockade, so predictive biomarkers are needed. Here, based on DNA methylation profiles and the objective response rates (ORRs) of PD-1/PD-L1 inhibition therapy, we identified 269 CpG sites and developed an initial CpG-based model by Lasso to predict ORRs. Notably, as measured by the area under the receiver operating characteristic curve (AUC), our model can produce better performance (AUC = 0.92) than both a model based on tumor mutational burden (TMB) (AUC = 0.77) and a previously reported TMB model (AUC = 0.71). In addition, most CpGs also have additional synergies with TMB, which can achieve a higher prediction accuracy when joined with TMB. Furthermore, we identified CpGs that are associated with TMB at the individual level. DNA methylation modules defined by protein networks, Kyoto Encylopedia of Genes and Genomes (KEGG) pathways, and ligand-receptor gene pairs are also associated with ORRs. This method suggested novel immuno-oncology targets that might be beneficial when combined with PD-1/PD-L1 blockade. Thus, DNA methylation studies might hold great potential for individualized PD1/PD-L1 blockade or combinatory therapy.

Keywords: PD-1/PD-L1 inhibition therapy, objective response rate, DNA methylation, biomarkers, Lasso model

INTRODUCTION

Cancer immunotherapies have increasingly become a promising treatment strategy in the past few years. These therapies are designed to help the immune system identify and destroy cancer cells by targeting immune checkpoints such as programmed cell death protein 1 (PD-1) and its ligand (PD-L1) (Mahoney et al., 2015). PD-1 is expressed on the surface of activated T lymphocyte cells, and its major role is to inhibit T cell activation by binding to the PD-L1 ligand on cancer cells, leading to immune suppression (Medina and Adams, 2016). A number of immune checkpoint-modulating drugs that target PD-1/PD-L1 have shown remarkable clinical benefits in multiple cancers. For instance, nivolumab and pembrolizumab, the first two monoclonal antibodies approved by the US Food and

Drug Administration (FDA) (Prasad and Kaestner, 2017), have already been registered for treatment of malignant melanoma (MM), advanced non-small-cell lung cancer (NSCLC), urothelial cancer, renal cell cancer, and head and neck squamous cell cancer (HNSCC) (Motzer et al., 2015; Robert et al., 2015; Reck et al., 2016; Bellmunt et al., 2017; Forster and Devlin, 2018). These drugs act by influencing the interaction between PD-1 and PD-L1, whose unobstructed interaction will downregulate T cells, causing cancer cells to evade immune surveillance (Prasad and Kaestner, 2017).

Compared with conventional therapy, inhibitors of PD-1 or PD-L1 can induce long-lasting responses in patients with metastatic cancer, but only one fourth to one third of patients have objective responses to immune checkpoint blockade therapy (Schachter et al., 2016). Additionally, these treatments are costly and might have some associated toxicities (Schmidt, 2017). Therefore, it is important to accurately identify the applicable population. Currently, emerging primary biomarkers used in response to immunotherapy are PD-1/PD-L1 protein expression, microsatellite instability (MSI), and tumor mutational burden (TMB) (Topalian et al., 2016; Chalmers et al., 2017; Chang et al., 2017). However, obvious limitations exist among these biomarkers due to low efficacy, antibody discrepancy, sampling bias, and strict requirements for cancer tissue. Achieving accurate forecasts and guiding clinical treatment remain critical challenges (Johnson et al., 2016).

The abnormal epigenomic landscape is one of the hallmarks of tumor initiation and progression (Esteller, 2008; Tsai and Baylin, 2011). In particular, aberrant patterns of DNA methylation can alter chromatin structure and gene transcription without altering the DNA sequence (Bird, 2007); these patterns have been extensively studied. In mammals, DNA methylation is almost exclusively found in CpG dinucleotides (CpGs). Recent work has revealed that DNA methylation affects tumorigenesis by regulating the tumor microenvironment (Xiao et al., 2016; Zhang et al., 2017). There are a multitude of DNA methylation biomarkers for the prognosis, diagnosis, and response to treatment in several types of cancer (Rodriguez-Paredes and Esteller, 2011). Based on the above evidence, we hypothesize that DNA methylation signatures could act as reliable immune checkpoint blockade biomarkers.

Ideally, abundant tumor molecule profiles along with patient objective response rates of immune inhibitors can be used to train reliable multiple biomarkers. However, in reality, only a small number of samples have both types of data. Alternatively, the tumor profiles are probably not the same ones whose response rates are assessed. For example, the research of Yarchoan et al. (2017) assessed the relationship between the tumor mutational burden and the objective response rate of PD-1/PD-L1 inhibition by pooling the response data from the published studies and the tumor mutational burden for each tumor type, which was provided by Foundation Medicine (Chalmers et al., 2017), and their analysis was that the sequenced tumor specimens may be different from those whose clinical responses were evaluated (Yarchoan et al., 2017). Similar to their method, we collected a large amount of DNA methylation profiles from 18 cancer types in The Cancer Genome Atlas (TCGA, https://tcga-data.ncbi.nih.gov/tcga/) and corresponding objective response rates from the largest published studies. We calculated the correlations between CpG probes and response rates in the 18 cancer types and then used CpG probes that were significantly correlated with response rates to construct a model for predicting the objective response rate by the Lasso regression method. We proposed that, compared with the model of predicting the response rate with TMB, the method with the CpG signatures was more accurate. Next, we utilized multimethod detection to verify the reliability of the DNA methylation signatures as surrogate biomarkers to predict the objective response rate of PD-1/PD-L1 inhibition.

MATERIALS AND METHODS

Data Availability

The objective response rate (ORR) data for PD-1/PD-L1 inhibitors were obtained from the study of Yarchoan et al. (2017), and the data sets of the samples of each cancer were retrieved from TCGA (https://tcga-data.ncbi.nih.gov/tcga/). Each data set contained DNA methylation profiles obtained by Illumina 450K methylation assays. According to the research of Yarchoan et al. (2017) and the cancer types of TCGA, 18 cancer types have validated both ORRs and 450K methylation array data. In this study, these 18 cancer data sets were analyzed (**Table 1**).

TABLE 1 | Objective response rates (ORRs) collection of 18 cancer types.

Tumor types	Abbreviation	ORR (literature)
Adrenocortical carcinoma	ACC	0.06 (Le Tourneau et al., 2017)
Bladder urothelial carcinoma	BLCA	0.182 (Rosenberg et al., 2016; Apolo et al., 2017; Bellmunt et al., 2017; Powles et al., 2017; Sharma et al., 2017)
Breast invasive carcinoma	BRCA	0.052 (Dirix et al., 2018)
Cervical squamous cell carcinoma and endocervical adenocarcinoma	CESC	0.208 (Hollebecque et al., 2017)
Esophageal carcinoma	ESCA	0.112 (Chung et al., 2016; Fuchs et al., 2017)
Glioblastoma multiforme	GBM	0.08 (Reardon et al., 2017a; Reardon et al., 2017b)
Head and neck squamous cell carcinoma	HNSC	0.16 (Ferris et al., 2016; Bauml et al., 2017)
Kidney renal clear cell carcinoma	KIRC	0.25 (Motzer et al., 2015)
Liver hepatocellular carcinoma	LIHC	0.2 (El-Khoueiry et al., 2017; Wainberg et al., 2017)
Lung adenocarcinoma	LUAD	0.19 (Borghaei et al., 2015)
Lung squamous cell carcinoma	LUSC	0.2 (Brahmer et al., 2015)
Mesothelioma	MESO	0.167 (Scherpereel et al., 2017)
Ovarian serous cystadenocarcinoma	OV	0.097 (Brahmer et al., 2012; Hamanishi et al., 2015; Disis et al., 2016)
Pancreatic adenocarcinoma	PAAD	0 (Brahmer et al., 2012)
Sarcoma	SARC	0.11 (Dangelo et al., 2017; Tawbi et al., 2017)
Skin cutaneous melanoma	SKCM	0.387 (Larkin et al., 2015; Robert et al., 2015)
Uterine corpus endometrial carcinoma	UCEC	0.13 (Fleming et al., 2017)
Uveal melanoma	UVM	0.036 (Algazi et al., 2016)

For independent verification to assess the robustness of our model, we collected the 450K methylation array data of NSCLC from the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE39279, which includes 444 patient samples.

To calculate the TMB of the 18 cancer types, 18 Mutation Annotation Format (MAF) files processed by MuSE (Fan et al., 2016) were downloaded from the GDC data portal (https://portal.gdc.cancer.gov/repository). The MAF files contained the somatic mutations of TCGA cohorts.

Three hundred twenty-four annotated KEGG pathways comprising 7,448 genes (Entrez Gene IDs) were retrieved from Kyoto Encylopedia of Genes and Genomes (KEGG) pathway database (https://www.genome.jp/kegg-bin/get_htext?hsa00001+3101). These data were used for pathway analysis.

A human protein-protein interaction (PPI) network was derived from the STRING database (STRING, http://www.string-db.org). The default score threshold of interactions is typically 400 (Franceschini et al., 2013). Therefore, interactions with scores lower than 400 were discarded. These PPIs were used to construct subnetworks for a given gene.

Identification of CpG Probes Associated With ORR

We used the β values reported by the 450K Illumina platform for each probe as the methylation level measurement for the targeted CpG site. The range of the β value is from 0 (no methylation) to 1 (100% methylation). A higher β value indicates a higher DNA methylation level. Each CpG value in a cancer type was represented by the mean β values in the tumor samples; then, the Spearman's rank correlation test was used to quantify the association strength between the methylation level of the CpGs and the ORRs of the 18 cancer types. Since Bonferroni adjustment for multiple comparisons of the ~480,000 CpGs is too conservative, especially with the small sample size (18 cancer types) in our research, we used a less stringent threshold of P value ≤0.001 and an absolute value [Spearman's rank correlation coefficient (Spearman's rho)] = 0.7 to obtain reliable ORR-associated CpG signatures. The annotation of each CpG, such as CpG's position in the genome and its corresponding gene, was derived from the GEO database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL13534).

Defining Methylation Levels of Functional Modules Based on Entropy

At a wide range of genomic positions, the CpG signals do not conform to a normal distribution but are distributed in a nearly bimodal distribution. Thus, too much information would be lost when simply averaging the β values.

In information theory, concept entropy is the average rate at which information is produced by a stochastic source of data. When the data source has a lower probability value (i.e., when a low-probability event occurs), the event carries more "information" ("surprisal") than when the source data have a higher probability value. The amount of information conveyed by each event defined in this way becomes a random variable whose expected value represents the information entropy. Generally, entropy refers to disorder or uncertainty. Here, we capture the methylation levels of various functional modules based on Shannon's entropy, which is described as follows:

$$\mathbf{H} = -\sum_{i=1}^{n} (\mathbf{p}_i \ln \mathbf{p}_i)$$

In this equation, pi is the β value of each CpG probe and n is the number of CpG probes within the functional modules (protein network, KEGG pathway, and ligand-receptor gene pairs). Likewise, we used the Spearman correlation test to quantify the strength of associations between each functional unit and the ORRs of 18 cancer types.

Construction of the CpG-Based Lasso Regression Model

To predict the objection response rate of PD-1/PD-L1 inhibition with reliable CpG signatures for clinical applications, additional selection and model construction are necessary. The Lasso algorithm is used to perform the variable selection procedure by estimating linear regression coefficients by L1-constrained least squares. It minimizes the sum of squared residuals, which is affected by the sum of the absolute values of the coefficients being less than the constant. Because of this constraint, Lasso regression tends to produce some coefficients that are precisely 0. Finally, a robust and interpretable model can be given. The original linear regression model can be written as follows:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon$$

The Lasso estimates for the constant term (α) and the regression coefficient (β) are as follows:

$$\left(\hat{\alpha}, \hat{\beta}\right) = \mathbf{argmin} \sum_{i=1}^{n} (\mathbf{y}_i - \boldsymbol{\alpha}_i - \sum_{j=1}^{p} \boldsymbol{\beta}_j \mathbf{x}_{ij})^2, \mathbf{s.t.} \sum_{j=1}^{p} |\boldsymbol{\beta}_j| \leq \boldsymbol{\lambda}$$

Here, y represents the ORR values of 18 cancers, x represents the β values of CpG probes that are significantly associated with ORR, and λ is a nonnegative adjustment parameter that controls the amount of shrinkage. The determination of λ can be estimated using the cross-validated (CV) method proposed by Efron and Tibshirani in 1997 (Efron and Tibshirani, 1997). In this study, the Lasso function in MATLAB was used to fit the equation, and the CV was set to 10.

Tumor Mutational Burden (TMB) Calculations

TMB is a measure of the number of somatic protein-coding base substitutions and insertion/deletion mutations occurring in a tumor specimen. To calculate the TMB, the total number of mutations counted is divided by the size of the genome examined. Here, we used 38Mb as the estimate of the exome size. The somatic mutations were counted from the MAF files of

TCGA, and the tumor mutational burden for each patient was estimated as follows:

$$TMB = \frac{n}{38}$$

In this equation, n is the total number of missense mutations of a patient.

The median TMB for each cancer type can then be estimated as follows:

$$Median\ TMB = \frac{N}{38}$$

In this equation, N is the median number of coding somatic missense mutations in a cancer type.

Next, in line with Yarchoan et al.'s work, a new linear correlation formula that evaluates the relationship between the TMB and ORR was constructed as follows:

$$ORR = 0.0768 * In(X) + 0.1313$$

Here, X is the median TMB of each cancer type.

Synergy Index Calculations

A synergy index (S) was calculated to determine the presence of the interactions of the β values of each ORR-associated CpG probe and TMB. The synergy index is equal to 1 (S = 1) in the absence of a synergistic interaction; in such a case, the joint effect of two predictive variables is equal to the sum of their independent effects (i.e., it is additive). A synergy index greater than 1 (S > 1) suggests the presence of a synergistic interaction; the observed joint effect is greater than that expected from the sum of the independent effects of the component variables (i.e., it is synergistic). Conversely, a synergy index less than 1 (S < 1) suggests an "antagonistic" effect or a negative interaction. Here, the synergy index was calculated via a logistic regression model.

RESULTS

Identifying CpGs Associated With the Objective Response Rate (ORR) of PD-1/PD-L1 Inhibition Therapy

Based on Yarchoan et al.'s extensive literature searches, we obtained 18 cancer types for which validated ORRs and the 450K methylation array data are both available. From **Table 1**, we can observe that most ORRs of cancer types are less than 0.2.

We first performed Spearman's rank correlation test to identify CpGs whose methylation level was associated with the ORRs of anti-PD-1/anti-PD-L1 therapy. We collected current global immuno-oncology targets as the gold standard to assess our result by the Kolmogorov–Smirnov (KS) test (Tang et al., 2018). The targets that were more enriched in high Spearman rank correlation coefficient (Spearman's rho) ORR-associated genes exhibited a smaller P value (derived from the KS test), which indicated that our result was reliable (P value = 0.0249). At the threshold of an absolute value (Spearman's rho) ≥0.7 and a P value ≤0.001, we identified 269 genome-wide significant CpGs corresponding to 191 genes (**Table 2** and **Supplementary Table S1**). Then, we investigated the number of CpGs enriched in these 191 genes. The more enriched, the more likely they can be considered marker genes of anti-PD-1/anti-PD-L1 therapy. We annotated the functions of the top enriched genes from the UniProt database (https://www.UniProt.org/) and the literature (**Table 3**). For example, HLA-E [human leukocyte antigen (HLA) class I histocompatibility antigen, alpha chain E] is the most enriched gene in our results, and some studies have indicated that HLA class I antigen expression can be utilized in select patients who may benefit from anti-PD-1/PD-L1-based immunotherapy (Sabbatino et al., 2016; Chowell et al., 2018). Therefore, we have reasons to infer that other enriched genes could also be considered potential markers for anti-PD-1/PD-L1 therapy.

We next examined the functional enrichment of these 191 genes using KEGG pathway analysis via cluster Profiler of R (Yu et al., 2012). Notably, we found that most of these genes were related to immunological KEGG pathways, such as antigen processing and presentation, natural killer (NK) cell-mediated cytotoxicity, and autoimmune thyroid disease (**Supplementary Table S2**). A recent study showed that the capacity of antigen presentation influences responses to checkpoint immunotherapy (Kvistborg and Yewdell, 2018), and tumor immunity is mediated mainly by NK cells (Ferrari de Andrade et al., 2018). Furthermore, we detected the signature genes that belong to multiple relevant immunological pathways. From **Figure 1**, we can clearly observe that HLA class I antigens are related to all these pathways, which highlights their importance in immunotherapy.

Construction of the CpG-Based ORR Prediction Model by Lasso

To predict the objection response rate of PD-1/PD-L1 inhibition with reliable signatures for clinical applications, we used 269 CpGs that were obtained in the above section as initial variables to construct a model to predict ORR values by the Lasso algorithm.

First, we considered whether our CpG-based Lasso regression model method was generalized and practicable for predicting the ORRs of 18 cancer types. Therefore, we adopted a "leave-one-out cross validation" method to assess the feasibility of our model. Leave-one-out cross validation has been shown to give an almost unbiased estimator of the generalization properties of statistical models. Briefly, 17 cancer type-related data sets were used as training data sets for constructing the model, and the remaining data set was used as an independent data set. Then, we repeated this process 18 times to obtain the predicted ORRs of 18 cancer types. The Spearman's rank correlation coefficient between the predicted and real ORRs was 0.75 (P value = 0.00029). This result indicated that our CpG-based Lasso regression model can be used to predict the ORRs of the 18 cancer types.

After the Lasso method was confirmed as being generalized and practicable, we used 269 CpGs and the ORR values of 18 cancer types to construct a prediction model by the Lasso algorithm. We chose the regression result when the mean square error (MSE) was minimum (MSE = 0.0042); there were eight CpG probe variables left: cg03749154, cg16051114 (DHCR24), cg04144714

TABLE 2 | List of the top 10 ORR-associated CpGs.

CpG	Gene symbol	Chromosome	Genomic coordinate	Spearman's rho	P value
cg02358190	MAST4	5	66187002	–0.92514	3.91E–08
cg04033580	C22orf45; UPB1	22	24891666	–0.8539	6.53E–06
cg13459303	TMEM176B; TMEM176A	7	1.5E+08	–0.82912	2.11E–05
cg24644201	CREB3L1	11	46299066	0.822922	2.74E–05
cg25626312	CREB3L1	11	46299204	0.81776	3.39E–05
cg03885527	PLIN2	9	19125654	–0.81363	4.00E–05
cg05690644	GDF6	8	97158015	–0.81053	4.52E–05
cg23393637		14	95513095	–0.81053	4.52E–05
cg26981651	RNF5; RNF5P1	6	32147670	–0.81053	4.52E–05

These are the top 10 ORR-associated CpGs, and the 269 ORR-associated CpGs corresponding to 191 genes are provided in ***Supplementary Table S1****.*

TABLE 3 | Top enriched genes and function.

Gene symbol	CpGs count	Function
HLA-E	6	HLA-E has a very specialized role in cell recognition by natural killer cells (NK cells)
PLEC	4	Interlinks intermediate filaments with microtubules and microfilaments and anchors intermediate filaments to desmosomes or hemidesmosomes
HIVEP3	4	Plays a role of transcription factor; binds to recognition signal sequences for somatic recombination of immunoglobulin and T-cell receptor gene segments
FOXD2- AS1	4	lncRNA FOXD2-AS1 promotes NSCLC progression through Wnt/β-catenin signaling (Rong et al., 2017)
FOXD2	4	Probable transcription factor involved in embryogenesis and somatogenesis
CREB3L1	4	Transcription factor involved in unfolded protein response (UPR)

NSCLC, non-small-cell lung cancer.

(LYST, MIR1537), cg20395773 (ZBTB38), cg17484237 (HAVCR2), cg15006881 (GDF6), cg24644201 (CREB3L1), and cg13038847. The CpG-based prediction model is as follows:

$$y_{ORR} = 0.793 - 0.526 \times x_{cg03749154} - 0.0269 \times x_{cg16051114} - 0.711 \times x_{cg04144714} + 0.263 \times x_{cg20395773} - 0.00086 \times x_{cg17484237} - 0.012 \times x_{cg15006881} + 1.058 \times x_{cg24644201} - 0.0603 \times x_{cg13038847}$$

To assess the performance of the CpG-based prediction model for the 18 cancer types, we calculated the difference between predicted ORR values and true ones. As shown in **Figure 2**, except for PAAD, SKCM, and KIRC, the difference for other cancer types was very small. Moreover, to assess the robustness of our prediction model, we evaluated its performance in an independent sample of NSCLC from the GEO database. The ORR value of this independent data set was predicted to be 0.245, which was close to the real value of 0.2. This result further demonstrated that our model was accurate and robust.

The CpG-Based Model Performs Better than the TMB-Based Model in ORR Prediction

In a study by Yarchoan et al., researchers evaluated the relationship between the TMB and the ORR. A linear correlation formula was constructed that can be used to make hypotheses with respect to the ORR rate in tumor types for which anti-PD-1/PD-L1 therapy has not been explored. Here, we compared the performance of our CpG-based model and our TMB-based model with respect to 18 cancer types.

First, we adopted the root-mean-square error (RMSE), the mean absolute error (MAE), and Spearman correlations to compare the performance of the above two prediction models. As shown in **Table 4**, compared with the TMB model, the CpG-based model predicted ORR more accurately. Moreover, ROC curves were plotted to assess the sensitivities and specificities of these two models. As shown in **Figure 3A,** for the 18 cancer types, our model performs better than the TMB model in both sensitivity and specificity when 0.2 is used as a cutoff. The average area under the ROC curve (AUC) of the CpG-based model was 0.92, which was greater than the AUC of the TMB model, which was 0.71. For each cancer type, the performance evaluation criteria for the two models are compared to the actual ORR value of the cancer. The smaller the difference, the better the model effect. Except for CESC, HNSC, LUAD, and UVM, CpG-based model performs better than the TMB-based model in 14 cancer types. For the other four types of cancer, although CpG-based model is less powerful, our prediction is very close to the actual ORRs (**Supplementary Table S3).**

To maintain data consistency between methylation and TMB, we recalculated the TMB of 18 cancer types from TCGA and constructed another linear correlation formula according to the results of Yarchoan et al.'s study. Then, we compared the performance of these two models as above (**Table 4**). As shown in **Figure 3B**, the AUC of our model was 0.92, which was greater than the AUC of the TMB model, which was 0.77. For each cancer type, except for ACC, BRCA, and UVM, CpG-based model performs better than the TMB (TCGA)-based model in 15 cancer types (**Supplementary Table S3**). This result further demonstrated that our CpG-based model was more accurate than the TMB-based model in ORR prediction.

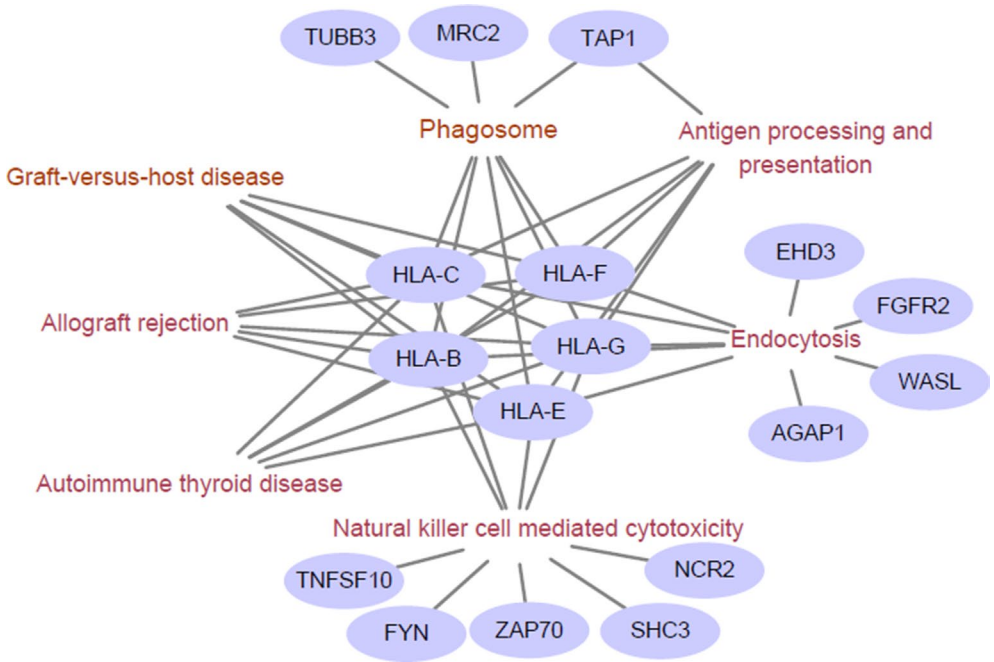


FIGURE 1 | Enriched immunological pathway and genes. The words in wine represent the Kyoto Encylopedia of Genes and Genomes (KEGG) pathway, and the purple circles represent objective response rate (ORR)-associated genes enriched in the KEGG pathway. The HLA class I antigens (HLA-B, HLA-C, HLA-E, HLA-G, and HLA-F) are related to all these pathways.

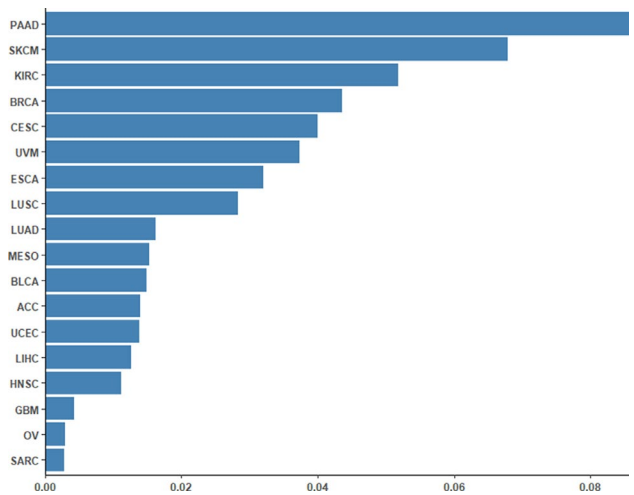


FIGURE 2 | Differences between the predicted ORRs and true ORRs of 18 cancer types. Except for PAAD, SKCM, and KIRC, most of the ORRs of the cancer types could be predicted fairly robustly.

Combining CpGs and TMB in ORR Prediction

TMB and DNA methylation describe different aspects of immunotherapy work against cancer. TMB reflects the mutation signatures in cancer, while DNA methylation affects the tumor microenvironment (TME), which plays an important role in supporting cancer progression and tumor immunity (Zhang et al., 2017; Sanmamed and Chen, 2018). Therefore, after confirming that the methylation level of a few CpGs performs better at ORR prediction than TMB, we tried to combine these two types of information by computing the synergy index (*S*) between each ORR-related CpG and TMB.

A synergistic index greater than 1 ($S > 1$) suggests the presence of a synergistic interaction between TMB and ORR-associated CpGs, so combining TMB information could enhance the predictive ability of these CpGs. Furthermore, we investigated the top 10 CpGs that, in conjunction with TMB, have a synergistic effect (**Table 5**). Notably, TNFSF10 and HIVEP3, which were identified as being strongly correlated withPD-1/PD-L1 inhibition therapy in the previous section [rho (TNFSF10) = −0.75; rho (HIVEP3) = 0.75], also displayed strong synergy with TMB. This result indicated that these CpGs could also be applied jointly with TMB to achieve a higher prediction performance.

TABLE 4 | Comparison of model performance.

Assessment index	CpG-based model	TMB-based model	TMB (TCGA)-based model
MAE	0.03	0.05	0.05
RMSE	0.04	0.07	0.06
Spearman correlation	0.93	0.58	0.69

TMB, tumor mutational burden; TCGA, The Cancer Genome Atlas; mean absolute error; RMSE, root-mean-square error.

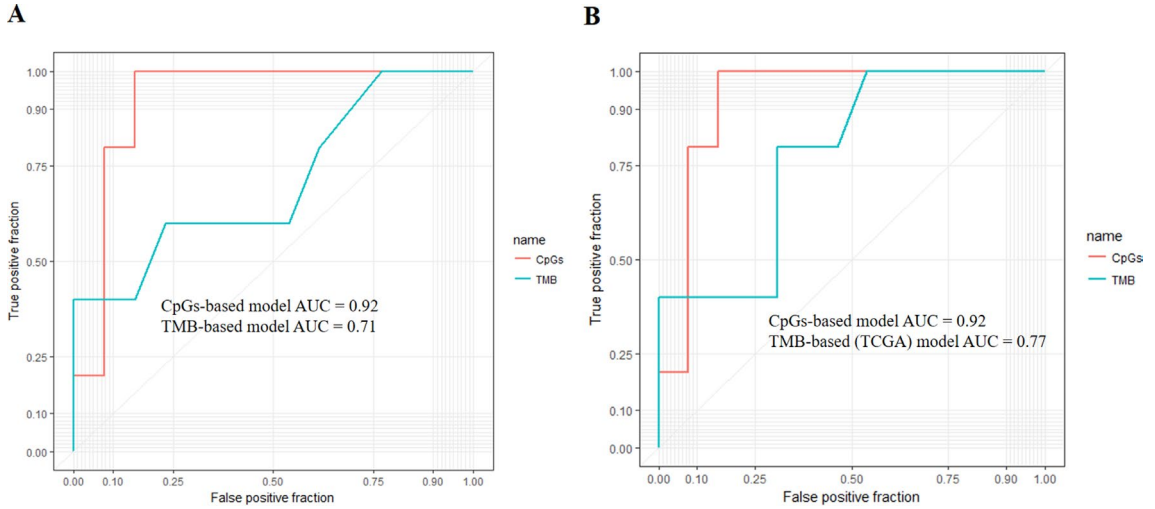


FIGURE 3 | (A) Performance comparison of the CpG-based model and tumor mutational burden (TMB)-based model. The area under the receiver operating characteristic curve (AUC) scores of the CpG-based model and TMB-based model were 0.92 and 0.71, respectively, which indicated that our model had better performance. **(B)** Performance comparison of the CpG-based model and TMB-based model using The Cancer Genome Atlas (TCGA) samples. The AUC scores of the CpG-based model and TMB (TCGA)-based model were 0.92 and 0.77, respectively, which indicated that our model had better performance.

TABLE 5 | List of the top 10 synergy sites.

CpG	Gene symbol	Synergy index
cg09248054	AGRN	37.17457358
cg22572614	TNFSF10	25.020374
cg23485436	KDM4B	24.3314227
cg25607920	HIVEP3	11.65557467
cg23902361	VAMP5	11.25511323
cg14116139		5.878589192
cg08405073	CCDC159	5.626048429
cg08405073	TMEM205	5.626048429
cg14615152	CSMD2	5.612506908
cg25577670	SVIL	5.466948003

The Methylation Level of CpGs Is Associated With TMB at the Individual Level

The above work involved mainly identifying CpG signatures of ORRs; these signatures are meaningful and could be used to construct a model to predict ORRs at the cancer type level. However, for clinical applications, we are more concerned about whether these signatures could also work for individuals. Since TMB has become a relatively mature biomarker of sensitivity to immune check points in individuals, we identified these 269 CpG signatures whose methylation levels were associated with TMB at the individual level. Most of these CpGs were significantly associated with TMB with an FDR = 0.0001 (**Supplementary Table S4**). Moreover, we investigated the top associated CpGs and found that specimens with relatively high methylation levels of these CpGs are more likely to have relatively high TMB (**Figure 4**). These CpG signatures could also become biomarkers of PD-1/PD-L1 inhibition therapy for individual patients.

Identification of DNA Methylation Modules Related to ORRs

A challenge of epigenetic studies is that DNA methylation changes can occur at a wide range of genomic positions, and their relationship between each single site and phenotype is not direct. A statistic to summarize the effects of environmental stimuli on gene regulation and the use of this feature to predict future medical events are highly desired. Here, we proposed a method to determine DNA methylation levels based on the entropy concept at different system levels, including protein networks, KEGG pathways, and ligand-receptor gene pairs, to represent coregulation units between two interactive cell types.

At the protein network level, we found 787 subnetworks that were significantly associated with ORRs at a P value <0.05 threshold (**Supplementary Table S5**). Then, we focused on the subnetwork that contained PD-1 (PDCD1) and PD-L1 (CD274) (**Figure 5A**). This subnetwork is mainly involved in two pathways: antigen processing and presentation (**Figure 5B**) and cell adhesion molecules (**Figure 5C**). β2-Microglobulin (B2M) is a component of the HLA class I complex and functions in immunosurveillance. Carolina et al. reported that mutations in B2M could impair the correct formation of the HLA-I complex, which subsequently affects the response to anti-PD-1/anti-PD-L1 therapies (Pereira et al., 2017). Here, based on entropy to quantify the level of DNA methylation in a subnetwork, we obtained a similar observation. Except for PD-1/PD-L1, we should also pay more attention to the other subnetwork genes that may inspire new immunotherapies.

At the KEGG pathway level, 37 KEGG pathways were significantly associated with ORRs at a P value <0.05 threshold. Among them, several KEGG pathways were related to immune processes (**Supplementary Table S6**), such as the B cell receptor signaling pathway, the T cell receptor signaling pathway, natural killer

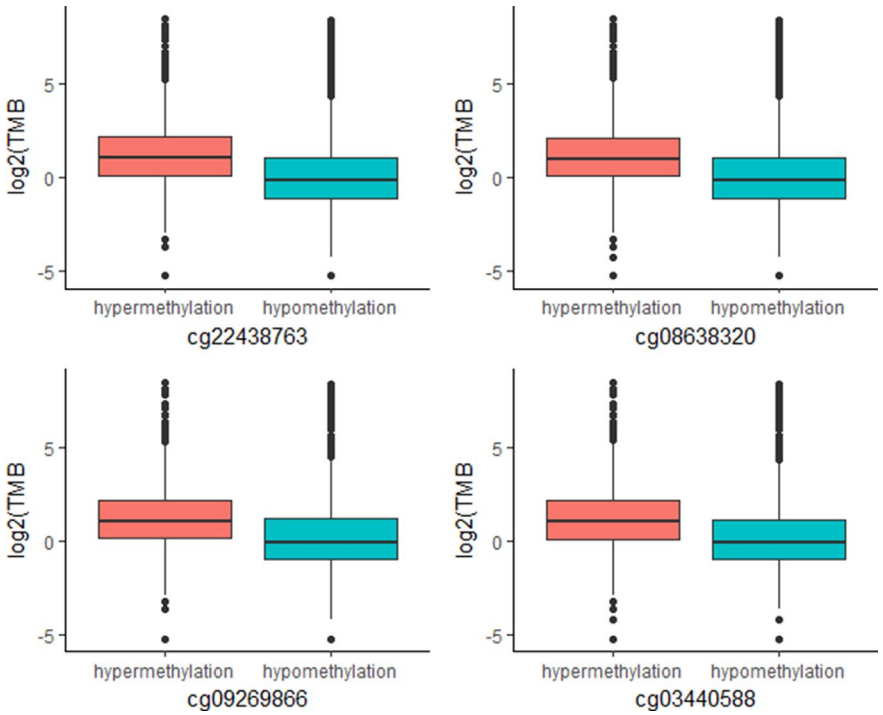


FIGURE 4 | Plot of mutation burden in specimens with hypermethylation and specimens with hypermethylation of top TMB-associated CpGs (*n (specimens)* = *5*,104).

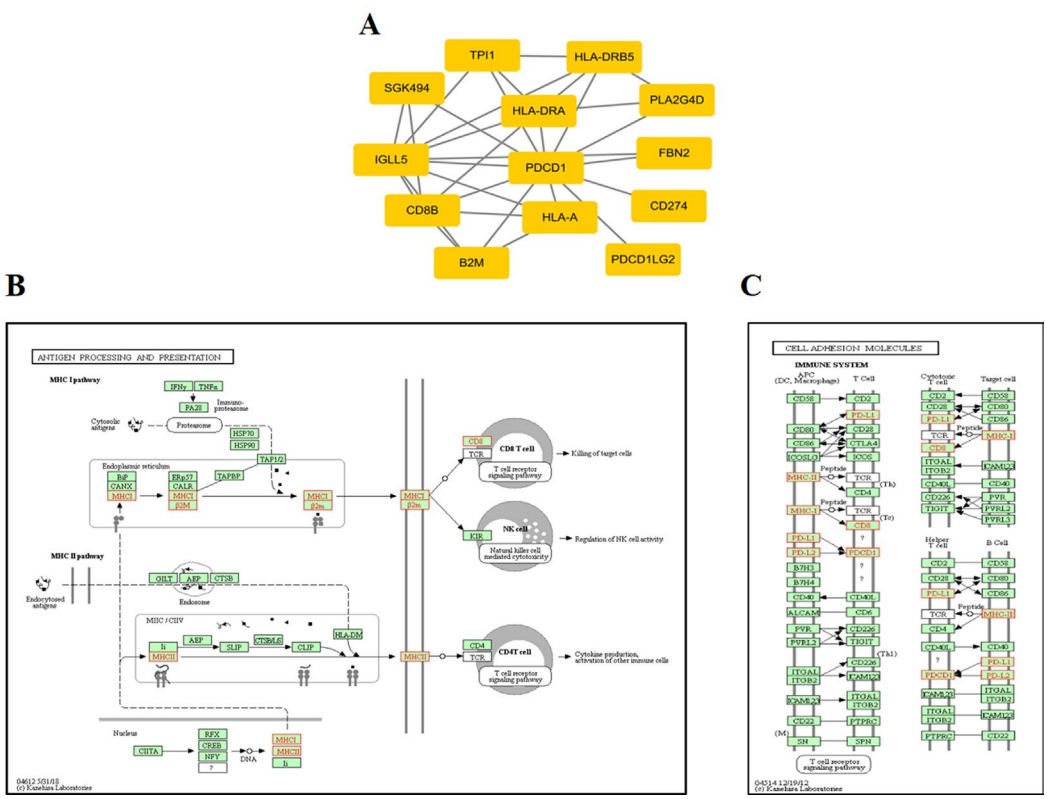


FIGURE 5 | ORR-associated subnetwork map of KEGG pathways. **(A)** PDCD1 and its interacting genes in protein-protein interaction (PPI). **(B)** The pathway of antigen processing and presentation. **(C)** The pathway of cell adhesion molecules (CAMs); the golden yellow color represents the genes in the subnetwork of PDCD1.

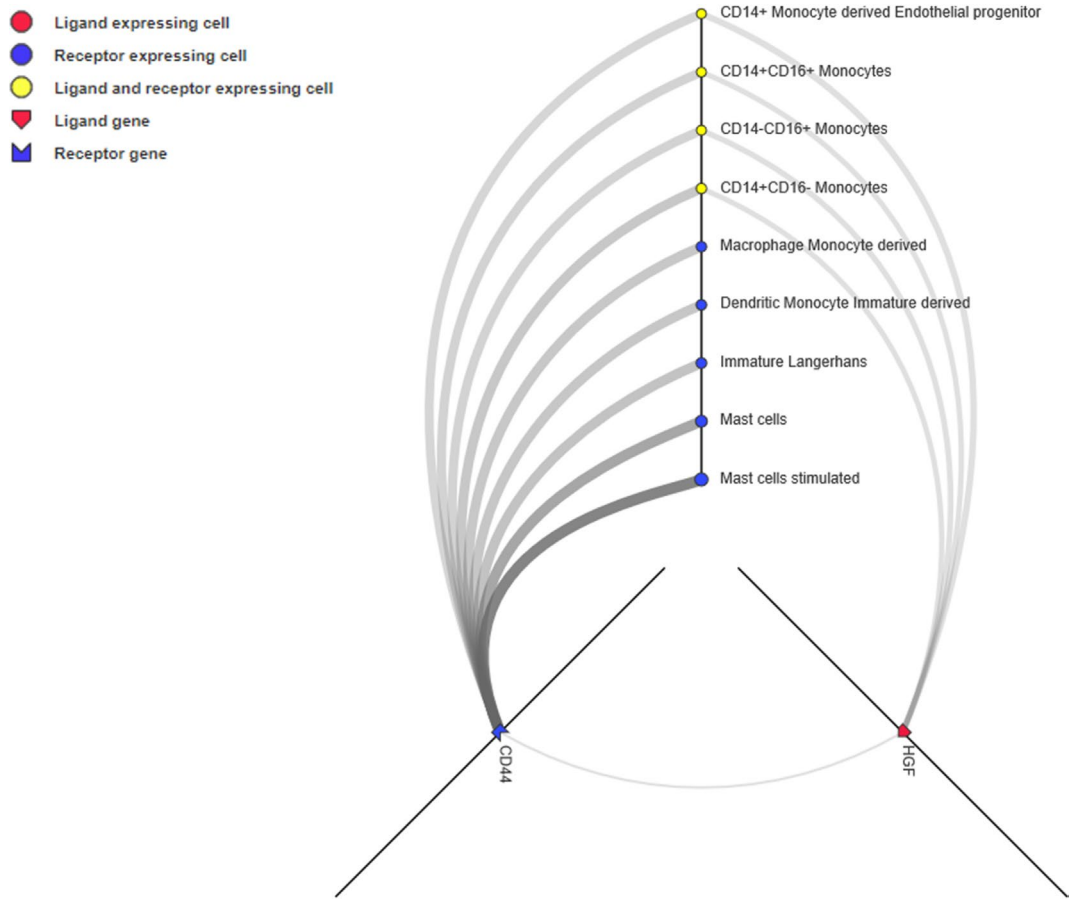


FIGURE 6 | CD44-HGF signaling network interface. In this network, both CD44 and hepatocyte growth factor (HGF) were expressed in monocytes (≥10 TPM). The interface is available at http://fantom.gsc.riken.jp/5/suppl/Ramilowski_et_al_2015/.

cell-mediated cytotoxicity, and autoimmune thyroid disease. These results were consistent with the previously enriched pathway by 269 CpGs. Moreover, although other KEGG pathways are not directly related to immunotherapy, they are also meaningful. For instance, riboflavin metabolism is strongly significantly associated with ORRs; previous research has shown that metabolites of vitamin B represent a class of antigens that are represented by MHC class I-like molecules (MR1s) for mucosal-associated invariant T (MAIT) cell immunosurveillance (Kjernielsen et al., 2012). Therefore, our results may provide new insights into PD-1/PD-L1 inhibition therapy.

Unlike the PPI network, which depicts the intracellular network, ligand-receptor mediated cell-to-cell communication across multiple cell types and tissues could inspire new immunotherapy techniques (Ramilowski et al., 2015). Ligands, receptors, and their interactions were retrieved from the CellPhoneDB (https://www.cellphonedb.org/) database. Including PD-1 and PDCD1, 103 ligand-receptor pairs were significantly associated with ORRs at a *P value ≤0.05 threshold (***Supplementary Table S7).** The ligand-receptors of CD44 and HGF was the most significantly associated with ORR. Thus, we observed the cell-to-cell networks of CD44-HGF (**Figure 6**). We noted that CD44-HGF was expressed in monocytes at notable levels [≥10 Transcripts Per Kilobase Million (TPM)]. Although there is still no evidence that CD44-HGF affects the response to anti-PD-1/anti-PD-L1 therapies, a recent study identified types of immune cells known as classical monocytes (CD14+CD16–HLA-DRhi) in the peripheral blood as potential biomarkers for responses to anti-PD-1 immune checkpoint therapy in metastatic melanoma (Goswami et al., 2018).

From the above analyses, based on the entropy concept, we identified various functional modules associated with ORRs from the protein network, KEGG pathways, and ligand-receptor gene pairs. Some of these modules have been reported by other research groups, which confirmed the reliability of the DNA methylation signatures as surrogate biomarkers to predict the objective response rate of PD-1/PD-L1 inhibition.

DISCUSSION

Compared with conventional therapies, immune check point inhibitor treatments represented by PD-1/PD-L1 have shown remarkable clinical benefits (Yarchoan et al., 2017), but predictive biomarkers are needed. In this study, using DNA methylation profiles and the objective response rates (ORR) of 18 cancer types, we successfully identified 269 CpG signatures related to

ORRs and developed an initial CpG-based objective response rate (ORR) prediction model by Lasso. We showed that these 269 CpG signatures (corresponding to 191 genes) can be considered potential immuno-oncology targets. Furthermore, the CpG-based ORR prediction model performed better than the TMB-based model. In the independent test of NSCLC data, our model also made accurate predictions. Moreover, we also identified CpGs that are associated with TMB at the individual level.

To further investigate the relationship between methylation and phenotype (i.e., ORR), we introduced a new method based on the entropy concept and identified various functional modules associated with ORR, from protein networks to KEGG pathways and ligand-receptor gene pairs, which may provide new insights into PD-1/PD-L1 inhibition therapy.

One limitation of our analysis is that the sequenced tumor samples were probably not the same for those whose ORRs were assessed, which would introduce deviation in our result. Matched clinical and genetic (i.e., DNA methylation profiles) data would help us develop a more robust and reliable model. The independent verification by bisulfite pyrosequencing of several most significant CpGs/genes can better demonstrate the accuracy of our conclusion. However, in the present study, we mainly focused on investigating the correlation between CpG methylation in genome and response to PD-1 or PD-L1 therapy and predicting ORR of cancer based on methylation level of several CpG sites in the patients. Based on statistical analysis and the evidence from the literature, it should be sufficient to draw a conclusion that such DNA methylation studies hold great potential for individualized PD1/PD-L1 blockade or combinatory therapy. Furthermore, CpG sites could also be applied jointly with other types of biomarkers, for instance, TMB, to achieve increased prediction performance to help oncologists select patients who are more likely to benefit from PD-1/PD-L1 inhibition therapy.

DATA AVAILABILITY

All datasets analyzed for this study are included in the manuscript and the supplementary files.

AUTHOR CONTRIBUTIONS

JX, X-YY, Z-HH and H-YZ conceived and supervised the study. GX, Z-JC, X-HZ, and F-JL analyzed the data. GX and Z-JC wrote the manuscript. JX, Z-HH, YC, Y-XZ, D-NT, and B-YH made manuscript revisions.

FUNDING

This research was funded by the grant from the National Instrumentation Program (No. 2013YQ190467), Shenzhen Science & Technology Program (JCYJ20151029154245758, CKFW2016082915204709), the Chinese Scientific and Technological Major Special Project (2012ZX09301003-002-003), Natural Science Foundation of China (91129708), the grant from State Key Lab of Space Medicine Fundamentals and Application (SMFA09A07, SMFA10A03, and SMFA13A04), Wu Jieping Medical Foundation (320.6750.18150) and the Fundamental Research Funds for the Central Universities (No. 2662017PY115).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00724/full#supplementary-material

REFERENCES

Algazi, A. P., Tsai, K. K., Shoushtari, A. N., Munhoz, R. R., Eroglu, Z., Piulats, J. M., et al. (2016). Clinical outcomes in metastatic uveal melanoma treated with PD-1 and PD-L1 antibodies. *Cancer* 122 (21), 3344–3353. doi: 10.1002/cncr.30258

Apolo, A. B., Infante, J. R., Balmanoukian, A., Patel, M. R., Wang, D., Kelly, K., et al. (2017). Avelumab, an anti-programmed death-ligand 1 antibody, in patients with refractory metastatic urothelial carcinoma: results from a multicenter, phase Ib study. *J. Clin. Oncol.* 35 (19), 2117–2124. doi: 10.1200/JCO.2016.71.6795

Bauml, J., Seiwert, T. Y., Pfister, D. G., Worden, F., Liu, S. V., Gilbert, J., et al. (2017). Pembrolizumab for platinum- and cetuximab-refractory head and neck cancer: results from a single-arm, phase II study. *J. Clin. Oncol.* 35 (14), 1542–1549. doi: 10.1200/JCO.2016.70.1524

Bellmunt, J., de Wit, R., Vaughn, D. J., Fradet, Y., Lee, J. L., Fong, L., et al. (2017). Pembrolizumab as second-line therapy for advanced urothelial carcinoma. *N. Engl. J. Med.* 376 (11), 1015–1026. doi: 10.1056/NEJMoa1613683

Bird, A. (2007). Perceptions of epigenetics. *Nature* 447 (7143), 396–398. doi: 10.1038/nature05913

Borghaei, H., Paz-Ares, L., Horn, L., Spigel, D. R., Steins, M., Ready, N. E., et al. (2015). Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N. Engl. J. Med.* 373 (17), 1627–1639. doi: 10.1056/NEJMoa1507643

Brahmer, J., Reckamp, K. L., Baas, P., Crino, L., Eberhardt, W. E., Poddubskaya, E., et al. (2015). Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N. Engl. J. Med.* 373 (2), 123–135. doi: 10.1056/NEJMoa1504627

Brahmer, J. R., Tykodi, S. S., Chow, L. Q. M., Hwu, W., Topalian, S. L., Hwu, P., et al. (2012). Safety and activity of anti–PD-L1 antibody in patients with advanced cancer. *N. Engl. J. Med.* 366 (26), 2455–2465. doi: 10.1056/NEJMoa1200694

Chalmers, Z. R., Connelly, C. F., Fabrizio, D., Gay, L., Ali, S. M., Ennis, R., et al. (2017). Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med.* 9 (1), 34. doi: 10.1186/s13073-017-0424-2

Chang, L., Chang, M., Chang, H. M., and Chang, F. (2017). Microsatellite instability: a predictive biomarker for cancer immunotherapy. *Appl. Immunohistochem. Mol. Morphol.* 26 (2), e15–e21. doi: 10.1097/PAI.0000000000000575

Chowell, D., Morris, L. G. T., Grigg, C. M., Weber, J. K., Samstein, R. M., Makarov, V., et al. (2018). Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science* 359 (6375), 582–587. doi: 10.1126/science.aao4572

Chung, H. C., Arkenau, H.-T., Wyrwicz, L., Oh, D.-Y., Lee, K.-W., Infante, J. R., et al. (2016). Avelumab (MSB0010718C; anti-PD-L1) in patients with advanced gastric or gastroesophageal junction cancer from JAVELIN solid tumor phase Ib trial: analysis of safety and clinical activity. *J. Clin. Oncol.* 34 (15_suppl), 4009–4009. doi: 10.1200/JCO.2016.34.15_suppl.4009

Dangelo, S. P., Mahoney, M. R., Van Tine, B. A., Atkins, J. N., Milhem, M. M., Tap, W. D., et al. (2017). A multi-center phase II study of nivolumab +/– ipilimumab for patients with metastatic sarcoma (Alliance A091401). *J. Clin. Oncol.* 35, 11007–11007. doi: 10.1200/JCO.2017.35.15_suppl.11007

Dirix, L. Y., Takacs, I., Jerusalem, G., Nikolinakos, P., Arkenau, H. T., Forero-Torres, A., et al. (2018). Avelumab, an anti-PD-L1 antibody, in patients with locally advanced or metastatic breast cancer: a phase 1b JAVELIN Solid Tumor study. *Breast Cancer Res. Treat.* 167 (3), 671–686. doi: 10.1007/s10549-017-4537-5

Disis, M. L., Patel, M. R., Pant, S., Hamilton, E. P., Lockhart, A. C., Kelly, K., et al. (2016). Avelumab (MSB0010718C; anti-PD-L1) in patients with recurrent/refractory ovarian cancer from the JAVELIN Solid Tumor phase Ib trial: safety and clinical activity. *J. Clin. Oncol.* 34, 5533–5533. doi: 10.1200/JCO.2016.34.15_suppl.5533

Efron, B., and Tibshirani, R. (1997). Improvements on cross-validation: the 632+ bootstrap method. *J. Am. Stat. Assoc.* 92 (438), 548–560. doi: 10.1080/01621459.1997.10474007

El-Khoueiry, A. B., Sangro, B., Yau, T., Crocenzi, T. S., Kudo, M., Hsu, C., et al. (2017). Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): an open-label, non-comparative, phase 1/2 dose escalation and expansion trial. *Lancet* 389 (10088), 2492–2502. doi: 10.1016/S0140-6736(17)31046-2

Esteller, M. (2008). Epigenetics in cancer. *N. Engl. J. Med.* 358 (11), 1148–1159. doi: 10.1056/NEJMra072067

Fan, Y., Xi, L., Hughes, D. S., Zhang, J., Zhang, J., Futreal, P. A., et al. (2016). MuSE: accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling from sequencing data. *Genome Biol.* 17 (1), 178. doi: 10.1186/s13059-016-1029-6

Ferrari de Andrade, L., Tay, R. E., Pan, D., Luoma, A. M., Ito, Y., Badrinath, S., et al. (2018). Antibody-mediated inhibition of MICA and MICB shedding promotes NK cell-driven tumor immunity. *Science* 359 (6383), 1537–1542. doi: 10.1126/science.aao0505

Ferris, R. L., Blumenschein, G., Jr., Fayette, J., Guigay, J., Colevas, A. D., Licitra, L., et al. (2016). Nivolumab for recurrent squamous-cell carcinoma of the head and neck. *N. Engl. J. Med.* 375 (19), 1856–1867. doi: 10.1056/NEJMoa1602252

Fleming, G. F., Emens, L. A., Eder, J. P., Hamilton, E. P., Liu, J. F., Liu, B., et al. (2017). Clinical activity, safety and biomarker results from a phase Ia study of atezolizumab (atezo) in advanced/recurrent endometrial cancer (rEC). *J. Clin. Oncol.* 35 (15_suppl), 5585–5585. doi: 10.1200/JCO.2017.35.15_suppl.5585

Forster, M. D., and Devlin, M. J. (2018). Immune checkpoint inhibition in head and neck cancer. *Front. Oncol.* 8, 310. doi: 10.3389/fonc.2018.00310

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., et al. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 41 (Database issue), D808–D815. doi: 10.1093/nar/gks1094

Fuchs, C. S., Doi, T., Jang, R. W.-J., Muro, K., Satoh, T., Machado, M., et al. (2017). KEYNOTE-059 cohort 1: efficacy and safety of pembrolizumab (pembro) monotherapy in patients with previously treated advanced gastric cancer. *J. Clin. Oncol.* 35 (15_suppl), 4003–4003. doi: 10.1200/JCO.2017.35.15_suppl.4003

Goswami, S., Basu, S., and Sharma, P. (2018). A potential biomarker for anti-PD-1 immunotherapy. *Nat. Med.* 24 (2), 123–124. doi: 10.1038/nm.4489

Hamanishi, J., Mandai, M., Ikeda, T., Minami, M., Kawaguchi, A., Murayama, T., et al. (2015). Safety and antitumor activity of anti-PD-1 antibody, nivolumab, in patients with platinum-resistant ovarian cancer. *J. Clin. Oncol.* 33 (34), 4015–4022. doi: 10.1200/JCO.2015.62.3397

Hollebecque, A., Meyer, T., Moore, K. N., Machiels, J.-P. H., Greve, J. D., López-Picazo, J. M., et al. (2017). An open-label, multicohort, phase I/II study of nivolumab in patients with virus-associated tumors (CheckMate 358): efficacy and safety in recurrent or metastatic (R/M) cervical, vaginal, and vulvar cancers. *J. Clin. Oncol.* 35 (15_suppl), 5504–5504. doi: 10.1200/JCO.2017.35.15_suppl.5504

Johnson, D. B., Frampton, G. M., Rioth, M. J., Yusko, E., Xu, Y., Guo, X., et al. (2016). Targeted next generation sequencing identifies markers of response to PD-1 blockade. *Cancer Immunol. Res.* 4 (11), 959–967. doi: 10.1158/2326-6066.CIR-16-0143

Kjernielsen, L., Patel, O., Corbett, A. J., Nours, J. L., Meehan, B., Liu, L., et al. (2012). MR1 presents microbial vitamin B metabolites to MAIT cells. *Nature* 491 (7426), 717–723. doi: 10.1038/nature11605

Kvistborg, P., and Yewdell, J. W. (2018). Enhancing responses to cancer immunotherapy. *Science* 359 (6375), 516–517. doi: 10.1126/science.aar6574

Larkin, J., Chiarion-Sileni, V., Gonzalez, R., Grob, J. J., Cowey, C. L., Lao, C. D., et al. (2015). Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. *N. Engl. J. Med.* 373 (1), 23–34. doi: 10.1056/NEJMoa1504030

Le Tourneau, C., Gulley, J., Zarwan, C., Hoimes, C., Wong, D. J., Bauer, S., et al. (2017). 913PAvelumab in patients with metastatic adrenocortical carcinoma (mACC): results from the JAVELIN solid tumor trial. *Ann. Oncol* 28 (suppl_5), 324. doi: 10.1093/annonc/mdx371.067

Mahoney, K. M., Rennert, P. D., and Freeman, G. J. (2015). Combination cancer immunotherapy and new immunomodulatory targets. *Nat. Rev. Drug Discov.* 14 (8), 561–584. doi: 10.1038/nrd4591

Medina, P. J., and Adams, V. R. (2016). PD-1 pathway inhibitors: immuno-oncology agents for restoring antitumor immune responses. *Pharmacotherapy* 36 (3), 317–334. doi: 10.1002/phar.1714

Motzer, R. J., Escudier, B., McDermott, D. F., George, S., Hammers, H. J., Srinivas, S., et al. (2015). Nivolumab versus everolimus in advanced renal-cell carcinoma. *N. Engl. J. Med.* 373 (19), 1803–1813. doi: 10.1056/NEJMoa1510665

Pereira, C., Gimenez-Xavier, P., Pros, E., Pajares, M. J., Moro, M., Gomez, A., et al. (2017). Genomic profiling of patient-derived xenografts for lung cancer identifies B2M inactivation impairing immunorecognition. *Clin. Cancer Res.* 23 (12), 3203–3213. doi: 10.1158/1078-0432.CCR-16-1946-T

Powles, T., Odonnell, P. H., Massard, C., Arkenau, H. T., Friedlander, T. W., Hoimes, C. J., et al. (2017). Efficacy and safety of durvalumab in locally advanced or metastatic urothelial carcinoma: updated results from a phase 1/2 open-label study. *JAMA Oncol.* 3 (9), e172411–e172411. doi: 10.1001/jamaoncol.2017.2411

Prasad, V., and Kaestner, V. (2017). Nivolumab and pembrolizumab: monoclonal antibodies against programmed cell death-1 (PD-1) that are interchangeable. *Semin. Oncol.* 44 (2), 132–135. doi: 10.1053/j.seminoncol.2017.06.007

Ramilowski, J. A., Goldberg, T., Harshbarger, J., Kloppmann, E., Lizio, M., Satagopam, V. P., et al. (2015). A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat. Commun.* 6, 7866. doi: 10.1038/ncomms8866

Reardon, D. A., Kaley, T. J., Dietrich, J., Clarke, J. L., Dunn, G. P., Lim, M., et al. (2017a). Phase 2 study to evaluate safety and efficacy of MEDI4736 (durvalumab [DUR]) in glioblastoma (GBM) patients: an update. *J. Clin. Oncol.* 35 (15_suppl), 2042–2042. doi: 10.1200/JCO.2017.35.15_suppl.2042

Reardon, D. A., Omuro, A., Brandes, A. A., Rieger, J., Wick, A., Sepulveda, J., et al. (2017b). OS10.3 Randomized phase 3 study evaluating the efficacy and safety of nivolumab vs bevacizumab in patients with recurrent glioblastoma: CheckMate 143. *Neuro. Oncol.* 19 (suppl_3), iii21–iii21. doi: 10.1093/neuonc/nox036.071

Reck, M., Rodriguez-Abreu, D., Robinson, A. G., Hui, R., Csoszi, T., Fulop, A., et al. (2016). Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N. Engl. J. Med.* 375 (19), 1823–1833. doi: 10.1056/NEJMoa1606774

Robert, C., Schachter, J., Long, G. V., Arance, A., Grob, J. J., Mortier, L., et al. (2015). Pembrolizumab versus ipilimumab in advanced melanoma. *N. Engl. J. Med.* 372 (26), 2521–2532. doi: 10.1056/NEJMoa1503093

Rodriguez-Paredes, M., and Esteller, M. (2011). Cancer epigenetics reaches mainstream oncology. *Nat. Med.* 17 (3), 330–339. doi: 10.1038/nm.2305

Rong, L., Zhao, R., and Lu, J. (2017). Highly expressed long non-coding RNA FOXD2-AS1 promotes non-small cell lung cancer progression via Wnt/beta-catenin signaling. *Biochem. Biophys. Res. Commun.* 484 (3), 586–591. doi: 10.1016/j.bbrc.2017.01.141

Rosenberg, J. E., Hoffman-Censits, J., Powles, T., van der Heijden, M. S., Balar, A. V., Necchi, A., et al. (2016). Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *Lancet* 387 (10031), 1909–1920. doi: 10.1016/S0140-6736(16)00561-4

Sabbatino, F., Villani, V., Yearley, J. H., Deshpande, V., Cai, L., Konstantinidis, I. T., et al. (2016). PD-L1 and HLA class I antigen expression and clinical course of the disease in intrahepatic cholangiocarcinoma. *Clin. Cancer Res.* 22 (2), 470–478. doi: 10.1158/1078-0432.CCR-15-0715

Sanmamed, M. F., and Chen, L. (2018). A paradigm shift in cancer immunotherapy: from enhancement to normalization. *Cell* 175 (2), 313–326. doi: 10.1016/j.cell.2018.09.035

Schachter, J., Ribas, A., Long, G. V., Arance, A., Grob, J. J., Mortier, L., et al. (2016). Pembrolizumab versus ipilimumab for advanced melanoma: final overall survival analysis of KEYNOTE-006. *J. Clin. Oncol.* 34 (15_suppl), 9504–9504. doi: 10.1200/JCO.2016.34.15_suppl.9504

Scherpereel, A., Mazieres, J., Greillier, L., Dô, P., Bylicki, O., Monnet, I., et al. (2017). Second- or third-line nivolumab (Nivo) versus nivo plus ipilimumab (Ipi) in malignant pleural mesothelioma (MPM) patients: results of the

IFCT-1501 MAPS2 randomized phase II trial. *J. Clin. Oncol.* 35 (18_suppl), LBA8507-LBA8507. doi: 10.1200/JCO.2017.35.18_suppl.LBA8507

Schmidt, C. (2017). The benefits of immunotherapy combinations. *Nature* 552 (7685), S67. doi: 10.1038/d41586-017-08702-7

Sharma, P., Retz, M., Siefkerradtke, A., Baron, A. D., Necchi, A., Bedke, J., et al. (2017). Nivolumab in metastatic urothelial carcinoma after platinum therapy (CheckMate 275): a multicentre, single-arm, phase 2 trial. *Lancet Oncol.* 18 (3), 312–322. doi: 10.1016/S1470-2045(17)30065-7

Tang, J., Pearce, L., O'Donnell-Tormey, J., and Hubbard-Lucey, V. M. (2018). Trends in the global immuno-oncology landscape. *Nat. Rev. Drug Discov.* 130 (176), 112. doi: 10.1038/nrd.2018.167

Tawbi, H. A., Burgess, M., Bolejack, V., Van Tine, B. A., Schuetze, S. M., Hu, J., et al. (2017). Pembrolizumab in advanced soft-tissue sarcoma and bone sarcoma (SARC028): a multicentre, two-cohort, single-arm, open-label, phase 2 trial. *Lancet Oncol.* 18 (11), 1493–1501. doi: 10.1016/S1470-2045(17)30624-1

Topalian, S. L., Taube, J. M., Anders, R. A., and Pardoll, D. M. (2016). Mechanism-driven biomarkers to guide immune checkpoint blockade in cancer therapy. *Nat. Rev. Cancer* 16 (5), 275–287. doi: 10.1038/nrc.2016.36

Tsai, H. C., and Baylin, S. B. (2011). Cancer epigenetics: linking basic biology to clinical medicine. *Cell Res.* 21 (3), 502–517. doi: 10.1038/cr.2011.24

Wainberg, Z. A., Segal, N. H., Jaeger, D., Lee, K.-H., Marshall, J., Antonia, S. J., et al. (2017). Safety and clinical activity of durvalumab monotherapy in patients with hepatocellular carcinoma (HCC). *J. Clin. Oncol.* 35 (15_suppl), 4071–4071. doi: 10.1200/JCO.2017.35.15_suppl.4071

Xiao, Q., Zhou, D., Rucki, A. A., Williams, J. K., Zhou, J., Mo, G., et al. (2016). Cancer-associated fibroblasts in pancreatic cancer are reprogrammed by tumor-induced alterations in genomic DNA methylation. *Cancer Res.* 76 (18), 5395–5404. doi: 10.1158/0008-5472.CAN-15-3264

Yarchoan, M., Hopkins, A., and Jaffee, E. M. (2017). Tumor mutational burden and response rate to PD-1 inhibition. *N. Engl. J. Med.* 377 (25), 2500–2501. doi: 10.1056/NEJMc1713444

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16 (5), 284–287. doi: 10.1089/omi.2011.0118

Zhang, M.-W., Fujiwara, K., Che, X., Zheng, S., and Zheng, L. (2017). DNA methylation in the tumor microenvironment. *J. Zhejiang Univ. Sci. B* 18 (5), 365–372. doi: 10.1631/jzus.B1600579

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling Editor declared a shared affiliation, though no other collaboration, with one of the authors ZH.

Copyright © 2019 Xue, Cui, Zhou, Zhu, Chen, Liang, Tang, Huang, Zhang, Hu, Yuan and Xiong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 05 September 2019
doi: 10.3389/fgene.2019.00774

A Hybrid Ensemble Approach for Identifying Robust Differentially Methylated Loci in Pan-Cancers

*Qi Tian[1], Jianxiao Zou[1], Yuan Fang[1], Zhongli Yu[1], Jianxiong Tang[1], Ying Song[1] and Shicai Fan[1,2]**

[1] School of Automation Engineering, University of Electronic Science and Technology of China, [2] Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu, China

DNA methylation is a widely investigated epigenetic mark that plays a vital role in tumorigenesis. Advancements in high-throughput assays, such as the Infinium 450K platform, provide genome-scale DNA methylation landscapes in single-CpG locus resolution, and the identification of differentially methylated loci has become an insightful approach to deepen our understanding of cancers. However, the situation with extremely unbalanced numbers of samples and loci (approximately 1:1,000) makes it rather difficult to explore differential methylation between the sick and the normal. In this article, a hybrid approach based on ensemble feature selection for identifying differentially methylated loci (HyDML) was proposed by incorporating instance perturbation and multiple function models. Experiments on data from The Cancer Genome Atlas showed that HyDML not only achieved effective DML identification, but also outperformed the single-feature selection approach in terms of classification performance and the robustness of feature selection. The intensive analysis of the DML indicated that different types of cancers have mutual patterns, and the stable DML sharing in pan-cancers is of the great potential to be biomarkers, which may strengthen the confidence of domain experts to implement biological validations.

Keywords: DNA methylation, differentially methylated loci, ensemble feature selection, robustness, pan-cancers

OPEN ACCESS

Edited by:
Yun Liu,
Fudan University, China

Reviewed by:
Osman A. El-Maarri,
University of Bonn, Germany
Daniel Vaiman,
Institut National de la Santé et de la Recherche Médicale (INSERM), France

***Correspondence:**
Shicai Fan
shicaifan@uestc.edu.cn

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *10 May 2019*
Accepted: *23 July 2019*
Published: *05 September 2019*

Citation:
Tian Q, Zou J, Fang Y, Yu Z, Tang J, Song Y and Fan S (2019) A Hybrid Ensemble Approach for Identifying Robust Differentially Methylated Loci in Pan-Cancers. Front. Genet. 10:774. doi: 10.3389/fgene.2019.00774

INTRODUCTION

DNA methylation is one of the essential epigenetic mechanisms, which plays a vital role in normal development and is closely correlated with the cell growth, differentiation, and transformation in eukaryotes (Robertson, 2005; Suzuki and Bird, 2008; Laird, 2010; Jones, 2012).Failure of proper maintenance of epigenetic marks, like abnormal DNA methylation, may result in inappropriate activation or inhibition of various signaling pathways, leading to diseased states, even cancers (Esteller, 2007; Hanahan and Weinberg, 2011; Dawson and Kouzarides, 2012; Aran and Hellman, 2013; Tolstorukov et al., 2013). For example, aberrant promoter hypermethylation that is associated with inappropriate gene silencing affects virtually every step in tumor progression (Jones and Baylin, 2002). So, the investigation of differential methylation, which displays the inherent difference between normal and tumor samples, could help us deepen our perception of oncogenesis and may assist in the early diagnosis of cancers (Tost, 2007; Deng et al., 2010).

High-throughput bisulfite sequencing provides a new stage for researchers to analyze methylation variability at single-base resolution, and the identification of differentially methylated loci (DML)

has become an insightful attempt for detection of tumor markers (Cokus et al., 2008; Down et al., 2008). In the early stage, obtaining methylation data is based on bisulfite sequence technique (BS-seq), and Lister et al. (2009) first use Fisher exact test to select differential methylation sites. Then, more R packages have been developed for identifying DML with this kind of data. BiSeq (Hebestreit et al., 2013) and DSS (Feng et al., 2014) concentrate on identifying DML through Wald tests, whereas MethylSig (Park et al., 2014) applies likelihood ratio tests for DML identification. Infinium HumanMethylation450 BeadChip is now widely used in methylation analysis for its advantages of lower cost and easier experimental protocol compared with BS-seq, like WGBS, and is suggested to be suitable for large-scale studies (Dedeurwaerder et al., 2011). For example, IMA achieves detection of site-level differential methylation using Wilcoxon rank-sum tests with HM450 data (Wang et al., 2012). Compared with IMA, based on the analysis of covariance, FastDMA performs better in identifying DML with higher computational efficiency (Wu et al., 2013). RnBeads provides a comprehensive pipeline for analysis and interpretation of DNA methylation with *t* statistics analysis based on linear model and empirical Bayes (Assenov et al., 2014). We consider that the identification of DML is to search for loci that can significantly distinguish between the normal and the sick, and therefore the essence of this problem can be regarded as applying feature selection to the identification of DML. Additionally, compared with the methods mentioned above, feature selection approaches can take the feature redundance and irrelevance into account, and this could be a benefit for selecting more significant DML.

However, considering that the HM450 data have a small number of samples but high dimensional features (approximately 1:1,000), the results from general feature selection methods for identifying DML will have poor robustness (Kim, 2009). The robustness (reproducibility or stability) of selected loci is extremely important for identifying DML, as domain experts tend to do subsequent analysis and validations with stable results. While feature selection has been considered a de facto standard in microarray data mining (Bolon-Canedo et al., 2014), how to identify robust DML with feature selection has received little attention. Recent advancements in ensemble feature selection provide a promising approach to solve the robustness problem in large-scale biological data (Saeys et al., 2008; Abeel et al., 2010; Liu et al., 2010; Yang et al., 2010; Haury et al., 2011; Yang et al., 2011; Yu et al., 2012). The rationale for this idea is combining single, less stable feature selectors to yield a more robust one, which is the same as ensemble learning: in a first step, a number of different feature selectors are used, and in a final phase, the output of these separate selectors is aggregated and returned as the final (ensemble) result. Specifically, there are two major means to achieve ensemble feature selection; one of them is data diversity (instance perturbation), which uses the same feature selection method on different data subsets from multiple sampling on the original data set, and the other is function diversity, which implements different feature selection methods on the original data set (Saeys et al., 2008; Yang et al., 2010; Awada et al., 2012; Yu et al., 2012).

In this article, we aggregate data diversity and function diversity to propose a hybrid ensemble approach for identification of DML (HyDML). Under the framework of ensemble feature selection, this newly proposed method not only can realize the effective identification of DML, but also can accommodate for the robustness of the results. Additionally, taking advantage of the large-scale Infinium 450K methylation data produced by The Cancer Genome Atlas (TCGA) project, we performed intensive analysis to look further into interrelationships between differential methylation and cancers and found that different cancers have common patterns, and robust DML sharing in pan-cancers is of the great potential to be biomarkers.

MATERIALS AND METHODS

Cancers and Samples

For feeding the algorithm and analysis, in total 13 cancers are selected with both normal and tumor samples. Specifically, these cancers are bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), head and neck squamous cell carcinoma (HNSC), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), prostate adenocarcinoma (PRAD), thyroid carcinoma (THCA), and uterine corpus endometrial carcinoma (UCEC). In all, there are 6,189 samples including 699 normal samples and 5491 tumor samples (**Table S1**).

DNA Methylation Data and Preprocess

We downloaded the DNA methylation data from TCGA data portal (https://tcga-data.nci.nih.gov/tcga/) for our selected samples. The methylation data are generated by Illumina Infinium HumanMethylation450k BeadChip technique. The Illumina Infinium assay utilizes a pair of probes for each CpG site, one probe for the methylated allele and the other for the unmethylated version. The methylation level is then estimated, based on the measured intensities of this pair of probes, as the ratio of methylated signal to the sum of methylated and unmethylated signal, which ranges from 0 (absent methylation) to 1 (completely methylated). To assess the ability of the selected DML to distinguish between the two types of samples (tumor and normal), we retrieved three independent test sets from the NCBI database. The three data sets are also obtained by HM 450 technique, including samples of breast (GSE52635), liver (GSE54503), and lung (GSE66836) cancer, as well as corresponding normal tissue data records (**Table S1**). For each type of cancer, the original methylation data record the methylation level at more than 450,000 loci. A series of preprocessing is required before implementing the selection of DML, which can reduce the computational complexity as well as improve the accuracy of the final results. The preprocessing steps for the methylation data are as follows: i) The 450k methylation chip uses two different types of probes (type I and type II) when measuring the locus methylation and results in two different types of data distribution. We use the SWAN algorithm to eliminate the abiotic variation caused by the measurement of the two probes while preserving the biological differences

of the samples (Maksimovic et al., 2012). ii) Eliminate batch effects caused by system bulk effects or abiotic differences using empirical Bayesian (EB) methods (Johnson et al., 2007). iii) Filter out some of the minimal variance loci to avoid dimensionality disasters and remove significantly unrelated redundant loci. After completing all of the preprocessing steps, approximately 350,000 feature sites are obtained for each cancer for subsequent feature selection. Considering polymorphisms (single-nucleotide polymorphisms), we chose to mark these sites in the results, and users can decide the stringency of probe filtering appropriate for their analysis.

Hybrid Ensemble Approach for Identification of DML

First, in order to obtain a diverse set of feature selectors, we perform multiple samplings on training samples to generate data subsets. To this end, we make use of resampling and cross-validation, integrating classifier training into the ensemble feature selection framework for selecting loci that are informative for classifying tumor and normal samples. In each sampling, the whole data set is divided into 10 pieces with the same number of samples, and each of them can be regarded as a test subset to validate subsequent classification performance, while the rest automatically becomes a training set for feature selection and classifier training (constructed with support vector machine) (Cortes and Vapnik, 1995). The instance level perturbation here can bring in the stability for feature selection after aggregating the result of each data subset, because the stable features are more likely to appear in different training subsets when the sample changes slightly. Then, generating functional diversity is achieved by using multiple feature selection methods on the same training set. With consideration of high dimensionality and small sample size of the 450k methylation data, embedded feature selection methods could be a practical choice for the appropriate computation complexity. Thus, we choose R packages "glmnet," "MDFS" and "rmcfs" as the basic feature selection approaches (Friedman et al., 2010; Draminski and Koronacki, 2018; Piliszek et al., 2018). Taking the advantages of combing L1 and L2 regularization (elastic net), glmnet can achieve variable extraction for the microarray data with high dimension but small number of samples. Combining linear model with elastic net for feature selection, the optimization function is as follows:

$$\arg\min_{w}\left\{\sum_{i=1}^{m}(y_i-w^T x_i)^2+\lambda\left[\alpha\sum_{j=1}^{p}|w_j|+(1-\alpha)\sum_{j=1}^{p}w_j^2\right]\right\}$$

where w represents the feature weight coefficient, m represents the number of samples, and p represents the total number of features in the data set. λ is used to balance the empirical risk and model complexity, whereas α is used to balance the regularization of L1 and L2. In MDFS, we apply feature selection with max information gain criterion, which measures the worth of a feature by computing the information gain values with respect to the class. For rmcfs, it relies on a Monte Carlo approach to select informative features and is capable of incorporating interdependencies between features. The three basic feature selection algorithms can be well adapted to the high-dimensional and small-sample characteristics of 450k methylation data, and the whole calculation amount is moderate, while classification performance can be guaranteed. For each data subset, aggregating the results of multiple feature selection methods could further enhance the stability. More formally, consider an ensemble feature selector $E = \{F_1, F_2, \ldots, F_s\}$ and each F_i provides a feature ranking $\mathbf{f_i} = \left(f_i^1, f_i^2, \ldots, f_i^N\right)$, f_i denotes the feature weight of each F_i and N represents the nth feature. Hence, a general aggregation formulation for the ensemble ranking **f**, obtained by weighted summing the ranks over all $\mathbf{f}_i$, is as follows:

$$\mathbf{f} = \sum_{i=1}^{s} acc_i^* \mathbf{f}_i$$

where acc_i donates the accuracy of the corresponding test set on the classifier trained by feature selector F_i, and **f** also can be regarded as the aggregation ranking for the ensemble feature selector. Here, s = 3, which represents the three basic feature selection methods, and we can get the preliminary DML at this level of aggregation. Then, taking the union set of obtained loci subsets is the second level of aggregation, and the corresponding formula representation is as follows:

$$\mathbf{f} = \sum_{i=1}^{s} \mathbf{f}_i$$

where s donates the number of data subsets, and $\mathbf{f}_i$ is the feature ranking of corresponding data subset. In this way, one aggregated ranking of all the features for each sampling can be yielded. We perform 10 iterations for generating more data subsets to further improve the stability of selected loci, and with the idea of bagging, the final DML set consisted of loci that appear more than five times in 10 iterations. The overall algorithm framework for one sampling is shown in **Figure 1**, and pseudo code flow is as follows:

ALGORITHM: HYDML

Require: methylation data **D**
Ensure: Divide data set D into $\{\boldsymbol{D}_1, \boldsymbol{D}_2, \ldots, \boldsymbol{D}_k, \ldots \boldsymbol{D}_{10}\}$ for 10-fold cross-validation;
1: begin
2: **for** k = 1 to 10 **do**. The data subset $\boldsymbol{D}_k$ is used as a **test set**, while other data subsets are used as a **training set** to produce DML with **multiple feature selection methods**; calculate $f_i^k \boldsymbol{f}_k$ for each feature in $\boldsymbol{D}_k$ with $\boldsymbol{acc}_i(\boldsymbol{i} = 1, 2, 3)$; filter out loci with the $\mathbf{f}_k < 0.01$; **end for**;
3: Take **union** set of $\{\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_{10}\}$ to obtain $\mathbf{F}_1$
4: **for** t = 1 to 10 **do**, step 2 and step 3; **end for**;
5: Aggregate $\boldsymbol{F}_1 \sim \boldsymbol{F}_{10}$ with bagging which filters out loci which appears less than five times; record as $\boldsymbol{F}$
6: **return** $\boldsymbol{F}$;
7: **End**

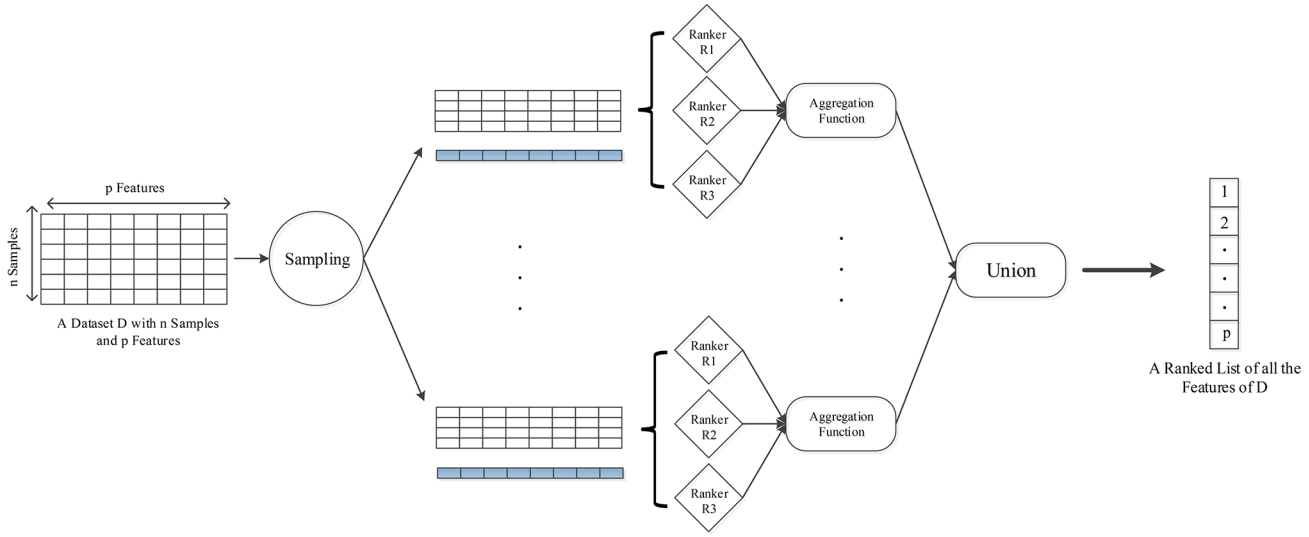


FIGURE 1 | The framework of HyDML for identifying differentially methylated loci using hybrid ensemble feature selection approach.

PERFORMANCE EVALUATION AND COMPARISON

Stability Measure

To measure the effect of our hybrid ensemble technique on the feature selection results, following Saeys et al. (2008), we take a similarity-based approach where feature stability is measured by comparing the signatures from the k feature selectors. The more similar all signatures are, the higher the stability measure will be. The overall stability can be defined as the average over all pairwise similarity comparisons between different signatures:

$$S_{tot} = \frac{2\sum_{i=1}^{k}\sum_{j=i+1}^{k} S\left(f_i, f_j\right)}{k\left(k-1\right)}$$

where f_i represents the signature obtained by the selection method on subsampling $i(1 \le i \le k)$; k is the number of data subsets; $S(f_i, f_j)$ is a similarity measure for feature subsets, which denotes the stability of f_i and f_j. Here, we use Jaccard index (Saeys et al., 2008) as $S(f_i, f_j)$:

$$S\left(f_i, f_j\right) = \frac{\left|f_i \bigcap f_j\right|}{\left|f_i \bigcup f_j\right|} = \frac{\sum_l I\left(f_i^l = f_j^l = 1\right)}{\sum_l I(f_i^l + f_j^l > 0)}$$

where the indicator function I(.) returns 1 if its argument is true, and zero otherwise. In the sequel, the overall stability S_{tot} is simply denoted by $S(f_i, f_j)$.

Classification Performance Measure

To evaluate the classification performance and perform comparisons, we use several characteristics of classification performance all derived from the confusion matrix. These characteristics are TP, TN, FP, and FP, which denote true-negatives, true-positives, false-negatives, and false-positives, respectively. All the performance metrics are calculated by these characteristics, including TPR (true-positive rate), FPR (false-negative rate), ACC (classification accuracy), Precision, Recall, and F1 score. We also include the area under the receive operating characteristic curve, which is defined by a function of sensitivity and specificity, further abbreviated as AUC.

RESULTS

Characteristics of Differentially Methylated Loci in 13 Cancers

For each of the 13 cancers, we finally obtained a set of DML, which varies from 5,700 in COAD to 14,516 in THCA (**Table S2**). Through t-SNE clustering (van der Maaten and Hinton, 2008), we found that these differential methylation sites were able to distinguish the difference between the normal and the sick, especially in COAD, ESCA, and KIRC (**Figure 2**). While very few samples were misclassified, it was probably due to the information compression since the original feature dimension is reduced by thousands of times during the t-SNE clustering process.

We first explored the distribution of DML in 22 pairs of autosomes for each cancer, which could help us to find out which chromosome gets potential extensive genetic variation when cancer occurs. To this end, we calculated the distribution density of the DML on each autosome, using ratio of the number of DML to the number of CpG sites determined by the 450K chip (**Figure S1A**). We can see from the results that chromosome 20 was

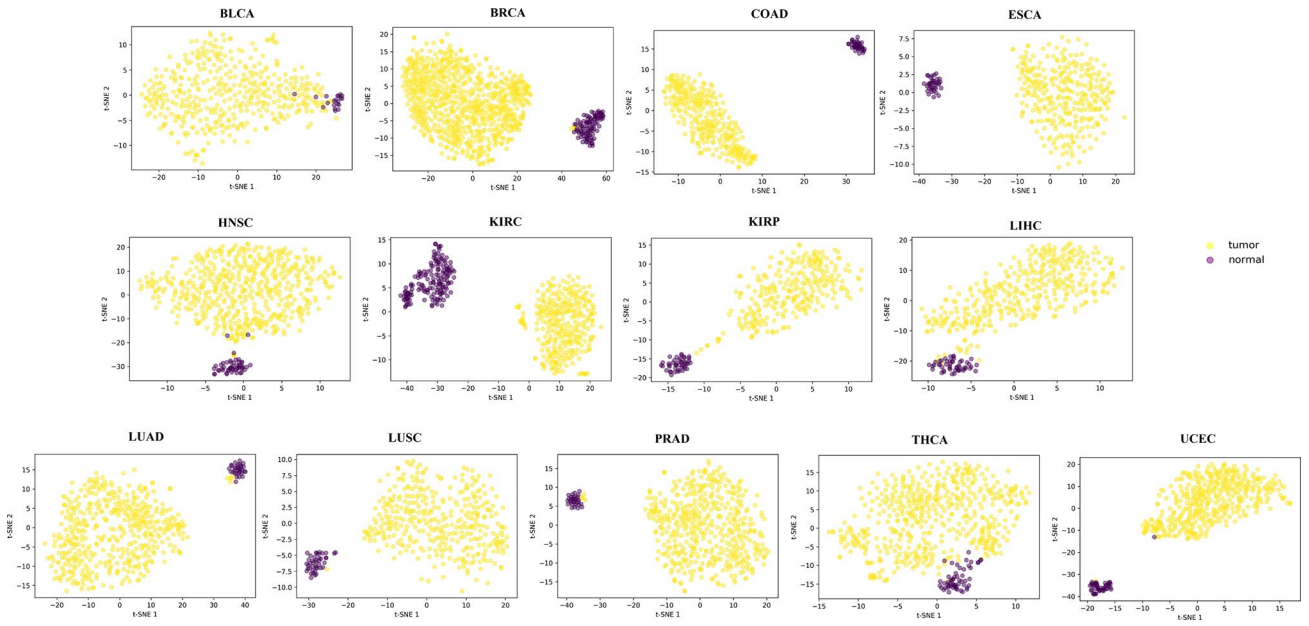


FIGURE 2 | The clustering results by t-SNE using the obtained differential methylation sites for each cancer.

enriched with more sites, whereas the DML were less distributed on chromosome 1, 9, oppositely. Combining the functional regions of genes on the chromosome, we further analyzed the distribution of DML in the promoter region (regions from 2,000 bps upstream to the transcription start site), gene body (excluding promoter region), and intergenic region for each cancer. Most of DML were located in nonpromoter regions (gene body and intergenic region; **Figure S1B**). However, considering that the promoter region occupied only a small part of the genome, the number of DML accounted for more than 20%, indicating that the abnormal methylation of this short functional region had an important impact on the tumorigenesis (Jones and Baylin, 2002; Baylin and Ohm, 2006). Most DML were distributed on CpG islands (**Figure S1C**), which has been reported that aberrant methylation of CpG islands was related to transcriptional gene silencing or activation of multiple oncogenes (Costello et al., 2000; Chan et al., 2002; Klutstein et al., 2016; Soozangar et al., 2018).

We also observed that biologically similar cancers shared more mutual DML through hierarchical clustering using similarity metric based on Jaccard index (**Figure S2**). Specifically, smoking- and drug addiction-related cancers, like LUSC and HNSC, were clustered together (Brennan et al., 1995; Johnson et al., 2005; Campbell et al., 2016). KIRC and KIRP were both due to renal lesion. High-risk cancers that were predisposed to women, such as BRCA and UCEC, shared more DML and were clustered together.

Robust Feature Selection Improves the Classification Performance

First, we compared our newly proposed method to its baseline methods, glmnet, rmcfs, and MDFS when the number of loci gradually decreased. This could help us analyze the robustness of the results from different feature selection methods as the features reduced, or if a feature selection method could identify more robust features, the decrement of features would not have a significant impact on the results. Here, for the three baseline methods, the feature sets were produced by a default configuration. Using the comprehensive classification metric, AUC, **Figure 3A** displays the trend of AUC change as the feature number reduced on PRAD data set. It can be observed that our ensemble approach clearly improved upon the baselines in terms of classification performance as the loci decreased. We also implemented the comparison on data of the other 12 cancers, and the results showed that the hybrid ensemble framework was superior to single-feature selection methods, thus demonstrating that the ensemble methods were better capable of eliminating noisy and irrelevant dimensions (**Figure S3**). We also compared the stability or robustness measure S_{tot} (based on Jaccard Index, see *Materials and Methods*), and the results in all 13 cancers showed the hybrid ensemble approach (HyDML) performed better than single-feature selection methods, which could be a benefit in performing subsequent analysis with the selected differential methylation sites (**Figure 3B**).

Moreover, three independent test sets from the NCBI database (BRCA: GSE52635; LIHC: GSE54503; LUAD: GSE66836) were used to compare HyDML with classical DML identification methods, including FastDMA and RnBeads, for analyzing the differences between the ensemble feature selection approach and the statistical test method. Using the original DML previously selected from the three cancers as training sets, we constructed a classification model based on SVM and performed the verification with the test sets. The results showed that DML selected by HyDML performed better than FastDMA and RnBeads (**Table 1**). Compared with the two classical DML

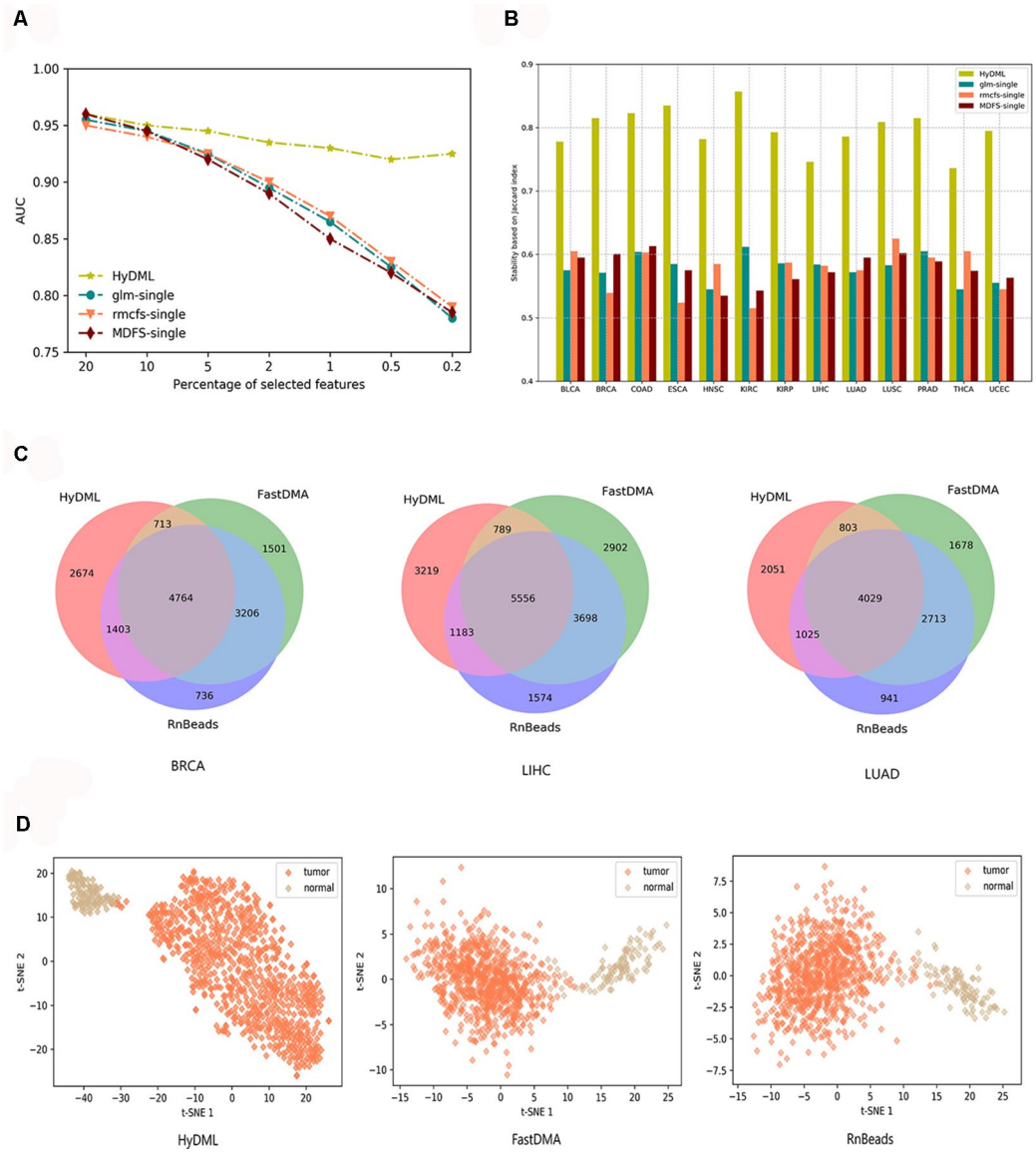


FIGURE 3 | Thorough classification performance and the robustness measure to compare different models in identifying differential methylation sites. **(A)** Classification performance of HyDML and its corresponding submethods as the selected features (loci) gradually reduce in PRAD. **(B)** Comparison of robustness measure using Jaccard index in 13 cancers for HyDML and its corresponding submethods. **(C)** Relationship of differential methylation sites obtained by HyDML/FastDMA/RnBeads in BRCA; LIHC and LUAD (corresponding to the three independent test sets) using a Venn picture. **(D)** T-SNE clustering results in BRCA using the unique differential methylation sites selected by each method.

finding approaches, the selected feature from HyDML showed better generalization ability in distinguishing the normal and tumor samples. Then, we analyzed the loci selected by the three methods to verify whether the loci were distinct from each other. Experiments on data of the three cancers showed that most DML were identical for the three methods, whereas FastDMA and RnBeads shared more mutual DML (**Figure 3C**). To capture the key differences of the three methods, we further studied the DML, which were uniquely selected by the corresponding method (the loci selected by one of the methods and not selected by the other two methods), through t-SNE clustering, and the results of BRCA showed that the uniquely selected DML from HyDML were more able to describe the difference between the normal and the sick (**Figure 3D**). The clustering results of the other two cancers can be obtained in **Figure S4**, and HyDML not surprisingly displayed better performance in classifying normal and tumor samples. This indicated that the differential methylation sites obtained by the hybrid ensemble approach were more likely to be reliable in biological validations. One evident reason for this was that HyDML takes the robustness of selected loci into account, and this could be rewarding to produce better DML in terms of analyzing the difference between the normal and the sick.

TABLE 1 | Classification performance comparison on three independent test sets.

	TPR	FPR	ACC	AUC	Precision	Recall	F1
GSE52635							
FastDMA	0.958	0.083	0.938	0.924	0.921	0.958	0.939
RnBeads	0.938	0.042	0.948	0.935	0.957	0.938	0.947
HyDML	**0.979**	**0.042**	**0.969**	**0.968**	**0.959**	**0.979**	**0.969**
GSE54503							
	TPR	FPR	ACC	AUC	Precision	Recall	F1
FastDMA	0.909	0.1667	0.8712	0.897	0.779	0.909	0.839
RnBeads	0.955	0.1515	0.9016	0.923	0.863	0.955	0.906
HyDML	**0.969**	**0.091**	**0.9408**	**0.962**	**0.914**	**0.969**	**0.941**
GSE66836							
	TPR	FPR	ACC	AUC	Precision	Recall	F1
FastDMA	0.909	0.316	0.886	0.876	0.961	0.909	0.934
RnBeads	0.915	0.263	0.896	0.893	0.968	0.915	0.940
HyDML	**0.951**	**0.158**	**0.94**	**0.943**	**0.981**	**0.951**	**0.966**

In bold font: best performance.

Pan-Cancer–Related DML Provide a Landscape of Commonality in Different Cancers

In order to further analyze the association between DNA methylation and cancer, we investigated the differential methylation sites that occurred in multiple cancers, which could help us reveal the pan-cancer–associated methylation patterns. First, we defined a selected site as a pan-cancer differentially methylated locus (pDML) if it occurred no less than 10 times in 13 cancers. We in total obtained 338 pDML, in which some of them presented as hypermethylated, whereas the others presented obvious hypomethylation, expressed by median value in normal and tumor samples (**Table S3**). By combining the methylation expression levels of pDML in tumor samples, different cancers reflected similarities in methylation variation (**Figure 4**). For example, LUAD and LUSC were clustered together as a result of carcinogenesis of lung tissues, and kidney disease–related cancer, such as KIRC and KIRP, were also shown to be similar in terms of pDML. This verified the methylation specificity expression caused by the differentiation of tissues, and even when the tissues were cancerous, there was a certain degree of difference in methylation variability between tissues, or the cancer subtypes of the same tissue had more similar methylation patterns.

In these pDML, we also found that, one probe, cg02829688, was significantly hypermethylated (the methylation level of loci in tumor samples was higher than that in normal samples) in all 13 cancers (**Figure 5**). Through the annotation files, we found that it was located at chr1:119527008 in a CpG island and belonged to a differentially methylated region (experimentally determined). Moreover, the corresponding upstream and downstream regions were located in a target gene, *TBX15*. It has been demonstrated that *TBX15* plays a vital role in multiple cancers, such as non–small cell lung cancer (Carvalho et al., 2013), thyroid cancer (Arribas et al., 2015), and ovarian carcinoma (Gozzi et al., 2016), and especially has been proved to be a methylation marker of prostate cancer (Kron et al., 2012). Moreover, Chelbi et al. (2011) identified a region located in the distal promoter of the *TBX15* that was differentially methylated and suggested that *TBX15* might be involved in the pathophysiology of placental diseases.

Using AME (McLeay and Bailey, 2010), the motif enrichment tool of MEME Suite, we detected sequence motifs that were enriched in the background sequences generated from the pDML, which were located in promoter regions and identified 84 motifs (**Table S4**). The motif of IRF3 was the most significantly enriched one (P = 5.55e-21) (**Figure 6A**), and the gene expression for IRF3 has been experimentally determined in multiple tissues (**Figure 6B**). IRF3 as a transcription factor has been reported as a regulator in type I interferon genes playing a vital role in mammalian response to pathogens and considered to be implicated in various biological pathological conditions, including cancer (Wang et al., 2017; Andrilenas et al., 2018). Baylin et al. (2006) also demonstrated that DNA methyltransferase inhibitors triggered viral defense and induced IRF3 to translocate to the nucleus and activated transcription of IFNβ1 to influence immune signaling in cancers (Chiappinelli et al., 2015).

Additionally, we had a deeper insight into the relationship between methylation and cancers through analyzing the corresponding biological pathways. Using the KEGG pathway database (Kanehisa and Goto, 2000), **Figure 7** showed the number of metabolic pathways for DML-associated genes in each cancer ($P < 0.05$). Then, we summarized the pathways that occurred in at least seven cancers and denoted as pan-cancer methylation-related pathways (PMPs) and obtained in total 11 PMPs, where 10 of them have been reported to be associated with cancers (**Table 2**). The only one PMP, neuroactive ligand-receptor interaction, has not been proven to be directly or indirectly associated with cancers, but further research is needed for deeper exploration.

DISCUSSION

Identifying DML is a promising approach to reveal the inherent intricacy between aberrant DNA methylation and tumorigenesis, and recent studies have paid more attention to this essential epigenetic mechanism. Taking advantage of the large-scale DNA methylation data produced by TCGA,

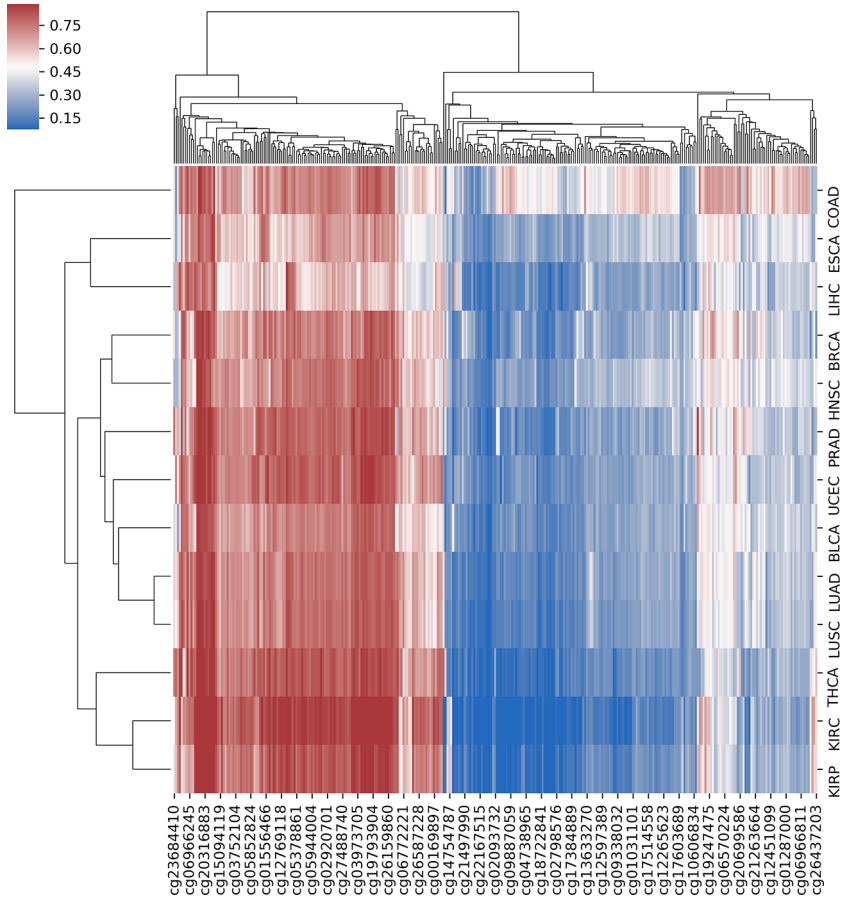


FIGURE 4 | The hierarchical clustering with heat map using all predefined pan-cancer differential methylation sites in 13 cancers.

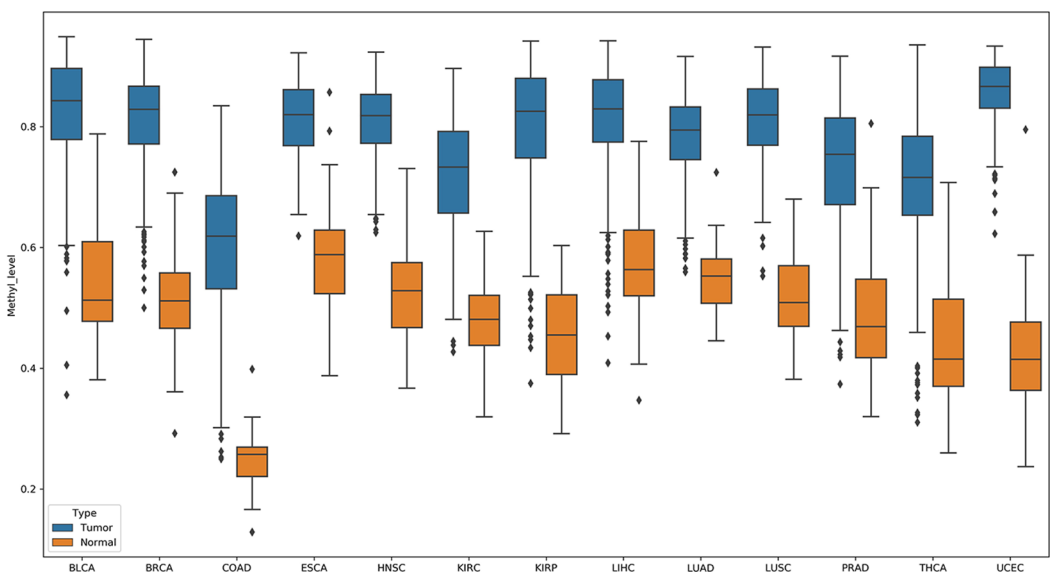


FIGURE 5 | Probe cg02829688 showed significant hypermethylation (the methylation level of loci in tumor samples were higher than those in normal samples) in all 13 cancers.

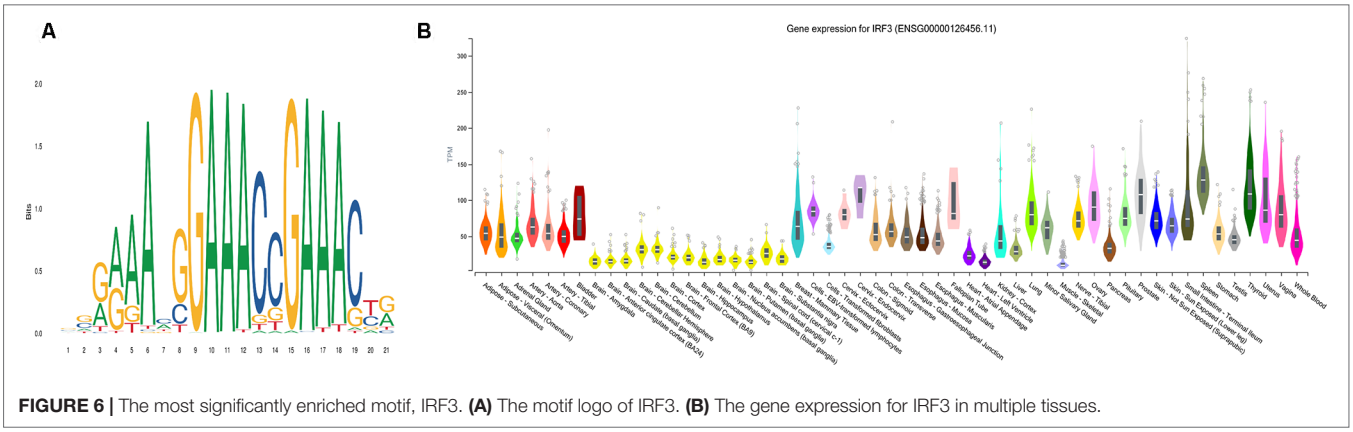


FIGURE 6 | The most significantly enriched motif, IRF3. **(A)** The motif logo of IRF3. **(B)** The gene expression for IRF3 in multiple tissues.

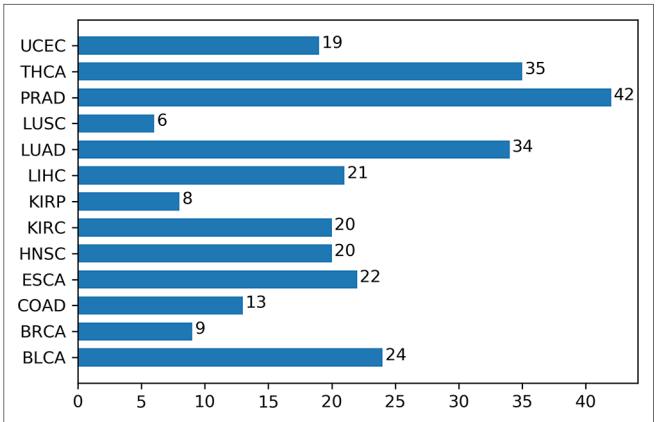


FIGURE 7 | The number of metabolic pathways for DML-associated genes in each cancer.

TABLE 2 | PMPs and their corresponding relations to cancer.

Pan-cancer methylation related pathways	Related to cancer?
Antigen processing and presentation (Chapman et al., 2013)	Yes
Allograft rejection (Leone et al., 2013)	Yes
cAMP signaling pathway (Fajardo et al., 2014)	Yes
Cell adhesion molecules (Okegawa et al., 2004)	Yes
Graft-versus-host disease (Curtis et al., 2005)	Yes
MicroRNAs in cancer (Wiemer, 2007)	Yes
Nicotine addiction (Benowitz, 2009)	Yes
Rap1 signaling pathway (Zhang et al., 2017)	Yes
Type II diabetes mellitus (Shlomai et al., 2016)	Yes
Autoimmune thyroid disease (Turken et al., 2003)	Yes
Neuroactive ligand-receptor interaction	Unknown

we investigated the differential methylation in 13 cancers with a newly proposed approach under hybrid ensemble feature selection framework. Compared with single-feature selection methods in identifying DML, HyDML could achieve identifying more robust loci, and the improvement of reproducibility of feature selection algorithm's results can enhance the confidence of researchers in experimental verification, especially in finding biomarkers. Compared with classical DML identification methods based on traditional statistic theory (such as FastDMA and RnBeads), feature selection–based approaches could select more informative loci that are closely related to the difference between the normal and the sick, as well as eliminating noisy and irrelevant loci, especially when dealing with microarray data of sparse samples and high-dimensional features. By t-SNE clustering, the results showed that the selected loci could distinguish between the normal and the sick well in each cancer, and the results from the independent test sets demonstrated that the classification model constructed by loci from HyDML had better generalization ability.

Additionally, comprehensive investigation of the pDML showed that different cancers shared some common patterns in methylation variability at CpG locus resolution and revealed the potential similarities in different cancers. We found that same tissues share more abnormal methylation patterns with different subtypes of tumorigenesis, such as KIRC and KIRP, and LUAD and LUSC. This may indicate that the tissue specificity of methylation is preserved even when the tissue is cancerous. We also found a locus (cg02829688), which was hypermethylated in 13 cancers, located in a functional region on the genome, and could be of great potential to be an oncogenesis biomarker. Enriched motifs analysis from the background sequences of pDML revealed the potential influence on transcription function by CpG methylation, and the most significantly enriched motif, IRF3, has been reported playing a vital role in tumorigenesis. Through pathway analysis, some pan-cancer–related pathways were also determined, which have been reported playing a vital role in start, development, and metastasis of tumors.

As an import epigenetic mark, DNA methylation has been widely investigated to deepen our understanding of its mechanism and correlation with human illness, and it is possible to analyze methylation at all levels with the massive data generated by high-throughput detection technology. However, how to effectively identify DML from high-throughput methylation data is still a tough challenge even if feature selection methods have been extensively explored in the context of gene expression data. Innovatively, combining the instance perturbation and function diversity, the newly proposed method HyDML achieved effective identification of DML, and this demonstrated that ensemble

feature selection could be used in dimension reduction for large-scale biological data. This will not only facilitate future early diagnosis of cancers based on the DNA methylation signatures but also enable additional investigations into the utilization of feature selection on other biomarker analysis domains. In the future, we will continue to study in depth the application of machine learning in biomarker identification and achieve better selection and prediction effect by combining more related information.

CONCLUSION

In this article, a hybrid ensemble approach is proposed by incorporating instance perturbation and multiple functions to identify differential methylation sites across 13 cancers from TCGA. The specially designed framework makes it possible to select robust differential methylation sites, which not only improves the accuracy of the classifier built by the selected sites, but also enhances the confidence of domain experts to implement biological validations. Further intensive analysis reveals that different cancer types have common methylation patterns, and part of the differential methylation sites shared in pan-cancers is of great potential to be crucial in the early diagnosis of cancers. All findings demonstrate that abnormal DNA methylation could be regarded as a marker that expresses the difference between the normal and the sick.

DATA AVAILABILITY

The data sets and materials for this study can be found in the following links:

HM 450 methylation data: https://tcga-data.nci.nih.gov/tcga/

Independent test sets:

1) For BRCA: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE52635
2) For LIHC: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE54503
3) For LUAD: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE66836

Source codes of HyDML, DML files, and single-nucleotide polymorphism files have been provided as an open source available at https://github.com/TQBio/HyDML.git.

AUTHOR CONTRIBUTIONS

QT, JZ, and SF conceived and designed the experiments. QT, ZY, YF, JT, YS, and SF performed the analysis and edited the manuscript. JZ and SF led the research and reviewed the manuscript. All authors read and approved the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (no. 61503061 and no. 61872063) and the Fundamental Research Funds for the Central Universities (no. ZYGX2016J102).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00774/full#supplementary-material

FIGURE S1 | (A) The distribution density of DML in 22 pairs of autosomal chromosomes in 13 cancers. **(B)** The distribution of DML in different functional regions in 13 cancers. **(C)** The distribution of DML in CpG island and non-CpG island in 13 cancers.

FIGURE S2 | Unsupervised hierarchical clustering of mutual DML in 13 cancers using similarity measure with Jaccard distance.

FIGURE S3 | The AUC changed when the number of selected loci gradually reduced in each cancer. All the results show that HyDML performed better than single-feature selection methods as it can select more robust loci for distinguish normal and tumor samples.

FIGURE S4 | The t-SNE clustering results with the loci that were uniquely selected by the three methods, HyDML, FastDMA, and RnBeads. Each row represents the loci from the corresponding cancer type, and each column represents the result of corresponding method.

REFERENCES

Abeel, T., Helleputte, T., Van de Peer, Y., Dupont, P., and Saeys, Y. (2010). Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. *Bioinformatics* 26 (3), 392–398. doi: 10.1093/bioinformatics/btp630

Andrilenas, K. K., Ramlall, V., Kurland, J., Leung, B., Harbaugh, A. G., and Siggers, T. (2018). DNA-binding landscape of IRF3, IRF5 and IRF7 dimers: implications for dimer-specific gene regulation. *Nucleic Acids Res*. 46 (5), 2509–2520. doi: 10.1093/nar/gky002

Aran, D., and Hellman, A. (2013). DNA methylation of transcriptional enhancers and cancer predisposition. *Cell* 154 (1), 11–13. doi: 10.1016/j.cell.2013.06.018

Arribas, J., Gimenez, E., Marcos, R., and Velazquez, A. (2015). Novel antiapoptotic effect of TBX15: overexpression of TBX15 reduces apoptosis in cancer cells. *Apoptosis* 20 (10), 1338–1346. doi: 10.1007/s10495-015-1155-8

Assenov, Y., Muller, F., Lutsik, P., Walter, J., Lengauer, T., and Bock, C. (2014). Comprehensive analysis of DNA methylation data with RnBeads. *Nat. Methods* 11 (11), 1138–1140. doi: 10.1038/nmeth.3115

Awada, W., Khoshgoftaar, T. M., Dittman, D., Wald, R., and Napolitano, A. (2012). A review of the stability of feature selection techniques for bioinformatics data. *2012 IEEE 13th International Conference on Information Reuse and Integration (IRI)*, 356–363. doi: 10.1109/IRI.2012.6303031

Baylin, S. B., and Ohm, J. E. (2006). Epigenetic gene silencing in cancer—a mechanism for early oncogenic pathway addiction? *Nat. Rev. Cancer* 6 (2), 107–116. doi: 10.1038/nrc1799

Benowitz, N. L. (2009). Pharmacology of nicotine: addiction, smoking-induced disease, and therapeutics. *Annu. Rev. Pharmacol*. 49, 57–71. doi: 10.1146/annurev.pharmtox.48.113006.094742

Bolon-Canedo, V., Sanchez-Marono, N., Alonso-Betanzos, A., Benitez, J. M., and Herrera, F. (2014). A review of microarray datasets and applied feature selection methods. *Inform. Sci.* 282, 111–135. doi: 10.1016/j.ins.2014.05.042

Brennan, J. A., Boyle, J. O., Koch, W. M., Goodman, S. N., Hruban, R. H., and Eby, Y. J. (1995). Association between cigarette-smoking and mutation of the P53 gene in squamous-cell carcinoma of the head and neck. *New Engl. J. Med*. 332 (11), 712–717. doi: 10.1056/NEJM199503163321104

Campbell, J. D., Alexandrov, A., Kim, J., Wala, J., Berger, A. H., and Pedamallu, C. S. (2016). Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat. Genet*. 48 (6), 607–60+. doi: 10.1038/ng.3564

Carvalho, R. H., Hou, J., Haberle, V., Aerts, J., Grosveld, F., and Lenhard, B. (2013). Genomewide DNA methylation analysis identifies novel methylated genes in non–small-cell lung carcinomas. *J. Thorac. Oncol*. 8 (5), 562–573. doi: 10.1097/JTO.0b013e3182863ed2

Chan, A. O. O., Broaddus, R. R., Houlihan, P. S., Issa, J. P. J., Hamilton, S. R., and Rashid, A. (2002). CpG island methylation in aberrant crypt foci of the colorectum. *Am. J. Pathol*. 160 (5), 1823–1830. doi: 10.1016/S0002-9440(10)61128-5

Chapman, J. R., Webster, A. C., and Wong, G. (2013). Cancer in the transplant recipient. *Csh Perspect. Med*. 3 (7). doi: 10.1101/cshperspect.a015677

Chelbi, S. T. Doridot, L., Mondon, F ., Dussour, C ., Rebourcet, R., and Vaiman, D (2011). Combination of promoter hypomethylation and PDX1 overexpression leads to TBX15 decrease in vascular IUGR placentas. *Epigenetics* 6, 2, 247–255. doi: 10.4161/epi.6.2.13791

Chiappinelli, K. B., Strissel, P. L., Desrichard, A., Li, H. L., Henke, C., and Akman, B. (2015). Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell* 162 (5), 974–986. doi: 10.1016/j.cell.2015.07.011

Cokus, S. J., Feng, S. H., Zhang, X. Y., Chen, Z. G., Merriman, B., and Haudenschild, C. D. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452 (7184), 215–219. doi: 10.1038/nature06745

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn* 20 (3), 273–297. doi: 10.1007/BF00994018

Costello, J. F., Fruhwald, M. C., Smiraglia, D. J., Rush, L. J., Robertson, G. P., and Gao, X. (2000). Aberrant CpG-island methylation has non-random and tumour-type–specific patterns. *Nat. Genet*. 24 (2), 132–138. doi: 10.1038/72785

Curtis, R. E., Metayer, C., Rizzo, J. D., Socie, G., Sobocinski, K. A., and Flowers, M. E. D. (2005). Impact of chronic GVHD therapy on the development of squamous-cell cancers after hematopoietic stem-cell transplantation: an international case-control study. *Blood* 105 (10), 3802–3811. doi: 10.1182/blood-2004-09-3411

Dawson, M. A., and Kouzarides, T. (2012). Cancer epigenetics: from mechanism to therapy. *Cell* 150 (1), 12–27. doi: 10.1016/j.cell.2012.06.013

Dedeurwaerder, S., Defrance, M., Calonne, E., Denis, H., Sotiriou, C., and Fuks, F. (2011). Evaluation of the infinium methylation 450K technology. *Epigenomics U.K.* 3 (6), 771–784. doi: 10.2217/epi.11.105

Deng, D. J., Liu, Z. J., and Du, Y. T. (2010). Epigenetic alterations as cancer diagnostic, prognostic, and predictive biomarkers. *Adv. Genet*. 71, 125–176. doi: 10.1016/B978-0-12-380864-6.00005-5

Down, T. A., Rakyan, V. K., Turner, D. J., Flicek, P., Li, H., and Kulesha, E. (2008). A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat. Biotechnol*. 26 (7), 779–785. doi: 10.1038/nbt1414

Draminski, M., and Koronacki, J. (2018). rmcfs: an R package for Monte Carlo feature selection and interdependency discovery. *J. Stat. Softw*. 85 (12). doi: 10.18637/jss.v085.i12

Esteller, M. (2007). Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat. Rev. Genet*. 8 (4), 286–298. doi: 10.1038/nrg2005

Fajardo, A. M., Piazza, G. A., and Tinsley, H. N. (2014). The role of cyclic nucleotide signaling pathways in cancer: targets for prevention and treatment. *Cancers* 6 (1), 436–458. doi: 10.3390/cancers6010436

Feng, H., Conneely, K. N., and Wu, H. (2014). A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Res*. 42 (8). doi: 10.1093/nar/gku154

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw*. 33 (1), 1–22. doi: 10.18637/jss.v033.i01

Gozzi, G., Chelbi, S. T., Manni, P., Alberti, L., Fonda, S., and Saponaro, S. (2016). Promoter methylation and downregulated expression of the TBX15 gene in ovarian carcinoma. *Oncol. Lett*. 12 (4), 2811–2819. doi: 10.3892/ol.2016.5019

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell* 144 (5), 646–674. doi: 10.1016/j.cell.2011.02.013

Haury, A. C., Gestraud, P., and Vert, J. P. (2011). The influence of feature selection methods on accuracy, stability and interpretability of molecular signatures. *PLoS One* 6 (12). doi: 10.1371/journal.pone.0028210

Hebestreit, K., Dugas, M., and Klein, H. U. (2013). Detection of significantly differentially methylated regions in targeted bisulfite sequencing data. *Bioinformatics* 29 (13), 1647–1653. doi: 10.1093/bioinformatics/btt263

Johnson, F. M., Saigal, B., Talpaz, M., and Donato, N. J. (2005). Dasatinib (BMS-354825) tyrosine kinase inhibitor suppresses invasion and induces cell cycle arrest and apoptosis of head and neck squamous cell carcinoma and non–small cell lung cancer cells. *Clin. Cancer Res*. 11 (19), 6924–6932. doi: 10.1158/1078-0432.CCR-05-0757

Johnson, W. E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8 (1), 118–127. doi: 10.1093/biostatistics/kxj037

Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet*. 13 (7), 484–492. doi: 10.1038/nrg3230

Jones, P. A., and Baylin, S. B. (2002). The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet*. 3 (6), 415–428. doi: 10.1038/nrg816

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 28 (1), 27–30. doi: 10.1093/nar/28.1.27

Kim, S. Y. (2009). Effects of sample size on robustness and prediction accuracy of a prognostic gene signature. *BMC Bioinf*. 10. doi: 10.1186/1471-2105-10-147

Klutstein, M., Nejman, D., Greenfield, R., and Cedar, H. (2016). DNA methylation in cancer and aging. *Cancer Res*. 76 (12), 3446–3450. doi: 10.1158/0008-5472.CAN-15-3278

Kron, K., Liu, L. Y., Trudel, D., Pethe, V., Trachtenberg, J., and Fleshner, N. (2012). Correlation of ERG expression and DNA methylation biomarkers with adverse clinicopathologic features of prostate cancer. *Clin. Cancer Res*. 18 (10), 2896–2904. doi: 10.1158/1078-0432.CCR-11-2901

Laird, P. W. (2010). Principles and challenges of genome-wide DNA methylation analysis. *Nat. Rev. Genet*. 11 (3), 191–203. doi: 10.1038/nrg2732

Leone, P., Shin, E. C., Perosa, F., Vacca, A., Dammacco, F., and Racanelli, V. (2013). MHC class I antigen processing and presenting machinery: organization, function, and defects in tumor cells. *JNCI-J. Natl. Cancer* I. 105 (16), 1172–1187. doi: 10.1093/jnci/djt184

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., and Tonti-Filippini, J. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462 (7271), 315–322. doi: 10.1038/nature08514

Liu, H. W., Liu, L., and Zhang, H. J. (2010). Ensemble gene selection by grouping for microarray data classification. *J. Biomed. Inform*. 43 (1), 81–87. doi: 10.1016/j.jbi.2009.08.010

Maksimovic, J., Gordon, L., and Oshlack, A. (2012). SWAN: subset-quantile within array normalization for illumina infinium humanmethylation450 beadchips. *Genome Biol*. 13 (6). doi: 10.1186/gb-2012-13-6-r44

McLeay, R. C., and Bailey, T. L. (2010). Motif enrichment analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinf*. 11. doi: 10.1186/1471-2105-11-165

Okegawa, T., Pong, R. C., Li, Y. M., and Hsieh, J. T. (2004). The role of cell adhesion molecule in cancer progression and its application in cancer therapy. *Acta Biochim. Pol*. 51 (2), 445–457.

Park, Y., Figueroa, M. E., Rozek, L. S., and Sartor, M. A. (2014). MethylSig: a whole genome DNA methylation analysis pipeline. *Bioinformatics* 30 (17), 2414–2422. doi: 10.1093/bioinformatics/btu339

Piliszek, R., Mnich, K., Tabaszewski, P., Migacz, S., Sułecki, A., and Rudnicki, W. R. (2018).Functions for MultiDimensional Feature Selection (MDFS): calculating multidimensional information gains, scoring variables, finding important variables, plotting selection results. This package includes an optional CUDA implementation that speeds up information gain calculation using NVIDIA GPGPUs. MDFS: MultiDimensional Feature Selection. https://cran.r-project.org/package=MDFS.

Robertson, K. D. (2005). DNA methylation and human disease. *Nat. Rev. Genet*. 6 (8), 597–610. doi: 10.1038/nrg1655

Saeys, Y., Abeel, T., and de Peer, Y. V. (2008). Robust feature selection using ensemble feature selection techniques. *Lect. Notes Artif. Int*. 5212, 313–31+. doi: 10.1007/978-3-540-87481-2_21

Shlomai, G., Neel, B., LeRoith, D., and Gallagher, E. J. (2016). Type 2 diabetes mellitus and cancer: the role of pharmacotherapy. *J. Clin. Oncol*. 34 (35), 4261–426+. doi: 10.1200/JCO.2016.67.4044

Soozangar, N., Sadeghi, M. R., Jeddi, F., Somi, M. H., Shirmohamadi, M., and Samadi, N. (2018). Comparison of genome-wide analysis techniques to DNA

methylation analysis in human cancer. *J. Cell. Physiol.* 233 (5), 3968–3981. doi: 10.1002/jcp.26176

Suzuki, M. M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* 9 (6), 465–476. doi: 10.1038/nrg2341

Tolstorukov, M. Y., Sansam, C. G., Lu, P., Koellhoffer, E. C., Helming, K. C., and Alver, B. H. (2013). Swi/Snf chromatin remodeling/tumor suppressor complex establishes nucleosome occupancy at target promoters. *P. Natl. Acad. Sci. U.S.A.* 110 (25), 10165–10170. doi: 10.1073/pnas.1302209110

Tost, J. J., D.m.r.t. (2007). "Analysis of DNA Methylation Patterns for the Early Diagnosis," in *Classification and Therapy of Human Cancers* (NY, USA: Nova Science Publishers), 87–133.

Turken, O., Demirbas, S., Onde, M. E., Sayan, O., and Kandemir, E. G. (2003). Breast cancer in association with thyroid disorders. *Breast Cancer Res.* 5 (5), R110–R113. doi: 10.1186/bcr609

van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. J. *Mach. Learn Res.* 9, 2579–2605.

Wang, C. M., Wang, Q. L., Xu, X. Q., Xie, B., Zhao, Y., and Li, N. (2017). The methyltransferase NSD3 promotes antiviral innate immunity via direct lysine methylation of IRF3. *J. Exp. Med.* 214 (12), 3597–3610. doi: 10.1084/jem.20170856

Wang, D., Yan, L., Hu, Q., Sucheston, L. E., Higgins, M. J., and Ambrosone, C. B. (2012). IMA: an R package for high-throughput analysis of Illumina's 450K Infinium methylation data. *Bioinformatics* 28 (5), 729–730. doi: 10.1093/bioinformatics/bts013

Wiemer, E. A. C. (2007). The role of microRNAs in cancer: no small matter. *Eur. J. Cancer* 43 (10), 1529–1544. doi: 10.1016/j.ejca.2007.04.002

Wu, D. G., Gu, J., and Zhang, M. Q. (2013). FastDMA: an infinium humanmethylation450 beadchip analyzer. *Plos One* 8 (9). doi: 10.1371/journal.pone.0074275

Yang, P. Y., Ho, J. W. K., Yang, Y. H., and Zhou, B. B. (2011). Gene-gene interaction filtering with ensemble of filters. *BMC Bioinf.* 12. doi: 10.1186/1471-2105-12-S1-S10

Yang, P. Y., Yang, Y. H., Zhou, B. B., and Zomaya, A. Y. (2010). A review of ensemble methods in bioinformatics. *Curr. Bioinf.* 5 (4), 296–308. doi: 10.2174/157489310794072508

Yu, L., Han, Y., and Berens, M. E. (2012). Stable gene selection from microarray data via sample weighting. *IEEE ACM T. Comput. Bi.* 9 (1), 262–272. doi: 10.1109/TCBB.2011.47

Zhang, Y. L., Wang, R. C., Cheng, K., Ring, B. Z., and Su, L. (2017). Roles of Rap1 signaling in tumor cell migration and invasion. *Cancer Biol. Med.* 14 (1), 90–99. doi: 10.20892/j.issn.2095-3941.2016.0086

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Tian, Zou, Fang, Yu, Tang, Song and Fan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 05 September 2019
doi: 10.3389/fgene.2019.00770

Maternal Smoking During Pregnancy Induces Persistent Epigenetic Changes Into Adolescence, Independent of Postnatal Smoke Exposure and Is Associated With Cardiometabolic Risk

*Sebastian Rauschert[1], Phillip E. Melton[2,3], Graham Burdge[4], Jeffrey M. Craig[5,6], Keith M. Godfrey[4,7], Joanna D. Holbrook[4], Karen Lillycrop[8], Trevor A. Mori[9], Lawrence J. Beilin[9], Wendy H. Oddy[10], Craig Pennell[11] and Rae-Chi Huang[1]**

[1] Telethon Kids Institute, University of Western Australia, Perth, WA, Australia, [2] Centre for Genetic Origins of Health and Disease, The University of Western Australia and Curtin University, Perth, WA, Australia, [3] School of Pharmacy and Biomedical Sciences, Curtin University, Bentley, WA, Australia, [4] Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, United Kingdom, [5] Early Life Epigenetics Group, MCRI, Royal Children's Hospital, Flemington Road, Parkville, VIC, Australia, [6] Centre for Molecular and Medical Research, School of Medicine, Deakin University, Geelong, VIC, Australia, [7] MRC Lifecourse Epidemiology Unit and NIHR Southampton Biomedical Research Centre, University of Southampton and University Hospital Southampton NHS Foundation Trust, Southampton, United Kingdom, [8] Centre for Biological Sciences, Faculty of Natural and Environmental Sciences, University of Southampton, Southampton, United Kingdom, [9] Medical School, Royal Perth Hospital Unit, University of Western Australia, Perth, WA, Australia, [10] Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS, Australia, [11] University of Newcastle, Newcastle, NSW, Australia

OPEN ACCESS

Edited by:
Yun Liu,
Fudan University, China

Reviewed by:
Zhiqing Huang,
Duke University, United States
Jorg Tost,
Commissariat à l'Energie Atomique et aux Energies Alternatives, France

***Correspondence:**
Rae-Chi Huang
Rae-Chi.Huang@telethonkids.org.au

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *26 April 2019*
Accepted: *22 July 2019*
Published: *05 September 2019*

Citation:
Rauschert S, Melton PE, Burdge G, Craig JM, Godfrey KM, Holbrook JD, Lillycrop K, Mori TA, Beilin LJ, Oddy WH, Pennell C and Huang R-C (2019) Maternal Smoking During Pregnancy Induces Persistent Epigenetic Changes Into Adolescence, Independent of Postnatal Smoke Exposure and Is Associated With Cardiometabolic Risk. Front. Genet. 10:770. doi: 10.3389/fgene.2019.00770

Background: Several studies have shown effects of current and maternal smoking during pregnancy on DNA methylation of CpG sites in newborns and later in life. Here, we hypothesized that there are long-term and persistent epigenetic effects following maternal smoking during pregnancy on adolescent offspring DNA methylation, independent of paternal and postnatal smoke exposure. Furthermore, we explored the association between DNA methylation and cardiometabolic risk factors at 17 years of age.

Materials and Methods: DNA methylation was measured using the Illumina HumanMethylation450K BeadChip in whole blood from 995 participants attending the 17-year follow-up of the Raine Study. Linear mixed effects models were used to identify differential methylated CpGs, adjusting for parental smoking during pregnancy, and paternal, passive, and adolescent smoke exposure. Additional models examined the association between DNA methylation and paternal, adolescent, and passive smoking over the life course. Offspring CpGs identified were analyzed against cardiometabolic risk factors (blood pressure, triacylglycerols (TG), high-density lipoproteins cholesterol (HDL-C), and body mass index).

Results: We identified 23 CpGs (genome-wide p level: 1.06×10^{-7}) that were associated with maternal smoking during pregnancy, including associated genes *AHRR* (cancer development), *FTO* (obesity), *CNTNAP2* (developmental processes), *CYP1A1*

(detoxification), *MYO1G* (cell signalling), and *FRMD4A* (nicotine dependence). A sensitivity analysis showed a dose-dependent relationship between maternal smoking and offspring methylation. These results changed little following adjustment for paternal, passive, or offspring smoking, and there were no CpGs identified that associated with these variables. Two of the 23 identified CpGs [cg00253568 (*FTO*) and cg00213123 (*CYP1A1*)] were associated with either TG (male and female), diastolic blood pressure (female only), or HDL-C (male only), after Bonferroni correction.

Discussion: This study demonstrates a critical timing of cigarette smoke exposure over the life course for establishing persistent changes in DNA methylation into adolescence in a dose-dependent manner. There were significant associations between offspring CpG methylation and adolescent cardiovascular risk factors, namely, TG, HDL-C, and diastolic blood pressure. Future studies on current smoking habits and DNA methylation should consider the importance of maternal smoking during pregnancy and explore how the persistent DNA methylation effects of *in utero* smoke exposure increase cardiometabolic risk.

Keywords: DNA methylation, maternal smoking during pregnancy, epigenetics, Raine Study, cardiometabolic health, adolescence

INTRODUCTION

Maternal smoking during pregnancy is associated with an increased risk in the offspring of chronic diseases including asthma, certain cancers, and cardiovascular disease in adulthood (Wakschlag et al., 2002; Hofhuis et al., 2003; DiFranza et al., 2004; Oken et al., 2005; Agrawal et al., 2010; Bhattacharya et al., 2014). Together with the association of antenatal exposure to maternal smoking with DNA methylation, this highlights the importance of the early environment on the development of diseases (Gillman, 2005; Hanson and Gluckman, 2008; Schulz, 2010; Nielsen et al., 2016). Furthermore, associations with maternal smoking during pregnancy have been associated with differential methylation of cytosine–phosphate–guanine (CpG) base pairs in newborns (Joubert et al., 2016), children (Rzehak et al., 2016), young adults (Lee et al., 2015), and middle aged adults (Sun et al., 2013).

An epigenome-wide DNA methylation meta-analysis by Joubert et al. with combined sample size of 6,685 newborns and 3,187 older children previously identified 2,965 methylated CpGs in the offspring that associated with maternal smoking during pregnancy (Joubert et al., 2016). Methylation levels at CpGs most strongly associated with maternal smoking were contained within genes also implicated in other studies including *MYO1G (myosin 1G)*, *CYP1A1 (cytochrome P450 family 1 subfamily A member 1)*, *GFI1 (growth factor independent 1 transcriptional repressor)*, *CNTNAP2 (contactin-associated protein-like 2)* (Rotroff et al., 2016; Rzehak et al., 2016; Tehranifar et al., 2018), and xenobiotics (*AHRR, aryl-hydrocarbon receptor repressor*) (Finkelstein and Jeong, 2017). These genes have been associated with tumorigenesis and metastasis (in the case of *MYO1G*, *CTNAP2*, *GFI1*, and *AHRR*), activation of compounds with carcinogenic properties (in the case of *CYP1A1* and *AHRR*), and autism (*CNTNAP2*), as well as mediating the effect of maternal smoking and birthweight (in the case of *GFI1*) (Finkelstein, 2017), thereby suggesting a possible epigenetic mechanism linking exposure to smoking during pregnancy with adverse outcomes such as obesity or cancer risk in the offspring.

Several studies have focused on the effect of current smoking on CpG methylation in adults and conclude that it strongly affects DNA methylation within the genes of *AHRR*, *GFI1*, and *MYO1G* and mediates risk of disease (Su et al., 2016; Bojesen et al., 2017; Wilson et al., 2017; Li et al., 2018). However, the studies did not account for the influence of maternal smoking during pregnancy (Wilson et al., 2017; Li et al., 2018). The overlap of CpGs sites that associate with both current (adolescent or adult) smoking and maternal smoking during pregnancy indicates the need for caution in attributing causation to postnatal smoke exposure (Lee et al., 2015; Joubert et al., 2016; Rzehak et al., 2016). Richmond et al. (2015) for example showed associations between maternal smoking during pregnancy and CpG methylation measured at three different timepoints. *In utero* smoke exposure was associated with CpGs within *AHRR*, *MYO1G*, *CYP1A1*, and *CNTNAP2* independent of current smoking of the adolescents. Furthermore, a recent study in 40-year-old women showed associations between CpG methylation levels in *FTO (fat mass and obesity-associated protein)*, *CYP1A1*, *MYO1G*, *AHRR*, *ANPEP* (*alanyl aminopeptidase, membrane)*, *ZNF536 (zinc finger protein 536)*, and *GFI1* and a history of exposure to maternal smoking *in utero*, which was independent of their own smoking status (Tehranifar et al., 2018). Socioeconomic status is potentially an important confounder in the association between CpG methylation and offspring smoking. Low socioeconomic status has, for example, been associated with both smoking in general (Rosemary et al., 2012) and offspring smoking (Gilman et al., 2003) and DNA methylation levels (McDade et al., 2019).

A recent study analyzed the association between eight CpGs, located in the *GFI1* gene region and cardiovascular health (Parmar et al., 2018). The authors found three out of the eight CpGs to be associated with maternal prenatal smoking and the remaining five to be associated with adolescents own smoking. They found the strongest associations between some of the CpGs with BMI, waist circumference, blood pressure, and triacylglycerol (TG), with the most consistent associations between CpGs and TG. This highlights the potential for maternal smoking to induce long lasting changes in association with both DNA methylation and cardiometabolic health.

The present study aimed to investigate whether there is an association between maternal smoking during pregnancy and DNA methylation in the offspring at 17 years of age and if methylation levels at the differentially methylated CpGs were associated with cardiometabolic risk factors, using data from the second generation (Gen2) of the Raine Study (www.rainestudy.org.au). We further determined if the relationships between methylation levels at these particular CpGs and maternal smoking were independent of paternal smoking, passive smoke exposure during childhood, and adolescent self-reported smoking. We hypothesized that there are long-term and persistent postnatal epigenetic effects following maternal smoking during pregnancy on adolescent offspring DNA methylation, relatively unaffected by smoke exposure from other sources. Furthermore, we hypothesize that these changes are associated with an adverse effect on cardiometabolic health.

METHODS

Study

The study design and initial characteristics of the Raine Study have been previously described (Newnham et al., 1993). From 1989 to 1999, a total of 2,900 pregnant women were enrolled to take part in this longitudinal cohort study. Recruitment took part at King Edward Memorial Hospital and surrounding private hospitals. The 2,868 live births (Gen2) have been followed up at 1, 2, 3, 5, 8, 10, 14, and 17 years during which anthropometric (e.g., height, weight, skinfolds), clinical, and biochemical data have been collected.

Ethics approval for conducting the epigenetic analysis at the 17-year follow-up was given by the Human Ethics Committee of the University of Western Australia. Informed and written consent was provided by the participants and their parents or carer.

The present analyses included 790 participants that had data for the variables of interest, being maternal educational level, family income, gestational weight gain, gestational age, maternal age, maternal prepregnancy BMI, birthweight, age, Caucasian ethnicity, and sex of the child. In separate analyses, we examined the potential confounding effects of offspring smoking ($n = 663$), passive smoking ($n = 513$), and paternal smoking ($n = 781$).

Smoking Variables

Maternal self-reported smoking during the 18th and 34th week of gestation and paternal smoking behavior (reported by the mother) at the 18th week of gestation were obtained by questionnaire. Smoking behavior of the adolescents at 17 years of age was self-reported in a confidential online questionnaire. Adolescents self-reported cigarette consumption over their lifetime (yes/no), in the past month (yes/no), and past 7 days (yes/no). Different smoking variables were derived for the analyses: Maternal smoking during pregnancy was coded as "never" versus "ever" smoking during pregnancy, based on the categorical variables for the number of cigarettes smoked daily at 18 and 34 weeks of gestation. Furthermore, maternal smoking at 18 and 34 weeks of gestation was analyzed in association with CpG in two separate statistical models, to ascertain if smoking during mid- or late-term pregnancy had a different effect on CpG methylation in the offspring. Smoking of the offspring was coded as a binary variable ("never" vs. "ever" smoked). Passive smoking exposure during childhood was defined by aggregating questionnaire data from the caregiver on smoking at eight intervening time points until 17 years (1, 2, 3, 5, 8, 10, 14, and 17) and coded as "never" versus "ever" exposed to passive smoking if the average number of cigarettes smoked over all time points was ≥1 (Le-Ha et al., 2013).

Cardiovascular and Anthropometric Variables

Height was measured using a stadiometer (Holtain, Crosswell, UK) to the nearest 0.1 cm. Weight was measured using a digital chair scale (Wedderburn, New South Wales, Australia) to the nearest 100 g. Body mass index (BMI) was calculated as weight (kilograms)/height (meters)^2. Waist circumference was measured to the nearest centimeter.

Venous blood samples were taken after an overnight fast. Serum insulin, glucose, high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), and triacylglycerols (TG) were measured in the PathWest Laboratory at Royal Perth Hospital as described (Huang et al., 2012). HOMA-IR (molar units) was calculated by insulin (mIU/L) × glucose (mmol/l)/22.5 (Matthews et al., 1985).

Laboratory Measures

DNA Methylation Profiling

Using whole-blood samples collected at age 17 years, epigenome-wide DNA methylation profiles for 1,192 (58 technical replicates) individuals were generated at the Centre for Molecular Medicine and Therapeutics, University of British Columbia using the Illumina Infinium HumanMethylation450 BeadChip array (Illumina San Diego, CA). Quality control was performed using the statistical software R and Bioconductor packages *shinyMethyl* (Fortin et al., 2014), *MethylAid* (Van Iterson et al., 2014), and *RnBeads* (Assenov et al., 2014).

Four participants with inconsistent results and identified as outliers ($n = 3$) or sex misclassification ($n = 1$) were removed. Sixty-five CpGs for which a common SNP disrupted the site leading to genotypic specific methylation levels, 11,648 sex chromosome CpGs and 10,777 CpGs with a detection $p > 0.05$ in any sample were removed. A further 160 probes with bead counts <3 in more than 5% of samples were removed. Batch effects persisted after beta-mixture quantile normalization (BMIQ) was applied (Teschendorff et al., 2013). Therefore, plate, slide, and

well number were included in all statistical models. As cellular heterogeneity can influence methylation profiles and drive some of the methylation differences detectable across individual blood samples, we adjusted for estimated cell counts using the Houseman estimating method (Houseman et al., 2014) as implemented in the R statistical package, *minfi* (Aryee et al., 2014) for six cell types (CD8T, CD4T, NK, B cell, monocytes, and granulocytes). Mapping of the CpG to the nearest gene was performed using the Illumina Infinium annotation genomic coordinates.

Genome-Wide Genotype Data

DNA was collected from blood samples from 74% of the adolescents who attended the 14-year follow-up and a further 5% who attended the 16-year follow-up, using standardized procedures. SNP data for this study were obtained from genome-wide genotype data as described previously (Jones et al., 2013). Briefly, genotyping was performed on the Illumina Human 660W Quad Array (Illumina, San Diego, California, USA), and exclusion criteria were low genotyping success (>3% missing), excessive heterozygosity, relation with another sample (identity by descent > 0.1875), ambiguous sex, and mislabeling. There were 1,494 individuals whose DNA samples passed the quality control criteria and were eligible for genetic analyses, and 965 of them had completed the AQ. Out of those, 753 samples overlapped with the 790 samples that had epigenetic information and nonmissing data in the covariates.

Statistical Analysis

All models were analyzed using R Version 3.4.3. A flow diagram of models and sample size is presented in **Figure 1**, showing the different analysis steps and the number of complete cases per analysis. Full results are presented in the **Supplemental Tables**.

Effect Modifier

Studies previously conducted to analyze the association between maternal smoking during pregnancy and offspring CpG methylation levels have controlled for a variety of effect modifiers with a high level of commonalities between studies. Potential maternal confounders included prepregnancy BMI and socioeconomic variables including family income and maternal education, gestational weight gain, maternal alcohol consumption during pregnancy, and maternal age. Offspring variables included birthweight, sex, and age. Analyses were adjusted for cell count using the reference-based Houseman approach (Sun et al., 2013), for batch effect and measurement derived variability utilizing linear mixed effect models with plate number as the random effect and the aforementioned variables. Hypothesis testing for differences between the maternal smoking exposed and the nonexposed group in the variables used in this study were performed using chi-squared tests for categorical and *t*-tests and Wilcoxon tests for continuous variables.

Models

Identification of CpGs Whose Methylation Levels Associated With Maternal Smoking

We used a linear mixed-effects model as our main model. The outcome was percent methylation at one single CpG per model, and the predictor was maternal smoking, adjusted for exact age of the adolescent, sex, and maternal, and offspring confounders as described previously ($n = 790$). The same model was run twice, with the second application also adjusted for blood cell count

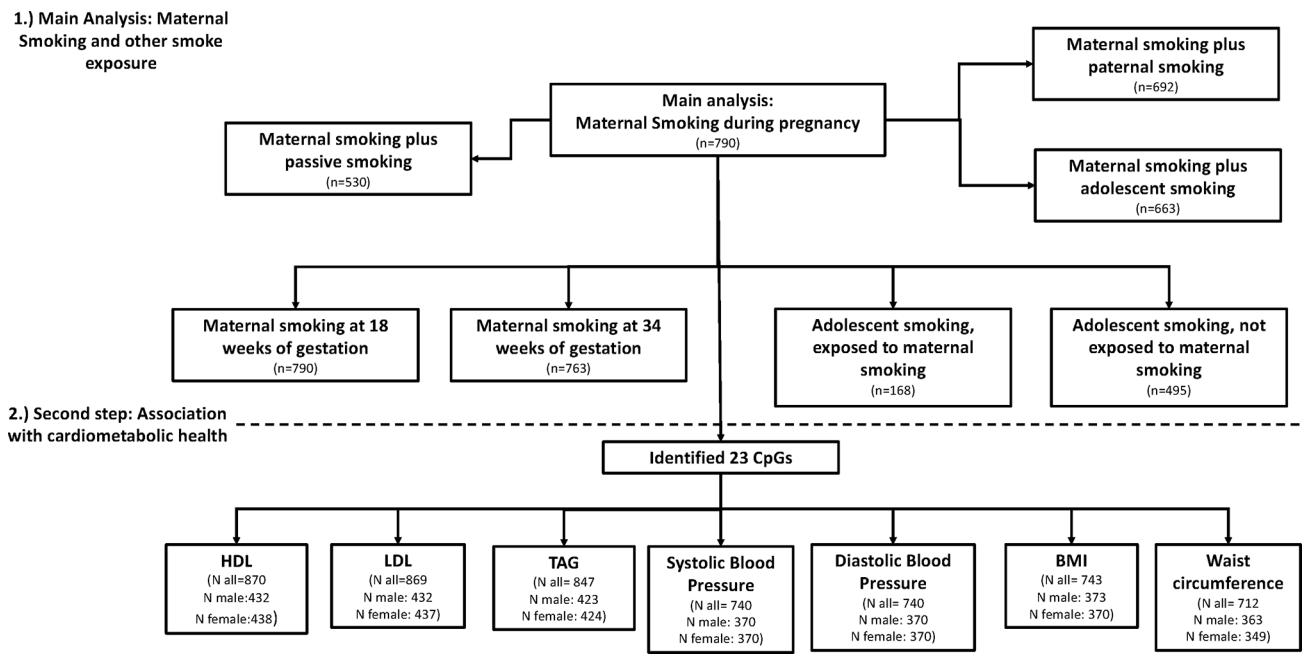


FIGURE 1 | Flow diagram of the analysis and sample size per model.

estimate (Houseman et al., 2014). This stepwise approach ensures that differences in the adjustment for cell count are accessible. To account for multiple testing, we utilized a conservative Bonferroni approach (genome-wide p level: 1.06×10^{-7}).

The main model with maternal smoking was then assessed for a sex interaction to examine if the effect of maternal smoking on the CpG methylation levels differed between male and female.

Additionally, we used the information on maternal average cigarette consumption during pregnancy (six categories: none, 1–5 daily; 6–10 daily; 11–15 daily; 16–20 daily; and 31 or more, with none as the reference group) to test for dose-dependent methylation in the identified sites.

Assessing the Effect of Paternal, Passive, and Adolescent Smoking on DNA Methylation

We included the covariates "ever smoking at 17 years of age" (yes/no) ($n = 663$), paternal smoking during pregnancy ($n = 781$) and passive smoking exposure during childhood ($n = 513$) (Le-Ha et al., 2013) to the main model predicting the CpGs associated with maternal smoking, to ascertain if they changed the effect size of maternal smoking. We also examined the effect of those variables in separate models as predictors.

For adolescent smokers, we split the data into those who were exposed to maternal smoking during pregnancy ($n = 168$) and those who were not exposed ($n = 495$) to explore any potential differences in effect sizes. The model was the same as described above, including confounders as described earlier.

Assessing the Genetic Effect on the Association Between Maternal Smoking and CpG Methylation

To assess if the association between maternal smoking and CpG methylation persists after taking SNPs into account, we utilized the *GEM* bioconductor package (Pan et al., 2016) to identify significant SNP–CpG associations in the Raine study. The *GEM* package is a computational efficient approach to identify methylation quantitative trait loci, perform DNA-methylation wide association studies, and assess the interaction of CpG methylation and SNPs on outcomes.

With *GEM*, those SNPs were identified, which were significantly associated with the CpGs associated with maternal smoking during pregnancy in this study. The identified SNPs were added to the respective main model with CpG as outcome and maternal smoking during pregnancy as predictor, adjusted for the aforementioned variables.

RESULTS

Characteristics of the Population

The characteristics of the participants are shown in **Table 1**. Of the 995 Caucasian participants, 30% were exposed to maternal smoking during pregnancy. Paternal, passive, and adolescent smoking rates were higher in the group exposed to maternal smoking. Of those exposed to maternal smoking during pregnancy, 60% had fathers who also smoked during the pregnancy period, 41% were exposed to passive smoking, and 30% reported smoking themselves. Of those not exposed to maternal smoking, 24% had fathers who smoked during the pregnancy period, 25% were exposed to passive smoking, and 40% reported smoking themselves.

TABLE 1 | Characteristics of the participants of the Raine Study. The p value refers to chi-square test results for categorical and t-test/Wilcoxon test for continuous variables.

	No maternal smoking	Maternal smoking	p value
N	**699**	**296**	
ADOLESCENT CHARACTERISTICS			
Sex of the child [*n*(%)]			0.17
Female	337 (48.21)	157 (53.04)	
Male	362 (51.79)	139 (46.96)	
Adolescent age			0.133
Mean	17.24	17.29	
SD	0.58	0.61	
Birthweight			2.09E–11
Mean	3,430.85	3,137.17	
SD	526.64	657.46	
Corrected gestational age (days)			0.00071
Mean	276.59	272.52	
SD	12.18	19.31	
Waist circumference (cm)			0.039
Mean	79.03	81.15	
SD	11.11	12.55	
Adolescent BMI (kg/m²)			0.001
Mean	22.8	24.12	
SD	4.16	5.19	
Adolescent HDL-C (mg/dl)			0.03337
Mean	1.3	1.27	
SD	0.29	0.3	
Adolescent LDL-C (mg/dl)			0.35
Mean	2.33	2.38	
SD	0.66	0.71	
Adolescent triacylglycerol (mg/dl)			0.012
Mean	1.04	1.15	
SD	0.5	0.67	
Adolescent systolic blood pressure (mmHg)			0.3247
Mean	114.99	115.69	
SD	11.37	10.73	
Adolescent diastolic blood pressure (mmHg)			0.6639
Mean	59.43	59.6	
SD	6.56	6.78	
Adolescent HOMA-IR			0.1503
Mean	2.04	2.11	
SD	2.85	1.82	
SMOKE EXPOSURE DURING OTHER PERIODS OF LIFE			
Paternal smoking during pregnancy [*n*(%)]			<2.2E–16
No	519 (74.25)	113 (38.18)	
Yes	174 (24.89)	178 (60.135)	
Missing	6 (0.86)	5 (1.69)	
Passive smoking [n(%)]			<2.2E–16
No	292 (41.77)	20 (6.76)	
Yes	134 (19.17)	122 (41.22)	
Missing	273 (39.06)	154 (52.03)	
Adolescent smoking			0.002541
Yes	381 (54.51)	119 (40.20)	
No	158 (22.60)	75 (25.34)	
Missing	160 (22.89)	102 (34.46)	

(*Continued*)

TABLE 1 | Continued

	No maternal smoking	Maternal smoking	*p* value
N	**699**	**296**	
MATERNAL AND FAMILIAL CHARACTERISTICS			
Maternal alcohol consumption during pregnancy [*n*(%)]			0.4107
no	354 (50.64)	147 (49.66)	
yes	344 (49.21)	149 (50.34)	
missing	1 (0.14)		
Maternal school level [*n*(%)]			1.65E−11
None	292 (41.77)	182 (61.49)	
Trade certificate or apprenticeship	55 (7.87)	27 (9.12)	
Professional registration (nondegree)	90 (12.88)	22 (7.43)	
College diploma or degree	129 (18.45)	40 (13.51)	
University degree	102 (14.59)	10 (3.38)	
Other	31 (4.43)	15 (5.07)	
Family income [*n*(%)]			2.91E-06
<24,000 AUD per annum	206 (29.47)	127 (42.91)	
>24,000 AUD per annum	475 (67.95)	148 (50)	
NA	18 (2.58)	21 (7.09)	
Maternal age (years)			1.62E−12
Mean	29.17	26.76	
SD	5.8	5.63	
Maternal prepregnancy BMI (kg/m²)			0.3591
Mean	22.66	22.25	
SD	4.45	4.51	
Maternal pregnancy weight gain ratio			0.9944
Mean	0.5	0.5	
SD	0.2	0.21	

Offspring exposed to maternal smoking during pregnancy included 49% from families in the lowest income bracket compared with 29% of those not exposed to maternal smoking. The group exposed to maternal smoking during pregnancy had significantly more mothers with a lower educational attainment ($p < 0.001$). Mothers who smoked during pregnancy were younger (26.76 ± 5.63 vs. 29.17 ± 5.8 years old) than those who did not smoke.

Those exposed to maternal smoking had significantly higher waist circumference ($p = 0.039$), BMI, and TG and lower HDL-C at 17 years of age compared to the nonexposed group.

For all smoking variables, those study participants of the original cohort who are not included in this study ($n = 1{,}317$) had higher numbers of smokers and smoke exposed individuals as well as lower socioeconomic status assessed by family income compared to the participants included in our study [$n = 995$ (**Supplement Table S1**)].

Epigenome-Wide DNA Methylation Analysis

Effects of Maternal Smoking on Offspring DNA Methylation

One identified CpG, namely, cg04224247 (*WWC3*), showed a bimodal distribution in the histogram, suggestive of a genotype driven rather than an environmental influence (Teh et al., 2014) and hence was excluded from further analysis.

Associations between any maternal smoking during pregnancy and methylation levels at individual CpG sites are shown in a forest plot (**Figure 2**). The analysis showed that inclusion or exclusion of cell count estimation based on the Houseman method did not change the number of CpGs whose methylation levels were associated with maternal smoking after Bonferroni correction (genome-wide p level: 1.06×10^{-7}, **Supplemental Tables S2**, **S3**, and **S4**). The smoking variables that combined data from 18 and 34 weeks did not differ in the direction of their association with CpGs in that methylation at the same CpGs was associated with maternal smoking at i) 18 weeks, ii) 34 weeks, and iii) combined 18 or 34 weeks (**Supplemental Tables S1**, **S5**, and **S6**).

The final model, including all confounding variables and batch number as random effect, showed methylation levels at 23 CpGs associated with maternal smoking during pregnancy after conservative Bonferroni correction; these 23 CpGs mapped to 10 genes (**Table 2**). Overall, seven CpGs (genes: *CNTNAP2*, *GFI1*, *WWC3*, *AHRR*, and *APOB*) showed hypomethylation in association with mothers who reported they were smoking during pregnancy, whereas 16 CpGs (associated genes: *CYP1A1/CYP1A2*, *MIR548T*, *AHRR*, *FRMD4A*, *FTO*, and *MYO1G*) were hypermethylated in those whose mothers smoked during pregnancy compared to offspring of nonsmokers. The highest percentage difference in DNA methylation between maternal smoking categories was 8.3% at cg12803068 (mapped to *MYO1G*). The remainder of the top 23 CpGs show methylation changes in association with *in utero* smoke exposure in the range of 1–6% (**Table 2**, coefficient times 100).

Inclusion or exclusion of cell count estimation in the model did not change either the number or direction of association CpGs falling under the Bonferroni threshold (genome-wide p level: 1.06×10^{-7}, **Supplemental Tables S2** and **S3**).

Dose Dependency With Maternal Smoking: Sensitivity Analysis

A sensitivity analysis for dose-dependent methylation with maternal smoking showed a significant trend towards hyper- or hypomethylation with an increasing number in cigarettes consumed (Wilcoxon significance test between groups $p < 0.05$). These results can be seen in the supplement (**Supplemental Figure 1**).

Effect of Other Sources of Smoke Exposure on Offspring DNA Methylation

Adjusting the main maternal smoking model for all additional smoking variables reduced the sample size due to missing values in the offspring self-reported smoking to 506 but did not change the direction of the estimates, as can be seen in the forest plot, comparing the models by beta coefficients (**Supplemental Figure 2** and **Supplemental Table S7**). Methylation levels at four CpGs (cg04180046, cg12803068, cg25949550, and cg05549655) were still significantly associated with maternal smoking (**Supplemental Table S6** and **S7**). These were also the most significant CpGs in the main model.

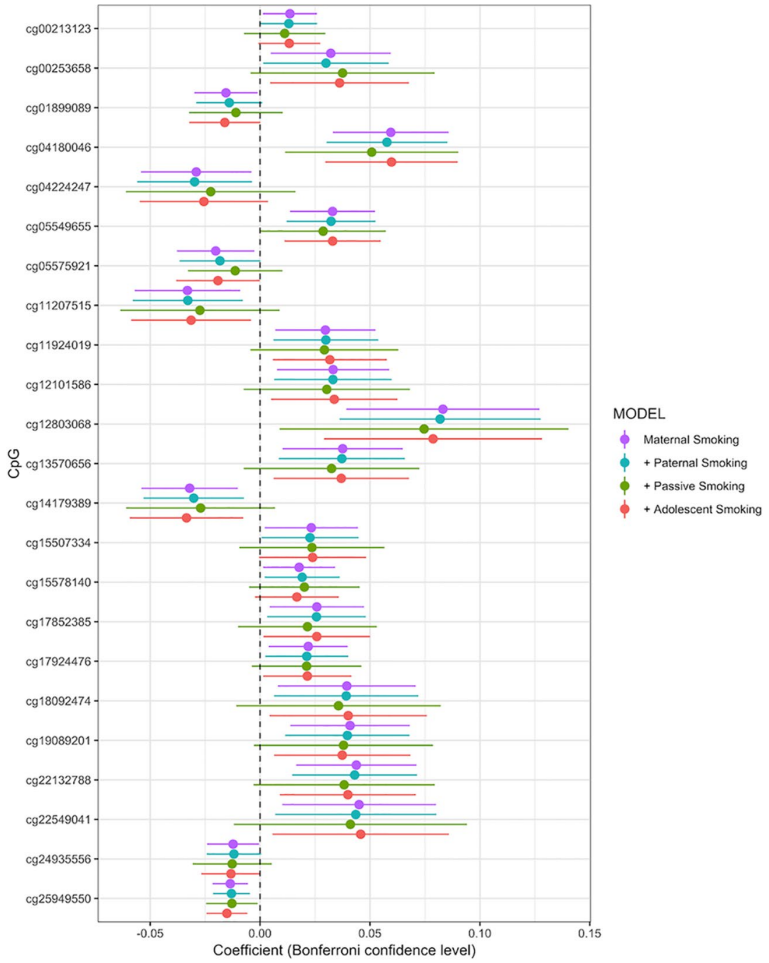


FIGURE 2 | Forrest plot for the epigenome-wide association study with maternal smoking as predictor and individual CpG sites as outcome, adjusted for sex, offspring age, age of the mother, birthweight, gestational weight gain, maternal alcohol consumption during pregnancy, maternal school level, maternal prepregnancy BMI, family income during pregnancy, cell count, and batch effects. Stratified by models without further adjustment (n = 790), adjustment for paternal smoking (n = 692), passive smoking (n = 530), and adolescent smoking (n = 663). *X-axis*: effect size from the linear mixed effects model and confidence interval; *Y-axis*: individual CpGs.

Stratified analysis for methylation levels at the identified 23 CpGs for male and female did not show significant effect size changes compared to the full model (**Supplemental Figure 3**). Furthermore, **Supplemental Figure 4** shows a barplot comparing the methylation change between those not exposed to any smoking (n = 381), those only smoking at 17 years (n = 158), those only exposed to maternal smoking during pregnancy (n = 119), and those exposed to both types of smoking (n = 75).

Effect of Paternal Smoking During Pregnancy

Paternal smoking (yes/no) during pregnancy (n = 692) was not significantly associated with methylation levels at any of the 23 CpGs detected in the maternal smoking EWAS (all genome-wide $p > 1.06 \times 10^{-7}$, **Supplemental Table S8**). Adding paternal smoking to the main model did not change the effect size for the effect of maternal smoking on CpG methylation levels. An EWAS of paternal smoke exposure did not detect any associations with CpG methylation levels at the genome-wide significance $p < 1.06 \times 10^{-7}$ (**Supplemental Table S8**).

Effect of Childhood Exposure to Smoking

No significant associations were detected for passive smoke exposure of the adolescent (n = 530) (Le-Ha et al., 2013) with DNA methylation at any CpG (all genome-wide $p > 1.06 \times 10^{-7}$) (**Table 2**, **Supplemental Table S9**). Adding passive childhood exposure to the main model did not change the effect size for the effect of maternal smoking on CpG methylation levels. An EWAS of passive smoke exposure did not show any associations with CpG methylation levels at the genome-wide significance $p < 1.06 \times 10^{-7}$ (**Supplemental Table S9**).

Effect of Adolescent Smoking

Adolescent reported smoking behavior (yes/no, n = 663) was not associated with the methylation level of any of the 23 CpGs detected in the maternal smoking EWAS (full results of the adolescent smoking EWAS: **Supplemental Table S10)**. The p values, effect sizes and standard errors are reported in **Table 2**. Adding active adolescent smoking to the main model did not change the effect size for the effect of maternal smoking on CpG methylation. An

TABLE 2 | CpGs associated with maternal smoking during pregnancy with UCSC gene annotation and model p-values, beta-coefficients and standard errors from the epigenome wide association study. The Bonferroni significance threshold is 1.06×10^{-7}.

Chromosome	ID	Function	Gene	Location	UCSC_CpG_Islands_Name	Relation_to_UCSC_CpG_Island	P.Value	Coefficient	Standard Error
7	cg04180046	intronic	MYO1G	Body	chr7:45002111-45002845	Island	8.93E-31	0.059451122	0.004899729
7	cg12803068	intronic	MYO1G	Body	chr7:45002111-45002845	S_Shore	1.04E-22	0.083162901	0.008174077
15	cg05549655	intergenic	CYP1A1,CYP1A2	TSS1500	chr15:75018186-75019336	Island	5.83E-19	0.032945838	0.003593891
7	cg25949550	intronic	CNTNAP2	Body	chr7:145813030-145814084	S_Shore	1.39E-18	–0.013562513	0.001496779
7	cg22132788	exonic	MYO1G	Body	chr7:45002111-45002845	Island	5.43E-17	0.043777558	0.005088082
7	cg19089201	UTR3	MYO1G	3'UTR	chr7:45002111-45002845	Island	2.76E-15	0.040922837	0.005057333
1	cg14179389	intronic	GFI1	Body	chr1:92945907-92952609	Island	1.53E-14	–0.032050625	0.004077831
7	cg11207515	intronic	CNTNAP2	Body			3.80E-13	–0.033065175	0.004462166
15	cg13570656	intergenic	CYP1A1,CYP1A2	TSS1500	chr15:75018186-75019336	Island	4.33E-13	0.037564996	0.005082323
15	cg11924019	intergenic	CYP1A1,CYP1A2	TSS1500	chr15:75018186-75019336	Island	6.07E-12	0.029699312	0.004240665
15	cg12101586	intergenic	CYP1A1,CYP1A2	TSS1500	chr15:75018186-75019336	Island	6.16E-12	0.033228312	0.004745949
15	cg22549041	intergenic	CYP1A1,CYP1A2	TSS1500	chr15:75018186-75019336	Island	1.02E-11	0.045041658	0.00650389
15	cg18092474	intergenic	CYP1A1,CYP1A2	TSS1500	chr15:75018186-75019336	Island	2.95E-11	0.03943739	0.005831697
5	cg17924476	intronic	AHRR	Body	chr5:320788-323010	S_Shore	1.16E-10	0.021888448	0.003342621
15	cg17852385	intergenic	CYP1A1,CYP1A2	TSS1500	chr15:75018186-75019336	Island	2.05E-10	0.025807923	0.003996831
16	cg00253658	intergenic	FTO, LOC100996338				4.50E-10	0.032150112	0.005079598
X	cg04224247	intronic	WWC3	5'UTR	chrX:9982513-9984583	Island	9.01E-10	–0.029045508	0.004673566
5	cg05575921	intronic	AHRR	Body	chr5:373842-374426	N_Shore	1.26E-09	–0.020194441	0.003278542
15	cg00213123	intergenic	CYP1A1,CYP1A2	TSS1500	chr15:75018186-75019336	Island	5.06E-09	0.013568664	0.002291017
10	cg15507334	upstream	FRMD4A	TSS200			5.57E-09	0.023297316	0.003944714
7	cg15578140	ncRNA_intronic	MIR548T	Body			8.41E-09	0.017756893	0.003043603
5	cg01899089	intronic	AHRR	Body	chr5:370185-370422	N_Shore	1.06E-08	–0.015519815	0.002679093
2	cg24935556	intergenic	APOB, LOC645949				3.37E-08	–0.012281317	0.002198374

EWAS of adolescent smoking did not detect any associations with CpG methylation levels at genome-wide significance $p < 1.06 \times 10^{-7}$ (**Supplemental Table S9**). We performed a stratified analysis for adolescents who were smoking, but were not exposed to maternal smoking during pregnancy, versus those who were exposed. Comparison of the beta coefficients suggested a stronger effect of maternal smoking than of adolescent smoking; for the group of not exposed adolescent smokers, the beta coefficients were smaller than the ones for those exposed to maternal smoking, but none of the CpGs were significantly associated with adolescent smoking (**Figure 3**, **Supplemental Tables S11** and **S12**).

Cardiometabolic Variables

Analyses that examined the methylation levels at the 23 CpGs associated with maternal smoking and cardiometabolic risk factors in the entire study population and separately for male and female, showed methylation levels at two CpGs (cg00253568 and cg00213123, located in the *FTO* and *CYP1A1* region) significantly associated with TG (cg00253568, full study population; coefficient, 1.97; standard error, 0.63; Bonferroni *p* value, 0.041), diastolic blood pressure (cg00253568, female subset; coefficient, 3.06; standard error, 0.91; Bonferroni *p* value, 0.021), and HDL-C (cg00213123, male subset; coefficient,

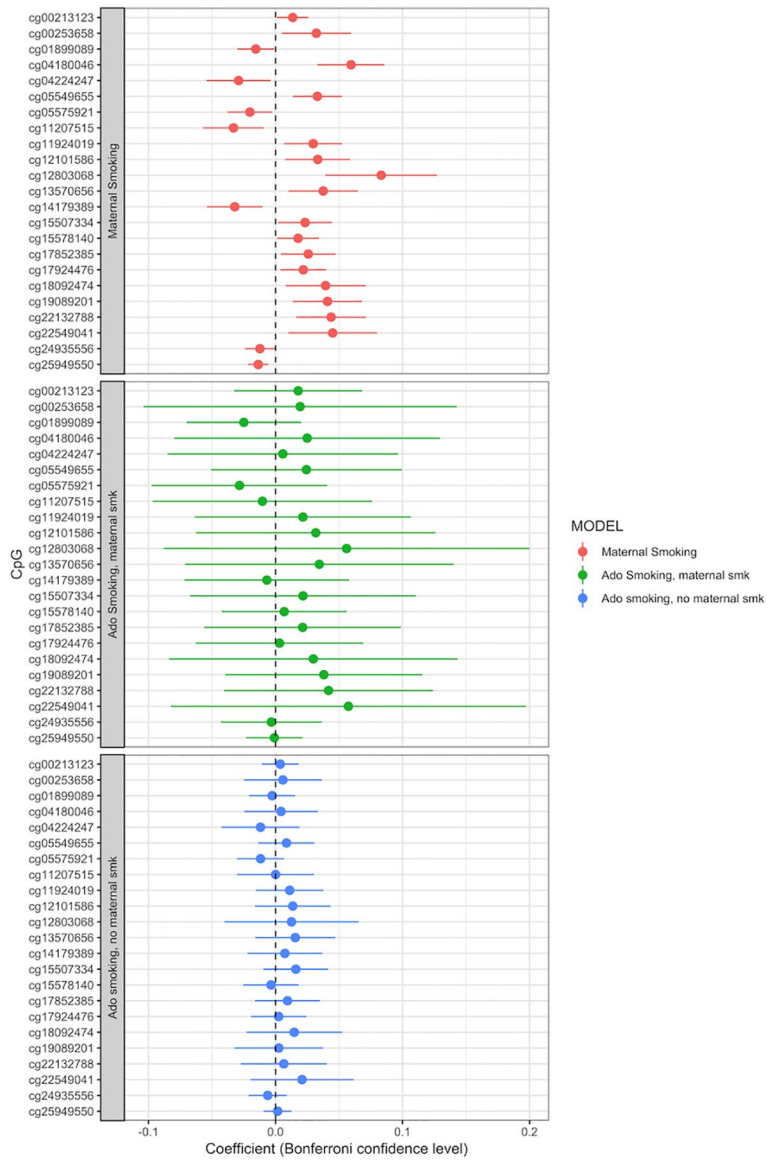


FIGURE 3 | Adolescent smoking, stratified by exposure to maternal smoking during pregnancy, and compared to the effect of maternal smoking during pregnancy (maternal smoking *n* = 790, adolescent smoking, maternal smoking *n* = 168, adolescent smoking, no maternal smoking *n* = 495). Models adjusted for sex, offspring age, age of the mother, birthweight, gestational weight gain, maternal alcohol consumption during pregnancy, maternal school level, maternal prepregnancy BMI, family income during pregnancy, cell count, and batch effects. *X-axis*: effect size from the linear mixed effects model and confidence interval; *Y-axis*: individual CpGs.

6.72; standard error, 2.04; Bonferroni p value, 0.025), whereas almost all of the 23 CpGs showed trends of either hyper- or hypomethylation in association with cardiometabolic variables, indicating a potentially long lasting effect of maternal smoking on cardiometabolic health of the offspring (Boxplots for cardiometabolic variables, stratified by exposure to maternal smoking: **Supplemental Table S5**).

Effect of SNPs on the Association Between CpG Methylation and Maternal Smoking

When adding the SNPs that were associated with the identified 23 CpGs in this study to the main model, the significant association between CpG methylation and maternal smoking persisted (**Supplemental Table S13**). Furthermore, the SNPs were not significantly associated with exposure to maternal smoking during pregnancy. This means that the association between DNA methylation and maternal smoking during pregnancy seems to be independent of SNPs, highlighting the potential importance of environmental influences on DNA methylation.

DISCUSSION

In this study, we showed associations between *in utero* maternal smoking exposure and CpG methylation in whole-blood DNA from adolescents, independent of paternal smoking during the period of pregnancy, cumulative passive smoke exposure, and adolescent smoking. Additionally, we showed a trend for dose-dependent effects of maternal smoking on offspring CpG methylation levels. The CpG methylation level associations with maternal smoking are in accordance with previous findings at birth (Joubert et al., 2016), during childhood (Rzehak et al., 2016), adolescence (Lee et al., 2015), and in middle age (Sun et al., 2013). Apart from cg24935556 (*APOB*), all CpGs identified in this study were identified in the meta-analysis by Joubert et al. (2016).

We did not detect associations between paternal smoking during pregnancy, adolescent smoking, or passive smoking exposure and DNA methylation. Our findings suggest that maternal smoking during pregnancy induces long-lasting DNA methylation changes in the offspring established by adolescence, which are not greatly modified by postnatal smoke exposure. Furthermore, we found that methylation levels at two CpGs (cg00253568 and cg00213123, located in the *FTO* and *CYP1A1* region) identified in association with maternal smoking during pregnancy were also associated with cardiometabolic health variables, suggesting that maternal smoking during pregnancy may induce changes that affect offspring cardiometabolic health.

Maternal Smoking

Differential hypermethylation associated with *AHRR*, *CNTNAP2*, *CYP1A1*, *FRMD4A*, *GFI1*, *MYO1G*, and *CYP1A1* has been shown previously in the same CpGs that we identified (Rotroff et al., 2016; Tehranifar et al., 2018). A study analyzing the associations between maternal smoking during pregnancy and adolescent CpG methylation levels in a discovery population of 132 and a replication cohort of n = 447 also showed methylation levels associated to maternal smoking within *MYO1G*, *CNTANAP2*, *GFI1*, *CYP1A1*, and *AHRR* but did not analyze the effect of any other sources of smoke exposure on CpG methylation levels (Lee et al., 2015). The majority of CpG sites in the meta-analysis by Joubert et al. (2016) were identified with the same direction of effect as in our study. Given Joubert et al. analyzed cord blood and our study used whole blood, these findings demonstrate consistent methylation patterns over different tissue types and time of sampling, which suggests lasting effects of maternal smoking during pregnancy on offspring DNA methylation. The specific methylation sites that we identified are consistent with previous reports in neonates (from cord blood) and middle-age populations [whole blood, lymphocytes (mononuclear)] (Philibert et al., 2013; Zeilinger et al., 2013). This establishes a high consistency of DNA methylation markers related to maternal smoking during pregnancy. Such lifetime persistence and consistency are essential prerequisites for using DNA methylation as a valid biomarker for exposure and potentially a predictor for future adverse health outcomes. Therefore, this study fills the gap in the literature confirming stable changes in DNA methylation in adolescence.

Overall, the findings from our study suggest that methylation changes are induced in early life and persist into adolescence. Maternal smoking during pregnancy potentially exposes the fetus to cigarette-related chemicals and toxins leading to an early life "programming" effect that persists into adolescence and potentially affects long-term health.

Paternal Smoking Effects

There are fewer studies examining the effect of paternal than maternal smoking on offspring health, although prepregnancy exposure to paternal smoking is associated with a higher risk of leukemia, childhood cancers, and asthma in the offspring (Jenkins et al., 2017).

Analyses in the ALSPAC study suggested associations between paternal prepregnancy smoking and offspring BMI (Pembrey et al., 2005). In this study, 166 fathers were identified who started smoking before their offspring was aged 11 years. Compared to the nonsmoking fathers and fathers with later onset of smoking, the male offspring of fathers who commenced smoking before they were 11 years old had a higher BMI at 9 years of age. When tested in the Raine study with paternal smoking during pregnancy and BMI of the offspring at 16 years of age, a significant association (p = 0.008) was found after adjusting for age and sex.

In another study, Jenkins et al. used the Illumina 450K BeadChip to examine the effect of paternal cigarette smoking on sperm DNA methylation (Jenkins et al., 2017) in 78 male never smokers compared to 78 smokers. They showed that 141 CpG loci were differentially methylated in the sperm of smokers and suggested transgenerational inheritance. In our study, we did not find any effects of paternal prepregnancy smoking on offspring whole-blood DNA methylation, possibly due to our sample being insufficiently statistically powered. It is also possible that the effect of paternal smoking might be less prominent and too small

to detect given our sample size. However, the inability to detect associated CpG methylation at genome-wide significance, while being able to detect a large number of CpGs with methylation levels associated maternal smoking, suggests a dominant effect of maternal smoke exposure.

Adolescent Smoking

There are some studies that have consistently shown cross-sectional associations of CpG methylation with current smoking in adults and adolescents. For example, the CpG cg05575921, mapped to *AHRR* in our study, has been associated with smoking in a recent study (Li et al., 2018). Similarly, a study that analyzed the effect of smoking on several timepoints and after smoking cessation (Wilson et al., 2017) showed that cg05575921 and five other CpGs related to *AHRR* associated with smoking. Another study by Lee et al. in a Korean population with a sample size of 100 (31 current, 30 former, and 39 never smoker) showed similar results, with the strongest association again being with cg05575921 (Lee et al., 2016).

A limitation of each of these studies that have examined the effect of adolescent smoking on DNA methylation is attributing the findings to current smoking without consideration of the possible effects of *in utero* exposure in the form of maternal smoking during pregnancy. This is very likely a complication, as offspring of maternal smokers are more likely to smoke themselves (Gilman et al., 2003; Rosemary et al., 2012). Within our study, 55% of adolescent smokers had mothers who smoked.

In the current study, we addressed this issue by running separate analyses for those adolescents who smoked themselves and were exposed to maternal smoking during pregnancy versus the ones who were not exposed. This showed that the beta coefficients in the adolescent smokers who were also exposed to maternal smoking were most similar to the CpGs associated with maternal smoking in our main analysis. Although the sample size is too small to show significant effects, this suggests a dominant effect of *in utero* smoke exposure. This tendency can be seen in **Supplemental Figure 4**, when comparing the methylation change between the study participants not exposed to any smoking, those smoking at 17 years, those exposed to maternal smoking during pregnancy, and those exposed to both types of smoking. In this barplot, the exposure to maternal smoking during pregnancy causes the majority of the methylation change, mostly equal to those exposed to maternal smoking and smoking themselves.

Passive Smoking

To our knowledge, there are no studies to date analyzing the effect of passive smoke exposure over the life course on DNA methylation, despite the evidence that passive smoking is associated with manifold diseases such as chronic obstructive pulmonary disease, wheeze, asthma, and food allergy, as well as cancer (Le-Ha et al., 2013; Saulyte et al., 2014; Vardavas et al., 2016). A study in the Avon Longitudinal Study of Parents and Children analyzed passive smoke exposure as paternal smoking during pregnancy and mothers exposure to smoking of her father and mother but did not assess the offspring' s postnatal passive smoke exposure (Richmond et al., 2018).

In our analysis, we did not detect associations between lifetime passive smoke exposure and CpG methylation in adolescence. The accuracy and reliability of measurement of passive exposure may be limited. However, validity is enhanced in the current study by repeated longitudinal measures, which act as internal validation and prospective collection of data. The consistent answering of the question on maternal and paternal smoking over eight follow-ups, rather than from a single time point, increases the likelihood of a valid measure.

Cardiometabolic Risk-Related Genes

We observed increased methylation, within the *FTO* gene (cg00253658; chr16:54210496), in the offspring of mothers who had smoked during pregnancy. Variants in this gene have previously been shown to associate with birthweight and the development of obesity and diabetes (Frayling et al., 2007) Their functional impact may be in modifying expression of the *IRX3* and *IRX5* genes, rather than *FTO* itself (Smemo et al., 2014). Others have found hypermethylation in the region of this gene in relation to maternal smoking, in African American and Hispanic populations although in a different CpG, namely cg03687532 (chr16:54228358) (Tehranifar et al., 2018). Furthermore, methylation levels are *FTO*, and *CYP1A1* mapped CpGs were significantly associated with TG, diastolic blood pressure and HDL-C in our study, suggesting correlation with early life environments (i.e., smoke exposure) and later cardiometabolic health.

Another study found that CpGs associated with exposure to maternal smoking during pregnancy were also associated with all cause as well as cardiovascular mortality. This study identified significant associations for all cause and cardiovascular specific mortality with the CpGs cg05575921 and cg06126421 (Zhang et al., 2016). In our study, cg05575921 was associated with maternal smoking during pregnancy but not significantly associated with any of the cardiometabolic risk factors. However, methylation levels at cg05575921 were associated with the lowest p value across the genome with systolic blood pressure in the male and female combined analysis (uncorrected $p = 0.02$, Bonferroni corrected = 0.47), the lowest p value in the female only association with systolic blood pressure (uncorrected $p = 0.006$, Bonferroni corrected = 0.14), among the top 5 lowest p values in the female only with diastolic blood pressure analysis (uncorrected $p = 0.017$, Bonferroni corrected = 0.39), and methylation levels at this CpG with the second lowest p value in the female only analysis with TG (uncorrected $p = 0.01$, Bonferroni corrected = 0.34). Considering the low sample sizes, especially of the female subset ($n = 370$), there may be a suggestive association.

The beforementioned study by Parmar et al. (2018) found a most significant association between CpG methylation levels and maternal prenatal smoking with waist circumference, TG, and blood pressure with cg14179389 (*GFI1*). As stated previously, this CpG was also among the 23 CpGs identified as having methylation levels significantly correlated with maternal smoking in our study, but methylation levels for cg14179389 were not significantly associated with cardiometabolic risk in our analysis. For the female subset in the TG analysis, however, methylation at this CpG had the lowest uncorrected p value (uncorrected

$p = 0.01$, Bonferroni corrected = 0.27). This sex-specific tendency seems to be in line with what Parmer et al. observed, as they stated that adjusting their model for sex, age, and adult own smoking strengthened the association. Furthermore, considering they found associations with a Bonferroni corrected $p \leq 0.01$ within a meta-analysis, accessing a sample size of 18,212 adults, our findings only showing tendencies with a maximum of $n = 870$ is not surprising.

Strengths and Limitations

Strengths of this study are the prospective and repeated measures (at eight time points) of cigarette smoke exposure in ~800 participants. The internal validation of cross-checking answers across time increases the reliability of the questionnaire data. Our findings accord with the same CpG sites that associate with smoking in studies that have used cotinine levels to confirm smoking status (Joubert et al., 2012; Philibert et al., 2013; Lee et al., 2016; Morales et al., 2016; Rotroff et al., 2016). DNA methylation sites identified in our study are in gene regions previously associated with maternal smoking and are in the same direction of association (Joubert et al., 2016).

While cotinine is considered the gold standard for evaluation of smoking, a number of studies have shown very high correlations between cotinine levels and questionnaire data, up to 97% (Patrick et al., 1994; Parazzini et al., 1996; Vartiainen et al., 2002; Dolcini et al., 2003). A subset of the Raine study mothers ($n = 238$) had cotinine measures available at 28 weeks of gestation, and, as previously shown, cotinine concentration significantly differed between the groups of reported number of cigarettes smoked, highlighting the validity of the Raine study maternal smoking questionnaire data (Stick et al., 1996).

A further strength of our study is the ability to adjust for a wide range of possible confounders, in particular socioeconomic status, which is associated with smoking behavior and DNA methylation (McDade et al., 2019). However, it is still possible that other unmeasured environmental factors in pregnancy or postnatally could be influencing or modifying some of these findings. Owing to the deeply phenotyped character of the Raine study, we were able to adjust all the models for multiple sources of smoke exposure, narrowing down to the specific effect of maternal smoking on DNA methylation in the offspring. The fact that we found associations between methylation levels at the identified CpGs and cardiometabolic health-related variables suggests correlations between smoke exposure and offspring health.

A further strength of our study is that we integrated genetic (SNPs) and epigenetic (CpG methylation) information and assessed if the association between CpG methylation and maternal smoking during pregnancy still persists when accounting for SNPs. To our knowledge, this was not done to this extent in any DNA methylation wide association study previous to ours.

The majority of CpG sites in the meta-analysis by Joubert et al. (2016) were identified with the same direction of effect as in our study. Given that Joubert et al. analyzed cord blood and our study used whole blood, these findings demonstrate consistent DNA methylation patterns over different sample types and time points in response to maternal smoking during pregnancy.

A potential limitation is that we only examined methylation from whole-blood DNA, which might not be the site of change in association with smoking. Few population studies have cell-sorted DNA methylation, and our findings suggest that some of these changes may be induced across multiple cell types. Furthermore, the sample sizes of some of the analysis that we conducted are below 200, making them potentially underpowered to detect small epigenetic changes.

It is known that up to 6% of the probes in the Illumina Methylation450 BeadChip kit could give false positives, due to known cross-reactivity. Furthermore, the array only covers 2% of the epigenome CpG DNA methylation sites (Kurdyukov and Bullock, 2016). To mitigate some of these limitations, we performed thorough preprocessing and QC steps to remove any problematic probes and samples. In addition, we accounted for batch effects in all our models and used a conservative Bonferroni correction for multiple testing to minimize any false positives that may have arisen due to technical issues from probes on the 450k array.

In our cohort, it is encouraging that we show similar associations between *in utero* smoke exposure and CpG methylation, both in amount and specific sites (Breton et al., 2017). However, performing independent methylation analysis such as pyrosequencing would have further strengthened the inferences. Lastly, for the dose–response relationship, the questionnaire variable for the number of cigarettes consumed needs to be analyzed with caution. With questionnaire data, there is always a chance of recall bias or underreporting, especially when it comes to behaviors such as cigarette or alcohol consumption.

CONCLUSIONS

We have shown associations between maternal smoking during pregnancy and offspring DNA methylation at 23 CpGs in adolescents at age 17 years. These associations were predominantly driven by maternal smoking and not modified by paternal, passive, or adolescent smoking. Furthermore, we are unable to detect genome-wide significant associations with paternal smoking and passive smoke exposure at any CpG sites. Our data that suggest DNA methylation changes in offspring are likely due to the direct effect of maternal smoking during pregnancy, rather than current, passive, or paternal smoking. Future studies on smoking habits and DNA methylation should adjust for maternal smoking, in addition to socioeconomic status of the mother and/or offspring, depending on the age of the offspring. The specific methylation sites that we identified are in agreement with previous reports in neonates (from cord blood) and middle aged populations [whole blood, lymphocytes (mononuclear)] (Philibert et al., 2013; Zeilinger et al., 2013). This establishes a high consistency of DNA methylation markers related to maternal smoking during pregnancy. Such persistence and consistency are essential prerequisites for using DNA methylation as a valid biomarker for exposure and potentially a predictor for future adverse health outcomes. Furthermore, we showed that maternal-smoking-induced methylation changes are associated with cardiometabolic variables, suggesting early life "programming" of later life cardiometabolic health.

DATA AVAILABILITY

The datasets used during and/or analyzed during the current study are available from the corresponding author on reasonable request.

ETHICS STATEMENT

Ethics approval for conducting the epigenetic analysis at the 17-year follow-up was given by the Human Ethics Committee of the University of Western Australia. Informed and written consent was provided by the participants and their parents or carer.

AUTHOR CONTRIBUTIONS

SR wrote the manuscript performed the analysis and interpreted the results. R-CH, PM, T-M, L-B, GB, JC, K-G, J-H, and KL contributed to the conception and design of the study, revised the manuscript and helped to interpret the results. CP and WO contributed to interpretation of the results and revised the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

The DNA methylation work was supported by NHMRC grant 1059711. R-CH and T-M are supported by NHMRC Fellowships (grant numbers 1053384 and 1042255, respectively).

This work was supported by resources provided by The Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia.

SR received support from the European LifeCycle project through the fellowship call of June 2018, Grant agreement No. 733206. K-G is supported by the UK Medical Research Council (MC_UU_12011/4), the National Institute for Health Research (as an NIHR Senior Investigator (NF-SI-0515-10042) and through the NIHR Southampton Biomedical Research Centre) and the European Union's Erasmus + Capacity-Building ENeA^{SEA} Project and Seventh Framework Programme (FP7/2007–2013), projects EarlyNutrition and ODIN under grant agreement numbers 289346 and 613977.

RCH, PM, and SR received further support through the NHMRC EU-collaborative grant with the number APP1142858—Early life stressors and lifecycle health.

ACKNOWLEDGMENTS

We acknowledge Raine Study participants and their families, The Raine Study Team for cohort coordination and data collection, the NHMRC for their long-term contribution to funding the study over the last 29 years, and the Telethon Kids Institute for the long-term support of the study. We also acknowledge The University of Western Australia, Curtin University, Women and Infants Research Foundation, Edith Cowan University, Murdoch University, The University of Notre Dame Australia, and the Raine Medical Research Foundation for providing funding for core management of the Raine Study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00770/full#supplementary-material

REFERENCES

Agrawal, A., Scherrer, J. F., Grant, J. D., Sartor, C. E., Pergadia, M. L., Duncan, A. E., et al. (2010). The effects of maternal smoking during pregnancy on offspring outcomes. *Prevent. Med.* 50 (1–2), 13–8. doi: 10.1016/j.ypmed.2009.12.009

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., et al. (2014). Minfi: a flexible and comprehensive bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 30 (10), 1363–9. doi: 10.1093/bioinformatics/btu049

Assenov, Y., Müller, F., Lutsik, P., Walter, J., Lengauer, T., and Bock, C. (2014). Comprehensive analysis of DNA methylation data with RnBeads. *Nat. Methods.* 11 (11), 1138. doi: 10.1038/nmeth.3115

Bhattacharya, S., Beasley, M., Pang, D., and Macfarlane, G. J. (2014). Maternal and perinatal risk factors for childhood cancer: record linkage study. *BMJ Open.* 4 (1), e003656. doi: 10.1136/bmjopen-2013-003656

Bojesen, S. E., Timpson, N., Relton, C., Davey Smith, G., and Nordestgaard, B. G. (2017). AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. *Thorax.* 72 (7), 646–53. doi: 10.1136/thoraxjnl-2016-208789

Breton, C. V., Marsit, C. J., Faustman, E., Nadeau, K., Goodrich, J. M., Dolinoy, D. C., et al. (2017). Small-magnitude effect sizes in epigenetic end points are important in children's environmental health studies: The Children's Environmental Health and Disease Prevention Research Center's Epigenetics Working Group. *Environ. Health Perspect.* 125 (4), 511–26. doi: 10.1289/EHP595

DiFranza, J. R., Aligne, C. A., and Weitzman, M. (2004). Prenatal and postnatal environmental tobacco smoke exposure and children's health. *Pediatrics.* 113 (Supplement 3), 1007–15.

Dolcini, M. M., Adler, N. E., Lee, P., and Bauman, K. E. (2003). An assessment of the validity of adolescent self-reported smoking using three biological indicators. *Nicotine Tob. Res.* 5 (4), 473–83. doi: 10.1080/14622200307247

Finkelstein, J. and Jeong I. C. (2017). Machine learning approaches to personalize early prediction of asthma exacerbations. *Ann. N. Y. Acad Sci.* 1387 (1), 153–65. doi: 10.1111/nyas.13218

Fortin, J. P., Fertig, E., and Hansen, K. (2014). shinyMethyl: interactive quality control of Illumina 450k DNA methylation arrays in R [version 2; peer review: 2 approved]. *F1000Res.* 3, 175. doi: 10.12688/f1000research.4680.2

Frayling, T. M., Timpson, N. J., Weedon, M. N., Zeggini, E., Freathy, R. M., Lindgren, C. M., et al. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science (New York, NY).* 316 (5826), 889–94.

Gillman, M. W. (2005). Developmental origins of health and disease. *N. Eng. J. Med.* 353 (17), 1848. doi: 10.1056/NEJMe058187

Gilman, S. E., Abrams, D. B., and Buka, S. L. (2003). Socioeconomic status over the life course and stages of cigarette use: initiation, regular use, and cessation. *J. Epidemiol. Commun. Health.* 57 (10), 802–8. doi: 10.1136/jech.57.10.802

Hanson, M. A., and Gluckman, P. D. (2008). Developmental origins of health and disease: new insights. *Basic Clin. Pharmacol. Toxicol.* 102 (2), 90–3. doi: 10.1111/j.1742-7843.2007.00186.x

Hofhuis, W., de Jongste, J. C., and Merkus, P. J. F. M. (2003). Adverse health effects of prenatal and postnatal tobacco smoke exposure on children. *Arch. Dis. Child.* 88 (12), 1086–90. doi: 10.1136/adc.88.12.1086

Houseman, E. A., Molitor, J., and Marsit, C. J. (2014). Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics (Oxford, England).* 30 (10), 1431–9. doi: 10.1093/bioinformatics/btu029

Huang, R.-C., Mori, T. A., Burrows, S., Ha, C. L., Oddy, W. H., Herbison, C., et al. (2012). Sex dimorphism in the relation between early adiposity and cardiometabolic risk in adolescents. *J. Clin. Endocrinol. Metab.* 97 (6), E1014–E22. doi: 10.1210/jc.2011-3007

Jenkins, T. G., James, E. R., Alonso, D. F., Hoidal, J. R., Murphy, P. J., Hotaling, J. M., et al. (2017). Cigarette smoking significantly alters sperm DNA methylation patterns. *Andrology.* 5 (6), 1089–99. doi: 10.1111/andr.12416

Jones, R. M., Cadby, G., Melton, P. E., Abraham, L. J., Whitehouse, A. J., and Moses, E. K. (2013). Genome-wide association study of autistic-like traits in a general population study of young adults. *Front. Hum. Neurosci.* 7, 658–. doi: 10.3389/fnhum.2013.00658

Joubert, B. R., Haberg, S. E., Nilsen, R. M., Wang, X., Vollset, S. E., Murphy, S. K., et al. (2012). 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environ. Health Perspect.* 120 (10), 1425–31. doi: 10.1289/ehp.1205412

Joubert, B. R., Felix, J. F., Yousefi, P., Bakulski, K. M., Just, A. C., Breton, C., et al. (2016). DNA Methylation in newborns and maternal smoking in pregnancy: Genome-Wide Consortium Meta-analysis. *Am. J. Hum. Genet.* 98 (4), 680–96. doi: 10.1016/j.ajhg.2016.02.019

Kurdyukov, S., and Bullock, M. (2016). DNA methylation analysis: choosing the right method. *Biology* 5 (1), 3. doi: 10.3390/biology5010003

Lee, K. W. K., Richmond, R., Hu, P., French, L., Shin, J., Bourdon, C., et al. (2015). Prenatal exposure to maternal cigarette smoking and DNA methylation: epigenome-wide association in a discovery sample of adolescents and replication in an independent cohort at birth through 17 years of age. *Environ. Health Perspect.* 123 (2), 193–9. doi: 10.1289/ehp.1408614

Lee, M. K., Hong, Y., Kim, S. Y., London, S. J., and Kim, W. J. (2016). DNA methylation and smoking in Korean adults: epigenome-wide association study. *Clin. Epigenetics.* 8, 103. doi: 10.1186/s13148-016-0266-6

Le-Ha, C., Beilin, L. J., Burrows, S., Huang, R. C., Oddy, W. H., Hands, B., et al. (2013). Gender difference in the relationship between passive smoking exposure and HDL-cholesterol levels in late adolescence. *J. Clin. Endocrinol. Metab.* 98 (5), 2126–35. doi: 10.1210/jc.2013-1016

Li, S., Wong, E. M., Bui, M., Nguyen, T. L., Joo, J. E., Stone, J., et al. (2018). Causal effect of smoking on DNA methylation in peripheral blood: a twin and family study. *Clin. Epigenetics.* 10, 18. doi: 10.1186/s13148-018-0452-9

Matthews, D. R., Hosker, J. P., Rudenski, A. S., Naylor, B. A., Treacher, D. F., and Turner, R. C. (1985). Homeostasis model assessment—insulin resistance and beta-cell function from fasting plasma-glucose and insulin concentrations in man. *Diabetologia.* 28 (7), 412–9. doi: 10.1007/BF00280883

McDade, T. W., Ryan, C. P., Jones, M. J., Hoke, M. K., Borja, J., Miller, G. E., et al. (2019). Genome-wide analysis of DNA methylation in relation to socioeconomic status during development and early adulthood. *Am. J. Phys. Anthropol.* 169 (1), 3–11. doi: 10.1002/ajpa.23800

Morales, E., Vilahur, N., Salas, L. A., Motta, V., Fernandez, M. F., Murcia, M., et al. (2016). Genome-wide DNA methylation study in human placenta identifies novel loci associated with maternal smoking during pregnancy. *Int. J. Epidemiol.* 45 (5), 1644–55. doi: 10.1093/ije/dyw196

Newnham, J. P., Evans, S. F., Michael, C. A., Stanley, F. J., and Landau, L. I. (1993). Effects of frequent ultrasound during pregnancy: a randomised controlled trial. *Lancet* 342 (8876), 887–91. doi: 10.1016/0140-6736(93)91944-H

Nielsen, C. H., Larsen, A., and Nielsen, A. L. (2016). DNA methylation alterations in response to prenatal exposure of maternal cigarette smoking: a persistent epigenetic impact on health from maternal lifestyle? *Arch. Toxicol.* 90 (2), 231–45. doi: 10.1007/s00204-014-1426-0

Oken, E., Huh, S. Y., Taveras, E. M., Rich-Edwards, J. W., and Gillman, M. W. (2005). Associations of maternal prenatal smoking with child adiposity and blood pressure. *Obes. Res.* 13 (11), 2021–8. doi: 10.1038/oby.2005.248

Pan, H., Holbrook, J. D., Karnani, N., and Kwoh, C. K. (2016). Gene, Environment and Methylation (GEM): a tool suite to efficiently navigate large scale epigenome wide association studies and integrate genotype and interaction between genotype and environment. *BMC Bioinform.* 17, 299. doi: 10.1186/s12859-016-1161-z

Parazzini, F., Davoli, E., Rabaiotti, M., Restelli, S., Stramare, L., Dindelli, M., et al. (1996). Validity of self-reported smoking habits in pregnancy: a saliva cotinine analysis. *Acta Obstet. Gynecol. Scand.* 75 (4), 352–4. doi: 10.3109/00016349609033330

Parmar, P., Lowry, E., Cugliari, G., Suderman, M., Wilson, R., Karhunen, V., et al. (2018). Association of maternal prenatal smoking GFI1-locus and cardio-metabolic phenotypes in 18,212 adults. *EBioMedicine* 38, 206–16. doi: 10.1016/j.ebiom.2018.10.066

Patrick, D. L., Cheadle, A., Thompson, D. C., Diehr, P., Koepsell, T., and Kinne, S. (1994). The validity of self-reported smoking: a review and meta-analysis. *Am. J. Public Health* 84 (7), 1086–93. doi: 10.2105/AJPH.84.7.1086

Pembrey, M. E., Bygren, L. O., Kaati, G., Edvinsson, S., Northstone, K., Sjöström, M., et al. (2005). Sex-specific, male-line transgenerational responses in humans. *Eur. J. Hum. Genet.* 14, 159. doi: 10.1038/sj.ejhg.5201538

Philibert, R. A., Beach, S. R., Lei, M. K., and Brody, G. H. (2013). Changes in DNA methylation at the aryl hydrocarbon receptor repressor may be a new biomarker for smoking. *Clin. Epigenetics.* 5 (1), 19. doi: 10.1186/1868-7083-5-19

Richmond, R. C., Simpkin, A. J., Woodward, G., Gaunt, T. R., Lyttleton, O., McArdle, W. L., et al. (2015). Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the Avon Longitudinal Study of Parents and Children (ALSPAC). *Hum. Mol. Genet.* 24 (8), 2201–17. doi: 10.1093/hmg/ddu739

Richmond, R. C., Suderman, M., Langdon, R., Relton, C. L., and Davey Smith, G. (2018). DNA methylation as a marker for prenatal smoke exposure in adults. *Int. J. Epidemiol.* 1, 47 (4), 1120–1130. doi: 10.1093/ije/dyy091

Rosemary, H., Linda, B., Amanda, A., A., F. J., and Marcus, M. (2012). Socioeconomic status and smoking: a review. *Ann. N. Y. Acad. Sci.* 1248 (1), 107–23. doi: 10.1111/j.1749-6632.2011.06202.x

Rotroff, D. M., Joubert, B. R., Marvel, S. W., Haberg, S. E., Wu, M. C., Nilsen, R. M., et al. (2016). Maternal smoking impacts key biological pathways in newborns through epigenetic modification in Utero. *BMC Genomics* 17 (1), 976. doi: 10.1186/s12864-016-3310-1

Rzehak, P., Saffery, R., Reischl, E., Covic, M., Wahl, S., Grote, V., et al. (2016). Maternal smoking during pregnancy and DNA-methylation in children at age 5.5 years: epigenome-wide-analysis in the European Childhood Obesity Project (CHOP)-Study. *PLoS One* 11 (5), e0155554. doi: 10.1371/journal.pone.0155554

Saulyte, J., Regueira, C., Montes-Martínez, A., Khudyakov, P., and Takkouche, B. (2014). Active or passive exposure to tobacco smoking and allergic rhinitis, allergic dermatitis, and food allergy in adults and children: a systematic review and meta-analysis. *PLOS Med.* 11 (3), e1001611. doi: 10.1371/journal.pmed.1001611

Schulz, L. C. (2010). The Dutch Hunger Winter and the developmental origins of health and disease. *Proc. Natl. Acad. Sci.* 107 (39), 16757–8. doi: 10.1073/pnas.1012911107

Smemo, S., Tena, J. J., Kim, K. H., Gamazon, E. R., Sakabe, N. J., Gomez-Marin, C., et al. (2014). Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* 507 (7492), 371–5. doi: 10.1038/nature13138

Stick, S. M., Burton, P. R., Gurrin, L., Sly, P. D., and LeSouëf, P. N. (1996). Effects of maternal smoking during pregnancy and a family history of asthma on respiratory function in newborn infants. *Lancet.* 348 (9034), 1060–4. doi: 10.1016/S0140-6736(96)04446-7

Su, D., Wang, X., Campbell, M. R., Porter, D. K., Pittman, G. S., Bennett, B. D., et al. (2016). Distinct epigenetic effects of tobacco smoking in whole blood and among leukocyte subtypes. *PLoS One* 11 (12), e0166486. doi: 10.1371/journal.pone.0166486

Sun, Y. V., Smith, A. K., Conneely, K. N., Chang, Q., Li, W., Lazarus, A., et al. (2013). Epigenomic association analysis identifies smoking-related DNA methylation sites in African Americans. *Hum. Genet.* 132 (9), 1027–37. doi: 10.1007/s00439-013-1311-6

Teh, A. L., Pan, H., Chen, L., Ong, M. L., Dogra, S., Wong, J., et al. (2014). The effect of genotype and in utero environment on interindividual variation in neonate DNA methylomes. *Genome Res.* 24 (7), 1064–74. doi: 10.1101/gr.171439.113

Tehranifar, P., Wu, H. C., McDonald, J. A., Jasmine, F., Santella, R. M., Gurvich, I., et al. (2018). Maternal cigarette smoking during pregnancy and offspring DNA methylation in midlife. *Epigenetics.* 13 (2), 129–34. doi: 10.1080/15592294.2017.1325065

Teschendorff, A. E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., et al. (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics (Oxford, England).* 29 (2), 189–96. doi: 10.1093/bioinformatics/bts680

Van Iterson, M., Tobi, E. W., Slieker, R. C., den Hollander, W., Luijk, R., Slagboom, P. E., et al. (2014). MethylAid: visual and interactive quality control of large Illumina 450k datasets. *Bioinformatics (Oxford, England).* 30 (23), 3435–7. doi: 10.1093/bioinformatics/btu566

Vardavas, C. I., Hohmann, C., Patelarou, E., Martinez, D., Henderson, A. J., Granell, R., et al. (2016). The independent role of prenatal and postnatal exposure to active and passive smoking on the development of early wheeze in children. *Eur. Respir. J.* 48, 115–124. doi: 10.1183/13993003.01016-2015

Vartiainen, E., Seppälä, T., Lillsunde, P., and Puska, P. (2002). Validation of self reported smoking by serum cotinine measurement in a community-based study. *J. Epidemiol. Commun. Health.* 56 (3), 167–70. doi: 10.1136/jech.56.3.167

Wakschlag, L. S., Pickett, K. E., Edwin Cook, J., Benowitz, N. L., and Leventhal, B. L. (2002). Maternal smoking during pregnancy and severe antisocial behavior in offspring: a review. *Am. J. Public Health* 92 (6), 966–74. doi: 10.2105/AJPH.92.6.966

Wilson, R., Wahl, S., Pfeiffer, L., Ward-Caviness, C. K., Kunze, S., Kretschmer, A., et al. (2017). The dynamics of smoking-related disturbed methylation: a two time-point study of methylation change in smokers, non-smokers and former smokers. *BMC Genomics* 18 (1), 805. doi: 10.1186/s12864-017-4198-0

Zeilinger, S., Kuhnel, B., Klopp, N., Baurecht, H., Kleinschmidt, A., Gieger, C., et al. (2013). Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One* 8 (5), e63812. doi: 10.1371/journal.pone.0063812

Zhang, Y., Schottker, B., Florath, I., Stock, C., Butterbach, K., Holleczek, B., et al. (2016). Smoking-associated DNA methylation biomarkers and their predictive value for all-cause and cardiovascular mortality. *Environ. Health Perspect.* 124 (1), 67–74. doi: 10.1289/ehp.1409020

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Rauschert, Melton, Burdge, Craig, Godfrey, Holbrook, Lillycrop, Mori, Beilin, Oddy, Pennell and Huang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 11 September 2019
doi: 10.3389/fgene.2019.00815

Predictive and Prognostic Value of Selected MicroRNAs in Luminal Breast Cancer

Maria Amorim[1,2], João Lobo[1,3,4†], Mário Fontes-Sousa[1,5†], Helena Estevão-Pereira[1,2], Sofia Salta[1], Paula Lopes[1,3], Nuno Coimbra[1,3,4], Luís Antunes[6], Susana Palma de Sousa[5], Rui Henrique[1,3,4‡] and Carmen Jerónimo[1,4‡]*

[1] Cancer Biology and Epigenetics Group, IPO Porto Research Center (CI-IPOP), Portuguese Oncology Institute of Porto (IPO Porto), Porto, Portugal, [2] Master in Oncology, Institute of Biomedical Sciences Abel Salazar–University of Porto (ICBAS-UP), Porto, Portugal, [3] Department of Pathology, Portuguese Oncology Institute of Porto, Porto, Portugal, [4] Department of Pathology and Molecular Immunology, Institute of Biomedical Sciences Abel Salazar–University of Porto (ICBAS-UP), Porto, Portugal, [5] Department of Medical Oncology, Portuguese Oncology Institute of Porto, Porto, Portugal, [6] Department of Epidemiology, Portuguese Oncology Institute of Porto, Porto, Portugal

OPEN ACCESS

Edited by:
Michael E. Symonds,
University of Nottingham,
United Kingdom

Reviewed by:
Nejat Dalay,
Istanbul University, Turkey
Mustafa Ozen,
Baylor College of Medicine,
United States

***Correspondence:**
Carmen Jerónimo
carmenjeronimo@ipoporto.min-saude.pt;
cljeronimo@icbas.up.pt

†These authors share second authorship
‡These authors share senior authorship

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: 16 January 2019
Accepted: 07 August 2019
Published: 11 September 2019

Citation:
Amorim M, Lobo J, Fontes-Sousa M, Estevão-Pereira H, Salta S, Lopes P, Coimbra N, Antunes L, Palma de Sousa S, Henrique R and Jerónimo C (2019) Predictive and Prognostic Value of Selected MicroRNAs in Luminal Breast Cancer. Front. Genet. 10:815. doi: 10.3389/fgene.2019.00815

Breast cancer (BrC) is the most frequent malignancy and the leading cause of cancer death among women worldwide. Approximately 70% of BrC are classified as luminal-like subtype, expressing the estrogen receptor. One of the most common and effective adjuvant therapies for this BrC subtype is endocrine therapy. However, its effectiveness is limited, with relapse occurring in up to 40% of patients. Because microRNAs have been associated with several mechanisms underlying endocrine resistance and sensitivity, they may serve as predictive and/or prognostic biomarkers in this setting. Hence, the main goal of this study was to investigate whether miRNAs deregulated in endocrine-resistant BrC may be clinically relevant as prognostic and predictive biomarkers in patients treated with adjuvant endocrine therapy. A global expression assay allowed for the identification of microRNAs differentially expressed between luminal BrC patients with or without recurrence after endocrine adjuvant therapy. Then, six microRNAs were chosen for validation using quantitative reverse transcription polymerase chain reaction in a larger set of tissue samples. Thus, *miR-30c-5p*, *miR-30b-5p*, *miR-182-5p*, and *miR-200b-3p* were found to be independent predictors of clinical benefit from endocrine therapy. Moreover, *miR-182-5p* and *miR-200b-3p* displayed independent prognostic value for disease recurrence in luminal BrC patients after endocrine therapy. Our results indicate that selected miRNAs' panels may constitute clinically useful ancillary tools for management of luminal BrC patients. Nevertheless, additional validation, ideally in a multicentric setting, is required to confirm our findings.

Keywords: Breast cancer, luminal subtype, endocrine therapy, endocrine resistance, biomarkers, microRNAs

INTRODUCTION

Breast cancer (BrC) is the second most common cancer worldwide and the most frequent cancer among women. Despite advances in screening, early diagnosis, and treatment strategies, BrC still constitutes the leading cause of cancer-related death among women (Bray et al., 2018). BrC is a highly heterogeneous disease with distinct biological features and clinical outcomes. Based on

gene expression profiling, BrC is often classified into four well-established intrinsic subtypes (**Table 1**) (Sørlie, 2004; Parker et al., 2009). However, due to logistic and economical constraints, surrogate approaches have been developed for routine clinical practice, using widely available immunohistochemistry (IHC) assays for estrogen receptor (ER), progesterone receptor (PR), and Ki-67 *index*, together with IHC and/or *in situ* hybridization for human epidermal growth factor 2 receptor (HER2) overexpression/amplification (Senkus et al., 2015).

In addition to surgery, therapeutic strategies for BrC patients include neoadjuvant, adjuvant, and palliative treatments. Adjuvant systemic therapy, aiming to prevent BrC recurrence by eradicating micrometastases present at diagnosis, includes three modalities: chemotherapy, anti-HER2 therapy (e.g., trastuzumab), and endocrine therapy (ET). ER and HER2 *status* are used as predictive factors to select patients for specific adjuvant therapies (**Table 1**). ET, which blocks ER activation, is recommended for patients with ER-positive disease, to stop or slow the growth of hormone-sensitive BrC (Curigliano et al., 2017). Most luminal A BrC tumors do not require adjuvant chemotherapy, except those with the highest risk of relapse, whereas most luminal B tumors, especially those with HER2 overexpression, benefit from chemotherapy in addition to trastuzumab (Slamon et al., 2011). Although ET results in substantial improvement of patients' outcome, resistance to treatment is a major hurdle (Zhang et al., 2014a), affecting 30–40% of ER-positive BrC patients, with all those treated in the metastatic setting eventually progressing (Normanno et al., 2005; Murphy and Dickler, 2016). According to the 3rd ESO–ESMO International Consensus Guidelines, endocrine resistance may be defined as primary endocrine resistance, when patients relapse within the first 2 years of adjuvant ET, or as secondary (acquired) endocrine resistance, when patients relapse while on adjuvant ET after the first 2 years of treatment or within 12 months after completing treatment (Cardoso et al., 2017).

MicroRNAs (miRNAs), a class of small (~22 nucleotides) non-coding single-stranded RNAs, have shown promise for assisting in clinical management of BrC as diagnostic, prognostic, or predictive biomarkers (Amorim et al., 2016), namely, through assessment in liquid biopsies (plasma, serum, and urine) (Schwarzenbach et al., 2014). Indeed, several studies have associated miRNAs deregulation with endocrine resistance and prognosis in luminal BrC (Rodriguez-Gonzalez et al., 2011; Muluhngwi and Klinge, 2015; Barbano et al., 2017; Muluhngwi and Klinge, 2017). Whereas decreased ER expression and endocrine resistance may be due to *miR-221/222* overexpression (Zhao et al., 2008; Rao et al., 2011; Wei et al., 2014; Song et al., 2017), *miR-342-3p* expression positively correlated with ER mRNA transcript levels, being downregulated in tamoxifen-refractory BrC (Cittelly et al., 2010). Moreover, miRNAs regulating growth, survival, and apoptosis of BrC cells may also be implicated in loss of responsiveness to ET by endowing tumor cells with alternative proliferative and survival stimuli (Thiantanawat et al., 2003). Indeed, *miR-519a* associated with worse prognosis in luminal BrC patients, directly targeting the transcripts of *cyclin dependent kinase inhibitor 1A* (*CDKN1A*) and *phosphatase and tensin homolog* (*PTEN*), allowing for enhanced signaling of the *phosphoinositide3-kinase* (PI3K) growth and survival pathway (Ward et al., 2014) and reducing sensitivity and tumor cell apoptosis in response to apoptotic stimuli (Breunig et al., 2017). Furthermore, miRNA-mediated endocrine resistance might be

TABLE 1 | Breast cancer molecular subtypes characterization (Perou et al., 2000; Sørlie et al., 2001; Oh et al., 2006; Eroles et al., 2012; Haque et al., 2012; Network, 2012; Howell, 2013; Zhang et al., 2014a; Senkus et al., 2015).

Breast cancer subtypes		Clinicopathological surrogate markers	Signature genes	Adjuvant systemic therapeutic options
Luminal A		ER^+ PR high[1] $HER2^-$ Ki-67 low[2]	*ESR1* and/or *PGR*, *KRT8*/18, *GATA3*, *XBP1*, *FOXA1,* and *ADH1B*	ET alone in most of cases + ChT if high tumor burden (≥N3, ≥T3)
Luminal B	**$HER2^-$**	ER^+ $HER2^-$ Ki-67 high or PR low	*ESR1* and/or *PGR*, *KRT8*/18, *FGFR1*, *ERBB1*, *MKI67* and/or *CCNE1*, *CCNB1,* and *MYBL2*	ET + ChT for the most of cases
	$HER2^+$	ER^+ $HER2^+$ Any Ki-67 Any PR		ChT + anti-HER2 + ET for all patients
Basal-like		ER^- PR^- $HER2^-$	*KRT5*/6, *KRT17*, *ERBB1* and/or *KIT*, FOXC1, TP63, *CDH3*, VIM, and *LAM*	ChT
HER2-enriched		$HER2^+$ ER^- PR^-	*ERBB2* and *GRB7*	ChT + anti-HER2

[1]Suggested cutoff value is 20%. [2]Ki-67 scores should be interpreted in the light of local laboratory median values. ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; ESR1, estrogen receptor 1; PGR, progesterone receptor; KRT, keratin; GATA3, GATA binding protein 3; XBP1, X-box binding protein 1; FOX, forkhead box; ADH1B, alcohol dehydrogenase 1B (Class I), beta polypeptide; FGFR1, fibroblast growth factor receptor 1; ERBB, Erb-B2 receptor tyrosine kinase; MKI67, marker of proliferation Ki-67; CCN, cyclin; MYBL2, MYB proto-oncogene like 2; MYBL2, MYB proto-oncogene like 2; KIT, KIT proto-oncogene receptor tyrosine kinase; TP63, tumor protein P63; CDH, cadherin; VIM, vimentin; LAM, laminin; GRB7, growth factor receptor bound protein 7; ChT, chemotherapy; ET, endocrine therapy; N, nodal stage; T, tumor size.

related with epithelial-to-mesenchymal transition (EMT) and metastatic potential of BrC cells, as members of the *miR-200 family (miR-200f)*, which act as major regulators of EMT, were found downregulated in endocrine-resistant BrC *vs.* endocrine-sensitive cell lines (Burk et al., 2008; Manavalan et al., 2013).

Herein, we aimed to identify miRNAs that might predict endocrine resistance in luminal BrC patients undergoing ET, by comparing expression levels between BrC samples of patients that developed endocrine resistance with those that did not, after long-term follow-up. Expression levels of the miRNAs identified might allow for stratification of luminal BrC cases into a low-risk patient subgroup, for which additional adjuvant systemic treatment can be safely omitted, and a high-risk group comprising patients at high risk for recurrence, allowing for detection of resistance to ET at an early stage.

MATERIALS AND METHODS

Patients and Samples Collection

For this study, 139 BrC tissue samples were prospectively collected, after informed consent, from patients with luminal BrC and without metastasis at diagnosis, aged between 41 and 75 years, submitted to adjuvant ET (with or without other adjuvant modalities), after first-line surgical treatment, from 1995 to 2002 at the Portuguese Oncology Institute of Porto (IPO-Porto). Furthermore, 26 normal breast tissue samples were collected from reduction mammoplasties of contralateral breast from BrC patients. All these specimens were obtained from patients without BrC hereditary syndrome and no evidence of preneoplastic/neoplastic lesions. After surgical resection, samples were immediately frozen at −80°C. Relevant clinical and pathological data were retrieved from patients' charts. Five-micrometer frozen sections were cut and stained with hematoxylin–eosin (H&E) staining for confirmation of BrC by an experienced pathologist, ensuring that samples contained at least 70% of tumor cells, and confirm that tissues obtained from reduction mammoplasties harbored normal epithelial cells. This study was approved by institutional ethical committee (CES-IPOFG-120/015).

BrC Subtyping

IHC was performed to identify the molecular subtype of each tumor tissue included in this study. Commercially available antibodies were used for ER (Clone 6F11, mouse, Leica), PR (Clone 16, mouse, Leica), HER2 (Clone 4B5, rabbit, Roche), and Ki-67 (Clone MIB-1, mouse, Dako). IHC was carried out in BenchMark ULTRA (Ventana, Roche) using ultraView Universal DAB Detection Kit (Ventana, Roche) according to the manufacturer's instructions. Each case was evaluated by an experienced pathologist; it was classified according to the College of American Pathologists recommendations (Fitzgibbons et al., 2014) and categorized according to ESMO guidelines (Senkus et al., 2015). Cutoffs for Ki-67 and PR expression were set at 15% and 25% of positive cells, respectively, according to the optimized protocols of Department of Pathology.

RNA Extraction From Fresh Frozen Tissues

Total RNA was extracted from fresh frozen tissues using the TRIzol® Reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's recommendations. RNA concentrations and purity ratios were ascertained using a NanoDrop Lite spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA), and RNA samples were stored at −80°C.

MiRNA cDNA Synthesis

cDNA synthesis was performed in a Veriti® Thermal Cycler (Applied Biosystems, Foster City, CA, USA) using miRCURY LNA™ Universal RT microRNA PCR (Exiqon, Vedbaek, Denmark) following the manufacturer's instructions. cDNA samples were then stored at −20°C.

Global Focus MiRNA PCR Panel

Global miRNAs' expression was evaluated using a Cancer Focus microRNA PCR Panel, 384-well (V4.R) (Exiqon). Each plate, besides containing 80 lyophilized LNA™ miRNA primer sets focusing on cancer-relevant human miRNAs, also contained interplate calibrators, candidate reference genes [miRNAs and small nuclear RNAs (snRNAs)], and one water blank. In each well, 0.05 µl of cDNA previously synthesized, 5 µl of SYBR® Green master mix (Exiqon), and 4.95 µl of nuclease-free water (Exiqon) were added. Quantitative reverse transcription polymerase chain reactions (RT-qPCR) were performed in the LightCycler 480 instrument (Roche Diagnostics, Manheim, Germany) according to the following conditions: 95°C for 10 min and 45 cycles at 95°C for 10 s and 60°C for 1 min.

The median values of *miR-103a-3p*, *miR-107*, *miR-191-5p,* and *SNORD38B* were used for normalization, as these genes were the most stably expressed candidate reference genes (**Supplementary Figure 1**). Differences in expression values for target miRNAs were calculated using the $2^{-\Delta\Delta CT}$ method. The selection of deregulated miRNAs for further validation was performed considering prominent fold change, good sensitivity for qRT-PCR detection (Ct values, in general, below 30), and novelty.

Individual Assays

Initially, cDNA samples were diluted 80× in sterile distilled water (B. Braun, Melsungen, Germany). Then, on ice, *per* each well of a 384-well plate, the following were added: 5 µl of NZYSpeedy qPCR Green Master Mix (2×) (NZYTECH, Portugal), 1 µl of miRNA specific primer mix (microRNA LNA™ PCR primer set, Exiqon), and 4 µl of previously diluted cDNA. Each amplification reaction was performed in triplicate on a LightCycler 480 instrument (Roche Diagnostics, Manheim, Germany). Each plate also contained two negative template controls. RT-qPCR protocol consisted of a denaturation step at 95°C for 2 min, followed by 40 amplification cycles at 95°C for 5 s and 60°C for 20 s. Melting curve analysis was performed according to the instrument's manufacturer's recommendations.

SNORD38B was used as a reference gene for data normalization, as this gene was the most stably expressed over

the whole range of the samples used for the global expression assay. Notwithstanding, the stability of *SNORD38B* expression was empirically validated in additional samples. Relative miRNA expression in each sample was calculated by the $2^{-\Delta\Delta CT}$ method (the target sequences of mature miRNAs analyzed are provided in **Supplementary Table 1**).

Statistical Analysis

Statistical analysis was performed using SPSS software (SPSS Version 24.0, Chicago, IL), and two-tailed p values were considered statistically significant when $p < 0.05$. Graphs were constructed using GraphPad 6 Prism (GraphPad Software, USA).

MiRNA Expression Analysis

Fold changes for single miRNAs were calculated using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001).

Association Between MiRNA Expression and Clinicopathological Features

To ascertain statistical significance for continuous variables, comparisons were made between independent samples and non-parametric Mann–Whitney U tests were performed. Spearman nonparametric correlation test was performed to assess the association between continuous variables. Chi-square test or Fisher's exact test were used as appropriate to compare proportions between two groups.

Survival Analysis

Some clinicopathological features were grouped, including pT stage (T1 and T2 and T3 and T4), pN stage (N0 and N1 and N2 and N3), and grade [grade (G)1 and G2 and G3] (Lakhani, 2012). Age was categorized into four groups (≤44, 45–64, 65–74, and ≥75), and miRNA expression levels were categorized according to 25th or 75th percentile. All survival analyses were restricted to 15 years of follow-up. Cox regression univariable and multivariable models were computed to assess standard clinicopathological variables and miRNA prognostic value. Hazard ratios (HRs) along with respective 95% confidence interval (95% CIs) were reported. Multivariable Cox models only included the statistically significant variables. Kaplan–Meier with log rank test was used to construct and compare survival curves according to categorized miRNA expression levels. Endocrine resistance-free survival (ERFS) was defined as the time between surgery and the recurrence dates. Recurrences occurring after 12 months of completing ET were not considered events for this analysis. Disease-free survival (DFS) was defined as the time between surgery date and recurrence date. Distant metastasis-free survival (DMFS) was defined as the time between surgery and the development of distant metastases. For prognostic assessment of miRNAs combined in panels, the miRNAs that remained statistically significant in multivariable analysis were differently combined, considering the same categories used in previous survival analysis (expression above or below P25). The best panels were selected based on the individual markers value in the Cox model: better HR, smaller 95% CI and p value, as well as value in stratified analysis.

RESULTS

Characteristics of Study Populations

The discovery cohort ($n = 16$), used for global expression assay analysis, consisted of four luminal A and four luminal B tumors from BrC patients who relapsed, and the same number of patients who did not relapse after adjuvant ET. Patients who relapsed during adjuvant ET or within the first 12 months of completing adjuvant ET were considered endocrine-resistant (**Table 2**).

The validation cohort was composed of a total of 149 subjects, comprising 123 luminal BrC and 26 normal breast tissues. Among 34 cancer patients that recurred during follow-up time,

TABLE 2 | Clinical and pathological data of luminal tumors included in the discovery cohort.

	Molecular subtype	Age at diagnosis	Grade	Stage	ChT	RT	Recurrency site	Endocrine-resistant
Patients who relapsed	**Luminal A**	82	G2	IIIA	NO	NO	Liver	YES
		41	G3	IIA	YES	YES	Bone	YES
		60	UNKN	IA	NO	YES	Contralateral breast	NO
		43	G2	IIB	YES	YES	Lymph nodes	NO
	Luminal B	65	G3	IIIC	YES	YES	Lung	YES
		63	G2	IIIA	NO	YES	Bone	YES
		67	G2	IIB	NO	NO	Bone	NO
		66	G3	IIIA	NO	NO	Locoregional	NO
Patients who did not relapse	**Luminal A**	70	G3	IIB	NO	YES	n.a.	n.a.
		68	G2	IIB	NO	YES		
		69	G2	IIIA	NO	NO		
		69	G2	IA	NO	YES		
	Luminal B	65	G3	IIIC	YES	YES		
		72	G3	IIIC	NO	YES		
		70	G1	IIB	NO	YES		
		73	G1	IIIC	NO	YES		

ChT, chemotherapy; RT, radiotherapy; UNKN, unknown; n.a., not applicable.

20 were considered endocrine-resistant. Clinical and pathological characteristics of patients and controls included in this study are shown in **Table 3**. Endocrine-sensitive and endocrine-resistant groups did not significantly differ concerning age distribution ($p = 0.136$). As expected, most of the endocrine-resistant cases were classified as luminal B ($p = 0.011$) and depicted high Ki-67 *index* ($p = 0.001$). Moreover, this group also showed a higher number of high-grade (G3) cases ($p = 0.027$). For the remaining clinicopathological features or treatment modalities, no significant differences were depicted.

TABLE 3 | Clinical and pathological data of luminal tumors and normal breast samples included in the validation cohort.

Clinipathological features	Endocrine-sensitive	Endocrine-resistant	NBr
Patients (n)	103	20	26
Age median (range)	61 (43–73)	59 (41–75)	54 (40–70)
	61 (41–75)		
Molecular subtype (%)			n.a.
Luminal A	47 (45.6)	3 (15.0)	
Luminal B	56 (54.4)	17 (85.0)	
Histological type (%)			n.a.
Invasive carcinoma of NST (IDC)	89 (86.4)	17 (85.0)	
Invasive lobular carcinoma	5 (4.8)	2 (10.0)	
Other special subtype carcinoma	1 (1.0)	1 (5.0)	
Mixed type carcinoma	8 (7.8)	0 (0.0)	
Progesterone receptor status (%)	85 (82.5)	13 (65.0)	n.a.
Positive	18 (17.5)	7 (35.0)	
Negative			
HER2 receptor status (%)			n.a.
Positive	9 (8.7)	5 (25.0)	
Negative	94 (91.3)	15 (75.0)	
Ki-67 index (%)			n.a.
<15%	78 (75.7)	6 (30.0)	
>15%	20 (19.4)	10 (50.0)	
UNKN	5 (4.9)	4 (20.0)	
Grade (%)			n.a.
G1	16 (15.5)	0 (0.0)	
G2	53 (51.5)	8 (40.0)	
G3	28 (27.2)	10 (50.0)	
Not determined	6 (5.8)	2 (10.0)	
Pathological T Stage (%)			n.a.
pT1	30 (29.1)	5 (25.0)	
pT2	50 (48.5)	13 (65.0)	
pT3	3 (2.9)	0 (0.0)	
pT4	5 (4.9)	1 (5.0)	
Not determined	15 (14.6)	1 (5.0)	
Pathological N Stage (%)			n.a.
pN0	40 (38.8)	8 (40.0)	
p N1	38 (36.9)	8 (40.0)	
p N2	7 (6.8)	3 (15.0)	
p N3	3 (2.9)	0 (0.0)	
Not determined	15 (14.6)	1 (5.0)	
Adjuvant RT			n.a.
Yes	76 (73.8)	17 (85.0)	
No	17 (16.5)	3 (15.0)	
Not determined	10 (9.7)	0 (0.0)	
Adjuvant ChT			n.a.
Yes	37 (35.9)	11 (55.0)	
No	50 (48.6)	7 (35.0)	
Not determined	16 (15.5)	2 (10.0)	

NBr, normal breast tissues; NST, no special type; IDC, invasive ductal carcinoma; HER2, human epidermal growth factor receptor 2; G, grade; RT, radiotherapy; ChT, chemotherapy; n.a., not applicable.

Global Focus MiRNA PCR Panel Analysis

In the global expression assay, one luminal A case with recurrence was excluded from the analysis, due to low RT-qPCR success rate (25% of the miRNAs did not amplify, and the remaining showed Ct values higher than 30). Likewise, 3 (*miR-202-3p*, *-206*, and *-20b-5p*) out of the 80 miRNAs were excluded due to low real-time PCR success rates. MiRNAs with fold variation values higher than 1 were selected, resulting in a panel comprising 56 miRNAs (**Table 4**).

Gene-Specific Assays

From the global expression assay analysis, *miR-30b-5p*, *miR-30c-5p*, *miR-181a-5p*, *miR-182-5p*, *miR-200b-3p*, and *miR-205-5p* were selected for further validation. All these miRNAs disclosed prominent fold change and good sensitivity for qRT-PCR detection, with different ranges of expression. *MiR-30b-5p* was chosen because several studies focused on other members of the *miR-30 family (miR-30f)* and, to the best of our knowledge, its predictive potential for ET had not been assessed previously (Cheng et al., 2012; Bockhorn et al., 2013; Zhang et al., 2014b; D'aiuto et al., 2015; Yang et al., 2017). *MiR-181a-5p* and *miR-200b-3p* were selected to confirm the reported association with endocrine resistance in *in vitro* studies (Hiscox et al., 2006; Maillot et al., 2009; Manavalan et al., 2011; Vesuna et al., 2012; Manavalan et al., 2013). Furthermore, *miR-182-5p* was also selected to better ascertain its role in endocrine resistance due to controversial results in global focus miRNA PCR panel, since it was overexpressed in luminal B tumors from recurrent patients and downregulated in luminal A tumors from recurrent patients. Finally, *miR-30c-5p* was chosen as a positive control since higher expression levels of this miRNA had been positively associated with benefit of ET, in multivariate analysis, in advanced ER-positive BrC (Rodriguez-Gonzalez et al., 2011).

Except for *miR-205-5p* expression ($p = 0.001$), *miR-181a-5p* ($p = 0.004$), *miR-182-5p* ($p < 0.001$), and *miR-200b-3p* ($p < 0.001$), expression levels were significantly higher in luminal BrC tissues than in normal breast tissues (**Figure 1**), whereas no differences were depicted for the levels of the remaining miRNAs. Nonetheless, *miR-30b-5p* ($p = 0.031$), *miR-30c-5p* ($p = 0.002$), and *miR-200b-3p* ($p = 0.021$) were significantly downregulated in endocrine-resistant BrC samples compared to endocrine-sensitive tumors (**Figure 2**).

Association Between MiRNA Expression and Clinicopathological Features

Higher *miR-30b-5p* and m*iR-30c-5p* expression levels were found in tumors lacking HER2 overexpression (HER2-negative) ($p = 0.010$, $p = 0.014$, respectively). Conversely, lower *miR-205-5p* expression levels were found in high grade (G3) BrC ($p = 0.009$) compared to G1/G2 BrC (**Figure 3**). Moreover, *miR-205-5p* expression levels inversely correlated with patients' age ($R = -0.200$, $p = 0.027$).

TABLE 4 | MiRNAs with fold variation values higher than 1 in the global expression assay.

LumA Rec *vs.* LumA NRec		LumB Rec *vs.* LumB NRec		Lum Rec *vs.* Lum NRec	
microRNA	**Fold Change**	**microRNA**	**Fold Change**	**microRNA**	**Fold Change**
miR-196a-5p	2.1281	**miR-9-5p[1]**	2.5978	**miR-9-5p[1]**	1.4448
miR-181b-5p	–1.0119	**miR-210-3p[1]**	1.7178	**miR-149-3p[1]**	1.23995
miR-130a-3p[1]	–1.0519	**miR-182-5p[2]**	1.6028	**miR-126-3p**	–1.0909
miR-29b-3p	–1.1169	**miR-7-5p[1]**	1.3978	**miR-1**	–1.1352
let-7b-5p	–1.1269	**miR-200c-3p**	1.2778	**miR-148a-3p**	–1.1419
let-7i-5p	–1.1369	**miR-31-5p[1]**	1.0928	**miR-30d-5p**	–1.2139
miR-106b-5p	–1.1419	**miR-221-3p**	1.0128	**miR-181a-5p[2]**	–1.4322
miR-132-3p[1]	–1.1519	**miR-125b-5p**	–1.0172	**miR-200a-3p**	–1.5732
miR-26b-5p	–1.1619	**miR-146a-5p**	–1.0372	**miR-205-5p[2]**	–2.3252
miR-19b-3p	–1.1769	**miR-181a-5p[2]**	–1.0622		
miR-192-5p[1]	–1.1969	**miR-205-5p[2]**	–1.1172		
let-7g-5p	–1.2019	**miR-1[1]**	–1.1472		
miR-16-5p	–1.2319	**miR-10b-5p**	–1.4022		
miR-15a-5p	–1.2619				
miR-106a-5p	–1.2669				
miR-20a-5p	–1.2769				
let-7a-5p	–1.3019				
miR-21-5p	–1.3169				
miR-214-3p	–1.3569				
miR-93-5p	–1.4119				
let-7f-5p	–1.4369				
miR-222-3p	–1.4419				
miR-200c-3p	–1.4719				
miR-155-5p	–1.5119				
let-7e-5p	–1.5119				
let-7d-5p	–1.5619				
miR-148a-3p	–1.6369				
miR-181a-5p[2]	–1.6519				
miR-23b-3p	–1.7569				
miR-23a-3p	–1.8069				
miR-19a-3p	–1.8519				
miR-1[1]	–1.8869				
miR-221-3p	–1.9319				
miR-195-5p	–1.9369				
miR-18a-5p[1]	–1.9919				
miR-30c-5p[2]	–2.0419				
miR-182-5p[2]	–2.1119				
miR-186-5p[1]	–2.1319				
miR-141-3p	–2.1619				
miR-17-5p[1]	–2.1919				
miR-30d-5p	–2.2769				
miR-30b-5p[2]	–2.4819				
miR-101-3p	–2.5319				
miR-200b-3p[2]	–3.0019				
miR-92b-3p[1]	–3.1069				
miR-200a-3p	–3.2169				
miR-205-5p[2]	–4.1269				

[1]Cps higher than 30. [2]miRNAs chosen for further validation. Lum, luminal; Rec, recurrent.

Survival Analyses

The median follow-up time was 180 months (17.4–180 months). At 15 years of follow-up, 70 (56.9% of total) patients were alive, of whom 66 (53.7% of total) had no evidence of cancer. Moreover, from the 53 patients (43.1% of total) who died, death was due to BrC in 30 (24.4% of total).

Overall, in univariable analysis, most standard clinicopathological parameters were significantly associated with ERFS. Specifically, patients with HER2 positivity (HR = 2.91, p = 0.039), high Ki-67 *index* (HR = 5.59, p = 0.001), high grade (G3) (HR = 2.84, p = 0.028), and luminal B subtype (HR = 4.48, p = 0.017) disclosed shorter ERFS. Importantly, the same was observed for patients with lower *miR-30c-5p*, *miR-30b-5p*, *miR-182-5p*, and *miR-200b-3p* levels (**Table 5**, **Figure 4**). In multivariable analysis, all miRNAs remained independent predictors of ERFS adjusted for Ki-67 *index* (**Table 5**). After stratification for Ki-67 *index*, *miR-30c-5p*, *miR-182-5p*, and *miR-200b-3p* only independently predicted shorter ERFS in highly proliferative tumors, whereas *miR-30b-5p* was significant in tumors with low proliferative (**Table 6**).

Regarding DFS, in addition to HER2 positivity (HR = 2.40, p = 0.039), high Ki-67 *index* (HR = 3.01, p = 0.003), and high grade (G3) (HR = 2.65, p = 0.006), lower *miR-30c-5p*, *miR-30b-5p*,

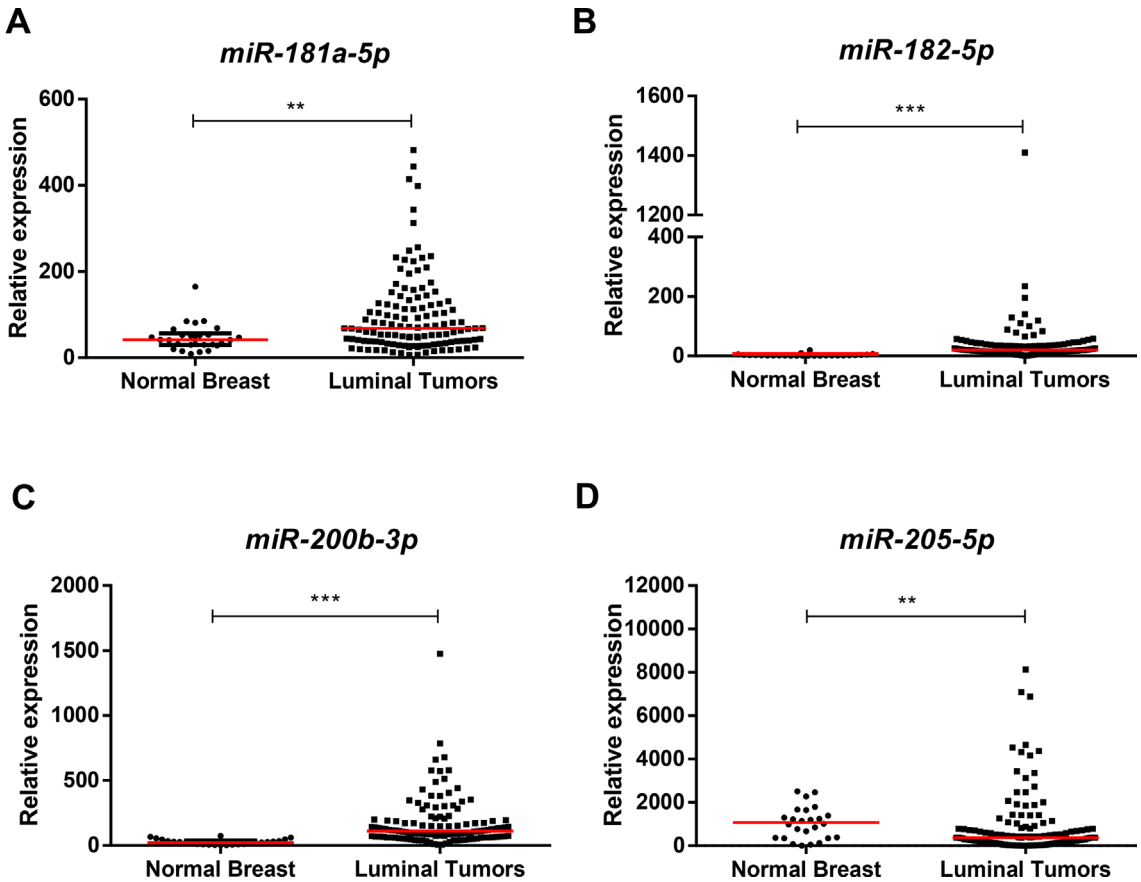


FIGURE 1 | Scatterplots of *miR-181a-5p* **(A)**, *miR-182-5p* **(B)**, *miR-200b-3p* **(C)**, and *miR-205-5p* **(D)** relative expression levels in luminal tumor tissues and normal breast tissues. A ** denotes *p* value <0.01 and a *** denotes *p* value <0.001 by non-parametric Mann–Whitney *U* test. *Y*-axis denotes $2^{-\Delta\Delta CT}$ values multiplied by 1000. Red horizontal lines represent median value.

miR-182-5p, *miR-200b-3p*, and *miR-205-5p* expression levels associated with decreased DFS in univariable analysis (**Table 5**, **Figure 5**). Nonetheless, in the multivariable model, only *miR-30c-5p*, *miR-200b-3p*, and *miR-182-5p* were disclosed as independent prognostic predictors adjusted for Ki-67 *index* (**Table 5**), and after stratification according for Ki-67 *index*, all miRNAs retained statistical significance in high Ki-67 *index* BrC patients (**Table 6**). Similarly, HER2 positivity (HR = 2.63, $p = 0.024$), high Ki-67 *index* (HR = 2.48, $p = 0.021$), and high grade (G3) (HR = 2.69, $p = 0.007$) associated with worse DMFS, along with lower *miR-182-5p* and *miR-200b-3p* expression levels, in univariate analysis (**Table 5**). However, only *miR-182-5p* retained statistical significance when adjusted for tumor grade in multivariable analysis (**Table 5**). After stratification by tumor grade, *miR-182-5p* showed prognostic value in patients harboring low/intermediate-grade tumors (**Table 6**).

Furthermore, the prognostic value of the miRNAs that individually predicted ERFS and DFS was assessed when combined in panels. For ERFS, the patients were grouped as expression above P25 for 3 or 4 miRNAs versus expression below P25 for 2 or more miRNAs. Thus, the combination of m*iR-30c-5p*, *miR-30b-5p*, *miR-182-5p*, and *miR-200b-3p* was shown as the best predictors of ERFS. Patients with miRNAs' expression below P25 displayed a shorter ERFS ($p < 0.001$), paralleling the results obtained in single miRNAs analysis (**Figure 6**, **Table 7**). In multivariable analysis, miRNAs combined in panel were found to be independent ERFS predictors after Ki-67 *index* stratification (**Table 7**). Regarding DFS, the best predictive panel was composed of *miR-182-5p* and *miR-200b-3p*. The patients were grouped as expression above P25 for both miRNAs versus expression below P25 for at least one miRNA. Patients with both miRNAs' expression levels above P25 showed longer DFS ($p < 0.001$) (**Figure 6**, **Table 7**). In multivariable analysis, miRNAs combined in panel remained independent DFS predictors, although only in cases with high Ki-67 *index* (**Table 7**).

DISCUSSION

BrC remains the most common malignancy in women and a major cause of morbidity and mortality (Bray et al., 2018). De-escalation of both systemic and local adjuvant treatment, paralleling trends in surgery, is critical to provide patient-tailored treatment and avoid harmful side effects (Hwang, 2014; Senkus

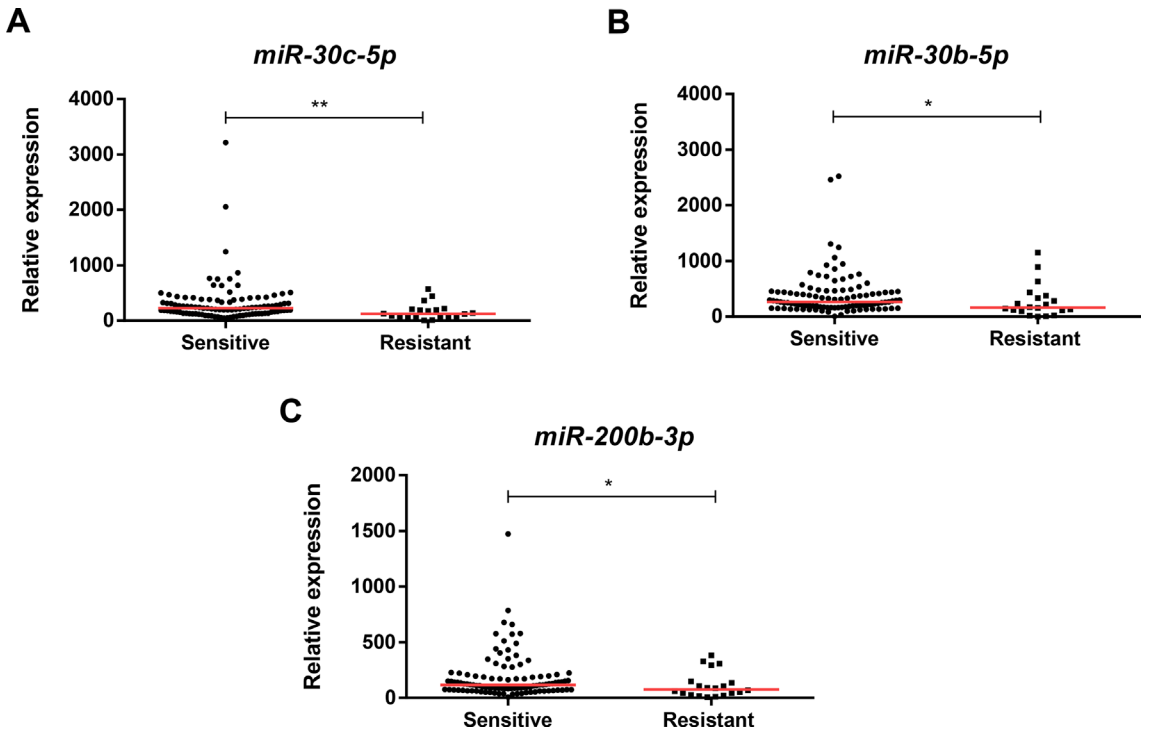


FIGURE 2 | Scatterplots of *miR-30c-5p* **(A)**, *miR-30b-5p* **(B)**, and *miR-200b-3p* **(C)** relative expression levels in tumor tissues from endocrine-sensitive and -resistant patients. A * denotes *p* value <0.05 and a ** denotes *p* value <0.01 by non-parametric Mann–Whitney *U* test. *Y*-axis denotes $2^{-\Delta\Delta CT}$ values multiplied by 1000. Red horizontal lines represent median value.

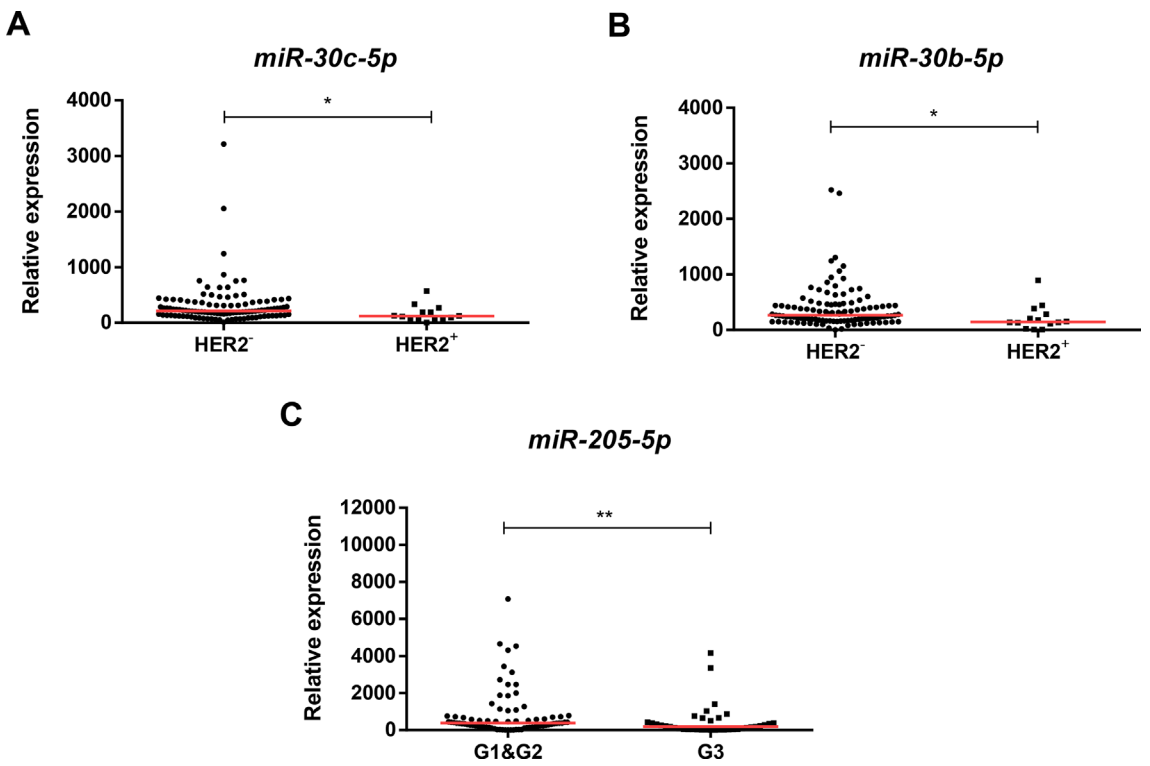


FIGURE 3 | Scatterplots of *miR-30c-5p* relative expression levels according to HER2 *status* **(A)**, *miR-30b-5p* relative expression according to HER2 *status* **(B)**, and *miR-205-5p* relative expression according to grade **(C)**. A * denotes *p* value <0.05 and a ** denotes *p* value <0.01 by non-parametric Mann–Whitney *U* test. Y-axis denotes $2^{-\Delta\Delta CT}$ values multiplied by 1000. Red horizontal lines represent median value.

TABLE 5 | Univariable and multivariable Cox regression models assessing the association between microRNAs expression levels and clinical outcome.

Model	Outcome	Variable	HR (95% CI)	*p* value
Univariate Analysis	**ERFS**	miR-30c-5p expression categorized		0.005
		≤P25	1	
		>P25	0.283 (0.117–0.683)	
		miR-30b-5p expression categorized		0.015
		≤P25	1	
		>P25	0.338 (0.141–0.812)	
		miR-182-5p expression categorized		0.001
		≤P25	1	
		>P25	0.207 (0.082–0.519)	
		miR-200b-3p expression categorized		0.003
		≤P25	1	
		>P25	0.245 (0.098–0.615)	
	DFS	miR-30c-5p expression categorized		0.013
		≤P25	1	
		>P25	0.422 (0.214–0.832)	
		miR-30b-5p expression categorized		0.025
		≤P25	1	
		>P25	0.458 (0.231–0.907)	
		miR-182-5p expression categorized		0.001
		≤P25	1	
		>P25	0.259 (0.120–0.558)	
		miR-200b-3p expression categorized		0.001
		≤P25	1	
		>P25	0.267 (0.127–0.562)	
		miR-205-5p expression categorized		0.043
		≤P25	1	
		>P25	0.494 (0.250–0.979)	
	DMFS	miR-182-5p expression categorized		0.017
		≤P25	1	
		>P25	0.356 (0.153–0.828)	
		miR-200b-3p expression categorized		0.012
		≤P25	1	
		>P25	0.354 (0.158–0.794)	
Multivariate analysis	**ERFS**	miR-30c-5p expression categorized[1]		0.006
		≤P25	1	
		>P25	0.223 (0.076–0.649)	
		miR-30b-5p expression categorized[1]		0.047
		≤P25	1	
		>P25	0.344 (0.120–0.987)	
		miR-182-5p expression categorized[1]		0.002
		≤P25	1	
		>P25	0.174 (0.057–0.529)	
		miR-200b-3p expression categorized[1]		0.003
		≤P25	1	
		>P25	0.184 (0.060–0.561)	
	DFS	miR-30c-5p expression categorized[1]		0.026
		≤P25	1	
		>P25	0.417 (0.193–0.902)	
		miR-182-5p expression categorized[1]		<0.001
		≤P25	1	
		>P25	0.190 (0.078–0.463)	
		miR-200b-3p expression categorized[1]		0.001
		≤P25	1	
		>P25	0.231 (0.094–0.564)	
	DMFS	miR-182-5p expression categorized[2]		0.020
		≤P25	1	
		>P25	0.302 (0.111–0.825)	

[1]Cox regression model adjusted for Ki-67 index. [2]Cox regression models adjusted for grade. ERFS, endocrine resistance-free survival; DFS, disease-free survival; DMFS, distant metastasis-free survival.

et al., 2015). Indeed, identification of luminal BrC patients with low recurrence risk after or while on ET, for which additional adjuvant systemic treatment can be safely omitted, is very important. On the other hand, identification of high-risk luminal BrC patients requiring more aggressive treatment regimens is critical to avoid recurrence and subsequent metastatic disease, which currently affects approximately 40% of luminal BrC patients after adjuvant ET (Guarneri and Conte, 2004; Normanno et al., 2005; Murphy

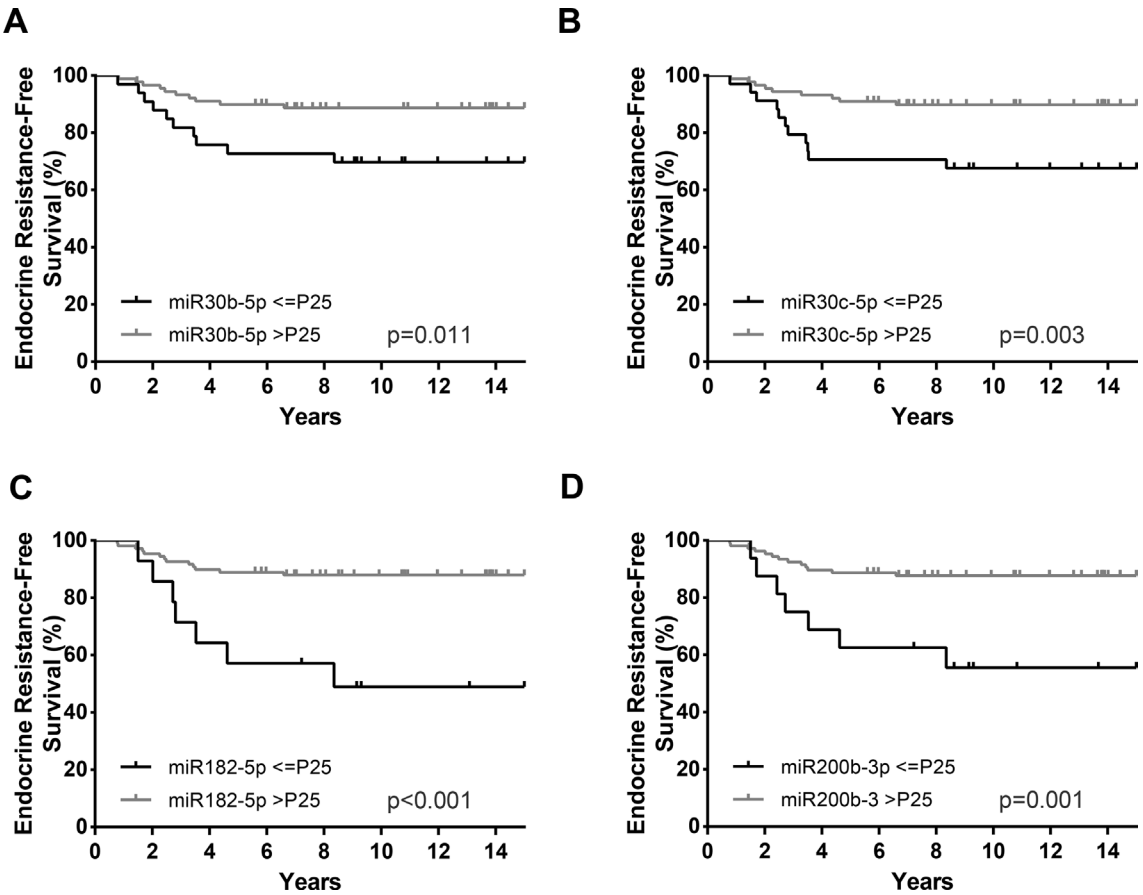


FIGURE 4 | Endocrine resistance-free survival (ERFS) curves (Kaplan–Meier with log rank test) of *miR-30b-5p* **(A)**, *miR-30c-5p* **(B)**, *miR-182-5p* **(C)**, and *miR-200b-3p* **(D)**. P25, Percentile 25.

and Dickler, 2016). Thus, identification of biomarkers providing predictive and prognostic information in this group of patients is clinically relevant. Assessment of specific miRNAs' expression deregulation, which has been associated with several mechanisms underlying endocrine resistance and sensitivity (Muluhngwi and Klinge, 2015; Muluhngwi and Klinge, 2017), might provide such kind of information. Nonetheless, most of those studies have been performed in cancer cell lines and display several limitations, including absence of epithelial–stromal and tumor–host interactions, that could modulate sensitivity *in vivo* (Shekhar et al., 2003). Conversely, tissue analysis from patients treated with ET may allow for broader insight into biologically and clinically relevant miRNAs that may serve as markers of response or resistance to ET. Thus, we focused on the identification of aberrantly expressed miRNAs in endocrine-resistant BrC, exploring its predictive and prognostic value in luminal BrC patients treated with adjuvant ET.

The first step of this study consisted on the profiling of miRNAs' expression patterns, looking for differences between endocrine-sensitive and endocrine-resistant luminal BrC. Hence, *miR-30c-5p*, *miR-30b-5p*, *miR-181a-5p*, *miR-182-5p*, *miR-200b-3p*, and *miR-205-5p* were selected for validation in a larger set of luminal BrC and normal breast tissues. Upregulation of *miR-181a-5p* and *miR-182-5p* and downregulation of *miR-205-5p* in this BrC tissue cohort was consistent with previous reports (Hui et al., 2009; Li et al., 2014a; Zhang and Fan, 2015), providing indirect validation of our methodological approach. However, *miR-200b-3p* downregulation in tumor compared to normal tissues has been previously reported (Ye et al., 2014; Yao et al., 2015). Nevertheless, these studies have used non-cancerous tissues from breasts harboring carcinoma as controls, which may not represent truly normal breast tissues. Our results also confirm the biomarker potential of *miR-30c-5p*, which was found downregulated in endocrine-resistant BrC patients and independently predicted ERFS in luminal BrC patients, particularly in highly proliferative tumors. Moreover, *miR-30c-5p* expression correlated with HER2 *status*, one of the most important predictive factors for ET sensitivity (Konecny et al., 2003). In fact, HER2 signaling activation has been widely implicated in endocrine resistance (Moon et al., 2011; AlFakeeh and Brezden-Masley, 2018). Moreover, *miR-200b-3p* expression levels displayed the same trend and, together with *miR-30b-5p* and *miR-182-5p*, also independently predicted ERFS in luminal BrC patients. Importantly, we were able to validate in primary BrC the association between *miR-200b-3p*

TABLE 6 | Cox regression models stratified according to the clinicopathological features with statistical significance in the multivariate analysis.

Outcome	Layering Variable	Variable	HR (95% CI)	*p* value
ERFS	Ki-67 index <15%	miR-30c-5p expression categorized		0.175
		≤P25	–	
		>P25		
	Ki-67 index >15%	miR-30c-5p expression categorized		0.007
		≤P25	1	
		>P25	0.171 (0.047–0.619)	
	Ki-67 index <15%	miR-30b-5p expression categorized		0.028
		≤P25	1	
		>P25	0.149 (0.027–0.813)	
	Ki-67 index >15%	miR-30b-5p expression categorized		0.334
		≤P25	–	
		>P25		
	Ki-67 index <15%	miR-182-5p expression categorized		0.537
		≤P25	–	
		>P25		
	Ki-67 index >15%	miR-182-5p expression categorized		0.003
		≤P25	1	
		>P25	0.137 (0.037–0.503)	
	Ki-67 index <15%	miR-200b-3p expression categorized		0.610
		≤P25	–	
		>P25		
	Ki-67 index >15%	miR-200b-3p expression categorized		0.002
		≤P25	1	
		>P25	0.121 (0.033–0.447)	
DFS	Ki-67 index <15%	miR-30c-5p expression categorized		0.247
		≤P25	–	
		>P25		
	Ki-67 index >15%	miR-30c-5p expression categorized		0.020
		≤P25	1	
		>P25	0.268 (0.088–0.815)	
	Ki-67 index <15%	miR-182-5p expression categorized		0.141
		≤P25	–	
		>P25		
	Ki-67 index >15%	miR-182-5p expression categorized		0.003
		≤P25	1	
		>P25	0.137 (0.037–0.503)	
	Ki-67 index <15%	miR-200b-3p expression categorized		0.202
		≤P25	–	
		>P25		
	Ki-67 index >15%	miR-200b-3p expression categorized		0.002
		≤P25	1	
		>P25	0.121 (0.033–0.447)	
DMFS	Grades 1 and 2	miR-182-5p expression categorized		0.022
		≤P25	1	
		>P25	0.255 (0.079–0.823)	
	Grade 3	miR-182-5p expression categorized		0.076
		≤P25	–	
		>P25		

ERFS, endocrine resistance-free survival; DFS, disease-free survival; DMFS, distant metastasis-free survival; HER2, human epidermal growth factor 2 receptor.

and endocrine-resistance, which was previously reported only in *in vitro* models (Manavalan et al., 2013). Interestingly, several members of *miR-30f* have been reported as markers of favorable prognosis in BrC (Cheng et al., 2012; Bockhorn et al., 2013; Zhang et al., 2014b; D'aiuto et al., 2015; Croset et al., 2018) and our study also revealed that *miR-30b-5p* might be predictive of response to ET. Finally, concerning *miR-182-5p*, our results extended previous observations on the correlation with clinical benefit from therapy with tamoxifen in advanced-stage BrC, only previously demonstrated in univariable analysis (Rodriguez-Gonzalez et al., 2011).

In addition to their predictive value, *miR-30b-5p* and *miR-30c-5p* lower expression levels also associated with decreased DFS, although in univariable analysis only. Indeed, the role of m*iR-30f* members as tumor suppressors in BrC has been previously reported (Bockhorn et al., 2013; Zhang et al., 2014b). Furthermore, decreased levels of *miR-30f* members in BrC patients have been associated with poor relapse-free survival (Croset et al., 2018). Importantly, lower *miR-182-5p* and *miR-200b-3p* expression levels independently associated with decreased DFS in highly proliferative tumors. The role of *miR-200b-3p* as a prognostic marker in BrC is not a novelty (Ye et al., 2014; Yao et al., 2015). Indeed, members of *miR-200f* are known to act as

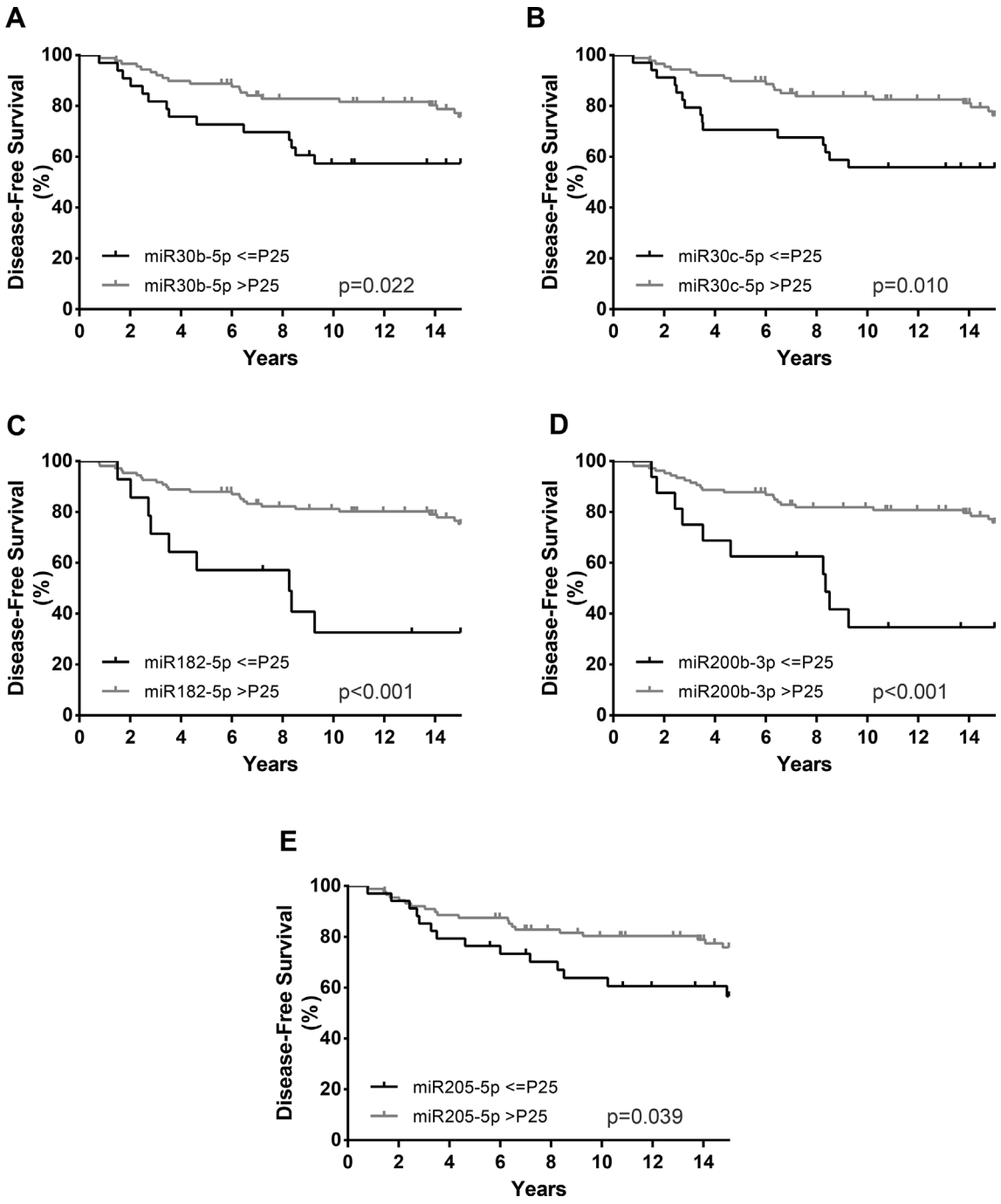


FIGURE 5 | Disease-free survival (DFS) curves (Kaplan–Meier with log rank test) of *miR-30b-5p* **(A)**, *miR-30c-5p* **(B)**, *miR-182-5p* **(C)**, *miR-200b-3p* **(D)**, and *miR-205-5p* **(E)**. P25, Percentile 25.

enforcers of epithelial phenotype through either Zinc finger E-box-binding homeobox (ZEB)-dependent or -independent pathways (Li et al., 2014b). Intriguingly, most *in vitro* studies consistently attributed an oncogenic role to *miR-182-5p* (Chiang et al., 2013; Zhan et al., 2017). Nonetheless, higher *miR-182-5p* expression levels were associated with poor clinical outcome in BrC patients (Song et al., 2016), contrarily to our findings. It should be recalled, however, that *miR-182-5p* is a member of a miRNA family comprising three homologous, coordinately expressed, miRNAs (*miR-183*, *miR-182*, and *miR-196*), which are clustered in chromosome 7q32.2 and that members of this cluster have been associated with both pro- and anti-metastatic behavior in BrC, suggesting that *miR-183/96/182* cluster members may have divergent functions that are regulated in a context- and tissue-dependent manner (Lowery et al., 2010;

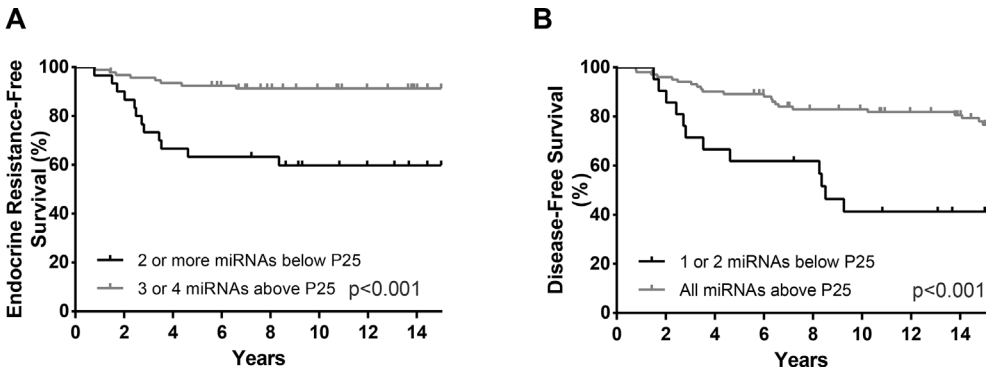


FIGURE 6 | ERFS curves (Kaplan–Meier with log rank test) of combined *miR-30b-5p*, *miR-30c-5p*, *miR-182-5p* and *miR200b-3p* **(A)** and DFS curves (Kaplan–Meier with log rank test) of combined *miR-182-5p* and *miR-200b-3p* panel **(B)**.

TABLE 7 | Univariable and multivariable Cox regression models assessing the association between combined microRNAs expression panel and clinical outcome.

Model	Outcome	Layering Variable	Variable	HR (95% CI)	*p* value
Univariate Analysis	ERFS	NA	**Combined miRNA panel**		<0.001
			2 or more miRNAs below P25	1	
			3 or 4 miRNAs above P25	0.183 (0.075–0.448)	
	DFS	NA	**Combined miRNA panel**		<0.001
			1 or 2 miRNAs below P25	1	
			All miRNAs above P25	0.283 (0.139–0.575)	
Multivariate Analysis	ERFS	NA	**Combined miRNA panel** [1]		<0.001
			2 or more miRNAs below P25	1	
			3 or 4 miRNAs above P25	0.126 (0.042–0.380)	
	DFS	NA	**Combined miRNA panel** [1]		<0.001
			1 or 2 miRNAs below P25	1	
			All miRNAs above P25	0.205 (0.088–0.476)	
Multivariate Analysis	ERFS	Ki-67 *index* <15%	**Combined miRNA panel**		0.018
			2 or more miRNAs below P25	1	
			3 or 4 miRNAs above P25	0.129 (0.024–0.703)	
		Ki-67 *index* >15%	**Combined miRNA panel**		0.002
			2 or more miRNAs below P25	1	
			3 or 4 miRNAs above P25	0.134 (0.037–0.485)	
	DFS	Ki-67 *index* <15%	**Combined miRNA panel**		0.184
			1 or 2 miRNAs below P25	–	
			All miRNAs above P25		
		Ki-67 *index* >15%	**Combined miRNA panel**		0.001
			1 or 2 miRNAs below P25	1	
			All miRNAs above P25	0.100 (0.027–0.367)	

[1]*Cox regression model adjusted for Ki-67 index. ERFS, endocrine resistance-free survival; DFS, disease-free survival; NA, not applicable.*

Li et al., 2014a; Hong et al., 2016). Furthermore, the 7q32.2 locus has been considered a metastasis suppressor locus, enduring genetic copy number losses in BrC progression (Png et al., 2011). Thus, the association between *miR-182-5p* downregulation and worse prognosis probably results from a complex molecular scenario and additional studies are required to discriminate which members of the *miR-183/96/182* cluster might contribute and to which extent to BrC prognosis.

BrC tissues displayed higher *miR-182-5p* and *miR-200b-3p* levels compared to normal breast, although *miR-182-5p* and *miR-200b-3p* downregulation associated with shorter DMFS. Because development of solid neoplasms results from multiple sequential steps in which malignant cells undergo widespread modifications allowing for successful migration and colonization of other organs, we are tempted to speculate whether a context-dependent role of these miRNAs might contribute to the emergence of a malignant phenotype. Indeed, decreased expression of *miR-200f* members might be associated with EMT initiation, enabling cells with invasive capabilities, whereas subsequent upregulation might be associated with mesenchymal-to-epithelial transition, facilitating colonization (Gravgaard et al., 2012; Hilmarsdottir et al., 2014).

Combined expression levels of *miR-30c-5p*, *miR-30b-5p*, *miR-182-5p*, and *miR-200b-3p* independently predicted ERFS, when adjusted for confounding factors (Ki-67 *index*). In fact, this combined miRNA panel was associated with ERFS in both low and highly proliferative tumors. In parallel, the *miR-182-5p*/

miR-200b-3p panel was shown to independently predict DFS in highly proliferative tumors. As previously reported in different tumor models, the combination of miRNAs in a panel might enable a more efficient diagnostic, predictive, and prognostic model overcoming the questionable value of single miRNAs (Sahlberg et al., 2015; Chen et al., 2018).

Although the retrospective design of the study and the relatively small number of samples of the discovery cohort constitute important limitations, our results suggest that a panel of miRNAs might be tested in primary tumor tissues to assess the likelihood of recurrence and resistance to ET in newly diagnosed luminal BrC. Nevertheless, these miRNAs need to be carefully validated, ideally in multicenter studies, to generate more conclusive results. Furthermore, *in vitro* studies, including gain- and loss-of-function assays following *in vitro* treatment with ET, are also critical to functionally characterize the role of these miRNAs. As a future perspective, we intend to evaluate the putative role of these miRNAs in tumor progression and dissemination. Additionally, we also intend to evaluate the potential role of these miRNAs in liquid biopsies, evaluating their potential as non-invasive biomarkers. Indeed, miRNAs in circulation would enable the repeated noninvasive monitoring of miRNA expression profile changes during treatment's course, which could allow for early detection of ET resistance and/or recurrence, potentially improving the management and care of luminal BrC patients.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Comissão de ética para a Saúde of Portuguese Oncology Institute of Porto, Portugal (CES-IPOFG-120/015) with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Comissão de ética para a Saúde of Portuguese Oncology Institute of Porto, Portugal.

AUTHOR CONTRIBUTIONS

MA prepared tissues for molecular analyses, including RNA extraction and cDNA synthesis, performed RT-qPCR assays, analyzed data, and drafted the manuscript. JL and NC collected normal breast tissues from reductive mammoplasty and assisted in histopathological evaluation of tissue samples. HE-P and SS contributed in data analysis and in the manuscript preparation. MF-S and SPS collected clinical follow-up data. PL performed IHC of all cases. LA assisted in the statistical analyses. RH performed histopathological evaluation of fresh frozen sections stained by H&E. RH and CJ designed and supervised the study and revised the manuscript. All the authors read and approved the final manuscript.

FUNDING

This work was supported by a grant from the Research Center of Portuguese Oncology Institute—Porto (PI 74-CI-IPOP-19-2016) and the Portuguese Society of Oncology—YOuR Project. SS is supported by a PhD fellowship IPO/ESTIMA-1 NORTE-01-0145-FEDER-000027. JL is supported by a PhD fellowship from FCT—Fundação para a Ciência e Tecnologia (SFRH/BD/132751/2017).

ACKNOWLEDGMENTS

The authors would like to acknowledge the IPO Porto's patients for their generous collaboration in providing the samples used in this study as well as to the Breast Cancer Clinic staff for their assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00815/full#supplementary-material

REFERENCES

AlFakeeh, A., and Brezden-Masley, C. (2018). Overcoming endocrine resistance in hormone receptor–positive breast cancer. *Curr. Oncol.* 25 (Suppl 1), S18. doi: 10.3747/co.25.3752

Amorim, M., Salta, S., Henrique, R., and Jerónimo, C. (2016). Decoding the usefulness of non-coding RNAs as breast cancer markers. *J. Transl. Med.* 14 (1), 265. doi: 10.1186/s12967-016-1025-3

Barbano, R., Pasculli, B., Rendina, M., Fontana, A., Fusilli, C., Copetti, M., et al. (2017). Stepwise analysis of MIR9 loci identifies miR-9-5p to be involved in oestrogen regulated pathways in breast cancer patients. *Sci. Rep.* 7, 45283. doi: 10.1038/srep45283

Bockhorn, J., Dalton, R., Nwachukwu, C., Huang, S., Prat, A., Yee, K., et al. (2013). MicroRNA-30c inhibits human breast tumour chemotherapy resistance by regulating TWF1 and IL-11. *Nat. Commun.* 4, 1393. doi: 10.1038/ncomms2393

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68 (6), 394–424. doi: 10.3322/caac.21492

Breunig, C., Pahl, J., Küblbeck, M., Miller, M., Antonelli, D., Erdem, N., et al. (2017). MicroRNA-519a-3p mediates apoptosis resistance in breast cancer cells and their escape from recognition by natural killer cells. *Cell Death Dis.* 8 (8), e2973. doi: 10.1038/cddis.2017.364

Burk, U., Schubert, J., Wellner, U., Schmalhofer, O., Vincan, E., Spaderna, S., et al. (2008). A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. *EMBO Rep.* 9 (6), 582–589. doi: 10.1038/embor.2008.74

Cardoso, F., Costa, A., Senkus, E., Aapro, M., André, F., Barrios, C., et al. (2017). 3rd ESO–ESMO international consensus guidelines for Advanced Breast Cancer (ABC 3). *Breast* 31, 244–259. doi: 10.1016/j.breast.2016.10.001

Chen, Z., Zhan, Y., Chi, J., Guo, S., Zhong, X., He, A., et al. (2018). Using microRNAs as novel predictors of urologic cancer survival: an integrated analysis. *EBioMedicine* 34, 94–107. doi: 10.1016/j.ebiom.2018.07.014

Cheng, C.-W., Wang, H.-W., Chang, C.-W., Chu, H.-W., Chen, C.-Y., Yu, J.-C., et al. (2012). MicroRNA-30a inhibits cell migration and invasion by downregulating vimentin expression and is a potential prognostic marker in breast cancer. *Breast Cancer Res. Treat.* 134 (3), 1081–1093. doi: 10.1007/s10549-012-2034-4

Chiang, C. H., Hou, M. F., and Hung, W. C. (2013). Up-regulation of miR-182 by beta-catenin in breast cancer increases tumorigenicity and invasiveness by targeting the matrix metalloproteinase inhibitor RECK. *Biochim. Biophys. Acta.* 1830 (4), 3067–3076. doi: 10.1016/j.bbagen.2013.01.009

Cittelly, D. M., Das, P. M., Spoelstra, N. S., Edgerton, S. M., Richer, J. K., Thor, A. D., et al. (2010). Downregulation of miR-342 is associated with tamoxifen resistant breast tumors. *Mol. Cancer* 9 (1), 1. doi: 10.1186/1476-4598-9-317

Croset, M., Pantano, F., Kan, C. W. S., Bonnelye, E., Descotes, F., Alix-Panabieres, C., et al. (2018). miRNA-30 family members inhibit breast cancer invasion,

osteomimicry, and bone destruction by directly targeting multiple bone metastasis-associated genes. *Cancer Res.* 78 (18), 5259–5273. doi: 10.1158/0008-5472.CAN-17-3058

Curigliano, G., Burstein, H. J., Winer, E. P., Gnant, M., Dubsky, P., Loibl, S., et al. (2017). De-escalating and escalating treatments for early-stage breast cancer: The St. Gallen International Expert Consensus Conference on the Primary Therapy of Early Breast Cancer 2017. *Ann. Oncol.* 28 (8), 1700–1712. doi: 10.1093/annonc/mdx308

D'aiuto, F., Callari, M., Dugo, M., Merlino, G., Musella, V., Miodini, P., et al. (2015). miR-30e* is an independent subtype-specific prognostic marker in breast cancer. *Br. J. Cancer* 113 (2), 290–298. doi: 10.1038/bjc.2015.206

Eroles, P., Bosch, A., Perez-Fidalgo, J. A., and Lluch, A. (2012). Molecular biology in breast cancer: intrinsic subtypes and signaling pathways. *Cancer Treat. Rev.* 38 (6), 698–707. doi: 10.1016/j.ctrv.2011.11.005

Fitzgibbons, P. L., Bradley, L. A., Fatheree, L. A., Alsabeh, R., Fulton, R. S., Goldsmith, J. D., et al. (2014). Principles of analytic validation of immunohistochemical assays: guideline from the college of american pathologists pathology and laboratory quality center. *Arch. Pathol. Lab. Med.* 138 (11), 1432–1443. doi: 10.5858/arpa.2013-0610-CP

Gravgaard, K. H., Lyng, M. B., Laenkholm, A. V., Sokilde, R., Nielsen, B. S., Litman, T., et al. (2012). The miRNA-200 family and miRNA-9 exhibit differential expression in primary versus corresponding metastatic tissue in breast cancer. *Breast Cancer Res. Treat.* 134 (1), 207–217. doi: 10.1007/s10549-012-1969-9

Guarneri, V., and Conte, P. F. (2004). The curability of breast cancer and the treatment of advanced disease. *Eur. J. Nucl. Med. Mol. Imaging* 31 Suppl 1 (1), S149–S161. doi: 10.1007/s00259-004-1538-5

Haque, R., Ahmed, S. A., Inzhakova, G., Shi, J., Avila, C., Polikoff, J., et al. (2012). Impact of breast cancer subtypes and treatment on survival: an analysis spanning two decades. *Cancer Epidemiol. Biomarkers Prev.* 21 (10), 1848–1855. doi: 10.1158/1055-9965.EPI-12-0474

Hilmarsdottir, B., Briem, E., Bergthorsson, J. T., Magnusson, M. K., and Gudjonsson, T. (2014). Functional role of the microRNA-200 family in breast morphogenesis and neoplasia. *Genes (Basel)* 5 (3), 804–820. doi: 10.3390/genes5030804

Hiscox, S., Jiang, W. G., Obermeier, K., Taylor, K., Morgan, L., Burmi, R., et al. (2006). Tamoxifen resistance in MCF7 cells promotes EMT-like behaviour and involves modulation of beta-catenin phosphorylation. *Int. J. Cancer* 118 (2), 290–301. doi: 10.1002/ijc.21355

Hong, Y., Liang, H., Uzair Ur, R., Wang, Y., Zhang, W., Zhou, Y., et al. (2016). miR-96 promotes cell proliferation, migration and invasion by targeting PTPN9 in breast cancer. *Sci. Rep.* 6, 37421. doi: 10.1038/srep37421

Howell, S. J. (2013). Advances in the treatment of luminal breast cancer. *Curr. Opin. Obstet. Gynecol.* 25 (1), 49–54. doi: 10.1097/GCO.0b013e32835c0410

Hui, A. B., Shi, W., Boutros, P. C., Miller, N., Pintilie, M., Fyles, T., et al. (2009). Robust global micro-RNA profiling with formalin-fixed paraffin-embedded breast cancer tissues. *Lab. Invest.* 89 (5), 597–606. doi: 10.1038/labinvest.2009.12

Hwang, E. S. (2014). Breast conservation: is the survival better for mastectomy? *J. Surg. Oncol.* 110 (1), 58–61. doi: 10.1002/jso.23622

Konecny, G., Pauletti, G., Pegram, M., Untch, M., Dandekar, S., Aguilar, Z., et al. (2003). Quantitative association between HER-2/neu and steroid hormone receptors in hormone receptor-positive primary breast cancer. *J. Natl. Cancer Inst.* 95 (2), 142–153. doi: 10.1093/jnci/95.2.142

Lakhani, S. R. (2012). *WHO classification of tumours of the breast*. International Agency for Research on Cancer, Lyon.

Li, P., Sheng, C., Huang, L., Zhang, H., Huang, L., Cheng, Z., et al. (2014a). MiR-183/-96/-182 cluster is up-regulated in most breast cancers and increases cell proliferation and migration. *Breast Cancer Res.* 16 (6), 1. doi: 10.1186/s13058-014-0473-z

Li, X., Roslan, S., Johnstone, C., Wright, J., Bracken, C., Anderson, M., et al. (2014b). MiR-200 can repress breast cancer metastasis through ZEB1-independent but moesin-dependent pathways. *Oncogene* 33 (31), 4077–4088. doi: 10.1038/onc.2013.370

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2–DDCT method. *Methods* 25 (4), 402–408. doi: 10.1006/meth.2001.1262

Lowery, A. J., Miller, N., Dwyer, R. M., and Kerin, M. J. (2010). Dysregulated miR-183 inhibits migration in breast cancer cells. *BMC Cancer* 10 (1), 502. doi: 10.1186/1471-2407-10-502

Maillot, G., Lacroix-Triki, M., Pierredon, S., Gratadou, L., Schmidt, S., Benes, V., et al. (2009). Widespread estrogen-dependent repression of micrornas involved in breast tumor cell growth. *Cancer Res.* 69 (21), 8332–8340. doi: 10.1158/0008-5472.CAN-09-2206

Manavalan, T. T., Teng, Y., Appana, S. N., Datta, S., Kalbfleisch, T. S., Li, Y., et al. (2011). Differential expression of microRNA expression in tamoxifen-sensitive MCF-7 versus tamoxifen-resistant LY2 human breast cancer cells. *Cancer Lett.* 313 (1), 26–43. doi: 10.1016/j.canlet.2011.08.018

Manavalan, T. T., Teng, Y., Litchfield, L. M., Muluhngwi, P., Al-Rayyan, N., and Klinge, C. M. (2013). Reduced expression of miR-200 family members contributes to antiestrogen resistance in LY2 human breast cancer cells. *PLoS One* 8 (4), e62334. doi: 10.1371/journal.pone.0062334

Moon, Y. W., Park, S., Sohn, J. H., Kang, D. R., Koo, J. S., Park, H. S., et al. (2011). Clinical significance of progesterone receptor and HER2 status in estrogen receptor-positive, operable breast cancer with adjuvant tamoxifen. *J. Cancer Res. Clin. Oncol.* 137 (7), 1123–1130. doi: 10.1007/s00432-011-0976-2

Muluhngwi, P., and Klinge, C. M. (2015). Roles for miRNAs in endocrine resistance in breast cancer. *Endocr. Relat. Cancer* 22 (5), R279–R300. doi: 10.1530/ERC-15-0355

Muluhngwi, P., and Klinge, C. M. (2017). Identification of miRNAs as biomarkers for acquired endocrine resistance in breast cancer. *Mol. Cell. Endocrinol.* 456, 76–86. doi: 10.1016/j.mce.2017.02.004

Murphy, C. G., and Dickler, M. N. (2016). Endocrine resistance in hormone-responsive breast cancer: mechanisms and therapeutic strategies. *Endocr. Relat. Cancer* 23 (8), R337–R352. doi: 10.1530/ERC-16-0121

Network C.G.A. (2012). Comprehensive molecular portraits of human breast tumours. *Nature* 490 (7418), 61–70. doi: 10.1038/nature11412

Normanno, N., Di Maio, M., De Maio, E., De Luca, A., de Matteis, A., Giordano, A., et al. (2005). Mechanisms of endocrine resistance and novel therapeutic strategies in breast cancer. *Endocr. Relat. Cancer* 12 (4), 721–747. doi: 10.1677/erc.1.00857

Oh, D. S., Troester, M. A., Usary, J., Hu, Z., He, X., Fan, C., et al. (2006). Estrogen-regulated genes predict survival in hormone receptor–positive breast cancers. *J. Clin. Oncol.* 24 (11), 1656–1664. doi: 10.1200/JCO.2005.03.2755

Parker, J. S., Mullins, M., Cheang, M. C., Leung, S., Voduc, D., Vickery, T., et al. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* 27 (8), 1160–1167. doi: 10.1200/JCO.2008.18.1370

Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., et al. (2000). Molecular portraits of human breast tumours. *Nature* 406 (6797), 747–752. doi: 10.1038/35021093

Png, K. J., Yoshida, M., Zhang, X. H.-F., Shu, W., Lee, H., Rimner, A., et al. (2011). MicroRNA-335 inhibits tumor reinitiation and is silenced through genetic and epigenetic mechanisms in human breast cancer. *Genes Dev.* 25 (3), 226–231. doi: 10.1101/gad.1974211

Rao, X., Di Leva, G., Li, M., Fang, F., Devlin, C., Hartman-Frey, C., et al. (2011). MicroRNA-221/222 confers breast cancer fulvestrant resistance by regulating multiple signaling pathways. *Oncogene* 30 (9), 1082–1097. doi: 10.1038/onc.2010.487

Rodriguez-Gonzalez, F. G., Sieuwerts, A. M., Smid, M., Look, M. P., Meijer-van Gelder, M. E., de Weerd, V., et al. (2011). MicroRNA-30c expression level is an independent predictor of clinical benefit of endocrine therapy in advanced estrogen receptor positive breast cancer. *Breast Cancer Res. Treat.* 127 (1), 43–51. doi: 10.1007/s10549-010-0940-x

Sahlberg, K. K., Bottai, G., Naume, B., Burwinkel, B., Calin, G. A., Børresen-Dale, A.-L., et al. (2015). A serum microRNA signature predicts tumor relapse and survival in triple-negative breast cancer patients. *Clin. Cancer Res.* 21 (5), 1207–1214. doi: 10.1158/1078-0432.CCR-14-2011

Schwarzenbach, H., Nishida, N., Calin, G. A., and Pantel, K. (2014). Clinical relevance of circulating cell-free microRNAs in cancer. *Nat. Rev. Clin. Oncol.* 11 (3), 145–156. doi: 10.1038/nrclinonc.2014.5

Senkus, E., Kyriakides, S., Ohno, S., Penault-Llorca, F., Poortmans, P., Rutgers, E., et al. (2015). Primary breast cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* 26 Suppl 5, v8–30. doi: 10.1093/annonc/mdv298

Shekhar, M. P., Pauley, R., and Heppner, G. (2003). Host microenvironment in breast cancer development: extracellular matrix-stromal cell contribution to neoplastic phenotype of epithelial cells in the breast. *Breast Cancer Res.* 5 (3), 130–135. doi: 10.1186/bcr580

Slamon, D., Eiermann, W., Robert, N., Pienkowski, T., Martin, M., Press, M., et al. (2011). Adjuvant trastuzumab in HER2-positive breast cancer. *N. Engl. J. Med.* 365 (14), 1273–1283. doi: 10.1056/NEJMoa0910383

Song, C., Zhang, L., Wang, J., Huang, Z., Li, X., Wu, M., et al. (2016). High expression of microRNA-183/182/96 cluster as a prognostic biomarker for breast cancer. *Sci. Rep.* 6, 24502. doi: 10.1038/srep24502

Song, J., Ouyang, Y., Che, J., Li, X., Zhao, Y., Yang, K., et al. (2017). Potential value of miR-221/222 as diagnostic, prognostic, and therapeutic biomarkers for diseases. *Front. Immunol.* 8, 56. doi: 10.3389/fimmu.2017.00056

Sørlie, T. (2004). Molecular portraits of breast cancer: tumour subtypes as distinct disease entities. *Eur. J. Cancer* 40 (18), 2667–2675. doi: 10.1016/j.ejca.2004.08.021

Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci.* 98 (19), 10869–10874. doi: 10.1073/pnas.191367098

Thiantanawat, A., Long, B. J., and Brodie, A. M. (2003). Signaling pathways of apoptosis activated by aromatase inhibitors and antiestrogens. *Cancer Res.* 63 (22), 8037–8050.

Vesuna, F., Lisok, A., Kimble, B., Domek, J., Kato, Y., van der Groep, P., et al. (2012). Twist contributes to hormone resistance in breast cancer by downregulating estrogen receptor-alpha. *Oncogene* 31 (27), 3223–3234. doi: 10.1038/onc.2011.483

Ward, A., Shukla, K., Balwierz, A., Soons, Z., König, R., Sahin, Ö., et al. (2014). MicroRNA-519a is a novel oncomir conferring tamoxifen resistance by targeting a network of tumour-suppressor genes in ER+ breast cancer. *J. Pathol.* 233 (4), 368–379. doi: 10.1002/path.4363

Wei, Y., Lai, X., Yu, S., Chen, S., Ma, Y., Zhang, Y., et al. (2014). Exosomal miR-221/222 enhances tamoxifen resistance in recipient ER-positive breast cancer cells. *Breast Cancer Res. Treat.* 147 (2), 423–431. doi: 10.1007/s10549-014-3037-0

Yang, S. J., Yang, S. Y., Wang, D. D., Chen, X., Shen, H. Y., Zhang, X. H., et al. (2017). The miR-30 family: versatile players in breast cancer. *Tumour Biol.* 39 (3), 1–13. doi: 10.1177/1010428317692204

Yao, Y., Hu, J., Shen, Z., Yao, R., Liu, S., Li, Y., et al. (2015). MiR-200b expression in breast cancer: a prognostic marker and act on cell proliferation and apoptosis by targeting Sp1. *J. Cell. Mol. Med.* 19 (4), 760–769. doi: 10.1111/jcmm.12432

Ye, F., Tang, H., Liu, Q., Xie, X., Wu, M., Liu, X., et al. (2014). miR-200b as a prognostic factor in breast cancer targets multiple members of RAB family. *J. Transl. Med.* 12 (1), 17. doi: 10.1186/1479-5876-12-17

Zhan, Y., Li, X., Liang, X., Li, L., Cao, B., Wang, B., et al. (2017). MicroRNA-182 drives colonization and macroscopic metastasis via targeting its suppressor SNAI1 in breast cancer. *Oncotarget* 8 (3), 4629. doi: 10.18632/oncotarget.13542

Zhang, H., and Fan, Q. (2015). MicroRNA-205 inhibits the proliferation and invasion of breast cancer by regulating AMOT expression. *Oncol. Rep.* 34 (4), 2163–2170. doi: 10.3892/or.2015.4148

Zhang, M. H., Man, H. T., Zhao, X. D., Dong, N., and Ma, S. L. (2014a). Estrogen receptor-positive breast cancer molecular signatures and therapeutic potentials (Review). *Biomed. Rep.* 2 (1), 41–52. doi: 10.3892/br.2013.187

Zhang, N., Wang, X., Huo, Q., Sun, M., Cai, C., Liu, Z., et al. (2014b). MicroRNA-30a suppresses breast tumor growth and metastasis by targeting metadherin. *Oncogene* 33 (24), 3119–3128. doi: 10.1038/onc.2013.286

Zhao, J.-J., Lin, J., Yang, H., Kong, W., He, L., Ma, X., et al. (2008). MicroRNA-221/222 negatively regulates estrogen receptorα and is associated with tamoxifen resistance in breast cancer. *J. Biol. Chem.* 283 (45), 31079–31086. doi: 10.1074/jbc.M806041200.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Amorim, Lobo, Fontes-Sousa, Estevão-Pereira, Salta, Lopes, Coimbra, Antunes, Palma de Sousa, Henrique and Jerónimo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Cell and Developmental Biology

REVIEW
published: 19 September 2019
doi: 10.3389/fcell.2019.00182

Epigenetic Biomarkers in the Management of Ovarian Cancer: Current Prospectives

Alka Singh[1], *Sameer Gupta*[2] *and Manisha Sachan*[1]*

[1] Department of Biotechnology, Motilal Nehru National Institute of Technology, Allahabad, India, [2] Department of Surgical Oncology, King George Medical University, Lucknow, India

OPEN ACCESS

Edited by:
Yun Liu,
Fudan University, China

Reviewed by:
Paola Parrella,
Casa Sollievo Della Sofferenza (IRCCS), Italy
Mariana Brait,
Johns Hopkins University, United States

***Correspondence:**
Manisha Sachan
manishas@mnnit.ac.in;
manishas77@rediffmail.com

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Cell and Developmental Biology

Received: *02 April 2019*
Accepted: *19 August 2019*
Published: *19 September 2019*

Citation:
Singh A, Gupta S and Sachan M (2019) Epigenetic Biomarkers in the Management of Ovarian Cancer: Current Prospectives. Front. Cell Dev. Biol. 7:182. doi: 10.3389/fcell.2019.00182

Ovarian cancer (OC) causes significant morbidity and mortality as neither detection nor screening of OC is currently feasible at an early stage. Difficulty to promptly diagnose OC in its early stage remains challenging due to non-specific symptoms in the early-stage of the disease, their presentation at an advanced stage and poor survival. Therefore, improved detection methods are urgently needed. In this article, we summarize the potential clinical utility of epigenetic signatures like DNA methylation, histone modifications, and microRNA dysregulation, which play important role in ovarian carcinogenesis and discuss its application in development of diagnostic, prognostic, and predictive biomarkers. Molecular characterization of epigenetic modification (methylation) in circulating cell free tumor DNA in body fluids offers novel, non-invasive approach for identification of potential promising cancer biomarkers, which can be performed at multiple time points and probably better reflects the prevailing molecular profile of cancer. Current status of epigenetic research in diagnosis of early OC and its management are discussed here with main focus on potential diagnostic biomarkers in tissue and body fluids. Rapid and point of care diagnostic applications of DNA methylation in liquid biopsy has been precluded as a result of cumbersome sample preparation with complicated conventional methods of isolation. New technologies which allow rapid identification of methylation signatures directly from blood will facilitate sample-to answer solutions thereby enabling next-generation point of care molecular diagnostics. To date, not a single epigenetic biomarker which could accurately detect ovarian cancer at an early stage in either tissue or body fluid has been reported. Taken together, the methodological drawbacks, heterogeneity associated with ovarian cancer and non-validation of the clinical utility of reported potential biomarkers in larger ovarian cancer populations has impeded the transition of epigenetic biomarkers from lab to clinical settings. Until addressed, clinical implementation as a diagnostic measure is a far way to go.

Keywords: biomarker, cell free DNA, diagnosis, DNA methylation, epigenetics, epithelial ovarian cancer

KEYPOINTS

- Prompt diagnosis remains challenging due to non-specific symptoms in the early-stage of the disease, their presentation at an advanced stage and poor survival.
- DNA methylation occurs very early in malignant transformation and their utility as biomarker holds great promise to overcome the false positive detection of ovarian cancer using current standard serum marker CA125.
- Not even a single report has suggested or demonstrated a good epigenetic marker for early and accurate detection of OC in either tissue or fluid. Thus, early detection still remains a huge unmet need. However, analysis of a panel of aberrant methylation based epigenetic markers in blood-based non-invasive assay could pave its way into clinical implementation.

INTRODUCTION

Ovarian cancer, a molecularly heterogeneous disease, remains the most lethal disease among gynecological malignancies. Representing as the third most frequent cancer among female gynecological system carcinoma, ovarian cancer is associated with the highest mortality rates. Despite constituting only 3% of all female cancer, the annual incidence of ovarian cancer worldwide is 220,000 with 21,290 estimated numbers of new cases and 14,600 estimated deaths annually (Siegel et al., 2015). Typical diagnosis of more than 70% of OC cases, at an advanced disease stage is one of the potent reasons for high fatality rate and carries poor prognosis with current therapies. The median age of disease presentation in ovarian cancer is 60 years and its lifetime risk is one in seventy with an overall lifetime mortality of one in ninety five (Cannistra, 2004; Howe et al., 2006).

Epithelial ovarian cancer (E0C) comprises 90% of all forms of OC cases and is characterized by heterogeneity at histopathological, clinical and molecular level. The exact cause for the ovarian malignancy still remains unknown. A strong familiar history either of ovarian or breast cancer has been described as important risk factors associated with OC. More than one-fifth of ovarian carcinomas (about 23%) have hereditary susceptibility and germline mutations of BRCA1 and BRCA2 tumor suppressor genes; in particular contribute to 65–85% of these cases (Ramus et al., 2007). An association of hormonal risk in postmenopausal women is suggested by over 50% of deaths. In addition, parity, pregnancy, lactation, tubal ligation, and oral contraceptive use are associated with reduced risk and have been found to be protective factors against disease development.

Rapid growth, non-specific clinical symptoms at early stage of the disease and lack of early diagnostic methods make prompt diagnosis challenging. As a result, EOC is typically diagnosed at an advanced stage (FIGO III/IV), when the tumor has spread beyond the pelvis and even unlikely to be completely removed by surgery. The long term survival rates for women with disseminated malignancies are low (10–30%). However, diagnosis of ovarian cancer at the localized stage (confinement of lesion still to the ovaries) is highly curable (over 95% 5 year survival rate; Siegel et al., 2011). To improve the overall survival of women diagnosed with EOC and to overcome the non-specific clinical manifestation of EOC, identification of molecular biomarkers of preclinical or early stage EOC tumors is critically needed. A better understanding of EOC genome portrait will help in the identification of promising biomarkers of clinical utility for early diagnosis of OC.

MOLECULAR CLASSIFICATION

The primary OC were classified into epithelial (60%), germ cell (30%), and sex-cord stromal tumors (8%), by the World Health Organization (WHO) classification and tumor morphology system (2014). A large majority of OC, almost 80–85%, are of epithelial origin. However, a small proportion accounting approximately 10% of all OC falls into germ cell and sex-cord stromal tumor categories (Devouassoux-Shisheboran and Genestie, 2015). Further on the basis of disease dissemination, the American Joint committee on Cancer/Tumor Node Metastasis (AJCC/TNM) and International Federation of Gynecology and Obstetrics (FIGO) staging systems, classified ovarian cancer into various stages. The confinement of tumors to the ovaries is represented by stage I and II whereas stage III is associated with local metastasis (usually lymph) and stage IV with distal organ metastases (Yarbro et al., 1999).

EOCs have been further sub-categorized based on following two criteria: (a) firstly, on the degree of proliferation, grade and extent of invasion into Benign (adenoma and cystadenoma), low malignant potential (LMP) and malignant (b) and secondly based on tumor histopathological grade and molecular characteristics, EOC malignant tumors are classified into serous (70%, most common), endometrioid (10–20%), clear cell (12%), mucinous (3%) and less commonly, transitional (6%), squamous, mixed, and undifferentiated (<1%) subtypes (Bowtell, 2010; Devouassoux-Shisheboran and Genestie, 2015; Earp and Cunningham, 2015; **Figure 1)** On the basis of histological type and grade, these tumors exhibit different genetic and epidemiological risk factors, pattern of spread, molecular abnormalities, response to targeted therapies and disease prognosis (Devouassoux-Shisheboran and Genestie, 2015; Earp and Cunningham, 2015).

Almost a decade ago, a dualistic classification system recognized Type I and Type II EOC tumors (Shih and Kurman, 2004; Vang et al., 2009). Type I EOCs are generally low grade serous carcinomas but also include mucinous, endometrioid, and clear cell subtype tumors. They are thought to arise from a low malignant potential precursor, are characterized as slow growing with low levels of chromosomal instability, intact DNA repair machinery and harbor mutations in KRAS, BRAF, and ERBB2 at a high frequency. Type II EOCs arise *de novo* and are comprised of high-grade serous carcinoma. These aggressive tumors also include malignant mixed mesodermal and undifferentiated carcinomas, are characterized by rapid growth with no identified precursor lesions, high levels of chromosomal aberrations along with high frequency of TP53,

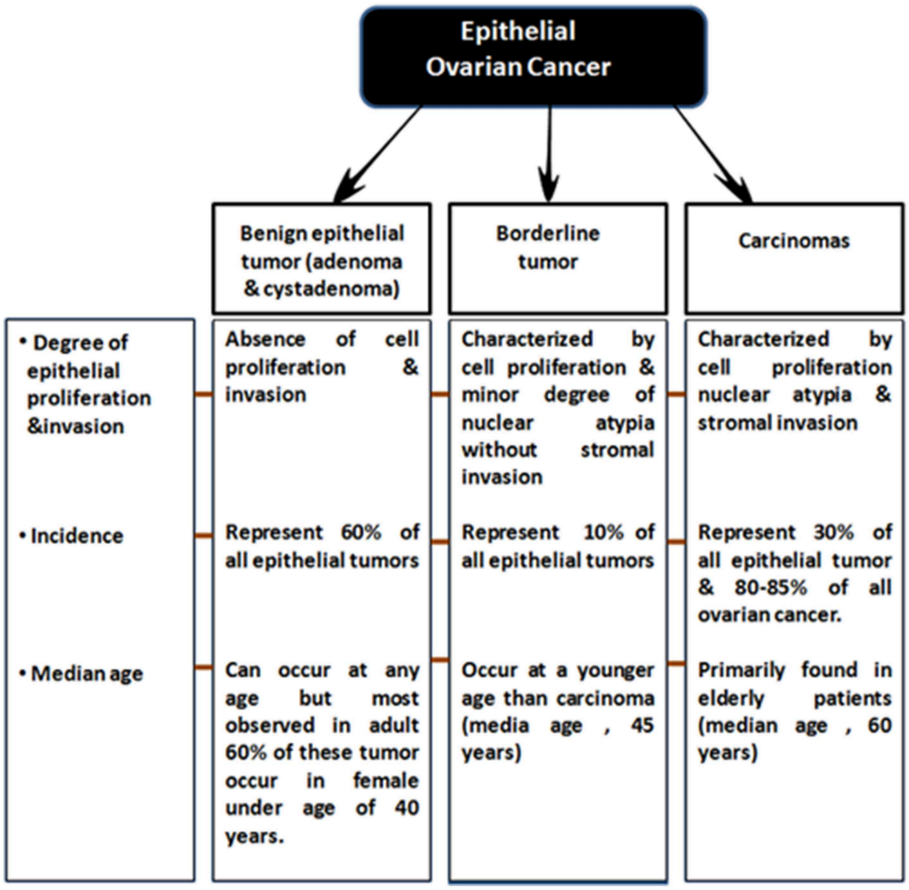


FIGURE 1 | Sub-classification of epithelial ovarian tumors.

BRCA1/2 mutations. They constitute 70% of EOC cases (Jayson et al., 2014; **Figure 2**).

The cells of origin of ovarian cancer are still debated. Two models with respect to the origin of ovarian cancer have been proposed: (1) origin from ovarian surface epithelium (OSE), (2) from the fallopian tube. Taken together, the pro-inflammatory environment due to ovulation events, expression pattern of ovarian inclusion cysts and biomarkers which are shared by OSE and malignant growth, form the basis of first model. On contrary, tubal precursor lesions, genetic evidence of BRCA1/2 mutation carriers and recent studies strongly implicate a non-ovarian origin and form the basis of the later model. To date, neither model has evidently revealed superiority over the other. Thus, it is speculated that the HGSOC could have arisen from two different sites which undergo similar changes and could be a possible reason for tumor heterogeneity (Klotz and Wimberger, 2017). It has also been postulated that aberrantly methylated Mullerian duct cells migrate into ovarian stroma where they are supported by the epigenetically/ genetically altered stromal environment, facilitating a cascade of events which culminate in ovarian carcinogenesis. Epigenetic profiling of endocervical glandular cells would facilitate in prediction of risk or early detection of ovarian cancer (Jones et al., 2010).

SCREENING AND EARLY DETECTION

OC is generally characterized by few non-specific early symptoms, presentation of the disease at a late stage and poor survival. Difficulty to diagnose it in its early stages still remains challenging. Early diagnosis, screening and personalized treatment is still the biggest unmet need to combat this devastating disease. Unavailability of early cancer-specific diagnostic markers and ubiquitous acquisition of drug resistance to targeted therapies are the most striking obstacles for the effective OC treatment.

Clinically, serum antigen-125 (CA125) is the most extensively studied, established and utilized diagnostic marker of EOC, despite its elevation marked by only 47% of early-stage EOC (Woolas et al., 1993). Additionally, aberrantly elevated serum CA125 have been reported in several benign conditions of endometriosis, pregnancy, peritonitis, pelvic inflammatory disease, uterine fibroids, menstrual cycle, liver cirrhosis. Its elevation is also associated with several malignancies such as lung and colorectal cancer (Jacobs and Bast, 1989). Moreover, poor specificity, high false positive rate, and low positive prediction value make CA125 alone unsuitable as an EOC diagnostic marker.

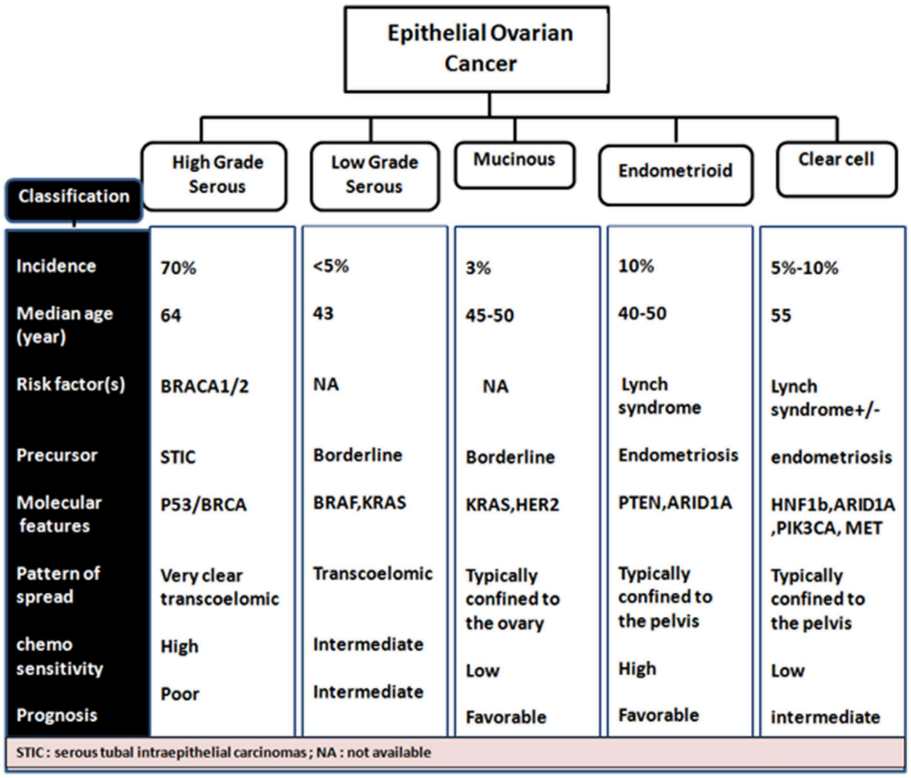


FIGURE 2 | Phenotypic- genotypic classification of epithelial ovarian cancer subtypes.

However, CA125 is more suitable markers for tumor recurrence (Clarke-Pearson, 2009).

For clinical needs to diagnose OC at an early stage, the conventional screening methods such as serum cancer antigen 125 (CA125) concentrations, transvaginal ultrasound probe and magnetic resonance imaging have not shown reliability in reducing population mortality or morbidity due to high false-negatives rates and lower sensitivity and specificity (Menon and Jacobs, 2000; Jacobs and Menon, 2004; Munkarah et al., 2007). Therefore, methods for early detection are critically required. Owing to the low incidence rate of OC amongst postmenopausal women, a logistic diagnostic screening test warrants the need of high sensitivity (>75%) and specificity (>99.6%) to attain a positive prediction value (PPV) of 10%. Novel biomarkers for early-stage diagnosis are being explored and it is more likely that a combination of biomarkers could achieve these required diagnostic criteria (Moore et al., 2010).

To determine the effect of screening on OC mortality, several randomized controlled trial in general population had been undertaken. Recently, both CA125 and transvaginal sonography (TVS) was evaluated in the Prostate, Lung, Colorectal, and Ovarian (PLCO) cancer screening trial, however no significant difference was observed in OC mortality between screening and conventional care arms (Buys et al., 2011). The United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS), being considered as the largest prospective randomized trial, comprised of over 200,000 asymptomatic postmenopausal women who were screened with TVS alone and combined TVS and CA125. Although improvement in specificity of detection was achieved on combining CA125 with TVS, however these trials failed to attain the requisite diagnostic accuracy of 99.6% specificity (Menon et al., 2009). CA125 together with HE4 has somewhat improved sensitivity and specificity of detection which correctly identified 76.4% of cancer samples and 95% of cancer negative samples. This accuracy was notably higher than either marker alone. However further validation is still required (Moore et al., 2010). According to the Guide to Clinical Preventive Services 2010–2011, it has been mentioned that neither of any screening test [serum antigen-125 (CA-125), ultrasound imaging, pelvic examination or any earlier diagnosis methods] was able to improve OC survival rates U. S. Preventive Services Task Force (2010).

The Risk of Malignancy Index, widely used at present, particularly UK, is a score based on ultrasound variables, menopausal status and CA125 (Jacobs et al., 1990). Its sensitivity is the determining criteria for a patient to be sent to experts by referring gynecologist provided objective assessment score is lower (78%) (Geomini et al., 2009). Transvaginal sonography (TVS) is based on a formal scoring model system. Though highly sensitive and being considered as an ideal method for second stage diagnosis, the major limitation associated with this method is its high dependency on individual expertise (Yazbek

et al., 2008). Therefore, in clinical practice to discriminate benign and malignant ovarian tumors is still a significant challenge. The availability of biomarker or their combination which can potentially detect ovarian cancer at its earliest stage with required sensitivity and specificity would help in improving clinical outcomes.

MARKERS FOR OVARIAN CANCER DIAGNOSIS AND MANAGEMENT

Protein Markers

As discussed before, a suitable screening test for OC early stage diagnosis will require high sensitivity and high specificity. Current practices for screening of OC include transvaginal ultrasonography, biomarker analysis, or a combination of both. To date, a number of potential biomarkers for early diagnosis of OC have been identified through intense research in proteomic and genomic. Here, we summarize a comprehensive account of recent researches on explored novel and robust serum based biomarkers for the non-invasive early stage screening of ovarian cancer (**Table 1**).

Although being considered as the "gold standard" biomarker for detection of OC, its clinical relevance mainly falls in evaluating disease recurrence. Other biochemical markers such as lysophosphatidic acid, human epididymis protein 4 (HE4), inhibins (which are members of TGF-β subfamily), Mesothelin (associated with migration and metastasis) (Huang et al., 2006), Osteopontin, and YKL-40 have been reported to be elevated in sera of patients with OC amongst various studies, which could be of diagnostic significance for improved cancer detection, most likely in various combination with one another and /or with CA125 (Rosenthal et al., 2006; Moore et al., 2010). The most promising molecular biomarker of all these, to date are HE4 and Mesothelin. So far, US FDA has only approved CA125 and HE4 for monitoring disease progression/recurrence, but not for screening purpose (Rosenthal et al., 2006).

For the triage of pelvic mass, the multivariate index assay OVA1, constituting measurements of 5-proteins: CA125-II, apolipoprotein A1, transthyretin, beta 2 microglobulin, and transferrin, has been approved by FDA since 2009. Although, the test had improved sensitivity but compromised in revealing diagnostic potential with its low specificity upon replacement of CA125 with the multivariate index assay (Nguyen et al., 2013). Elevated levels of Kallikrein 6 and 7 (KLK6 and KLK7) was reported in sera of ovarian carcinoma subtypes, depicting their potential to improve early detection of OC. Other biomarkers with potential clinical significance for early diagnosis in women with EOC include GSTT1, Prostasin (PRSS8), KLK6, KLK7, FOLR1, and ALDH1, which are currently under research and clinical trials (Sarojini et al., 2012).

Evaluation of several prediagnostic multimarker panels along with PLCO screening trial has identified promising biomarkers which are able to distinguish ovarian cancer cases from normal control groups; for instance, a four biomarker panel consisting of CA-125, HE4, CEA, and VCAM-1 effectively discriminated early stage OC from healthy controls with sensitivity of 86% at 98% specificity (Lin et al., 2009). Another panel constituting of CA-125, ApoA1, TTR, and H418, was able to differentiate OC patients at early stage of disease from cancer-free healthy control samples with 74% sensitivity at 97% specificity (Zhang et al., 2004). Still to date, no panel of biomarkers that has been examined amongst numerous studies could outperform CA125 alone, in distinguishing between the two groups. The sensitivity and specificity of serum based non-invasive biomarkers for improved ovarian cancer detection from various studies as well as the currently active/completed clinical trials evaluating potent biochemical markers of clinical significance for early diagnosis of EOC are summarized in **Tables 2**, **3** respectively.

Genetic Marker

About 23% of ovarian tumors have been associated with hereditary conditions and the genetic abnormalities in about 65–85% of hereditary ovarian carcinomas is the germline mutation in BRCA (breast cancer early onset genes BRCA1 and BRCA2) genes which are essential for DNA repair as well as in maintaining genomic stability and integrity. The cumulative lifetime risk of EOC for a woman with BRCA1 and BRCA2 mutation is 39–46% and 12–20%, respectively (Ramus et al., 2007). Lifetime risk to develop breast cancer and ovarian cancer is enhanced up to 85% and up to 54% respectively in the carriers of BRCA1 and BRCA2 mutations. Association of several tumor suppressor genes and oncogenes (tumor suppressor gene TP53 in Li- Fraumeni syndrome, mismatch repair genes (MMR) in Lynch syndrome, genes involved with double strand break repair system: BARD1, CHEK2, RAD51, and PALB2) with hereditary ovarian cancer has been reported. Till date, around 16 genes have been reported to be associated with hereditary ovarian carcinogenesis while several other mutations are yet unknown and need to be further explored (Toss et al., 2015).

Epigenetic Marker

Epigenetics is the mechanism for the regulation of gene expression without any alternation in the primary DNA sequence (Jones and Laird, 1999; Jones and Baylin, 2002; Feinberg and Tycko, 2004). DNA methylation, modification of histone proteins and miRNAs are the key modulator in regulating several cellular processes such as cell differentiation, embryogenesis, inactivation of X chromosome, genome imprinting, and many others (Jones, 2001; Reik and Lewis, 2005; Kacem and Feil, 2009; Portela and Esteller, 2010). The epigenetic alternations involve interplay between DNA methylation, histone modification and micro RNA expression to modulate gene expression during development and cancer progression. (1) The global hypomethylation, largely of repetitive DNA which results in demethylation of several oncogenes and (2) localized hypermethylation at promoters of various tumor suppressor genes leading to their transcriptional silencing, are two opposite epigenetic phenomenon involved in tumorigenesis (Sharma et al., 2010). DNA methyltransferase (DNMT) mediated methylation of deoxycytosine located within the CpG dinucleotides is the best known and widely studied epigenetic mechanism leading to transcription repression in cancer (Bird and Wolffe, 1999; Hendrich and Bird, 2000; Bird,

TABLE 1 | Novel tumor biochemical markers for early detection of ovarian cancer.

Biochemical marker	Nature	Origin/function	Status	Source	Clinical utility	References
CA125	Glycoprotein antigen	Expressed by fetal amniotic epithelium and coelomic epithelium	Elevated Early stage: 47% Advanced stage: >90%	Serum	Valuable marker for tumor recurrence Limitations: Unsuitable for early detection due to insufficient sensitivity and being elevated in other conditions	Clarke-Pearson, 2009; Moore et al., 2010
HE4	Protease	Serum maturation	Elevated Serous: 93% Endometrioid: 100%	Serum	Histotype differentiation and screening	Hough et al., 2000; Lu et al., 2004; Moore et al., 2008
Mesothelin	Glycosylphosphatidylinositol-linked cell surface molecule	*Expressed by mesothelial cells *Involved in metastasis	Elevated Early stage: 60%	Urine	Early screening	McIntosh et al., 2004; Tang et al., 2013
Transthyretin	An acute phase reactant and major carrier of serum thyroxine	Tumor development	Downregulated in EOC patients	Serum	Early stage screening	Mählck and Grankvist, 1994; Schweigert and Sehouli, 2005; Nosov et al., 2009
ApoA1	Constituent of high density lipoproteins	Prevents tumor development	Downregulated in ovarian cancer patients	Serum	Early stage screening	Gadomska et al., 2005; Kim et al., 2012
Kallikrein	Family of Serine proteases *Human KLK family: 15 members *Chromosome position: 19q13.3–4	Regulates proteolytic cascades	Elevated: 12KLK/15	Serum	Elevated KLK-6 and−10 in OC cases with low level of CA125. Useful marker for OC detection	Borgoño and Diamandis, 2004; Rosen et al., 2005
Osteopontin	An adhesive glycoprotein	*Synthesized by osteoblasts and vascular endothelial cells *Associated with bone remodeling and immune function	Elevated in invasive and borderline ovarian cancer tumors	Plasma	Early stage screening	Kim et al., 2002

2002). DNA methylation is known to be the earliest event during carcinogenesis and plays a crucial role in silencing of tumor suppressor genes (Sharma et al., 2010; Teschendorff and Widschwendter, 2012; Teschendorff et al., 2012, 2016; Bartlett et al., 2016). Promoter methylation mediated epigenetic silencing of gene is regulated by the recruitment of MBD (methyl CpG binding proteins such as MeCP2, MBD1, MBD2, and MBD4) which in turn regulates chromatin state by recruiting histone modifying and chromatin-remodeling complexes (repressors) at the site of methylation, which subsequently generates condensed chromatin structure and results in transcriptional repression (Esteller, 2007; Lopez-Serra and Esteller, 2008). On contrary, epigenetic activation of gene is regulated by recruitment of Cfp1 and histone methyltransferase Setd1 which aids in generating an open chromatin structure by creating domains which are enriched with active histone marks (acetylation and H3K4 trimethylation) (Thomson et al., 2010, p. 1; Jones and Baylin, 2007; **Supplementary Figure 1**). Increasing evidences has revealed the significant role of DNA methylation in cancer development and progression, right from transcriptional silencing of tumor suppressor genes to the activation of oncogenes and consequently promoting metastasis (Costello and Plass, 2001; Herman and Baylin, 2003; Wilting and Dannenberg, 2012). Apparently, it is quite evident now that DNA methylation plays an equal or possibly even greater role than the genetic lesion such as mutations, deletion and translocations which have been associated for long, with malignant transformations and carcinogenesis (Chan T. A. et al., 2008). For instance, though the familial breast cancer susceptibility gene 1 (BRCA1) mutations contributes to 5–10% of EOC, promoter hypermethylation of non-mutated BRCA1 allele is the second disruptive event to the development of this cancer (Barton et al., 2008).

Tissue Biomarkers

Diagnosis

So far, several methylation based signatures have been reported in EOC. Here, we summarize an overview of some of the extensively studied potential biomarkers of diagnostic utility in ovarian cancer (**Table 6**). In ovarian cancer, a large number of tumor suppressor genes have been identified to be silenced by promoter hypermethylation and downregulated includes DAPK, LOT1, TMS1/ASC, and PAR4 (pro-apoptotic function and cell cycle regulation), p16, SPARC, ANGPTL2, and CTGF (tumor suppressor activity), ICAM-1 and CDH1 (cell adhesion), PEG31 (role in imprinting) and many others (**Tables 4**, **5**). In

TABLE 2 | Specificity and sensitivity of early detection biomarkers for ovarian cancer from various studies.

Biochemical biomarker	Source	Population tested	Clinical prediction		References
			Sensitivity	Specificity	
HE4	Serum	147 Cancer (111 ovarian cancer cases), 285 Benign, 66 controls	79.3%	98.9%	Molina et al., 2011
HE4 + CA125	Serum	383 Benign and 89 Cancer	100% 92.30%	74.20% 76.0 %	Moore et al., 2011
Osteopontin	Plasma	46 Benign, 47 cancer, 51 ovarian cancer, 107 controls	-	-	Kim et al., 2002
Prostasin + CA125	Serum	64 cancer, 137 control	92%	94%	Mok et al., 2001
KLK6	Serum	141 Benign, 146 ovarian cancer, 97 controls	21–26%	95%	Diamandis et al., 2003
KLK6+ CA125	Serum	141 Benign, 146 ovarian cancer, 97 controls	42%	90%	Diamandis et al., 2003
B7-H4	Serum	1,023 cancer, 997 Benign (236 ovarian cancer cases, 260 controls)	45%	97%	Simon et al., 2006
B7-H4 + CA125	Serum	1,023 cancer, 997 Benign (236 ovarian cancer cases, 260 controls)	65%	97%	Simon et al., 2007
CA125/IL-6/IL-8/VEGF/EGF	Serum	44 Early-stage cancers 37 Benign, 45 controls	84%	95%	Gorelik, 2005
CA125/IL-6/G-CSF/VEGF/EGF	Serum	44 Early-stage cancers 37 Benign, 45 controls	86.5%	93%	Gorelik, 2005
CA125/HE4/Glycodelin/ PLAUR/MUC1/PAI-1	Serum	200 Cancers (133 stage I/II), 396 Healthy controls	80.5%	96.5%	Havrilesky et al., 2008
Leptin/Prolactin/ Osteopontin/IGF2	Serum	100 Cancers, 106 controls	95%	94%	Mor et al., 2005
CA125 HE4 Mesothelin	Serum	143 Cancers, 124 benign, 344 controls	78% 68–82% 31–44%	98% 98% 98%	Shah et al., 2009
Leptin/Prolactin/ Osteopontin/IGF2/MIF/ CA125	Serum	Training: 113 cancers, 181 controls Test: 43 cancers, 181 Controls	95.3%	99.4%	Visintin et al., 2008
CA125/ CA19- 9 /EGFR /CRP/ Myoglobin/APOA1/ APOC3/MIP1A/ IL-6/IL-18/ Tenascin C	Serum	115 Cancers, 93 benign 24 Controls, 13 non-ovarian cancers	91.3%	88.5%	Amonkar et al., 2009
CA-125, HE4, SI	Serum	74 cancer, 137 controls	84%	98.5%	Andersen et al., 2010
RIM, ROMA, CA-125, HE4,	Serum	445 Benign, 31 borderline, and 162 malignant tumors	Postmenopausal 89, 91, 92, and 72% Premenopausal 87, 87, 96, and 83%	Postmenopausal 80, 77, 80, and 92% Premenopausal 90, 81, 60, 91%	Lycke et al., 2018
CA-125 HE4 TTR	Serum	130 Benign, 126 ovarian cancer, 400 controls	64.29% 46.4% 78.6%	53.57% 43.3% 68.8%	Zheng et al., 2018
CA-125, HE4, TK1	Serum	75 ovarian cancer, 40 Benign, 35 controls	94.18%	79.53%	Xi et al., 2017
CA-125+HE4, HE4+FOLR1	Serum	150 benign controls, 216 ovarian cancer, 20 controls	67% 65%	95% 95%	Leung et al., 2016
CA-125, ApoA1, TTR	Serum	200 cancer, 142 controls	74%	97%	Zhang et al., 2004
CA125, HE4, MMP-7, CA72-4	Serum	142 stage I cancer, 217 controls	83.2%	98%	Simmons et al., 2016
CA-125, CA 72–4, MCSF	Plasma	123 cancer, 224 controls	70%	98%	Edgell et al., 2010
CA-125, TTR, ApoA1	Serum	20 cancer, 82 controls	89%	92%	Su et al., 2007
CA-125, HE4, CEA, VCAM-1	Serum	456 cancer, 2,000 controls	86–93%	98%	Lin et al., 2009
CA-125, CRP, SAA, IL-6, IL-8	Plasma	150 cancer, 212 controls	94.1%	91.3%	Edgell et al., 2010
S100A4	Serum	160 cancer, 52 Benign, 52 controls	78%	92%	Lv et al., 2018
KPNA2	Serum	162 cancer, 48 controls	71.4%	81.2%	Huang et al., 2017
Septin-9 Clusterin	Plasma	137 EOC, 12 borderline,51 benign, 58 controls	82.5% 71.5%	50.0% 41.4%	Lyu et al., 2018

TABLE 3 | Clinical trials (currently active or completed) for evaluating novel biomarkers of ovarian cancer.

Biochemical marker	Setting	Phase	Samples (*n*)	Status	Clinical trial no.	Primary clinical outcome	References
All biomarkers	Adnexal mass	1	500 (E)	Completed	NCT01466049	Screening	NA
HE4 + CA125	Pelvic mass	0	566	Completed	NCT00315692	Cancer vs. benign disease	Moore et al., 2009
CA125	Low risk women	1	9500 (E)	Recruiting	NCT00539162	Rate of increase in CA125 over time	NA
HE4 + CA125	Adnexal mass	1	512	Completed	NCT00987649	Initial cancer risk assessment	NA
CA125+ HE4	High risk women	1	1208 (E)	Active, not recruiting	NCT01121640	PPV of screening protocol	NA
CA125	High risk women	2	2400 (E)	Withdrawn	NCT00080639	Screening	NA
Mesothelin	Low risk women	0	250 (E)	Unknown	NCT000155740	Screening	NA
FOLR1	Stage I ovarian cancer	2	50 (E)	Terminated	NCT01511055	Sensitivity and specificity of Intraoperative imaging (IOI) with folate	NA
CA125 + TVU	Ovarian disease	0	750 (E)	Terminated	NCT01292733	CA125 measurement in blood over time	NA
CA125± TVU	Postmenopausal	0	48230	Completed	NCT00058032	Screening post menopausal women	Menon et al., 2009; Jacobs et al., 2011
CA125	High risk women	0	2430	Unknown	NCT00039559	Sensitivity and specificity of early detection for ovarian cancer	NA
CA125+ TVU	High genetic risk women	0	5000 (E)	Completed	NCT00033488	Screening women at high genetic risk for ovarian cancer	NA
CA125	High risk women	0	6000 (E)	Recruiting	NCT00005095	Screening	NA
Combined methods	Ovarian neoplasms	0	36000	Unknown	NCT01178736	Low- cost screening	NA
Interventional	High risk women	0	1500	Recruiting	NCT00849199	Genetic testing, screening	NA
All biomarkers	High risk women	0	250 (E)	Unknown	NCT00854399	Overall survival	NA
Tumor markers	High risk women	0	5000	Completed	NCT00267072	Early stage detection	NA
DNA markers	Ovarian cancer	0	118 (E)	Active, not recruiting	NCT00879840	Assessment of screening modalities	NA
BRCA1/2 Mutation	Ovarian neoplasms	0	1500	Completed	NCT00001468	Identifying BRCA1/2 mutation	NA
BRCA	Epithelial ovarian cancer	0	600 (E)	Completed	NCT03229122	Genetic Testing of BRCA	NA
All biomarkers	High risk women	0	500(E)	Recruiting	NCT03150121	Identification of uterine lavage-based biomarkers for early detection	NA
All biomarkers	High risk women	0	6000	Recruiting	NCT00005095	Early stage detection and prevention	NA
CA125	High genetic risk women	0	40	Completed	NCT00043472	Screening	NA
DNA markers	Women with serous epithelial ovarian cancer	0	250	Not yet recruiting	NCT03622385	Early detection of high grade serous epithelial ovarian cancer	NA

TVU, transvaginal ultrasonography; (w), women; (E), estimated enrollment; IOI, intraoperative imaging. Source: http://clinicaltrials.gov/.

ovarian cancer, some of the most frequently methylated genes include OPCML (tumor suppressor activity), TES (involved in regulation of cell motility) and RASSF1A (tumor suppressor activity as well as an inhibitor of the anaphase-promoting complex) (Barton et al., 2008). Promoter methylation of HOXA10 and HOXA11, which are involved in very early ovarian tumor initiation effectively distinguished normal and malignant ovaries (Fiegl et al., 2008; Widschwendter et al., 2009). Methylation induced silencing of PTEN has also been frequently observed in primary epithelial ovarian carcinomas (Kurose et al., 2001). CTGF (encodes the connective tissue growth factor) (Kikuchi et al., 2007; Barbolina et al., 2009), CCBE1 (hypothesized to be involved in regulation of cell motility) (Barton et al., 2010), HIC1 (a p53 target gene) (Rathi et al., 2002), CDH13 (Makarla et al., 2005), and CDH1 (the loss of which correlates with the upregulation of matrix metalloproteinases and metastasis- promoting protein a 5-integrin) (Sawada et al., 2008) act as metastasis suppressors.

TABLE 4 | List of most frequently epigenetically dysregulated genes in ovarian cancer.

S. No	Gene	Function	Epigenetic event	References
1.	Survivin	Apoptotic inhibitor	Downregulated	Mirza et al., 2002
2.	GATA 4	Transcription factor	Downregulated	Caslini et al., 2006; Cai et al., 2009
3.	APC	Regulation of cell adhesion	Downregulated	Tam et al., 2007
4.	ARHI	Ras homolog, maternally imprinted putative tumor suppressor. Negative regulator of cancer growth and progression	Downregulated	Yu et al., 2003; Fu et al., 2015
5.	BRCA 1	DNA damage response	Downregulated	Baldwin et al., 2000; Wilcox et al., 2005; Press et al., 2008
6.	DAPK	Apoptosis	Downregulated	Bai et al., 2004
7.	Estrogen receptor β	Transcriptional activator	Downregulated	Suzuki et al., 2008
8.	hMLH 1	DNA mismatch repair	Downregulated	Meng et al., 2008; Zhang H. et al., 2008
9.	hMSH 2	DNA mismatch repair	Downregulated	Zhang H. et al., 2008
10.	ICAM 1	Cell/matrix adhesion	Downregulated	Arnold et al., 2001
11.	LOT1	Apoptosis	Downregulated	Abdollahi et al., 2003; Kamikihara et al., 2005
12.	OPCML	Cell adhesion	Downregulated	Sellar et al., 2003; Teodoridis et al., 2005; Zhang J. et al., 2006
13.	PACE-4	Serine protease	Downregulated	Fu et al., 2003
14.	RASSFIA	Microtubule stability	Downregulated	Agathanggelou et al., 2001; Yoon et al., 2001
15.	PEG 3	Apoptosis	Downregulated	Feng W. et al., 2008
16.	DLEC 1	Unknown	Downregulated	Kwong et al., 2006
17.	ARLTS 1	Premature termination of translation	Downregulated	Petrocca et al., 2006
18.	TCEAL 7	Cell death regulation	Downregulated	Chien et al., 2005
19.	P16	Cell cycle control	Downregulated	Milde-Langosch et al., 1998; Katsaros et al., 2004
20.	TMS1	Apoptosis	Downregulated	Akahira et al., 2004a
21.	WT1	Transcription factor	Downregulated	Kaneuchi et al., 2005
22.	14-3-3 SIGMA	Regulation of cell growth and differentiation	Downregulated	Kaneuchi et al., 2004
23.	DR 4	Apoptosis	Downregulated	Horak et al., 2005
24.	FBXO 32	Apoptosis (?)	Downregulated	Chou et al., 2010
25.	IGFBP-3	Antiproliferative, pro-apoptosis, and invasion suppressor	Downregulated	Torng et al., 2009
26.	SFRP5	Modulator of Wnt signaling	Downregulated	Su et al., 2010
27.	CCBE 1	Migration and cell survival	Downregulated	Barton et al., 2010
28.	RUNX3	Transcription factor	Downregulated	Zhang et al., 2009
29.	CHFR	Cell cycle control	Downregulated	Gao et al., 2009
30.	Maspin(SERPINBS)	Protease inhibitor	Expressed	Rose et al., 2006
31.	FANCF	DNA repair(?)	Up regulated	Taniguchi et al., 2003
32.	Synuclein gamma	Unknown	Up regulated	Gupta et al., 2003
33.	TUBB3	Formation of microtubules	Up regulated	Izutsu et al., 2008
34.	CLDN3	Tight junction protein	Up regulated	Honda et al., 2007
35.	HOXA10	Transcription factor	Up regulated	Cheng et al., 2010
36.	FBXW7	Ubiquitin ligase mediates ubiquitylation of oncoproteins	Downregulated	Kitade et al., 2016
37.	SPARC	Membrane-associated glycoprotein, normal development, anti-proliferative, and de-adhesive properties angiogenesis, ECM production	Downregulated	Socha et al., 2009
38.	HIC1	Transcriptional repressor	Downregulated	Pieretti et al., 1995; Rathi et al., 2002; Teodoridis et al., 2005; Tam et al., 2007
39.	Rarβ	Transcriptional regulator of cell growth	Downregulated	Makarla et al., 2005; Tam et al., 2007
40.	GSTP1	Intra cellular detoxification	Downregulated	Makarla et al., 2005; Bol et al., 2010

(Continued)

TABLE 4 | Continued

S. No	Gene	Function	Epigenetic event	References
41.	TBX15	Development of mesodermal derivative	Downregulated	Gozzi et al., 2016
42.	HUSIF 1	Modulate heparin binding growth factor signaling promote Wnt signaling pathway	Downregulated	Staub et al., 2007
43.	SFRP1	Inhibitor of Wnt/β-catenin signaling pathway	Downregulated	Su et al., 2009; Kardum et al., 2017
44.	RunX1T1	Putative SMAD4 target/TGFβ/SMAD4 signaling	Downregulated	Yeh et al., 2011
45.	ANGPTL2	Secreted glycoprotein involved in angiogenesis	Downregulated	Kikuchi et al., 2008
46.	CTGF	Adhesion molecule, motility modulator	Downregulated	Kikuchi et al., 2007
47.	FOXD3	Transcriptional regulator of development, cell maintenance, and lineage specification	Downregulated	Luo et al., 2019
48.	NISCH	Encodes imidazoline receptor Nischarin, regulates cellular migration, and invasion upon interacting with PAK1, LIMK, Rac1, and LKB1	Downregulated	Li J. et al., 2015
49.	ABCA1	A TGF-β target, regulator of cholesterol efflux and metabolism	Downregulated	Chou et al., 2015, p. 1
50.	TIMP2	A EZH2 target, endogenous regulator of matrix metalloproteinases, repressor of metastasis	Downregulated	Yi et al., 2017
51.	PCDH17	Transmembrane protein belonging to cadherin superfamily, potential calcium-dependent cell-adhesion protein	Downregulated	Baranova et al., 2018
52.	LDOC1	A nuclear transcription factor, regulator of NFκB Signaling	Downregulated	Buchholtz et al., 2014
53.	RGS2	Regulator of GTPase activity of G protein subunits. Negative regulator of angiotensin-activated signaling pathway	Downregulated	Cacan, 2017
54.	PRTFDC1	Unknown	Downregulated	Cai et al., 2007
55.	DDR	Subclass of RTKs, associated with cell differentiation, proliferation, adhesion, migration, and invasion	Downregulated	Chung et al., 2017
56.	ARNTL	Circadian gene	Downregulated	Yeh et al., 2014
57.	GULP1	Apoptosis, lipid homeostasis, regulator of Arf6-mediated signaling	Downregulated	Maldonado et al., 2018, p. 1
58.	TGFB1	Adhesion, essential for function of microfibrils and interacts with fibronectin and integrins	Downregulated	Kang et al., 2010
59.	SALL2	Cellular quiescence factor, neural development	Downregulated	Sung et al., 2013
60.	PDZ-LIM	Ubiquitination of nuclear p65, inflammation	Downregulated	Zhao et al., 2016
61.	KLF11	Inhibitory regulator of TGFβ signaling, promotes apoptosis	Downregulated	Wang et al., 2015
62.	GBGT1	Encodes glycosyltransferase which plays role in synthesis of Forssman glycolipid	Downregulated	Jacob et al., 2014

Methylation induced repression of these suppressors correlates with invasive EOC.

Several studies have identified the association of tumor-specific gene methylation with molecular, clinical, and pathological characteristics of epithelial ovarian carcinomas. For instance, highest degree of promoter methylation of SFN (an inhibitor of cell cycle progression), TMS1 and WT1 has been demonstrated in clear-cell ovarian tumors than in other histological types (Kaneuchi et al., 2004, p. 14; Terasawa et al., 2004; Kaneuchi et al., 2005; Teodoridis et al., 2005). Another finding suggests that promoter methylation of RASSF1A, APC, GSTP1, and MGMT correlates with the presence of invasive ovarian carcinomas (Makarla et al., 2005). Hypermethylation of FOXD3 correlated with tumor suppressive role (inhibition of proliferation, migration and promotion of apoptosis) in ovarian cancer cells and thus could serve as a potential therapeutic target for diagnosis of ovarian cancer (Luo et al., 2019).

Using a high–throughput approach to screen genes that showed highest differential methylation between ovarian cancer and normal tissue, Melnikov et al. identified 10 genes to be informative in tissue samples which include: BRCA1, EP300, NR3C1, MLH1, DNAJC15, CDKN1C, TP73, PGR, THBS1, and TMS1. A maximum sensitivity of 69% with 70% specificity was attained on testing the potential of several combinations of these genes to discriminate normal from cancer tissue. Since, all tumors analyzed were of advanced stage (either stage IIIA or higher), therefore, the potential of this panel to diagnose EOC at an early stage is unknown (Melnikov et al., 2009). Ibanez de Caceres et al. demonstrated that hypermethylation of atleast one of the six genes in panel (BRCA1, RASSF1A, APC, $p14^{arf}$, $p16^{ink4a}$, and DAPK) could be detected in 70/ 71 (99%) of EOCs using methylation specific PCR. Furthermore, none of the normal non-neoplastic tissue showed methylation, revealing a specificity of 100%. Additionally, across all histological subtypes, grades, stages as well as age, hypermethylation of TSGs was observed (Ibanez de Caceres et al., 2004). Taken together, these results support hypermethylation of these tumor suppressor genes as a relatively early event in ovarian carcinogenesis and could serve as

TABLE 5 | List of hypermethylated genes in ovarian cancer.

Gene	Frequency of hypermethylation in ovarian cancer	Ovarian cancer subtype	Frequency of hypermethylation in normal tissue	Method	References
RASSF1A	15.6–50%	S, M, E, CC	0–13%	MSP	Yoon et al., 2001; Makarla et al., 2005; Choi et al., 2006; Bol et al., 2010
DAPK	50–67% (full)	S, CC, E, M, CS, PDA	50% (Partial)	MSP	Collins et al., 2006; Häfner et al., 2011
p16(CDKN2A)	16.9–42%	S, M, E	0–25%	MSP	Strathdee et al., 2001; Rathi et al., 2002; Tam et al., 2007
HIC1	15.9–51.7%	Not specific	12.5–19%; 11.1%	MSP	Strathdee et al., 2001; Ongenaert et al., 2008
OPCML	46.5–83.3%	Not specific	0%	Restriction enzyme cut analysis, MSP	Czekierdowski et al., 2006a; Zhang J. et al., 2006
MLH1	10%	S, E, M, CC, MIX	0% ADJ NLS	MSP	Strathdee et al., 2001
TERT	29.8%	S, M, E, CC	30%	qMSP	Widschwendter A. et al., 2004
PTEN	16.9%	E, S, M, CC	0%	MSP	Ongenaert et al., 2008
ING1	24%	S, M, E, CC, PDA	0%	MSP	Shen et al., 2005
ITGA8	13.3%	S, E, M, CC, SCC	0% END cyst	MSP	Cai et al., 2007
MGMT	9%	S, M, CC, E, UN	16%	MSP	Makarla et al., 2005
MINT25	16%	S, E, M, CC, MIX	0% ADJ NLS	MSP	Strathdee et al., 2001
APC	18–47.2%	S, M, CC, E, UN	0–25%	MSP	Rathi et al., 2002; Makarla et al., 2005; Ongenaert et al., 2008
BRCA1	10–24%	S, E, M, CC, MIX	0–5.5% ADJ NLS	MSP	Strathdee et al., 2001; Rathi et al., 2002; Ibanez de Caceres et al., 2004
CASP8	3%	S, E, M, CC, MIX	0% ADJ NLS	MSP	Strathdee et al., 2001
CDH1 (E-cadherin)	26–29%	S, M, CC, E, UN	6%	MSP	Rathi et al., 2002; Makarla et al., 2005
CDH13 (H-cadherin)	18–22%	S,M, CC, E, UN	8–13%	MSP	Rathi et al., 2002; Makarla et al., 2005
DCR1	43%	NS	0%	MSP	Shivapurkar et al., 2004
GPR150	26.6%	S, E, M, CC, SCC	0% END cyst	MSP	Cai et al., 2007
Htr (TERC)	24%	S, E, M, CC, MIX	0% ADJ NLS	MSP	Strathdee et al., 2001

ADJ NLS, Adjacent normals; CC, Clear cell; CS, Carcinosarcoma; E, Endometroid; END, Endometrial; M, Mucinous; MIX, Mixed; MSP, Methylation-specific PCR; NS, Not specified; PDA, Poorly differentiated adenocarcinoma; QMSP, Quantitative methylation-specific PCR; S, Serous; SCC, Squamous cell carcinoma; UN, Undifferentiated.

a potential biomarker for detection and accurate discrimination of EOC at early stage.

Using 7- genes panel [secreted frizzled receptor proteins 1, 2 4, 5 (SFRP1, 2, 4, 5), SRY box1 (SOX1), paired box gene 1(PAX1), and LIM homeobox transcription factor 1, alpha (LMX1A)], Sui et al. investigated methylation in 126 primary ovarian tumors, 75 benign ovarian tumors and 14 borderline ovarian tumors and in 26 OC serum samples. Their findings indicated that promoter methylation of any one of SOX1, PAX1, and SFRP1 could distinguish EOC patients from normal control with a sensitivity of 73.08% and a specificity of 75%. Though these test scores are higher than those of CA125 alone, however it is probably not high enough to warrant its implementation as a diagnostic test for individual patients. Moreover, as no specification of tumor stage within the studied group was provided, the performance of this panel in detection of EOC at an early stage therefore remains unclear (Su et al., 2009).

Hypomethylation induced abnormal expression of several oncogenes such as CLDN4 (encodes an integral component of tight junctions) (Honda et al., 2006; Litkouhi et al., 2007), MAL (mal, T-cell differentiation protein) (Lee et al., 2010), BORIS (a cancer testis antigen family candidate oncogenes) (Woloszynska-Read et al., 2007), and IGF2 (an imprinted gene involved in other malignancies) (Murphy et al., 2006) has been demonstrated in ovarian carcinomas. Promoter hypomethylation induced upregulation of other cancer-associated genes in ovarian cancer includes maspin (SERPINB5) (Rose et al., 2006), MCJ (Strathdee et al., 2004, 2005), and SNCG (synucelin-γ) (Gupta et al., 2003; Czekierdowski et al., 2006b), which encodes an activator of the MAPK and Elk-1 signaling cascades. Hypomethylation of SNCG, MASPIN, and CLDN4 correlates with advanced-stage and metastasis while that of BORIS is linked with disease presence.

Hypomethylation of Sat2 (satellite 2) DNA in the juxtacentromeric region of chromosome 1 and 16 has been

TABLE 6 | Epigenetic biomarkers for ovarian cancer detection.

Epigenetic marker	Source	Sensitivity /specificity	Technique	References
GTF2A1 and HAAO	Tumor tissue	Presence of malignancy 95/88%; 89/82%	qMSP	Huang et al., 2009
HOXA9 and SCGB3A1	Tumor tissue	Early stage carcinoma (18/24, $p = 0.002$) (5/25, $p = 0.020$)	MSP	Wu et al., 2007
RASSF1A and HIC1	Tumor tissue	Early stage ovarian carcinomas 34/100, 2/68, OR = 0.3 34/100, 10/68, OR = 0.4	MSP	Feng Q. et al., 2008
RASSF1A, GSTP1, MGMT, APC	Tumor tissue	Presence of invasive tumors RASSF1A: 30% vs. 0%; GSTP1: 9% vs. 0%; MGMT: 9% vs. 0%; APC:22% vs. 0%;	MSP	Barton et al., 2008
SPARC	Tumor tissue	Association with tumor grade Methylation frequency: 68%	MSP	Socha et al., 2009
CDH13, CRABP1, HOXA9, and SCGB3A1	Tumor tissue	Histological subtype differentiation CDH13, CRABP1, HOXA9 and SCGB3A1 (P = 0.041, P < 0.001, P = 0.007, P < 0.001)	MSP	Wu et al., 2007
CTGF	Tumor tissue	Inversely correlated with invasive disease	cDNA microarray analysis	Barbolina et al., 2009
CCBE1	Tumor tissue	Inversely correlated with metastasis 6/11 (55%) in OC cell lines 38/81 (41%) OC tumors.	Small-interfering RNA (siRNA)-mediated knockdown	Barton et al., 2010
HIC1	Tumor tissue	Presence of malignancy Methylation frequency: 35%	MSP	Rathi et al., 2002
CDH1	Tumor tissue	Inversely correlated with metastasis	Small-interfering RNA (siRNA)-mediated knockdown	Sawada et al., 2008
SFN, TMS1, and WTI	Tumor tissue	Methylation exclusive for Clear cell subtype	MSP	Barton et al., 2008
hMLH1, CDKN2A, and MGMT	Matched tumors	Associated with development of Synchronous endometrial and ovarian cancer Methylation frequency: 39, 41, and 48%	MSP	Furlan, 2006
*14–3-3*s	Tumor tissue	Advanced stage ovarian carcinomas	MSP, quantitative reverse transcription-PCR	Akahira et al., 2004b
HOXA11	Tumor tissue	Presence of malignancy, Suboptimal tumor debulking and poor outcome (relative risk for death = 3.4)	MethyLight assay	Fiegl et al., 2008
10 gene panel	Tumor tissue	Presence of serous adenocarcinoma (69.4% sensitivity and 70.2% specificity)	Microarray Mediated Methylation Assay (MethDet test)	Melnikov et al., 2009
Polycomb group target genes in particular HOXA9	Normal endometrium	Hoxa9 hypermethylation association with increased risk (12.3 fold) of ovarian cancer	MethyLight assay	Widschwendter et al., 2009
SNCG, MASPIN, and CLDN4	Tumor tissue	Advanced stage ovarian carcinomas	Small-interfering RNA (siRNA)-mediated knockdown, qRT-PCR	Gupta et al., 2003; Rose et al., 2006; Choi et al., 2007; Honda et al., 2007
PCDH17	Tumor tissue	Presence of malignancy	Next-generation sequencing, Methylation-sensitive high-resolution melting Analysis	Baranova et al., 2018
EGFL7 and RASSF1	Tumor tissue	Early stage disease detection and progression	DNA methylation microarray assay, Bisulfite pyrosequencing	Rattanapan et al., 2018
LDOC1	Ovarian cancer cell line	Early stage disease detection	RT-PCR and real-time PCR	Buchholtz et al., 2014
GPR150, ITGA8, and HOXD11	Ovarian cancer cell line	Tumor marker	Methylation-sensitive-representational difference analysis (MS-RDA) and MSP	Cai et al., 2007

(Continued)

TABLE 6 | Continued

Epigenetic marker	Source	Sensitivity /specificity	Technique	References
TGFBI	Tumor tissue and ovarian cancer cell line	Presence of malignancy	Real-time RT-PCR, MS-PCR, and bisulfite sequencing	Kang et al., 2010
DAPK1 and SOX1	Tumor tissue	Early stage disease	MethyLight	Kaur et al., 2016
BORIS	Tumor tissue	Presence of malignancy	qRT-PCR	Woloszynska-Read et al., 2007
long-intergenic non-coding RNA (lincRNA) gene (LOC134466)	Tumor tissue and ovarian cancer cell line	Potential novel diagnostic biomarker for high grade (Type II) serous ovarian carcinoma (HGSOC)	MeDIP-Chip and Sequenom massARRAY methylation analysis	Gloss et al., 2012

reported in ovarian cancer (Qu et al., 1999). A significant increase in hypomethylation of chromosome 1 Sat2 and chromosome 1 satellite α from non-neoplastic tissue toward ovarian cancer tissue was observed. Higher hypomethylation levels were observed in serous and endometrioid tumors in comparison to mucinous. Moreover, extensive hypomethylation was prevalent in high grade or advanced stage tumors (Widschwendter M. et al., 2004). Taken together, consistent higher expression levels along with hypomethylation of L1 and human endogenous retrovirus-W retrotransposons (repetitive sequences widely distributed throughout the genome) has been reported in malignant ovarian tumors against normal control samples (Menendez et al., 2004). It has been hypothesized that promotion in homologous recombination as a result of increased hypomethylation, leads to chromosomal aberrations associated with carcinogenesis (Kolomietz et al., 2002; Symer et al., 2002).

Prognosis

Potential prognostic biomarker includes FBXO32, which correlates with advanced stage and shorter disease free survival (Chou et al., 2010), Ribosomal DNA (18S and 28S) linked with prolonged disease free survival (Chan, 2005), IGFBP-3, correlates with disease progression and death in early stage EOC (Wiley et al., 2006b) and HOXA11, association with postsurgical residual tumor and poor outcome (Fiegl et al., 2008). Methylation of ≥1 gene of SFRP1, SFRP2, and SOX1 correlated with short disease free survival while SOX1, LMX1A, and SFRP1 methylation was associated with recurrence and short overall survival (Su et al., 2009). A progression-free survival prediction accuracy of 95% is reported by Wei et al. with hMLH1, IGFP3, and NEUROD1 among a panel of 112 highly discriminatory loci (Wei, 2006). Furthermore, detection of prognostic epigenetic biomarker has also been described in plasma as well as peritoneal fluid. Methylation of hMLH1, analyzed in 138 plasma samples predicted poor survival (hazard ratio: 1.99) (Gifford, 2004) while CDH1, CDH13, and APC (out of a 15 gene panel) analyzed in peritoneal fluid from 57 ovarian cancer patients could predict overall survival (Suehiro et al., 2008). Huang et al. recently reported that the epigenetic loss of heparin sulfate 3-O-sulfation makes ovarian cancer cells sensitive to oncogenic signals and could predict prognosis, thereby reflecting the utility of HS3ST2 for targeted therapy (Huang et al., 2018).

Recently using genome-wide methylation data analysis, five-methylation signature (SLC39A14, PREX2, KCNIP2, CORO6, and EFNB1) were reported as novel independent prognostic biomarker for patients with ovarian serous cystadenocarcinoma, which significantly associated with OS of patients. Moreover, these signatures exhibited higher sensitivity and specificity to predict OSC prognosis (AUC = 0.715), which reflects their clinical significance in improving outcome prediction. Furthermore, these 5- methylation signatures were more accurate over known biomarkers in predicting prognostic survival of OSC patients (Guo T. Y. et al., 2018). Promoter methylation of BRCA1 has been reported to be associated significantly with increased PFS of patients with OC undergoing adjuvant platinum–taxane-based chemotherapy ($P = 0.008$) as well for the patients with disease recurrence (PFS = 18.5 months over 12.8 months for patients without BRCA1 promoter methylation), thereby reflecting that promoter methylation of BRCA1 could be a better predictive marker of response to platinum–taxane-based chemotherapy in sporadic Epithelial ovarian carcinoma (Ignatov et al., 2014).

Another study highlights the potential of CDH1, DLEC1, and SFRP5 gene methylation panel as a prognostic biomarker in advanced stage OC patients. Presence of two or more methylated genes in patients significantly correlated with disease recurrence (hazard ratio: 1.91; $p = 0.002$) and shorter overall survival and disease free survival (hazard ratio: 1.96; $p = 0.006$) (Lin et al., 2018). Liu et al. reported the prognostic potential of C/EBPβ (a transcription factor) which augments chemoresistance of ovarian cancer cells by maintaining an open chromatin state via reprogramming H3K79 methylation of multiple drug-resistance genes upon direct interaction with DOT1L (DNA methyltransferase), thus provides a new insight for more precise therapeutics options in OC by identifying and targeting the key regulators of epigenetics (Liu et al., 2018).

Several recent researches have suggested the hypermethylation and reduced expression is prognostic for shorter progression free survival. For instance, using genome wide array analyses, Hafner et al. reported 220 differentially methylated region with short and long PFS. Validation experiments on a large cohort of type II EOC revealed the association of RUNX3/CAMK2N1 with poor clinical outcome (Lower PFS), indicating the prognostic potential of these genes (Häfner et al., 2016). Few studies have highlighted the tight link between promoter methylation and metastasis. For instance, stimulation of ovarian cancer cell lines by TGFβ, which is a key player in metastasis, extensively change promoter methylation of genes that are associated with

EMT (Epithelial-mesenchymal transition) and progression of cancer (Cardenas et al., 2014). Deng et al. reported the tumor suppressive role of IQGAP2 which suppresses the ovarian cancer progression via suppressing Epithelial-mesenchymal transition by regulating Wnt/β signaling, thereby providing a potential biomarker and therapeutic strategy to combat ovarian cancer diagnosis (Deng et al., 2016).

Brachova et al. studied the association of oncomorphic TP53 mutation on patient outcome diagnosed with advanced EOC. Oncomorphic TP53 mutation correlated with worse progression free survival, higher risk of recurrence and higher rate of platinum resistance (Brachova et al., 2015). Dai et al. explored the association of methylation-based prognostic biomarkers within key ovarian cancer-related pathways with progression free survival to platinum based chemotherapy in HGSOC. NKD1, VEGFB, and PRDX2 were identified as the best predictors of progression free survival (PFS: HR = 2.3 $p = 3.3 \times 10–5$; Overall Survival: HR = 1.9, $p = 0.007$). Further validation using independent TCGA data set revealed the significant association of VEGFA, VEGFB, and VEGFC promoter methylation with progression free survival (Dai et al., 2013).

Promoter hypomethylation and expression of PRAME correlates with increased survival in high grade serous ovarian carcinoma (Zhang et al., 2016). Promoter hypomethylation and increased expression of proto-oncogenes is predictive for more aggressiveness and metastasis of disease and thereby lower survival, which is evident from recent studies on GABRP, SLC6A12, MGAT3, CT45, CA9, MUC13, and AGR2 (Sung et al., 2014a,b,c, 2017a,b; Zhang et al., 2015; Kohler et al., 2016). Hypomethylation of Sat2 DNA (Chr 1) was associated relapse and poor prognosis (Widschwendter M. et al., 2004), and LINE1 was linked with poorer overall survival and lower progression free survival (Pattamadilok et al., 2008; **Table 7**).

Another important study by Wei et al. reported 112 methylated loci which were prognostic for reduced PFS and could predict PFS with an accuracy of 95% using Significance Analysis of Microarray and Prediction Analysis of Microarray algorithm (Wei, 2006). Twenty-two hypermethylated loci were identified by global methylation profiling of 485 tumor samples of clear-cell ovarian cancer in a recent study. These hypermethylated loci were associated with 9 genes (VWA1, FOXP1, FGFRL1, LINC00340, KCNH2, ANK1, ATXN2, NDRG21, and SLC16A11). Further, methylation induced silencing of KCNH2 (HERG, a potassium channel) could be a better prognostic factor for poor survival provided increased proliferation was mediated by overexpression of Eag family members. However, further validation on larger cohort is still warranted (Cicek et al., 2013). Huang et al. identified 63 differentially methylated regions of prognostic relevance which significantly correlated with poor PFS. Further, epigenetic silencing of regulators of hedgehog signaling pathway ZIC1 and ZIC4 was associated with increased proliferation, migration, and invasion. Additionally, promoter hypermethylation of ZIC1 significantly correlated with poor survival and thus could serve as prognostic determinant for patient outcome (Huang et al., 2013).

Another study describes that the global methylome status of HGSOC PDX (patient-derived xenografts) resembled with global methylation in corresponding patient tumor over several generations and could be efficiently modulated by demethylating agents. C-terminal Src kinase (CSK), a novel epigenetically regulated gene and associated pathways were also identified. Low CSK methylation significantly correlated with improved PFS and OS in HGSOC patients (Tomar et al., 2016). Koestler et al. using integrative global methylation and single nucleotide polymorphisms analysis identified DNA methylation marks (13 unique CpGs and 17 unique SNPs) which could mediate EOC genetic risk (Koestler et al., 2014).

Recently, Sharma et al. investigated epigenetic regulation of POTE gene family, which is localized to autosomal pericentromeric region. POTE gene family is over-expressed in HGSOC. Epigenetic silencing of POTE gene was functionally verified by experiments involving treatment with Decitabine and DNMT knockout cell lines. In addition expression of individual gene in POTE gene family correlated with chemoresistance and poor clinical outcome in HGSOC patients. Furthermore, several epigenetic alternations (pericentromeric activation, global and locus-specific L1 hypomethylation, and locus-specific 5' CpG hypomethylation) served as a determinant for regulation of epigenetic activation of POTE gene (Sharma et al., 2019).

In conclusion, these studies provides insight to the association of several potential methylation based prognostic biomarkers with clinical outcome in ovarian carcinoma and further suggest that these reports on epigenome wide interrogation of DNA methylation warrants detailed functional analysis of loci sufficiently discriminating OC with normal state. New targets identified through comprehensive methylome analysis in OC have significant translational potential to pave the design of future clinical investigations and therapeutics.

Predictive

Methylation mediated transcriptional repression of specific drug-response genes results in acquisition of drug resistance and significantly extends its impact on different facets of chemotherapeutic actions: membrane entry/exit, drug metabolism, response to cellular injury, DNA repair, apoptosis etc., in cancer cells. Hypermethylated genes such as hMLH1, ASS1 (arginine biosynthesis-related gene), ESR2 (encoding ER-β), and SFRP5 (encodes an inhibitor of oncogenic WNT signaling pathway) have been implicated in platinum resistance. Three studies well defined in ovarian cancer includes: Methylation of either BRCA1, GSTP1, or MGMT significantly correlates with improved response to chemotherapy ($p = 0.013$) (Teodoridis et al., 2005). Hypermethylation of RASSF1A and CABIN1 have been reported to correlate with response to adjuvant therapy. Patients who responded to therapy had moderately higher frequencies of RASSF1A hypermethylation (OR = 0.4) and significantly higher frequencies of CABIN1 hypermethylation (OR = 0.1) (Feng Q. et al., 2008). Strathdee et al. demonstrated that high levels of MCJ methylation significantly correlated with poor response to therapy (p = 0.027) and poor overall survival ($p = 0.023$; HR = 2.9) (Strathdee et al., 2005). Hypomethylation induced upregulation of ABCG2 (multidrug transporter) MAL (determinant of platinum resistance) and TUBB3 (determinant of taxane resistance) genes have been described in advanced

TABLE 7 | Epigenetic biomarkers for ovarian cancer prognostication.

Epigenetic marker	Source	Clinical prediction	References
182-gene panel	Tumor tissue	HR 2.5 for PFS	Wei et al., 2002
112-gene panel	Tumor tissue	Prediction accuracy of 95% for shorter Disease free survival	Wei, 2006
SFRP1, *SFRP2,* and *SOX1*	Tumor tissue	Correlates with shorter Disease free survival	Su et al., 2009
SOX1, *LMX1A,* and *SFRP1*	Tumor tissue	Correlates with shorter overall survival	Su et al., 2009
hMLH1	Plasma	If hypermethylated HR:1.99 for OS	Gifford, 2004
HOXA10, HOXA11	Tumor tissue	RR for death:3.4, if HOXA11 methylated	Fiegl et al., 2008, p. 11
18S and 28S rDNA	Tumor tissue	Prolonged DFS	Chan, 2005
EN2	Tumor tissue	Short progression free survival	McGrath et al., 2018
MYLK3	Tumor tissue	Higher methylation level significantly predicted better overall survival with least residual disease	Phelps et al., 2017
FBXO32	Tumor tissue	Advanced stage and short DFS	Chou et al., 2010
Panel of IGFFBP3, p16, BRCA1, GSTP1, ER-α, hMLH1	Tumor tissue	Seven fold increased risk of short DFS HR: 6.53 for disease progression	Wiley et al., 2006a
RUNX3, CAMK2N1	Tumor tissue and ovarian cancer cell line	Short overall survival	Häfner et al., 2016
ABCA1	Tumor tissue and ovarian cancer cell line	Shorter overall survival	Chou et al., 2015
GULP1	ovarian cancer cell line	Residual disease, worse overall survival, and disease specific survival	Maldonado et al., 2018, p. 1
FZD4, DVL1, and ROCK1	Tumor tissue	Correlated with early disease relapse	Dai et al., 2011
DNA hypomethylated genes			
15 gene panel	Peritoneal fluid	Short overall survival	Muller, 2004
Sat 2 DNA (Chr1)	Tumor tissue	RR for relapse:4.1, RR for death:9.4 if region methylated	Widschwendter M. et al., 2004
LINE1	Tumor tissue	Lower methylation level significantly predicted poor OS and PFS	Pattamadilok et al., 2008
ATG4A, HIST1H2BN	Tumor tissue	Poor progression free survival	Liao et al., 2014

ovarian carcinoma cases with drug-acquired chemoresistance (Izutsu et al., 2008; Balch et al., 2010; Lee et al., 2010; **Table 8**).

Recently Pulliam et al. demonstrated the combinatorial effect of DNA methyltransferase inhibitor (DNMTi) guadecitabine and the Poly (ADP-ribose) polymerase (PARP) inhibitors (PARPi) talazoparib in resensitizing PARPi resistant breast and ovarian cancer irrespective of BRCA status. Synergistic effect of guadecitabine and talazoparib increased ROS accumulation, and further sensitized the breast and ovarian cancer cells toward PARPi sensitivity by subsequent activation of cAMP/PKA signaling which in turn promoted PARP activation. Furthermore, DNMTi augmented PARP "trapping" by talazoparib. The finding of this complementary model supports further clinical exploration of this combination therapy in PARPi-resistant cancers (Pulliam et al., 2018). Another study using integrated global methylation analysis on extreme chemoresponsive HGSOC patients identified four genes of clinical relevance (FZD10, FAM83A, MYO18B, and MKX) as epigenetic marker of platinum based chemoresponse, of which, FZD10 was reported as functionally validated marker of platinum sensitivity (Tomar et al., 2017). Promoter methylation of OPCML was significantly associated with poor overall survival of OC patients and thus could be of use in predicting disease prognosis (Zhou et al., 2014).

A recent study has described induction of hypomethylation in resistant ovarian cancer patients upon treatment with cisplatin, though, in the intergenic regions, the loss of methylation was primarily observed (Lund et al., 2017). Hypomethylation of developmental genes MSX1 and TMEM88 correlated with platinum resistance in patients with ovarian cancer (Bonito et al., 2016; de Leon et al., 2016). Stimulation of EMT by non-coding RNA HOTAIR has been reported to be regulated by DNA methylation and is indicative of resistance to carboplatin (Teschendorff et al., 2015). Likewise, another study highlights promotion of platinum resistance by TET. Induction of EMT by TET is mediated by demethylation of Vimentin promoter in ovarian carcinoma (Han et al., 2017).

A recent study has described how methylome-targeting strategies could bring forth anti-tumor effect. Guadecitabine-mediated induction of global hypomethylation not only affects metabolic and immune responses but also activates tumor suppressor genes which eventually contribute to platinum drug re-sensitization in ovarian cancer. This might offer utility in improving survival outcomes of patients with ovarian cancer

TABLE 8 | Epigenetic biomarkers for ovarian cancer prediction.

Epigenetic marker	Source	Clinical prediction	References
Methylation of >1 of BRCA1, GSTP1, and MGMT	Tumor tissue	Association with improved response to chemotherapy	Teodoridis et al., 2005
RASSF1A, CABIN1	Tumor tissue	Association with response to chemotherapy	Feng Q. et al., 2008
ASS1	Tumor tissue	Determinant of Platinum resistance	Nicholson et al., 2009
HSulf1	Tumor tissue	Determinant of Platinum resistance	Staub et al., 2007
SFRP5	Tumor tissue	Determinant of Platinum resistance	Su et al., 2010
hMLH1	Plasma	Relapse of Chemoresistant tumor	Gifford, 2004
ESR2	Tumor tissue	Determinant of Platinum resistance	Yap et al., 2009
MCJ	Tumor tissue	Association with response to chemotherapy and overall survival	Strathdee et al., 2005
TUBB3	Tumor tissue	Taxane resistance	Izutsu et al., 2008
MSX1	Ovarian cancer cell line	Sensitivity to platinum drug	Bonito et al., 2016
TBX2	Tumor tissue	Sensitivity to platinum drug	Tasaka et al., 2017
MAL	Tumor tissue	Platinum resistance	Lee et al., 2010

(Fang et al., 2018). Another recent study has highlighted the tumor suppressor role of ZNF671 and its methylation could act as a predictor for early recurrence of serous ovarian carcinoma (Mase et al., 2019). Another important study by M. Keita et al. has for the first time reported the exclusive association of massive DNA hypomethylation with poorly differentiated tumors, which correlates with disease aggressiveness and progression. This report also raises concern over the adverse effect of use of demethylating agents which probably aid the activation of oncogenes and prometastatic genes (Keita et al., 2013).

In conclusion, it is speculated that the combinatorial therapies utilizing epigenetic inhibitors holds promise and would be most effective for chemo-resensitization of resistant tumors, possibly by restoration of pathways associated with drug response, and thus would subsequently implicate improved survival outcomes as well as personalized treatment for this devastating disease.

Histone Modifications in Ovarian Cancer

Compared with DNA methylation, the evidence on chromatin modification in development of ovarian cancer is limited. Histone modification mediated regulation of cell cycle regulatory proteins such as cyclin B1 (Valls et al., 2005), p21 (Richon et al., 2000), and ADAM19 (Chan M. W. et al., 2008) have been described in various reports. Association of histone modifications with aberrant class III β tubulin protein expression (Izutsu et al., 2008), reduction of PACE3 expression (Fu et al., 2003) and silencing of survivin (Mirza et al., 2002) has been reported in ovarian tumorigenesis. Upregulation of tumor suppressor Rb and CDKN1 (cyclin-dependent kinase inhibitor) by histone acetylation was described by Strait et al. (2002). Moreover, the overexpression of HDACs 1–3 in ovarian cancer has been reported to be associated with high grade tumors and resulting poor prognosis (Weichert et al., 2008). On the other note, the derepression of claudin-3 and claudin-4 was found to be associated with loss of trimethylated histone 3 lysine 27 (H3K27me3) (Kwon et al., 2010). The transcriptional repression of osteoprotegerin (OPG) has been reported to be mediated by reduced histone 3 lysine 4 trimethylation (H3K4me3) and increased H3K27me3 (Lu et al., 2009). Similarly, the association of transcriptional silencing of GATA4 and GATA6 with hypoacetylation of histones H3 and H4 and loss of trimethylated histone 3l ysine 4 (H3K4me3) has been described by Caslini et al. (2006).

A very recent report has provided insight into the mechanism associated with development and progression of OC. Early Loss of E3 ubiquitin ligase RNF20 and histone H2B monoubiquitylation (H2Bub1) has been reported to drive ovarian tumorigenesis by altering chromatin accessibility and thereby activating immune signaling pathways (IL6), and this loss has been defined by majority of high grade serous ovarian carcinomas tumors (Hooda et al., 2019). Cacan et al. reported that the loss of FAS expression which contributes to drug resistance is mediated by histone deacetylase 1 (HDAC1) in chemoresistant OC cells (Cacan, 2016). Recently Tang et al. highlighted the repression of histone H3 lysine 27 trimethylation (H3K27me3) which was mitigated by AMP-activated protein kinase (AMPK) phosphorylation upon treatment with metformin thus implicated the antitumor effect of metformin and suggested its utility in the treatment of EOC patients who are not diabetic (Tang et al., 2018).

In another study, the mechanism associated with upregulation of ABCB1 was conferred to chromatin remodeling (via p300 mediated H3K9ac and AR complex binding to ARE4) which in turn leads to the development of taxol resistant phenotype. It was shown that the upregulation of p300 and GCN5 (HATs) was associated with overexpression of ABCB1 and resistance to taxol and PI3K/AKT pathway which is activated by taxol, mediates the regulation of the expression of p300 and AR. These results further reveal the significance of AKT/p300/AR axis as a novel treatment strategy in combating taxol resistance (Sun et al., 2019). Using ChIP-seq approach, Curry et al. identified genome-wide bivalent domains (H3K27me3 and H3K4me3) at gene promoter in tumor samples which were collected pre and post platinum resistance acquisition, and showed that these representative poised gene sets are pre-disposed to hypermethylation induced epigenetic silencing during acquisition of drug resistance, thus provides novel insights to prevent emergence of drug resistance (Curry et al., 2018).

Yi et al. reported that Enhancer of zeste homolog 2 (EZH2) mediates repression of tissue inhibitor of metalloproteinases

2(TIMP2) by H3K27me3 and DNA methylation thereby facilitating ovarian cancer metastasis (Yi et al., 2017). In similar context, another study highlighted silencing of ARHI in ovarian cancer which was synergistically mediated by Enhancer of zeste homolog 2 (EZH2) induced H3K27me3 and DNA methylation. Furthermore increased EZH2 expression correlated with worse overall survival rates, implicating prognostic potential of EZH2 in EOC (Fu et al., 2015). Repression of Regulator of G-protein signaling 2 (RGS2) via histone deacetylases (HDACs) and DNA methyltransferase I in chemoresistant OC cells has been reported recently by Cacan et al., and utility of their inhibitors might serve as a novel approach to overcome chemoresistance in ovarian cancer (Cacan, 2017).

Clinical Application of Epigenetic Biomarker in Liquid Biopsies for Ovarian Cancer Management

Cell Free DNA Biology

Advancement in the understanding underlying molecular pathogenesis of cancer, along with advancements in molecular techniques has facilitated the study of molecular alternations associated with cancer development at an early stage in body fluids. Circulating cell free DNA which are believed to have derived from tumor cells, reflect specific genetic and epigenetic alternations, and thus may offer potential non-invasive viable biomarkers for several cancer, capable of providing valuable information regarding disease progression and response to therapy in real time.

In 1948, the existence of cell free DNA was first described by Mandel and Métais. Cell free DNA are derived from necrotic and apoptotic cells, commonly released by all cell types. Further, numerous subsequent studies confirmed that the tumor-specific pattern of alterations, such as chromosomal abnormality, somatic mutations, resistance mutation, aberrant methylation and copy number variations could be found in cfDNA, which can serve as potential target for diagnosis of cancer through non-invasive approach (Leon et al., 1977; Polivka et al., 2015; **Figure 3**).

Numerous studies support the detection of methylation signature in almost any body fluid (such as serum, plasma, smears, nipple fluid aspirate, and vaginal fluid etc.). As sampling of blood can be considered as minimal invasive process, thus serves as an ideal substrate for methylation analysis. The average concentrations of circulating cell free DNA in healthy subjects is 30 ng/ml. However, in cancer patients, the average concentration of cell free serum DNA is higher, approximately 180 ng/ml as dying cancer cells release tumor DNA into the blood (Gormally et al., 2007). The average length of circulating cfDNA, which are usually fragmented, is 140 to 170 bp and of which, only a fraction of few thousand amplifiable copies of cfDNA /ml of blood, might be of diagnostic relevance (Gormally et al., 2007; Polivka et al., 2015). The levels of circulating cell free DNA in serum is abnormally high in early as well as advanced-stage tumors (Perlin and Moquin, 1972; Leon et al., 1977). For this phenomenon, the proposed two primary mechanisms includes: either cells in cancer tissue undergoes *in situ* apoptosis and/or necrosis or cells might detach from tumors and extravasate into bloodstream where they undergo lysis (**Figure 4**).

Since its first validation, the potential application of circulating DNA in research settings and for non-invasive management of cancer as "liquid biopsy" is expanding with improvement in molecular and genomic techniques. Numerous studies have demonstrated that tumor specific aberrant methylation can also be detected in cfDNA of patients with different tumor types such as lung, prostate, breast and colorectal cancer and further confirmed altered methylation as an independent diagnostic/ prognostic marker (Board et al., 2008; Brock et al., 2008; Lofton-Day et al., 2008; Vlassov et al., 2010). Warren et al. developed a highly sensitive non-invasive test for screening of colorectal cancer based on methylation of SEPT9 in plasma which could specifically detect all stages and locations of colorectal cancers (Warren et al., 2011). Hypermethylation of Vim gene is strongly correlated with the occurrence of colorectal cancer. Similarly hypermethylation of SHOX2 in sputum has been used as biomarker for distinguishing malignant and benign lung diseases (Kneip et al., 2011). Gstp1 methylation status in urine is strongly correlated with early onset of prostate cancer (Belinsky, 2004).

Numerous reports have highlighted the potential of DNA methylation based biomarkers for non-invasive detection of cancer utilizing cell free DNA. Recently, using integrated methylome analysis Wei et al. reported hypermethylation of SPG20, a putative STAT3 target, for non-invasive detection of gastric cancer at an early stage (Wei et al., 2019). Yang et al. explored the potential of eight gene panel for non-invasive detection of lung cancer using qMSP and revealed that the promoter methylation of any of the eight gene could detect the disease with a sensitivity of 72% with 91% specificity, reflecting the utility of plasma DNA methylation as a novel approach for detection of lung cancer at early stage (Yang et al., 2018).

Similarly, promoter methylation of OPCML and HOXD9 assessed in serum cell free DNA using methylation-sensitive high-resolution melting, was detected with a sensitivity of 62.50% with specificity of 100%, thus could serve as a non-invasive differential biomarker to prevent misdiagnosis of cholangiocarcinoma (CCA) and other biliary diseases (Wasenang et al., 2019). Further, for the management of pancreatic cancer and its early detection Eissa et al. analyzed the methylation of ADAMTS1 and BNC1 in cfDNA using qMSP, which exhibited a sensitivity of 94.8% and specificity of 91.6% with a AUC of 0.95 reflecting diagnostic potential of this blood based two-gene panel in detection of pancreatic cancer at an early stage (Eissa et al., 2019). Methylation of APC, FOXA1, and RASSF1A in cell free DNA served as a best performing cassette in terms of diagnostic and prognostic value, revealing a sensitivity, specificity and accuracy over 70% suggesting its putative utility in management of breast cancer (Salta et al., 2018).

Other studies using genome-wide methylation profiling of serum/plasma cell-free DNA have identified potential biomarkers for clinical utility. For instance, Xu et al. using MeDIP-seq approach reported 10 significant differentially methylated genes as potent biomarker for lung cancer clinical application (Xu et al., 2019). Similarly, using genome-wide methylome profiling and

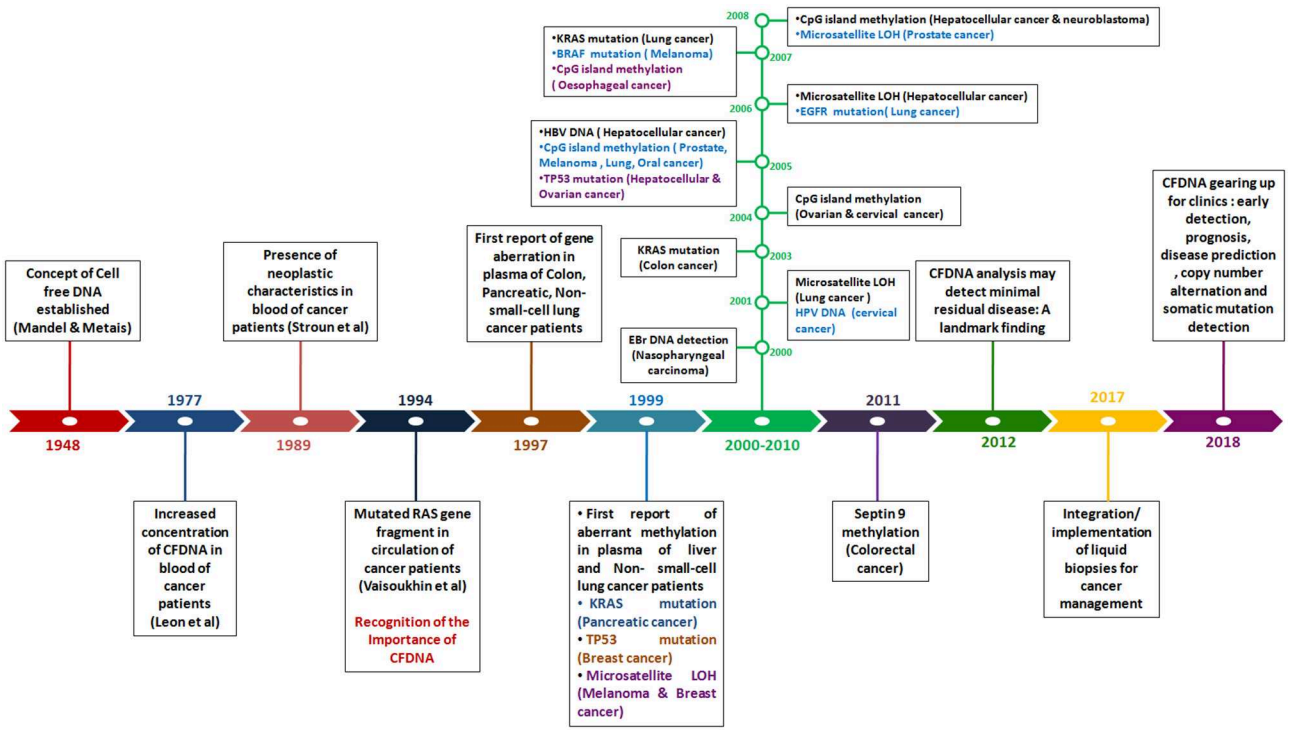


FIGURE 3 | Timeline reflecting the detection of genetic and epigenetic alternations in Cell-free DNA in blood of patients with different cancer type.

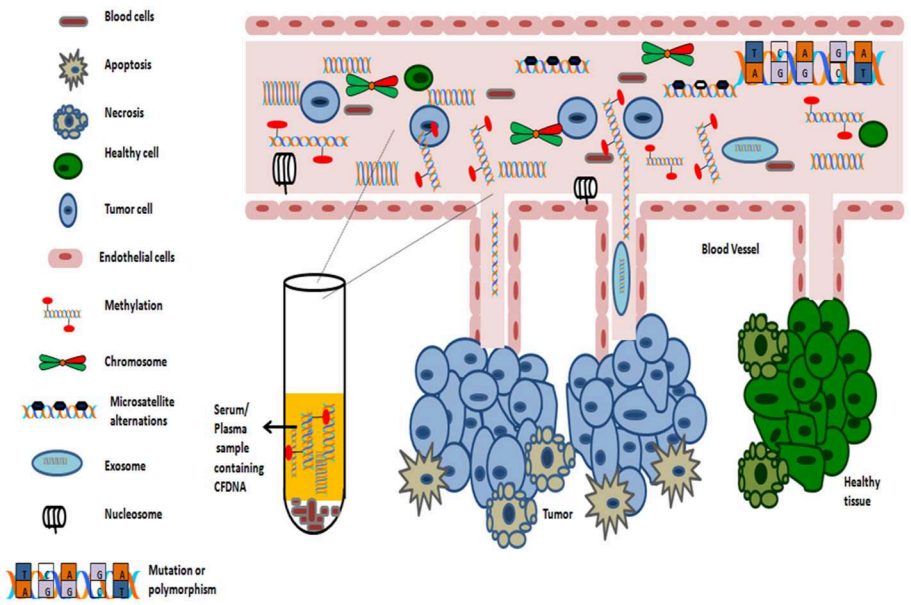


FIGURE 4 | Origin of Cell-free DNA in blood.

Sequenom MassARRAY approach, it was reported that promoter methylation of CASZ1, CDH13, and ING2 could serve as a potent noninvasive biomarker for detection of esophageal cancer at early stage (Wang H.Q. et al., 2018).

Challenges

The analysis of blood borne cell-free DNA has tremendous potential to enable rapid, non-invasive molecular diagnosis of cancer. They are of great clinical relevance as they provide

specific targets for initial diagnosis, permit monitoring of treatment efficacy as well as information about tumor profile and its dynamics which are critical for treatment decisions (De Mattos-Arruda et al., 2014; Lewis et al., 2015).

The advantages of analyzing tumor specific DNA methylation in cell free serum DNA includes, improved sensitivity as cfDNA can be easily amplified by PCR, fewer false positive rate as methylation pattern is generally conserved throughout the progression of disease, stability during sample collection as abnormal DNA methylation is chemically as well as biologically stable and remains relatively unaffected by physiological condition at the time of sample collection, increased technical sensitivity and specificity for gene specific assays as well as offers assay design advantages over genetic alternation that might be interspersed throughout a given gene. Furthermore, DNA methylation is a positively detectable signal, unlike a loss of signal as in chromosomal deletions (Wittenberger et al., 2014).

Several limitations in the methylation detection of cell free serum DNA includes extremely low amount of available cfDNA, missing bisulfite conversions as they are usually fragmented, low sensitivity demonstrated by a single marker and time-consuming, complicated and expensive conventional techniques for cfDNA isolation. The most commonly used technique for methylation detection is MSP PCR (methylation specific PCR) which is a bisulfite-conversion- based method. The limitation of bisulfite conversion of cfDNA is the missing DNA. Because of the technical difficulties of DNA methylation analysis, only few DNA methylation based markers has been identified to date, which apply only to a fraction of gynecological cancers including breast, ovarian and endometrial cancers (Wittenberger et al., 2014; Lewis et al., 2015).

The two technological challenges to be addressed include (1) the detection of low abundant tumor-specific DNA methylation patterns through methylation specific PCR priming or probing with high signal-to-noise ratio (2) the determination of methylation status of consecutive sites in individual DNA molecules with single base-pair resolution. This requires methylation-independent priming and sequence analysis of combined PCR product. Clinically the major problem associated with DNA methylation assays is to detect scarcely abundant alleles within high background levels of non-target molecules. However, with the advent of digital MethyLight assay together with rapid advances in next generation sequencing based technologies, these issues can be overcome. One example of this novel approach is the development of the PraenaTest™ (LifeCodexx, Germany) (Weisenberger et al., 2008).

Serum Based Epigenetic Biomarker

Tumor-specific methylation-based biomarkers might possibly prove valuable for monitoring disease prognosis and different pathological determinants; however, non-invasive analysis and characterization of biomarkers in body fluids offers more feasibility in early screening and detection of the disease as well in monitoring the response to therapy. Numerous studies have reported aberrant methylation in ovarian cancer as discussed earlier; there are relatively few reports of serum/plasma methylation biomarkers for earlier detection of OC. Various studies that demonstrated striking detection sensitivities and specificities in non-invasive assays, thereby supporting the promising utility of these biomarkers for early screening and detection of OC has been summarized in **Table 9.**

MicroRNAs in Ovarian Cancer

Aberrant expression of microRNAs has been confirmed in ovarian carcinogenesis. A decrease in mRNA levels of the miR-processing enzymes in OC malignant cases against normal controls, strongly implicates an overall tumor suppressive role of miRs in ovarian tumorigenesis (Merritt et al., 2008; Pampalakis et al., 2010). Overexpression of Drosha and Dicer was significantly associated with better survival, while low expression of Drosha was associated with suboptimal surgical cytoreduction and low expression of Dicer with advanced tumor stage, thereby further implicating the tumor suppressive role of microRNAs in OC (Merritt et al., 2008; Faggad et al., 2010). With respect to ovarian cancer, the potential targets for several upregulated miRs includes pro-apoptotic, metastasis-suppressing or antiproliferation gene products while those for the downregulated miRs includes growth signaling, prometastatic- or anti-apoptosis-associated proteins. A list of upregulated/downregulated miRs involved in ovarian cancer development is shown in **Table 10**. A list of aberrantly expressed miRs which could serve as a promising biomarker for detection of ovarian cancer has been summarized in **Table 11**. Chao et al. reported that in advanced stage cancer, miR-187 regulates carcinogenesis through Dab2 dependent EMT (epithelial-to-mesenchymal transition) (Chao et al., 2012, p. 2). Furthermore, other studies have described miR-199a, miR-200a, miR-200b, miR-200c, and miR-214 as significantly overexpressed and miR-100 and miRNAlet-7i as significantly downregulated in ovarian tumors (Iorio et al., 2007; Yang H. et al., 2008; Yang N. et al., 2008). Several miRNA signatures that could distinguish ovarian tumors based on histological subtypes has been studied such as miR-200b and miR-141 was observed to be overexpressed in serous and endometrioid subtypes; upregulated of miR-21, miR-203, and miR-205 correlated with endometrioid histotype; downregulated miR-145 correlated with serous and clear cell subtype, while downregulated miR 222 was associated with endometrioid and clear cell subtype (Iorio et al., 2007).

Recently, Braga et al. described methylation of miR-9-1, miR-9-3, and miR-130b which strongly correlated with progression of OC (Braga et al., 2018a). Different histotype of ovarian carcinomas reflect differential expression of specific miRNAs which might serve as a valid biomarker. Agostini et al. reported significant overexpression of miR-192, miR-194, and miR-215 in mucinous subtype of ovarian carcinomas. However their expression was downregulated in other subtypes and sex cord-stromal tumors (Agostini et al., 2018).

A list of promising aberrantly expressed miRs which could be of prognostic and predictive relevance in ovarian cancer has been summarized in **Table 11**. A lower ratio of miR-221 to miR-222 significantly correlated with worse overall survival in predominantly high grade, advanced stage sporadic ovarian carcinomas (Wurz et al., 2010). Downregulation of miR-141,

TABLE 9 | Non-invasive epigenetic DNA methylation biomarkers for ovarian cancer.

Epigenetic marker(s)	Source	Patient/sample	Clinical prediction	Technology	References
Methylation of ≥1 gene of BRCA1, RASSF1A, APC, p14ARF, p16INK4A, or DAPK	Serum and cytologically negative peritoneal fluid	50 serum, 40 peritoneal fluid from EOC patients along with 40 control serum and peritoneal fluid samples	Presence of malignancy Sensitivity: 41/50 (82% for serum) 28/30 (93% for peritoneal fluid), Specificity: 100% (all tumor stages).	Methylation-specific-PCR (MSP)	Ibanez de Caceres et al., 2004
DAPK	Whole peripheral blood DNA	26 peripheral blood samples	Sensitivity: 14/16 (54%) for DAPK-methylation-positive samples: Specificity: 10/10 (100%) for DAPK methylation negative	Methylation-specific PCR (MSP)	Collins et al., 2006
Methylation of ≥1 gene of SOX1, PAX1 or SFRP1	Serum	46 (26 ovarian cancer cases and 20 patients with a benign condition	Sensitivity: 73.08% Specificity: 75%	Methylation-specific PCR (MSP)	Su et al., 2009
Methylation of 7-gene panel (APC, RASSF1A, CDH1, RUNX3, TFPI2, SFRP5, and OPCML	Serum	202 patients (87 EOC cases, 53 benign cases and 62 controls)	Sensitivity of 7-gene panel: 85.3% Specificity of 7-gene panel: 90.5% in stage I EOC Sensitivity of CA125: 56.1% Specificity of CA125: 64.15%	Multiplex methylation-specific PCR (MSP)	Zhang et al., 2013
10-gene panel (Combination of BRCA1, HIC1,PAX5,PGR, THBS1)	Plasma	66 (33 cancer cases and 33 control)	Presence of malignancy Sensitivity: 61% Specificity: 85%	Microarray based multiplex assay(MethDet56 technique)	Melnikov et al., 2009
Several gene panel (RASSF1A and PGR-PROX) (RASSF1A, CALCA and EP300)	Serum	90 (30 EOC cases, 30 cases with Benign disease along with 30 controls non neoplastic samples)	Methylation of RASSF1A and PGR-PROX Sensitivity: 80.0% Specificity:73.3% Methylation of RASSF1A, CALCA and EP300 Sensitivity: 90.0% Specificity: 86.7%.	Microarray based Assay(MethDet 56)	Liggett et al., 2011
OPCML	Serum	194 (71 EOC, 43 benign and 80 controls non neoplastic samples)	Sensitivity: 87.18% Specificity: 93.75% Accuracy: 90.14%	Nested Methylation -specific PCR (MSP)	Wang et al., 2017
RASSF1A	Plasma	53 samples including OC tumors, adjacent tumor cell free tissues and paired plasma circulating tumor DNA	Sensitivity: 33/53 (62.3%), RASSF1A methylation of paired plasma CtDNA showed slight concordant with primary tumor samples ($P = 0.227$, 2-sided Pearson $\chi 2$ test, $k = 0.156$). Significantly correlates with overall survival	Real-time methylation specific-PCR (real-time MSP) and methylation-sensitive high-resolution melting analysis (MS-HRMA)	Giannopoulou et al., 2017
ESR1	Plasma	Group A: 66 OC cases Group B: 53 OC case along with 50 paired plasma samples	Sensitivity of detection: 38%. ESR1 methylation predicted better clinical outcomes: overall survival ($P = 0.027$), progression-free survival ($P = 0.041$)	Real-time methylation-specific PCR (real-time MSP) assay	Giannopoulou et al., 2018
3-gene panel	Serum	For assay development: 151 cases and for validation study 250 cases with different conditions in 3 sets	Pre-chemotherapy Sensitivity: 41.4% Specificity: 90.7% Post chemotherapy Responders: 78% non-responders: 86%	Targeted ultra-high coverage bisulfite sequencing	Widschwendter et al., 2017

miR−200a, miR-200b, miR-200c, and miR-429 correlated with poor progression free survival. Moreover, multivariate analysis of relevant clinicopathological variables such as debulking status, stage and grade of tumor revealed the correlation of miR-429 expression with recurrence-free survival (Leskela et al., 2010). Downregulated miR-422b and miR-34c correlated with decreased disease-specific survival in HGSOC patients with BRCA1/2 abnormalities (Lee et al., 2009).

In ovarian cancer, overexpression of miR-214 has been specifically associated with the degradation of PTEN mRNA which further leads to the activation of Akt pathway and has been correlated with platinum resistance (Yang H. et al., 2008). Downregulation of miR-Let7i has been reported in platinum-resistance ovarian tumors; however its gain of function resulted in restoration of drug sensitivity of chemoresistance OC cells (Yang N. et al., 2008).

TABLE 10 | List of dysregulated miRNAs in ovarian cancer.

	Mechanism(s)	miR(s)	Targets	Consequence(s) leading to tumorigenesis	References
Upregulated	c-myc activation	130a	MCSF, GAX and HOXA5	Chemotherapy resistance, angiogenesis, and dedifferentiation	Taylor and Gercel-Taylor, 2008; Eitan et al., 2009
	Copy gain	27a and 451	ZBTB10, Myt-1, HMGB2, HOXA2, and CYP1B1	Multidrug export, oncogenic signaling and reduced apoptotic potential	Shibata et al., 2006; Zhang L. et al., 2006
		213	APP and SATB2	Chemoresistance	Boren et al., 2009
		199a, 200a, b, c and 335	TGFβ, ZEB1, ZEB2, BAP1, GATA4, GATA6, TNC, FN1, EXOC5, and TUBB3	Mesenchymal–epithelial transition	Weisenberger et al., 2008; Su et al., 2009; Liggett et al., 2011; Giannopoulou et al., 2017; Wang et al., 2017
	Hypomethylation	203	p63, SOCS-3, ABL1, MCEF, and ADAMTS6	Unknown	Iorio et al., 2007; Lee et al., 2009
		205	ZEB1, ZEB2, E2F1, ERBB3, PKCe, and SHIP2	Mesenchymal–epithelial transition, oncogenic signal transduction	Iorio et al., 2007; Yu et al., 2008; Lee et al., 2009
		21	PDCD4†, RPS7†, NCAPG†, TPM1, and PTEN	Reduced apoptotic potential and anchorage independence	Iorio et al., 2007; Laios et al., 2008; Sorrentino et al., 2008; Eitan et al., 2009; Lee et al., 2009; Resnick et al., 2009; Yang et al., 2009
	Unknown	340	PAM, RTN3, PPL, RNF34, and ZNF513	Chemoresistance	Boren et al., 2009
		221/222	KIT, AIP1, p21, p57, TCF12, RIMS3, and ARNT	Cell cycle progression and angiogenesis	Merritt et al., 2008; Vlassov et al., 2010; Kneip et al., 2011
Downregulated	C/EBPα Downregulation	1	FOXP1, HDAC4 c-Met, Pim1 and HAND2	Tumor growth, cell motility and proliferation	Iorio et al., 2007
	DNA methylation and copy loss	137	CDK6, MITF, KLF12, and PDLIM3	Cell cycle progression and dedifferentiation	Iorio et al., 2007; Zhang L. et al., 2008
		140	c-SRK, MMP13 and FGF2	Oncogenic signaling	Iorio et al., 2007; Zhang L. et al., 2008
		150	c-Myb, MAK9, Akt3, and MAP2K4	Oncogenic signaling	Zhang L. et al., 2008
		551a	LPHN1, ERBB4, and ZFP36	Oncogenic signaling	Dahiya et al., 2008
		9	NF-kB†, Bcl2†, Bcl6, FGF†, and b-Raf	Oncogenic signaling	Leon et al., 1977; Weisenberger et al., 2008; Faggad et al., 2010; Kneip et al., 2011
		184	TTK69, K10, and Sax(A)	Dedifferentiation	Zhang L. et al., 2008
	Unknown	30b and d	CTGF	Invasion/metastasis	Laios et al., 2008
		98	HMGA2, LIN28B, and HIC2	Oncogenic signaling and cancer stemness	Dahiya et al., 2008
		517a and b	CREAP-1, MAPKAPK5, NFKBIE, and PTK2B	Chemoresistance and oncogenic signaling	Lee et al., 2009
		Let-7i	HMGA2, LIN28Bm TRIM71, and IGF2BP1	Chemoresistance	Yang N. et al., 2008
		662	NEGR1, MKX, and CSF3	Unknown	Dahiya et al., 2008

Several studies have recently highlighted the diagnostic and prognostic relevance of several miRNAs, their association with overall survival of patients and have shown that they could serve as putative biomarker as well as therapeutic target for ovarian cancer management. For instance, Li et al. have reported tumor suppressive role of miR-542-3p, which directly targets CDK14 and was observed significantly downregulated in EOC tissue and OC cell lines (Li et al., 2019).

Si et al. highlighting the therapeutic significance of miR-27a in OC, reported miR-27a mediated regulation of proliferation, chemosensitivity and invasion of OC by targeting Cullin 5 (CUL5) (Si et al., 2019, p. 5). Another study by Jia et al. reported the tumor suppressive role of miR-34 in regulation of tumor proliferation via inducting autophagy and apoptosis and suppression of cell invasion by targeting Notch 1 (Jia et al., 2019, p. 1). Wang et al. utilizing integrated meta-analysis approach

TABLE 11 | List of misexpressed miRNAs in ovarian cancer.

Epigenetic marker	Alternations in miRNAs	Source	Methodology	Clinical prediction	References
Dicer, Drosha mRNA	Downregulated	Tumor tissue and ovarian cancer cell lines	Quantitative RT-PCR, gene expression array	Advanced tumor stage and Suboptimal tumor debulking	Merritt et al., 2008; Faggad et al., 2010; Pampalakis et al., 2010
let-7i, miR-221,−30c,−152 and−193 miR-185,−106a,−181a, −210,−423,−103,−107 and let-7c	Downregulated Upregulated	Tumor tissue from Endometrial cancer, normal endometrial and atypical hyperplasia	Quantitative RT-PCR, microarray analysis	Association with Endometrial cancer development	Boren et al., 2008
miR-124-1,−124-2,−124-3,−127,−132,−137,−193A, 375 and−339	Downregulated	Tumor tissue	Quantitative RT-PCR, MSP, direct Sanger sequencing	EOC metastasis (including peritoneal macro-metastases)	Loginov et al., 2018
let-7i	Downregulated	Tumor tissue and ovarian cancer cell lines	miRNA Microarray, Stem-loop real-time RT-PCR (TaqMan miRNA assay)	Associated with increased resistance to chemotherapy drugs, cis-platinum, and short progression-free survival	Yang N. et al., 2008
miR-30c,−130a, and−335	Downregulated	Paclitaxel and cisplatin resistant cancer cell lines	miRNA Microarray, qPCR, Northern blots	Association with development of chemoresistance	Sorrentino et al., 2008
miR- 199b-5p	Downregulated	Cisplatin-sensitive and -resistant ovarian cancer cell lines	miRCURY LNA^{TM} microRNA array and Q-PCR	Development of acquired chemoresistance through the activation of JAG1-Notch1 signaling cascade	Liu et al., 2014
miR-34a, miR-34b*/c	Downregulated	Tumor tissue	Quantitative RT-PCR, *in- situ* hybridization	Associated with motility and invasion by regulation of MET, progression of disease to advanced stages	Corney et al., 2010
miR−34a,−200a,−200b, −449b,−509-3p,−509-3-5p, −513a-5p and −574-5p	Upregulated	Tumor tissue	MicroRNA microarray	Differentially expressed in Stage I disease	Eitan et al., 2009
miR-302b,−22, and−373 miR-148b and−211	Upregulated Downregulated	Tumor tissue and ovarian cancer cell lines	MicroRNA microarray, Quantitative Real time PCR (Taqman based)	Discriminates serous vs. non-serous disease	Iorio et al., 2007
miR-7, 34c-5p, 146b-5p and 449a	Upregulated	Tumor tissue	Massively parallel pyrosequencing, TaqMan qRT-PCR assays	Serous adenocarcinoma	Wyman et al., 2009
miR-23p,−125a-3p,−125a-5p,−130a,−146b-5p,−193a-3p,−193a-5p,−423-5p,−451 and−491-5p	Upregulated	Tumor tissue	MicroRNA microarray	Differentially expressed in Stage III disease	Eitan et al., 2009
miRs 100, 199a, 200a, and 214	Upregulated	Tumor tissue and ovarian cancer cell line	miRNA array and Northern blot analysis, quantitative reverse transcription-PCR.	Late clinical stage and high-grade tumors, negative regulation of PTEN by miR-214 thereby inducing cell survival and cisplatin resistance	Yang H. et al., 2008
miR-302b,−325,−299-5p,−222, and−324-3p miR-212 and−150	Upregulated Downregulated	Tumor tissue and ovarian cancer cell lines	MicroRNA microarray, Quantitative Real time PCR (Taqman based)	Differentiates Serous vs. endometrioid disease	Iorio et al., 2007
miR-325,−22,−302c,−299–5p,−373, and−196b miR-9 and−18	Upregulated Downregulated	Tumor tissue and ovarian cancer cell lines	MicroRNA microarray, Quantitative Real time PCR (Taqman based)	Associated with Poor differentiation	Iorio et al., 2007
miR-30a,−30a*, and−486-5p	Upregulated	Tumor tissue and ovarian cancer cell lines	MicroRNA microarray, Quantitative Real time PCR (Taqman based)	Clear cell disease	Iorio et al., 2007

(Continued)

TABLE 11 | Continued

Epigenetic marker	Alternations in miRNAs	Source	Methodology	Clinical prediction	References
Methylation of let-7a-3	Downregulated	Tumor tissue	Real-time methylation-specific PCR and real-time reverse Transcription-PCR, direct Sanger sequencing	Favorable prognosis	Lu et al., 2007, p. 3
miR-449b	Upregulated	Tumor tissue	MicroRNA microarray	Good prognosis	Eitan et al., 2009
let-7e	Upregulated	Paclitaxel and cisplatin resistant cancer cell lines	miRNA Microarray, qPCR, Northern blots	Associated with resistance to Paclitaxel	Sorrentino et al., 2008
miR-200a,−200b, and−429	Upregulated	Primary tumor and ovarian cancer cell line	Real-time reverse transcription-PCR	Long disease free survival and delayed recurrence, prognostic marker in advanced ovarian cancer	Cochrane et al., 2009; Hu et al., 2009
miRs 100, 199a, 200a, and 214	Upregulated	Tumor tissue and ovarian cancer cell line	miRNA array and Northern blot analysis, quantitative reverse transcription-PCR.	Late clinical stage and high-grade tumors	Yang H. et al., 2008
Methylation of miR-34a	Downregulated	Tumor tissue	Quantitative reverse transcription-PCR, MethyLight assay	Inversely associated with grade, p53 mutation, and dualistic tumor type. Reduced progression free survival and worsen overall survival.	Schmid et al., 2016
Methylation of miR-199a-3p	Downregulated	Tumor tissue and ovarian cancer cell line	Methylation-specific PCR and bisulphite sequencing	Tumor aggressiveness and enhanced cisplatin resistance through promoting DDR1 expression	Deng et al., 2017
Methylation of 10 miRNA genes (miR-124-2,−124-3,−125B-1,−127,−129-2,−137, −193A,−203A,−339,−375)	Downregulated	Tumors tissue and matched peritoneal metastases	Methylation-specific PCR	Involved in metastasis	Loginov et al., 2018
(miR-34b/c, miR-9-1, miR-124-3, miR-129-2, and miR-107)	Downregulated	Tumor tissue	Methylation-specific PCR	Associated with clinical grade and metastasis. High sensitivity and specificity reveals its diagnostic potential (87–94%, AUC = 0.92).	Braga et al., 2018b
miR-150	Downregulated	Tumor tissue	Real time PCR	Correlated with shorter progression free survival	Wilczynski et al., 2018
miR-4443 and miR-5195-3p	Downregulated	Tumor tissue	Real time PCR	Correlates with metastasis and tumor progression	Ebrahimi and Reiisi, 2019
miR-148a	Downregulated	Plasma samples and ovarian cancer cell line	Real time PCR	Associated with poor prognosis, tumor growth and metastasis	Gong et al., 2016
(hsa-miR-135, 150,−340, 625, 1908, 3187,−96,−196b,−449c, and−1275)	Downregulated	Tumor tissue	Small RNA sequencing, quantitative RT-PCR	Associated with survival	Chen et al., 2018
miR-595	Downregulated	Tumor tissue	qRT-PCR	Correlated with shorter overall survival	Zhou et al., 2017
miR-498	Downregulated	Tumor tissue	qPCR	Correlated with shorter overall survival and progression free survival	Cong et al., 2015

(Continued)

TABLE 11 | Continued

Epigenetic marker	Alternations in miRNAs	Source	Methodology	Clinical prediction	References
miR-9	Downregulated	Tumor tissue and ovarian cancer cell line	Real time PCR, luciferase reporter assay, Western Blot, Methylation study, RNAi technique, and cytotoxicity Assay	Resistance to paclitaxel by targeting CCNG1.	Li X. et al., 2015
miR-508-3p	Downregulated	Tumor tissue and ovarian cancer cell line	System biology approach, qRT-PCR, methylation PCR, RNA sequencing, immuno blot	Predictor for mesenchymal subtype and metastasis	Zhao et al., 2019
hsa-miR-1273g-3p	Downregulated	Serum samples of recurrent epithelial ovarian cancer patients	Microarray and qRT-PCR	Prognostic biomarker for recurrence	Günel et al., 2018

have shown the oncogenic role of miRNA-27a by mediating FOXO1 and its inhibition could serve as a new strategy in combating ovarian cancer (Wang Z. et al., 2018, p. 1). Hu et al. identified miR-934 as an oncogene in OC by directly targeting BRMS1L, and thus could serve as a therapeutic marker (Hu et al., 2019). It has been reported that miR-1294 was identified to be downregulated in EOC and correlated with tumor progression and shorter overall survival, thereby could serve as an independent prognostic indicator (Guo W. et al., 2018).

Liu et al. provided insights into the oncogenic role of microRNA-96 (miR-96-5p) in ovarian cancer. Its significant overexpression was found in tissue as well as serum samples. Overexpression of miR-96-5p was correlated with increased proliferation and migration by suppressing Caveolae1 (CAV1) and inhibiting AKT signaling pathway and its downstream proteins (Cyclin D1 and P70), thus implying that miR-96-5p could serve as a promising therapeutic target for ovarian cancer (Liu et al., 2019, p. 1). Similarly, Chaluvally-Raghavan et al. reported that miR551b-3p which is an oncogenic microRNA, directly upregulates STAT3 expression and further deregulates proliferation and metastasis *in vivo* and *in vitro*. Reduced expression of STAT3 in OC cells *in vitro* and *in vivo* via anti-miR551b-3p leads to reduction in growth of ovarian tumor *in vivo*, thereby implying that it could serve as promising therapeutic target in future for ovarian cancer (Chaluvally-Raghavan et al., 2016).

In another study, miR-152 mediated suppression of tumor proliferation along with promotion of apoptosis via repression of ERBB3 was reported, thus demonstrating miR-152 as a potential therapeutic target (Li et al., 2017, p. 3). Liu et al. reported association of miR-506 with better response to therapy as well as long PFS and overall survival in OC patients. Further, it sensitized cancer cells to chemotherapy by directly targeting RAD51 and thus could be of therapeutic importance (Liu et al., 2015).

10 miRs which were identified using genome wide MicroRNA expression profiling were capable to discriminate malignant tissue samples from normal with a sensitivity of 97% and specificity of 92% (Wang et al., 2014). Biamonte et al. have reported tumor suppressive role of miR-let-7g and significant association of its reduced expression in both tissue as well as serum with chemoresistance in advanced stage EOC patient which reflects its potential as a predictive biomarker to monitor response to chemotherapy (Biamonte et al., 2019). Kobayashi et al. have shown significant overexpression of serum miR-1290 in advance stage HGSOC in comparison to early stage. Moreover, it was capable to discriminate patients with HGSOC from patients with malignancies of other histological subtypes with a sensitivity of 47% and specificity of 85% (AUC = 0.76), thus reflecting diagnostic potential of miR-1290 for HGSOC (Kobayashi et al., 2018).

Mahmoud et al. examined the diagnostic significance of serum miR-21 and reported that its upregulation was significantly negatively correlated with Programmed Cell Death-4 (PDCD4) expression in EOC patients (Mahmoud et al., 2018). Another study highlighted significantly elevated expression of serum exosomal miR-93, miR-145, and miR-200c in OC. Moreover, the sensitivity for miR-145 and miR-200c was 91.6 and 90.0% which was far superior in comparison with CA125, thus these serum exosomal microRNAs could be of diagnostic relevance for preoperative diagnosis of OC (Kim et al., 2019). miR-21 was observed significantly overexpressed in the sera of EOC patients and its elevated expression correlated with shorter overall survival (Xu et al., 2013). Further, downregulation of serum miR-25 and miR-93 and upregulation of miR-7 and miR-429 have been reported in OC patients. In addition, the sensitivity and specificity achieved by these four serum miRs were 93 and 92% to discriminate cancer patients from non-neoplastic control samples, deciphering their diagnostic significance in EOC. Moreover serum miR-429 correlated with overall survival and could serve as an independent prognostic indicator (Meng et al., 2015). Findings from another study reveal the relevance of serum miR-141 and miR-200c in OC diagnosis and prognosis. Both of these miRs were identified to be overexpressed in serum of EOC patients; however miR-200c displayed a descending expression

trend across tumor stage (early to advance) while an escalating expression trend was observed in case of miR-141. Moreover, the sensitivity for miR-141 and miR-200c were 0.69 and 0.72 with a specificity of 0.72 and 0.70, respectively, to discriminate cancerous samples from normal control [AUC = 0.75 and 0.79, respectively]. Furthermore, high serum miR-200c correlated with higher survival rate. On contrary, low serum miR-141 correlated with higher survival rate (Gao and Wu, 2015).

Langhe et al. using Exiqon platform explored a 4-miR panel in serum of EOC patients for their diagnostic utility and found that these miRs were significantly downregulated in EOC patients. Furthermore these miRs target WNT signaling, AKT/mTOR signaling and TLR-4/MyD88 to regulate ovarian cancer progression and resistance (Langhe et al., 2015). Overexpression of serum miR-200a, miR-200b, and miR-200c which have been observed in EOC patients, correlates with aggressive disease progression and could be indicative of disease prognosis and patient survival (Zuberi et al., 2015). Higher serum concentration of exosomal miR-200b and miR-200c correlated with shorter overall survival, which suggests its prognostic relevance. (Meng et al., 2016b) Serum miR-200a, miR-200b, and miR-200c differentiated cancerous and benign tumors with 83% sensitivity and 100% specificity, which reflect that these miRs could be of diagnostic utility (Meng et al., 2016a).

These miRs though hold great potential for their utility in ovarian cancer management; however its therapeutical implementation still remains a challenge. To address this, well-designed clinical study as well as validated methodologies is essentially warranted.

Expert Commentary

It is now well-established that DNA methylation occurs very early in malignant transformation and their utility as biomarker holds great promise to overcome the false positive detection of ovarian cancer using current standard serum marker CA125. In this review, we highlight the recent epigenetic biomarkers analyzed in tissue and body fluids for early detection of OC. Strikingly; to date no single epigenetic biomarker facilitating early diagnosis of OC has made transition to the clinics. The probable reasons for this could be: the heterogeneous nature of EOC, difference in sample processing, assay design, technique used and approach could explain the variations observed in methylation frequencies amongst various studies for individual genes. Most of the studies for methylation analysis of genes were conducted on small sample size and in particular the normal control samples were insufficient to conclude the specificity of the assay. Therefore, further studies on larger sample size are necessary to be conducted to determine the potential of methylation if it could serve as biomarker for early EOC screening or not. Another limitation is the absence of standardized reference value for methylation analysis when trying to analyze if a particular locus is hyper or hypomethylated. To overcome this, currently, methylation cut off points which are based on already published reports or consensus are used.

The majority of the reports highlight the methylation status of gene or genes in a panel. No epigenetic biomarker screening study has been performed till date. However, for the detection figures approaching current screening modalities (89.5% sensitivity and 99.8% specificity) has been achieved by Ibanez de Caceres et al. (2004) with 82% sensitivity and 100% specificity (Ibanez de Caceres et al., 2004). All 30 control cases showed 0 false positive rate and further replication of the study on the basis of this sample size would give a false positive rate between 0 and 11.4% (95% confidence interval), thereby indicating that perfect specificity would unlikely hold up in the follow-up studies. In view of these considerations regarding the study of Ibanez de Caceres et al. are left to follow-up studies to shed light on. However, none of the report has been further validated undertaking follow up studies on a larger cohort and prospective study design thereby limiting the utility of the reported findings.

Molecular analysis of epigenetic modification (methylation) in circulating cell free tumor DNA in fluids serves as a novel, non-invasive approach for identification of potential promising cancer biomarkers, which can be performed at multiple time points and probably better reflects the prevailing molecular profile of cancer. Very few studies analyzing the methylation status of genes in blood-based assay for ovarian cancer diagnosis has been reported. Careful precision handling and processing of liquid biopsy for cell free DNA extraction is critically needed.

Future Prospects

Over the last decade, an exponential progress in DNA methylation based biomarker development has been witnessed. Owing to the stability of DNA and methylation pattern, a number of cfDNA as well as tissue based screening assay has paved its way into clinics. The commercial success of several tests based on DNA methylation biomarkers for early detection of colon, lung and prostate cancer and prediction of bladder cancer along with various markers under validation study shows that the time for transition into clinics can be relatively rapid. New technologies which allow rapid identification of methylation signatures directly from blood will facilitate sample-to answer solutions thereby enabling next-generation point of care molecular diagnostics. Moreover, ongoing work on liquid biopsies together with the recent advanced technologies such as digital PCR, bisulfite sequencing, methyl immune-precipitation coupled with next-generation sequencing, and methylation arrays along with advanced statistical data analysis may mitigate the problematic issues for the development of non-invasive method thereby overcoming the existing challenges to personalized medicine.

AUTHOR CONTRIBUTIONS

AS wrote the manuscript. SG and MS edited the final version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2019.00182/full#supplementary-material

REFERENCES

Abdollahi, A., Pisarcik, D., Roberts, D., Weinstein, J., Cairns, P., and Hamilton, T. C. (2003). LOT1 (PLAGL1/ZAC1), the candidate tumor suppressor gene at chromosome 6q24-25, is epigenetically regulated in cancer. *J. Biol. Chem.* 278, 6041–6049. doi: 10.1074/jbc.M210361200

Agathanggelou, A., Honorio, S., Macartney, D. P., Martinez, A., Dallol, A., Rader, J., et al. (2001). Methylation associated inactivation of RASSF1A from region 3p21.3 in lung, breast and ovarian tumours. *Oncogene* 20, 1509–1518. doi: 10.1038/sj.onc.1204175

Agostini, A., Brunetti, M., Davidson, B., Trop,é, C. G., Eriksson, A. G. Z., Heim, S., et al. (2018). The microRNA miR-192/215 family is upregulated in mucinous ovarian carcinomas. *Sci. Rep.* 8:11069. doi: 10.1038/s41598-018-29332-7

Akahira, J., Sugihashi, Y., Ito, K., Niikura, H., Okamura, K., and Yaegashi, N. (2004a). Promoter methylation status and expression of TMS1 gene in human epithelial ovarian cancer. *Cancer Sci.* 95, 40–43. doi: 10.1111/j.1349-7006.2004.tb03168.x

Akahira, J., Sugihashi, Y., Suzuki, T., Ito, K., Niikura, H., Moriya, T., et al. (2004b). Decreased expression of 14-3-3 sigma is associated with advanced disease in human epithelial ovarian cancer: its correlation with aberrant DNA methylation. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 10, 2687–2693. doi: 10.1158/1078-0432.CCR-03-0510

Amonkar, S. D., Bertenshaw, G. P., Chen, T.-H., Bergstrom, K. J., Zhao, J., Seshaiah, P., et al. (2009). Development and preliminary evaluation of a multivariate index assay for ovarian cancer. *PLoS ONE* 4:e4599. doi: 10.1371/journal.pone.0004599

Andersen, M. R., Goff, B. A., Lowe, K. A., Scholler, N., Bergan, L., Drescher, C. W., et al. (2010). Use of a symptom index, CA125, and HE4 to predict ovarian cancer. *Gynecol. Oncol.* 116, 378–383. doi: 10.1016/j.ygyno.2009.10.087

Arnold, J. M., Cummings, M., Purdie, D., and Chenevix-Trench, G. (2001). Reduced expression of intercellular adhesion molecule-1 in ovarian adenocarcinomas. *Br. J. Cancer* 85, 1351–1358. doi: 10.1054/bjoc.2001.2075

Bai, T., Tanaka, T., Yukawa, K., Maeda, M., and Umesaki, N. (2004). Reduced expression of death-associated protein kinase in human uterine and ovarian carcinoma cells. *Oncol. Rep.* 11, 661–665. doi: 10.3892/or.11.3.661

Balch, C., Matei, D. E., Huang, T. H.-M., and Nephew, K. P. (2010). Role of epigenomics in ovarian and endometrial cancers. *Epigenomics* 2, 419–447. doi: 10.2217/epi.10.19

Baldwin, R. L., Nemeth, E., Tran, H., Shvartsman, H., Cass, I., Narod, S., et al. (2000). BRCA1 promoter region hypermethylation in ovarian carcinoma: a population-based study. *Cancer Res.* 60, 5329–5333.

Baranova, I., Kovarikova, H., Laco, J., Dvorak, O., Sedlakova, I., Palicka, V., et al. (2018). Aberrant methylation of PCDH17 gene in high-grade serous ovarian carcinoma. *Cancer Biomark.* 23, 125–133. doi: 10.3233/CBM-181493

Barbolina, M. V., Adley, B. P., Kelly, D. L., Shepard, J., Fought, A. J., Scholtens, D., et al. (2009). Downregulation of connective tissue growth factor by three-dimensional matrix enhances ovarian carcinoma cell invasion. *Int. J. Cancer* 125, 816–825. doi: 10.1002/ijc.24347

Bartlett, T. E., Chindera, K., McDermott, J., Breeze, C. E., Cooke, W. R., Jones, A., et al. (2016). Epigenetic reprogramming of fallopian tube fimbriae in BRCA mutation carriers defines early ovarian cancer evolution. *Nat. Commun.* 7:11620. doi: 10.1038/ncomms11620

Barton, C. A., Gloss, B. S., Qu, W., Statham, A. L., Hacker, N. F., Sutherland, R. L., et al. (2010). Collagen and calcium-binding EGF domains 1 is frequently inactivated in ovarian cancer by aberrant promoter hypermethylation and modulates cell migration and survival. *Br. J. Cancer* 102, 87–96. doi: 10.1038/sj.bjc.6605429

Barton, C. A., Hacker, N. F., Clark, S. J., and O'Brien, P. M. (2008). DNA methylation changes in ovarian cancer: implications for early diagnosis, prognosis and treatment. *Gynecol. Oncol.* 109, 129–139. doi: 10.1016/j.ygyno.2007.12.017

Belinsky, S. A. (2004). Gene-promoter hypermethylation as a biomarker in lung cancer. *Nat. Rev. Cancer* 4, 707–717. doi: 10.1038/nrc1432

Biamonte, F., Santamaria, G., Sacco, A., Perrone, F. M., Di Cello, A., Battaglia, A. M., et al. (2019). MicroRNA let-7g acts as tumor suppressor and predictive biomarker for chemoresistance in human epithelial ovarian cancer. *Sci. Rep.* 9:5668. doi: 10.1038/s41598-019-42221-x

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21. doi: 10.1101/gad.947102

Bird, A. P., and Wolffe, A. P. (1999). Methylation-induced repression–belts, braces, and chromatin. *Cell* 99, 451–454. doi: 10.1016/S0092-8674(00)81532-9

Board, R. E., Knight, L., Greystoke, A., Blackhall, F. H., Hughes, A., Dive, C., et al. (2008). DNA methylation in circulating tumour DNA as a biomarker for cancer. *Biomark. Insights* 2, 307–319. doi: 10.1177/117727190700200003

Bol, G. M., Suijkerbuijk, K. P. M., Bart, J., Vooijs, M., van der Wall, E., and van Diest, P. J. (2010). Methylation profiles of hereditary and sporadic ovarian cancer. *Histopathology* 57, 363–370. doi: 10.1111/j.1365-2559.2010.03642.x

Bonito, N. A., Borley, J., Wilhelm-Benartzi, C. S., Ghaem-Maghami, S., and Brown, R. (2016). Epigenetic regulation of the homeobox gene MSX1 associates with platinum-resistant disease in high-grade serous epithelial ovarian cancer. *Clin. Cancer Res.* 22, 3097–3104. doi: 10.1158/1078-0432.CCR-15-1669

Boren, T., Xiong, Y., Hakam, A., Wenham, R., Apte, S., Chan, G., et al. (2009). MicroRNAs and their target messenger RNAs associated with ovarian cancer response to chemotherapy. *Gynecol. Oncol.* 113, 249–255. doi: 10.1016/j.ygyno.2009.01.014

Boren, T., Xiong, Y., Hakam, A., Wenham, R., Apte, S., Wei, Z., et al. (2008). MicroRNAs and their target messenger RNAs associated with endometrial carcinogenesis. *Gynecol. Oncol.* 110, 206–215. doi: 10.1016/j.ygyno.2008.03.023

Borgoño, C. A., and Diamandis, E. P. (2004). The emerging roles of human tissue kallikreins in cancer. *Nat. Rev. Cancer* 4, 876–890. doi: 10.1038/nrc1474

Bowtell, D. D. L. (2010). The genesis and evolution of high-grade serous ovarian cancer. *Nat. Rev. Cancer* 10, 803–808. doi: 10.1038/nrc2946

Brachova, P., Mueting, S. R., Carlson, M. J., Goodheart, M. J., Button, A. M., Mott, S. L., et al. (2015). TP53 oncomorphic mutations predict resistance to platinum- and taxane-based standard chemotherapy in patients diagnosed with advanced serous ovarian carcinoma. *Int. J. Oncol.* 46, 607–618. doi: 10.3892/ijo.2014.2747

Braga, E. A., Loginov, V. I., Burdennyi, A. M., Filippova, E. A., Pronina, I. V., Kurevlev, S. V., et al. (2018a). Five Hypermethylated MicroRNA genes as potential markers of ovarian cancer. *Bull. Exp. Biol. Med.* 164, 351–355. doi: 10.1007/s10517-018-3988-y

Braga, E. A., Loginov, V. I., Filippova, E. A., Burdennyi, A. M., Pronina, I. V., Kazubskaya, T. P., et al. (2018b). Diagnostic value of a group of MicroRNA genes hypermethylated in ovarian carcinoma. *Bull. Exp. Biol. Med.* 166, 253–256. doi: 10.1007/s10517-018-4326-0

Brock, M. V., Hooker, C. M., Ota-Machida, E., Han, Y., Guo, M., Ames, S., et al. (2008). DNA methylation markers and early recurrence in stage I lung cancer. *N. Engl. J. Med.* 358, 1118–1128. doi: 10.1056/NEJMoa0706550

Buchholtz, M.-L., Brüning, A., Mylonas, I., and Jückstock, J. (2014). Epigenetic silencing of the LDOC1 tumor suppressor gene in ovarian cancer cells. *Arch. Gynecol. Obstet.* 290, 149–154. doi: 10.1007/s00404-014-3177-9

Buys, S. S., Partridge, E., Black, A., Johnson, C. C., Lamerato, L., Isaacs, C., et al. (2011). Effect of screening on ovarian cancer mortalitys (PLCO) cancer screening randomized controlled trial. *JAMA* 305, 2295–2303. doi: 10.1001/jama.2011.766

Cacan, E. (2016). Histone deacetylase-1-mediated suppression of FAS in chemoresistant ovarian cancer cells. *Anticancer Res.* 36, 2819–2826.

Cacan, E. (2017). Epigenetic regulation of RGS2 (Regulator of G-protein signaling 2) in chemoresistant ovarian cancer cells. *J. Chemother.* 29, 173–178. doi: 10.1080/1120009X.2016.1277007

Cai, K. Q., Caslini, C., Capo-chichi, C. D., Slater, C., Smith, E. R., Wu, H., et al. (2009). Loss of GATA4 and GATA6 expression specifies ovarian cancer histological subtypes and precedes neoplastic transformation of ovarian surface epithelia. *PLoS ONE* 4:e6454. doi: 10.1371/journal.pone.0006454

Cai, L., Abe, M., Izumi, S., Imura, M., Yasugi, T., and Ushijima, T. (2007). Identification of PRTFDC1 silencing and aberrant promoter methylation of GPR150, ITGA8 and HOXD11 in ovarian cancers. *Life Sci.* 80, 1458–1465. doi: 10.1016/j.lfs.2007.01.015

Cannistra, S. A. (2004). Cancer of the ovary. *N. Engl. J. Med.* 351, 2519–2529. doi: 10.1056/NEJMra041842

Cardenas, H., Vieth, E., Lee, J., Segar, M., Liu, Y., Nephew, K. P., et al. (2014). TGF-β induces global changes in DNA methylation during the epithelial-to-mesenchymal transition in ovarian cancer cells. *Epigenetics* 9, 1461–1472. doi: 10.4161/15592294.2014.971608

Caslini, C., Capo-Chichi, C. D., Roland, I. H., Nicolas, E., Yeung, A. T., and Xu, X. X. (2006). Histone modifications silence the GATA transcription factor genes in ovarian cancer. *Oncogene* 25:5446. doi: 10.1038/sj.onc.1209533

Chaluvally-Raghavan, P., Jeong, K. J., Pradeep, S., Silva, A. M., Yu, S., Liu, W., et al. (2016). Direct upregulation of STAT3 by MicroRNA-551b-3p deregulates growth and metastasis of ovarian cancer. *Cell Rep.* 15, 1493–1504. doi: 10.1016/j.celrep.2016.04.034

Chan, M. W., Huang, Y.-W., Hartman-Frey, C., Kuo, C.-T., Deatherage, D., Qin, H., et al. (2008). Aberrant transforming growth factor beta1 signaling and SMAD4 nuclear translocation confer epigenetic repression of ADAM19 in ovarian cancer. *Neoplasia* 10, 908–919. doi: 10.1593/neo.08540

Chan, M. W. Y. (2005). Hypermethylation of 18S and 28S Ribosomal DNAs predicts progression-free survival in patients with ovarian cancer. *Clin. Cancer Res.* 11, 7376–7383. doi: 10.1158/1078-0432.CCR-05-1100

Chan, T. A., Glockner, S., Yi, J. M., Chen, W., Van Neste, L., Cope, L., et al. (2008). Convergence of mutation and epigenetic alterations identifies common genes in cancer that predict for poor prognosis. *PLoS Med.* 5:e114. doi: 10.1371/journal.pmed.0050114

Chao, A., Lin, C.-Y., Lee, Y.-S., Tsai, C.-L., Wei, P.-C., Hsueh, S., et al. (2012). Regulation of ovarian cancer progression by microRNA-187 through targeting disabled homolog-2. *Oncogene* 31, 764–775. doi: 10.1038/onc.2011.269

Chen, S. F., Liu, Z., Chaurasiya, S., Dellinger, T. H., Lu, J., Wu, X., et al. (2018). Identification of core aberrantly expressed microRNAs in serous ovarian carcinoma. *Oncotarget* 9, 20451–20466. doi: 10.18632/oncotarget.24942

Cheng, W., Jiang, Y., Liu, C., Shen, O., Tang, W., and Wang, X. (2010). Identification of aberrant promoter hypomethylation of HOXA10 in ovarian cancer. *J. Cancer Res. Clin. Oncol.* 136, 1221–1227. doi: 10.1007/s00432-010-0772-4

Chien, J., Staub, J., Avula, R., Zhang, H., Liu, W., Hartmann, L. C., et al. (2005). Epigenetic silencing of TCEAL7 (Bex4) in ovarian cancer. *Oncogene* 24, 5089–5100. doi: 10.1038/sj.onc.1208700

Choi, Y.-L., Kang, S. Y., Shin, Y. K., Choi, J. S., Kim, S. H., Lee, S.-J., et al. (2006). Aberrant hypermethylation of RASSF1A promoter in ovarian borderline tumors and carcinomas. *Virchows Arch. Int. J. Pathol.* 448, 331–336. doi: 10.1007/s00428-005-0091-3

Choi, Y.-L., Kim, J., Kwon, M. J., Choi, J. S., Kim, T.-J., Bae, D.-S., et al. (2007). Expression profile of tight junction protein claudin 3 and claudin 4 in ovarian serous adenocarcinoma with prognostic correlation. *Histol. Histopathol.* 22, 1185–1195. doi: 10.14670/HH-22.1185

Chou, J.-L., Huang, R.-L., Shay, J., Chen, L.-Y., Lin, S.-J., Yan, P. S., et al. (2015). Hypermethylation of the TGF-β target, ABCA1 is associated with poor prognosis in ovarian cancer patients. *Clin. Epigenetics* 7:1. doi: 10.1186/s13148-014-0036-2

Chou, J.-L., Su, H.-Y., Chen, L.-Y., Liao, Y.-P., Hartman-Frey, C., Lai, Y.-H., et al. (2010). Promoter hypermethylation of FBXO32, a novel TGF-beta/SMAD4 target gene and tumor suppressor, is associated with poor prognosis in human ovarian cancer. *Lab. Investig. J. Tech. Methods Pathol.* 90, 414–425. doi: 10.1038/labinvest.2009.138

Chung, V. Y., Tan, T. Z., Huang, R.-L., Lai, H.-C., and Huang, R. Y.-J. (2017). Loss of discoidin domain receptor 1 (DDR1) via CpG methylation during EMT in epithelial ovarian cancer. *Gene* 635, 9–15. doi: 10.1016/j.gene.2017.09.001

Cicek, M. S., Koestler, D. C., Fridley, B. L., Kalli, K. R., Armasu, S. M., Larson, M. C., et al. (2013). Epigenome-wide ovarian cancer analysis identifies a methylation profile differentiating clear-cell histology with epigenetic silencing of the HERG K+ channel. *Hum. Mol. Genet.* 22, 3038–3047. doi: 10.1093/hmg/ddt160

Clarke-Pearson, D. L. (2009). Clinical practice. Screening for ovarian cancer. *N. Engl. J. Med.* 361, 170–177. doi: 10.1056/NEJMcp0901926

Cochrane, D. R., Spoelstra, N. S., Howe, E. N., Nordeen, S. K., and Richer, J. K. (2009). MicroRNA-200c mitigates invasiveness and restores sensitivity to microtubule-targeting chemotherapeutic agents. *Mol. Cancer Ther.* 8, 1055–1066. doi: 10.1158/1535-7163.MCT-08-1046

Collins, Y., Dicioccio, R., Keitz, B., Lele, S., and Odunsi, K. (2006). Methylation of death-associated protein kinase in ovarian carcinomas. *Int. J. Gynecol. Cancer Off. J. Int. Gynecol. Cancer Soc.* 16 (Suppl. 1), 195–199. doi: 10.1136/ijgc-00009577-200602001-00031

Cong, J., Liu, R., Wang, X., Wang, J., Wang, H., and Hou, J. (2015). Low miR-498 expression levels are associated with poor prognosis in ovarian cancer. *Eur. Rev. Med. Pharmacol. Sci.* 19, 4762–4765.

Corney, D. C., Hwang, C.-I., Matoso, A., Vogt, M., Flesken-Nikitin, A., Godwin, A. K., et al. (2010). Frequent downregulation of miR-34 family in human ovarian cancers. *Clin. Cancer Res.* 16, 1119–1128. doi: 10.1158/1078-0432.CCR-09-2642

Costello, J. F., and Plass, C. (2001). Methylation matters. *J. Med. Genet.* 38, 285–303. doi: 10.1136/jmg.38.5.285

Curry, E., Zeller, C., Masrour, N., Patten, D. K., Gallon, J., Wilhelm-Benartzi, C. S., et al. (2018). Genes predisposed to DNA hypermethylation during acquired resistance to chemotherapy are identified in ovarian tumors by bivalent chromatin domains at initial diagnosis. *Cancer Res.* 78, 1383–1391. doi: 10.1158/0008-5472.CAN-17-1650

Czekierdowski, A., Czekierdowska, S., Szymanski, M., Wielgos, M., Kaminski, P., and Kotarski, J. (2006a). Opioid-binding protein/cell adhesion molecule-like (OPCML) gene and promoter methylation status in women with ovarian cancer. *Neuro Endocrinol. Lett.* 27, 609–613.

Czekierdowski, A., Czekierdowska, S., Wielgos, M., Smolen, A., Kaminski, P., and Kotarski, J. (2006b). The role of CpG islands hypomethylation and abnormal expression of neuronal protein synuclein-gamma (SNCG) in ovarian cancer. *Neuro Endocrinol. Lett.* 27, 381–386.

Dahiya, N., Sherman-Baust, C. A., Wang, T.-L., Davidson, B., Shih, I.-M., Zhang, Y., et al. (2008). MicroRNA expression and identification of putative miRNA targets in ovarian cancer. *PLoS ONE* 3:e2436. doi: 10.1371/journal.pone.0002436

Dai, W., Teodoridis, J. M., Zeller, C., Graham, J., Hersey, J., Flanagan, J. M., et al. (2011). Systematic CpG Islands methylation profiling of genes in the Wnt pathway in epithelial ovarian cancer identifies biomarkers of progression-free survival. *Clin. Cancer Res.* 17, 4052–4062. doi: 10.1158/1078-0432.CCR-10-3021

Dai, W., Zeller, C., Masrour, N., Siddiqui, N., Paul, J., and Brown, R. (2013). Promoter CpG Island methylation of genes in key cancer pathways associates with clinical outcome in high-grade serous ovarian cancer. *Clin. Cancer Res.* 19, 5788–5797. doi: 10.1158/1078-0432.CCR-13-1217

de Leon, M., Cardenas, H., Vieth, E., Emerson, R., Segar, M., Liu, Y., et al. (2016). Transmembrane protein 88 (TMEM88) promoter hypomethylation is associated with platinum resistance in ovarian cancer. *Gynecol. Oncol.* 142, 539–547. doi: 10.1016/j.ygyno.2016.06.017

De Mattos-Arruda, L., Weigelt, B., Cortes, J., Won, H. H., Ng, C. K. Y., Nuciforo, P., et al. (2014). Capturing intra-tumor genetic heterogeneity by *de novo* mutation profiling of circulating cell-free tumor DNA: a proof-of-principle. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* 25, 1729–1735. doi: 10.1093/annonc/mdu239

Deng, Y., Zhao, F., Hui, L., Li, X., Zhang, D., Lin, W., et al. (2017). Suppressing miR-199a-3p by promoter methylation contributes to tumor aggressiveness and cisplatin resistance of ovarian cancer through promoting DDR1 expression. *J. Ovarian Res.* 10. doi: 10.1186/s13048-017-0333-4

Deng, Z., Wang, L., Hou, H., Zhou, J., and Li, X. (2016). Epigenetic regulation of IQGAP2 promotes ovarian cancer progression via activating Wnt/β-catenin signaling. *Int. J. Oncol.* 48, 153–160. doi: 10.3892/ijo.2015.3228

Devouassoux-Shisheboran, M., and Genestie, C. (2015). Pathobiology of ovarian carcinomas. *Chin. J. Cancer* 34, 50–55. doi: 10.5732/cjc.014.10273

Diamandis, E. P., Scorilas, A., Fracchioli, S., Van Gramberen, M., De Bruijn, H., Henrik, A., et al. (2003). Human kallikrein 6 (hK6): a new potential serum biomarker for diagnosis and prognosis of ovarian carcinoma. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* 21, 1035–1043. doi: 10.1200/JCO.2003.02.022

Earp, M. A., and Cunningham, J. M. (2015). DNA methylation changes in epithelial ovarian cancer histotypes. *Genomics* 106, 311–321. doi: 10.1016/j.ygeno.2015.09.001

Ebrahimi, S. O., and Reiisi, S. (2019). Downregulation of miR-4443 and miR-5195-3p in ovarian cancer tissue contributes to metastasis and tumorigenesis. *Arch. Gynecol. Obstet.* 299, 1453–1458. doi: 10.1007/s00404-019-05107-x

Edgell, T., Martin-Roussety, G., Barker, G., Autelitano, D. J., Allen, D., Grant, P., et al. (2010). Phase II biomarker trial of a multimarker diagnostic for ovarian cancer. *J. Cancer Res. Clin. Oncol.* 136, 1079–1088. doi: 10.1007/s00432-009-0755-5

Eissa, M. A. L., Lerner, L., Abdelfatah, E., Shankar, N., Canner, J. K., Hasan, N. M., et al. (2019). Promoter methylation of ADAMTS1 and BNC1 as potential biomarkers for early detection of pancreatic cancer in blood. *Clin. Epigenetics* 11:59. doi: 10.1186/s13148-019-0650-0

Eitan, R., Kushnir, M., Lithwick-Yanai, G., David, M. B., Hoshen, M., Glezerman, M., et al. (2009). Tumor microRNA expression patterns associated with

resistance to platinum based chemotherapy and survival in ovarian cancer patients. *Gynecol. Oncol.* 114, 253–259. doi: 10.1016/j.ygyno.2009.04.024

Esteller, M. (2007). Epigenetic gene silencing in cancer: the DNA hypermethylome. *Hum. Mol. Genet.* 16, R50–R59. doi: 10.1093/hmg/ddm018

Faggad, A., Budczies, J., Tchernitsa, O., Darb-Esfahani, S., Sehouli, J., Müller, B. M., et al. (2010). Prognostic significance of Dicer expression in ovarian cancer-link to global microRNA changes and oestrogen receptor expression. *J. Pathol.* 220, 382–391. doi: 10.1002/path.2658

Fang, F., Cardenas, H., Huang, H., Jiang, G., Perkins, S. M., Zhang, C., et al. (2018). Genomic and epigenomic signatures in ovarian cancer associated with resensitization to platinum drugs. *Cancer Res.* 78, 631–644. doi: 10.1158/0008-5472.CAN-17-1492

Feinberg, A. P., and Tycko, B. (2004). The history of cancer epigenetics. *Nat. Rev. Cancer* 4, 143–153. doi: 10.1038/nrc1279

Feng, Q., Deftereos, G., Hawes, S. E., Stern, J. E., Willner, J. B., Swisher, E. M., et al. (2008). DNA hypermethylation, Her-2/neu overexpression and p53 mutations in ovarian carcinoma. *Gynecol. Oncol.* 111, 320–329. doi: 10.1016/j.ygyno.2008.07.036

Feng, W., Marquez, R. T., Lu, Z., Liu, J., Lu, K. H., Issa, J.-P. J., et al. (2008). Imprinted tumor suppressor genes ARHI and PEG3 are the most frequently down-regulated in human ovarian cancers by loss of heterozygosity and promoter methylation. *Cancer* 112, 1489–1502. doi: 10.1002/cncr.23323

Fiegl, H., Windbichler, G., Mueller-Holzner, E., Goebel, G., Lechner, M., Jacobs, I. J., et al. (2008). HOXA11 DNA methylation–a novel prognostic biomarker in ovarian cancer. *Int. J. Cancer* 123, 725–729. doi: 10.1002/ijc.23563

Fu, Y., Campbell, E. J., Shepherd, T. G., and Nachtigal, M. W. (2003). Epigenetic regulation of proprotein convertase PACE4 gene expression in human ovarian cancer cells. *Mol. Cancer Res.* 1, 569–576.

Fu, Y., Chen, J., Pang, B., Li, C., Zhao, J., and Shen, K. (2015). EZH2-Induced H3K27me3 is Associated with epigenetic repression of the ARHI tumor-suppressor gene in ovarian cancer. *Cell Biochem. Biophys.* 71, 105–112. doi: 10.1007/s12013-014-0168-1

Furlan, D. (2006). The high frequency of *de novo* promoter methylation in synchronous primary endometrial and ovarian carcinomas. *Clin. Cancer Res.* 12, 3329–3336. doi: 10.1158/1078-0432.CCR-05-2679

Gadomska, H., Grzechocinska, B., Janecki, J., Nowicka, G., Powolny, M., and Marianowski, L. (2005). Serum lipids concentration in women with benign and malignant ovarian tumours. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 120, 87–90. doi: 10.1016/j.ejogrb.2004.02.045

Gao, Y., Lou, G., Zhang, G.-M., Sun, X.-W., Ma, Y.-Y., Yang, Y.-M., et al. (2009). CHFR promoter hypermethylation and reduced CHFR mRNA expression in ovarian cancer. *Int. J. Biol. Markers* 24, 83–89. doi: 10.1177/172460080902400204

Gao, Y., and Wu, J. (2015). MicroRNA-200c and microRNA-141 as potential diagnostic and prognostic biomarkers for ovarian cancer. *Tumor Biol.* 36, 4843–4850. doi: 10.1007/s13277-015-3138-3

Geomini, P., Kruitwagen, R., Bremer, G. L., Cnossen, J., and Mol, B. W. J. (2009). The accuracy of risk scores in predicting ovarian malignancy: a systematic review. *Obstet. Gynecol.* 113, 384–394. doi: 10.1097/AOG.0b013e318195ad17

Giannopoulou, L., Chebouti, I., Pavlakis, K., Kasimir-Bauer, S., and Lianidou, E. S. (2017). RASSF1A promoter methylation in high-grade serous ovarian cancer: a direct comparison study in primary tumors, adjacent morphologically tumor cell-free tissues and paired circulating tumor DNA. *Oncotarget* 8:21429. doi: 10.18632/oncotarget.15249

Giannopoulou, L., Mastoraki, S., Buderath, P., Strati, A., Pavlakis, K., Kasimir-Bauer, S., et al. (2018). ESR1 methylation in primary tumors and paired circulating tumor DNA of patients with high-grade serous ovarian cancer. *Gynecol. Oncol.* 150, 355–360. doi: 10.1016/j.ygyno.2018.05.026

Gifford, G. (2004). The Acquisition of hMLH1 Methylation in Plasma DNA after chemotherapy predicts poor survival for ovarian cancer patients. *Clin. Cancer Res.* 10, 4420–4426. doi: 10.1158/1078-0432.CCR-03-0732

Gloss, B. S., Patterson, K. I., Barton, C. A., Gonzalez, M., Scurry, J. P., Hacker, N. F., et al. (2012). Integrative genome-wide expression and promoter DNA methylation profiling identifies a potential novel panel of ovarian cancer epigenetic biomarkers. *Cancer Lett.* 318, 76–85. doi: 10.1016/j.canlet.2011.12.003

Gong, L., Wang, C., Gao, Y., and Wang, J. (2016). Decreased expression of microRNA-148a predicts poor prognosis in ovarian cancer and associates with tumor growth and metastasis. *Biomed. Pharmacother.* 83, 58–63. doi: 10.1016/j.biopha.2016.05.049

Gorelik, E. (2005). Multiplexed immunobead-based cytokine profiling for early detection of ovarian cancer. *Cancer Epidemiol. Biomarkers Prev.* 14, 981–987. doi: 10.1158/1055-9965.EPI-04-0404

Gormally, E., Caboux, E., Vineis, P., and Hainaut, P. (2007). Circulating free DNA in plasma or serum as biomarker of carcinogenesis: practical aspects and biological significance. *Mutat. Res. Mutat. Res.* 635, 105–117. doi: 10.1016/j.mrrev.2006.11.002

Gozzi, G., Chelbi, S. T., Manni, P., Alberti, L., Fonda, S., Saponaro, S., et al. (2016). Promoter methylation and downregulated expression of the TBX15 gene in ovarian carcinoma. *Oncol. Lett.* 12, 2811–2819. doi: 10.3892/ol.2016.5019

Günel, T., Gumusoglu, E., Dogan, B., Ertem, F. B., Hosseini, M. K., Cevik, N., et al. (2018). Potential biomarker of circulating hsa-miR-1273g-3p level for detection of recurrent epithelial ovarian cancer. *Arch. Gynecol. Obstet.* 298, 1173–1180. doi: 10.1007/s00404-018-4913-3

Guo, T. Y., Xu, H.-Y., Chen, W. J., Wu, M.-X., and Dai, X. (2018). Downregulation of miR-1294 associates with prognosis and tumor progression in epithelial ovarian cancer. *Eur. Rev. Med. Pharmacol. Sci.* 22, 7646–7652.

Guo, W., Zhu, L., Yu, M., Zhu, R., Chen, Q., and Wang, Q. (2018). A five-DNA methylation signature act as a novel prognostic biomarker in patients with ovarian serous cystadenocarcinoma. *Clin. Epigenetics* 10:142. doi: 10.1186/s13148-018-0574-0

Gupta, A., Godwin, A. K., Vanderveer, L., Lu, A., and Liu, J. (2003). Hypomethylation of the synuclein gamma gene CpG island promotes its aberrant expression in breast carcinoma and ovarian carcinoma. *Cancer Res.* 63, 664–673.

Häfner, N., Diebolder, H., Jansen, L., Hoppe, I., Dürst, M., and Runnebaum, I. B. (2011). Hypermethylated DAPK in serum DNA of women with uterine leiomyoma is a biomarker not restricted to cancer. *Gynecol. Oncol.* 121, 224–229. doi: 10.1016/j.ygyno.2010.11.018

Häfner, N., Steinbach, D., Jansen, L., Diebolder, H., Dürst, M., and Runnebaum, I. B. (2016). RUNX3 and CAMK2N1 hypermethylation as prognostic marker for epithelial ovarian cancer: prognostic methylation markers for EOC. *Int. J. Cancer* 138, 217–228. doi: 10.1002/ijc.29690

Han, X., Zhou, Y., You, Y., Lu, J., Wang, L., Hou, H., et al. (2017). TET1 promotes cisplatin-resistance via demethylating the vimentin promoter in ovarian cancer: TET1 in ovarian cancer. *Cell Biol. Int.* 41, 405–414. doi: 10.1002/cbin.10734

Havrilesky, L. J., Whitehead, C. M., Rubatt, J. M., Cheek, R. L., Groelke, J., He, Q., et al. (2008). Evaluation of biomarker panels for early stage ovarian cancer detection and monitoring for disease recurrence. *Gynecol. Oncol.* 110, 374–382. doi: 10.1016/j.ygyno.2008.04.041

Hendrich, B., and Bird, A. (2000). Mammalian methyltransferases and methyl-CpG-binding domains: proteins involved in DNA methylation. *Curr. Top. Microbiol. Immunol.* 249, 55–74. doi: 10.1007/978-3-642-59696-4_4

Herman, J. G., and Baylin, S. B. (2003). Gene silencing in cancer in association with promoter hypermethylation. *N. Engl. J. Med.* 349, 2042–2054. doi: 10.1056/NEJMra023075

Honda, H., Pazin, M. J., D'Souza, T., Ji, H., and Morin, P. J. (2007). Regulation of the CLDN3 gene in ovarian cancer cells. *Cancer Biol. Ther.* 6, 1733–1742. doi: 10.4161/cbt.6.11.4832

Honda, H., Pazin, M. J., Ji, H., Wernyj, R. P., and Morin, P. J. (2006). Crucial roles of Sp1 and epigenetic modifications in the regulation of the CLDN4 promoter in ovarian cancer cells. *J. Biol. Chem.* 281, 21433–21444. doi: 10.1074/jbc.M603767200

Hooda, J., Novak, M., Salomon, M. P., Matsuba, C., Ramos, R. I., MacDuffie, E., et al. (2019). Early loss of histone H2B monoubiquitylation alters chromatin accessibility and activates key immune pathways that facilitate progression of ovarian cancer. *Cancer Res.* 79, 760–772. doi: 10.1158/0008-5472.CAN-18-2297

Horak, P., Pils, D., Haller, G., Pribill, I., Roessler, M., Tomek, S., et al. (2005). Contribution of epigenetic silencing of tumor necrosis factor-related apoptosis inducing ligand receptor 1 (DR4) to TRAIL resistance and ovarian cancer. *Mol. Cancer Res.* 3, 335–343. doi: 10.1158/1541-7786.MCR-04-0136

Hough, C. D., Sherman-Baust, C. A., Pizer, E. S., Montz, F. J., Im, D. D., Rosenshein, N. B., et al. (2000). Large-scale serial analysis of gene expression reveals genes differentially expressed in ovarian cancer. *Cancer Res.* 60, 6281–6287.

Howe, H. L., Wu, X., Ries, L. A. G., Cokkinides, V., Ahmed, F., Jemal, A., et al. (2006). Annual report to the nation on the status of cancer, 1975–2003, featuring cancer among U.S. Hispanic/Latino populations. *Cancer* 107, 1711–1742. doi: 10.1002/cncr.22193

Hu, X., Macdonald, D. M., Huettner, P. C., Feng, Z., El Naqa, I. M., Schwarz, J. K., et al. (2009). A miR-200 microRNA cluster as prognostic marker in advanced ovarian cancer. *Gynecol. Oncol.* 114, 457–464. doi: 10.1016/j.ygyno.2009.05.022

Hu, Y., Zhang, Q., Cui, J., Liao, Z.-J., Jiao, M., Zhang, Y.-B., et al. (2019). Oncogene miR-934 promotes ovarian cancer cell proliferation and inhibits cell apoptosis through targeting BRMS1L. *Eur. Rev. Med. Pharmacol. Sci.* 23, 5595–5602. doi: 10.26355/eurrev_201907_18293

Huang, C.-Y., Cheng, W.-F., Lee, C.-N., Su, Y.-N., Chien, S.-C., Tzeng, Y.-L., et al. (2006). Serum mesothelin in epithelial ovarian carcinoma: a new screening marker and prognostic factor. *Anticancer Res.* 26, 4721–4728.

Huang, L., Zhou, Y., Cao, X.-P., Lin, J.-X., Zhang, L., Huang, S.-T., et al. (2017). KPNA2 is a potential diagnostic serum biomarker for epithelial ovarian cancer and correlates with poor prognosis. *Tumor Biol.* 39:101042831770628. doi: 10.1177/1010428317706289

Huang, R.-L., Chen, H.-J., Chen, L.-Y., Chao, T.-K., Lin, W.-Y., Liew, P.-L., et al. (2018). Epigenetic loss of heparan sulfate 3-O-sulfation sensitizes ovarian carcinoma to oncogenic signals and predicts prognosis: less 3-O-sulfated HS sensitizes prognostic signals. *Int. J. Cancer* 143, 1943–1953. doi: 10.1002/ijc.31580

Huang, R.-L., Gu, F., Kirma, N. B., Ruan, J., Chen, C.-L., Wang, H.-C., et al. (2013). Comprehensive methylome analysis of ovarian tumors reveals hedgehog signaling pathway regulators as prognostic DNA methylation biomarkers. *Epigenetics* 8, 624–634. doi: 10.4161/epi.24816

Huang, Y.-W., Jansen, R. A., Fabbri, E., Potter, D., Liyanarachchi, S., Chan, M. W. Y., et al. (2009). Identification of candidate epigenetic biomarkers for ovarian cancer detection. *Oncol. Rep.* 22, 853–861. doi: 10.3892/or_00000509

Ibanez de Caceres, I., Battagli, C., Esteller, M., Herman, J. G., Dulaimi, E., Edelson, M. I., et al. (2004). Tumor cell-specific BRCA1 and RASSF1A hypermethylation in serum, plasma, and peritoneal fluid from ovarian cancer patients. *Cancer Res.* 64, 6476–6481. doi: 10.1158/0008-5472.CAN-04-1529

Ignatov, T., Eggemann, H., Costa, S. D., Roessner, A., Kalinski, T., and Ignatov, A. (2014). BRCA1 promoter methylation is a marker of better response to platinum–taxane-based therapy in sporadic epithelial ovarian cancer. *J. Cancer Res. Clin. Oncol.* 140, 1457–1463. doi: 10.1007/s00432-014-1704-5

Iorio, M. V., Visone, R., Di Leva, G., Donati, V., Petrocca, F., Casalini, P., et al. (2007). MicroRNA signatures in human ovarian cancer. *Cancer Res.* 67, 8699–8707. doi: 10.1158/0008-5472.CAN-07-1936

Izutsu, N., Maesawa, C., Shibazaki, M., Oikawa, H., Shoji, T., Sugiyama, T., et al. (2008). Epigenetic modification is involved in aberrant expression of class III beta-tubulin, TUBB3, in ovarian cancer cells. *Int. J. Oncol.* 32, 1227–1235. doi: 10.3892/ijo_32_6_1227

Jacob, F., Hitchins, M. P., Fedier, A., Brennan, K., Nixdorf, S., Hacker, N. F., et al. (2014). Expression of GBGT1 is epigenetically regulated by DNA methylation in ovarian cancer cells. *BMC Mol. Biol.* 15:24. doi: 10.1186/1471-2199-15-24

Jacobs, I., and Bast, R. C. (1989). The CA 125 tumour-associated antigen: a review of the literature. *Hum. Reprod. Oxf. Engl.* 4, 1–12. doi: 10.1093/oxfordjournals.humrep.a136832

Jacobs, I., Gentry-Maharaj, A., Burnell, M., Manchanda, R., Singh, N., Sharma, A., et al. (2011). Sensitivity of transvaginal ultrasound screening for endometrial cancer in postmenopausal women: a case-control study within the UKCTOCS cohort. *Lancet Oncol.* 12, 38–48. doi: 10.1016/S1470-2045(10)70268-0

Jacobs, I., Oram, D., Fairbanks, J., Turner, J., Frost, C., and Grudzinskas, J. G. (1990). A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. *Br. J. Obstet. Gynaecol.* 97, 922–929. doi: 10.1111/j.1471-0528.1990.tb02448.x

Jacobs, I. J., and Menon, U. (2004). Progress and challenges in screening for early detection of ovarian cancer. *Mol. Cell. Proteomics* 3, 355–366. doi: 10.1074/mcp.R400006-MCP200

Jayson, G. C., Kohn, E. C., Kitchener, H. C., and Ledermann, J. A. (2014). Ovarian cancer. *Lancet* 384, 1376–1388. doi: 10.1016/S0140-6736(13)62146-7

Jia, Y., Lin, R., Jin, H., Si, L., Jian, W., Yu, Q., et al. (2019). MicroRNA-34 suppresses proliferation of human ovarian cancer cells by triggering autophagy and apoptosis and inhibits cell invasion by targeting Notch 1. *Biochimie* 160, 193–199. doi: 10.1016/j.biochi.2019.03.011

Jones, A., Lechner, M., Fourkala, E.-O., Kristeleit, R., and Widschwendter, M. (2010). Emerging promise of epigenetics and DNA methylation for the diagnosis and management of women's cancers. *Epigenomics* 2, 9–38. doi: 10.2217/epi.09.47

Jones, P. A. (2001). The role of DNA methylation in mammalian epigenetics. *Science* 293, 1068–1070. doi: 10.1126/science.1063852

Jones, P. A., and Baylin, S. B. (2002). The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* 3:415. doi: 10.1038/nrg816

Jones, P. A., and Baylin, S. B. (2007). The epigenomics of cancer. *Cell* 128, 683–692. doi: 10.1016/j.cell.2007.01.029

Jones, P. A., and Laird, P. W. (1999). Cancer-epigenetics comes of age. *Nat. Genet.* 21, 163–167. doi: 10.1038/5947

Kacem, S., and Feil, R. (2009). Chromatin mechanisms in genomic imprinting. *Mamm. Genome* 20, 544–556. doi: 10.1007/s00335-009-9223-4

Kamikihara, T., Arima, T., Kato, K., Matsuda, T., Kato, H., Douchi, T., et al. (2005). Epigenetic silencing of the imprinted geneZAC by DNA methylation is an early event in the progression of human ovarian cancer. *Int. J. Cancer* 115, 690–700. doi: 10.1002/ijc.20971

Kaneuchi, M., Sasaki, M., Tanaka, Y., Shiina, H., Verma, M., Ebina, Y., et al. (2004). Expression and methylation status of 14-3-3 sigma gene can characterize the different histological features of ovarian cancer. *Biochem. Biophys. Res. Commun.* 316, 1156–1162. doi: 10.1016/j.bbrc.2004.02.171

Kaneuchi, M., Sasaki, M., Tanaka, Y., Shiina, H., Yamada, H., Yamamoto, R., et al. (2005). WT1 and WT1-AS genes are inactivated by promoter methylation in ovarian clear cell adenocarcinoma. *Cancer* 104, 1924–1930. doi: 10.1002/cncr.21397

Kang, S., Dong, S. M., and Park, N.-H. (2010). Frequent promoter hypermethylation of TGFBI in epithelial ovarian cancer. *Gynecol. Oncol.* 118, 58–63. doi: 10.1016/j.ygyno.2010.03.025

Kardum, V., Karin, V., Glibo, M., Skrtic, A., Martic, T. N., Ibisevic, N., et al. (2017). Methylation-associated silencing of SFRP1 gene in high-grade serous ovarian carcinomas. *Ann. Diagn. Pathol.* 31, 45–49. doi: 10.1016/j.anndiagpath.2017.07.002

Katsaros, D., Cho, W., Singal, R., Fracchioli, S., Rigault De La Longrais, I. A., Arisio, R., et al. (2004). Methylation of tumor suppressor gene p16 and prognosis of epithelial ovarian cancer. *Gynecol. Oncol.* 94, 685–692. doi: 10.1016/j.ygyno.2004.06.018

Kaur, M., Singh, A., Singh, K., Gupta, S., and Sachan, M. (2016). Development of a multiplex MethyLight assay for the detection of DAPK1 and SOX1 methylation in epithelial ovarian cancer in a north Indian population. *Genes Genet. Syst.* 91, 175–181. doi: 10.1266/ggs.15-00051

Keita, M., Wang, Z.-Q., Pelletier, J.-F., Bachvarova, M., Plante, M., Gregoire, J., et al. (2013). Global methylation profiling in serous ovarian cancer is indicative for distinct aberrant DNA methylation signatures associated with tumor aggressiveness and disease progression. *Gynecol. Oncol.* 128, 356–363. doi: 10.1016/j.ygyno.2012.11.036

Kikuchi, R., Tsuda, H., Kanai, Y., Kasamatsu, T., Sengoku, K., Hirohashi, S., et al. (2007). Promoter hypermethylation contributes to frequent inactivation of a putative conditional tumor suppressor gene connective tissue growth factor in ovarian cancer. *Cancer Res.* 67, 7095–7105. doi: 10.1158/0008-5472.CAN-06-4567

Kikuchi, R., Tsuda, H., Kozaki, K., Kanai, Y., Kasamatsu, T., Sengoku, K., et al. (2008). Frequent inactivation of a putative tumor suppressor, angiopoietin-like protein 2, in ovarian cancer. *Cancer Res.* 68, 5067–5075. doi: 10.1158/0008-5472.CAN-08-0062

Kim, J.-H., Skates, S. J., Uede, T., Wong, K., Schorge, J. O., Feltmate, C. M., et al. (2002). Osteopontin as a potential diagnostic biomarker for ovarian cancer. *JAMA* 287, 1671–1679. doi: 10.1001/jama.287.13.1671

Kim, S., Choi, M. C., Jeong, J.-Y., Hwang, S., Jung, S. G., Joo, W. D., et al. (2019). Serum exosomal miRNA-145 and miRNA-200c as promising biomarkers for preoperative diagnosis of ovarian carcinomas. *J. Cancer* 10, 1958–1967. doi: 10.7150/jca.30231

Kim, Y.-W., Bae, S. M., Lim, H., Kim, Y. J., and Ahn, W. S. (2012). Development of multiplexed bead-based immunoassays for the detection of early stage ovarian cancer using a combination of serum biomarkers. *PLoS ONE* 7:e44960. doi: 10.1371/journal.pone.0044960

Kitade, S., Onoyama, I., Kobayashi, H., Yagi, H., Yoshida, S., Kato, M., et al. (2016). FBXW7 is involved in the acquisition of the malignant phenotype

in epithelial ovarian tumors. *Cancer Sci.* 107, 1399–1405. doi: 10.1111/cas.13026

Klotz, D. M., and Wimberger, P. (2017). Cells of origin of ovarian cancer: ovarian surface epithelium or fallopian tube? *Arch. Gynecol. Obstet.* 296, 1055–1062. doi: 10.1007/s00404-017-4529-z

Kneip, C., Schmidt, B., Seegebarth, A., Weickmann, S., Fleischhacker, M., Liebenberg, V., et al. (2011). SHOX2 DNA methylation is a biomarker for the diagnosis of lung cancer in plasma. *J. Thorac. Oncol.* 6, 1632–1638. doi: 10.1097/JTO.0b013e318220ef9a

Kobayashi, M., Sawada, K., Nakamura, K., Yoshimura, A., Miyamoto, M., Shimizu, A., et al. (2018). Exosomal miR-1290 is a potential biomarker of high-grade serous ovarian carcinoma and can discriminate patients from those with malignancies of other histological types. *J. Ovarian Res.* 11:81. doi: 10.1186/s13048-018-0458-0

Koestler, D. C., Chalise, P., Cicek, M. S., Cunningham, J. M., Armasu, S., Larson, M. C., et al. (2014). Integrative genomic analysis identifies epigenetic marks that mediate genetic risk for epithelial ovarian cancer. *BMC Med. Genomics* 7:260. doi: 10.1158/1538-7445.AM2014-260

Kohler, R. S., Anugraham, M., López, M. N., Xiao, C., Schoetzau, A., Hettich, T., et al. (2016). Epigenetic activation of *MGAT3* and corresponding bisecting GlcNAc shortens the survival of cancer patients. *Oncotarget* 7:51674–51686. doi: 10.18632/oncotarget.10543

Kolomietz, E., Meyn, M. S., Pandita, A., and Squire, J. A. (2002). The role of Alu repeat clusters as mediators of recurrent chromosomal aberrations in tumors. *Genes. Chromosomes Cancer* 35, 97–112. doi: 10.1002/gcc.10111

Kurose, K., Zhou, X.-P., Araki, T., Cannistra, S. A., Maher, E. R., and Eng, C. (2001). Frequent loss of PTEN Expression Is Linked to Elevated Phosphorylated Akt levels, but not associated with p27 and Cyclin D1 expression, in primary epithelial ovarian carcinomas. *Am. J. Pathol.* 158, 2097–2106. doi: 10.1016/S0002-9440(10)64681-0

Kwon, M. J., Kim, S.-S., Choi, Y.-L., Jung, H. S., Balch, C., Kim, S.-H., et al. (2010). Derepression of CLDN3 and CLDN4 during ovarian tumorigenesis is associated with loss of repressive histone modifications. *Carcinogenesis* 31, 974–983. doi: 10.1093/carcin/bgp336

Kwong, J., Lee, J.-Y., Wong, K.-K., Zhou, X., Wong, D. T. W., Lo, K.-W., et al. (2006). Candidate tumor-suppressor gene DLEC1 is frequently downregulated by promoter hypermethylation and histone hypoacetylation in human epithelial ovarian cancer. *Neoplasia* 8, 268–278. doi: 10.1593/neo.05502

Laios, A., O'Toole, S., Flavin, R., Martin, C., Kelly, L., Ring, M., et al. (2008). Potential role of miR-9 and miR-223 in recurrent ovarian cancer. *Mol. Cancer* 7:35. doi: 10.1186/1476-4598-7-35

Langhe, R., Norris, L., Saadeh, F. A., Blackshields, G., Varley, R., Harrison, A., et al. (2015). A novel serum microRNA panel to discriminate benign from malignant ovarian disease. *Cancer Lett.* 356, 628–636. doi: 10.1016/j.canlet.2014.10.010

Lee, C.-H., Subramanian, S., Beck, A. H., Espinosa, I., Senz, J., Zhu, S. X., et al. (2009). MicroRNA profiling of BRCA1/2 mutation-carrying and non-mutation-carrying high-grade serous carcinomas of ovary. *PLoS ONE* 4:e7314. doi: 10.1371/journal.pone.0007314

Lee, P. S., Teaberry, V. S., Bland, A. E., Huang, Z., Whitaker, R. S., Baba, T., et al. (2010). Elevated MAL expression is accompanied by promoter hypomethylation and platinum resistance in epithelial ovarian cancer. *Int. J. Cancer* 126, 1378–1389. doi: 10.1002/ijc.24797

Leon, S. A., Shapiro, B., Sklaroff, D. M., and Yaros, M. J. (1977). Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Res.* 37, 646–650.

Leskela, S., Leandro-Garcia, L. J., Mendiola, M., Barriuso, J., Inglada-Perez, L., Munoz, I., et al. (2010). The miR-200 family controls -tubulin III expression and is associated with paclitaxel-based treatment response and progression-free survival in ovarian cancer patients. *Endocr. Relat. Cancer* 18, 85–95. doi: 10.1677/ERC-10-0148

Leung, F., Bernardini, M. Q., Brown, M. D., Zheng, Y., Molina, R., Bast, R. C., et al. (2016). Validation of a novel biomarker panel for the detection of ovarian cancer. *Cancer Epidemiol. Biomarkers Prev.* 25, 1333–1340. doi: 10.1158/1055-9965.EPI-15-1299

Lewis, J. M., Heineck, D. P., and Heller, M. J. (2015). Detecting cancer biomarkers in blood: challenges for new molecular diagnostic and point-of-care tests using cell-free nucleic acids. *Expert Rev. Mol. Diagn.* 15, 1187–1200. doi: 10.1586/14737159.2015.1069709

Li, J., He, X., Dong, R., Wang, Y., Yu, J., and Qiu, H. (2015). Frequent loss of NISCH promotes tumor proliferation and invasion in ovarian cancer via inhibiting the FAK signal pathway. *Mol. Cancer Ther.* 14, 1202–1212. doi: 10.1158/1535-7163.MCT-14-0911

Li, J., Shao, W., and Feng, H. (2019). MiR-542-3p, a microRNA targeting CDK14, suppresses cell proliferation, invasiveness, and tumorigenesis of epithelial ovarian cancer. *Biomed. Pharmacother.* 110, 850–856. doi: 10.1016/j.biopha.2018.11.104

Li, L.-W., Xiao, H.-Q., Ma, R., Yang, M., Li, W., and Lou, G. (2017). miR-152 is involved in the proliferation and metastasis of ovarian cancer through repression of ERBB3. *Int. J. Mol. Med.* 41, 1529–1535. doi: 10.3892/ijmm.2017.3324

Li, X., Pan, Q., Wan, X., Mao, Y., Lu, W., Xie, X., et al. (2015). Methylation-associated Has-miR-9 deregulation in paclitaxel- resistant epithelial ovarian carcinoma. *BMC Cancer* 15:1. doi: 10.1186/s12885-015-1509-1

Liao, Y.-P., Chen, L.-Y., Huang, R.-L., Su, P.-H., Chan, M. W. Y., Chang, C.-C., et al. (2014). Hypomethylation signature of tumor-initiating cells predicts poor prognosis of ovarian cancer patients. *Hum. Mol. Genet.* 23, 1894–1906. doi: 10.1093/hmg/ddt583

Liggett, T. E., Melnikov, A., Yi, Q., Replogle, C., Hu, W., Rotmensch, J., et al. (2011). Distinctive DNA methylation patterns of cell-free plasma DNA in women with malignant ovarian tumors. *Gynecol. Oncol.* 120, 113–120. doi: 10.1016/j.ygyno.2010.09.019

Lin, B., White, J. T., Wu, J., Lele, S., Old, L. J., Hood, L., et al. (2009). Deep depletion of abundant serum proteins reveals low-abundant proteins as potential biomarkers for human ovarian cancer. *Proteomics Clin. Appl.* 3, 853–861. doi: 10.1002/prca.200800141

Lin, H.-W., Fu, C.-F., Chang, M.-C., Lu, T.-P., Lin, H.-P., Chiang, Y.-C., et al. (2018). *CDH1*, *DLEC1* and *SFRP5* methylation panel as a prognostic marker for advanced epithelial ovarian cancer. *Epigenomics* 10, 1397–1413. doi: 10.2217/epi-2018-0035

Litkouhi, B., Kwong, J., Lo, C.-M., Smedley, J. G., McClane, B. A., Aponte, M., et al. (2007). Claudin-4 overexpression in epithelial ovarian cancer is associated with hypomethylation and is a potential target for modulation of tight junction barrier function using a C-terminal fragment of *Clostridium perfringens* enterotoxin. *Neoplasia* 9, 304–314. doi: 10.1593/neo.07118

Liu, B., Zhang, J., and Yang, D. (2019). miR-96-5p promotes the proliferation and migration of ovarian cancer cells by suppressing Caveolae1. *J. Ovarian Res.* 12:57. doi: 10.1186/s13048-019-0533-1

Liu, D., Zhang, X.-X., Li, M.-C., Cao, C.-H., Wan, D.-Y., Xi, B.-X., et al. (2018). C/EBPβ enhances platinum resistance of ovarian cancer cells by reprogramming H3K79 methylation. *Nat. Commun.* 9:1739. doi: 10.1038/s41467-018-03590-5

Liu, G., Yang, D., Rupaimoole, R., Pecot, C. V., Sun, Y., Mangala, L. S., et al. (2015). Augmentation of response to chemotherapy by microRNA-506 through regulation of RAD51 in serous ovarian cancers. *J. Natl. Cancer Inst.* 107:djv108. doi: 10.1093/jnci/djv108

Liu, M. X., Siu, M. K. Y., Liu, S. S., Yam, J. W. P., Ngan, H. Y. S., and Chan, D. W. (2014). Epigenetic silencing of microRNA-199b-5p is associated with acquired chemoresistance via activation of JAG1-Notch1 signaling in ovarian cancer. *Oncotarget* 5, 944–958. doi: 10.18632/oncotarget.1458

Lofton-Day, C., Model, F., Devos, T., Tetzner, R., Distler, J., Schuster, M., et al. (2008). DNA methylation biomarkers for blood-based colorectal cancer screening. *Clin. Chem.* 54, 414–423. doi: 10.1373/clinchem.2007.095992

Loginov, V. I., Pronina, I. V., Burdennyy, A. M., Filippova, E. A., Kazubskaya, T. P., Kushlinsky, D. N., et al. (2018). Novel miRNA genes deregulated by aberrant methylation in ovarian carcinoma are involved in metastasis. *Gene* 662, 28–36. doi: 10.1016/j.gene.2018.04.005

Lopez-Serra, L., and Esteller, M. (2008). Proteins that bind methylated DNA and human cancer: reading the wrong words. *Br. J. Cancer* 98, 1881–1885. doi: 10.1038/sj.bjc.6604374

Lu, K. H., Patterson, A. P., Wang, L., Marquez, R. T., Atkinson, E. N., Baggerly, K. A., et al. (2004). Selection of potential markers for epithelial ovarian cancer with gene expression arrays and recursive descent partition analysis. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 10, 3291–3300. doi: 10.1158/1078-0432.CCR-03-0409

Lu, L., Katsaros, D., Rigault de la Longrais, I. A., Sochirca, O., and Yu, H. (2007). Hypermethylation of let-7a-3 in Epithelial Ovarian Cancer Is Associated with

Low Insulin-like Growth Factor-II Expression and Favorable Prognosis. *Cancer Res.* 67, 10117–10122. doi: 10.1158/0008-5472.CAN-07-2544

Lu, T.-Y., Kao, C.-F., Lin, C.-T., Huang, D.-Y., Chiu, C.-Y., Huang, Y.-S., et al. (2009). DNA methylation and histone modification regulate silencing of OPG during tumor progression. *J. Cell. Biochem.* 108, 315–325. doi: 10.1002/jcb.22256

Lund, R. J., Huhtinen, K., Salmi, J., Rantala, J., Nguyen, E. V., Moulder, R., et al. (2017). DNA methylation and transcriptome changes associated with cisplatin resistance in ovarian cancer. *Sci. Rep.* 7:1469. doi: 10.1038/s41598-017-01624-4

Luo, G., Chen, C., Wang, J., Yue, H., Tian, Y., Yang, P., et al. (2019). FOXD3 may be a new cellular target biomarker as a hypermethylation gene in human ovarian cancer. *Cancer Cell Int.* 19:44. doi: 10.1186/s12935-019-0755-8

Lv, Y., Niu, Z., Guo, X., Yuan, F., and Liu, Y. (2018). Serum S100 calcium binding protein A4 (S100A4, metatasin) as a diagnostic and prognostic biomarker in epithelial ovarian cancer. *Br. J. Biomed. Sci.* 75, 88–91. doi: 10.1080/09674845.2017.1394052

Lycke, M., Kristjansdottir, B., and Sundfeldt, K. (2018). A multicenter clinical trial validating the performance of HE4, CA125, risk of ovarian malignancy algorithm and risk of malignancy index. *Gynecol. Oncol.* 151, 159–165. doi: 10.1016/j.ygyno.2018.08.025

Lyu, N., Wang, Y., Wang, J., Zhang, Z., and Kong, W. (2018). Study on early diagnosis of epithelial ovarian cancer by analysis of plasma septin-9 and clusterin level. *J. Cancer Res. Ther.* 14:444. doi: 10.4103/0973-1482.181178

Mählck, C. G., and Grankvist, K. (1994). Plasma prealbumin in women with epithelial ovarian carcinoma. *Gynecol. Obstet. Invest.* 37, 135–140. doi: 10.1159/000292542

Mahmoud, E. H., Fawzy, A., and, A., Elshimy, R. A. (2018). Serum MicroRNA-21 negatively relates to expression of programmed cell death-4 in patients with epithelial ovarian cancer. *Asian Pac. J. Cancer Prev.* 19, 33–38.

Makarla, P. B., Saboorian, M. H., Ashfaq, R., Toyooka, K. O., Toyooka, S., Minna, J. D., et al. (2005). Promoter hypermethylation profile of ovarian epithelial neoplasms. *Clin. Cancer Res.* 11, 5365–5369. doi: 10.1158/1078-0432.CCR-04-2455

Maldonado, L., Brait, M., Izumchenko, E., Begum, S., Chatterjee, A., Sen, T., et al. (2018). Integrated transcriptomic and epigenomic analysis of ovarian cancer reveals epigenetically silenced GULP1. *Cancer Lett.* 433, 242–251. doi: 10.1016/j.canlet.2018.06.030

Mase, S., Shinjo, K., Totani, H., Katsushima, K., Arakawa, A., Takahashi, S., et al. (2019). *ZNF671* DNA methylation as a molecular predictor for the early recurrence of serous ovarian cancer. *Cancer Sci.* 110, 1105–1116. doi: 10.1111/cas.13936

McGrath, S. E., Annels, N., Madhuri, T. K., Tailor, A., Butler-Manuel, S. A., Morgan, R., et al. (2018). Engrailed-2 (EN2) – a novel biomarker in epithelial ovarian cancer. *BMC Cancer* 18:943. doi: 10.1186/s12885-018-4816-5

McIntosh, M. W., Drescher, C., Karlan, B., Scholler, N., Urban, N., Hellstrom, K. E., et al. (2004). Combining CA 125 and SMR serum markers for diagnosis and early detection of ovarian carcinoma. *Gynecol. Oncol.* 95, 9–15. doi: 10.1016/j.ygyno.2004.07.039

Melnikov, A., Scholtens, D., Godwin, A., and Levenson, V. (2009). Differential methylation profile of ovarian cancer in tissues and plasma. *J. Mol. Diagn.* 11, 60–65. doi: 10.2353/jmoldx.2009.080072

Menendez, L., Benigno, B. B., and McDonald, J. F. (2004). L1 and HERV-W retrotransposons are hypomethylated in human ovarian carcinomas. *Mol. Cancer* 3:12. doi: 10.1186/1476-4598-3-12

Meng, C.-F., Dai, D.-Q., and Guo, K.-J. (2008). [Effects of 5-Aza-2′-deoxycytidine and trichostatin A on DNA methylation and expression of hMLH1 in ovariancancer cell line COC1/DDP]. *Ai Zheng* 27, 1251–1255.

Meng, X., Joosse, S. A., Müller, V., Trillsch, F., Milde-Langosch, K., Mahner, S., et al. (2015). Diagnostic and prognostic potential of serum miR-7, miR-16, miR-25, miR-93, miR-182, miR-376a and miR-429 in ovarian cancer patients. *Br. J. Cancer* 113, 1358–1366. doi: 10.1038/bjc.2015.340

Meng, X., Müller, V., Milde-Langosch, K., Trillsch, F., Pantel, K., and Schwarzenbach, H. (2016a). "Circulating Cell-Free miR-373, miR-200a, miR-200b and miR-200c in Patients with Epithelial Ovarian Cancer," in *Circulating Nucleic Acids in Serum and Plasma – CNAPS IX*, eds P. B. Gahan, M. Fleischhacker, and B. Schmidt (Cham: Springer International Publishing), 3–8. doi: 10.1007/978-3-319-42044-8_1

Meng, X., Müller, V., Milde-Langosch, K., Trillsch, F., Pantel, K., and Schwarzenbach, H. (2016b). Diagnostic and prognostic relevance of circulating exosomal miR-373, miR-200a, miR-200b and miR-200c in patients with epithelial ovarian cancer. *Oncotarget* 7. doi: 10.18632/oncotarget.7850

Menon, U., Gentry-Maharaj, A., Hallett, R., Ryan, A., Burnell, M., Sharma, A., et al. (2009). Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: results of the prevalence screen of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *Lancet Oncol.* 10, 327–340. doi: 10.1016/S1470-2045(09)70026-9

Menon, U., and Jacobs, I. J. (2000). Recent developments in ovarian cancer screening. *Curr. Opin. Obstet. Gynecol.* 12, 39–42. doi: 10.1097/00001703-200002000-00007

Merritt, W. M., Lin, Y. G., Han, L. Y., Kamat, A. A., Spannuth, W. A., Schmandt, R., et al. (2008). Dicer, Drosha, and outcomes in patients with ovarian cancer. *N. Engl. J. Med.* 359, 2641–2650. doi: 10.1056/NEJMoa0803785

Milde-Langosch, K., Ocon, E., Becker, G., and Löning, T. (1998). p16/MTS1 inactivation in ovarian carcinomas: high frequency of reduced protein expression associated with hyper-methylation or mutation in endometrioid and mucinous tumors. *Int. J. Cancer* 79, 61–65.

Mirza, A., McGuirk, M., Hockenberry, T. N., Wu, Q., Ashar, H., Black, S., et al. (2002). Human survivin is negatively regulated by wild-type p53 and participates in p53-dependent apoptotic pathway. *Oncogene* 21, 2613–2622. doi: 10.1038/sj.onc.1205353

Mok, S. C., Chao, J., Skates, S., Wong, K., Yiu, G. K., Muto, M. G., et al. (2001). Prostasin, a potential serum marker for ovarian cancer: identification through microarray technology. *J. Natl. Cancer Inst.* 93, 1458–1464. doi: 10.1093/jnci/93.19.1458

Molina, R., Escudero, J. M., Aug,é, J. M., Filella, X., Foj, L., Torn,é, A., et al. (2011). HE4 a novel tumour marker for ovarian cancer: comparison with CA 125 and ROMA algorithm in patients with gynaecological diseases. *Tumour Biol. J. Int. Soc. Oncodevelopmental Biol. Med.* 32, 1087–1095. doi: 10.1007/s13277-011-0204-3

Moore, R. G., Brown, A. K., Miller, M. C., Skates, S., Allard, W. J., Verch, T., et al. (2008). The use of multiple novel tumor biomarkers for the detection of ovarian carcinoma in patients with a pelvic mass. *Gynecol. Oncol.* 108, 402–408. doi: 10.1016/j.ygyno.2007.10.017

Moore, R. G., MacLaughlan, S., and Bast, R. C. (2010). Current state of biomarker development for clinical application in epithelial ovarian cancer. *Gynecol. Oncol.* 116, 240–245. doi: 10.1016/j.ygyno.2009.09.041

Moore, R. G., McMeekin, D. S., Brown, A. K., DiSilvestro, P., Miller, M. C., Allard, W. J., et al. (2009). A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. *Gynecol. Oncol.* 112, 40–46. doi: 10.1016/j.ygyno.2008.08.031

Moore, R. G., Miller, M. C., Disilvestro, P., Landrum, L. M., Gajewski, W., Ball, J. J., et al. (2011). Evaluation of the diagnostic accuracy of the risk of ovarian malignancy algorithm in women with a pelvic mass. *Obstet. Gynecol.* 118, 280–288. doi: 10.1097/AOG.0b013e318224fce2

Mor, G., Visintin, I., Lai, Y., Zhao, H., Schwartz, P., Rutherford, T., et al. (2005). Serum protein markers for early detection of ovarian cancer. *Proc. Natl. Acad. Sci.U.S.A.* 102, 7677–7682. doi: 10.1073/pnas.0502178102

Muller, H. M. (2004). Analysis of methylated genes in peritoneal fluids of ovarian cancer patients: a new prognostic tool. *Clin. Chem.* 50, 2171–2173. doi: 10.1373/clinchem.2004.034090

Munkarah, A., Chatterjee, M., and Tainsky, M. A. (2007). Update on ovarian cancer screening: *Curr. Opin. Obstet. Gynecol.* 19, 22–26. doi: 10.1097/GCO.0b013e328011ec99

Murphy, S. K., Huang, Z., Wen, Y., Spillman, M. A., Whitaker, R. S., Simel, L. R., et al. (2006). Frequent IGF2/H19 domain epigenetic alterations and elevated IGF2 expression in epithelial ovarian cancer. *Mol. Cancer Res.* 4, 283–292. doi: 10.1158/1541-7786.MCR-05-0138

Nguyen, L., Cardenas-Goicoechea, S. J., Gordon, P., Curtin, C., Momeni, M., Chuang, L., et al. (2013). Biomarkers for early detection of ovarian cancer. *Womens Health* 9, 171–187. doi: 10.2217/WHE.13.2

Nicholson, L. J., Smith, P. R., Hiller, L., Szlosarek, P. W., Kimberley, C., Sehouli, J., et al. (2009). Epigenetic silencing of argininosuccinate synthetase confers resistance to platinum-induced cell death but collateral sensitivity to arginine auxotrophy in ovarian cancer. *Int. J. Cancer* 125, 1454–1463. doi: 10.1002/ijc.24546

Nosov, V., Su, F., Amneus, M., Birrer, M., Robins, T., Kotlerman, J., et al. (2009). Validation of serum biomarkers for detection of early-stage ovarian cancer. *Am. J. Obstet. Gynecol.* 200, 639.e1–5. doi: 10.1016/j.ajog.2008.12.042

Ongenaert, M., Van Neste, L., De Meyer, T., Menschaert, G., Bekaert, S., and Van Criekinge, W. (2008). PubMeth: a cancer methylation database combining text-mining and expert annotation. *Nucl. Acids Res.* 36, D842–D846. doi: 10.1093/nar/gkm788

Pampalakis, G., Diamandis, E. P., Katsaros, D., and Sotiropoulou, G. (2010). Down-regulation of dicer expression in ovarian cancer tissues. *Clin. Biochem.* 43, 324–327. doi: 10.1016/j.clinbiochem.2009.09.014

Pattamadilok, J., Huapai, N., Rattanatanyong, P., Vasurattana, A., Triratanachat, S., Tresukosol, D., et al. (2008). LINE-1 hypomethylation level as a potential prognostic factor for epithelial ovarian cancer. *Int. J. Gynecol. Cancer* 18, 711–717. doi: 10.1111/j.1525-1438.2007.01117.x

Perlin, E., and Moquin, R. B. (1972). Serum DNA levels in patients with malignant disease. *Am. J. Clin. Pathol.* 58, 601–602. doi: 10.1093/ajcp/58.5.601

Petrocca, F., Iliopoulos, D., Qin, H. R., Nicoloso, M. S., Yendamuri, S., Wojcik, S. E., et al. (2006). Alterations of the tumor suppressor gene ARLTS1 in ovarian cancer. *Cancer Res.* 66, 10287–10291. doi: 10.1158/0008-5472.CAN-06-2289

Phelps, D. L., Borley, J. V., Flower, K. J., Dina, R., Darb-Esfahani, S., Braicu, I., et al. (2017). Methylation of MYLK3 gene promoter region: a biomarker to stratify surgical care in ovarian cancer in a multicentre study. *Br. J. Cancer* 116, 1287–1293. doi: 10.1038/bjc.2017.83

Pieretti, M., Cavalieri, C., Conway, P. S., Gallion, H. H., Powell, D. E., and Turker, M. S. (1995). Genetic alterations distinguish different types of ovarian tumors. *Int. J. Cancer* 64, 434–440. doi: 10.1002/ijc.2910640614

Polivka, J., Pesta, M., and Janku, F. (2015). Testing for oncogenic molecular aberrations in cell-free DNA-based liquid biopsies in the clinic: are we there yet? *Expert Rev. Mol. Diagn.* 15, 1631–1644. doi: 10.1586/14737159.2015.1110021

Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. *Nat. Biotechnol.* 28, 1057–1068. doi: 10.1038/nbt.1685

Press, J. Z., De Luca, A., Boyd, N., Young, S., Troussard, A., Ridge, Y., et al. (2008). Ovarian carcinomas with genetic and epigenetic BRCA1 loss have distinct molecular abnormalities. *BMC Cancer* 8:17. doi: 10.1186/1471-2407-8-17

Pulliam, N., Fang, F., Ozes, A. R., Tang, J., Adewuyi, A., Keer, H., et al. (2018). An effective epigenetic-PARP inhibitor combination therapy for breast and ovarian cancers independent of BRCA mutations. *Clin. Cancer Res.* 24, 3163–3175. doi: 10.1158/1078-0432.CCR-18-0204

Qu, G., Dubeau, L., Narayan, A., Yu, M. C., and Ehrlich, M. (1999). Satellite DNA hypomethylation vs. overall genomic hypomethylation in ovarian epithelial tumors of different malignant potential. *Mutat. Res.* 423, 91–101. doi: 10.1016/S0027-5107(98)00229-2

Ramus, S. J., Harrington, P. A., Pye, C., DiCioccio, R. A., Cox, M. J., Garlinghouse-Jones, K., et al. (2007). Contribution of *BRCA1* and *BRCA2* mutations to inherited ovarian cancer. *Hum. Mutat.* 28, 1207–1215. doi: 10.1002/humu.20599

Rathi, A., Virmani, A. K., Schorge, J. O., Elias, K. J., Maruyama, R., Minna, J. D., et al. (2002). Methylation profiles of sporadic ovarian tumors and nonmalignant ovaries from high-risk women. *Clin. Cancer Res.* 8, 3324–3331.

Rattanapan, Y., Korkiatsakul, V., Kongruang, A., Chareonsirisuthigul, T., Rerkamnuaychoke, B., Wongkularb, A., et al. (2018). EGFL7 and RASSF1 promoter hypermethylation in epithelial ovarian cancer. *Cancer Genet.* 224–225, 37–40. doi: 10.1016/j.cancergen.2018.04.117

Reik, W., and Lewis, A. (2005). Co-evolution of X-chromosome inactivation and imprinting in mammals. *Nat. Rev. Genet.* 6, 403–410. doi: 10.1038/nrg1602

Resnick, K. E., Alder, H., Hagan, J. P., Richardson, D. L., Croce, C. M., and Cohn, D. E. (2009). The detection of differentially expressed microRNAs from the serum of ovarian cancer patients using a novel real-time PCR platform. *Gynecol. Oncol.* 112, 55–59. doi: 10.1016/j.ygyno.2008.08.036

Richon, V. M., Sandhoff, T. W., Rifkind, R. A., and Marks, P. A. (2000). Histone deacetylase inhibitor selectively induces p21WAF1 expression and gene-associated histone acetylation. *Proc. Natl. Acad. Sci. U.S.A.* 97, 10014–10019. doi: 10.1073/pnas.180316197

Rose, S. L., Fitzgerald, M. P., White, N. O., Hitchler, M. J., Futscher, B. W., De Geest, K., et al. (2006). Epigenetic regulation of maspin expression in human ovarian carcinoma cells. *Gynecol. Oncol.* 102, 319–324. doi: 10.1016/j.ygyno.2005.12.025

Rosen, D. G., Wang, L., Atkinson, J. N., Yu, Y., Lu, K. H., Diamandis, E. P., et al. (2005). Potential markers that complement expression of CA125 in epithelial ovarian cancer. *Gynecol. Oncol.* 99, 267–277. doi: 10.1016/j.ygyno.2005.06.040

Rosenthal, A. N., Menon, U., and Jacobs, I. J. (2006). Screening for ovarian cancer. *Clin. Obstet. Gynecol.* 49, 433–447. doi: 10.1097/00003081-200609000-00004

Salta, S., P., Nunes, S., Fontes-Sousa, M., Lopes, P., Freitas, M., et al. (2018). A DNA methylation-based test for breast cancer detection in circulating cell-free DNA. *J. Clin. Med.* 7:420. doi: 10.3390/jcm7110420

Sarojini, S., Tamir, A., Lim, H., Li, S., Zhang, S., Goy, A., et al. (2012). Early detection biomarkers for ovarian cancer. *J. Oncol.* 2012, 1–15. doi: 10.1155/2012/709049

Sawada, K., Mitra, A. K., Radjabi, A. R., Bhaskar, V., Kistner, E. O., Tretiakova, M., et al. (2008). Loss of E-cadherin promotes ovarian cancer metastasis via alpha 5-integrin, which is a therapeutic target. *Cancer Res.* 68, 2329–2339. doi: 10.1158/0008-5472.CAN-07-5167

Schmid, G., Notaro, S., Reimer, D., Abdel-Azim, S., Duggan-Peer, M., Holly, J., et al. (2016). Expression and promotor hypermethylation of miR-34a in the various histological subtypes of ovarian cancer. *BMC Cancer* 16:102. doi: 10.1186/s12885-016-2135-2

Schweigert, F. J., and Sehouli, J. (2005). Transthyretin, a biomarker for nutritional status and ovarian cancer. *Cancer Res.* 65:1114; author reply 1114.

Sellar, G. C., Watt, K. P., Rabiasz, G. J., Stronach, E. A., Li, L., Miller, E. P., et al. (2003). OPCML at 11q25 is epigenetically inactivated and has tumor-suppressor function in epithelial ovarian cancer. *Nat. Genet.* 34, 337–343. doi: 10.1038/ng1183

Shah, C. A., Lowe, K. A., Paley, P., Wallace, E., Anderson, G. L., McIntosh, M. W., et al. (2009). Influence of ovarian cancer risk status on the diagnostic performance of the serum biomarkers mesothelin, HE4, and CA125. *Cancer Epidemiol. Biomarkers Prev.* 18, 1365–1372. doi: 10.1158/1055-9965.EPI-08-1034

Sharma, A., Albahrani, M., Zhang, W., Kufel, C. N., James, S. R., Odunsi, K., et al. (2019). Epigenetic activation of POTE genes in ovarian cancer. *Epigenetics* 14, 185–197. doi: 10.1080/15592294.2019.1581590

Sharma, S., Kelly, T. K., and Jones, P. A. (2010). Epigenetics in cancer. *Carcinogenesis* 31, 27–36. doi: 10.1093/carcin/bgp220

Shen, D.-H., Chan, K. Y.-K., Khoo, U.-S., Ngan, H. Y.-S., Xue, W.-C., Chiu, P.-M., et al. (2005). Epigenetic and genetic alterations of p33ING1b in ovarian cancer. *Carcinogenesis* 26, 855–863. doi: 10.1093/carcin/bgi011

Shibata, D., Mori, Y., Cai, K., Zhang, L., Yin, J., Elahi, A., et al. (2006). RAB32 hypermethylation and microsatellite instability in gastric and endometrial adenocarcinomas. *Int. J. Cancer* 119, 801–806. doi: 10.1002/ijc.21912

Shih, I.-M., and Kurman, R. J. (2004). Ovarian tumorigenesis: a proposed model based on morphological and molecular genetic analysis. *Am. J. Pathol.* 164, 1511–1518. doi: 10.1016/S0002-9440(10)63708-X

Shivapurkar, N., Toyooka, S., Toyooka, K. O., Reddy, J., Miyajima, K., Suzuki, M., et al. (2004). Aberrant methylation of trail decoy receptor genes is frequent in multiple tumor types. *Int. J. Cancer* 109, 786–792. doi: 10.1002/ijc.20041

Si, L., Jia, Y., Lin, R., Jian, W., Yu, Q., and Yang, S. (2019). 10.1016/j.biochi.2019.03.011. *Arch. Biochem. Biophys.* 668, 9–15. doi: 10.1016/j.abb.2019.04.009

Siegel, R., Ward, E., Brawley, O., and Jemal, A. (2011). Cancer statistics, 2011: The impact of eliminating socioeconomic and racial disparities on premature cancer deaths. *CA Cancer J. Clin.* 61, 212–236. doi: 10.3322/caac.20121

Siegel, R. L., Miller, K. D., and Jemal, A. (2015). Cancer Statistics, 2015. *CA Cancer J. Clin.* 65, 5–29. doi: 10.3322/caac.21254

Simmons, A. R., Clarke, C. H., Badgwell, D. B., Lu, Z., Sokoll, L. J., Lu, K. H., et al. (2016). Validation of a biomarker panel and longitudinal biomarker performance for early detection of ovarian cancer. *Int. J. Gynecol. Cancer* 26, 1070–1077. doi: 10.1097/IGC.0000000000000737

Simon, I., Katsaros, D., Rigault de la Longrais, I., Massobrio, M., Scorilas, A., Kim, N. W., et al. (2007). B7-H4 is over-expressed in early-stage ovarian cancer and is independent of CA125 expression. *Gynecol. Oncol.* 106, 334–341. doi: 10.1016/j.ygyno.2007.03.035

Simon, I., Zhuo, S., Corral, L., Diamandis, E. P., Sarno, M. J., Wolfert, R. L., et al. (2006). B7-h4 is a novel membrane-bound protein and a candidate serum and tissue biomarker for ovarian cancer. *Cancer Res.* 66, 1570–1575. doi: 10.1158/0008-5472.CAN-04-3550

Socha, M. J., Said, N., Dai, Y., Kwong, J., Ramalingam, P., Trieu, V., et al. (2009). Aberrant Promoter Methylation of Sparc in Ovarian Cancer. *Neoplasia* 11, 126–135. doi: 10.1593/neo.81146

Sorrentino, A., Liu, C.-G., Addario, A., Peschle, C., Scambia, G., and Ferlini, C. (2008). Role of microRNAs in drug-resistant ovarian cancer cells. *Gynecol. Oncol.* 111, 478–486. doi: 10.1016/j.ygyno.2008.08.017

Staub, J., Chien, J., Pan, Y., Qian, X., Narita, K., Aletti, G., et al. (2007). Epigenetic silencing of HSulf-1 in ovarian cancer:implications in chemoresistance. *Oncogene* 26, 4969–4978. doi: 10.1038/sj.onc.1210300

Strait, K. A., Dabbas, B., Hammond, E. H., Warnick, C. T., Iistrup, S. J., and Ford, C. D. (2002). Cell cycle blockade and differentiation of ovarian cancer cells by the histone deacetylase inhibitor trichostatin A are associated with changes in p21, Rb, and Id proteins. *Mol. Cancer Ther.* 1, 1181–1190.

Strathdee, G., Appleton, K., Illand, M., Millan, D. W., Sargent, J., Paul, J., et al. (2001). Primary ovarian carcinomas display multiple methylator phenotypes involving known tumor suppressor genes. *Am. J. Pathol.* 158, 1121–1127. doi: 10.1016/S0002-9440(10)64059-X

Strathdee, G., Davies, B. R., Vass, J. K., Siddiqui, N., and Brown, R. (2004). Cell type-specific methylation of an intronic CpG island controls expression of the MCJ gene. *Carcinogenesis* 25, 693–701. doi: 10.1093/carcin/bgh066

Strathdee, G., Vass, J. K., Oien, K. A., Siddiqui, N., Curto-Garcia, J., and Brown, R. (2005). Demethylation of the MCJ gene in stage III/IV epithelial ovarian cancer and response to chemotherapy. *Gynecol. Oncol.* 97, 898–903. doi: 10.1016/j.ygyno.2005.03.023

Su, F., Lang, J., Kumar, A., Ng, C., Hsieh, B., Suchard, M. A., et al. (2007). Validation of candidate serum ovarian cancer biomarkers for early detection. *Biomark. Insights* 2, 369–375. doi: 10.1177/117727190700200011

Su, H.-Y., Lai, H.-C., Lin, Y.-W., Chou, Y.-C., Liu, C.-Y., and Yu, M.-H. (2009). An epigenetic marker panel for screening and prognostic prediction of ovarian cancer. *Int. J. Cancer* 124, 387–393. doi: 10.1002/ijc.23957

Su, H.-Y., Lai, H.-C., Lin, Y.-W., Liu, C.-Y., Chen, C.-K., Chou, Y.-C., et al. (2010). Epigenetic silencing of SFRP5 is related to malignant phenotype and chemoresistance of ovarian cancer through Wnt signaling pathway. *Int. J. Cancer* 127, 555–567. doi: 10.1002/ijc.25083

Suehiro, Y., Okada, T., Okada, T., Anno, K., Okayama, N., Ueno, K., et al. (2008). Aneuploidy predicts outcome in patients with endometrial carcinoma and is related to lack of CDH13 hypermethylation. *Clin. Cancer Res.* 14, 3354–3361. doi: 10.1158/1078-0432.CCR-07-4609

Sun, N., Kohli, A., Huang, S., Chang, T., and Chao, C. C. -K. (2019). Androgen receptor transcriptional activity and chromatin modifications on the *ABCB1/MDR* gene are critical for taxol resistance in ovarian cancer cells. *J. Cell. Physiol.* 234, 8760–8775. doi: 10.1002/jcp.27535

Sung, C. K., Li, D., Andrews, E., Drapkin, R., and Benjamin, T. (2013). Promoter methylation of the SALL2 tumor suppressor gene in ovarian cancers. *Mol. Oncol.* 7, 419–427. doi: 10.1016/j.molonc.2012.11.005

Sung, H. Y., Choi, E. N., Lyu, D., Park, A. K., Ju, W., and Ahn, J.-H. (2014a). Aberrant hypomethylation-mediated AGR2 overexpression induces an aggressive phenotype in ovarian cancer cells. *Oncol. Rep.* 32, 815–820. doi: 10.3892/or.2014.3243

Sung, H. Y., Ju, W., and Ahn, J.-H. (2014b). DNA hypomethylation-mediated overexpression of carbonic anhydrase 9 induces an aggressive phenotype in ovarian cancer cells. *Yonsei Med. J.* 55:1656. doi: 10.3349/ymj.2014.55.6.1656

Sung, H. Y., Park, A. K., Ju, W., and Ahn, J.-H. (2014c). overexpression of mucin 13 due to promoter methylation promotes aggressive behavior in ovarian cancer cells. *Yonsei Med. J.* 55:1206. doi: 10.3349/ymj.2014.55.5.1206

Sung, H. Y., Yang, S.-D., Ju, W., and Ahn, J.-H. (2017a). Aberrant epigenetic regulation of GABRP associates with aggressive phenotype of ovarian cancer. *Exp. Mol. Med.* 49:e335. doi: 10.1038/emm.2017.62

Sung, H. Y., Yang, S.-D., Park, A. K., Ju, W., and Ahn, J.-H. (2017b). Aberrant hypomethylation of *solute carrier family 6 member 12* promoter induces metastasis of ovarian cancer. *Yonsei Med. J.* 58:27. doi: 10.3349/ymj.2017.58.1.27

Suzuki, F., Akahira, J.-I., Miura, I., Suzuki, T., Ito, K., Hayashi, S.-I., et al. (2008). Loss of estrogen receptor beta isoform expression and its correlation with aberrant DNA methylation of the 5'-untranslated region in human epithelial ovarian carcinoma. *Cancer Sci.* 99, 2365–2372. doi: 10.1111/j.1349-7006.2008.00988.x

Symer, D. E., Connelly, C., Szak, S. T., Caputo, E. M., Cost, G. J., Parmigiani, G., et al. (2002). Human l1 retrotransposition is associated with genetic instability *in vivo*. *Cell* 110, 327–338. doi: 10.1016/S0092-8674(02)00839-5

Tam, K. F., Liu, V. W. S., Liu, S. S., Tsang, P. C. K., Cheung, A. N. Y., Yip, A. M. W., et al. (2007). Methylation profile in benign, borderline and malignant ovarian tumors. *J. Cancer Res. Clin. Oncol.* 133, 331–341. doi: 10.1007/s00432-006-0178-5

Tang, G., Guo, J., Zhu, Y., Huang, Z., Liu, T., Cai, J., et al. (2018). Metformin inhibits ovarian cancer via decreasing H3K27 trimethylation. *Int. J. Oncol.* 52, 1899–1911. doi: 10.3892/ijo.2018.4343

Tang, Z., Qian, M., and Ho, M. (2013). The role of mesothelin in tumor progression and targeted therapy. *Anticancer Agents Med. Chem.* 13, 276–280. doi: 10.2174/1871520611313020014

Taniguchi, T., Tischkowitz, M., Ameziane, N., Hodgson, S. V., Mathew, C. G., Joenje, H., et al. (2003). Disruption of the Fanconi anemia-BRCA pathway in cisplatin-sensitive ovarian tumors. *Nat. Med.* 9, 568–574. doi: 10.1038/nm852

Tasaka, R., Fukuda, T., Shimomura, M., Inoue, Y., Wada, T., Kawanishi, M., et al. (2017). TBX2 expression is associated with platinum-sensitivity of ovarian serous carcinoma. *Oncol. Lett.* 15, 3085–3090. doi: 10.3892/ol.2017.7719

Taylor, D. D., and Gercel-Taylor, C. (2008). MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer. *Gynecol. Oncol.* 110, 13–21. doi: 10.1016/j.ygyno.2008.04.033

Teodoridis, J. M., Hall, J., Marsh, S., Kannall, H. D., Smyth, C., Curto, J., et al. (2005). CpG island methylation of DNA damage response genes in advanced ovarian cancer. *Cancer Res.* 65, 8961–8967. doi: 10.1158/0008-5472.CAN-05-1187

Terasawa, K., Sagae, S., Toyota, M., Tsukada, K., Ogi, K., Satoh, A., et al. (2004). Epigenetic inactivation of TMS1/ASC in ovarian cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 10, 2000–2006. doi: 10.1158/1078-0432.CCR-0932-03

Teschendorff, A. E., Gao, Y., Jones, A., Ruebner, M., Beckmann, M. W., Wachter, D. L., et al. (2016). DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. *Nat. Commun.* 7:10478. doi: 10.1038/ncomms10478

Teschendorff, A. E., Jones, A., Fiegl, H., Sargent, A., Zhuang, J. J., Kitchener, H. C., et al. (2012). Epigenetic variability in cells of normal cytology is associated with the risk of future morphological transformation. *Genome Med.* 4:24. doi: 10.1186/gm323

Teschendorff, A. E., Lee, S.-H., Jones, A., Fiegl, H., Kalwa, M., Wagner, W., et al. (2015). HOTAIR and its surrogate DNA methylation signature indicate carboplatin resistance in ovarian cancer. *Genome Med.* 7:108. doi: 10.1186/s13073-015-0233-4

Teschendorff, A. E., and Widschwendter, M. (2012). Differential variability improves the identification of cancer risk markers in DNA methylation studies profiling precursor cancer lesions. *Bioinformatics* 28, 1487–1494. doi: 10.1093/bioinformatics/bts170

Thomson, J. P., Skene, P. J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., et al. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* 464, 1082–1086. doi: 10.1038/nature08924

Tomar, T., Alkema, N. G., Schreuder, L., Meersma, G. J., de Meyer, T., van Criekinge, W., et al. (2017). Methylome analysis of extreme chemoresponsive patients identifies novel markers of platinum sensitivity in high-grade serous ovarian cancer. *BMC Med.* 15:116. doi: 10.1186/s12916-017-0870-0

Tomar, T., de Jong, S., Alkema, N. G., Hoekman, R. L., Meersma, G. J., Klip, H. G., et al. (2016). Genome-wide methylation profiling of ovarian cancer patient-derived xenografts treated with the demethylating agent decitabine identifies novel epigenetically regulated genes and pathways. *Genome Med.* 8:107. doi: 10.1186/s13073-016-0361-5

Torng, P.-L., Lin, C.-W., Chan, M. W., Yang, H.-W., Huang, S.-C., and Lin, C.-T. (2009). Promoter methylation of IGFBP-3 and p53 expression in ovarian endometrioid carcinoma. *Mol. Cancer* 8:120. doi: 10.1186/1476-4598-8-120

Toss, A., Tomasello, C., Razzaboni, E., Contu, G., Grandi, G., Cagnacci, A., et al. (2015). Hereditary ovarian cancer: not only *BRCA* 1 and 2 genes. *BioMed Res. Int.* 2015, 1–11. doi: 10.1155/2015/341723

U. S. Preventive Services Task Force (2010). *The Guide to Clinical Preventive Services 2010 - 2011: Recommendations of the U.S. Preventive Services Task Force*. Rockville, MD: Agency for Healthcare Research and Quality (US) Available online at: http://www.ncbi.nlm.nih.gov/books/NBK56707/ (accessed December 19, 2017).

Valls, E., Sánchez-Molina, S., and Martínez-Balbás, M. A. (2005). Role of histone modifications in marking and activating genes through mitosis. *J. Biol. Chem.* 280, 42592–42600. doi: 10.1074/jbc.M507407200

Vang, R., Shih, I.-M., and Kurman, R. J. (2009). Ovarian low-grade and high-grade serous carcinoma: pathogenesis, clinicopathologic and molecular biologic features, and diagnostic problems. *Adv. Anat. Pathol.* 16, 267–282. doi: 10.1097/PAP.0b013e3181b4fffa

Visintin, I., Feng, Z., Longton, G., Ward, D. C., Alvero, A. B., Lai, Y., et al. (2008). Diagnostic markers for early detection of ovarian cancer. *Clin. Cancer Res.* 14, 1065–1072. doi: 10.1158/1078-0432.CCR-07-1569

Vlassov, V. V., Laktionov, P. P., and Rykova, E. Y. (2010). Circulating nucleic acids as a potential source for cancer biomarkers. *Curr. Mol. Med.* 10, 142–165. doi: 10.2174/156652410790963295

Wang, B., Yu, L., Luo, X., Huang, L., Li, Q., Shao, X., et al. (2017). Detection of OPCML methylation, a possible epigenetic marker, from free serum circulating DNA to improve the diagnosis of early-stage ovarian epithelial cancer. *Oncol. Lett.* 14, 217–223. doi: 10.3892/ol.2017.6111

Wang, G., Li, X., Tian, W., Wang, Y., Wu, D., Sun, Z., et al. (2015). Promoter DNA methylation is associated with *KLF11* expression in epithelial ovarian cancer: promoter Dna Methylation Is Associated With *KLF11* Expression. *Genes. Chromosomes Cancer* 54, 453–462. doi: 10.1002/gcc.22257

Wang, H.-Q., Yang, C.-Y., Wang, S.-Y., Wang, T., Han, J.-L., Wei, K., et al. (2018). Cell-free plasma hypermethylated CASZ1, CDH13 and ING2 are promising biomarkers of esophageal cancer. *J. Biomed. Res.* 32, 424–433. doi: 10.7555/JBR.32.20170065

Wang, L., Zhu, M.-J., Ren, A.-M., Wu, H.-F., Han, W.-M., Tan, R.-Y., et al. (2014). A Ten-MicroRNA signature identified from a genome-wide MicroRNA expression profiling in human epithelial ovarian cancer. *PLoS ONE* 9:e96472. doi: 10.1371/journal.pone.0096472

Wang, Z., Ji, G., Wu, Q., Feng, S., Zhao, Y., Cao, Z., et al. (2018). Integrated microarray meta-analysis identifies miRNA-27a as an oncogene in ovarian cancer by inhibiting FOXO1. *Life Sci.* 210, 263–270. doi: 10.1016/j.lfs.2018.08.043

Warren, J. D., Xiong, W., Bunker, A. M., Vaughn, C. P., Furtado, L. V., Roberts, W. L., et al. (2011). Septin 9 methylated DNA is a sensitive and specific blood test for colorectal cancer. *BMC Med.* 9:133. doi: 10.1186/1741-7015-9-133

Wasenang, W., Chaiyarit, P., Proungvitaya, S., and Limpaiboon, T. (2019). Serum cell-free DNA methylation of OPCML and HOXD9 as a biomarker that may aid in differential diagnosis between cholangiocarcinoma and other biliary diseases. *Clin. Epigenetics* 11:39. doi: 10.1186/s13148-019-0634-0

Wei, K.-L., Chou, J.-L., Chen, Y.-C., Jin, H., Chuang, Y.-M., Wu, C.-S., et al. (2019). Methylomics analysis identifies a putative STAT3 target, SPG20, as a noninvasive epigenetic biomarker for early detection of gastric cancer. *PLoS ONE* 14:e0218338. doi: 10.1371/journal.pone.0218338

Wei, S. H. (2006). Prognostic DNA methylation biomarkers in ovarian cancer. *Clin. Cancer Res.* 12, 2788–2794. doi: 10.1158/1078-0432.CCR-05-1551

Wei, S. H., Chen, C.-M., Strathdee, G., Harnsomburana, J., Shyu, C.-R., Rahmatpanah, F., et al. (2002). Methylation microarray analysis of late-stage ovarian carcinomas distinguishes progression-free survival in patients and identifies candidate epigenetic markers. *Clin. Cancer Res.* 8, 2246–2252.

Weichert, W., Denkert, C., Noske, A., Darb-Esfahani, S., Dietel, M., Kalloger, S. E., et al. (2008). Expression of class I histone deacetylases indicates poor prognosis in endometrioid subtypes of ovarian and endometrial carcinomas. *Neoplasia* 10, 1021–1027. doi: 10.1593/neo.08474

Weisenberger, D. J., Trinh, B. N., Campan, M., Sharma, S., Long, T. I., Ananthnarayan, S., et al. (2008). DNA methylation analysis by digital bisulfite genomic sequencing and digital MethyLight. *Nucleic Acids Res.* 36, 4689–4698. doi: 10.1093/nar/gkn455

Widschwendter, A., Müller, H. M., Hubalek, M. M., Wiedemair, A., Fiegl, H., Goebel, G., et al. (2004). Methylation status and expression of human telomerase reverse transcriptase in ovarian and cervical cancer. *Gynecol. Oncol.* 93, 407–416. doi: 10.1016/j.ygyno.2004.01.036

Widschwendter, M., Apostolidou, S., Jones, A. A., Fourkala, E. O., Arora, R., Pearce, C. L., et al. (2009). HOXA methylation in normal endometrium from premenopausal women is associated with the presence of ovarian cancer: a proof of principle study. *Int. J. Cancer* 125, 2214–2218. doi: 10.1002/ijc.24599

Widschwendter, M., Jiang, G., Woods, C., Müller, H. M., Fiegl, H., Goebel, G., et al. (2004). DNA hypomethylation and ovarian cancer biology. *Cancer Res.* 64, 4472–4480. doi: 10.1158/0008-5472.CAN-04-0238

Widschwendter, M., Zikan, M., Wahl, B., Lempiäinen, H., Paprotka, T., Evans, I., et al. (2017). The potential of circulating tumor DNA methylation analysis for the early detection and management of ovarian cancer. *Genome Med.* 9:116. doi: 10.1186/s13073-017-0500-7

Wilcox, C. B., Baysal, B. E., Gallion, H. H., Strange, M. A., and DeLoia, J. A. (2005). High-resolution methylation analysis of the BRCA1 promoter in ovarian tumors. *Cancer Genet. Cytogenet.* 159, 114–122. doi: 10.1016/j.cancergencyto.2004.12.017

Wilczynski, M., Zytko, E., Danielska, J., Szymanska, B., and Dzieniecka, M., Nowak, M., et al. (2018). Clinical significance of miRNA-21,−103,−129,−150 in serous ovarian cancer. *Arch. Gynecol. Obstet.* 297, 741–748. doi: 10.1007/s00404-018-4660-5

Wiley, A., Katsaros, D., Chen, H., Rigault de la Longrais, I. A., Beeghly, A., Puopolo, M., et al. (2006a). Aberrant promoter methylation of multiple genes in malignant ovarian tumors and in ovarian tumors with low malignant potential. *Cancer* 107, 299–308. doi: 10.1002/cncr.21992

Wiley, A., Katsaros, D., Fracchioli, S., and Yu, H. (2006b). Methylation of the insulin-like growth factor binding protein-3 gene and prognosis of epithelial ovarian cancer. *Int. J. Gynecol. Cancer* 16, 210–218. doi: 10.1111/j.1525-1438.2006.00299.x

Wilting, R. H., and Dannenberg, J.-H. (2012). Epigenetic mechanisms in tumorigenesis, tumor cell heterogeneity and drug resistance. *Drug Resist. Updat.* 15, 21–38. doi: 10.1016/j.drup.2012.01.008

Wittenberger, T., Sleigh, S., Reisel, D., Zikan, M., Wahl, B., Alunni-Fabbroni, M., et al. (2014). DNA methylation markers for early detection of women's cancer: promise and challenges. *Epigenomics* 6, 311–327. doi: 10.2217/epi.14.20

Woloszynska-Read, A., James, S. R., Link, P. A., Yu, J., Odunsi, K., and Karpf, A. R. (2007). DNA methylation-dependent regulation of BORIS/CTCFL expression in ovarian cancer. *Cancer Immun.* 7, 21.

Woolas, R. P., Xu, F. J., Jacobs, I. J., Yu, Y. H., Daly, L., Berchuck, A., et al. (1993). Elevation of multiple serum markers in patients with stage I ovarian cancer. *J. Natl. Cancer Inst.* 85, 1748–1751. doi: 10.1093/jnci/85.21.1748

Wu, Q., Lothe, R. A., Ahlquist, T., Silins, I., Tropé, C. G., Micci, F., et al. (2007). DNA methylation profiling of ovarian carcinomas and their *in vitro* models identifies HOXA9, HOXB5, SCGB3A1, and CRABP1 as novel targets. *Mol. Cancer* 6:45. doi: 10.1186/1476-4598-6-45

Wurz, K., Garcia, R. L., Goff, B. A., Mitchell, P. S., Lee, J. H., Tewari, M., et al. (2010). MiR-221 and MiR-222 alterations in sporadic ovarian carcinoma: relationship to CDKN1B, CDKNIC and overall survival. *Genes. Chromosomes Cancer.* 49, 577–584. doi: 10.1002/gcc.20768

Wyman, S. K., Parkin, R. K., Mitchell, P. S., Fritz, B. R., O'Briant, K., Godwin, A. K., et al. (2009). Repertoire of microRNAs in epithelial ovarian cancer as determined by next generation sequencing of small RNA cDNA Libraries. *PLoS ONE* 4:e5311. doi: 10.1371/journal.pone.0005311

Xi, Q.-P., Pu, D.-H., and Lu, W.-N. (2017). Research on application value of combined detection of serum CA125, HE4 and TK1 in the diagnosis of ovarian cancer. *Eur. Rev. Med. Pharmacol. Sci.* 21, 4536–4541.

Xu, W., Lu, J., Zhao, Q., Wu, J., Sun, J., Han, B., et al. (2019). Genome-wide plasma cell-free DNA methylation profiling identifies potential biomarkers for lung cancer. *Dis. Markers* 2019, 1–7. doi: 10.1155/2019/4108474

Xu, Y.-Z., Xi, Q.-H., Ge, W.-L., and Zhang, X.-Q. (2013). Identification of serum MicroRNA-21 as a biomarker for early detection and prognosis in human epithelial ovarian cancer. *Asian Pac. J. Cancer Prev.* 14, 1057–1060. doi: 10.7314/APJCP.2013.14.2.1057

Yang, H., Kong, W., He, L., Zhao, J.-J., O'Donnell, J. D., Wang, J., et al. (2008). MicroRNA expression profiling in human ovarian cancer: miR-214 induces cell survival and cisplatin resistance by targeting PTEN. *Cancer Res.* 68, 425–433. doi: 10.1158/0008-5472.CAN-07-2488

Yang, N., Kaur, S., Volinia, S., Greshock, J., Lassus, H., Hasegawa, K., et al. (2008). MicroRNA microarray identifies Let-7i as a novel biomarker and therapeutic target in human epithelial ovarian cancer. *Cancer Res.* 68, 10307–10314. doi: 10.1158/0008-5472.CAN-08-1954

Yang, Y., Chaerkady, R., Beer, M. A., Mendell, J. T., and Pandey, A. (2009). Identification of miR-21 targets in breast cancer cells using a quantitative proteomic approach. *Proteomics* 9, 1374–1384. doi: 10.1002/pmic.200800551

Yang, Z., Qi, W., Sun, L., Zhou, H., Zhou, B., and Hu, Y. (2018). 10.1186/s13148-019-0634-0. *Adv. Clin. Exp. Med.* 28, 361–366. doi: 10.17219/acem/84935

Yap, O. W. S., Bhat, G., Liu, L., and Tollefsbol, T. O. (2009). Epigenetic modifications of the Estrogen receptor beta gene in epithelial ovarian cancer cells. *Anticancer Res.* 29, 139–144.

Yarbro, J. W., Page, D. L., Fielding, L. P., Partridge, E. E., and Murphy, G. P. (1999). American Joint Committee on Cancer prognostic factors consensus conference. *Cancer* 86, 2436–2446.

Yazbek, J., Raju, S. K., Ben-Nagi, J., Holland, T. K., Hillaby, K., and Jurkovic, D. (2008). Effect of quality of gynaecological ultrasonography on management of patients with suspected ovarian cancer: a randomised controlled trial. *Lancet Oncol.* 9, 124–131. doi: 10.1016/S1470-2045(08)70005-6

Yeh, C.-M., Shay, J., Zeng, T.-C., Chou, J.-L., Huang, T. H.-M., Lai, H.-C., et al. (2014). Epigenetic silencing of ARNTL, a circadian gene and potential tumor suppressor in ovarian cancer. *Int. J. Oncol.* 45, 2101–2107. doi: 10.3892/ijo.2014.2627

Yeh, K.-T., Chen, T.-H., Yang, H.-W., Chou, J.-L., Chen, L.-Y., Yeh, C.-M., et al. (2011). Aberrant TGFβ/SMAD4 signaling contributes to epigenetic silencing of a putative tumor suppressor, RunX1T1 in ovarian cancer. *Epigenetics* 6, 727–739. doi: 10.4161/epi.6.6.15856

Yi, X., Guo, J., Guo, J., Sun, S., Yang, P., Wang, J., et al. (2017). EZH2-mediated epigenetic silencing of TIMP2 promotes ovarian cancer migration and invasion. *Sci. Rep.* 7:3568. doi: 10.1038/s41598-017-03362-z

Yoon, J. H., Dammann, R., and Pfeifer, G. P. (2001). Hypermethylation of the CpG island of the RASSF1A gene in ovarian and renal cell carcinomas. *Int. J. Cancer* 94, 212–217. doi: 10.1002/ijc.1466

Yu, J., Ryan, D. G., Getsios, S., Oliveira-Fernandes, M., Fatima, A., and Lavker, R. M. (2008). MicroRNA-184 antagonizes microRNA-205 to maintain SHIP2 levels in epithelia. *Proc. Natl. Acad. Sci. U.S.A.* 105, 19300–19305. doi: 10.1073/pnas.0803992105

Yu, Y., Fujii, S., Yuan, J., Luo, R. Z., Wang, L., Bao, J., et al. (2003). Epigenetic regulation of ARHI in breast and ovarian cancer cells. *Ann. N. Y. Acad. Sci.* 983, 268–277. doi: 10.1111/j.1749-6632.2003.tb05981.x

Zhang, H., Zhang, S., Cui, J., Zhang, A., Shen, L., and Yu, H. (2008). Expression and promoter methylation status of mismatch repair gene hMLH1 and hMSH2 in epithelial ovarian cancer. *Aust. N. Z. J. Obstet. Gynaecol.* 48, 505–509. doi: 10.1111/j.1479-828X.2008.00892.x

Zhang, J., Ye, F., Chen, H., Ye, D., Lu, W., and Xie, X. (2006). [Deletion of OPCML gene and promoter methylation in ovarian epithelial carcinoma]. *Zhongguo Yi Xue Ke Xue Yuan Xue Bao* 28, 173–177.

Zhang, L., Huang, J., Yang, N., Greshock, J., Megraw, M. S., Giannakakis, A., et al. (2006). microRNAs exhibit high frequency genomic alterations in human cancer. *Proc. Natl. Acad. Sci. U.S.A.* 103, 9136–9141. doi: 10.1073/pnas.0508889103

Zhang, L., Volinia, S., Bonome, T., Calin, G. A., Greshock, J., Yang, N., et al. (2008). Genomic and epigenetic alterations deregulate microRNA expression in human epithelial ovarian cancer. *Proc. Natl. Acad. Sci. U.S.A.* 105, 7004–7009. doi: 10.1073/pnas.0801615105

Zhang, Q., Hu, G., Yang, Q., Dong, R., Xie, X., Ma, D., et al. (2013). A multiplex methylation-specific PCR assay for the detection of early-stage ovarian cancer using cell-free serum DNA. *Gynecol. Oncol.* 130, 132–139. doi: 10.1016/j.ygyno.2013.04.048

Zhang, S., Wei, L., Zhang, A., Zhang, L., and Yu, H. (2009). RUNX3 gene methylation in epithelial ovarian cancer tissues and ovarian cancer cell lines. *Omics J. Integr. Biol.* 13, 307–311. doi: 10.1089/omi.2009.0030

Zhang, W., Barger, C. J., Eng, K. H., Klinkebiel, D., Link, P. A., Omilian, A., et al. (2016). PRAME expression and promoter hypomethylation in epithelial ovarian cancer. *Oncotarget* 7, 45352–45369. doi: 10.18632/oncotarget.9977

Zhang, W., Barger, C. J., Link, P. A., Mhawech-Fauceglia, P., Miller, A., Akers, S. N., et al. (2015). DNA hypomethylation-mediated activation of *Cancer/Testis Antigen 45* (*CT45*) genes is associated with disease progression and reduced survival in epithelial ovarian cancer. *Epigenetics* 10, 736–748. doi: 10.1080/15592294.2015.1062206

Zhang, Z., Bast, R. C., Yu, Y., Li, J., Sokoll, L. J., Rai, A. J., et al. (2004). Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res.* 64, 5882–5890. doi: 10.1158/0008-5472.CAN-04-0746

Zhao, L., Wang, W., Xu, L., Yi, T., Zhao, X., Wei, Y., et al. (2019). Integrative network biology analysis identifies miR-508-3p as the determinant for the mesenchymal identity and a strong prognostic biomarker of ovarian cancer. *Oncogene* 38, 2305–2319. doi: 10.1038/s41388-018-0577-5

Zhao, L., Yu, C., Zhou, S., Lau, W. B., Lau, B., Luo, Z., et al. (2016). Epigenetic repression of PDZ-LIM domain-containing protein 2 promotes ovarian cancer via NOS2-derived nitric oxide signaling. *Oncotarget* 7, 1408–1420. doi: 10.18632/oncotarget.6368

Zheng, X., Chen, S., Li, L., Liu, X., Liu, X., Dai, S., et al. (2018). Evaluation of HE4 and TTR for diagnosis of ovarian cancer: comparison with CA-125. *J. Gynecol. Obstet. Hum. Reprod.* 47, 227–230. doi: 10.1016/j.jogoh.2018.03.010

Zhou, F., Tao, G., Chen, X., Xie, W., Liu, M., and Cao, X. (2014). Methylation of OPCML promoter in ovarian cancer tissues predicts poor patient survival. *Clin. Chem. Lab. Med.* 52, 735–742. doi: 10.1515/cclm-2013-0736

Zhou, Q.-H., Zhao, Y.-M., Jia, L.-L., and Zhang, Y. (2017). Mir-595 is a significant indicator of poor patient prognosis in epithelial ovarian cancer. *Eur. Rev. Med. Pharmacol. Sci.* 21, 4278–4282.

Zuberi, M., Mir, R., Das, J., Ahmad, I., Javid, J., Yadav, P., et al. (2015). Expression of serum miR-200a, miR-200b, and miR-200c as candidate biomarkers in epithelial ovarian cancer and their association with clinicopathological features. *Clin. Transl. Oncol.* 17, 779–787. doi: 10.1007/s12094-015-1303-1

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Singh, Gupta and Sachan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 27 September 2019
doi: 10.3389/fgene.2019.00910

A Novel DNA Methylation-Based Signature Can Predict the Responses of MGMT Promoter Unmethylated Glioblastomas to Temozolomide

Rui-Chao Chai[1,2†‡], Yu-Zhou Chang[2,3†], Qiang-Wei Wang[1†‡], Ke-Nan Zhang[1,2,3‡], Jing-Jun Li[1‡], Hua Huang[1‡], Fan Wu[1‡], Yu-Qing Liu[1*‡] and Yong-Zhi Wang[1,2,3*‡]*

[1] Department of Molecular Neuropathology, Beijing Neurosurgical Institute, Beijing Tiantan Hospital, Capital Medical University, Beijing, China, [2] China National Clinical Research Center for Neurological Diseases, Beijing Tiantan Hospital, Capital Medical University, Beijing, China, [3] Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, Beijing, China

OPEN ACCESS

Edited by:
Jiucun Wang,
Fudan University,
China

Reviewed by:
Mariana Brait,
Johns Hopkins University,
United States
David D. Eisenstat,
University of Alberta,
Canada

***Correspondence:**
Yong-Zhi Wang
yongzhiwang_bni@163.com
Yu-Qing Liu
liuyuqing0704@163.com;
Fan Wu
wufan0510284@163.com;

†These authors have contributed equally to this work

‡Members of Chinese Giloma Genome Atlas Network (CGGA)

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: 24 April 2019
Accepted: 28 August 2019
Published: 27 September 2019

Citation:
Chai R-C, Chang Y-Z, Wang Q-W, Zhang K-N, Li J-J, Huang H, Wu F, Liu Y-Q and Wang Y-Z (2019) A Novel DNA Methylation-Based Signature Can Predict the Responses of MGMT Promoter Unmethylated Glioblastomas to Temozolomide. Front. Genet. 10:910. doi: 10.3389/fgene.2019.00910

Glioblastoma (GBM) is the most malignant glioma, with a median overall survival (OS) of 14–16 months. Temozolomide (TMZ) is the first-line chemotherapy drug for glioma, but whether TMZ should be withheld from patients with GBMs that lack O^6-methylguanine-DNA methyltransferase (*MGMT*) promoter methylation is still under debate. DNA methylation profiling holds great promise for further stratifying the responses of *MGMT* promoter unmethylated GBMs to TMZ. In this study, we studied 147 TMZ-treated *MGMT* promoter unmethylated GBM, whose methylation information was obtained from the HumanMethylation27 (HM-27K) BeadChips (n = 107) and the HumanMethylation450 (HM-450K) BeadChips (n = 40) for training and validation, respectively. In the training set, we performed univariate Cox regression and identified that 3,565 CpGs were significantly associated with the OS of the TMZ-treated *MGMT* promoter unmethylated GBMs. Functional analysis indicated that the genes corresponding to these CpGs were enriched in the biological processes or pathways of mitochondrial translation, cell cycle, and DNA repair. Based on these CpGs, we developed a 31-CpGs methylation signature utilizing the least absolute shrinkage and selection operator (LASSO) Cox regression algorithm. In both training and validation datasets, the signature identified the TMZ-sensitive GBMs in the *MGMT* promoter unmethylated GBMs, and only the patients in the low-risk group appear to benefit from the TMZ treatment. Furthermore, these identified TMZ-sensitive *MGMT* promoter unmethylated GBMs have a similar OS when compared with the *MGMT* promoter methylated GBMs after TMZ treatment in both two datasets. Multivariate Cox regression demonstrated the independent prognostic value of the signature in TMZ-treated *MGMT* promoter unmethylated GBMs. Moreover, we also noticed that the hallmark of epithelial–mesenchymal transition, ECM related biological processes and pathways were highly enriched in the MGMT unmethylated GBMs with the high-risk score, indicating that enhanced ECM activities could be involved in the TMZ-resistance of GBM.

In conclusion, our findings promote our understanding of the roles of DNA methylation in *MGMT* umethylated GBMs and offer a very promising TMZ-sensitivity predictive signature for these GBMs that could be tested prospectively.

Keywords: glioblastoma, DNA methylation, temozolomide, MGMT, signature

INTRODUCTION

Glioma is the most common type of malignant brain tumor in adults (Jiang et al., 2016; Chai et al., 2019b). Glioblastoma (GBM, WHO IV) is the most malignant glioma, accounting for 50–60% of total glioma (Louis et al., 2016). Currently, the prognosis for patients with GBM is still dismal, with a median overall survival (OS) of 14–16 months (Jiang et al., 2016; Louis et al., 2016; Chai et al., 2019a). The alkylating agent temozolomide (TMZ) is the first-line chemotherapy drug for glioma. TMZ is used concurrently with radiation and then provided as monotherapy during adjuvant treatment. The promoter methylation level of the O^6-methylguanine-DNA methyltransferase (*MGMT*), a ubiquitous DNA repair enzyme which can rapidly reverse alkylation at the O^6 position, has been acknowledged as a predictive marker for TMZ sensitivity (Hegi et al., 2005; Chai et al., 2019a; Chai et al., 2019e). *MGMT* promoter methylated GBM displays higher sensitivity to TMZ treatment than *MGMT* promoter unmethylated GBM (Hegi et al., 2005; Wick et al., 2014; Chai et al., 2019a). However, we noticed that the prognosis for TMZ treated *MGMT* promoter unmethylated GBM is still largely heterogeneous, indicating that some other factors may also affect the sensitivity of *MGMT* promoter unmethylated GBM to TMZ treatment. Thus, further stratification of these GBM is urgently needed.

In the central nervous system, DNA methylation profiling has been used as a robust and reproducible method to further stratify the tumors into different subgroups (Sturm et al., 2012; Pajtler et al., 2015; Sturm et al., 2016). Moreover, general DNA methylation or a group of CpGs methylation profiling could also serve as biomarkers to evaluate drug- or radio-therapeutic sensitivity in various diseases, including tumors (Kumar et al., 2018; Zhao et al., 2018b; Chen et al., 2019b). In a recent study, a five-CpG DNA methylation score has shown its value in predicting metastatic-lethal outcomes of males suffering localized prostate cancer, treated with radical prostatectomy (Zhao et al., 2018b). The rapid accumulation of DNA methylation datasets makes it also possible to further stratify the glioma and may uncover novel biomarkers for management of gliomas. Recently, DNA methylation profiling of 23 DNA damage response (DDR) genes was shown to be associated with benefit from RT or TMZ therapy in IDH mutant low-grade glioma (Bady et al., 2018). Nevertheless, whether a group of CpGs DNA methylation profiling can predict the TMZ sensitivity of *MGMT* promoter unmethylated GBM remains unclear.

Here, we aimed to identify TMZ-sensitive GBMs in the entity of *MGMT* promoter unmethylated GBMs, using DNA methylation profiling. We adopted 107 and 40 TMZ treated *MGMT* promoter unmethylated GBMs as the training set and the validation set, respectively. We identified a list of CpGs whose methylation levels are significantly associated with the OS of TMZ-treated MGMT promoter unmethylated GBMs by univariate Cox regression analyses. Based on this, we developed a 31-CpGs TMZ therapeutic prognosis risk signature in the *MGMT* promoter unmethylated GBMs. This risk signature could successfully identify a subgroup of TMZ treated *MGMT* promoter unmethylated GBMs which have a similar prognosis when compared with the TMZ treated *MGMT* promoter methylated GBMs.

MATERIALS AND METHODS

Samples Information

A total of 376 cases were enrolled in this study according to the following criteria: (a) diagnosed with GBM; (b) the DNA methylation data could be obtained; (c) the TMZ treatment option is available. The DNA methylation data and corresponding clinicopathological features for these cases were obtained from The Cancer Genome Atlas (TCGA) (http://cancergenome.nih.gov/). Within the 376 cases, the DNA methylation information of 279 cases (the 27K cohort) was collected from the HumanMethylation27 (HM-27K) BeadChips dataset, and the other 97 cases (the 450K cohort) were obtained from the HumanMethylation450 (HM-450K) BeadChips dataset. Clinicopathological information for all cases is summarized in **Supplementary Table 1**.

Of all 279 cases in the 27K cohort, 107 cases who received TMZ treatment and also with unmethylated *MGMT* were used to investigate the TMZ therapeutic prognosis value of CpGs methylation levels, and we also developed a risk signature using these cases. Of the 97 cases in the 450K cohort, 40 TMZ treated cases with unmethylated *MGMT* were used as the validation cohort. Clinicopathological information for these 147 cases is summarized in **Table 1**. There is no statistically significant difference for the clinicopathological features between the training and validation cohorts.

Analytical Approach

The approach and workflow for the selection of TMZ therapeutic prognosis associated CpGs, functional annotation for the genes corresponding to these CpGs, development and validation of a TMZ therapeutic prognostic risk signature, analysis of the correlation between the risk signature and other clinicopathological features, and the functional analysis of genes associated with the risk signature are summarized in **Figure 1**.

TABLE 1 | Clinicopathological characteristics for *MGMT* unmethylated GBM patients who received TMZ.

		The 27K cohort		The 450K cohort		P-value
		Number	Percentage	Number	Percentage	
Total		107	100.00%	40	100.00%	
Age		20–89 (57)		23–78 (58)		0.99[a]
	< median	50	46.73%	19	47.50%	
	≥ median	57	53.27%	21	52.50%	
Gender						0.53[b]
	Male	69	64.49%	28	70.00%	
	Female	38	35.51%	12	30.00%	
IDH						0.32[b]
	Mutant	4	3.74%	0	0.00%	
	Wildtype	90	84.11%	37	92.50%	
	NA	13	12.15%	3	7.50%	
TCGA defined subgroup						0.10[b]
	Neural	10	9.35%	2	5.00%	
	Proneural	18	16.82%	6	15.00%	
	Classical	30	28.04%	12	30.00%	
	Mesenchymal	43	40.19%	12	30.00%	
	NA	6	5.61%	8	20.00%	
Chr 7 gain/Chr 10 loss						0.37[b]
	Combined alteration	73	68.22%	28	70.00%	
	No combined alteration	29	27.10%	12	30.00%	
	NA	5	4.67%	0	0.00%	

[a]*P-value is calculated by the nonparametric test.* [b]*P-value is calculated by the Chi-square tests.*

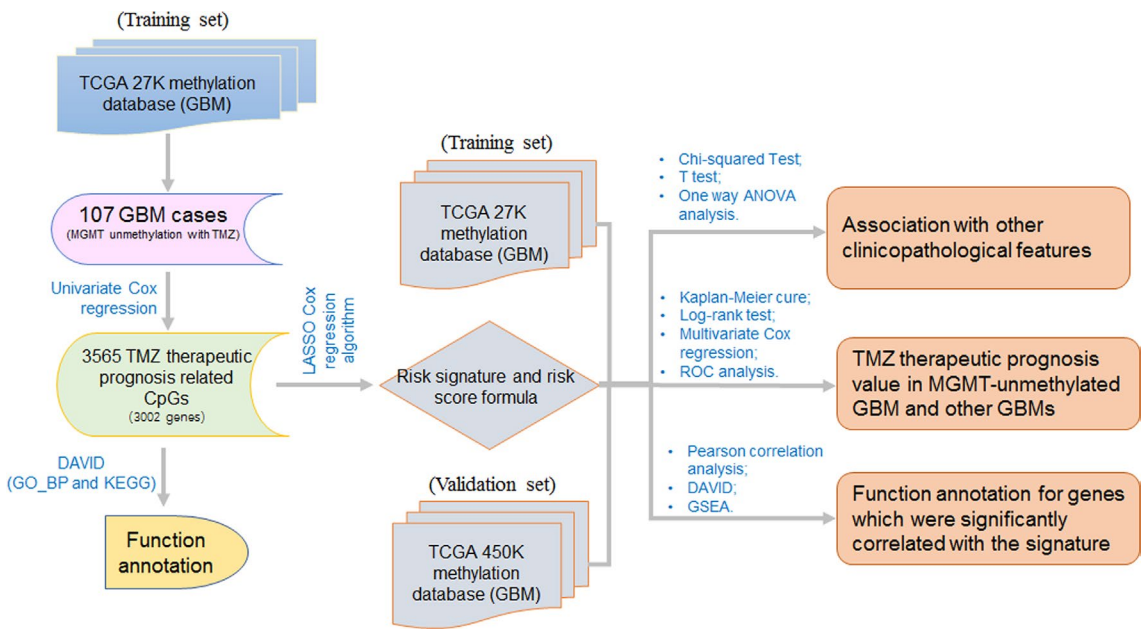


FIGURE 1 | The workflow for this study. The workflow for the selection of TMZ therapeutic prognosis related CpGs, development and validation of a TMZ therapeutic prognostic risk signature, and the functional analysis of genes that are correlated with the signature risk score.

Identification of the Risk Signature

We performed univariate Cox regression analysis of the CpGs methylation to identify CpGs significantly correlated with the prognosis of TMZ treated *MGMT* unmethylated GBM in the 27K cohort. Then, we used the least absolute shrinkage and selection operator (LASSO) Cox regression algorithm to develop an optimal risk signature with the minimum number of CpGs (Dai et al., 2018; Zhou et al., 2018; Chai et al., 2019d). Finally, a set of 31 CpGs and their coefficients were determined by the minimum criteria, which involves selecting the best penalty

parameter λ associated with the smallest 10-fold cross validation within the training set. The risk score for the risk signature was calculated using the formula:

$$\text{Risk score} = \Sigma_{i=1}^{n} Coef_i * x_i$$

where *Coef* is the coefficient and x_i is the beta-value of each selected CpGs. In both groups (cohorts), we used the beta-value [beta-value = the methylated signal/(methylated signal + unmethylated signal)] to represent the methylation level of each CpGs. Since the Risk score was calculated as a weighted sum of the methylation level of all selected CpG sites (Chai et al., 2019d; Chen et al., 2019a), we just used the original beta value of each CpG sites to calculate the risk scores.

We did not directly compare the samples in two different groups (cohorts). In order to avoid the bias caused by the different arrays, we only compared the methylation levels among samples in the same cohort. We first developed the risk signature in 107 samples used HumanMethylation27 (HM-27K) BeadChips. Then, we used another 40 samples to validate the prognostic value of the proposed signature. Patients were divided into "high-risk" and "low-risk" groups using the respective median risk score as the cutoff value in both the training and validation datasets.

Bioinformatic Analysis

Significance analysis of microarray (SAM) was performed to identify differentially expressed genes within the low- and high-risk scores. We performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses with the Database for Annotation, Visualization, and Integrated Discovery (http://david.abcc.ncifcrf.gov/home.jsp) to functionally annotate genes corresponding to the CpGs with prognosis of TMZ treated MGMT unmethylated GBM and genes that were differentially expressed between the low- and high-risk groups in the 27K cohort. Gene Set Enrichment Analysis (GSEA) was performed to investigate the functions of genes that were differentially expressed between the low- and high-risk groups in the 27K cohort.

Statistical Analysis

We used the nonparametric test to compare the distribution of age between the low- and high-risk groups, and Chi-square tests were used to compare the distribution of other clinicopathological features. A one-way analysis of variance was performed to compare the risk scores in patients grouped by the TCGA defined subtypes. Student's *t* test was performed to compare the risk scores in patients grouped by other clinical or molecular-pathological characteristics.

Univariate and multivariate Cox regression analysis was performed to determine the prognostic value of the risk score and various clinical and molecular–pathological characteristics.

The Kaplan–Meier method with a two-sided log-rank test was used to compare the OS of patients stratified by the risk scores or other clinicopathological features. All statistical analyses were conducted using R v3.4.1 (https://www.r-project.org/), SPSS 16.0 (SPSS, Inc., Chicago, IL, USA) and Prism 7 (GraphPad Software, Inc., La Jolla, CA, USA).

RESULTS

A Set of CpGs' Methylation Profile Could Predict the TMZ Therapeutic Response of *MGMT* Unmethylated GBMs

To assess the TMZ therapeutic response value of the methylation of CpGs, we performed univariate Cox regression analysis of all CpGs methylation levels in the 107 TMZ treated *MGMT* unmethylated GBMs of the 27K cohort. We found that the methylation levels of 3,565 CpGs were significantly correlated with the OS of these GBMs (**Supplementary Table 2**). Based on the methylation profile of these genes, we could divide the 107 TMZ treated *MGMT* unmethylated GBMs into 3 clusters (Cluster A–C) in the heatmap (**Figure 2A**). We observed that patients in the Cluster A had significantly shorter survival than patients in the Cluster B and C, and the patients in the Cluster B and C had a similar OS with the TMZ treated *MGMT* methylated GBM patients (**Figure 2B**).

We also investigated the functions of the respective genes for the 3,565 CpGs. Three thousand one hundred eighty-two of these CpGs methylation levels were found to have a HR < 1 and were considered protective-associated, and the remaining 383 CpGs methylation levels with a HR >1 were considered risk-associated. GO terms of biological progress (BP) and KEGG pathway analysis indicated that the genes corresponding to the protective-associated CpGs were enriched in the processes of mitochondrial translation, protein modification, cell cycle, DNA repair, others, and pathways in cancer (**Figures 2C**, **D**). In contrast, the genes corresponding to the risk-associated CpGs were mainly enriched in the cellular membrane-associated biological processes and pathways (**Figures 2C**, **D**).

Identification of a 31-CpGs Panel as a TMZ Therapeutic Prognosis Risk Signature in *MGMT* Unmethylated GBMs

We next sought to develop a representative "risk signature" with a small number of CpGs to predict the TMZ therapeutic responses of the *MGMT* unmethylated GBMs. We applied the LASSO Cox regression algorithm to the 3,565 CpGs in 107 GBMs of the 27K cohort (**Figure 3A**). Finally, a total of 31 CpGs were contained in the risk signature, and the respective genes and the coefficients of these CpGs were also shown (**Figure 3B** and **Supplementary Table 3**). Twenty-four of the 31 CpGs are located in the CpG islands of prospective genes, and 5 of the other 7 CpGs are located within 200 bp of the transcription start site of the prospective genes (**Supplementary Table 3**). Most of the genes corresponding to the 31 CpGs have been reported to be involved in the tumorigenesis or prognosis of cancer, including *ATOH1, ATPIG1, ELL3, RBM15B, GATA4, TXN, DLX5, THSD4, Polr2d, LGALS3BP, HIST1H3D, FLRT1, IFI35* and *OSBPL5*. Among these genes, hypermethylation of *THSD4* has been reported to be associated with the prediction of prognosis in GBM

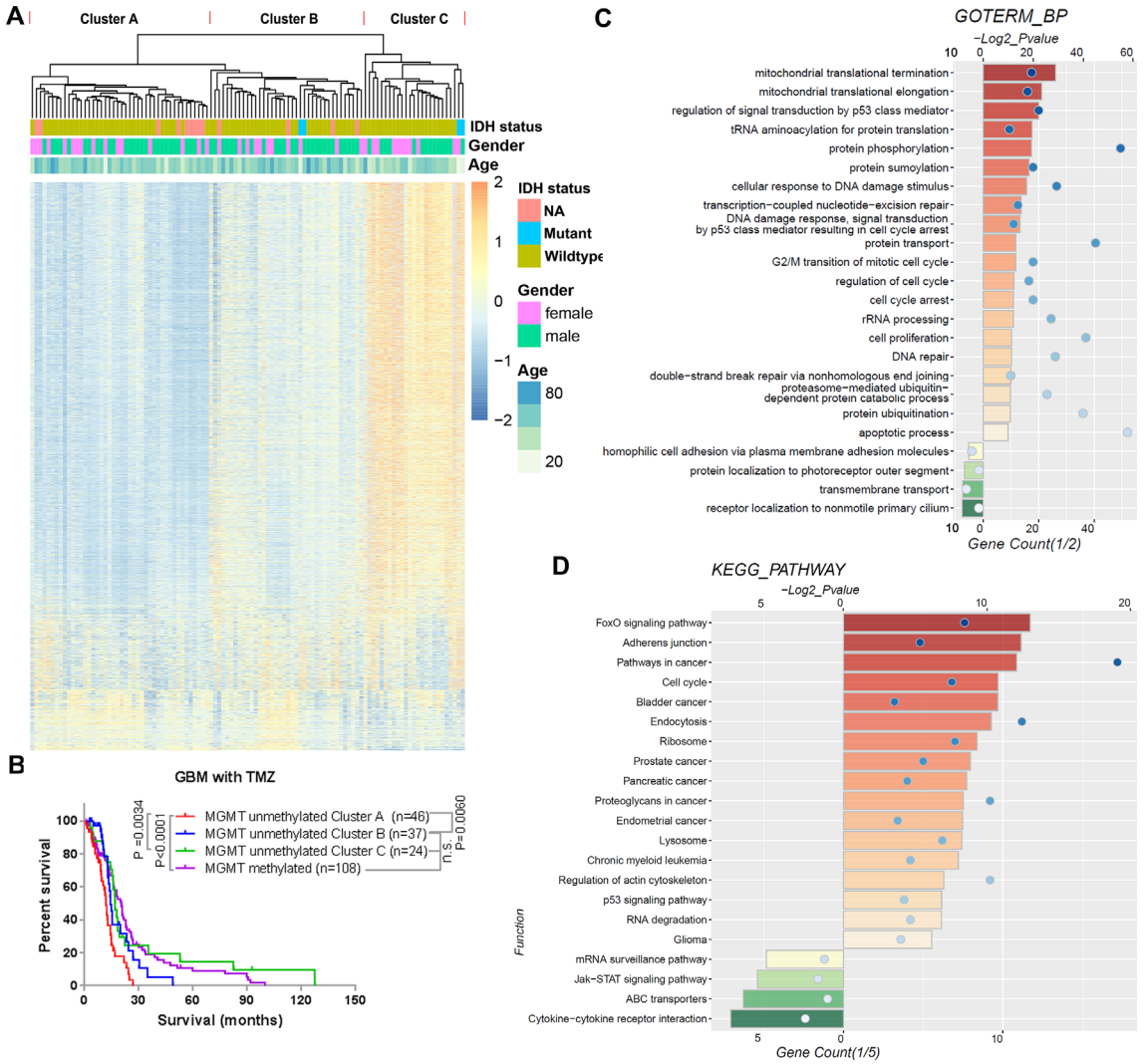


FIGURE 2 | TMZ therapeutic prognosis associated CpGs' methylation profile in *MGMT* unmethylated GBMs. **(A)** Heatmap showing the methylation levels of the 3,565 GpGs associated with the overall survival of TMZ treated patients with *MGMT* unmethylated GBMs. The *MGMT* unmethylated GBMs could be clustered into 3 clusters (Cluster A–C) according to the CpGs methylation levels. **(B)** Kaplan–Meier overall survival (OS) curves of TMZ treated *MGMT* unmethylated GBM patients (stratified by Cluster A–C) and TMZ treated *MGMT* methylated GBM patients. **(C, D)** GO biological process terms **(C)** and KEGG pathways **(D)** enriched among the genes positively and negatively corresponding to the 3,565 GpGs.

(Ma et al., 2015), and *Polr2d* expression is associated with the therapy response of GBM (Serao et al., 2011).

We divided patients into high-risk and low-risk groups using their median risk-score as the cutoffs. We observed significant differences between the low- and high- risk groups with respect to *IDH* status (P = 0.0431), age (P = 0.0069) and TCGA defined subtype (0.0047), but no differences in gender or chromosome 7 gain combined with chromosome 10 loss (chr 7 gain and chr 10 loss) (**Figure 3B** and **Supplementary Table 4**).

Then we investigated the relationship between the risk signature and OS of TMZ treated *MGMT* unmethylated GBM patients. The data showed that patients with low-risk-scores had significantly longer OS than patients with high-risk-scores in both the training (P < 0.0001) and validation (P = 0.0331) datasets (**Figures 3C, D**). In addition, although the OS of *MGMT* methylated GBM patients was significantly longer than that of *MGMT* unmethylated GBM patients (**Supplementary Figure 1**), we noticed that the OS of *MGMT* unmethylated GBM patients in the low-risk group was similar to that of *MGMT* methylated GBM patients in both the training and validation datasets (**Figures 3C, D**).

Association of the Risk Signature and Other Clinicopathological Features

Considering that the TMZ therapeutic prognosis value of the risk signature may be associated with other known clinicopathological features, we examined this in the *MGMT* unmethylated GBMs.

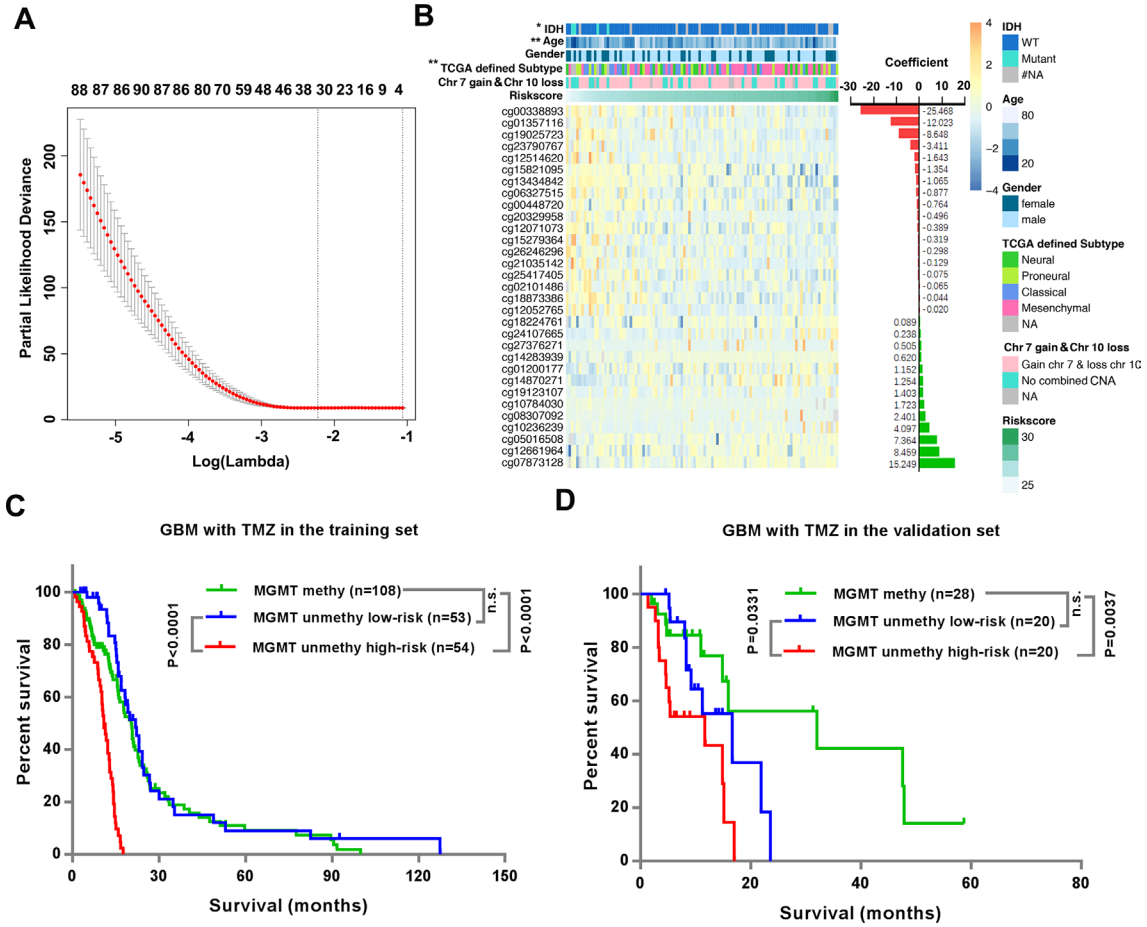


FIGURE 3 | Identification of the risk signature could stratify the TMZ therapeutic prognosis of the MGMT unmethylated GBM. **(A)** Ten-fold cross validation for tuning parameter selection in the LASSO model. The minimum criterion was indicated by the dashed vertical line (left). **(B)** Heatmap shows the association of risk scores and clinicopathological features based on the methylation profile of the 31 CpGs in the signature. The coefficients were calculated by multivariate Cox regression analysis using LASSO. **(C–D)** Kaplan–Meier overall survival (OS) curves for TMZ treated patients with *MGMT* methylated GBMs, TMZ treated patients with *MGMT* unmethylated GBMs with low- or high-risk significance scores in the training set **(C)** and validation set **(D)**, respectively.

We observed that the risk scores were only significantly different between patients stratified by age ($P < 0.05$), rather than gender, chr 7 gain and chr 10 loss, and the TCGA defined subtypes (**Supplementary Figure 2**). We did not compare the risk scores in patients with different *IDH* status, as there were only four *IDH*-mutant patients.

We also performed univariate and multivariate Cox regression analyses in the TMZ treated *MGMT* unmethylated GBMs of both the training and validation datasets. By univariate analysis, the risk score [hazard ratio (HR) = 12.674 (7.661–20.968) in the training set; HR = 1.685 (1.058–2.682) in the validation set] and age [HR = 1.029 (1.009–1.048) in the training set; HR = 1.075 (1.023–1.13) in the validation set] were significantly correlated with the OS in both two datasets (**Table 2**). When including these factors into the multivariate Cox regression analysis, the risk score remained significantly associated with the OS in the training [HR = 12.748 (7.767–21.173)] and validation [HR = 2.157 (1.139–4.086)] datasets (**Table 2**). These results indicated that the risk score can independently predict the TMZ therapeutic prognosis of patients with *MGMT* unmethylated GBMs.

We also investigated the association of risk scores and clinicopathological features in all GBM. We found that the risk scores were only significantly different between patients with different *IDH* status ($P < 0.0001$) or between Proneural subtype and Mesenchymal subtype ($P < 0.01$), but not between patients stratified by age, gender, *MGMT* promoter methylation status, chr 7 gain and chr 10 loss, or treated with or without TMZ (**Supplementary Figure 3**).

Prognosis Value of the Risk Signature in Stratified GBMs

To further understand the TMZ therapeutic prognostic value of the risk signature in *MGMT* unmethylated GBMs, we compared the OS of *MGMT* unmethylated GBMs patients stratified by TMZ treatment status in the low-risk and high-risk groups respectively. The results indicated that patients with TMZ treatment had

TABLE 2 | Univariate and multivariate Cox regression analyses for the risk score in the training and validation set, respectively.

		Univariate Cox regression				Multivariate Cox regression			
		P-value[a]	HR	95% CI		P-Value[b]	HR	95% CI	
				Lower	Higher			Lower	Higher
The training set	Age	**0.028**	1.029	1.009	1.048	0.858	0.998	0.974	1.022
	Gender	0.15	0.705	0.438	1.134	–	–	–	–
	Risk score	**<0.001**	12.674	7.661	20.968	**<0.001**	12.748	7.676	21.173
The validation set	Age	**0.004**	1.075	1.023	1.13	**0.03**	1.093	1.03	1.16
	Gender	0.654	0.802	0.305	2.107	–	–	–	–
	Risk score	**0.028**	1.685	1.058	2.682	**0.018**	2.157	1.139	4.086

[a]The P-value is the sig. value in the univariate cox regression, and the method is Enter; [b]The p-value is the sig. value in the multivariate cox regression analysis, and the method is Enter. P-value <0.05 are highlighted by bold front.

longer OS than that of patients without TMZ treatment in the low-risk group of both the training set (P < 0.0001, **Figure 4A**) and the validation set (P = 0.0456, **Figure 4F**). In contrast, there was no significant difference between patients with or without TMZ treatment in the high-risk group (**Figures 4B**, **G**).

We also investigated the prognostic value of the risk signature in other stratified GBMs. We respectively stratify the GBM patients into four subgroups according to *MGMT* status and TMZ treatment option. In the training set, the risk signature could not stratify the prognosis of three subgroups (TMZ non-treated MGMT unmethylated GBM, TMZ treated MGMT methylated GBM, and TMZ non-treated MGMT methylated GBM) (**Figures 4C–E**). Similar results could also be observed in the validation set except TMZ non-treated *MGMT* unmethylated GBM (**Figures 4G**, **H**).

The Potential Functions Underlying the TMZ Therapeutic Prognostic Value of the Risk Signature

To determine the functional differences between the high-risk and low-risk cases of the TMZ treated *MGMT* unmethylated GBM in the 27K cohort, we identified the differentially (P < 0.05) expressed genes by SAM (**Figure 5A**). GO and KEGG analyses revealed that extracellular matrix related biological processes and signaling pathways were significantly enriched in the high-risk group (**Figures 5B**, **C**). In contrast, the biological processes of T cell differentiation, nervous system development, and transcription were significantly enriched in the low-risk group (**Figure 5B**). Meanwhile, GSEA also indicated that the high-risk cases showed enrichment of "regulation of endothelial cell apoptotic process," "extracellular structure organization,"

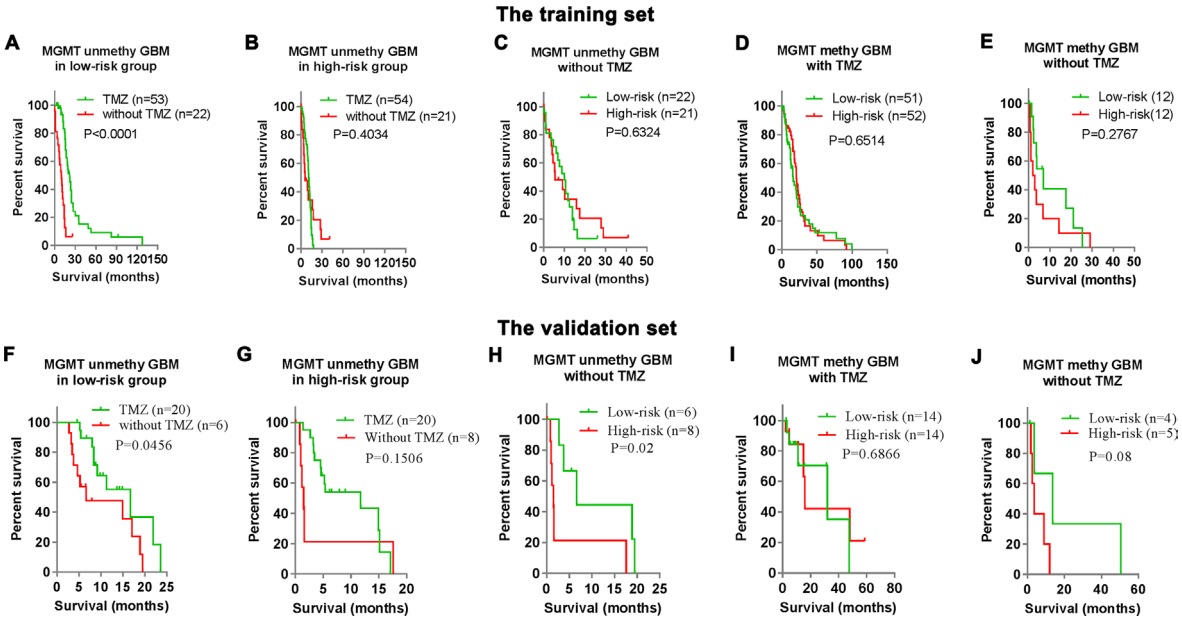


FIGURE 4 | Clinical outcomes prediction of the signature in patients with stratified GBMs. **(A–B)** Kaplan–Meier overall survival (OS) curves for *MGMT* unmethylated GBM patients with or without TMZ treatment in the low-risk group **(A)** and high-risk groups **(B)** of the training set. **(C–E)** Kaplan–Meier overall survival (OS) curves for stratified GBM patients **(C)** *MGMT* unmethylated GBM without TMZ; **(D)** *MGMT* methylated GBM with TMZ; **(E)** *MGMT* methylated GBM without TMZ) with low- or high-risk scores in the training set. **(F–J)** Kaplan–Meier overall survival (OS) curves for stratified GBM patients in the validation set.

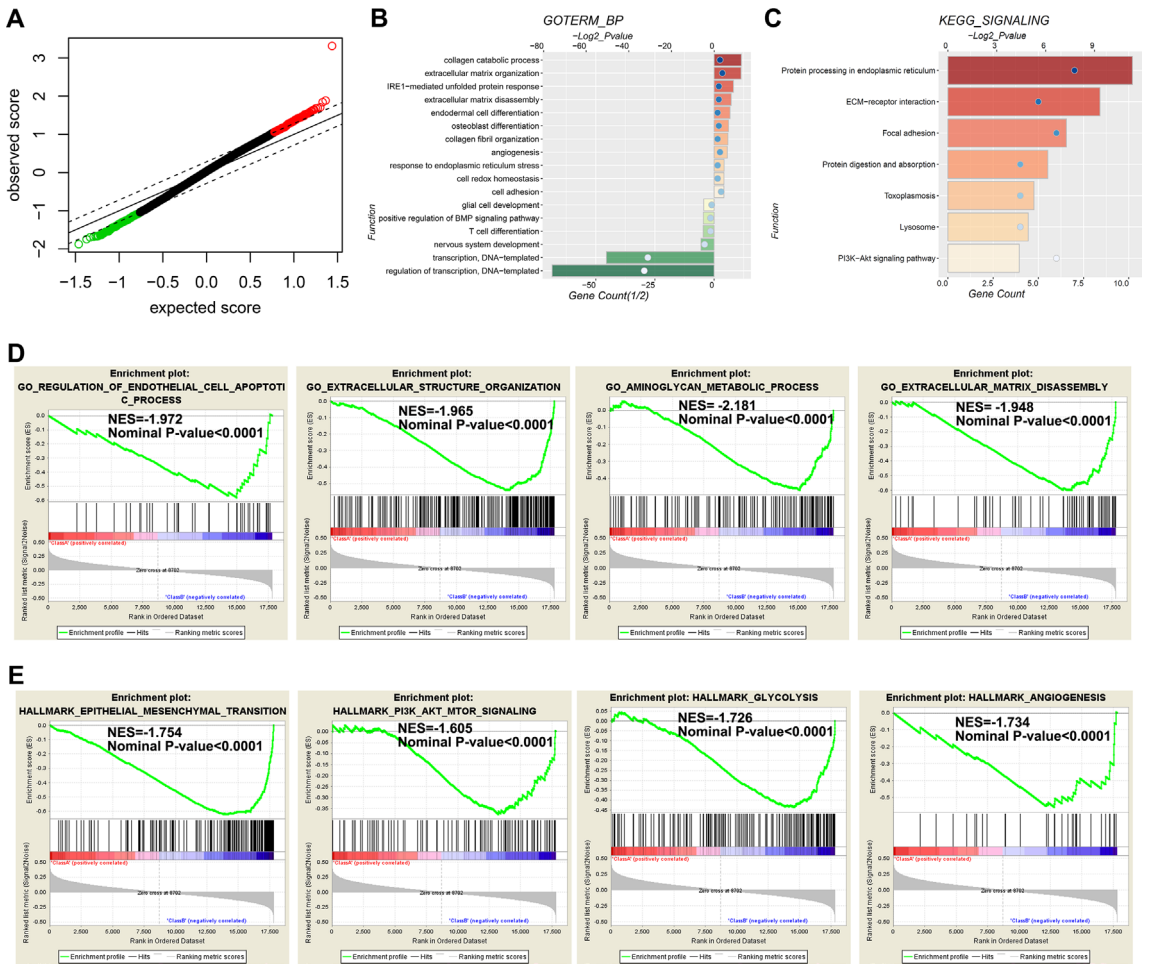


FIGURE 5 | Functional annotation for genes differentially expressed between low- and high-risk groups. **(A)** The differential genes between low- and high-risk groups are shown by green (enriched in the low-risk group) and red (enriched in the high-risk group) dots. **(B–C)** Go analysis **(B)** and KEGG analysis **(C)** are used to evaluate differential genes between low-and high-risk groups. **(D** and **E)** GSEA analysis reveals the biological processes **(D)** and cancer hallmarks **(E)** enriched in the high-risk groups.

"aminoglycan metabolic process," and "extracellular matrix disassembly biological progresses" (**Figure 5D**). Moreover, the hallmarks of "epithelial–mesenchymal transition," "PI3K-AKT-mTOR signaling," "glycolysis", and "angiogenesis" also enriched in the high-risk cases (**Figure 5E**). The results indicated that the extracellular matrix related functions and mesenchymal phenotype could contribute to the TMZ-resistant of glioma.

DISCUSSION

Undoubtedly, *MGMT* promoter methylation status is critical for the chemotherapeutic management of glioma, especially for GBM (Hegi et al., 2005; Chai et al., 2019a; Chai et al., 2019e). However, whether TMZ should be withheld from patients with GBMs that lack *MGMT* promoter methylation is still under debate, and some of these patients indeed benefit from the treatment (Wick et al., 2014). Thus, it is critical to uncover novel biomarkers to identify TMZ-sensitive individuals with *MGMT* promoter unmethylated GBMs. In this study, we successfully developed a 31-CpG methylation signature which could identify the TMZ-sensitive GBMs in the *MGMT* promoter unmethylated GBMs from both the training and validation datasets, and OS of these TMZ-sensitive GBMs is similar to that of the *MGMT* promoter methylated GBMs after TMZ treatment in both two datasets. Considering the robust and reproducible nature of DNA methylation in the classification of brain tumors, this signature has great value in predicting the TMZ sensitivity of the GBMs that lack *MGMT* promoter methylation.

In this study, we systematically investigated 107 *MGMT* promoter unmethylated GBMs to obtain the TMZ therapeutic prognostic value of each of the CpGs that were included in the HM-27K BeadChip, and we identified that 3,565 CpGs are significantly associated with the OS of these GBMs. Previous studies have indicated that abnormal metabolism could alter the response of tumor cells to chemotherapy through inhibiting the activities of DNA repair enzymes (Gusyatiner and Hegi, 2018). DNA instability and DNA injury repair have been linked to the

chemo-resistance of cancer cells (Kanai et al., 2012; Roos et al., 2018; Zhao et al., 2018a; Ha Thi et al., 2019; Zhang et al., 2019). Here we also investigated the functions of genes corresponding to the 3,565 CpGs, and the results indicated that biological processes or pathways of mitochondrial translation, cell cycle and DNA repair could be involved in the TMZ-sensitivity of *MGMT* promoter unmethylated GBMs. Given that DNA proliferation rate is positively correlated to the sensitivity to chemotherapy (Li et al., 2017; Krell et al., 2019; Qiang et al., 2018), our finding supports that transcriptional activities of genes enriched in mitochondria, DNA injury and repair, and cell cycle processes could be important in the sensitivity of GBM cells to TMZ chemotherapy.

The extracellular matrix (ECM) components and their partners, including the glycosaminoglycans, glycoproteins, and proteoglycans, play a crucial role in the glioma invasion through promoting tumor cell migration and angiogenesis (Ferrer et al., 2018). The up-regulation of ECM partners, such as CD44, has been acknowledged as a marker for the "proneural–mesenchymal transition" of GBM cells (Yang et al., 2017). Here, we noticed that not only the hallmark of epithelial–mesenchymal transition but also ECM related biological processes and pathways were highly enriched in the *MGMT* unmethylated GBMs with the high-risk score, indicating that enhanced ECM activities could be involved in the TMZ-resistance of GBMs. This may be associated with the roles of ECM in regulating the extracellular microenvironments and intracellular signaling pathways (Wang et al., 2018). Chemokine (C-X-C motif) ligand 12 (CXCL12) and its receptors CXCR4 and CXCR7, which are stored in or attached to the ECM, are extremely important in forming a more invasive and resistant phenotype of glioma (Gatti et al., 2013; Zhao et al., 2018a). Recently, we also identified that the glycoprotein ADAMTS4, which is important for the upregulation of integrins, is also a novel immune-related biomarker for the primary GBM (Zhao et al., 2019). Transforming growth factor-beta (TGF-β), an ECM-bound bioactive factor, is involved in both the activation of NF-κB signaling and mesenchymal transition of GBM (Song et al., 2018; Batlle and Massague, 2019). Both of these two processes have been involved in the TMZ-resistance of GBM (Ming et al., 2017; Yang et al., 2017; Chai et al., 2019c; Chai et al., 2019d). All of these emphasize the value of the ECM in glioma TMZ sensitivity. Thus, the ECM and microenvironment should not be neglected in drug development, especially in developing an ideal *in vitro* drug screening model for glioma.

Chr 7 gain and Chr 10 loss is quite common in GBM (Bady et al., 2016; Chai et al., 2019a). Patients with high-grade gliomas harboring deletions of chromosomes 9p and 10q may benefit more from TMZ treatment (Wemmert et al., 2005), and the MGMT resides on chromosome 10q. Here, we also investigated the association between the risk signature and deletion of one copy of chromosome 10, and the results indicated that the predictive value of the risk signature was not affected by the status of Chr 7 gain and Chr 10 loss. This finding excludes the possibility that the predictive value of the risk signature may be caused by the unbalanced *MGMT* expression between GBM with or without Chr 7 gain and Chr 10 loss. Moreover, we have reported that chromosome 10/10q deletion does not significantly affect *MGMT* expression of GBM in the TCGA dataset (Chai et al., 2019a).

In conclusion, our findings reveal the predictive value of DNA methylation profiling in GBMs with an unmethylated *MGMT* promoter. The developed 31-CpG methylation signature could accurately predict the TMZ-sensitivity of *MGMT* promoter unmethylated GBMs. Though the risk signature still needs to be confirmed in future prospective studies with specific test kits, our current findings can promote our understanding of the roles of DNA methylation in GBMs with an unmethylated *MGMT* promoter and also offer a very promising TMZ-sensitivity predictive signature for these GBMs.

DATA AVAILABILITY STATEMENT

All methylation and clinical data used in this study were available from the TCGA database (http://cancergenome.nih.gov). Other information is available through contacting the corresponding authors.

AUTHOR CONTRIBUTIONS

R-CC conceived and designed the study, R-CC, Y-ZC, and Q-WW crafted the literature search, figures, and tables and were responsible for the writing and critical reading of the manuscript. K-NZ, J-JL, and HH contributed to the data analysis and the critical reading of the manuscript. FW, Y-QL, and Y-ZW supervised the analysis and contributed to the critical reading of the manuscript.

FUNDING

This work was supported by the National Key Research and Development Program of China (2018YFC0115604), the National Natural Science Foundation of China (81773208, 81802994), the National Natural Science Foundation of China (NSFC)/Research Grants Council (RGC) Joint Research Scheme (81761168038), Beijing Municipal Administration of Hospitals' Mission Plan (SML20180501, 2018.03-2022.02).

ACKNOWLEDGMENTS

The authors gratefully acknowledge contributions from the TCGA Network.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00910/full#supplementary-material

REFERENCES

Bady, P., Delorenzi, M., and Hegi, M. E. (2016). Sensitivity analysis of the MGMT-STP27 Model and impact of genetic and epigenetic context to predict the MGMT methylation status in gliomas and other tumors. *J. Mol. Diagn.* 18 (3), 350–361. doi: 10.1016/j.jmoldx.2015.11.009

Bady, P., Kurscheid, S., Delorenzi, M., Gorlia, T., van den Bent, M. J., Hoang-Xuan, K., et al. (2018). The DNA methylome of DDR genes and benefit from RT or TMZ in IDH mutant low-grade glioma treated in EORTC 22033. *Acta Neuropathol.* 135 (4), 601–615. doi: 10.1007/s00401-018-1810-6

Batlle, E., and Massague, J. (2019). Transforming growth factor-beta signaling in immunity and cancer. *Immunity* 50 (4), 924–940. doi: 10.1016/j.immuni.2019.03.024

Chai, R. C., Liu, Y. Q., Zhang, K. N., Wu, F., Zhao, Z., Wang, K. Y., et al. (2019a). A novel analytical model of MGMT methylation pyrosequencing offers improved predictive performance in patients with gliomas. *Mod. Pathol.* 32 (1), 4–15. doi: 10.1038/s41379-018-0143-2

Chai, R. C., Wang, N., Chang, Y. Z., Zhang, K. N., Li, J. J., Niu, J. J., et al. (2019b). Systematically profiling the expression of eIF3 subunits in glioma reveals the expression of eIF3i has prognostic value in IDH-mutant lower grade glioma. *Cancer Cell Int.* 19, 155. doi: 10.1186/s12935-019-0867-1

Chai, R. C., Wu, F., Wang, Q. X., Zhang, S., Zhang, K. N., Liu, Y. Q., et al. (2019c). m(6)A RNA methylation regulators contribute to malignant progression and have clinical prognostic impact in gliomas. *Aging* (Albany N.Y.) 11 (4), 1204–1225. doi: 10.18632/aging.101829

Chai, R. C., Zhang, K. N., Chang, Y. Z., Wu, F., Liu, Y. Q., Zhao, Z., et al. (2019d). Systematically characterize the clinical and biological significances of 1p19q genes in 1p/19q non-codeletion glioma. *Carcinogenesis*. doi: 10.1093/carcin/bgz102

Chai, R. C., Zhang, K. N., Liu, Y. Q., Wu, F., Zhao, Z., Wang, K. Y., et al. (2019e). Combinations of four or more CpGs methylation present equivalent predictive value for MGMT expression and temozolomide therapeutic prognosis in gliomas. *CNS Neurosci. Ther.* 25 (3), 314–322. doi: 10.1111/cns.13040

Chen, D., Song, Y., Zhang, F., Wang, X., Zhu, E., Zhang, X., et al. (2019a). Genome-wide analysis of lung adenocarcinoma identifies novel prognostic factors and a prognostic score. *Front. Genet.* 10, 493. doi: 10.3389/fgene.2019.00493

Chen, S., Pu, W., Guo, S., Jin, L., He, D., and Wang, J. (2019b). Genome-wide DNA methylation profiles reveal common epigenetic patterns of interferon-related genes in multiple autoimmune diseases. *Front.Genet.* 10 (223). doi: 10.3389/fgene.2019.00223

Dai, W., Feng, Y., Mo, S., Xiang, W., Li, Q., Wang, R., et al. (2018). Transcriptome profiling reveals an integrated mRNA-lncRNA signature with predictive value of early relapse in colon cancer. *Carcinogenesis* 39 (10), 1235–1244. doi: 10.1093/carcin/bgy087

Ferrer, V. P., Moura Neto, V., and Mentlein, R. (2018). Glioma infiltration and extracellular matrix: key players and modulators. *Glia* 66 (8), 1542–1565. doi: 10.1002/glia.23309

Gatti, M., Pattarozzi, A., Bajetto, A., Wurth, R., Daga, A., Fiaschi, P., et al. (2013). Inhibition of CXCL12/CXCR4 autocrine/paracrine loop reduces viability of human glioblastoma stem-like cells affecting self-renewal activity. *Toxicology* 314 (2-3), 209–220. doi: 10.1016/j.tox.2013.10.003

Gusyatiner, O., and Hegi, M. E. (2018). Glioma epigenetics: from subclassification to novel treatment options. *Semin. Cancer Biol.* 51, 50–58. doi: 10.1016/j.semcancer.2017.11.010

Ha Thi, H. T., Kim, H. Y., Lee, Y. J., Kim, S. J., and Hong, S. (2019). SMAD7 in keratinocytes promotes skin carcinogenesis by activating ATM-dependent DNA repair and an EGFR-mediated cell proliferation pathway. *Carcinogenesis* 40 (1), 112–120. doi: 10.1093/carcin/bgy121

Hegi, M. E., Diserens, A. C., Gorlia, T., Hamou, M. F., de Tribolet, N., Weller, M., et al. (2005). MGMT gene silencing and benefit from temozolomide in glioblastoma. *N. Engl. J. Med.* 352 (10), 997–1003. doi: 10.1056/NEJMoa043331

Jiang, T., Mao, Y., Ma, W., Mao, Q., You, Y., Yang, X., et al. (2016). CGCG clinical practice guidelines for the management of adult diffuse gliomas. *Cancer Lett.* 375 (2), 263–273. doi: 10.1016/j.canlet.2016.01.024

Kanai, R., Rabkin, S. D., Yip, S., Sgubin, D., Zaupa, C. M., Hirose, Y., et al. (2012). Oncolytic virus-mediated manipulation of DNA damage responses: synergy with chemotherapy in killing glioblastoma stem cells. *J. Natl. Cancer Inst.* 104 (1), 42–55. doi: 10.1093/jnci/djr509

Krell, A., Wolter, M., Stojcheva, N., Hertler, C., Liesenberg, F., Zapatka, M., et al. (2019). MiR-16-5p is frequently down-regulated in astrocytic gliomas and modulates glioma cell proliferation, apoptosis and response to cytotoxic therapy. *Neuropathol. Appl. Neurobiol.* 45 (5), 441–458 doi: 10.1111/nan.12532

Kumar, R., Liu, A. P. Y., Orr, B. A., Northcott, P. A., and Robinson, G. W. (2018). Advances in the classification of pediatric brain tumors through DNA methylation profiling: from research tool to frontline diagnostic. *Cancer* 124 (21), 4168–4180. doi: 10.1002/cncr.31583

Li, T., Li, S., Chen, D., Chen, B., Yu, T., Zhao, F., et al. (2017). Transcriptomic analyses of RNA-binding proteins reveal eIF3c promotes cell proliferation in hepatocellular carcinoma. *Cancer Sci.* 108 (5), 877–885. doi: 10.1111/cas.13209

Louis, D. N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W. K., et al. (2016). The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol.* 131 (6), 803–820. doi: 10.1007/s00401-016-1545-1

Ma, J., Hou, X., Li, M., Ren, H., Fang, S., Wang, X., et al. (2015). Genome-wide methylation profiling reveals new biomarkers for prognosis prediction of glioblastoma. *J. Cancer Res. Ther.* 11 Suppl 2, C212–C215. doi: 10.4103/0973-1482.168188

Ming, J., Sun, B., Li, Z., Lin, L., Meng, X., Han, B., et al. (2017). Aspirin inhibits the SHH/GLI1 signaling pathway and sensitizes malignant glioma cells to temozolomide therapy. *Aging (Albany N.Y.)* 9 (4), 1233–1247. doi: 10.18632/aging.101224

Pajtler, K. W., Witt, H., Sill, M., Jones, D. T., Hovestadt, V., Kratochwil, F., et al. (2015). Molecular classification of ependymal tumors across All CNS compartments, histopathological grades, and age groups. *Cancer Cell* 27 (5), 728–743. doi: 10.1016/j.ccell.2015.04.002

Qiang, Z., Jun-Jie, L., Hai, W., Hong, L., Bing-Xi, L., Lei, C., et al. (2018). TPD52L2 impacts proliferation, invasiveness and apoptosis of glioblastoma cells *via* modulation of wnt/beta-catenin/snail signaling. *Carcinogenesis* 39 (2), 214–224. doi: 10.1093/carcin/bgx125

Roos, W. P., Frohnapfel, L., Quiros, S., Ringel, F., and Kaina, B. (2018). XRCC3 contributes to temozolomide resistance of glioblastoma cells by promoting DNA double-strand break repair. *Cancer Lett.* 424, 119–126. doi: 10.1016/j.canlet.2018.03.025

Serao, N. V., Delfino, K. R., Southey, B. R., Beever, J. E., and Rodriguez-Zas, S. L. (2011). Cell cycle and aging, morphogenesis, and response to stimuli genes are individualized biomarkers of glioblastoma progression and survival. *BMC Med. Genomics* 4, 49. doi: 10.1186/1755-8794-4-49

Song, Y., Chen, Y., Li, Y., Lyu, X., Cui, J., Cheng, Y., et al. (2018). Metformin inhibits TGF-beta1-induced epithelial-to-mesenchymal transition-like process and stem-like properties in GBM *via* AKT/mTOR/ZEB1 pathway. *Oncotarget* 9 (6), 7023–7035. doi: 10.18632/oncotarget.23317

Sturm, D., Orr, B. A., Toprak, U. H., Hovestadt, V., Jones, D. T. W., Capper, D., et al. (2016). New brain tumor entities emerge from molecular classification of CNS-PNETs. *Cell* 164 (5), 1060–1072. doi: 10.1016/j.cell.2016.01.015

Sturm, D., Witt, H., Hovestadt, V., Khuong-Quang, D. A., Jones, D. T., Konermann, C., et al. (2012). Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell* 22 (4), 425–437. doi: 10.1016/j.ccr.2012.08.024

Wang, X., Li, X., Dai, X., Zhang, X., Zhang, J., Xu, T., et al. (2018). Coaxial extrusion bioprinted shell-core hydrogel microfibers mimic glioma microenvironment and enhance the drug resistance of cancer cells. *Colloids Surf. B Biointerfaces* 171, 291–299. doi: 10.1016/j.colsurfb.2018.07.042

Wemmert, S., Ketter, R., Rahnenfuhrer, J., Beerenwinkel, N., Strowitzki, M., Feiden, W., et al. (2005). Patients with high-grade gliomas harboring deletions of chromosomes 9p and 10q benefit from temozolomide treatment. *Neoplasia* 7 (10), 883–893. doi: 10.1593/neo.05307

Wick, W., Weller, M., van den Bent, M., Sanson, M., Weiler, M., von Deimling, A., et al. (2014). MGMT testing—the challenges for biomarker-based glioma treatment. *Nat. Rev. Neurol.* 10 (7), 372–385. doi: 10.1038/nrneurol.2014.100

Yang, F., Liu, X., Liu, Y., Liu, Y., Zhang, C., Wang, Z., et al. (2017). miR-181d/MALT1 regulatory axis attenuates mesenchymal phenotype through NF-kappaB pathways in glioblastoma. *Cancer Lett.* 396, 1–9. doi: 10.1016/j.canlet.2017.03.002

Zhang, Y., Li, J., Yi, K., Feng, J., Cong, Z., Wang, Z., et al. (2019). Elevated signature of a gene module coexpressed with CDC20 marks genomic instability in glioma. *Proc. Natl. Acad. Sci. U S A* 116 (14), 6975–6984. doi: 10.1073/pnas.1814060116

Zhao, K., Yao, Y., Luo, X., Lin, B., Huang, Y., Zhou, Y., et al. (2018a). LYG-202 inhibits activation of endothelial cells and angiogenesis through CXCL12/CXCR7 pathway in breast cancer. *Carcinogenesis* 39 (4), 588–600. doi: 10.1093/carcin/bgy007

Zhao, S., Leonardson, A., Geybels, M. S., McDaniel, A. S., Yu, M., Kolb, S., et al. (2018b). A five-CpG DNA methylation score to predict metastatic-lethal outcomes in men treated with radical prostatectomy for localized prostate cancer. *Prostate*. doi: 10.1002/pros.23667

Zhao, Z., Zhang, K. N., Chai, R. C., Wang, K. Y., Huang, R. Y., Li, G. Z., et al. (2019). ADAMTSL4, a secreted glycoprotein, is a novel immune-related biomarker for primary glioblastoma multiforme. *Dis. Markers* 2019, 1802620. doi: 10.1155/2019/1802620

Zhou, Z., Huang, R., Chai, R., Zhou, X., Hu, Z., Wang, W., et al. (2018). Identification of an energy metabolism-related signature associated with clinical prognosis in diffuse glioma. *Aging (Albany N.Y.)* 10 (11), 3185–3209. doi: 10.18632/aging.101625

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Chai, Chang, Wang, Zhang, Li, Huang, Wu, Liu and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 02 October 2019
doi: 10.3389/fgene.2019.00908

Comprehensive RNA-Seq Data Analysis Identifies Key mRNAs and lncRNAs in Atrial Fibrillation

Dong-Mei Wu[1,2†], Zheng-Kun Zhou[2†], Shao-Hua Fan[1,2†], Zi-Hui Zheng[3†], Xin Wen[1,2], Xin-Rui Han[1,2], Shan Wang[1,2], Yong-Jian Wang[1,2], Zi-Feng Zhang[1,2], Qun Shan[1,2], Meng-Qiu Li[1,2], Bin Hu[1,2], Jun Lu[1,2], Gui-Quan Chen[4]*, Xiao-Wu Hong[5]* and Yuan-Lin Zheng[1,2]**

[1] Key Laboratory for Biotechnology on Medicinal Plants of Jiangsu Province, School of Life Science, Jiangsu Normal University, Xuzhou, China, [2] College of Health Sciences, Jiangsu Normal University, Xuzhou, China, [3] State Key Laboratory Cultivation Base For TCM Quality and Efficacy, School of Medicine and Life Science, Nanjing University of Chinese Medicine, Nanjing, China, [4] State Key Laboratory of Pharmaceutical Biotechnology, MOE Key Laboratory of Model Animal for Disease Study, Model Animal Research Center, Nanjing University, Nanjing, China, [5] Department of Immunology, School of Basic Medical Sciences, Fudan University, Shanghai, China

OPEN ACCESS

Edited by:
Kyoko Yokomori,
University of California,
United States

Reviewed by:
Apiwat Mutirangura,
Chulalongkorn University,
Thailand
Abhijit Shukla,
Memorial Sloan Kettering
Cancer Center,
United States

***Correspondence:**
Jun Lu
lu-jun75@163.com
Gui-Quan Chen
chenguiquan@nju.edu.cn
Xiao-Wu Hong
xiaowuhong@fudan.edu.cn
Yuan-Lin Zheng
ylzheng@jsnu.edu.cn

†These authors share first authorship

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: 12 May 2019
Accepted: 28 August 2019
Published: 02 October 2019

Citation:
Wu D-M, Zhou Z-K, Fan S-H, Zheng Z-H, Wen X, Han X-R, Wang S, Wang Y-J, Zhang Z-F, Shan Q, Li M-Q, Hu B, Lu J, Chen G-Q, Hong X-W and Zheng Y-L (2019) Comprehensive RNA-Seq Data Analysis Identifies Key mRNAs and lncRNAs in Atrial Fibrillation. Front. Genet. 10:908. doi: 10.3389/fgene.2019.00908

Long non-coding RNAs (lncRNAs) are an emerging class of RNA species that may play a critical regulatory role in gene expression. However, the association between lncRNAs and atrial fibrillation (AF) is still not fully understood. In this study, we used RNA sequencing data to identify and quantify the both protein coding genes (PCGs) and lncRNAs. The high enrichment of these up-regulated genes in biological functions concerning response to virus and inflammatory response suggested that chronic viral infection may lead to activated inflammatory pathways, thereby alter the electrophysiology, structure, and autonomic remodeling of the atria. In contrast, the downregulated GO terms were related to the response to saccharides. To identify key lncRNAs involved in AF, we predicted lncRNAs regulating expression of the adjacent PCGs, and characterized biological function of the dysregulated lncRNAs. We found that two lncRNAs, ETF1P2, and AP001053.11, could interact with protein-coding genes (PCGs), which were implicated in AF. In conclusion, we identified key PCGs and lncRNAs, which may be implicated in AF, which not only improves our understanding of the roles of lncRNAs in AF, but also provides potentially functional lncRNAs for AF researchers.

Keywords: long non-coding RNAs, atrial fibrillation, RNA-Seq, genes, protein coding genes

INTRODUCTION

Atrial fibrillation (AF), one of the most common serious arrhythmia worldwide, whose extreme complications such as heart failure and embolic stroke are often of high risks and associated with increasing morbidity and mortality (Conen et al., 2011). Atrial remodeling, both electrical and structural, are important characteristics in AF (Li et al., 2017; Allessie et al., 2002). AF could bring permanent changes such as enlarged left and right atrial size. Moreover, increasing left atrial volume has been stated as a risk factor of cardioembolic stroke, and it is critical to interpret the mechanism behind this to improve the stroke prevention strategy.

The etiology of AF has not been fully elucidated as a varying range of factors would contribute to AF, such as family history, unhealthy lifestyle, high blood pressure and other diseases (Shi et al., 2013). With the development of next-generation sequencing technologies, non-coding

RNAs (ncRNAs) emerged as the epicenter for researchers to further explore the genetic cause behind AF. ncRNAs, which can be subdivided into small ncRNAs (< 200 nt) and long ncRNAs (lncRNAs), are not translated in proteins, but some of them are capable of regulating various cellular processes such as the expression of certain genes. Evidences have verified that many lncRNAs, often generated from transcriptional units, play a critical role in several cardiovascular diseases (Su et al., 2018), and it is of great importance to survey how they function in AF and how they are connected with atrial remodeling. Several researches are conducted to explore how lncRNAs acted as regulators in atrial electrical remodeling, revealing that TCONS_00075467 may help decrease AF vulnerability through suppressing the electrical remodeling (Li et al., 2017). Recent reports have also unveiled that lncRNAs can act as modulators of miRNA levels in various cardiac diseases (Greco et al., 2018). Also, inflammation and AF are confirmed to have a close relationship. Abundant inflammatory markers and higher ratios of neutrophil and lymphocyte are often observed in patients with AF (Hu et al., 2015), and AF subsequently triggers more inflammatory response, which in turns results in worsened conditions. Exploring active lncRNAs in inflammation would shed light on the prevention, diagnosis, and therapeutic strategies of AF, and help elucidate underlying mechanisms.

In the present study, we identified differentially expressed lncRNAs and mRNAs in patients with AF and predicted lncRNA function in a co-expression-based manner. Prediction of cis-acting lncRNAs and functional annotation of dysregulated lncRNAs screened out some critical lncRNAs implicated in AF, which not only improves our understanding of the roles of lncRNAs in AF, but also provides potentially functional lncRNAs for AF researchers. In addition, as evidence proves that a variety of inflammation-associated cytokines and chemokines are involved in the pathogenesis of AF (Schnabel et al., 2009), we further investigate whether our findings are related to cytokines and chemokines in certain aspects.

MATERIALS AND METHODS

Data Collection

We collected RNA sequencing data of 6 cases with AF and 6 controls from Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra) database (Leinonen et al., 2011) with an accession number SRP093226, which was provided by previous study (Yu et al., 2017). We uncompressed the SRA files by fastq-dump with the option *–split-files*, which generated two paired fastq files.

Read Mapping and Gene Expression Quantification

For each sample, we first mapped the RNA-seq reads to UCSC hg19 human reference genome (www.genome.ucsc.edu) using hisat2 (Kim et al., 2015), and then sorted the SAM files by samtools. With the gene annotation from GENCODE v19 (Harrow et al., 2012), the gene expression was estimated by the StringTie (Pertea et al., 2015) and ballgown pipeline. We considered genes with biotypes, including 'processed_transcript', 'pseudogene', 'lincRNA', '3prime_overlapping_ncrna', 'antisense', 'sense_intronic', and 'sense_overlapping', as lncRNAs.

Differential Expression Analysis

The FPKM-based expression was used to identify differentially expressed genes. The gene expression values were first transformed to $\log_2$ (FPKM + 1), and then tested for differential expression by t test. The differentially expressed genes were identified with the thresholds of *P*-value <0.05 and fold change >2 or <1/2.

Gene Ontology Enrichment Analysis

The Gene Ontology (GO) enrichment analysis was implemented in R with clusterProfiler package (Yu et al., 2012), which used overrepresentation enrichment analysis (ORA) to identify enriched GO terms. The GO terms were deemed to be significantly enriched if the adjusted $P < 0.05$ and the gene count in each GO term was more than 3.

Functional Annotation of lncRNAs

The biological function of lncRNAs was annotated by overrepresentation enrichment analysis (ORA) of co-expressed protein-coding genes (PCGs). The PCGs were defined as co-expressed genes with a given lncRNA if the $P < 0.0001$ for the correlation coefficient test.

Identification of Cis-Acting lncRNAs

As the lncRNAs may regulate the expression levels of the corresponding adjacent PCGs by cis-acting manner, the cis-acting lncRNAs were identified if the lncRNA and its adjacent PCGs (within one million base pairs) exhibit highly correlated expression (Pearson correlation coefficient > 0.5 or < −0.5).

RESULTS

RNA Sequencing Method Reveals Diverse RNAs in Both AF and Control Groups

The analysis of sequencing data with 6 AF patients and 6 controls allowed us to identify 15,147 genes in total (FPKM > 1 in at least one sample), which consisted of 29 RNA categories, including protein-coding genes (PCGs), pseudogenes, antisense RNAs and etc. (**Figure 1A**). The PCGs, pseudogenes, and antisense RNAs accounted for about 90% of the total identified RNAs. For each RNA category, the number of genes in AF was not observed to be higher or lower than that in control (Wilcoxon rank-sum test, $P < 0.05$). In addition, we also considered genes with seven specific biotypes as lncRNAs (See Material and Methods). Given a threshold of FPKM >1 in at least 25% samples (n = 3), we identified 9,233 PCGs, 2,213 lncRNAs, and 961 other ncRNA genes, which were then used for downstream analysis (**Figure 1B**).

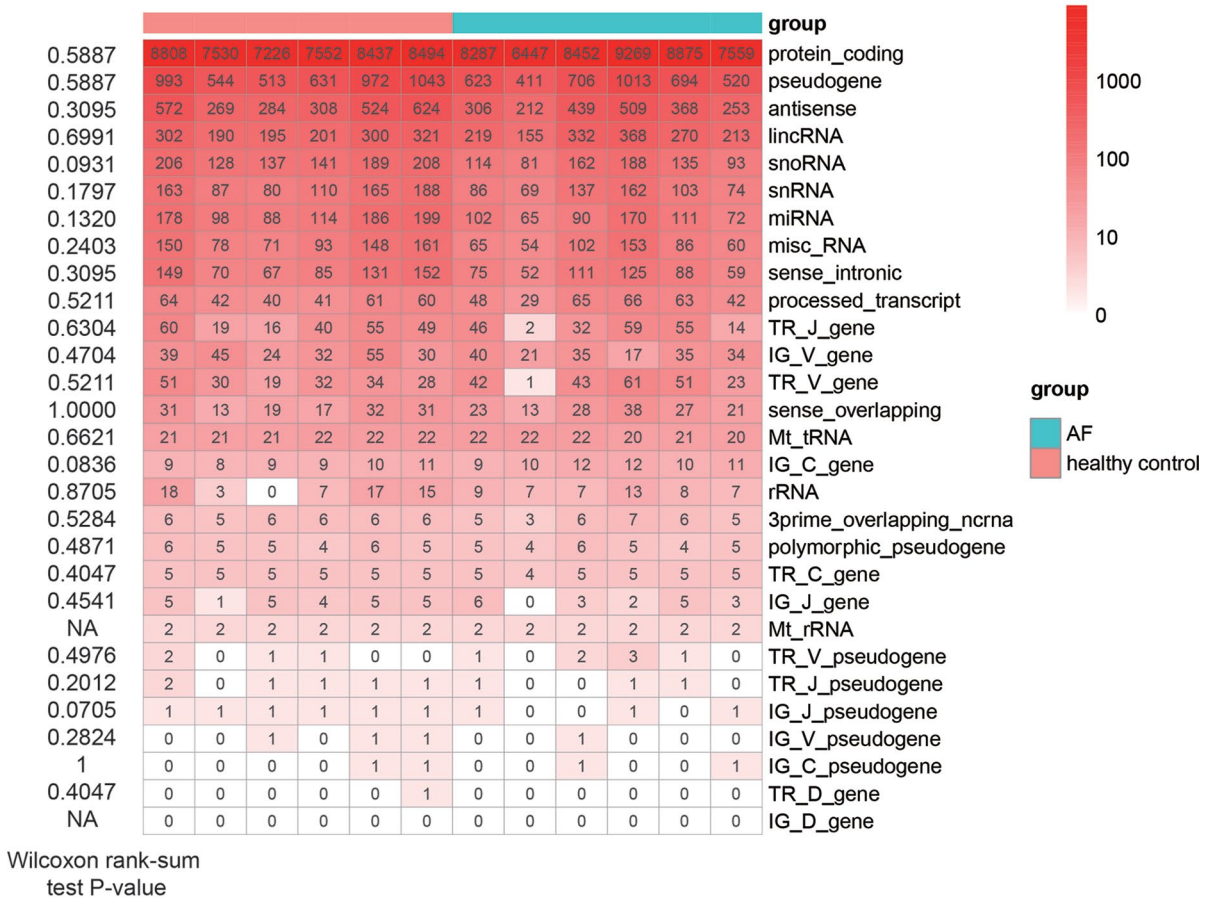


FIGURE 1 | The overview of genes identified by RNA sequencing method. **(A)** The number of genes identified in each sample (FPKM > 1). The Wilcoxon rank-sum test *P*-values that compare the gene counts in AF with those in normal controls are listed on the left of the heat-map. **(B)** The pie chart displays the number of genes, including PCGs, lncRNAs, and other ncRNAs.

Identification of Dysregulated mRNAs and lncRNAs in AF

Differential expression analysis was conducted to identify dysregulated genes in AF using the gene expression profiles. Specifically, a total of 946 genes, including 327 up- and 619 down-regulated genes, were differentially expressed in AF as compared with the healthy controls (t-test, $P < 0.05$ and fold change >2 or $<^1/_2$, **Figure 2A**, **Supplementary Table S1**). To investigate the distinction of the dysregulated genes between AF and healthy controls, we performed hierarchical clustering analysis of the dysregulated gene expression profiles, and found that the samples with AF could be clearly distinguished from the healthy controls (**Figure 2B**), indicating that gene expression profiles between AF and healthy controls had marked differences. Among the dysregulated genes, the proportion of PCGs was significantly higher in the up-regulated genes than in the down-regulated genes (**Figure 2C**, 249/327 vs. 156/619, proportion test, $P < 0.0001$). In contrast, the proportion of lncRNAs was observed higher in the down-regulated genes than in the up-regulated genes (**Figure 2C**, 315/619 vs. 63/327, proportion test, $P < 0.0001$).

Furthermore, we selected the top-five up- and down-regulated genes (**Figure 2D**), and found that the top-five up-regulated genes were all PCGs, while only one down-regulated gene encoded protein. Notably, three of the top-five upregulated genes, *GIMAP8*, *TNFAIP8L2*, and *RNASEL*, were involved in inflammatory response, suggesting that the dysregulation of inflammatory response may be the an important indicator for AF. On the other hand, the *PTX3* had an antiangiogenic role, and its downregulation may lead to enhanced angiogenesis. In addition, lncRNAs, HOTAIRM1, RP11-262H14.1, and RP11-84A19.4, have been reported to be dysregulated in AF by previous studies (Yu et al., 2017; Qian et al., 2019). These results indicated that differential expression analysis could uncover some key genes in AF.

Gene Ontology-Based Enrichment Analysis of the Dysregulated Genes

To investigate some key biological functions involved in AF, we performed overrepresentation enrichment analysis (ORA) on the up- and down-regulated genes, respectively. We found that the up-regulated genes were highly enriched in biological functions related to response to virus, such as defense response to virus, response to virus, viral life cycle, regulation of viral process, and regulation of viral life cycle, and related to inflammatory response, such as positive regulation of I-kappaB kinase/NF-kappaB signaling, and regulation of chemotaxis (**Figure 3A**, adjusted $P < 0.05$). These results indicated that chronic viral infection may lead to activated inflammatory pathways, thereby alter the electrophysiology, structure, and autonomic remodeling of the atria (Chiang et al., 2013).

Among the down-regulated GO terms, biological functions related to the response to saccharides (**Figure 3B**), such as response to lipopolysaccharide, response to glucose, response to hexose, response to monosaccharide, and response to carbohydrate were significantly enriched by the down-regulated genes. Notably, the weakened response to glucose in blood may reduce the insulin level, thereby lead to hyperglycemia, which further demonstrate the close association between hyperglycemia and AF (Rigalleau et al., 2002).

Prediction of lncRNAs Regulating Expression of the Adjacent PCGs

It has been widely recognized that lncRNAs could regulate the expression of the adjacent PCGs by cis-acting manner (Kornienko et al., 2013). To identify these cis-acting lncRNAs, we first searched the adjacent dysregulated PCGs within one million base pairs for each dysregulated lncRNA, and found 187 lncRNA-PCG pairs. The expression levels between the lncRNA and its corresponding PCGs were highly correlated (**Figure 4A**). Particularly, the expression levels of about half of the lncRNA-PCG pairs were negatively correlated, indicating that the lncRNAs may suppress the expression of the adjacent PCGs. With a threshold at Pearson correlation coefficient > 0.5 or < −0.5, we identified 71 lncRNA-PCG pairs, composed of 58 PCGs and 63 lncRNAs, with potential regulatory relationship (**Supplementary Table S2**). Among the cis-acting lncRNAs, we found that pseudogene (46%), antisense (34%), and lincRNA (11%) were the major categories (**Figure 4B**).

To identify key lncRNAs involved in regulating gene expression, we selected seven lncRNAs, *AL021707.2*, *CTD-2622I13.3*, *ETF1P2*, *RP11-4K3__A.5*, *RP11-95J11.1*, *ZNF137P*, and *H2AFZP1*, that regulated multiple PCGs. Notably, we found that *ETF1P2*, a pseudogene locating within 7q36, was negatively correlated with two adjacent PCGs with similar functions, *GIMAP2* and *GIMAP4* (**Figure 4C**), which participated in the regulation of T helper cell differentiation (Filen et al., 2009), indicating that the pseudogene *ETF1P2* may be the upstream regulator of T helper cell differentiation.

Functional Annotation of the Dysregulated lncRNAs by Co-Expressed PCGs

As co-expressed genes are more likely to be co-regulated, sharing similar functions, or involved in similar biological processes (Stuart et al., 2003), we predicted the function of lncRNAs by performing overrepresentation enrichment analysis on the co-expressed PCGs to identify enriched GO terms (**Supplementary Table S3**). We found that a large number of lncRNAs (n > 20) had the biological functions termed transcription corepressor activity, proximal promoter sequence-specific DNA binding, and RNA polymerase II proximal promoter sequence-specific DNA binding (**Figure 5A**). Specifically, 38 lncRNAs were characterized with transcription corepressor activity, and highly correlated with five PCGs (Pearson correlation coefficient > 0.5), *SF1*, *MNT*, *NR1D1*, *SKI*, *DNAJB1*, and *YY1*, with the same GO term (**Figures 5B**, **C**). In addition, we also found that one lncRNA, *AP001053.11*, may participate in inflammatory response related GO terms, such as chemokine binding, chemokine receptor activity, cytokine binding, and cytokine receptor activity (**Figure 5D**). Notably, three chemokine receptor, *CX3CR1*, *CCR2*, and *CCR5*, were highly correlated with *AP001053.11* (Pearson correlation coefficient > 0.9), further suggesting a critical role of *AP001053.11* in regulation of chemokine receptor activity.

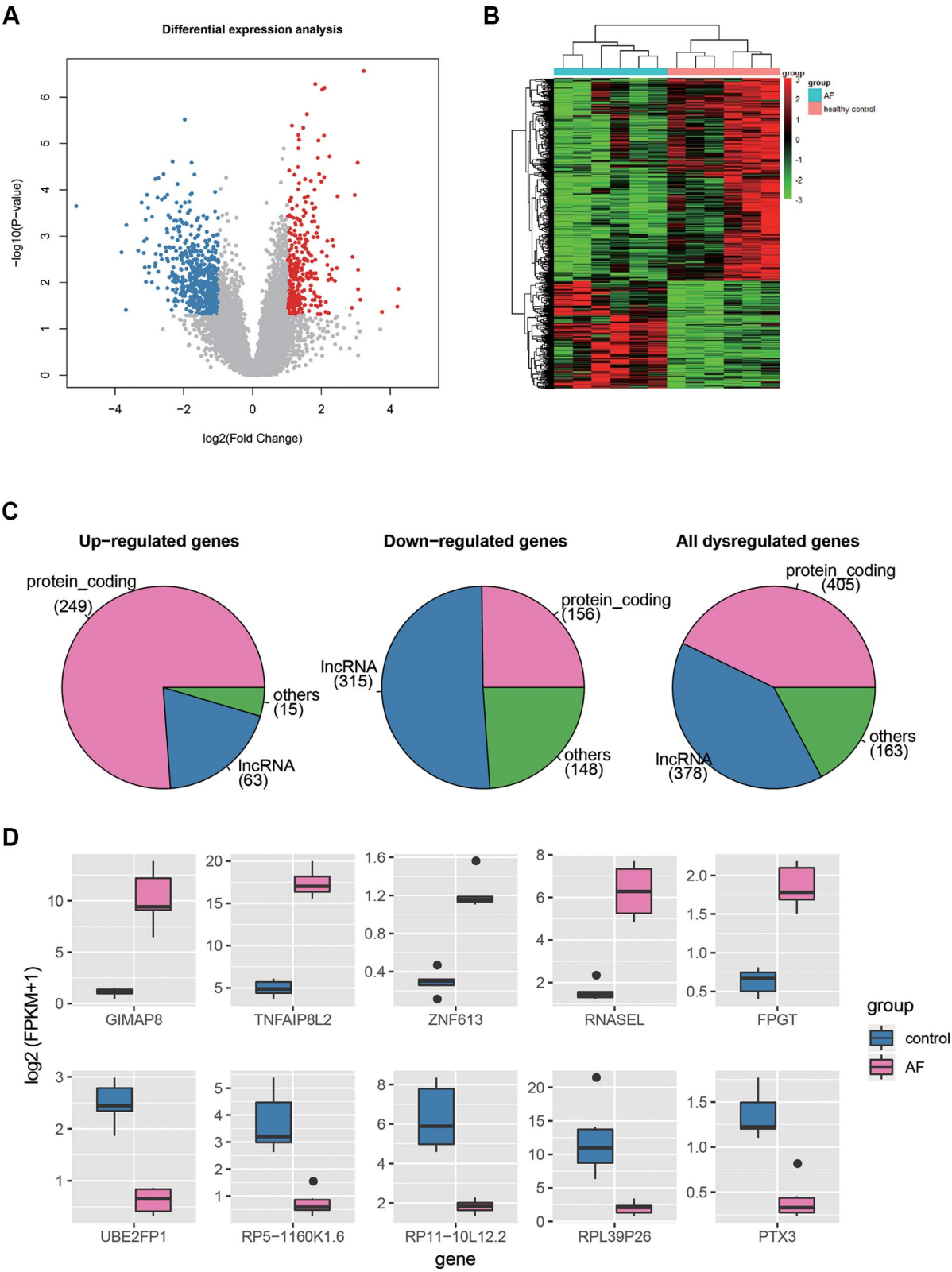


FIGURE 2 | The differentially expressed genes in AF. **(A)** The volcano plot displays the up-regulated (red dots) and down-regulated (blue dots) genes. **(B)** The heat-map shows the scaled gene expression of dysregulated genes. **(C)** The number of PCGs, lncRNAs, and other ncRNAs in up-regulated, down-regulated, and all dysregulated genes. **(D)** The expression levels of the top-five up-regulated and down-regulated genes in AF and control.

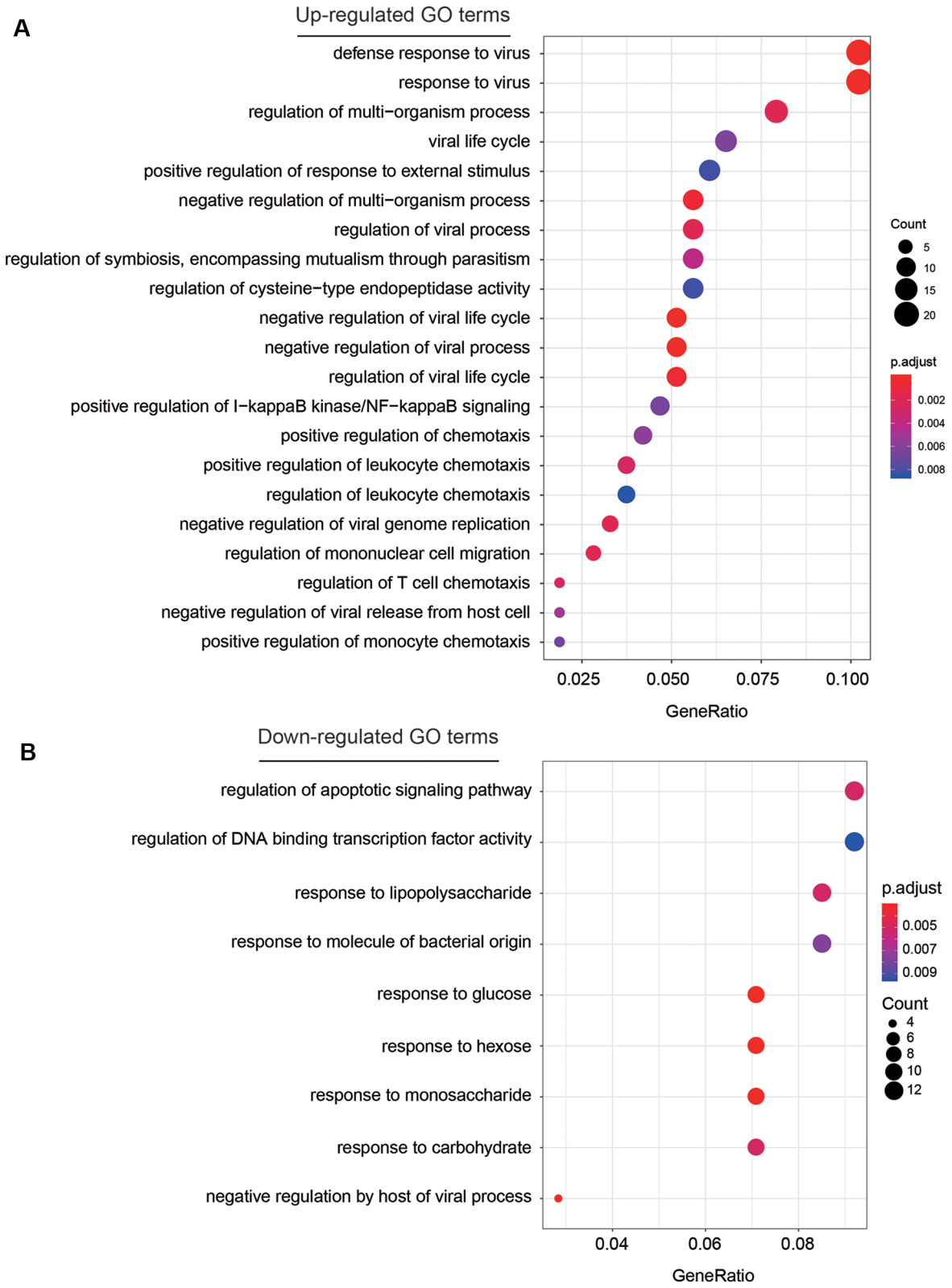


FIGURE 3 | The GO terms enriched by dysregulated genes. The GO terms enriched by up-regulated and down-regulated genes are displayed in **(A)** and **(B)**, respectively. The more the gene count, the larger size the circle.

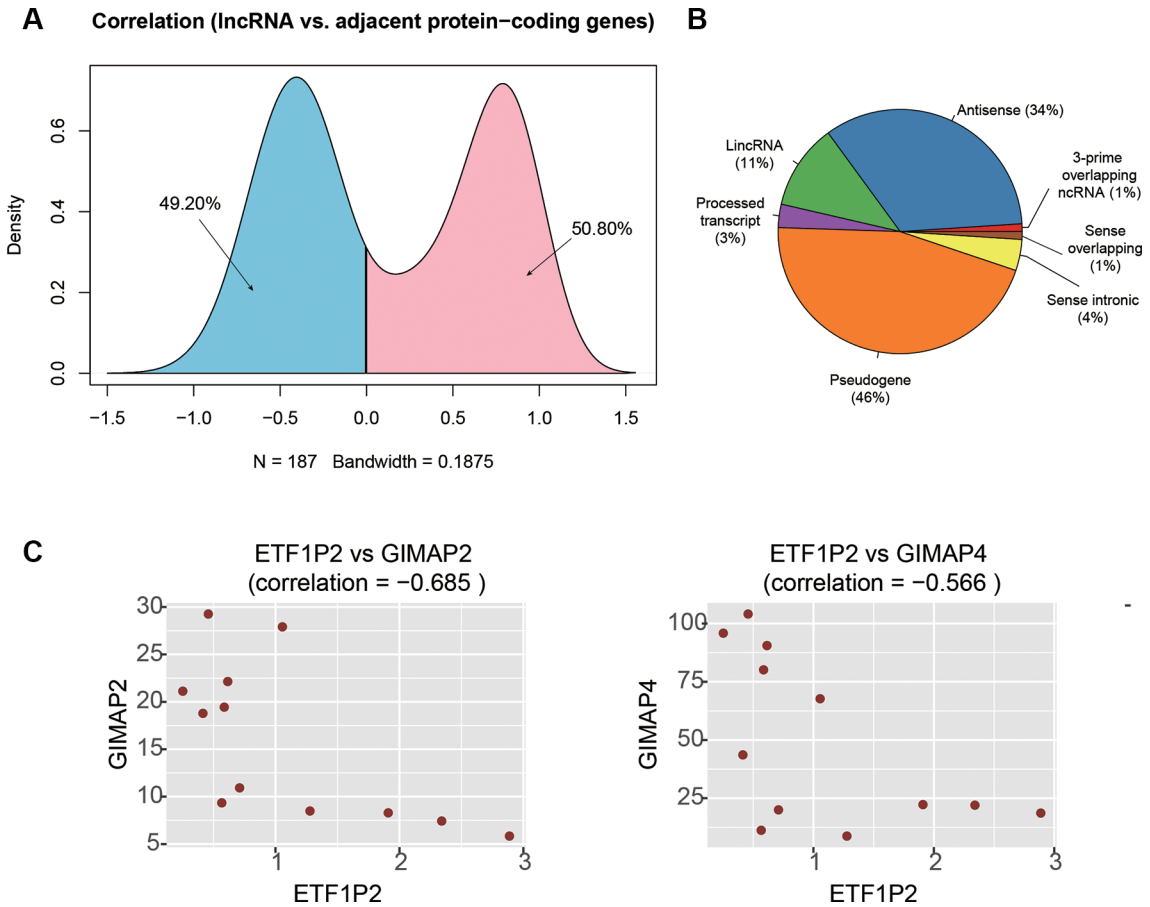


FIGURE 4 | The cis-acting lncRNA candidates involved in AF. **(A)** The density of the correlation coefficients between lncRNAs and the corresponding adjacent PCGs. **(B)** The distribution of RNA biotypes for the cis-acting lnRNA candidates. **(C)** The correlation coefficients between ETF1P2 and GIMAP2, and between ETF1P2 and GIMAP4.

DISCUSSION

LncRNAs are an emerging class of RNA species that may play a critical regulatory role in gene expression. LncRNAs can serve as diagnostic biomarkers or therapeutic targets for many diseases (Ishii et al., 2006; Chen et al., 2008; Chubb et al., 2008). However, the association between lncRNAs and AF is still not fully understood.

In this study, we used RNA sequencing data to identify and quantify the both PCGs and lncRNAs, and conducted differential expression analysis to identify dysregulated genes in AF. Specifically, a total of 946 genes, including 327 up- and 619 down-regulated genes, were differentially expressed in AF as compared with the healthy controls (t-test, P <0.05 and fold change >2 or $<^1/_2$, **Figure 2A**, **Supplementary Table S1**). The hierarchical clustering analysis of those dysregulated gene expression profiles showed that the samples with AF could be clearly distinguished from the healthy controls (**Figure 2B**), indicating that gene expression profiles between AF and healthy controls had marked differences. Furthermore, we found that three of the top-five upregulated genes, *GIMAP8*, *TNFAIP8L2*, and *RNASEL*, were involved in inflammatory response, which was in accordance with the conclusion that the infiltration of immune cells and proteins that mediate inflammatory response in cardiac tissue and circulatory processes is associated with AF by previous studies (Yamashita et al., 2010; Harada et al., 2015). On the other hand, the *PTX3* had an antiangiogenic role, and its downregulation may lead to enhanced angiogenesis, which has been reported to be associated with AF (Berntsson et al., 2019).

To investigate some key biological functions involved in AF, we performed ORA on the dysregulated genes. The significant enrichment of these up-regulated genes in biological functions related to response to virus and inflammatory response suggested that chronic viral infection may lead to activated inflammatory pathways, thereby alter the electrophysiology, structure, and autonomic remodeling of the atria (Chiang et al., 2013). In contrast, the downregulated GO terms were related to the response to saccharides (**Figure 3B**), which gave us a hint that the weakened response to glucose in blood may reduce the insulin level, thereby lead to hyperglycemia as previous study reported (Rigalleau et al., 2002).

To identify key lncRNAs involved in AF, we predicted lncRNA-regulated expression of the adjacent PCGs, and characterized

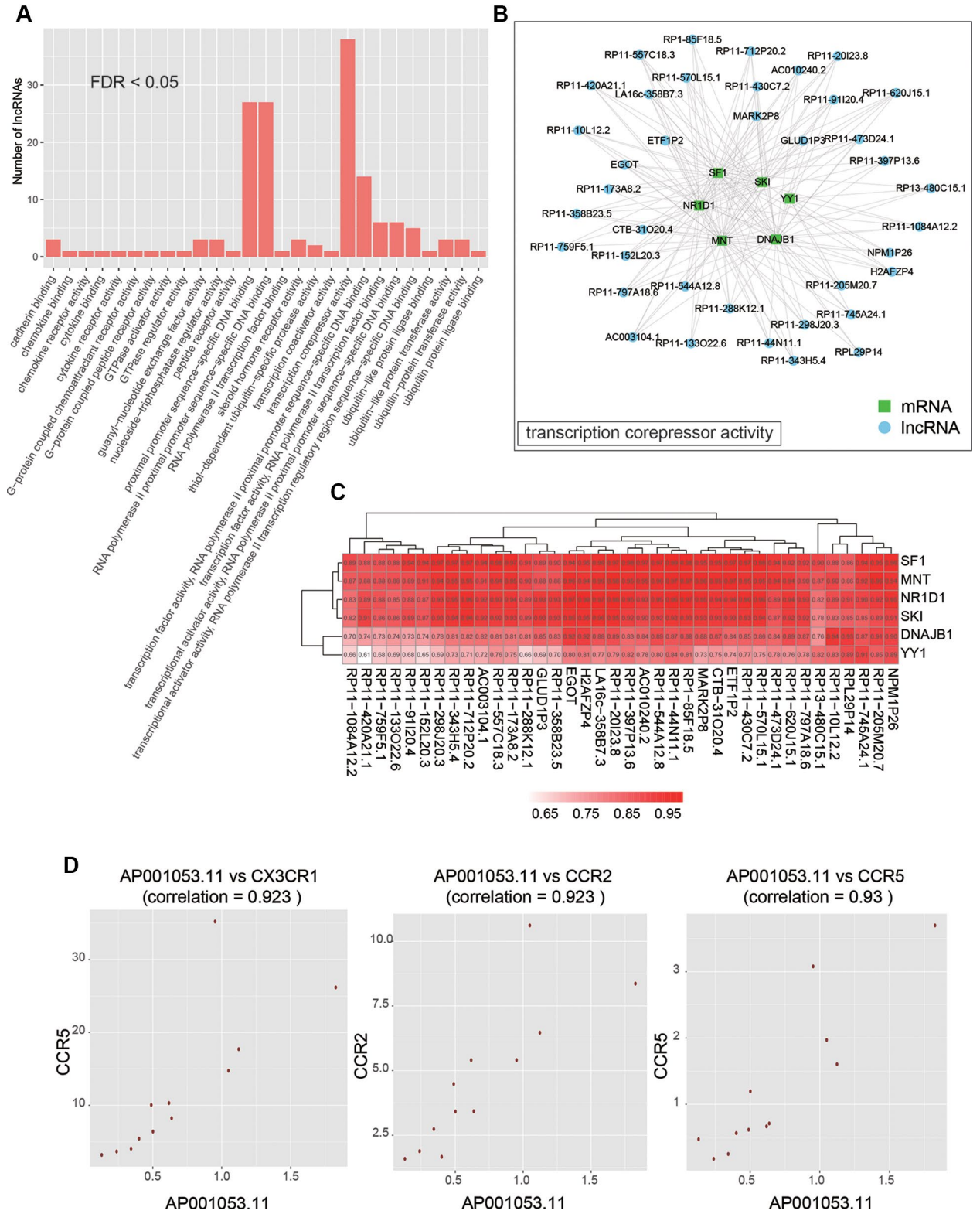


FIGURE 5 | The functional annotation of lncRNAs by co-expressed PCGs. **(A)** The overview of the GO terms for the annotation of dysregulated lncRNAs (FDR < 0.05). **(B)** The PCGs and lncRNAs involved in transcription corepressor activity. **(C)** The correlation coefficients between PCGs and lncRNAs involved in transcription corepressor activity. **(D)** The scatter plots display the correlation between AP001053.11 and one of three chemokine receptors, including CX3CR1, CCR2, and CCR5.

biological function of the dysregulated lncRNAs. We found that *ETF1P2*, a pseudogene locating within 7q36, was negatively correlated with two adjacent PCGs with similar functions, *GIMAP2* and *GIMAP4* (**Figure 4C**), which participated in regulation of T helper cell differentiation (Filen et al., 2009), indicating that the pseudogene *ETF1P2* may be an upstream regulator of T helper cell differentiation. Moreover, we also found that one lncRNA, *AP001053.11*, may participate in inflammatory-response-related GO terms by co-expression-based functional annotation. Notably, three chemokine receptor, *CX3CR1*, *CCR2*, and *CCR5*, were highly correlated with *AP001053.11* (Pearson correlation coefficient > 0.9), further suggesting that *AP001053.11* may be implicated in AF *via* the regulation of chemokine receptor activity.

In addition, there are also some limitations in this study. Firstly, more samples were needed to support our findings about the key lncRNAs. We will collect more samples with AF and healthy donors in the near future, which can overcome this limitation. Secondly, some experimental validation would be required for future verification of the functional lncRNAs. We hope to conduct further research with larger sample size, experimental validation and improved methodology for data analysis in the near future.

In conclusion, we have identified key PCGs and lncRNAs, which may be implicated in AF, which not only improves our understanding of the roles of lncRNAs in AF, but also provides potentially functional lncRNAs for AF researchers.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the SRP093226.

AUTHOR CONTRIBUTIONS

Conception and design: D-MW, JL, G-QC, X-WH, Y-LZ; Administrative support: JL, G-QC, Y-LZ; Provision of study materials or patients: D-MW, S-HF, Z-HZ; Collection and assembly of data: D-MW, S-HF, XW, X-RH, SW, Y-JW, Z-FZ; Data analysis and interpretation: D-MW, QS, M-QL; Manuscript writing: All authors; Final approval of manuscript: All authors.

FUNDING

This work was supported by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD); the 2016 "333 Project" Award of Jiangsu Province, the 2013 "Qinglan Project" of the Young and Middle-aged Academic Leader of Jiangsu College and University, the National Natural Science Foundation of China (81871249, 81571055, 81400902, 81271225, 81171012, and 30950031), the Major Fundamental Research Program of the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (13KJA180001), and grants from the Cultivate National Science Fund for Distinguished Young Scholars of Jiangsu Normal University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00908/full#supplementary-material

REFERENCES

Allessie, M., Ausma, J., and Schotten, U. (2002). Electrical, contractile and structural remodeling during atrial fibrillation. *Cardiovasc. Res.* 54 (2), 230–246. doi: 10.1016/S0008-6363(02)00258-4

Berntsson, J., Smith, J. G., Johnson, L. S. B., Soderholm, M., Borne, Y., Melander, O., et al. (2019). Increased vascular endothelial growth factor D is associated with atrial fibrillation and ischaemic stroke. *Heart* 105 (7), 553–558. doi: 10.1136/heartjnl-2018-313684

Chen, W. L., Lin, J. W., Huang, H. J., Wang, S. M., Su, M. T., Lee-Chen, G. J., et al. (2008). SCA8 mRNA expression suggests an antisense regulation of KLHL1 and correlates to SCA8 pathology. *Brain Res.* 1233, 176–184. doi: 10.1016/j.brainres.2008.07.096

Chiang, C. H., Huang, C. C., Chan, W. L., Huang, P. H., Chen, Y. C., Chen, T. J., et al. (2013). Herpes simplex virus infection and risk of atrial fibrillation: a nationwide study. *Int. J. Cardiol.* 164 (2), 201–204. doi: 10.1016/j.ijcard.2011.06.126

Chubb, J. E., Bradshaw, N. J., Soares, D. C., Porteous, D. J., and Millar, J. K. (2008). The DISC locus in psychiatric illness. *Mol. Psychiatry* 13 (1), 36–64. doi: 10.1038/sj.mp.4002106

Conen, D., Chae, C. U., Glynn, R. J., Tedrow, U. B., Everett, B. M., Buring, J. E., et al. (2011). Risk of death and cardiovascular events in initially healthy women with new-onset atrial fibrillation. *JAMA* 305 (20), 2080–2087. doi: 10.1001/jama.2011.659

Filen, J. J., Filen, S., Moulder, R., Tuomela, S., Ahlfors, H., West, A., et al. (2009). Quantitative proteomics reveals GIMAP family proteins 1 and 4 to be differentially regulated during human T helper cell differentiation. *Mol. Cell Proteomics* 8 (1), 32–44. doi: 10.1074/mcp.M800139-MCP200

Greco, S., Salgado Somoza, A., Devaux, Y., and Martelli, F. (2018). Long noncoding RNAs and cardiac disease. *Antioxid. Redox Signal.* 29 (9), 880–901. doi: 10.1089/ars.2017.7126

Harada, M., Van Wagoner, D. R., and Nattel, S. (2015). Role of inflammation in atrial fibrillation pathophysiology and management. *Circ. J.* 79 (3), 495–502. doi: 10.1253/circj.CJ-15-0138

Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* 22 (9), 1760–1774. doi: 10.1101/gr.135350.111

Hu, Y. F., Chen, Y. J., Lin, Y. J., and Chen, S. A. (2015). Inflammation and the pathogenesis of atrial fibrillation. *Nat. Rev. Cardiol.* 12 (4), 230–243. doi: 10.1038/nrcardio.2015.2

Ishii, N., Ozaki, K., Sato, H., Mizuno, H., Saito, S., Takahashi, A., et al. (2006). Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction. *J. Hum. Genet.* 51 (12), 1087–1099. doi: 10.1007/s10038-006-0070-9

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12 (4), 357–360. doi: 10.1038/nmeth.3317

Kornienko, A. E., Guenzl, P. M., Barlow, D. P., and Pauler, F. M. (2013). Gene regulation by the act of long non-coding RNA transcription. *BMC Biol.* 11, 59. doi: 10.1186/1741-7007-11-59

Leinonen, R., Sugawara, H., Shumway, M., and International nucleotide sequence database C (2011). The sequence read archive. *Nucleic Acids Res.* 39 (Database issue), D19–D21. doi: 10.1093/nar/gkq1019

Li, Z., Wang, X., Wang, W., Du, J., Wei, J., Zhang, Y., et al. (2017). Altered long non-coding RNA expression profile in rabbit atria with atrial fibrillation: TCONS_00075467 modulates atrial electrical remodeling by sponging

miR-328 to regulate CACNA1C. *J. Mol. Cell Cardiol.* 108, 73–85. doi: 10.1016/j.yjmcc.2017.05.009

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33 (3), 290–295. doi: 10.1038/nbt.3122

Qian, C., Li, H., Chang, D., Wei, B., and Wang, Y. (2019). Identification of functional lncRNAs in atrial fibrillation by integrative analysis of the lncRNA-mRNA network based on competing endogenous RNAs hypothesis. *J. Cell Physiol.* 234 (7), 11620–11630. doi: 10.1002/jcp.27819

Rigalleau, V., Baillet, L., Hocini, M., and Gin, H. (2002). Atrial fibrillation can cause major hyperglycemia. *Diabetes Metab.* 28 (3), 239–240.

Schnabel, R. B., Larson, M. G., Yamamoto, J. F., Kathiresan, S., Rong, J., Levy, D., et al. (2009). Relation of multiple inflammatory biomarkers to incident atrial fibrillation. *Am. J. Cardiol.* 104 (1), 92–96. doi: 10.1016/j.amjcard.2009.02.053

Shi, K. H., Tao, H., Yang, J. J., Wu, J. X., Xu, S. S., and Zhan, H. Y. (2013). Role of microRNAs in atrial fibrillation: new insights and perspectives. *Cell Signal.* 25 (11), 2079–2084. doi: 10.1016/j.cellsig.2013.06.009

Stuart, J. M., Segal, E., Koller, D., and Kim, S. K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science* 302 (5643), 249–255. doi: 10.1126/science.1087447

Su, Y., Li, L., Zhao, S., Yue, Y., and Yang, S. (2018). The long noncoding RNA expression profiles of paroxysmal atrial fibrillation identified by microarray analysis. *Gene* 642, 125–134. doi: 10.1016/j.gene.2017.11.025

Yamashita, T., Sekiguchi, A., Iwasaki, Y. K., Date, T., Sagara, K., Tanabe, H., et al. (2010). Recruitment of immune cells across atrial endocardium in human atrial fibrillation. *Circ. J.* 74 (2), 262–270. doi: 10.1253/circj.CJ-09-0644

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16 (5), 284–287. doi: 10.1089/omi.2011.0118

Yu, X. J., Zou, L. H., Jin, J. H., Xiao, F., Li, L., Liu, N., et al. (2017). Long noncoding RNAs and novel inflammatory genes determined by RNA sequencing in human lymphocytes are up-regulated in permanent atrial fibrillation. *Am. J. Transl. Res.* 9 (5), 2314–2326.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Wu, Zhou, Fan, Zheng, Wen, Han, Wang, Wang, Zhang, Shan, Li, Hu, Lu, Chen, Hong and Zheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

REVIEW
published: 08 October 2019
doi: 10.3389/fgene.2019.00962

Essential Role of Histone Replacement and Modifications in Male Fertility

Tong Wang[1†], *Hui Gao*[1†], *Wei Li*[1,2*] *and Chao Liu*[1,2*]

[1] State Key Laboratory of Stem Cell and Reproductive Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, China, [2] College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China

Spermiogenesis is a complex cellular differentiation process that the germ cells undergo a distinct morphological change, and the protamines replace the core histones to facilitate chromatin compaction in the sperm head. Recent studies show the essential roles of epigenetic events during the histone-to-protamine transition. Defects in either the replacement or the modification of histones might cause male infertility with azoospermia, oligospermia or teratozoospermia. Here, we summarize recent advances in our knowledge of how epigenetic regulators, such as histone variants, histone modification and their related chromatin remodelers, facilitate the histone-to-protamine transition during spermiogenesis. Understanding the molecular mechanism underlying the modification and replacement of histones during spermiogenesis will enable the identification of epigenetic biomarkers of male infertility, and shed light on potential therapies for these patients in the future.

Keywords: spermiogenesis, histone-to-protamine transition, histone variants, histone modification, male infertility

OPEN ACCESS

Edited by:
Kyoko Yokomori, University of California, United States

Reviewed by:
Alexander Kouzmenko, Tokiwa Foundation, Japan
Abhijit Shukla, Memorial Sloan Kettering Cancer Center, United States

***Correspondence:**
Wei Li
leways@ioz.ac.cn
Chao Liu
liuchao@ioz.ac.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *08 May 2019*
Accepted: *10 September 2019*
Published: *08 October 2019*

Citation:
Wang T, Gao H, Li W and Liu C (2019) Essential Role of Histone Replacement and Modifications in Male Fertility. Front. Genet. 10:962. doi: 10.3389/fgene.2019.00962

INTRODUCTION

Spermatogenesis is the process of male gamete production with successive cellular differentiation, which can be subdivided into spermatogonial mitosis, spermatocytic meiosis and spermiogenesis (Roosen-Runge, 1962; Hess and Renato De Franca, 2008). During spermatogenesis, SSC (spermatogonial stem cells) undergo self-renewal and differentiate into spermatogonia that perform meiosis to generate haploid germ cells and ensure the genetic diversity through meiotic recombination (Rathke et al., 2014; Bao and Bedford, 2016). Then, the haploid germ cells undergo spermiogenesis with a distinct morphological change and chromatin compaction in the sperm nuclei to prevent the paternal genome from mutagenesis and damage (Govin et al., 2004; Bao and Bedford, 2016). During the nuclear chromatin re-organization in spermiogenesis, the majority of the somatic histones are firstly replaced by testis-specific histone variants, and transition proteins (TPs) are subsequently incorporated in the nuclei of spermatids, protamines (PRMs) further replace TPs in the late spermatids to pack the genome into the highly condensed sperm nucleus (Rathke et al., 2014; Bao and Bedford, 2016). During the histone-to-protamine transition, the histone variants and specific histone modifications play essential roles by modulating the chromatin compaction and higher-order chromatin structure (**Table 1**) (Boskovic and Torres-Padilla, 2013; Bao and Bedford, 2016; Hada et al., 2017; Hao et al., 2019). Defects in either the replacement or the modification of histones might result in azoospermia, oligospermia or teratozoospermia, which leads to male infertility (**Table 2**). The focus of this review is on recent advances in our knowledge of how epigenetic regulators, such as histone variants, histone modification and their

TABLE 1 | The main histone variants and modifications during the histone-to-protamine transition.

Type	Histone	Name	Period	Function	Reference
Histone variants	H1	H1T	Spermatocytes to elongating spermatids	Maintain open chromatin configuration and no detectable phenotype in H1t-null testis	Delucia et al., 1994; Khadake and Rao, 1995; Drabent et al., 2000; Fantz et al., 2001
		H1T2	Round spermatids and elongating spermatids	Indispensable for the replacement of histones with protamines and chromatin condensation	Martianov et al., 2005; Tanaka et al., 2005
		HILS1	Elongating and elongated spermatids	Contribute to the open chromatin structure	Yan et al., 2003; Mishra et al., 2018
	H2A	TH2A	Spermatocytes to elongated spermatids	Contribute to the open chromatin structure and cooperate with TH2B to regulate TP2 incorporation	Padavattan et al., 2015; Shinagawa et al., 2015; Padavattan et al., 2017
		H2AL2	Elongating and elongated spermatids	Assemble open nucleosomes and allow TPs incorporation	Govin et al., 2007; Barral et al., 2017
		H2A.B	Spermatocytes to round spermatids	Destabilize chromatin and modulate the dynamics of H2AL2 removal and TP1 incorporation and	Soboleva et al., 2012; Soboleva et al., 2017; Anuar et al., 2019
	H2B	Th2B	Spermatocytes, round spermatids and elongating spermatids	Destabilize chromatin and regulate the TPs and PRMs incorporation	Meistrich et al., 1985; Montellier et al., 2013
	H3	H3.3	All types of germ cell	Contribute to the open chromatin structure, modulate TP1 removal and PRM1 incorporation	Bramlage et al., 1997; Couldrey et al., 1999; Van Der Heijden et al., 2007; Thakar et al., 2009; Chen et al., 2013; Tang et al., 2015
		H3T	Spermatocytes, round spermatids and elongating spermatids	Contribute to the open chromatin and required for spermatogonial differentiation and ensures entry into meiosis	Tachiwana et al., 2010; Ueda et al., 2017
Histone modifications	Acetylation	H4K5/8/12ac	Spermatogonia, spermatocytes and elongating spermatids	Essential for destabilization and remodeling of nucleosomes, TPs incorporation	Hazzouri et al., 2000;Gaucher et al., 2012; Qian et al., 2013; Bell et al., 2014; Dong et al., 2017; Ketchum et al., 2018
		H4K16ac	Elongating spermatids		
	Ubiquitination	UbH2A	Spermatocytes and elongating spermatids	Essential for the recruitment of the MOF acetyltransferase complex to modulate H4K16ac and histone removal	Chen et al., 1998; Baarends et al., 1999; Lu et al., 2010; Gou et al., 2017; Meng et al., 2019; Wang et al., 2019
		UbH2B	Spermatocytes and elongating spermatids		
	Methylation	H3K4me3	Spermatogonia, spermatocytes, round spermatids and elongating spermatids	Essential for the recruitment of PYGO2 to recognize HAT to facilitate H3 acetylation; recruit PHF7 to catalyze H2A ubiquitination to facilitate the histone removal	Godmann et al., 2007; Song et al., 2011; Nair et al., 2008; Wang et al., 2019
		H3K9me1/2/3	Spermatogonia, Round spermatids and elongating spermatids	Regulate the *Tnps* and *Prms* genes expression	Okada et al., 2007
		H3K36me3	Spermatocytes and round spermatids	Regulate the *Tnps* and *Prms* genes expression	Zuo et al., 2018
		H3K79me3	Elongating spermatids	Correlate with histone H4 hyperacetylation to regulate histone-to-protamine transition	Dottermusch-Heidel et al., 2014
	Phosphorylation	γH2AX	Spermatocytes elongating spermatids	Require for the normal quantities of H3, H4 and PRM2 precursor and intermediate	Li et al., 2005; Spiridonov et al., 2005; Jha et al., 2017
		H4S1	Spermatocyte, round spermatids and elongating spermatids	Essential for chromatin compaction and concomitantly histone accessibility	Krishnamoorthy et al., 2006; Zhang et al., 2016
	Other	Crotonylation	Elongating spermatids	Facilitate TP1 and PRM2 incorporation	Liu et al., 2017b
		PARsylation	Elongating spermatids	Require for histone removal and TP1 incorporation	Meyer-Ficca et al., 2009; Meyer-Ficca et al., 2011; Meyer-Ficca et al., 2015

TABLE 2 | Mouse models related with the histone-to-protamine transition.

Gene	Phenotype	Function	Reference
H1t	Fertility and no spermatogenesis abnormalities	Dispensable for histone-to-protamine transition	Drabent et al., 2000; Fantz et al., 2001
H1t2	Reduced fertility with delayed nuclear condensation and aberrant elongation of spermatids	Indispensable for the incorporation of PRMs and proper chromatin condensation	Martianov et al., 2005; Tanaka et al., 2005
Th2b	Fertility with normal spermatogenesis in *Th2b*-null mice Infertility with abnormal spermatozoa in TH2B C-terminus modified mice	Destabilize chromatin and regulate TPs and PRMs incorporation	Montellier et al., 2013
Th2a/Th2b	Infertility with accumulated spermatocytes at interkinesis and abnormal spermatozoa	Indispensable for cohesin release and TP2 incorporation	Shinagawa et al., 2015
H2al2	Infertility and the sperm chromatin show a compaction defects	Assemble open nucleosomes and allow TPs incorporation	Barral et al., 2017
H2a.b	Reduced fertility with abnormal spermatozoa	Destabilize chromatin and modulate the dynamics of H2AL2 removal and TP1 incorporation	Anuar et al., 2019
H3f3a	Reduced fertility with dysmorphic spermatozoa	Require for normal development of some spermatids	Tang et al., 2015
H3f3b	Infertility with abnormal spermatozoa and reduced sperm count	Indispensable for spermatogenesis related genes expression, TP1 removal and PRM1 incorporation	Yuen et al., 2014
H3t	Infertility with azoospermia	Require for spermatogonial differentiation and ensures entry into meiosis	Ueda et al., 2017
Epc1	Infertility with abnormal round spermatids to elongating spermatids transition	Require for round spermatids maturation by regulating histone acetylation and TP2 incorporation	Dong et al., 2017
Tip60	The germ cell is arrested at the RS stage	Contribute to round spermatids maturation by regulating histone acetylation and TP2 incorporation	Dong et al., 2017
Sirt1	Reduced fertility with abnormal spermatozoa and decreased sperm count	Require for acetylation of H4K5, H4K8 and H4K12 histone, and TP2 incorporation	Bell et al., 2014
Brdt	Infertility with complete absence of post-meiotic cells in *Brdt*-null mice Infertility with abnormal spermatids in $Brdt^{\Delta BD1/\Delta BD1}$ mice	Control the chromatin organization and meiotic sex chromosome inactivation; the first bromodomain of BRDT is essential to link histone removal and TPs, PRMs incorporation	Shang et al., 2007; Dhar et al., 2012; Gaucher et al., 2012; Manterola et al., 2018
Pa200	Reduced fertility with abnormal spermatozoa and decreased sperm count	Recognize acetylated histones and mediate the core histones for acetylation dependent degradation through proteasomes	Khor et al., 2006; Qian et al., 2013
Rnf8	Infertility with abnormal spermatozoa and reduced sperm count	Require for histone ubiquitination and modulate H4K16ac to facilitate histone removal and TPs, PRMs incorporation	Lu et al., 2010
Miwi	Infertility with abnormal spermatozoa and reduced sperm count in *Miwi* D-box mutations mice	Essential for nuclear translocation of RNF8 and facilitates the histone ubiquitination and further histone removal	Gou et al., 2017
L3mbtl2	Reduced fertility with abnormal spermatozoa and decreased sperm count	Require for the RNF8-UbH2A pathway and further PRM1 incorporation	Meng et al., 2019
Phf7	Infertility with abnormal spermatozoa and decreased sperm count	Recognize the H3K4me3/me2 and catalyze H2A ubiquitination to facilitate the histone removal	Wang et al., 2019
Pygo2	Infertile with abnormal spermatozoa and decreased sperm count in *Pygo2* reduced mice	Recognize H3K4me3 and recruit HAT to facilitate H3 acetylation and expression of *Prms*, *Tnp2*, and *H1fnt*.	Nair et al., 2008
Jhdm2a	Infertile with abnormal spermatozoa and decreased sperm count	Control H3K9 methylation at the promoter of *Tnp1* and *Prm1* genes and regulate their expression	Okada et al., 2007
Setd2	Infertility with round spermatid arrest	Catalyze H3K36me3 and facilitate the activation of *Tnp*s and *Prms* genes	Zuo et al., 2018
Tssk6	Infertility with abnormal spermatozoa	Mediate γH2AX to possess normal quantities of histone H3, H4 and PRM2 precursor and intermediate	Spiridonov et al., 2005; Jha et al., 2017
Cdyl	Reduced fertility with decreased sperm count and motility	Regulate histone crotonylation to facilitate TP1 and PRM2 incorporation	Liu et al., 2017b
Parp11	Infertility with teratozoospermia	Modulate PARsylation to facilitate chromatin condensation	Meyer-Ficca et al., 2015
Parg110	Reduced fertility with poor sperm chromatin quality	Dispensable for histone removal and TP1 incorporation	Meyer-Ficca et al., 2009; Meyer-Ficca et al., 2011
Tnp1	Reduced fertility with subtle abnormal spermatozoa and decreased spermatozoa	Dispensable for histone displacement as the compensation by TP2 and PRM2 precursor	Yu et al., 2000
Tnp2	Reduced fertility with abnormal spermatozoa	Dispensable for histone displacement but necessary for maintaining the normal processing of PRM2 and the completion of chromatin condensation	Zhao et al., 2001

(Continued)

TABLE 2 | Continued

Gene	Phenotype	Function	Reference
Tnp1/Tnp2	Infertile with abnormal spermatozoa and decreased sperm count	Indispensable for PRM2 incorporation and chromatin condensation	Shirley et al., 2004
Prm1	Infertile with abnormal spermatozoa and decreased sperm count	Indispensable for spermiogenesis and chromatin condensation	Cho et al., 2001
Prm2	Infertile with abnormal spermatozoa and decreased sperm count		
Camk4	Infertile with abnormal spermatozoa and decreased sperm count	Mediate the phosphorylation of PRM2 and facilitate basic nuclear proteins removal	Wu et al., 2000

related chromatin remodelers, regulate the highly orchestrated chromatin re-organization and facilitate the histone-to-protamine transition during spermiogenesis.

HISTONE VARIANTS

In eukaryotes, nucleosomes are the packing units of DNA, which contain four types of canonical histones (H2A, H2B, H3, and H4) and the linker histone H1 (Talbert and Henikoff, 2010; Kowalski and Palyga, 2012). While canonical histone expression is typically coupled to DNA replication, some non-canonical histones (histone variants) that are distinct form their canonical paralogues in amino acid sequence, are constitutively expressed and have roles in a wide range of processes (Talbert and Henikoff, 2010). Many histones variants are expressed during spermiogenesis and modulate the chromatin structure to facilitate the histone-to-protamine replacement (Mccarrey et al., 2005; Govin et al., 2007). Here, we summarize the recent advances in our understanding of the role of histone variants during the histone-to-protamine transition.

H1 VARIANTS

Linker histones contribute to form and stabilize the higher-order chromatin structure (Bednar et al., 1998). In mammals, there are about 11 different subtypes of histone H1 (Happel and Doenecke, 2009). Among these, H1T, H1T2, and HILS1 are testis-specific H1 variants (**Figure 1**) (Happel and Doenecke, 2009).

H1T is exclusively detected as early as mid- to late pachytene spermatocytes, and maintains high expression levels in the elongating spermatids (**Figure 1**) (Drabent et al., 1996; Drabent et al., 2003). Biochemical and biophysical studies found that, distinct from other somatic H1 variants, H1T binds less tightly to H1-depleted nucleosomes, suggesting it may maintain a relatively open chromatin configuration to facilitate histone replacement during spermiogenesis (Delucia et al., 1994; Khadake and Rao, 1995). Unexpectedly, *H1t*-null mice are fertile and exhibit no spermatogenesis abnormalities, and the histone-to-protamine transition in *H1t*-deficient testis is normal (Drabent et al., 2000; Fantz et al., 2001). Although the expression of some canonical subtypes, including H1.1, H1.2, and H1.4, is enhanced in *H1t*-null mice, elevated levels of H1.1 or H1.2 could not be observed in the *H1t*-deficient spermatids (Drabent et al., 2003), indicating some other types of H1 variants may play redundant roles in the histone-to-protamine transition.

H1T2 selectively localizes at the apical pole in the nucleus of round and elongating spermatids but not in mature spermatozoa (**Figure 1**) (Martianov et al., 2005). Distinct from H1T, H1T2 is critical for spermiogenesis, as homozygous *H1t2*-mutant males are infertile due to delayed nuclear condensation and aberrant elongation of spermatids. Further analysis shows the protamine levels are substantially reduced in *H1t2*-null spermatozoa (Martianov et al., 2005; Tanaka et al., 2005), indicating H1T2 is necessary for the incorporation of protamines, and proper chromitin condensation during the histone-to-protamine transition.

HILS1 is strongly expressed in the nuclei of elongating and elongated spermatids (**Figure 1**) (Yan et al., 2003). HILS1 is the least conserved H1 variant, and a poor condenser of chromatin compared with somatic H1, demonstrating the idea that HILS1 may have a distinct role in the histone-to-protamine transition (Yan et al., 2003; Mishra et al., 2018). In *Drosophila*, *Mst77F* encodes a linker histone-like protein that is similar with the mammalian HILS1 protein and expressed in elongating spermatids (Raja and Renkawitz-Pohl, 2005). The disruption of *Mst77F* cause male sterile as producing spermatozoa with malformed heads. Although the histone-to-protamine transition occurs independently of *Mst77F*, the nuclei of spermatid fail to properly condense after the histone-to-protamine replacement in *Mst77F* mutant male (Kimura and Loppin, 2016). However, the functional roles of HILS1 in mammalian spermiogenesis need further investigation.

H2A VARIANTS

Multiple testis-specific H2A variants have been identified in mammals, including TH2A, H2AL1, H2AL2, H2AL3 and H2A.B (Trostleweige et al., 1982; Govin et al., 2007; Soboleva et al., 2012).

TH2A is present and actively synthesized in early primary spermatocytes and gradually disappears during condensation of spermatid nuclei (**Figure 1**) (Shires et al., 1976; Trostleweige et al., 1982). TH2A could contribute to the open chromatin structure, as crystal structures of nucleosome core particles (NCPs) with TH2A show the H-bonding interactions between the TH2A/TH2A′ L1 loops are lost and the histone dimer-DNA contacts are dramatically decreased (Padavattan et al., 2015; Padavattan et al., 2017). Although a *Th2a*-knockout mouse model has yet

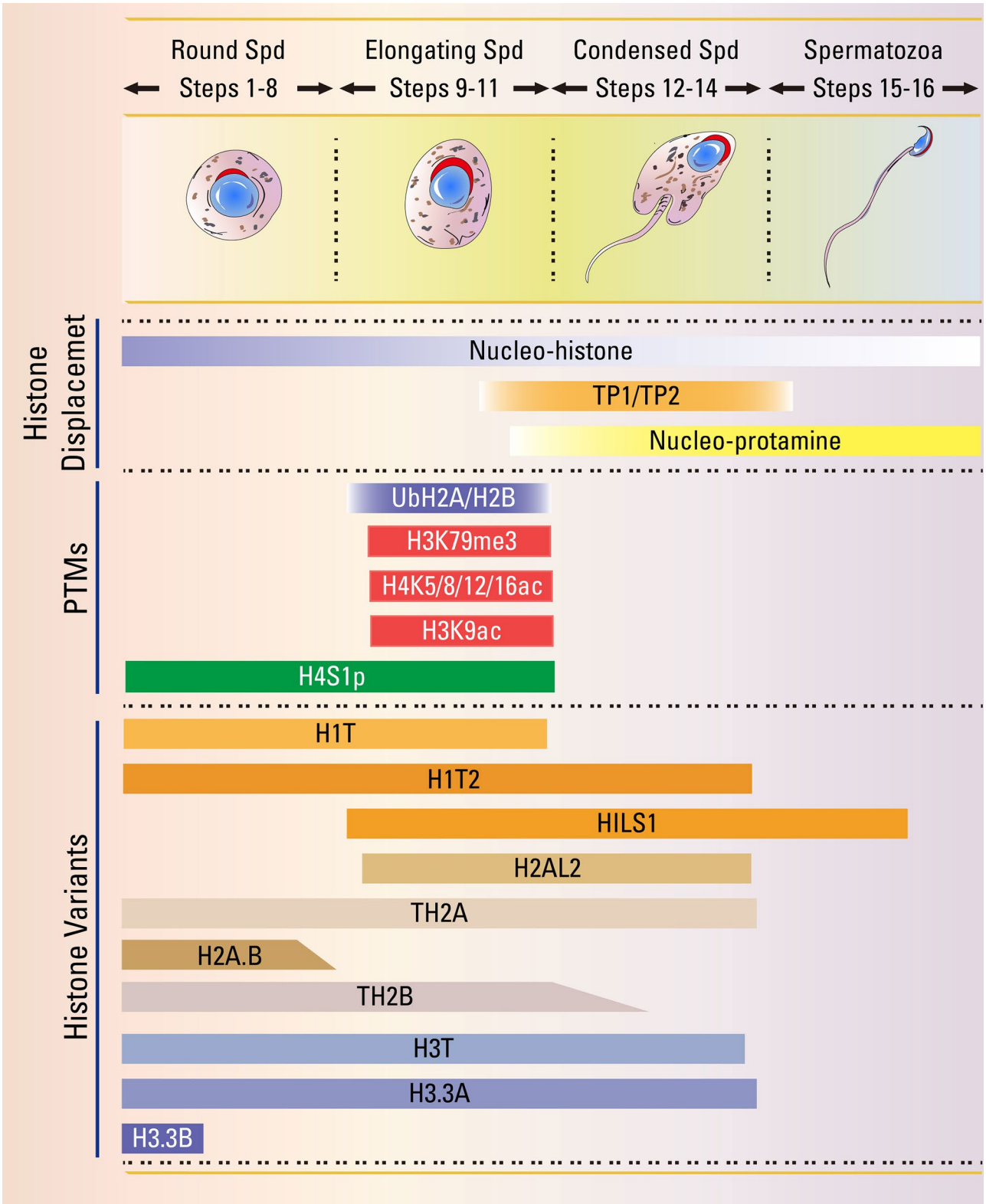


FIGURE 1 | Summary of critical time points and epigenetic events during the histone-to-protamine transition. The haploid germ cells undergo a dramatic morphological change, and nuclear chromatin re-organization starts from round spermatid (Spd) to mature spermatozoa. Detailed studies of the indicated histone variants and histone modification might establish the precise epigenetic events of spermiogenesis.

to be established, mice with knockouts of the testis-specific H2B variants *Th2a* and *Th2b* exhibit male infertility with few sperm in the epididymis (Shinagawa et al., 2015). In this double-knockout mouse, impaired chromatin incorporation of transition protein 2 (TP2) and elevated H2B could be observed in the mutant testis, suggesting the TH2A and TH2B may regulate the function in chromatin dynamics or the total histone levels to facilitate the histone replacement during spermatogenesis (Shinagawa et al., 2015). As the *Th2b*-null male mice show normal spermatogenesis and fertility (Montellier et al., 2013), the histone replacement defect in *Th2a*/*Th2b* double-knockout male mice is probably caused by the depletion of *Th2a* or their synergistic effect.

In late-developing post-meiotic male germ cells, H2AL2 is specifically expressed in condensing spermatids that correlates with the expression of TPs (**Figure 1**) (Govin et al., 2007). By comparing *H2al2*-null mice to wild-type mice, H2AL2 was demonstrated to be required to load TPs onto the nucleosome and for efficient PRMs assembly during the histone-to-protamine transition. Additionally, the nucleosome reconstitution assays revealed that the incorporation of H2A.L.2 can drastically modulate the nucleosome structure to facilitate TPs invading the nucleosomes and further transformation (Barral et al., 2017). Thus, H2AL2 could assemble open nucleosomes and allow TPs invading, which further promotes protamine processing and sperm genome compaction.

H2A.B is spatially and temporally regulated during spermatogenesis and detectable from the pachytene stage to the round spermatids (**Figure 1**) (Soboleva et al., 2012; Soboleva et al., 2017). *In vitro* studies show that H2A.B is able to destabilize chromatin and has unfolding properties to chromatin (Soboleva et al., 2012), indicating H2A.B might promote chromatin reorganization and further histones displacement by TPs. Male *H2a.b*-null male mice are subfertile due to the production of abnormal spermatozoa and clogged seminiferous tubules (Anuar et al., 2019). In *H2a.b*-null elongating spermatids, H2AL2 could not be detected in pericentric heterochromatin, and the replacement of TP1 by protamines appears to be delayed (Anuar et al., 2019). These results indicate H2A.B might modulate the dynamics of H2AL2 and TP1 chromatin incorporation and removal to participate in the histone-to-protamine transition.

H2B VARIANTS

The testis-specific histone variant TH2B is one of the earliest histone variants identified in testis (Shires et al., 1975). TH2B massively replaces somatic H2B during meiosis and remains the main type of H2B in round and elongating spermatids (Meistrich et al., 1985; Montellier et al., 2013), suggesting TH2B might be indispensable for meiotic and post-meiotic germ cells. The crystal structure analysis shows the TH2B could not form the water-mediated hydrogen bonds with H4R78 (Urahama et al., 2014), which may affect the stability of the TH2B nucleosome and facilitate histone replacement during spermiogenesis. In a *Th2b* mutant mouse, which contains modified C-terminus of the TH2B protein and causes a dominant-negative effect, males were infertile and severe abnormalities were seen in the elongating spermatids, which affected subnucleosomal transitional states during histone replacement (Boskovic and Torres-Padilla, 2013; Montellier et al., 2013). In contrast, *Th2b*-null mice are fertile and show normal spermatogenesis process, indicating a compensatory mechanism that rescues deficiency of TH2B in the histone-to-protamine transition. Indeed, in *Th2b*-null testis, the expression of somatic H2B was significantly increased and elevated methylation of H4R35, H4R55, H4R67, and H2BR72 could be detected in *Th2b*-null spermatids. As H4R35, H4R55, H4R67, and H2BR72 participate in the interactions of histone–DNA and histone–histone, and their methylation may impair these intranucleosomal interactions (Hoghoughi et al., 2018). Thus, the elevated somatic H2B and histone modification in *Th2b*-null spermatids might rescue the *Th2b* deficiency in testis (Montellier et al., 2013; Bao and Bedford, 2016).

In humans, H2BFWT is a testis-specific histone, is synthesized and aggregated in testes, and single nucleotide polymorphisms (SNPs) in this gene is highly associated with male infertility (Churikov et al., 2004; Lee et al., 2009; Ying et al., 2012; Rafatmanesh et al., 2018; Teimouri et al., 2018). And spermatid-specific H2B (ssH2B) and H2BL1 have been identified and are strongly enriched in round or elongating spermatids, similar to that of TPs and protamines (Moss and Orth, 1993; Unni et al., 1995; Govin et al., 2007). However, the functional roles of these H2B variants in the histone-to-protamine transition still need to be further elucidated.

H3 VARIANTS

In addition to the two canonical histones H3.1 and H3.2, three additional H3 variants have been identified and expressed in mammal testes, including H3.3, H3T and H3.5 (Rathke et al., 2014; Bao and Bedford, 2016).

H3.3 differs from canonical H3.1 with five amino acids, is expressed throughout mouse seminiferous tubules, and accumulates in the XY body of spermatocytes (Bramlage et al., 1997; Van Der Heijden et al., 2007). Biochemical and biophysical studies show that H3.3 contributes to an open chromatin configuration and promotes transcription through disrupting the higher-order chromatin structure (Thakar et al., 2009; Chen et al., 2013). H3.3 could be encoded by two gene paralogs in mammal, *H3f3a* and *H3f3b*, and the depletion of either *H3f3a* or *H3f3b* causes male infertility. The disruption of *H3f3a* produces abnormal spermatozoa (Couldrey et al., 1999; Tang et al., 2015), and the loss of *H3f3b* leads to growth defects and death at birth, with surviving *H3f3b*-null males showing complete infertility (Yuen et al., 2014). In *H3f3b*-null germ cells, the TP1 is abnormally deposited in elongating spermatids while PRM1 could not be observed in in elongated spermatids and mature spermatozoa, indicating that *H3f3b* is required for chromatin reorganization and the histone-to-protamine transition (Yuen et al., 2014). H3T (H3.4) is exclusively expressed in the spermatocyte and diminishes in the elongating spermatids (Ueda et al., 2017). Biochemical studies clearly indicate that, in the H3T nucleosome, the DNA around the entry-exit regions shows more flexible than that of the H3.1-containing nucleosome, and that

the H3T-containing polynucleosome could formed more open configuration than that of H3.1 (Tachiwana et al., 2010). However, the disruption of H3T leads to sterile males with azoospermia, as spermatocyte and spermatids are absent in the *H3t*-null testes (Ueda et al., 2017). Thus, the function of H3T in the later stage of spermatogenesis need further investigated by using spatially and temporally specific knockout mouse models.

H3.5 is highly expressed in human testis and specifically observed in spermatogonia and spermatocytes (Shiraishi et al., 2017). *In vitro* studies reveal that the H3.5-specific L103 residue, reduces the hydrophobic interaction with histone H4 in the H3.5-containing nucleosome, which corresponds to the H3.3 Phe104 residue (Urahama et al., 2016). H3.5 is significantly reduced in non-obstructive azoospermia (NOA) patients (Shiraishi et al., 2017), whereas the precise roles of H3.5 in spermatogenesis remain largely unknown.

HISTONE MODIFICATION

Covalent conjugation of different post-translational modification of histones has a dramatic effect on the chromatin conformation by affecting the stability of the nucleosome and the histone-DNA interaction (Bao and Bedford, 2016). Many types of histone modifications have been identified to facilitate the histone-to-protamine transition, including acetylation, ubiquitination, methylation, and phosphorylation (Luense et al., 2016).

ACETYLATION

Hyperacetylated histones could facilitate histone eviction, and the acetylation of H2A, H2B, H3, H4 and histone variants have been detected in mammal testis (Grimes and Henderson, 1984a; Grimes and Henderson, 1984b; Oliva and Mezquita, 1986; Oliva et al., 1987). In *Drosophila*, inactivation of histone acetyltransferases by anacardic acid prevents the histones degradation and further a protamine incorporation during spermiogenesis (Awe and Renkawitz-Pohl, 2010), suggesting that histone acetylation is essential for the histone-to-protamine replacement.

H4 acetylation (H4K5ac, H4K8ac, H4K12ac, and H4K16ac) shows a spatial distribution pattern during spermatogenesis and is indispensable for the histone-to-protamine transition (Bao and Bedford, 2016; Ketchum et al., 2018). H4K5ac, H4K8ac and H4K12ac are expressed in spermatogonia and pre-leptotene spermatocytes, disappear in leptotene to pachytene spermatocytes, reappeared in elongating spermatids, and finally disappeared in condensing spermatids (**Figure 1**) (Hazzouri et al., 2000; Ketchum et al., 2018). In contrast, H4K16ac could only be detected in elongating spermatids (**Figure 1**) (Ketchum et al., 2018). *In vitro* analysis shows that H4 acetylation is essential for destabilization and remodeling of nucleosomes, and the incorporation of H4K16ac into nucleosomes prevents the formation of compact chromatin fibers and influence chromatin forming cross-fiber interactions (Tse et al., 1998; Shogren-Knaak et al., 2006; Kan et al., 2009). These findings indicate that H4 acetylation modulates higher order chromatin structure to facilitate the histone-to-protamine transition. EPC1 (Enhancer Of Polycomb Homolog 1) and TIP60 (Tat-interactive protein, 60 kDa), which are two components for the mammalian NuA4 (nucleosome acetyltransferase of H4) complexes (**Figure 2**) (Doyon et al., 2004), are co-localized to the nuclear periphery near the acrosomes in both round spermatids and elongating spermatids (Dong et al., 2017). The depletion of either *Epc1* or *Tip60* perturbs histone hyperacetylation, especially H4 acetylation, and affects histone replacement during spermiogenesis (Dong et al., 2017). Another gene that may play a role in acetylation is SIRT1 (Sirtuin 1), a member of the NAD+-dependent deacetylase. Germ cell-specific *Sirt1* knockout mice display reduced male fertility due to decreased spermatozoa number and increased proportion of abnormal spermatozoa (Bell et al., 2014; Liu et al., 2017a). In *Sirt1*-null elongating and elongated spermatids, acetylation levels of H4K5, H4K8 and H4K12 are decreased and TP2 could not co-localize in the nucleus, leading to a chromatin condensation defect in *Sirt1*-null spermatozoa (Bell et al., 2014). Thus, SIRT1 may modulate other factors to promote H4 acetylation and the histone-to-protamine transition.

The histone acetylation might be recognized by some chromatin remodelers to confer downstream signaling, and the double bromodomain and extra-terminal domain (BET) proteins have been identified to be critical epigenetic readers binding to acetylated histones and modulating changes in chromatin structure and organization during spermiogenesis (Berkovits and Wolgemuth, 2013). BRDT is a testis-specific BET member protein, which is expressed specifically in spermatocytes and spermatids, and contains two bromodomains that specifically recognize acetylated lysine residues (Shang et al., 2007; Dhar et al., 2012; Manterola et al., 2018). BRDT binds the hyperacetylated histone H4 tail and co-localizes with acetylated H4 in elongating spermatids (Pivot-Pajot et al., 2003; Govin et al., 2006). Remodeling assays have shown BRDT regulated the chromatin reorganization dependent acetylation in round spermatids (Dhar et al., 2012). In mice, the disruption of the first bromodomain in BRDT resulted in male sterility by producing the morphologically abnormal spermatids (Shang et al., 2007). In elongating spermatids with BRDT containing a knockout of bromodomain 1 (BD1), TPs and protamines remained in the cytoplasm and histone replacement did not occur, suggesting BRDT is required for the histone-to-protamine transition by mediating the replacement of acetylated histones (**Figure 2**) (Gaucher et al., 2012). Furthermore, BRDT was found to bind with the N-terminus of SMARCE1 (SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1), a member of the SWI/SNF family of ATP-dependent chromatin remodeling complexes (Dhar et al., 2012), indicating BRDT may cooperate with SMARCE1 to facilitate the histone-to-protamine transition during spermiogenesis (**Figure 2**).

Proteasomes catalyze ATP- and polyubiquitin-dependent protein degradation, and they are made up of a 20S catalytic core particle (CP) and regulatory particle (RP). The 20S CP could be activated by cooperation with various RPs, such as PA700/19S, PA28α/β, PA28γ, and PA200 (Stadtmueller and Hill, 2011). PA200 is highly expressed in the testis, and the disruption of PA200

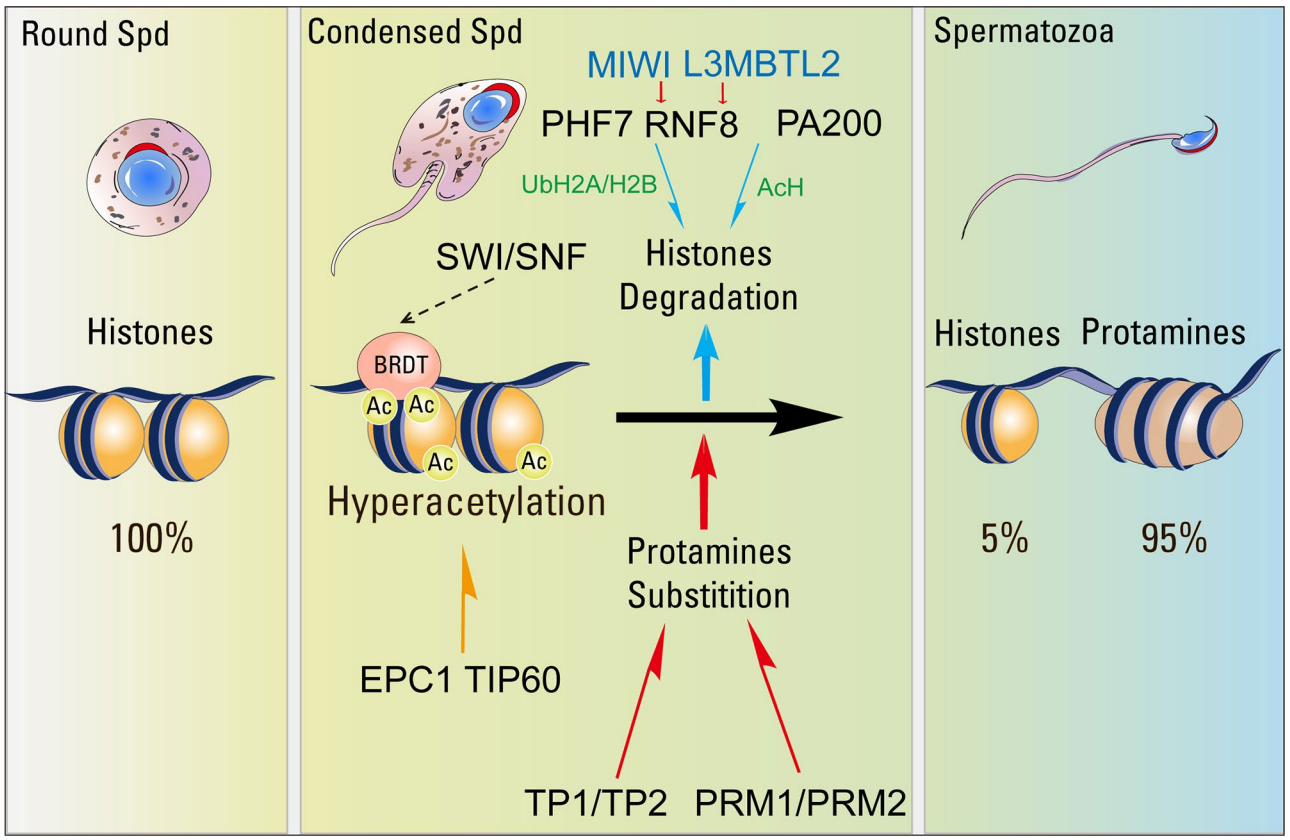


FIGURE 2 | The key factors related to the histone-to-protamine transition. Global incorporation of various H2A, H2B and H3 histone variants creates highly unstable nucleosomes, which then undergo histone hyperacetylation by EPC1/TIP60 or some other nucleosome acetyltransferase complexes. Acetylation at critical lysines further destabilizes the nucleosomes, while tail acetylation generates a platform for the recruitment of BRDT. BRDT interacts with the SWI/SNF family protein then starts the process of histone eviction and replacement by TPs. Evicted acetylated histones would then be recognized by PA200 and degraded by proteasomes during spermatogenesis. RNF8 could catalyze the ubiquitination of H2A. Ubiquitinated H2A and H2B control H4K16ac by regulating the association of MOF to the chromatin and facilitates histone removal in elongating spermatids. MIWI binds to RNF8 in the cytoplasm of early spermatids (Spd) through a Piwi-interacting RNA (piRNA)-independent manner, and promotes the nuclear translocation of RNF8 in late spermatids to catalyze histone ubiquitination and trigger histone removal. L3MBTL2 could interact with RNF8 and facilitate RNF8-dependent histone ubiquitination-related histone removal. PHF7 could recognize the H3K4me3/me2 and catalyze H2A ubiquitination to facilitate histone removal in elongating spermatids.

results in male infertility and severe defects in spermatogenesis (Ustrell et al., 2005; Khor et al., 2006). During spermiogenesis, PA200 regulatory could directly recognize acetylated histones through a bromodomain-like module and promote their ubiquitin-independent degradation. In *Pa200*-null spermatids, results showed that H2B, H3 and elevated H4K16ac could be detected at the end of the elongation stage (Qian et al., 2013). Thus, PA200 specifically recognizes acetylated histones and mediates the core histones for acetylation dependent degradation through proteasomes during spermatogenesis (**Figure 2**).

UBIQUITINATION

Ubiquitin is a 76 amino acid protein that is attached to target proteins to regulate several cellular processes, such as protein degradation, cell signaling, autophagy, DNA damage responses and so on (Hershko and Ciechanover, 1998; Pickart, 2001; Welchman et al., 2005; Komander and Rape, 2012). Ubiquitinated H2A and H2B are enriched in spermatocytes and elongating spermatids (Chen et al., 1998; Baarends et al., 1999). RNF8 is an ubiquitin E3 ligase that participates in DDR (DNA damage repair) by catalyzing the ubiquitination of H2A to promote the recruitment of some DNA damage response factors on the damage sites (Ma et al., 2011). The disruption of *Rnf8* causes significant late-stage developmental defects in spermatids due to problematic histone-to-protamine replacement, with the canonical histones being detectable in *Rnf8*-deficient mature spermatozoa (Lu et al., 2010). In *Rnf8*-null mice, both ubiquitinated H2A and H2B are decreased in the testes and H4K16ac is dramatically decreased as well (Lu et al., 2010). Further studies showed that ubiquitinated H2A and H2B were essential for the efficient recruitment of the MOF (males absent on the first) acetyltransferase complex, which is highly expressed in elongating spermatids and responsible for H4K16 acetylation in the chromatin (Akhtar and Becker, 2000; Lu et al., 2010). Thus, RNF8 catalyzed histone ubiquitination could modulate H4K16ac by regulating the localization of MOF on the chromatin and facilitate histone removal in the elongating spermatids.

The RNF8-dependent histone ubiquitination during spermiogenesis could also be modulated by PIWI protein, which is specifically expressed during germline development and enlists piRNAs (Piwi-interacting RNAs) to repress TE (transposable elements) and protect the germ cell genome integrity (Juliano et al., 2011; Siomi et al., 2011; Gou et al., 2017). In mice, *Miwi*, *Mili*, and *Miwi2*, the *Piwi* paralogs, have been identified in the testis and are required for male fertility (Deng and Lin, 2002; Kuramochi-Miyagawa et al., 2004; Carmell et al., 2007). During spermiogenesis, MIWI binds to RNF8 in the cytoplasm of early spermatids through a piRNAs-independent manner, and APC/C mediated MIWI degradation in late spermatids is essential for nuclear translocation of RNF8, which catalyzes histone ubiquitination and further facilitates histone removal (Gou et al., 2017). In both humans and mice, mutations in the conserved destruction box (D-box) of HIWI and MIWI proteins, which lead to their stabilization, cause male infertility due to impaired histone ubiquitination and histone-to-protamine transition (Gou et al., 2017). Except MIWI, L3MBTL2 (Lethal 3 malignant brain tumor like 2), a member of the MBT-domain proteins that is implicated in chromatin compaction, could also interact with RNF8. The depletion of *L3mbtl2* in germ cells affected male fertility by producing abnormal spermatozoa and the decrease of sperm counts. *L3mbtl2* deficiency also caused the reduction of in levels of the RNF8 and histone ubiquitination in elongating spermatids, which further influenced the PRM1 deposition and chromatin condensation during spermiogenesis (Meng et al., 2019).

PHF7 (PHD Finger Protein 7), which contains PHD (plant homeodomain) and RING finger domain, has been identified as a novel H2A ubiquitination E3 ligase in mouse testis (Hou et al., 2012; Wang et al., 2019). PHF7 is specifically located in the elongating spermatid nuclei, and the disruption of *Phf7* led to male mouse infertility as reduction of sperm count and the increased proportion of abnormal spermatozoa (Wang et al., 2019). PHF7 could recognize the H3K4me3/me2 through its PHD domain and catalyze H2A ubiquitination by its RING domain. In *Phf7*-null spermatids, the H2A ubiquitination was dramatically decreased that resulted in the histone retention and protamine replacement defect (**Figure 2**) (Wang et al., 2019). Therefore, PHF7 has dual roles during the histone-to-protamine transition that works as an epigenetic reader by recognizing H3K4me3/me2 and as an epigenetic writer through catalyzing H2A ubiquitination to promote histone removal.

METHYLATION

Multiple histone methylation have been identified in elongating spermatids, for instance H3K4me2, H3K4me3, H3K9me2, H3K9me3, H3K27me3, H3K79me2, and H3K79me3 (Godmann et al., 2007; Song et al., 2011; De Vries et al., 2012; Dottermusch-Heidel et al., 2014). Among them, the methylation of H3K4 and plus acetylation might help to achieve a more-open chromatin configuration, whereas H3K9 and H3K27 methylation are known to be associated with a more-repressed chromatin configuration (Rathke et al., 2014), indicating a balance of "opened" and "closed" chromatin regions during the histone-to-protamine transition. As some histone methyltransferases and demethylases are detectable during spermiogenesis (Godmann et al., 2007; Liu et al., 2010; Ushijima et al., 2012), the histone methylation may be dynamically regulated in testis. Although few mouse models exist that allow precise detection of methylation activity that directly regulates histone replacement during spermiogenesis, some studies have revealed that histone methylation may modulate the histone-to-protamine transition through some other ways. PYGO2 (Pygopus homolog 2) comprises a C-terminal PHD finger, which can recognize the H3K4me3 and is specifically located in the elongating spermatid nuclei. In mice, the reduction of *Pygo2* influenced the *Tnp, Prm* genes expression and caused the abnormal nuclear condensation, which further led to male sterility (Nair et al., 2008). Furthermore, PYGO2 associates with a histone acetyltransferase (HAT) activity, and the acetylation of H3 is disrupted in *Pygo2* reduced elongating spermatids (Nair et al., 2008), indicating PYGO2 may recognize H3K4me3 through its PHD domain and could recruit HAT to facilitate H3 acetylation and further histone-to-protamine transitions. As described before, PHF7 could recognize the H3K4me3/me2 and catalyze H2A ubiquitination to facilitate the histone-to-protamine transitions (Wang et al., 2019). The predominant histone methyltransferase SETD2 (SET domain–containing 2) catalyzes the H3K36me3, and knocking out *Setd2* in mouse germ cells causes aberrant spermiogenesis, resulting in complete male infertility. Moreover, the disruption of SETD2 causes complete loss of H3K36me3 and impaired activation of *Tnp* and *Prm* genes (Zuo et al., 2018), indicating H3K36me3 may regulate the histone-to-protamine transition by activating *Tnp* and *Prm* genes expression. Contrarily, JHDM2A (JmjC-domain-containing histone demethylase 2A) is an H3K9me2/1-specific demethylase. The loss of *Jhdm2a* in mice exhibits post-meiotic chromatin condensation defects and leads to male infertility. Although global H3K9 methylation has no effect in *Jhdm2a*-null testis, JHDM2A directly binds to and controls H3K9 methylation at the promoter of *Tnp1* and *Prm1* genes, which further regulates the sperm genome packaging and chromatin condensation (Okada et al., 2007).

PHOSPHORYLATION

Histone phosphorylation is involved in various cellular processes (Rossetto et al., 2012; Bao and Bedford, 2016), and dynamic histone phosphorylation have been observed during spermatogenesis (Govin et al., 2010; Bao and Bedford, 2016). The phosphorylation of histone H2AX at residue Ser139 (γH2AX) plays important roles in many biological processes, such as meiotic recombination and male sex chromosome inactivation in germ cells (Li et al., 2005). γH2AX is detectable in elongating spermatids, and TSSK6 has been identified to be responsible for the H2AX phosphorylation during spermiogenesis (Jha et al., 2017). In mice, targeted deletion of *Tssk6* leads to male sterility caused by the impairment in morphology and motility of spermatozoa (Spiridonov et al., 2005). In spermatozoa, the loss of TSSK6 blocks γH2AX formation, resulting in elevated H3, H4 and the precursor and intermediate of PRM2 (Jha et al., 2017). These results indicate that TSSK6 may mediate γH2AX to participate in the histone-to-protamine transition. H4S1 phosphorylation is highly expressed in mouse spermatocyte,

round and elongating spermatids (Krishnamoorthy et al., 2006; Zhang et al., 2016). H4S1 phosphorylation has been found to be essential for chromatin compaction and concomitantly histone accessibility (Krishnamoorthy et al., 2006; Wendt and Shilatifard, 2006), suggesting that H4S1 phosphorylation is required for histone replacement during spermiogenesis. Outside the canonical histones, many phosphorylated residues have been identified, using mass spectrometry analyses, that exist on different testis-specific histone variants, such as H1T, HILS1, TH2A, TH2B (Sarg et al., 2009; Pentakota et al., 2014; Mishra et al., 2015; Luense et al., 2016; Hada et al., 2017). Although many core histones and histone variants phosphorylation have been identified in germ cells, their physiological roles need further investigation.

OTHER MODIFICATIONS

A variety of histone lysine modifications have been identified, including butyrylation, crotonylation, malonylation, propionylation, and succinylation (Tan et al., 2011; Sabari et al., 2017). Kcr (Lysine crotonylation) is a newly identified histone modification and is detectable in elongating spermatids, which regulated testis-specific genes activation in post-meiotic germ cells (Tan et al., 2011). The CDYL (chromodomain Y-like) protein, which contains a C-terminal CoAP domain that interacts with CoA to achieve its crotonyltransferase activity, may suppress the histone Kcr by converting crotonyl-CoA to β-hydroxybutyryl-CoA. Accordingly, *Cdyl*-deficient male mice show reduced fertility, decreased epididymal sperm count and sperm cell motility, and dysregulated histone Kcr (Liu et al., 2017b). In the *Cdyl*-deficient mouse testes, further analysis showed that the elevated TP1 and PRM2 were localized in a chromatin-free regions (Liu et al., 2017b), suggesting that histone crotonylation is essential for the histone-to-protamine transition during spermiogenesis.

Poly-ADP-ribosylation (PARsylation) is a common protein PTM (post-translational modification) observed in higher eukaryotes and involved in many different fundamental cellular functions. All of core histones and the linker histone H1 can be ADP-ribosylated (Gagne et al., 2006; Messner and Hottiger, 2011), which could be catalyzed by poly(ADP-ribose) polymerases, such as PARP1 and PARP2, and resolved by PARG (PAR glycohydrolase) (Gibson and Kraus, 2012). The PARP1, PARP2 and PARsylation proteins are specifically detected in elongating spermatids (Meyer-Ficca et al., 2005), and the perturbed PARsylation causes reduced male fertility with abnormal retention of core histones, H1T and HILS1 in mature spermatozoa (Meyer-Ficca et al., 2009; Meyer-Ficca et al., 2011; Meyer-Ficca et al., 2015). Thus, PARsylation is essential for the histone-to-protamine replacement, yet the precise PARsylation histone sites need further characterization.

TRANSITION PROTEINS

Between histone eviction and protamine incorporation in the nuclei of spermatids, about ninety percent of the chromatin components consist of TPs, which are arginine- and lysine-rich proteins encoded by *Tnp1* and *Tnp2* (Meistrich et al., 2003). However, the functional roles of each TP are still controversial (Rathke et al., 2014). TP1 could reduce the melting temperature of DNA and relax the DNA from core particles of nucleosome, whereas TP2 tends to compact the nucleosomal DNA by increasing its melting temperature, indicating TP2 may promote DNA condensation while TP1 facilitates the eviction of the histones (Singh and Rao, 1988; Akama et al., 1998; Kolthur-Seetharam et al., 2009; Rathke et al., 2014). However, a separate study that shown that neither TP1 nor TP2 leads to the conformation changes in supercoiled DNA (Levesque et al., 1998). These differences might reveal their unique roles during mammal spermiogenesis, as single knockout of either *Tnp1* or *Tnp2* leads to little morphological alteration of spermatozoa in mouse models. Elevated TP2 and TP1 proteins could be observed in *Tnp1*-null and *Tnp2*-null spermatids, respectively (Yu et al., 2000; Zhao et al., 2001). Thus, TP1 and TP2 may compensate for each other *in vivo*. Indeed, *Tnp1* and *Tnp2* double-knockout mice show severe abnormal spermiogenesis with a general decrease in sperm motility and abnormal sperm morphology (Shirley et al., 2004). The chromatin condensation is perturbed in the *Tnp1* and *Tnp2* double-knockout mice as severe histones retention is detectable, indicating TPs function redundantly yet have unique roles in the histone-to-protamine transition (Shirley et al., 2004; Zhao et al., 2004; Bao and Bedford, 2016).

PROTAMINES

Protamines are basic proteins that replace TPs in late spermatids (Rathke et al., 2014; Bao and Bedford, 2016). Two protamine genes (*Prm1* and *Prm2*) localize on the same chromosome in both humans and mice (Balhorn, 2007). Protamines tightly interact with DNA *via* a central arginine-rich DNA-binding domain (Balhorn, 2007). Unlike *Tnp* genes, the disruption of either *Prm1* or *Prm2* leads to the male infertility (Cho et al., 2001). Protamines have multiple PTM sites, and a total of 11 PTMs have been identified on the protamines of mouse spermatozoa, including acetylation, phosphorylation and methylation (Brunner et al., 2014). One site of interest is PRM2 S55, which is a candidate phosphorylated substrate residue of CAMK4 (Ca^{2+}/calmodulin-dependent protein kinase IV) (Wu et al., 2000). Targeted *Camk4* knockout male mice are infertile, and the transition protein displacement by PRM2 is perturbed as a specific loss of PRM2 and prolonged retention of TP2 in *Camk4*-null spermatids. *In vitro*, PRM2 could be phosphorylated by CAMK4, implicating CAMK4 mediated PRM2 phosphorylation is required for the protamine incorporation during spermiogenesis (Wu et al., 2000). Thus, the specific post-translational modifications on protamines may also be essential for the histone-to-protamine transition.

CONCLUSION AND FUTURE PERSPECTIVES

During the histone-to-protamine transition, many epigenetic regulators work together to facilitate paternal genome re-organization and packaging into the highly condensed nuclei of spermatozoa,

through histone variation, specific histone modification and their related chromatin remodelers. Any defects during the histone-to-protamine transition would lead to male infertility (Bao and Bedford, 2016). While the morphological changes during spermiogenesis are well characterized, the precise molecular mechanisms underlying the chromatin re-organization, in particular the transition from histones to protamines, are still unclear. It's difficult to characterize the dynamic processes that occur during histone eviction, transition protein incorporation and protamine insertion. Moreover, 10% of the spermatozoa population in the epididymis has not yet completed the histone-to-protamine transition (Yoshida et al., 2018). These problems may be ascribed to a lack of experimental methods, which could fully recapitulate germ cell development *in vitro*. Further physiological insights may be gained by developing an *in vitro* germ-cell culture system that more accurately recapitulates the *in vivo* histone-to-protamine transition.

Many histone variants modulate histone replacement by regulating the chromatin structure; therefore, nucleosomes containing these histone variants often maintain a relatively decondensed and open chromatin configuration, facilitating histone replacement during spermiogenesis. The redundant function of histone variants in modulating chromatin configuration ensures that defects in some histone variants have a limited effect on spermatogenesis. Indeed, some mutant histone variants in mouse models are dispensable for male fertility, and mice may show elevated levels of compensatory histones or histone variants. However, the redundant function of histone variants makes it difficult to explore the precise role of each histone variant in histone replacement.

Although many histone modifications have been identified during the histone-to-protamine transition, many studies are descriptive and correlative. The direct manipulation of histone modification sites to reveal function is still urgently needed. With the development of gene editing tools, for example the CRISPR/Cas9 system, mouse models disrupting these histone modifications may be generated and used to elucidate function and *in vivo* relevance in the future. The following open-ended questions still need to be answered to provide in-depth investigation in the field.

(1) In addition to the histone variants and modifications mentioned above, what other novel histone variants and modifications participate in the histone-to-protamine transition? How can we identify them?
(2) How and where do histone variants replace canonical histones? What signal is needed to initiate replacement?
(3) As histone variants and modifications are identified that participate in the histone-to-protamine transition, how do we establish an epigenetic modulating network for this process? Which type of histone code is the initiating code?
(4) Histone hyperacetylation works as a determining event during the histone-to-protamine transition. Is this histone hyperacetylation an initial signal or an indirect consequence of prior events?
(5) What's the relationship between these histone variants and modifications? What's the mechanism underlying the cross talk between them?
(6) Chromatin assembly is modulated by histone chaperones or other chromatin remodelers. What's the functional role of histone chaperones during the histone-to-protamine transition? Which histone variants or modifications send a signal to the chaperones?
(7) How do the transition proteins replace the histone? How do protamines replace the transition proteins? What are the detailed functional roles of transition proteins?
(8) Are there still post-translational modifications that need to be discovered in order to more accurately describe how histone modification plays a role in spermatozoa maturation?

These questions and their underlying ideas need further investigation and refining to help us more thoroughly understand the complex molecular relationships and exact regulating mechanisms of the histone-to-protamine transition.

AUTHOR CONTRIBUTIONS

WT and HG wrote the manuscript and drew the figures; CL and WL proposed the idea and revised the manuscript. All authors listed have made a substantial, direct and intellectual contribution to the work and approved it for publication.

FUNDING

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA16020701), the National Key R&D Program of China (grant 2016YFA0500901, 2018YFC1004202), the National Natural Science Foundation of China (grants 31771501, 91649202) and the Youth Innovation Promotion Association CAS (2018109).

ACKNOWLEDGMENTS

We thank Tracey Baas for critical reading of the manuscript.

REFERENCES

Akama, K., Sato, H., Hasegawa, S., Shimada, I., and Nakano, M. (1998). Transition protein 1 from boar late spermatid nuclei having DNA-melting activity is a dimeric protein. *Biochem. Mol. Biol. Int.* 44, 315–323. doi: 10.1080/15216549800201332

Akhtar, A., and Becker, P. B. (2000). Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in *Drosophila*. *Mol. Cell* 5, 367–375. doi: 10.1016/S1097-2765(00)80431-1

Anuar, N. D., Kurscheid, S., Field, M., Zhang, L., Rebar, E., Gregory, P., et al. (2019). Gene editing of the multi-copy H2A.B gene and its importance for fertility. *Genome Biol.* 20, 23. doi: 10.1186/s13059-019-1633-3

Awe, S., and Renkawitz-Pohl, R. (2010). Histone H4 acetylation is essential to proceed from a histone- to a protamine-based chromatin structure in spermatid nuclei of *Drosophila melanogaster*. *Syst. Biol. in Reprod. Med.* 56, 44–61. doi: 10.3109/19396360903490790

Baarends, W. M., Hoogerbrugge, T. W., Roest, H. P., Ooms, M., Vreeburg, J., Hoeijmakers, J. H. J., et al. (1999). Histone ubiquitination and chromatin

remodeling in mouse spermatogenesis. *Dev. Biol.* 207, 322–333. doi: 10.1006/dbio.1998.9155

Balhorn, R. (2007). The protamine family of sperm nuclear proteins. *Genome Biol.* 8, 1–8. doi: 10.1186/gb-2007-8-9-227

Bao, J., and Bedford, M. T. (2016). Epigenetic regulation of the histone-to-protamine transition during spermiogenesis. *Reproduction* 151, R55–R70. doi: 10.1530/REP-15-0562

Barral, S., Morozumi, Y., Tanaka, H., Montellier, E., Govin, J., De Dieuleveult, M., et al. (2017). Histone Variant H2A.L.2 Guides transition protein-dependent protamine assembly in male germ cells. *Mol. Cell* 66, 89–101 e108. doi: 10.1016/j.molcel.2017.02.025

Bednar, J., Horowitz, R. A., Grigoryev, S. A., Carruthers, L. M., Hansen, J. C., Koster, A. J., et al. (1998). Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14173–14178. doi: 10.1073/pnas.95.24.14173

Bell, E. L., Nagamori, I., Williams, E. O., Del Rosario, A. M., Bryson, B. D., Watson, N., et al. (2014). SirT1 is required in the male germ cell for differentiation and fecundity in mice. *Development* 141, 3495–3504. doi: 10.1242/dev.110627

Berkovits, B. D., and Wolgemuth, D. J. (2013). The role of the double bromodomain-containing BET genes during mammalian spermatogenesis. *Gametogenesis* 102, 293–326. doi: 10.1016/B978-0-12-416024-8.00011-8

Boskovic, A., and Torres-Padilla, M. E. (2013). How mammals pack their sperm: a variant matter. *Genes Dev.* 27, 1635–1639. doi: 10.1101/gad.226167.113

Bramlage, B., Kosciessa, U., and Doenecke, D. (1997). Differential expression of the murine histone genes H3.3A and H3.3B. *Differentiation* 62, 13–20. doi: 10.1046/j.1432-0436.1997.6210013.x

Brunner, A. M., Nanni, P., and Mansuy, I. M. (2014). Epigenetic marking of sperm by post-translational modification of histones and protamines. *Epigenet. Chromatin* 7, 1–12. doi: 10.1186/1756-8935-7-2

Carmell, M. A., Girard, A., Van De Kant, H. J. G., Bourc'his, D., Bestor, T. H., De Rooij, D. G., et al. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev. Cell* 12, 503–514. doi: 10.1016/j.devcel.2007.03.001

Chen, H. Y., Sun, J. M., Zhang, Y., Davie, J. R., and Meistrich, M. L. (1998). Ubiquitination of histone H3 in elongating spermatids of rat testes. *J. Biol. Chem.* 273, 13165–13169. doi: 10.1074/jbc.273.21.13165

Chen, P., Zhao, J. C., Wang, Y., Wang, M., Long, H. Z., Liang, D., et al. (2013). H3.3 actively marks enhancers and primes gene transcription via opening higher-ordered chromatin. *Genes Dev.* 27, 2109–2124. doi: 10.1101/gad.222174.113

Cho, C., Willis, W. D., Goulding, E. H., Haesook, J. H., Choi, Y. C., Hecht, N. B., and Eddy, E. M. (2001). Haploinsufficiency of protamine-1 or-2 causes infertility in mice. *Nat. Genet.* 28, 82–86. doi: 10.1038/ng0501-82

Churikov, D., Siino, J., Svetlova, M., Zhang, K. L., Gineitis, A., Bradbury, E. M., et al. (2004). Novel human testis-specific histone H2B encoded by the interrupted gene on the X chromosome. *Genomics* 84, 745–756. doi: 10.1016/j.ygeno.2004.06.001

Couldrey, C., Carlton, M. B. L., Nolan, P. M., Colledge, W. H., and Evans, M. J. (1999). A retroviral gene trap insertion into the histone 3.3A gene causes partial neonatal lethality, stunted growth, neuromuscular deficits and male sub-fertility in transgenic mice. *Hum. Mol. Genet.* 8, 2489–2495. doi: 10.1093/hmg/8.13.2489

De Vries, M., Ramos, L., Housein, Z., and De Boer, P. (2012). Chromatin remodelling initiation during human spermiogenesis. *Biol. Open* 1, 446–457. doi: 10.1242/bio.2012844

Delucia, F., Faraonemennella, M. R., Derme, M., Quesada, P., Caiafa, P., and Farina, B. (1994). Histone-induced condensation of rat testis chromatin—testis-specific H1t versus somatic H1 variants. *Biochem. Biophys. Res. Commun.* 198, 32–39. doi: 10.1006/bbrc.1994.1005

Deng, W., and Lin, H. F. (2002). Miwi, a murine homolog of Piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Dev. Cell* 2, 819–830. doi: 10.1016/S1534-5807(02)00165-X

Dhar, S., Thota, A., and Rao, M. R. S. (2012). Insights into role of bromodomain, testis-specific (Brdt) in acetylated histone H4-dependent chromatin remodeling in mammalian spermiogenesis. *J. Biol. Chem.* 287, 6387–6405. doi: 10.1074/jbc.M111.288167

Dong, Y., Isono, K., Ohbo, K., Endo, T. A., Ohara, O., Maekawa, M., et al. (2017). EPC1/TIP60-mediated histone acetylation facilitates spermiogenesis in mice. *Mol. Cell. Biol.* 37, e00082-17. doi: 10.1128/MCB.00082-17

Dottermusch-Heidel, C., Klaus, E. S., Gonzalez, N. H., Bhushan, S., Meinhardt, A., Bergmann, M., et al. (2014). H3K79 methylation directly precedes the histone-to-protamine transition in mammalian spermatids and is sensitive to bacterial infections. *Andrology* 2, 655–665. doi: 10.1111/j.2047-2927.2014.00248.x

Doyon, Y., Selleck, W., Lane, W. S., Tan, S., and Cote, J. (2004). Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. *Mol. Cell. Biol.* 24, 1884–1896. doi: 10.1128/MCB.24.5.1884-1896.2004

Drabent, B., Benavente, R., and Hoyer-Fender, S. (2003). Histone H1t is not replaced by H1.1 or H1.2 in pachytene spermatocytes or spermatids of H1t-deficient mice. *Cytogenet. Genome Res.* 103, 307–313. doi: 10.1159/000076818

Drabent, B., Bode, C., Bramlage, B., and Doenecke, D. (1996). Expression of the mouse testicular histone gene H1t during spermatogenesis. *Histochem. Cell Biol.* 106, 247–251. doi: 10.1007/BF02484408

Drabent, B., Saftig, P., Bode, C., and Doenecke, D. (2000). Spermatogenesis proceeds normally in mice without linker histone H1t. *Histochem. Cell Biol.* 113, 433–442.

Fantz, D. A., Hatfield, W. R., Horvath, G., Kistler, M. K., and Kistler, W. S. (2001). Mice with a targeted disruption of the H1t gene are fertile and undergo normal changes in structural chromosomal proteins during spermiogenesis. *Biol. Reprod.* 64, 425–431. doi: 10.1095/biolreprod64.2.425

Gagne, J. P., Hendzel, M. J., Droit, A., and Poirier, G. G. (2006). The expanding role of poly(ADP-ribose) metabolism: current challenges and new perspectives. *Curr. Opin. Cell Biol.* 18, 145–151. doi: 10.1016/j.ceb.2006.02.013

Gaucher, J., Boussouar, F., Montellier, E., Curtet, S., Buchou, T., Bertrand, S., et al. (2012). Bromodomain-dependent stage-specific male genome programming by Brdt. *EMBO J.* 31, 3809–3820. doi: 10.1038/emboj.2012.233

Gibson, B. A., and Kraus, W. L. (2012). New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs. *Nat. Rev. Mol. Cell Biol.* 13, 411–424. doi: 10.1038/nrm3376

Godmann, M., Auger, V., Ferraroni-Aguiar, V., Di Sauro, A., Sette, C., Behr, R., et al. (2007). Dynamic regulation of histone h3 methylation at lysine 4 in mammalian spermatogenesis. *Biol. Reprod.* 77, 754–764. doi: 10.1095/biolreprod.107.062265

Gou, L. T., Kang, J. Y., Dai, P., Wang, X., Li, F., Zhao, S., et al. (2017). Ubiquitination-deficient mutations in human Piwi cause male infertility by impairing histone-to-protamine exchange during spermiogenesis. *Obstet. Gynecol. Survey* 72, 540–541. doi: 10.1097/OGX.0000000000000482

Govin, J., Caron, C., Lestrat, C., Rousseaux, S., and Khochbin, S. (2004). The role of histones in chromatin remodelling during mammalian spermiogenesis. *Eur. J. Biochem.* 271, 3459–3469. doi: 10.1111/j.1432-1033.2004.04266.x

Govin, J., Dorsey, J., Gaucher, J., Rousseaux, S., Khochbin, S., and Berger, S. L. (2010). Systematic screen reveals new functional dynamics of histones H3 and H4 during gametogenesis. *Genes Dev.* 24, 1772–1786. doi: 10.1101/gad.1954910

Govin, J., Escoffier, E., Rousseaux, S., Kuhn, L., Ferro, M., Thevenon, J., et al. (2007). Pericentric heterochromatin reprogramming by new histone variants during mouse spermiogenesis. *J. Cell Biol.* 176, 283–294. doi: 10.1083/jcb.200604141

Govin, J., Lestrat, C., Caron, C., Pivot-Pajot, C., Rousseaux, S., and Khochbin, S. (2006). Histone acetylation-mediated chromatin compaction during mouse spermatogenesis. *Cancer Ther.* 57, 155–15+. doi: 10.1007/3-540-37633-X_9

Grimes, S. R., and Henderson, N. (1984a). Acetylation of rat testis histones H2b and Th2b. *Dev. Biol.* 101, 516–521. doi: 10.1016/0012-1606(84)90165-9

Grimes, S. R. Jr., and Henderson, N. (1984b). Hyperacetylation of histone H4 in rat testis spermatids. *Exp. Cell Res.* 152, 91–97. doi: 10.1016/0014-4827(84)90232-5

Hada, M., Masuda, K., Yamaguchi, K., Shirahige, K., and Okada, Y. (2017). Identification of a variant-specific phosphorylation of TH2A during spermiogenesis. *Sci. Rep.* 7, 1–13. doi: 10.1038/srep46228

Hao, S. L., Ni, F. D., and Yang, W. X. (2019). The dynamics and regulation of chromatin remodeling during spermiogenesis. *Gene* 706, 201–210. doi: 10.1016/j.gene.2019.05.027

Happel, N., and Doenecke, D. (2009). Histone H1 and its isoforms: contribution to chromatin structure and function. *Gene* 431, 1–12. doi: 10.1016/j.gene.2008.11.003

Hazzouri, M., Pivot-Pajot, C., Faure, A. K., Usson, Y., Pelletier, R., Sele, B., et al. (2000). Regulated hyperacetylation of core histones during mouse

spermatogenesis: involvement of histone-deacetylases. *Eur. J. Cell Biol.* 79, 950–960. doi: 10.1078/0171-9335-00123

Hershko, A., and Ciechanover, A. (1998). The ubiquitin system. *Annu. Rev. Biochem.* 67, 425–479. doi: 10.1146/annurev.biochem.67.1.425

Hess, R. A., and Renato De Franca, L. (2008). Spermatogenesis and cycle of the seminiferous epithelium. *Adv. Exp. Med. Biol.* 636, 1–15. doi: 10.1007/978-0-387-09597-4_1

Hoghoughi, N., Barral, S., Vargas, A., Rousseaux, S., and Khochbin, S. (2018). Histone variants: essential actors in male genome programming. *J. Biochem.* 163, 97–103. doi: 10.1093/jb/mvx079

Hou, X., Zhang, W., Xiao, Z., Gan, H., Lin, X., Liao, S., et al. (2012). Mining and characterization of ubiquitin E3 ligases expressed in the mouse testis. *BMC Genomics* 13, 495. doi: 10.1186/1471-2164-13-495

Jha, K. N., Tripurani, S. K., and Johnson, G. R. (2017). TSSK6 is required for gamma H2AX formation and the histone-to-protamine transition during spermiogenesis. *J. Cell Sci.* 130, 1835–1844. doi: 10.1242/jcs.202721

Juliano, C., Wang, J. Q., and Lin, H. F. (2011). Uniting germline and stem cells: the function of Piwi proteins and the piRNA pathway in diverse organisms. *Ann. Rev. Genet.* 45, 447–469. doi: 10.1146/annurev-genet-110410-132541

Kan, P. Y., Caterino, T. L., and Hayes, J. J. (2009). The H4 tail domain participates in intra- and internucleosome interactions with protein and DNA during folding and oligomerization of nucleosome arrays. *Mol. Cell. Biol.* 29, 538–546. doi: 10.1128/MCB.01343-08

Ketchum, C. C., Larsen, C. D., Mcneil, A., Meyer-Ficca, M. L., and Meyer, R. G. (2018). Early histone H4 acetylation during chromatin remodeling in equine spermatogenesis. *Biol. Reprod.* 98, 115–129. doi: 10.1093/biolre/iox159

Khadake, J. R., and Rao, M. R. S. (1995). DNA-condensing and chromatin-condensing properties of rat testes Hla and hit compared to those of rat-liver Hlbdec—Hlt is a poor condenser of chromatin. *Biochemistry* 34, 15792–15801. doi: 10.1021/bi00048a025

Khor, B., Bredemeyer, A. L., Huang, C. Y., Turnbull, I. R., Evans, R., Maggi, L. B., et al. (2006). Proteasome activator PA200 is required for normal spermatogenesis. *Mol. Cell. Biol.* 26, 2999–3007. doi: 10.1128/MCB.26.8.2999-3007.2006

Kimura, S., and Loppin, B. (2016). The *Drosophila* chromosomal protein Mst77F is processed to generate an essential component of mature sperm chromatin. *Open Biol.* 6, 1–12. doi: 10.1098/rsob.160207

Kolthur-Seetharam, U., Pradeepa, M. M., Gupta, N., Narayanaswamy, R., and Rao, M.R.S. (2009). Spatiotemporal organization of AT- and GC-rich DNA and their association with transition proteins TP1 and TP2 in rat condensing spermatids. *J. Histochem. Cytochem.* 57, 951–962. doi: 10.1369/jhc.2009.953414

Komander, D., and Rape, M. (2012). The ubiquitin code. *Ann. Rev. Biochem.* 81, 203–229. doi: 10.1146/annurev-biochem-060310-170328

Kowalski, A., and Palyga, J. (2012). Linker histone subtypes and their allelic variants. *Cell Biol. Int.* 36, 981–996. doi: 10.1042/CBI20120133

Krishnamoorthy, T., Chen, X., Govin, J., Cheung, W. L., Dorsey, J., Schindler, K., et al. (2006). Phosphorylation of histone H4 Ser1 regulates sporulation in yeast and is conserved in fly and mouse spermatogenesis. *Genes Dev.* 20, 2580–2592. doi: 10.1101/gad.1457006

Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T. W., Isobe, T., Asada, N., Fujita, Y., et al. (2004). Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* 131, 839–849. doi: 10.1242/dev.00973

Lee, J., Park, H. S., Kim, H. H., Yun, Y. J., Lee, D. R., and Lee, S. (2009). Functional polymorphism in H2BFWT-5'UTR is associated with susceptibility to male infertility. *J. Cell. Mol. Med.* 13, 1942–1951. doi: 10.1111/j.1582-4934.2009.00830.x

Levesque, D., Veilleux, S., Caron, N., and Boissonneault, G. (1998). Architectural DNA-binding properties of the spermatidal transition proteins 1 and 2. *Biochem. Biophys. Res. Commun.* 252, 602–609. doi: 10.1006/bbrc.1998.9687

Li, A., Eirin-Lopez, J., and Ausio, J. (2005). H2AX: tailoring histone H2A for chromatin-dependent genomic integrity. *Biochem. Cell Biol.* 83, 505–515. doi: 10.1139/o05-114

Liu, C., Song, Z. H., Wang, L. N., Yu, H. Y., Liu, W. X., Shang, Y. L., et al. (2017a). Sirt1 regulates acrosome biogenesis by modulating autophagic flux during spermiogenesis in mice. *Development* 144, 441–451. doi: 10.1242/dev.147074

Liu, S. M., Yu, H. J., Liu, Y. Q., Liu, X. H., Zhang, Y., Bu, C., et al. (2017b). Chromodomain protein CDYL acts as a crotonyl-CoA hydratase to regulate histone crotonylation and spermatogenesis. *Mol. Cell* 67, 853–85+. doi: 10.1016/j.molcel.2017.07.011

Liu, Z. L., Zhou, S. L., Liao, L., Chen, X., Meistrich, M., and Xu, J. M. (2010). Jmjd1a Demethylase-regulated histone modification is essential for cAMP-response element modulator-regulated gene expression and spermatogenesis. *J. Biol. Chem.* 285, 2758–2770. doi: 10.1074/jbc.M109.066845

Lu, L. Y., Wu, J. X., Ye, L., Gavrilina, G. B., Saunders, T. L., and Yu, X. C. (2010). RNF8-dependent histone modifications regulate nucleosome removal during spermatogenesis. *Dev. Cell* 18, 371–384. doi: 10.1016/j.devcel.2010.01.010

Luense, L. J., Wang, X. S., Schon, S. B., Weller, A. H., Shiao, E. L., Bryant, J. M., et al. (2016). Comprehensive analysis of histone post-translational modifications in mouse and human male germ cells. *Epigenet. Chromatin* 9, 1–15. doi: 10.1186/s13072-016-0072-6

Ma, T., Keller, J. A., and Yu, X. C. (2011). RNF8-dependent histone ubiquitination during DNA damage response and spermatogenesis. *Acta Biochim. Biophys. Sin.* 43, 339–345. doi: 10.1093/abbs/gmr016

Manterola, M., Brown, T. M., Oh, M. Y., Garyn, C., Gonzalez, B. J., and Wolgemuth, D. J. (2018). BRDT is an essential epigenetic regulator for proper chromatin organization, silencing of sex chromosomes and crossover formation in male meiosis. *PLoS Genet.* 14, 1–30. doi: 10.1371/journal.pgen.1007209

Martianov, I., Brancorsini, S., Catena, R., Gansmuller, A., Kotaja, N., Parvinen, M., et al. (2005). Polar nuclear localization of H1T2, a histone H1 variant, required for spermatid elongation and DNA condensation during spermiogenesis. *Proc. Natl Acad. Sci. U. S. A.* 102, 2808–2813. doi: 10.1073/pnas.0406060102

Mccarrey, J. R., Geyer, C. B., and Yoshioka, H. (2005). Epigenetic regulation of testis-specific gene expression. *Testicular Cell Dyn. Endocr. Signal.* 1061, 226–242. doi: 10.1196/annals.1336.025

Meistrich, M. L., Bucci, L. R., Trostleweige, P. K., and Brock, W. A. (1985). Histone variants in rat spermatogonia and primary spermatocytes. *Dev. Biol.* 112, 230–240. doi: 10.1016/0012-1606(85)90137-X

Meistrich, M. L., Mohapatra, B., Shirley, C. R., and Zhao, M. (2003). Roles of transition nuclear proteins in spermiogenesis. *Chromosoma* 111, 483–488. doi: 10.1007/s00412-002-0227-z

Meng, C., Liao, J., Zhao, D., Huang, H., Qin, J., Lee, T. L., et al. (2019). L3MBTL2 regulates chromatin remodeling during spermatogenesis. *Cell Death Differ.* 13, 1–14. doi: 10.1038/s41418-019-0283-z

Messner, S., and Hottiger, M. O. (2011). Histone ADP-ribosylation in DNA repair, replication and transcription. *Trends Cell Biol.* 21, 534–542. doi: 10.1016/j.tcb.2011.06.001

Meyer-Ficca, M. L., Ihara, M., Bader, J. J., Leu, N. A., Beneke, S., and Meyer, R. G. (2015). Spermatid head elongation with normal nuclear shaping requires ADP-ribosyltransferase PARP11 (ARTD11) in Mice. *Biol. Reprod.* 92, 1–13. doi: 10.1095/biolreprod.114.123661

Meyer-Ficca, M. L., Ihara, M., Lonchar, J. D., Meistrich, M. L., Austin, C. A., Min, W., et al. (2011). Poly(ADP-ribose) metabolism is essential for proper nucleoprotein exchange during mouse spermiogenesis. *Biol. Reprod.* 84, 218–228. doi: 10.1095/biolreprod.110.087361

Meyer-Ficca, M. L., Lonchar, J., Credidio, C., Ihara, M., Li, Y., Wang, Z. Q., et al. (2009). Disruption of poly(ADP-ribose) homeostasis affects spermiogenesis and sperm chromatin integrity in mice. *Biol. Reprod.* 81, 46–55. doi: 10.1095/biolreprod.108.075390

Meyer-Ficca, M. L., Scherthan, H., Burkle, A., and Meyer, R. G. (2005). Poly(ADP-ribosyl)ation during chromatin remodeling steps in rat spermiogenesis. *Chromosoma* 114, 67–74. doi: 10.1007/s00412-005-0344-6

Mishra, L. N., Gupta, N., and Rao, S. M. R. (2015). Mapping of post-translational modifications of spermatid-specific linker histone H1-like protein, HILS1. *J. Proteomics* 128, 218–230. doi: 10.1016/j.jprot.2015.08.001

Mishra, L. N., Shalini, V., Gupta, N., Ghosh, K., Suthar, N., Bhaduri, U., et al. (2018). Spermatid-specific linker histone HILS1 is a poor condenser of DNA and chromatin and preferentially associates with LINE-1 elements. *Epigenet. Chromatin* 11, 1–21. doi: 10.1186/s13072-018-0214-0

Montellier, E., Boussouar, F., Rousseaux, S., Zhang, K., Buchou, T., Fenaille, F., et al. (2013). Chromatin-to-nucleoprotamine transition is controlled by the histone H2B variant TH2B. *Genes Dev.* 27, 1680–1692. doi: 10.1101/gad.220095.113

Moss, S. B., and Orth, J. M. (1993). Localization of a spermatid-specific histone-2b protein in mouse spermiogenic cells. *Biol. Reprod.* 48, 1047–1056. doi: 10.1095/biolreprod48.5.1047

Nair, M., Nagamori, I., Sun, P., Mishra, D. P., Rheaume, C., Li, B., et al. (2008). Nuclear regulator Pygo2 controls spermiogenesis and histone H3 acetylation. *Dev. Biol.* 320, 446–455. doi: 10.1016/j.ydbio.2008.05.553

Okada, Y., Scott, G., Ray, M. K., Mishina, Y., and Zhang, Y. (2007). Histone demethylase JHDM2A is critical for Tnp1 and Prm1 transcription and spermatogenesis. *Nature* 450, 119–11+. doi: 10.1038/nature06236

Oliva, R., Bazett-Jones, D., Mezquita, C., and Dixon, G. H. (1987). Factors affecting nucleosome disassembly by protamines *in vitro.* Histone hyperacetylation and chromatin structure, time dependence, and the size of the sperm nuclear proteins. *J. Biol. Chem.* 262, 17016–17025.

Oliva, R., and Mezquita, C. (1986). Marked differences in the ability of distinct protamines to disassemble nucleosomal core particles in vitro. *Biochemistry* 25, 6508–6511. doi: 10.1021/bi00369a025

Padavattan, S., Shinagawa, T., Hasegawa, K., Kumasaka, T., Ishii, S., and Kumarevel, T. (2015). Structural and functional analyses of nucleosome complexes with mouse histone variants TH2a and TH2b, involved in reprogramming. *Biochem. Biophys. Res. Commun.* 464, 929–935. doi: 10.1016/j.bbrc.2015.07.070

Padavattan, S., Thiruselvam, V., Shinagawa, T., Hasegawa, K., Kumasaka, T., Ishii, S., et al. (2017). Structural analyses of the nucleosome complexes with human testis-specific histone variants, hTh2a and hTh2b. *Biophys. Chem.* 221, 41–48. doi: 10.1016/j.bpc.2016.11.013

Pentakota, S. K., Sandhya, S., Sikarwar, A. P., Chandra, N., and Rao, M. R. S. (2014). Mapping post-translational modifications of mammalian testicular specific histone variant TH2B in tetraploid and haploid germ cells and their implications on the dynamics of nucleosome structure. *J. Proteome Res.* 13, 5603–5617. doi: 10.1021/pr500597a

Pickart, C. M. (2001). Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* 70, 503–533. doi: 10.1146/annurev.biochem.70.1.503

Pivot-Pajot, C., Caron, C., Govin, M., Vion, A., Rousseaux, S., and Khochbin, S. (2003). Acetylation-dependent chromatin reorganization by BRDT, a testis-specific bromodomain-containing protein. *Mol. Cell. Biol.* 23, 5354–5365. doi: 10.1128/MCB.23.15.5354-5365.2003

Qian, M. X., Pang, Y., Liu, C. H., Haratake, K., Du, B. Y., Ji, D. Y., et al. (2013). Acetylation-mediated proteasomal degradation of core histones during DNA repair and spermatogenesis. *Cell* 153, 1012–1024. doi: 10.1016/j.cell.2013.04.032

Rafatmanesh, A., Nikzad, H., Ebrahimi, A., Karimian, M., and Zamani, T. (2018). Association of the c.-9C > T and c.368A > G transitions in H2BFWT gene with male infertility in an Iranian population. *Andrologia* 50, 1–6. doi: 10.1111/and.12805

Raja, S. J., and Renkawitz-Pohl, R. (2005). Replacement by *Drosophila melanogaster* protamines and Mst77F of histones during chromatin condensation in late spermatids and role of sesame in the removal of these proteins from the male pronucleus. *Mol. Cell. Biol.* 25, 6165–6177. doi: 10.1128/MCB.25.14.6165-6177.2005

Rathke, C., Baarends, W. M., Awe, S., and Renkawitz-Pohl, R. (2014). Chromatin dynamics during spermiogenesis. *Biochim. Biophys. Acta* 1839, 155–168. doi: 10.1016/j.bbagrm.2013.08.004

Roosen-Runge, E. C. (1962). The process of spermatogenesis in mammals. *Biol. Rev. Camb Philos. Soc.* 37, 343–377. doi: 10.1111/j.1469-185X.1962.tb01616.x

Rossetto, D., Avvakumov, N., and Cote, J. (2012). Histone phosphorylation A chromatin modification involved in diverse nuclear events. *Epigenetics* 7, 1098–1108. doi: 10.4161/epi.21975

Sabari, B. R., Zhang, D., Allis, C. D., and Zhao, Y. M. (2017). Metabolic regulation of gene expression through histone acylations. *Nat. Rev. Mol. Cell Biol.* 18, 90–101. doi: 10.1038/nrm.2016.140

Sarg, B., Chwatal, S., Talasz, H., and Lindner, H. H. (2009). Testis-specific linker histone H1t is multiply phosphorylated during spermatogenesis identification of phosphorylation siTES. *J. Biol. Chem.* 284, 3610–3618. doi: 10.1074/jbc.M805925200

Shang, E. Y., Nickerson, H. D., Wen, D. C., Wang, X. Y., and Wolgemuth, D. J. (2007). The first bromodomain of Brdt, a testis-specific member of the BET sub-family of double-bromodomain-containing proteins, is essential for male germ cell differentiation. *Development* 134, 3507–3515. doi: 10.1242/dev.004481

Shinagawa, T., Huynh, L. M., Takagi, T., Tsukamoto, D., Tomaru, C., Kwak, H. G., et al. (2015). Disruption of Th2a and Th2b genes causes defects in spermatogenesis. *Development* 142, 1287–1292. doi: 10.1242/dev.121830

Shiraishi, K., Shindo, A., Harada, A., Ohkawa, Y., Kurumizaka, H., Kimura, H., et al. (2017). Roles of histone H3.5 in human spermatogenesis and spermatogenic disorders. *J. Urol.* 197, E85–E85. doi: 10.1016/j.juro.2017.02.277

Shires, A., Carpenter, M. P., and Chalkley, R. (1975). New histones found in mature mammalian testes. *Proc. Natl. Acad. Sci. U. S. A.* 72, 2714–2718. doi: 10.1073/pnas.72.7.2714

Shires, A., Carpenter, M. P., and Chalkley, R. (1976). Cysteine-containing H2b-like histone found in mature mammalian testis. *J. Biol. Chem.* 251, 4155–4158.

Shirley, C. R., Hayashi, S., Mounsey, S., Yanagimachi, R., and Meistrich, M. L. (2004). Abnormalities and reduced reproductive potential of sperm from Tnp1- and Tnp2-null double mutant mice. *Biol. Reprod.* 71, 1220–1229. doi: 10.1095/biolreprod.104.029363

Shogren-Knaak, M., Ishii, H., Sun, J. M., Pazin, M. J., Davie, J. R., and Peterson, C. L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* 311, 844–847. doi: 10.1126/science.1124000

Singh, J., and Rao, M. R. S. (1988). Interaction of rat testis protein, Tp, with nucleosome core particle. *Biochem. Int.* 17, 701–710.

Siomi, M. C., Sato, K., Pezic, D., and Aravin, A. A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* 12, 246–258. doi: 10.1038/nrm3089

Soboleva, T. A., Nekrasov, M., Pahwa, A., Williams, R., Huttley, G. A., and Tremethick, D. J. (2012). A unique H2A histone variant occupies the transcriptional start site of active genes. *Nat. Struct. Mol. Biol.* 19, 25–U37. doi: 10.1038/nsmb.2161

Soboleva, T. A., Parker, B. J., Nekrasov, M., Hart-Smith, G., Tay, Y. J., Tng, W. Q., et al. (2017). A new link between transcriptional initiation and pre-mRNA splicing: the RNA binding histone variant H2A.B. *PloS Genet.* 13, 1–31. doi: 10.1371/journal.pgen.1006633

Song, N., Liu, J., An, S. C., Nishino, T., Hishikawa, Y., and Koji, T. (2011). Immunohistochemical analysis of histone H3 modifications in germ cells during mouse spermatogenesis. *Acta Histochem. Cytochem.* 44, 183–190. doi: 10.1267/ahc.11027

Spiridonov, N. A., Wong, L., Zerfas, P. M., Starost, M. F., Pack, S. D., Paweletz, C. P., et al. (2005). Identification and characterization of SSTK, a serine/threonine protein kinase essential for male fertility. *Mol. Cell. Biol.* 25, 4250–4261. doi: 10.1128/MCB.25.10.4250-4261.2005

Stadtmueller, B. M., and Hill, C. P. (2011). Proteasome Activators. *Mol. Cell* 41, 8–19. doi: 10.1016/j.molcel.2010.12.020

Tachiwana, H., Kagawa, W., Osakabe, A., Kawaguchi, K., Shiga, T., Hayashi-Takanaka, Y., et al. (2010). Structural basis of instability of the nucleosome containing a testis-specific histone variant, human H3T. *Proc. Natl. Acad. Sci. U. S. A.* 107, 10454–10459. doi: 10.1073/pnas.1003064107

Talbert, P. B., and Henikoff, S. (2010). Histone variants - ancient wrap artists of the epigenome. *Nat. Rev. Mol. Cell Biol.* 11, 264–275. doi: 10.1038/nrm2861

Tan, M. J., Luo, H., Lee, S., Jin, F. L., Yang, J. S., Montellier, E., et al. (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* 146, 1015–1027. doi: 10.1016/j.cell.2011.08.008

Tanaka, H., Iguchi, N., Isotani, A., Kitamura, K., Toyama, Y., Matsuoka, Y., et al. (2005). HANP1/H1T2, a novel histone H1-like protein involved in nuclear formation and sperm fertility. *Mol. Cell. Biol.* 25, 7107–7119. doi: 10.1128/MCB.25.16.7107-7119.2005

Tang, M. C. W., Jacobs, S. A., Mattiske, D. M., Soh, Y. M., Graham, A. N., Tran, A., et al. (2015). Contribution of the two genes encoding histone variant H3.3 to viability and fertility in mice. *PloS Genet.* 11, 1–23. doi: 10.1371/journal.pgen.1004964

Teimouri, M., Najaran, H., Hosseinzadeh, A., and Mazoochi, T. (2018). Association between two common transitions of H2BFWT gene and male infertility: a case-control, meta, and structural analysis. *Andrology* 6, 306–316. doi: 10.1111/andr.12464

Thakar, A., Gupta, P., Ishibashi, T., Finn, R., Silva-Moreno, B., Uchiyama, S., Fukui, et al. (2009). H2A.Z and H3.3 histone variants affect nucleosome structure: biochemical and biophysical studies. *Biochemistry* 48, 10852–10857. doi: 10.1021/bi901129e

Trostleweige, P. K., Meistrich, M. L., Brock, W. A., Nishioka, K., and Bremer, J. W. (1982). Isolation and characterization of Th2a, a germ cell-specific variant of histone-2a in rat testis. *J. Biol. Chem.* 257, 5560–5567.

Tse, C., Sera, T., Wolffe, A. P., and Hansen, J. C. (1998). Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of

nucleosomal arrays by RNA polymerase III. *Mol. Cell. Biol.* 18, 4629–4638. doi: 10.1128/MCB.18.8.4629

Ueda, J., Harada, A., Urahama, T., Machida, S., Maehara, K., Hada, M., et al. (2017). Testis-specific histone variant H3t gene is essential for entry into spermatogenesis. *Cell Rep.* 18, 593–600. doi: 10.1016/j.celrep.2016.12.065

Unni, E., Zhang, Y., Kangasniemi, M., Saperstein, W., Moss, S. B., and Meistrich, M. L. (1995). Stage-specific distribution of the spermatid-specific histone 2b in the rat testis. *Biol. Reprod.* 53, 820–826. doi: 10.1095/biolreprod53.4.820

Urahama, T., Harada, A., Maehara, K., Horikoshi, N., Sato, K., Sato, Y., et al.(2016). Histone H3.5 forms an unstable nucleosome and accumulates around transcription start sites in human testis. *Epigenet. Chromatin* 9, 1–16. doi: 10.1186/s13072-016-0051-y

Urahama, T., Horikoshi, N., Osakabe, A., Tachiwana, H., and Kurumizaka, H. (2014). Structure of human nucleosome containing the testis-specific histone variant TSH2B. *Acta Crystallogr. Sect. F-Struct. Biol. Commun.* 70, 444–449. doi: 10.1107/S2053230X14004695

Ushijima, Y., Inoue, Y. H., Konishi, T., Kitazawa, D., Yoshida, H., Shimaji, K., et al. (2012). Roles of histone H3K9 methyltransferases during *Drosophila* spermatogenesis. *Chrom. Res.* 20, 319–331. doi: 10.1007/s10577-012-9276-1

Ustrell, V., Pratt, G., Gorbea, C., and Rechsteiner, M. (2005). Purification and assay of proteasome activator PA200. *Methods Enzymol.* 398, 321–329. doi: 10.1016/S0076-6879(05)98026-9

Van Der Heijden, G. W., Derijck, A. a. H. A., Posfai, E., Giele, M., Pelczar, P., Ramos, L., et al. (2007). Chromosome-wide nucleosome replacement and H3.3 incorporation during mammalian meiotic sex chromosome inactivation. *Nat. Genet.* 39, 251–258. doi: 10.1038/ng1949

Wang, X., Kang, J. Y., Wei, L., Yang, X., Sun, H., Yang, S., et al. (2019). PHF7 is a novel histone H2A E3 ligase prior to histone-to-protamine exchange during spermiogenesis. *Development* 146, 1–11. doi: 10.1242/dev.175547

Welchman, R. L., Gordon, C., and Mayer, R. J. (2005). Ubiquitin and ubiquitin-like proteins as multifunctional signals. *Nat. Rev. Mol. Cell. Biol.* 6, 599–609. doi: 10.1038/nrm1700

Wendt, K. D., and Shilatifard, A. (2006). Packing for the germy: the role of histone H4 Ser1 phosphorylation in chromatin compaction and germ cell development. *Genes Dev.* 20, 2487–2491. doi: 10.1101/gad.1477706

Wu, J. Y., Ribar, T. J., Cummings, D. E., Burton, K. A., Mcknight, G. S., and Means, A. R. (2000). Spermiogenesis and exchange of basic nuclear proteins are impaired in male germ cells lacking Camk4. *Nat. Genet.* 25, 448–452. doi: 10.1038/78153

Yan, W., Ma, L., Burns, K. H., and Matzuk, M. M. (2003). HILS1 is a spermatid-specific linker histone H1-like protein implicated in chromatin remodeling during mammalian spermiogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 100, 10546–10551. doi: 10.1073/pnas.1837812100

Ying, H. Q., Scott, M. B., and Zhou-Cun, A. (2012). Relationship of SNP of H2BFWT gene to male infertility in a Chinese population with idiopathic spermatogenesis impairment. *Biomarkers* 17, 402–406. doi: 10.3109/1354750X.2012.677066

Yoshida, K., Muratani, M., Araki, H., Miura, F., Suzuki, T., Dohmae, N., et al. (2018). Mapping of histone-binding sites in histone replacement-completed spermatozoa. *Nat. Commun.* 9, 1–11. doi: 10.1038/s41467-018-06243-9

Yu, Y. E., Zhang, Y., Unni, E., Shirley, C. R., Deng, J. M., Russell, L. D., et al. (2000). Abnormal spermatogenesis and reduced fertility in transition nuclear protein 1-deficient mice. *Proc. Natl. Acad. Sci. U. S. A.* 97, 4683–4688. doi: 10.1073/pnas.97.9.4683

Yuen, B.T.K., Bush, K. M., Barrilleaux, B. L., Cotterman, R., and Knoepfler, P. S. (2014). Histone H3.3 regulates dynamic chromatin states during spermatogenesis. *Development* 141, 3483–3494. doi: 10.1242/dev.106450

Zhang, Z. H., Mu, S. M., Guo, M. S., Wu, J. L., Li, Y. Q., Zhang, H., et al. (2016). Dynamics of histone H2A, H4 and HS1ph during spermatogenesis with a focus on chromatin condensation and maturity of spermatozoa. *Sci. Rep.* 6, 1–11. doi: 10.1038/srep25089

Zhao, M., Shirley, C. R., Mounsey, S., and Meistrich, M. L. (2004). Nucleoprotein transitions during spermiogenesis in mice with transition nuclear protein Tnp1 and Tnp2 mutations. *Biol. Reprod.* 71, 1016–1025. doi: 10.1095/biolreprod.104.028191

Zhao, M., Shirley, C. R., Yu, Y. E., Mohapatra, B., Zhang, Y., Unni, E., et al. (2001). Targeted disruption of the transition protein 2 gene affects sperm chromatin structure and reduces fertility in mice. *Mol. Cell Biol.* 21, 7243–7255. doi: 10.1128/MCB.21.21.7243-7255.2001

Zuo, X. L., Rong, B. W., Li, L., Lv, R. T., Lan, F., and Tong, M. H. (2018). The histone methyltransferase SETD2 is required for expression of acrosin-binding protein 1 and protamines and essential for spermiogenesis in mice. *J. Biol. Chem.* 293, 9188–9197. doi: 10.1074/jbc.RA118.002851

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Wang, Gao, Li and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

REVIEW
published: 09 October 2019
doi: 10.3389/fgene.2019.00950

Epigenetic Biomarkers in Cardiovascular Diseases

Carolina Soler-Botija[1,2]*, *Carolina Gálvez-Montón*[1,2] *and Antoni Bayés-Genís*[1,2,3,4]

[1] *Heart Failure and Cardiac Regeneration (ICREC) Research Program, Health Science Research Institute Germans Trias i Pujol (IGTP), Badalona, Spain,* [2] *CIBERCV, Instituto de Salud Carlos III, Madrid, Spain,* [3] *Cardiology Service, HUGTiP, Badalona, Spain,* [4] *Department of Medicine, Barcelona Autonomous University (UAB), Badalona, Spain*

Cardiovascular diseases are the number one cause of death worldwide and greatly impact quality of life and medical costs. Enormous effort has been made in research to obtain new tools for efficient and quick diagnosis and predicting the prognosis of these diseases. Discoveries of epigenetic mechanisms have related several pathologies, including cardiovascular diseases, to epigenetic dysregulation. This has implications on disease progression and is the basis for new preventive strategies. Advances in methodology and big data analysis have identified novel mechanisms and targets involved in numerous diseases, allowing more individualized epigenetic maps for personalized diagnosis and treatment. This paves the way for what is called pharmacoepigenetics, which predicts the drug response and develops a tailored therapy based on differences in the epigenetic basis of each patient. Similarly, epigenetic biomarkers have emerged as a promising instrument for the consistent diagnosis and prognosis of cardiovascular diseases. Their good accessibility and feasible methods of detection make them suitable for use in clinical practice. However, multicenter studies with a large sample population are required to determine with certainty which epigenetic biomarkers are reliable for clinical routine. Therefore, this review focuses on current discoveries regarding epigenetic biomarkers and its controversy aiming to improve the diagnosis, prognosis, and therapy in cardiovascular patients.

Keywords: epigenetics, biomarker, microRNA, cardiovascular diseases, myocardial infarction, heart failure, atherosclerosis, hypertension

OPEN ACCESS

Edited by:
Yun Liu,
Fudan University, China

Reviewed by:
Daniel B. Lipka,
German Cancer Research Center (DKFZ), Germany
Jeffrey Mark Craig,
Murdoch Childrens Research Institute (MCRI), Australia

***Correspondence:**
Carolina Soler-Botija
csoler@igtp.cat

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *12 May 2019*
Accepted: *05 September 2019*
Published: *09 October 2019*

Citation:
Soler-Botija C, Gálvez-Montón C and Bayés-Genís A (2019) Epigenetic Biomarkers in Cardiovascular Diseases. Front. Genet. 10:950. doi: 10.3389/fgene.2019.00950

INTRODUCTION

Cardiovascular diseases (CVDs) are one of the leading causes of mortality in developed countries. Cardiovascular diseases refer to disorders affecting the structures or function of the heart and blood vessels, including hypertension, atherosclerosis, myocardial infarction (MI), ischemia/reperfusion injury, stroke, and heart failure (HF), among others (Wang et al., 2016a; Thomas et al.,

Abbreviations: AMI, acute myocardial infarction; ApoE, apolipoprotein E; BNP B-type natriuretic peptide; CK, creatine kinase; cTnI, cardiac troponin I; cTnT, cardiac troponin T; DOT1L, disruptor of telomeric silencing-1; ENaC, epithelial sodium channel; EZH2, enhancer of zeste homolog 2; GEO, Gene Expression Omnibus; HDAC, histone deacetylase; HF, heart failure; HFrEF, heart failure with reduced ejection fraction; HFpEF, heart failure with preserved ejection fraction; hs-cTnT, high-sensitivity cardiac troponin T; hs-CRP, high-sensitivity C-reactive protein; lncRNAs, long noncoding RNAs; LV, left ventricular; MI, myocardial infarction; miRNAs, microRNAs; ncRNAs, noncoding RNAs; NSTEMI, non- ST-segment elevation myocardial infarction; STEMI, ST-segment elevation myocardial infarction; pmiRNAs, platelet miRNAs; piRNAs, p-element-induced wimpy testis (PIWI)-interacting RNAs; tRNA, transfer RNA; ZEB1, zinc finger E-box binding homeobox 1.

2018). Mechanisms underlying the complex pathophysiology that leads to CVDs are of great interest but still far from clear. Progress in the field of epigenetics have opened a new world for the comprehension and management of human diseases, including the prevalence of CVDs, based on the role of genetics and its environmental interaction in pathological conditions (Jaenisch and Bird, 2003). Significant evidence suggests that the environment and lifestyle can define epigenetic patterns throughout life. These epigenetic patterns are a cellular memory of further environmental exposure. Epigenetic modifications are reversible, different among cell types, and can potentially lead to disease susceptibility by producing long-term changes in gene transcription (Fraga et al., 2005; Beekman et al., 2010).

Epigenetic modifications include DNA methylation and posttranslational modifications of histone tails. However, in this review, posttranscriptional regulation of gene expression by noncoding RNAs (ncRNAs) is also considered a part of the epigenetic machinery. MicroRNAs (miRNAs) are small ncRNAs that contribute to regulation of the expression of different epigenetic regulators such as DNA methyltransferases (DNMTs) and histone deacetylases (HDACs), among others. Similarly, DNA methylation and histone modifications can regulate the expression of some miRNAs, forming a feedback loop. Thus, miRNAs and epigenetic regulators cooperate to modulate the expression of mutual targets. Therefore, although miRNAs are not strictly considered epigenetic factors, they contribute to the modulation of gene expression through epigenetics. Disruption of this complex regulation may participate in the development of different diseases (Iorio et al., 2010; Hoareau-Aveilla and Meggetto, 2017; Moutinho and Esteller, 2017; Wang et al., 2017a) (**Figure 1**). DNA and histone proteins comprise the chromatin, which can be remodeled into a tightly condensed state (heterochromatin) or an open conformation (euchromatin) that would allow access to transcription factors or DNA binding proteins, allowing the regulation of gene expression (Kouzarides, 2007). Thus, epigenetics involves changes in gene expression due to chromatin adjustments that change the accessibility of DNA without changing its sequence, leading to silencing or downregulation/upregulation of gene expression (Baccarelli et al., 2010). Chromatin modifications, such as DNA methylation, consist of the transfer of a methyl group to carbon 5 of the cytosine residues [5-methylcytosine (5mC)] in CpG dinucleotides sites. CpG dinucleotides are localized throughout the genome but are more abundant in certain regions, such as gene promoters, forming so-called CpG islands. CpG methylation causes transcriptional repression by directly blocking transcription factor access to the DNA or indirectly *via* chromatin-modifying proteins (methyl-binding proteins) that recognize the methylated regions and recruit corepressors. DNA methyltransferases catalyze DNA methylation by recognizing

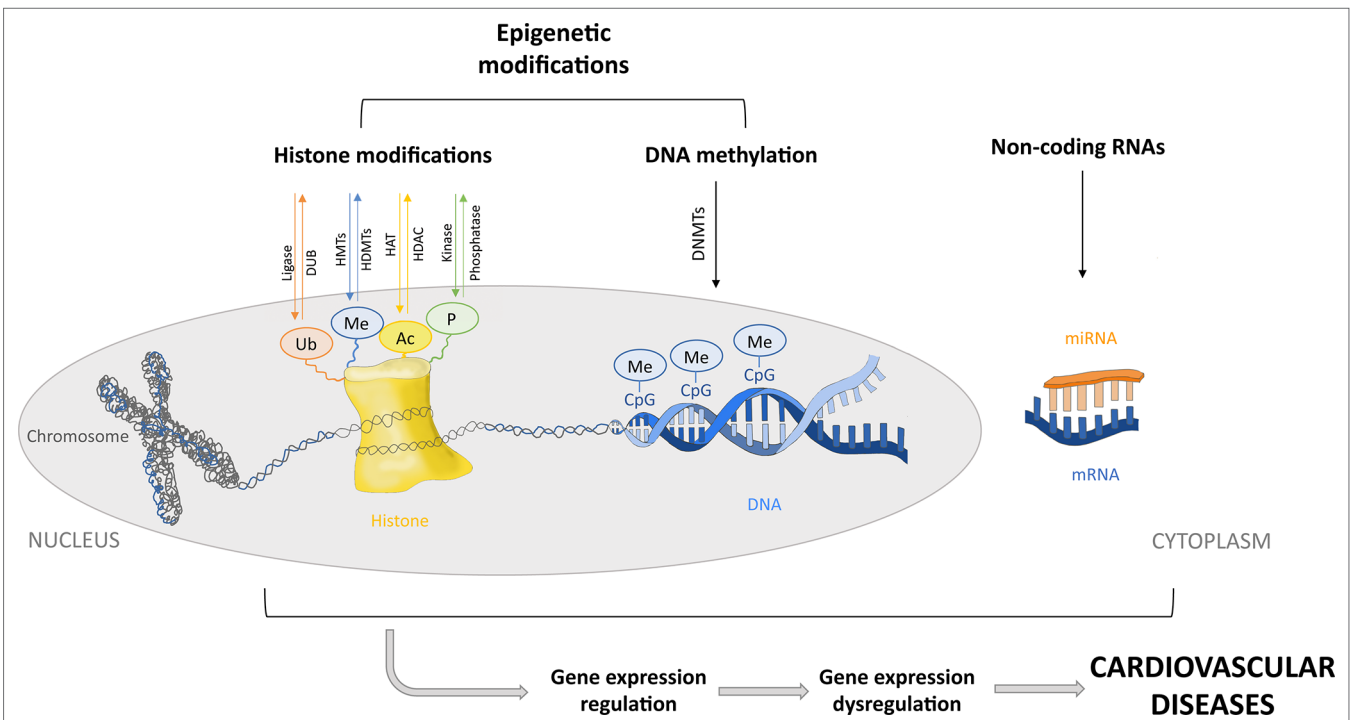


FIGURE 1 | Epigenetic regulatory mechanisms. Posttranslational modifications of histone tails by acetylation, deacetylation, ubiquitination, methylation, and phosphorylation. DNA methylation by DNA methyltransferases (DNMTs). Posttranscriptional regulation of gene expression by microRNAs. Epigenetic modifications involve silencing or downregulation/upregulation of gene expression. Dysregulation of the epigenetic machinery could lead to gene expression dysregulation and cardiovascular diseases. Ubiquitin (Ub), methionine (Me), acetyl group (Ac), phosphate (P), deubiquitinating enzyme (DUB), histone methyltransferase (HMTs), histone demethylase (HDMTs), histone acetyltransferase (HAT), histone deacetylase (HDAC), a cytosine followed by a guanine (CpG), microRNAs (miRNAs), and messenger RNA (mRNA).

and maintaining hypermethylated DNA during replication (DNMT1) or by *de novo* methylation (DNMT3a and DNMT3b). Moreover, gene bodies of actively transcribed genes normally show slightly higher DNA methylation levels as compared to gene bodies of nontranscribed genes. In contrast, hypomethylation is usually found in enhancer regions and promoters (Costantino et al., 2018). Posttranslational modification of histone tails is another epigenetic modification that regulates gene expression by chromatin remodeling. Histone acetylation, deacetylation, methylation, phosphorylation and ubiquitination change DNA accessibility, regulating gene transcription. The acetylation of histone tails is regulated by histone acetyltransferases (HATs) and HDACs. Histone acetyltransferase enzymes acetylate the lysine residues of the histones, whereas HDACs deacetylate them, promoting gene activation or silencing, respectively. Histone methylation is regulated by histone methyltransferases (HMTs) and histone demethylases (HDMT). Methylation occurs at the lysine or arginine residues and can activate or repress gene transcription depending on the degree of methylation and which residue is methylated (Li et al., 2017c; Sabia et al., 2017). The serine, threonine, and tyrosine residues of histone tails can also be phosphorylated and dephosphorylated by protein kinases and phosphatases, respectively. Histone tail phosphorylation modulates chromatin structure, taking part in transcription, DNA repair, and chromatin compaction in cell division and apoptosis (Rossetto et al., 2012). Lastly, histone tail ubiquitination is sequentially catalyzed by ligases enzymes, which attach ubiquitin to lysine residues. Ubiquitination and deubiquitination are involved in the activation of transcription and are usually associated with histone methylation. Their effect on repressing or activating transcription generally depends on what histone is modified (Cao and Yan, 2012). Finally, miRNAs regulate gene expression *via* degradation of the transcript or repression of translation when binding to the 3′-untranslated region of the target mRNA. Thus, miRNA represses mRNA translation without changing the DNA sequence of the gene. MicroRNA binding to mRNA is imperfect, so each miRNA has multiple targets. This allows the regulation of a great part of the human genome (Bartel, 2009). The miRNAs are 19-25 nucleotides in length, encoded in the genome and transcribed into primary miRNA (pri-miRNA). Pri-miRNAs derive into miRNAs precursors (pre-miRNA) by the nuclear RNase III called Dorsha and are transferred to the cytoplasm and processed by the endonuclease Dicer to generate a double-stranded miRNA duplex. This product is incorporated into an RNA-induced silencer complex (RISC)–loading complex. Then, one strand is removed from the complex, and the other strand forms a mature RISC, serving as a template for target mRNAs (Sato et al., 2011; Nishiguchi et al., 2015).

Due to this important function in gene regulation, epigenetic modifications and miRNA may play a crucial role in the development of pathological conditions, including CVDs. Understanding the epigenetic machinery underlying cardiac disorders and how these epigenetic mechanisms can be introduced into diagnostics (i.e., biomarkers) and therapies is fundamental to improving the quality of life of patients. In medicine, a biomarker is defined as a measurable characteristic that indicates a particular physiological or pathological state or a response to a therapeutic treatment (Strimbu and Tavel, 2010). Ideally, biomarkers should have easy accessibility, predictable detection, and reliability (Sun et al., 2017). It is mandatory to present a specific measurable change that clearly associates with a diagnosis or a predictable outcome. Thus, biomarkers provide information to physicians when evaluating the probability of developing a disease, making a diagnosis, evaluating the severity of a disease and its progression; during therapeutic decision making; or when monitoring a patient's response and may result in significant cost reduction (Baccarelli et al., 2010). Their classification can be based on their application (predisposition, diagnosis, monitoring, safety, prognostic, or predictive biomarkers). Predisposition biomarkers determine how likely it is for a patient to develop a certain disease and are usually utilized when there is a personal or family history that indicates a disease risk, and the results can help guide medical care. Diagnostic biomarkers are used to detect or confirm the existence of a health disorder and may assist its early detection. Monitoring biomarkers evaluate the status of a disease or determine exposure to an environmental agent or medical product. Safety biomarkers indicate the probability, presence, or extent of toxicity of a certain medical product or environmental agent. Prognostic biomarkers indicate how a disease may progress in patients who already have the particular disease. These biomarkers do not predict the treatment response but can be useful when selecting patients for treatment. Predictive biomarkers identify patients who are most likely to have a favorable or unfavorable response to a specific treatment. Thus, they can predict treatment success or undesired side effects in a particular patient. A particular disease can have different biological mechanisms in different patients. Predictive biomarkers can be associated with the specific mechanism of a health disorder. This facilitates a targeted therapy, which uses drugs specific for a particular biological mechanism associated with a disease, increasing its effectiveness (FDA-NIH Biomarker Working Group, 2016). Specifically, epigenetic biomarkers belonging to most of these classifications are discussed in this review, with a focus on CVDs. Among the epigenetic biomarkers, miRNAs are the most attractive, as they can be detected in small sample volumes, are stable, and can be obtained from plasma, serum, saliva, and urine. Interestingly, they are highly conserved, and this allows a reliable comparison between patients and animal models of disease (Matsumoto et al., 2013). Therefore, although all epigenetic mechanisms are being intensively investigated, miRNAs are evaluated the most for their use as predictive biomarkers. This review presents an overview of current research on epigenetic biomarkers in CVDs and how this knowledge can benefit the diagnosis, prognosis, and therapy for cardiovascular patients.

EPIGENETIC BIOMARKERS IN CVDS

Over the last few years, numerous studies have linked cardiovascular risk factors to epigenetic modifications in human patients. Modification of the epigenetic environment alters cardiovascular homeostasis and impacts cardiovascular disorders. The function of epigenetic mechanisms in the regulation of

gene expression is well known, although the role of epigenetic marks in CVDs is not clearly understood. Thus, the exploration of epigenetic biomarkers may lead to a deep comprehension of the molecular mechanisms and pathways associated with CVDs. In this section, we focus on major CVDs, such as hypertension, atherosclerosis, MI, and HF, and the epigenetic biomarkers associated with them.

Hypertension

Arterial hypertension is a multifactorial disease with several mechanisms and metabolic systems involved in its pathogenesis. Genetic factors and environmental background may lead to alterations in multiple pathways that can eventually trigger development of the disease (Franceschini and Le, 2014). Intrauterine alterations, such as malnutrition, starvation, obesity, alcohol, drugs, nicotine, or environmental toxins, are some of the environmental factors directly related to hypertension development in the progeny (Bogdarina et al., 2007; Nuyt and Alexander, 2009). In addition, individuals who have aerobic training present with lower blood pressure than nontrained individuals (Fagard, 2006). This has an important impact on CVD risk factor control and is a nonpharmacological way to treat patients. There are also epigenetic factors that can influence the appearance of hypertension in adults, such as hypermethylation of genes, including superoxide dismutase-2 (*SOD2*) or *Granulysin*, or increased levels of histone acetylation at the promoter of the endothelial oxide synthetase gene (*eNOS*) (Wang et al., 2018b). Environmental factors are important to determining an individual's predisposition to developing major cardiovascular risk factors by means of epigenetic modifications, and identification of the epigenetic mechanisms that participate in hypertension development may help generate new treatments. This is of great interest because hypertension is a key risk factor for CVDs, including MI, HF, stroke, and end-stage renal disease (**Table 1** and **Figure 2**).

Essential hypertension is a multifactorial disease with no identifiable cause that is affected by environmental and epigenetic factors. Environmental stressors cause acetylation of histone 3 in the neurons of the area postrema, leading to an increase in pressure that results in hypertension (Irmak and Sizlan, 2006). Low activity of the 11 beta-hydroxysteroid dehydrogenase 2 (HSD11B2) induces hypertension. In a study performed in patients with essential hypertension or glucocorticoid-induced hypertension, the *HSD11B2* promoter was highly methylated. These changes may reflect a global status, with methylation of gene promoter being a potentially useful molecular biomarker to characterize hypertensive patients (Alikhani-Koopaei et al., 2004; Friso et al., 2008). Moreover, a polymorphism in the disruptor of telomeric silencing-1 gene (*DOT1L*), which encodes a methyltransferase that enhances methylation of histone 3 (H3K79) in the renal epithelial sodium channel gene (*ENaC*) promoter, is associated with blood pressure regulation (Duarte et al., 2012). It has also been reported that a *DOT1A* and *ALL1* (fused gene from chromosome 9 [Af9]) interaction is associated with H3K79 hypermethylation of the *ENaC* promoter, suppressing its transcriptional activity. This interaction is disrupted by aldosterone and causes hypomethylation of H3K79 at specific regions, disinhibiting the *ENaC* promoter and leading to hypertension. Thus, the Dot1a-Af9 pathway may also be involved in the control of genes implicated in hypertension (Zhang et al., 2009). Hypomethylation of the α-adducin gene (*ADD1*) promoter has been found to be connected to the risk of essential hypertension. However, differences between females and males have been found (Zhang et al., 2013a). Moreover, histone 3 (H3K4 or H3K9) demethylation is induced by lysine-specific demethylase-1 (LSD1), which modifies gene transcription. Hypermethylation of histone 3 has been associated with hypertension, increased vascular contraction, and decreased relaxation *via* the nitric oxide-cGMP (NO-cGMP) pathway in heterozygous *LSD1* knockout mice fed a high-salt diet (Pojoga et al., 2011). Histone deacetylation is also important in the development of pulmonary arterial hypertension. HDAC1 and HDAC5 protein levels have been demonstrated to be elevated in the lungs of patients and hypoxic rats. Inhibition of these proteins by valproic acid and suberoylanilide hydroxamic acid diminished the development of hypoxia-induced pulmonary hypertension in rats. Thus, HDAC1 and HDAC5 levels could be useful predictive biomarkers for the treatment of pulmonary hypertension in patients (Zhao et al., 2012).

In a study evaluating alterations in the global DNA methylation status of patients with essential hypertension, the level of the epigenetic marker 5mC was lower in hypertensive patients than in healthy people (Smolarek et al., 2010). In an *in vivo* model of hypertension using Dahl salt-sensitive rats, the levels of 5mC and 5-hydroxymethylcytosine (5hmC) were evaluated in the outer renal medulla. In response to salt administration, the 5mC levels were significantly higher for genes with low transcription and 5hmC levels higher in genes with higher expression. This study revealed important features of 5mC and 5hmC for understanding the role of epigenetic modifications in the regulation of hypertension (Liu et al., 2014).

Rivière et al. (2011) analyzed the regulation of somatic angiotensin-converting enzyme gene (*sACE*) expression by promoter methylation. *sACE* regulates blood pressure by catalyzing the conversion of angiotensin I into angiotensin II, a potent vasopressor. Hypermethylation of *sACE* promoter in cultures of human endothelial cells and rats was associated with transcriptional repression, suggesting an epigenetic mechanism in hypertension regulation (Rivière et al., 2011). More recently, Fan et al. (2017) demonstrated opposite results in patients with essential hypertension. The authors indicated that hypermethylation of the *ACE2* promoter may increase essential hypertension risk, with variabilities in CpG islands methylation in males and females (Fan et al., 2017).

Moreover, a genome-wide methylation study on essential hypertension revealed that changes in the DNA methylation of leukocytes are involved in the pathogenesis of hypertension. They found increased methylation in the gene encoding sulfatase 1 (*SULF1*), which is involved in apoptosis, and decreased methylation in the gene encoding prolylcarboxypeptidase (*PRCP*), a regulator of angiotensin II and III cleavage (Wang et al., 2013b). Another genome-wide study of blood pressure characteristics found new genetic variants that influence blood pressure and are

TABLE 1 | Epigenetic biomarkers in hypertension.

Epigenetic modification	Biomarker	Regulation in hypertension	Sample source	Study type	References
DNA methylation	*HSD11B2* promoter	Highly methylated	Rat's urine and tissues and human cell lines	Experimental: *in vitro* and rat model	(Alikhani-Koopaei et al., 2004)
	SERPIN3 CpG island	Hypomethylation	Placental tissue	Clinical	(Chelbi et al., 2007)
	HSD11B2 promoter	Highly methylated	Blood and urine	Clinical	(Friso et al., 2008)
	5mC	Lower levels	Blood	Clinical	(Smolarek et al., 2010)
	NKCC1 promoter	Hypomethylation	Aorta, heart and kidney	Experimental: spontaneously hypertensive rodent model	(Lee et al., 2010; Cho et al., 2011)
	sACE promoter	Hypermethylation	Blood	Clinical and experimental: *in vitro*	(Rivière et al., 2011)
	ERα promoter	Methylation	Uterine arteries	Clinical	(Dasgupta et al., 2012)
	SULF1, *PRCP*	*SULF1*: hypermethylation; *PRCP*: hypomethylation	Blood	Clinical	(Wang et al., 2013b)
	ADD1 promoter	Hypomethylation	Plasma	Clinical	(Zhang et al., 2013a)
	5mC, 5hmC	Higher levels	Tissue	Experimental: Dahl salt-sensitive rats	(Liu et al., 2014)
	AGT promoter	Demethylation	H295R cells and visceral adipose tissue	Experimental: *in vitro* and rat model	(Wang et al., 2014a)
	DSCR3	Hypermethylation	Maternal blood and placental tissue	Clinical	(Kim et al., 2015)
	miRNA-34a gene promoter	Hypomethylation	Placental tissue	Clinical	(Rezaei et al., 2018)
	ACE2 promoter	Hypermethylation	Plasma	Clinical	(Fan et al., 2017)
	CBS promoter	Hypermethylation	Maternal blood and placental tissue	Clinical	(Kim et al., 2015)
	MTHFD1 promoter	Hypermethylation	Plasma	Clinical	(Xu et al., 2019)
Histone modifications	H3K79	Hypermethylation	NA	Clinical	(Rodriguez-Iturbe, 2006; Duarte et al., 2012)
	Histone 3	Acetylation	Germ cells	Review	(Irmak and Sizlan, 2006)
	H3K79	DNA methylation	Bibliography	Review	(Zhang et al., 2009)
	HDAC8	Inhibition	mDCT cells and tissues	Experimental: rat models of salt-sensitive hypertension	(Mu et al., 2011)
	H3K4 or H3K9	Hypermethylation	Tissue, plasma, and urine	Experimental: LSD1 knockout mice with a high-salt diet	(Pojoga et al., 2011)
	HDAC1, HDAC5	High levels	Lung tissue and adventitial fibroblasts	Clinical and experimental: *in vitro* and hypoxic rat	(Zhao et al., 2012)
miRNA	miR-18a, miR-210, miR-152, miR-363, miR-377, miR-411, miR-518b, miR-542-3p	miR-18a, miR-363, miR-377, miR-411, miR-542-3p: underexpression; miR-210, miR-152, miR-518b: overexpression	Placental tissue	Clinical	(Zhu et al., 2009)
	22 miRNAs	15 upregulated and 7 downregulated	Serum	Clinical	(Yang et al., 2011)
	let-7b, miR-302*, miR-104, miR-128a, miR-182*, miR-133b	Overexpression	Placental tissue	Clinical	(Noack et al., 2011)
	miR-92b, miR-197, miR-342-3p, miR-296-5p, miR-26b, miR-25, miR-296-3p, miR-26a, miR-198, miR-202, miR-191, miR-95, miR-204, miR-21, miR-223	miR-92b, miR-197, miR-342-3p, miR-296-5p, miR-26b, miR-25, miR-296-3p, miR-26a, miR-198, miR-202, miR-191, miR-95, miR-204: overexpression; miR-21, miR-223: underexpression	Placental tissue	Clinical	(Choi et al., 2013)
	miR-9, miR-126	Lower levels	Peripheral blood mononuclear cells	Clinical	(Kontaraki et al., 2014)
	miR1233	Higher levels	Serum	Clinical	(Ura et al., 2014)

(*Continued*)

TABLE 1 | Continued

Epigenetic modification	Biomarker	Regulation in hypertension	Sample source	Study type	References
	miR-18a, miR-19b1, miR-92a1, miR-210	miR-210: upregulation; miR-18a, miR-19b1, and miR-92a1: downregulation	Plasma and placental tissue	Clinical	(Xu et al., 2014)
	miR-505	Upregulation	Plasma	Clinical	(Yang et al., 2014)
	miR-106a, miR-18b, miR-20b, miR-19b-2, miR-92a-2, miR-363	Dysregulation	Placental tissue	Clinical	(Zhang et al., 2015a)
	miR-515-5p, miR-518b, miR-518f-5p, miR-519d, miR-520h	Downregulation	Placental tissue	Clinical	(Hromadnikova et al., 2015)
	miR-335, miR-584	Upregulation	Placental tissue and HTR8/Svneo cells	Clinical and experimental: *in vitro*	(Jiang F. et al., 2015)
	miR-125b	Overexpression	Plasma and placental tissue	Clinical	(Yang et al., 2016b)
	miR-215, miR-155, miR-650, miR-210, miR-21, miR-18a, miR-19b1	MiR-215, miR-155, miR-650, miR-210, miR-21: upregulation; miR-18a, miR-19b1: downregulation	Plasma	Clinical	(Jairajpuri et al., 2017)
	miR-204-5p	Higher levels	Serum	Clinical	(Mei et al., 2017)
	let-7b*, let-7f-1*, miR-1183, miR-23c, miR-425*	miR-1183: upregulation; let-7b*, miR-23c, miR-425*, let-7f-1*: downregulation	Plasma and placental tissue	Clinical	(Gunel et al., 2017)
	miR-145	Downregulation	Placental tissue	Clinical	(Han et al., 2017)
	miR-202-3p	Upregulation	Placental tissue	Clinical	(Singh et al., 2017)
	let-7	Higher	Plasma	Clinical	(Huang et al., 2017b)
	miRNA	Dysregulation	Bibliography: Maternal serum and placental tissue	Bibliography review	(Laganà et al., 2018)
	miR-19a	Upregulation	Plasma and lung tissue	Clinical	(Chen and Li, 2017)
	miR-21	Upregulation	Peripheral blood mononuclear cells	Clinical	(Parthenakis et al., 2017)
	miR-21	Upregulation	Bibliography review	Bibliography review	(Sekar et al., 2017)
	miR-510	Upregulation	Serum	Clinical	(Krishnan et al., 2017)
	miR-206	Lower levels	Serum	Clinical	(Jin et al., 2017)
	miR-424(322)	Upregulation	Plasma	Clinical	(Baptista et al., 2018)
	miR-199a-3p, miR-208a-3p, miR-122-5p, miR-223-3p	Downregulation	Serum	Clinical	(Zhang et al., 2018c)
	miR-431-5p	Upregulation	Tissue	Experimental: mice made hypertensive and *in vitro*	(Huo et al., 2019)
	miR-143, NR_034083, NR_104181,	miR-143: upregulation; NR_034083: downregulation and NR_104181 and	Peripheral blood leucocytes	Clinical	(Chen et al., 2018b)

NA, not available.

strongly associated with local CpG island methylation. This study demonstrated the role of DNA methylation in the regulation of blood pressure (Kato et al., 2015).

The pathogenesis of hypertension is affected by alterations in ion flux mechanisms. Hypomethylation of the Na/K/2Cl cotransporter 1 gene (*NKCC1*) promoter results in overexpression in a rodent model with spontaneous hypertension (Lee et al., 2010). DNA methyltransferase activity maintained hypomethylation in the *NKCC1* promoter, playing an important role in *NKCC1* upregulation during the course of the disease. This encourages evaluation of the *NKCC1* methylation status in hypertensive patients (Cho et al., 2011). Furthermore, WNK4 is a serine-threonine kinase that negatively regulates the Na(+)-Cl(−)-cotransporter (NCC) and ENaC. This would affect the distal nephron, increasing the reabsorption of sodium. Stimulation of β(2)-adrenergic receptor (β(2)AR) in salt intake conditions would reduce *WNK4* transcription, resulting in inhibition of HDAC8 activity and increased histone acetylation. In the rat models of salt-sensitive hypertension, salt diet repressed renal WNK4 expression, activating the NCC and inducing salt-dependent hypertension. Thus, *WNK4* transcription is epigenetically modulated in the course of salt-sensitive hypertension, with the β(2)AR-WNK4 pathway as a potential therapeutic target for this disease (Mu et al., 2011).

Goyal et al. (2010) demonstrated that a low protein diet in pregnant mice leads to alterations in DNA methylation, miRNA,

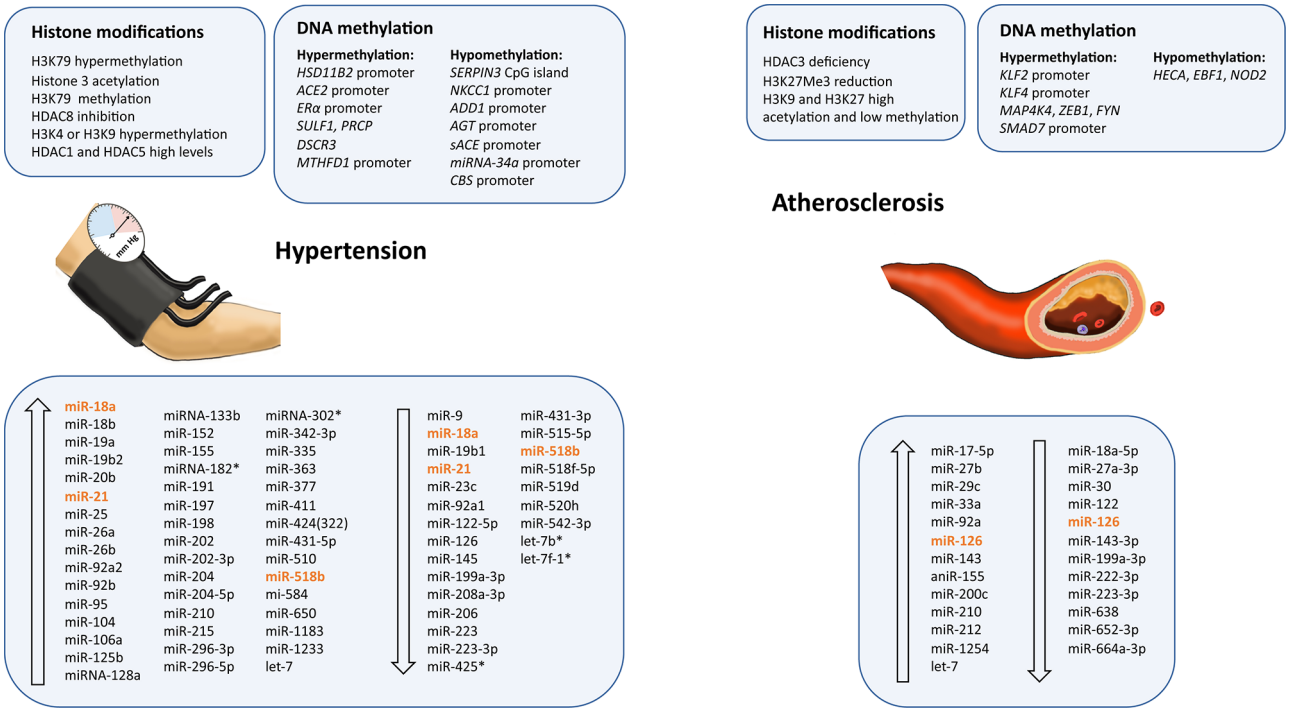


FIGURE 2 | Epigenetic modifications and microRNAs biomarkers dysregulated in atherosclerosis and hypertension. Ascending arrows indicate higher levels or upregulation, and descending arrows denote lower levels or downregulation, both compared to control conditions. Those miRNAs presenting opposite results are shown in orange.

and gene expression in the brain renin–angiotensin system, a key regulator of hypertension in adults (Goyal et al., 2010). Along the same lines, in a study carried out *in vitro* and in a rat model, DNA demethylation of the angiotensinogen gene (*AGT*) promoter activated its expression. AGT is an important substrate of the renin–angiotensin–aldosterone system and an important target in hypertension research. Elevated concentrations of circulating aldosterone and high consumption of salt stimulate the AGT gene expression in adipose-induced hypertension (Wang et al., 2014a). In addition, cystathionine β-synthase (CBS), an important enzyme in the metabolism of plasma homocysteine, is associated with hypertension and stroke. Hypermethylation of the *CBS* promoter has been demonstrated to increase the risk of both diseases, especially in male patients (Wang et al., 2019a). Similarly, hypermethylation of the methylenetetrahydrofolate dehydrogenase 1 gene (*MTHFD1*) promoter, which is also associated with homocysteine metabolism, was observed in hypertensive patients, and proposed as a potential diagnostic biomarker in patients with essential hypertension (Xu et al., 2019).

In addition to the previous classic epigenetic modifications, miRNAs often regulate hypertension and are attractive biomarkers for the disease. The miR-9 and miR-126 expression levels are significantly lower in hypertensive patients than healthy individuals and are related to hypertension prognosis and organ damage. Thus, miR-9 and miR-126 may be possible biomarkers in essential hypertension (Kontaraki et al., 2014). Moreover, ncRNAs, such as miR-143, miR-145, and NR_104181, are significantly higher in essential hypertensive patients than controls, whereas NR_027032 and NR_034083 are significantly reduced. After evaluating cardiovascular risk factors, they concluded that lower expression levels of NR_034083 and higher expression levels of NR_104181 and miR-143 were risk factors for essential hypertension (Chen et al., 2018b). Another study evaluated the correlation between miRNA let-7 expression and subclinical atherosclerosis in untreated patients with newly diagnosed essential hypertension and found increased levels in hypertensive patients, suggesting that plasma let-7 could be an indicator for monitoring end-organ damage and a biomarker for atherosclerosis in these patients (Huang et al., 2017b). Similarly, upregulation of miR-505, miR-19a, miR-21, miR-510, or miR-424(322) in blood from hypertensive patients suggests a possible use for miR-510 as a diagnostic biomarker and therapeutic target (Yang et al., 2014; Chen and Li, 2017; Krishnan et al., 2017; Parthenakis et al., 2017; Sekar et al., 2017; Baptista et al., 2018). Lower levels of the combination of miR-199a-3p, miR-208a-3p, miR-122-5p, and miR-223-3p have also been shown to be suitable for diagnosis of hypertension (Zhang et al., 2018c). Decreased miR-206 levels might also be especially useful in the detection of pulmonary hypertension in patients with left heart disease (Jin et al., 2017). Furthermore, a study in hypertensive mice produced by infusion of angiotensin II concluded that miR-431-5p knockdown delays the increase in blood pressure induced by angiotensin II and reduces vascular injury. This demonstrates its potential as a target for the treatment of hypertension and vascular injury (Huo et al., 2019).

Preeclampsia is an important pregnancy-induced syndrome characterized by hypertension and proteinuria. Chronic hypoxia is a common pregnancy stress that increases the risk of preeclampsia and is associated with changes in methylation of the estrogen receptor α gene (*ERα*) promoter. ERα is involved in adjustments to the uterine blood flow, and promoter methylation results in gene repression in uterine arteries, increasing blood pressure (Dasgupta et al., 2012). Preeclampsia also modifies the expression profile of several serine protease inhibitors (SERPINs) in the placenta. Specifically, *SERPIN3* CpG islands have a significantly low level of methylation in preeclampsia, providing a new potential marker for early diagnosis (Chelbi et al., 2007). Another study demonstrated a positive association between placenta global DNA methylation and hypertension in preeclampsia (Kulkarni et al., 2011). Next-generation sequencing technology and microarray assay analyses of the miRNA expression pattern in preeclamptic placentas versus healthy placentas have revealed that miRNAs expression is dysregulated in preeclampsia (Zhu et al., 2009; Noack et al., 2011; Yang et al., 2011; Choi et al., 2013; Xu et al., 2014; Hromadnikova et al., 2015; Zhang et al., 2015a; Gunel et al., 2017; Han et al., 2017). These results were in agreement with those found in the miRNA database from cell and tissue analyses. Thus, circulating miRNAs in the serum of pregnant women could be used as biomarkers for the diagnosis and prognosis of preeclampsia. To further demonstrate that miRNAs could be good predictors of preeclampsia, as well as its severity, circulating miRNA signatures were evaluated in women divided into groups based on preeclampsia severity. MiR-21, miR-29a, miR-125b, miR-155, miR-202-3p, miR-204-5p, miR-210, miR-215, miR-335, miR-518b, miR-584, miR-650, and miR-1233 were upregulated, whereas miR-15b, miR-18a, miR-19b1, and miR-144 were downregulated in women with severe preeclampsia compared to mild preeclampsia (Ura et al., 2014; Jiang et al., 2015; Yang et al., 2016b; Jairajpuri et al., 2017; Mei et al., 2017; Singh et al., 2017). In addition, a recent data recompilation supported a direct association between high or low expression of miRNAs in pregnancy serum and placenta in preeclamptic pregnancies (Laganà et al., 2018). Interestingly, an association has also been demonstrated between hypomethylation of the miR-34a promoter and preeclampsia severity (Rezaei et al., 2018). Another study analyzed the concentrations of Down syndrome critical region 3 (*DSCR3*), Ras association domain family 1 isoform A (*RASSF1A*), and sex-determining region Y (*SRY*) cell-free fetal DNA in maternal plasma from preeclamptic pregnancies and found that all of the markers significantly correlated with gestational age. The authors demonstrated that *DSCR3* is a novel epigenetic biomarker and an alternative to *RASSF1A* for the prediction of early-onset preeclampsia (Kim et al., 2015). However, no association was found between the methylation status of the cortisol-controlling gene (*HSD11B2*), tumor suppressor gene (*RUNX3*), or long interspersed nucleotide element-1 gene (*LINE-1*) and hypertensive disorders of pregnancy when placental DNA methylation was analyzed (Majchrzak-Celińska et al., 2017).

Atherosclerosis

Atherosclerosis is a chronic inflammatory disease characterized by the accumulation of cholesterol in the walls of large- and medium-sized arteries, the accumulation of extracellular matrix and lipids, and smooth muscle cell proliferation. This process leads to the infiltration of immune cells (mostly macrophages) and endothelial dysfunction, forming a plaque, and eventually developing into acute cardiovascular events, such as MI, peripheral vascular disease, aneurysms, and stroke (Wissler, 1991). Proatherogenic stimuli, such as low-density lipoprotein (LDL) cholesterol and oxidized LDL, have been suggested to stimulate a long-term epigenetic reprogramming of innate immune system cells. This induces a constant activation, even after the removal of atherosclerotic stimuli (Bekkering et al., 2016). Emerging evidence supports epigenetic modifications being involved in the initiation and progression of atherosclerosis, playing an important role in plaque development and vulnerability, and highlighting the importance of epigenetic biomarkers as predictors of CVDs (**Table 2** and **Figure 2**) (Xu et al., 2018).

Regarding histone modifications, HDAC3 is reported to have a protective effect in apolipoprotein E deficient (apoE−/−) mice. HDAC3 maintains the endothelial integrity, and its deficiency results in atherosclerosis (Zampetaki et al., 2010). Similarly, increased histone acetylation has been proposed to play some role in the progression of atherogenesis by modulating the expressions of proatherogenic genes (Choi et al., 2005). Histone deacetylases are upregulated in aortic smooth muscle cells when they were stimulated with mitogens. In contrast, inhibition of HDACs reduces aortic smooth muscle cell proliferation by changing cell cycle genes expression. This suggests a protective effect against atherosclerosis (Findeisen et al., 2011). Investigations of the association between changes in lysine 27 trimethylation of histone 3 (H3K27Me3), and atherosclerotic plaque development revealed a reduction in global levels of H3K27Me3 modification in vessels with advanced atherosclerotic plaques. This does not correlate with a reduction in the corresponding HMT, enhancer of zeste homolog 2 (EZH2). There was a relationship between the repression of H3K27Me3 mark in the vessels with advanced atherosclerotic plaques and the dynamic differentiation and proliferation of smooth muscle cells associated with atherosclerotic disease (Wierda et al., 2015). Histone acetylation, methylation, and the expression of their corresponding transferases in the atherosclerotic plaques of patients with carotid artery stenosis have been analyzed. Greißel et al. (2016) analyzed the expression of HATs GCN5L, P300, MYST1, and MYST2 and HMTs MLL2/4, SET7/9, hSET1A, SUV39H1, SUV39H2, ESET/SETDB1, EHMT1, EZH2, and G9a and described an enhancement in histone acetylation on H3K9 and H3K27 in the smooth muscle cells from severe atherosclerotic lesions that correlated with plaque severity. In addition, H3K9 and H3K27 methylation were significantly lower in atherosclerotic plaques and significantly associated with disease severity (Greißel et al., 2016).

DNA methylation is also involved in atherosclerosis. To identify CpG methylation profiles in the progression of atherosclerosis in the human aorta, Valencia-Morales et al. (2015) performed DNA methylation microarray analyses. They detected a correlation between histological pathology and the differential methylation of numerous autosomal genes in vascular tissue, providing potential biomarkers of damage severity and

TABLE 2 | Epigenetic biomarkers in atherosclerosis.

Epigenetic modification	Biomarker	Regulation in atherosclerosis	Sample source	Study type	References
DNA methylation	*KLF2* promoter	Methylation	HUVEC cells	Experimental: *in vitro*	(Kumar et al., 2013)
	KLF4 promoter	Methylation	HAEC cells	Experimental: *in vitro*	(Jiang et al., 2014)
	HECA, *EBF1*, *NOD2*, *MAP4K4*, *ZEB1*, *FYN*	*HECA*, *EBF1*, *NOD2*: Hypomethylated; *MAP4K4*, *ZEB1*, *FYN*: Hypermethylated	Human aortic intima and HEK293 cells	Clinical and experimental: *in vitro*	(Yamada et al., 2014)
	TIMP1, *ABCA1*, *ACAT1* promoters	Altered methylation status	Peripheral blood	Clinical	(Ma et al., 2016)
	SMAD7 promoter	Hypermethylation	Peripheral blood and atherosclerotic plaques	Clinical	(Wei et al., 2018)
	5mC, 5-hmC	Higher levels	Peripheral blood	Clinical	(Jiang et al., 2019)
Histone modifications	HDAC3	Deficiency	Aorta and HUVEC cells	Experimental: apoE–/– mice and *in vitro*	(Zampetaki et al., 2010)
	H3K27Me3	Reduction in H3K27Me3 modification	Perirenal aortic tissue patches	Clinical	(Wierda et al., 2015)
	H3K9, H3K27	Higher histone acetylation and lower histone methylation	Carotid tissue	Clinical	(Greißel et al., 2016)
miRNA	miR-130a, miR-27b, miR-210	Higher levels	Serum and intima tissue	Clinical	(Li et al., 2011)
	miR-17-5p	Higher levels	Plasma	Clinical	(Chen et al., 2015a)
	miR-143-3p, miR-222-3p	Lower levels	Microparticles	Clinical	(de Gonzalo-Calvo et al., 2016)
	miR-30	Lower levels	Plasma	Clinical	(Huang et al., 2016b)
	miR-92a	Higher levels	Plasma	Clinical	(Huang et al., 2017a)
	miR-18a-5p, miR-27a-3p, miR-199a-3p, miR-223-3p, miR-652-3p	Lower levels	Plasma	Clinical	(Vegter et al., 2017)
	miR-33a	Higher levels	Plasma	Clinical	(Kim et al., 2017)
	miR-126	Lower levels	Plasma	Experimental: apoE–/– mice	(Hao and Fan, 2017)
	miR-212	Overexpression	Serum	Clinical	(Jeong et al., 2017)
	miRNA let-7	Higher levels	Plasma	Clinical	(Huang et al., 2017b)
	miR-1254	Higher levels	Plasma	Clinical	(de Gonzalo-Calvo et al., 2018)
	miR-200c	Overexpression	Carotid plaques and plasma	Clinical	(Magenta et al., 2018)
	miR-29c	Higher levels	Plasma	Clinical	(Huang et al., 2018)
	miR-221, miR-222	Lower expression levels	Serum	Clinical	(Yilmaz et al., 2018)
	miR-638	Lower levels	Serum	Clinical	(Luque et al., 2018)
	miR-122	Higher levels	Serum	Clinical	(Wang and Yu, 2018)
	miR-221, miR-222	Higher levels in tissue samples and lower levels in whole blood	Coronary artery atherosclerotic plaques, and internal mammary arteries and whole blood	Clinical	(Bildirici et al., 2018)
	miR-664a-3p	Downregulation	Serum	Clinical	(Li et al., 2018b)
	miR-155	Higher levels	Serum	Clinical	(Qiu and Ma, 2018)
	miR-19A, miR-19B, miR-126, miR-155	Differential levels	GEO dataset	High throughput	(Mao et al., 2018)
	miR-126, miR-143	Higher levels	Plasma	Clinical	(Gao et al., 2019)

treatment targets (Valencia-Morales et al., 2015). Genes such as *Drosophila* headcase (*HECA*), early B-cell factor 1 (*EBF1*), and nucleotide-binding oligomerization domain containing 2 (*NOD2*) are significantly hypomethylated, whereas mitogen-activated protein kinase kinase kinase kinase 4 (*MAP4K4*), zinc finger E-box binding homeobox 1 (*ZEB1*), and proto-oncogene tyrosine-protein kinase (*FYN*) are hypermethylated in atheromatous plaque lesions compared to the plaque-free intima (Yamada et al., 2014). Another study described differentially methylated regions in genes associated with atherosclerosis in swine aorta endothelial cells (Jiang et al., 2015). Low-density lipoprotein cholesterol risk factor upregulates DNMT1, which methylates and represses the Krüppel-like factor 2 gene (*KLF2*) promoter. KLF2 is a transcription factor essential for endothelium homeostasis, and its repression results in endothelial dysfunction (Kumar et al., 2013). Similarly, DNMT3a upregulation in human

aortic endothelial cells exposed to disturbed flow induces the methylation and repression of the Krüppel-like factor 4 gene (*KLF4*) promoter, increasing regional atherosusceptibility (Jiang et al., 2014). In an attempt to determine biomarkers of atherosclerosis in the primary stages, the DNA methylation status was determined in a selection of gene promoters associated with the disease. They analyzed the promoter methylation of ATP binding cassette subfamily A member 1 (*ABCA1*), TIMP metallopeptidase inhibitor 1 (*TIMP1*), and acetyl-CoA acetyltransferase 1 (*ACAT1*) and observed significant alterations in the peripheral blood of atherosclerosis patients (Ma et al., 2016). A recent study found that *SMAD7* expression is decreased and its promoter highly methylated in atherosclerotic plaques compared to normal artery walls. There was also increased DNA methylation of the *SMAD7* promoter in the peripheral blood of atherosclerosis patients. Thus, the *SMAD7* promoter is hypermethylated in atherosclerosis patients and their atherosclerotic plaques, with a positive association with homocysteine levels (Wei et al., 2018). Moreover, increased 5mC and 5-hmC levels, which indicate DNA methylation and hydroxymethylation, respectively, have been demonstrated in peripheral blood mononuclear cells from elderly patients with coronary heart disease. These results positively correlate with the severity of coronary atherosclerosis (Jiang et al., 2019).

MicroRNAs have also been identified as attractive epigenetic biomarkers for atherosclerosis. Li et al. (2011) examined miRNA levels in serum samples and the intima of atherosclerosis obliterans patients and compared them to controls. They observed increased levels of miR-27b, miR-130a, and miR-210 in serum and sclerotic tissue from patients, proposing these miRNAs as epigenetic biomarkers for early stages of the disease (Li et al., 2011). Later, a study with a reduced number of patients suggested that elevated levels of circulating miR-17-5p may be a useful biomarker in the diagnosis of coronary atherosclerosis (Chen et al., 2015a).

Microparticles secreted by human coronary artery smooth muscle cells are a different source of cardiovascular biomarkers. These extracellular vesicles can contain miRNAs, such as miR-21-5p, miR-143-3p, miR-145-5p, miR-221-3p, and miR-222-3p. Lower levels of miR-143-3p and miR-222-3p have been found in microparticles derived from atherosclerotic plaque areas compared to nonatherosclerotic areas (de Gonzalo-Calvo et al., 2016).

Huang et al. (2016b) evaluated the expression of miR-30 in patients with essential hypertension compared to control individuals. They observed a reduction in miR-30 levels in the hypertensive patients and in the increased carotid intima-media thickness group. Thus, the authors suggested that circulating miR-30 may be a useful noninvasive atherosclerosis biomarker for patients with essential hypertension (Huang et al., 2016b). Later, the authors also identified higher levels of miR-92a as a possible biomarker of atherosclerosis in the same type of patients (Huang et al., 2017a).With the aim of investigating correlations between circulating miRNAs specific for HF and atherosclerosis in HF patients, Vegter et al. (2017) assessed miRNAs levels and related them to biomarkers associated with atherosclerotic disease and rehospitalizations of cardiovascular patients. They demonstrated a consistent trend between a high number of atherosclerosis manifestations and lower levels of miR-18a-5p, miR-27a-3p, miR-199a-3p, miR-223-3p, and miR-652-3p. Thus, lower levels of circulating miRNAs in HF patients with atherosclerotic disease and an elevated probability of cardiovascular-related rehospitalization were described (Vegter et al., 2017). High levels of miR-33a have also been demonstrated to be a potential cause of cholesterol accumulation and to exacerbate vessel walls inflammation in atherosclerotic disease. Thus, plasma miR-33a has been proposed as a suitable biomarker in atherosclerosis (Kim et al., 2017).

In an attempt to identify more atherosclerosis biomarkers, Hao and Fan (2017) performed microarray analysis using the plasma from apoE−/− mice and discovered that a reduction in miR-126 levels is a good indicator of atherosclerotic disease. They also determined that miR-126 is involved in the mitogen-associated protein kinase (MAPK) signaling pathway, reducing cytokine release and progressing atherosclerotic pathogenesis (Hao and Fan, 2017). In contrast, Gao et al. (2019) determined that higher expression levels of miR-126 and miR-143 correlate with the presence and severity of cerebral atherosclerosis (Gao et al., 2019). In another study, the authors evaluated the synergy of circulating miRNAs with cardiovascular risk factors to estimate the presence of atherosclerosis in ischemic stroke patients. They identified miR-212 as a novel marker that enhances the estimation of atherosclerosis presence in combination with hemoglobin A_{1c}, high-density lipoprotein cholesterol, and lipoprotein(a) (Jeong et al., 2017). Another candidate biomarker for atherosclerosis is miR-200c. The authors analyzed plaque instability in the carotid arteries of patients undergoing carotid endarterectomy by examining the expression of miR-200c. Higher expression of miR-200c positively correlated with instability biomarkers, such as monocyte chemoattractant protein-1, cyclooxygenase-2, interleukin 6 (IL-6), metalloproteinases, and miR-33a/b, and negatively correlated with stability biomarkers, such as ZEB1, endothelial nitric oxide synthase, forkhead boxO1, and Sirtuin1. Thus, miR-200c could be a biomarker of atherosclerotic plaque progression and clinically useful for identifying patients at high embolic risk (Magenta et al., 2018). Along the same lines, lower serum levels of miR-638 may be a suitable biomarker of plaque vulnerability and ischemic stroke in individuals with high cardiovascular risk (Luque et al., 2018). With the intention to explore the role of miRNAs associated with carotid atherosclerosis, Mao et al. (2018) analyzed the genes differentially expressed between primary and advanced atherosclerotic plaques using two public datasets from the Gene Expression Omnibus (GEO) databases. The authors found a total of 23 miRNAs and focused on miR-19A, miR-19B, miR-126, and miR-155, which may be considered biomarkers of carotid atherosclerosis (Mao et al., 2018). In addition, Li et al. (2018b) identified downregulation of specific circulating miR-664a-3p as a biomarker of atherosclerosis in patients with obstructive sleep apnea and enlarged maximum carotid intima-media thickness (Li et al., 2018b).

Circulating miR-221 and miR-222 could also be suitable biomarkers for the diagnosis of atherosclerosis, as lower levels of these miRNAs correlate with the disease (Bildirici et al., 2018;

Yilmaz et al., 2018). However, higher levels have been found in samples from coronary atherosclerotic plaques and internal mammary arteries (Bildirici et al., 2018). On the other hand, higher circulating levels of miR-29c, miR-122, and miR-155 in coronary atherosclerosis patients might allow noninvasive detection of the disease and its severity (Huang et al., 2018; Qiu and Ma, 2018; Wang and Yu, 2018). In another interesting study that assessed whether atherosclerosis of different arterial territories, not including the coronary artery, is associated with specific circulating miRNAs, the investigators were able to identify specific miRNA profiles for each territory with atherosclerotic disease. These findings may provide a pathophysiological understanding and be useful for selecting potential biomarkers for clinical practice (Pereira-da-Silva et al., 2018).

Myocardial Infarction

Acute MI (AMI) is a threatening disease worldwide. Early and accurate differential diagnosis is critical for immediate medical intervention and improved prognosis (Reed et al., 2017). In particular, it is important to notice that patients with ST-segment elevation MI (STEMI) have different requirements than patients with non-STEMI (NSTEMI). For the first group, reperfusion therapy should be administered quickly to reduce infarct size and mortality (Authors/Task Force members et al., 2014). However, in NSTEMI patients, revascularization strategies are recommended based on individual clinical characteristics (Reed et al., 2017). Therefore, biomarkers with the capacity to diagnose and personalize a therapeutic schedule in AMI would be of great interest. Currently, the favored diagnostic biomarkers of AMI are cardiac troponin I (cTnI) and T (cTnT), both of which are released from necrotic cardiomyocytes within 2 to 4 h post-MI (Babuin and Jaffe, 2005), with maximum levels at 24 to 48 h and lasting for more than 1 week (Jaffe et al., 2006). For this reason, small repeat infarctions after the main infarction are difficult to detect. Thus, it is fundamental to identify biomarkers for very early diagnosis of STEMI and for monitoring the entire pathological process of AMI (**Table 3** and **Figure 3**).

Regarding methylation as an indicator of MI, Talens et al. (2012) investigated the association between MI and DNA methylation at six loci described to be sensitive to prenatal nutrition. As a result, the researchers demonstrated that the risk of MI in women is associated with DNA hypermethylation at *INS* and *GNASAS*-specific loci (Talens et al., 2012). Moreover, microarray analyses investigating whole-genome DNA methylation using cases from the EPICOR study and EPIC-NL cohort (Fiorito et al., 2014) identified a hypomethylated region in the zinc finger and BTB domain-containing protein 12 (*ZBTB12*) and *LINE-1*, concluding that it is possible to detect specific methylation profiles in white blood cells a few years before MI occurs. This provides a promising early biomarker of MI (Guarrera et al., 2015). Another example is the hypermethylation of the aldehyde dehydrogenase 2 gene (*ALDH2*) promoter, which is associated with myocardial injury after MI in rats. The hypermethylation downregulates *ALDH2*, inhibiting its cardioprotective role (Wang et al., 2015). Rask-Andersen et al. (2016) performed an epigenome-wide association study to identify disease-specific alterations in DNA methylation. The authors observed differential DNA methylation at 211 CpG sites in individuals with MI, and some of these sites represented genes related to cardiac function, CVD, cardiogenesis, and recovery after ischemic injury. Their results highlight genes that might be important in the pathogenesis of MI or in recovery (Rask-Andersen et al., 2016). Along the same lines, a genome-wide DNA methylation and gene ontology analysis of white blood cells from a population-based study identified four differentially methylated sites in individuals who had a previous MI. Interestingly, they found a correlation between differences in DNA methylation in blood cells and the levels of growth differentiation factor 15 (GDF-15), which was overexpressed in the myocardium of MI patients (Ek et al., 2016). Later, a genome-wide DNA methylation study of whole blood samples from MI patients and controls identified two methylated regions in zinc finger homeobox 3 (*ZFHX3*) and SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily a, member 4 (*SMARCA4*) that were independently related to MI (Nakatochi et al., 2017).

Histone modifications are also involved in the pathological process of MI. To investigate the role of the HAT p300 in adverse left ventricular (LV) remodeling, Miyamoto et al. (2006) generated transgenic mice overexpressing wild-type p300 or its mutant in the heart. They subjected these mice to surgical MI and demonstrated that cardiac overexpression of p300 stimulated adverse LV remodeling. They concluded that the HAT activity of p300 is fundamental for the pathological course of MI (Miyamoto et al., 2006). Moreover, the class III deacetylase sirtuin 1 (SIRT1) is well known to confer a cardioprotective effect and is downregulated after cardiac injury. To understand the underlying mechanism, primary rat neonatal ventricular myocytes were exposed to ischemic or oxidative stress, leading to upregulation of the histone H3K9 methyltransferase SUV39H and downregulation of *SIRT1*. In addition, inhibition of SUV39H activity by chaetocin in wild-type mice and *SUV39H*-knockout mice protected against induced MI. SUV39H and heterochromatin protein 1 gamma cooperate to methylate the *SIRT1* promoter and repress its transcription. Thus, the authors described a role for SUV39H linking SIRT1 repression to MI (Yang et al., 2017a). To examine the role of HDAC4 in the modulation of cardiac function after an MI, Zhang et al. (2018b) generated a myocyte-specific activated HDAC4-transgenic mouse. They found that HDAC4 overexpression increases myocardial fibrosis and hypertrophy, leading to cardiac dysfunction. Furthermore, the overexpression of activated HDAC4 aggravated cardiac dysfunction and increased adverse remodeling and apoptosis in the infarcted myocardium. Thus, HDAC4 is an indicator of heart injury (Zhang et al., 2018b). More recently, the role of HDAC6 in the development of HF following MI was investigated using a rat model. The authors found that the deacetylase activity of HDAC6 is increased after MI (Nagata et al., 2019).

Abundant research has focused on miRNAs as novel biomarkers for MI. MiR-1 levels have been analyzed in plasma from patients with AMI and found to be significantly elevated, but decreased to normal levels with medication (Ai et al., 2010; Long et al., 2012a). MiR-1, miR-126, and cTnI expression levels exhibited a similar tendency. Thus, circulating miR-1 and

TABLE 3 | Epigenetic biomarkers in myocardial infarction.

Epigenetic modification	Biomarker	Regulation in myocardial infarction	Sample source	Study type	References
DNA methylation	*INS*, *GNASAS*	Hypermethylation	Leukocytes	Clinical	(Talens et al., 2012)
	LINE-1, *ZBTB12*	Hypomethylation	White blood cells	Clinical	(Guarrera et al., 2015)
	ALDH2 promoter	Hypermethylation	Experimental: rat model of MI	Experimental: rat model of MI	(Wang et al., 2015)
	ZFHX3, SMARCA4	Methylation	Whole blood	Clinical	(Nakatochi et al., 2017)
Histone modifications	p300	Overexpression	Myocardium	Experimental: mouse model of MI in HATmut p300-Tg mice	(Miyamoto et al., 2006)
	SUV39H, SIRT1	SUV39H upregulation and SIRT1 downregulation	H9C2 cells primary rat neonatal ventricular myocytes	Experimental: mouse model of MI in SUV39H–/– mice	(Yang et al., 2017a)
	HDAC4	Overexpression	Myocardium	Experimental: mouse model of MI in MHC-HDAC4-Tg mice	(Zhang et al., 2018b)
	HDAC6	Higher levels	Myocardium	Experimental: rat model of MI	(Nagata et al., 2019)
miRNA	miR-1	Higher levels	Plasma	Clinical	(Ai et al., 2010)
	miR-31, miR-126, miR-214, miR-499-5p	miR-31, miR-214: upregulation; miR-126, miR-499-5p: downregulation	Myocardium	Experimental: rat model of MI	(Shi et al., 2010)
	miR-499	Higher levels	Tissues and plasma	Clinical	(Adachi et al., 2010)
	miR-1, miR-133a, miR-133b, miR-499-5p, miR-122, miR-375	miR-1, miR-133a, miR-133b, miR-499-5p: upregulation; miR-122, miR-375: downregulation	Plasma	Clinical and experimental: mouse model of MI	(D'Alessandra et al., 2010)
	miR-1, miR-126	miR-1: increased; miR-126: decreased	Plasma	Clinical	(Long et al., 2012a)
	miR-133a	Higher levels	Plasma	Clinical	(Eitel et al., 2012)
	miR-30a, miR-195, let-7b	miR-30a, miR-195: increased; let-7: decreased	Plasma	Clinical	(Long et al., 2012b)
	miR-499-5p	Higher levels	Plasma	Clinical	(Olivieri et al., 2013)
	miR-1, miR-133a, miR-208b, miR-499	Higher levels	Plasma	Clinical	(Li et al., 2013)
	miR-150	Downregulation	plasma	Clinical	(Devaux et al., 2013)
	miR-133a	Higher levels	Plasma	Clinical	(Wang et al., 2013a)
	miR-21-5p, miR-361-5p, miR-519e-5p	miR-21-5p, miR-361-5p: increased; miR-519e-5p: reduced	Plasma	Clinical	(Wang et al., 2014b)
	miR-208a, miR-499	Higher levels in serum; miR-499: lower levels in scar, miR-208a: unchanged in scar	Serum and heart tissues	Experimental: mouse model of MI	(Xiao et al., 2014)
	miR-208b, miR-34a	Higher levels	Plasma	Clinical	(Lv et al., 2014)
	miR-328, miR-134	Higher levels	Plasma	Clinical	(He et al., 2014)
	miR-133, miR-1291, miR-663b	Higher levels	Plasma	Clinical	(Peng et al., 2014)
	miR-497	Upregulation	Plasma	Clinical	(Li et al., 2014b)
	miR-1	Higher levels	Plasma	Clinical	(Li et al., 2014a)
	miR-19a	Higher levels	Plasma	Clinical	(Zhong et al., 2014)
	miR-486-3p, miR-150-3p, miR-126-3p, miR-26a-5p, and miR-191-5p	miR-486-3p, miR-150-3p: upregulation; miR-126-3p, miR-26a-5p, miR-191-5p: downregulation	Serum	Clinical	(Hsu et al., 2014)
	miR-145	Higher levels	Serum	Clinical	(Dong et al., 2015)

(*Continued*)

TABLE 3 | Continued

Epigenetic modification	Biomarker	Regulation in myocardial infarction	Sample source	Study type	References
	hsa-miR-493-5p, hsa-miR-369-3p, hsa-miR-495, hsa-miR-3615, hsa-miR-433, hsa-miR-877-3p, hsa-miR-1306-3p, hsv1-miR-H2, hsa-miR-3130-5p, hcmv-miR-UL22A	hsa-miR-493-5p, hsa-miR-369-3p, hsa-miR-495, hsa-miR-3615, hsa-miR-433: upregulation, hsa-miR-877-3p, hsa-miR-1306-3p, hsv1-miR-H2, hsa-miR-3130-5p, hcmv-miR-UL22A: downregulation	Plasma	Clinical	(Liang et al., 2015)
	miR-499	Higher levels	Plasma	Clinical	(Zhang et al., 2015b)
	miR-486, miR-150	Higher levels	Plasma	Clinical	(Zhang et al., 2015c)
	miR-499	Higher levels	Plasma	Clinical	(Chen et al., 2015b)
	miR-146a, miR-21	Higher levels	Plasma	Clinical	(Liu et al., 2015a)
	miR-1, miR-208, miR-499	Higher levels	Plasma	Clinical	(Liu et al., 2015b)
	miR-208a	Higher levels	Plasma	Clinical	(Białek et al., 2015)
	miR-208	Overexpression	Plasma	Clinical	(Han et al., 2015)
	miR-122-5p	Higher levels	Plasma	Clinical	(Yao et al., 2015)
	miR-21	Higher levels	Plasma	Clinical	(Zhang et al., 2016)
	miR-99a	Downregulation	Plasma	Clinical	(Yang et al., 2016a)
	miR-19b-3p, miR-134-5p and miR-186-5p	Higher levels	Plasma	Clinical	(Wang et al., 2016b)
	miR-106a-5p, miR-424-5p, let-7g-5p, miR-144-3p, miR-660-5p	Higher levels	Blood	Clinical	(Bye et al., 2016)
	miR-19b-3p, miR-134-5p and miR-186-5p	Overexpression	Plasma	Clinical	(Wang et al., 2016b)
	miR-125b-5p, miR-30d-5p	Overexpression	Plasma	Clinical	(Jia et al., 2016)
	miR-423-5p, miR-30d	Overexpression	Plasma	Clinical	(Eryilmaz et al., 2016)
	miR-221-3p	Overexpression	Plasma	Clinical	(Coskunpinar et al., 2016)
	miR-208a	Overexpression in myocardium and high levels in serum	Myocardium and serum	Experimental: rat model of MI	(Feng et al., 2016)
	miR-133b, miR-22-5p	Upregulation	Serum/plasma	Clinical	(Maciejak et al., 2016)
	miR-103a	Higher levels in plasma	Plasma and peripheral blood mononuclear cells	Clinical and experimental: *in vitro*	(Huang et al., 2016a)
	miR-122-5p/133b	High ratio	Serum	Clinical	(Cortez-Dias et al., 2016)
	miR499a-5p	Higher levels	Plasma	Clinical	(O'Sullivan et al., 2016)
	miR-181a	Higher levels	Plasma	Clinical	(Zhu et al., 2016)
	miR-145	Decreased	Plasma	Clinical	(Zhang et al., 2017b)
	miR-133a	Higher levels	Plasma	Clinical	(Yuan et al., 2016)
	miR-208b	Higher levels	Plasma	Clinical	(Liu et al., 2017)
	miR-1, miR-92a, miR-99a, miR-143, miR-223	miR-143: increased; miR-1, miR-92a, miR-99a, miR-223: decreased	Monocytes	Clinical	(Parahuleva et al., 2017)
	miR-92a	Higher levels	Plasma	Clinical	(Zhang et al., 2017c)
	miR-208b	Overexpression	Plasma	Clinical	(Zhang and Xie, 2017)
	miR-124	Higher levels	Peripheral blood	Clinical	(Guo et al., 2017)
	miR-1, miR-21, miR-29b and miR-92a	miR-1, miR-21, miR-29b: increased	Plasma	Clinical	(Grabmaier et al., 2017)
	miR-874-3p	Downregulation	Plasma	Clinical	(Yan et al., 2017)
	pmiR-126	Lower levels	Platelet	Clinical	(Li et al., 2017b)
	miR-133a	Lower levels	Serum/Plasma	Clinical	(Zhu et al., 2018)
	miR-21	Upregulation	Serum	Clinical	(Wang et al., 2017b)
	miR-4478	Higher levels	Serum	Clinical	(Gholikhani-Darbroud et al., 2017)
	miR-23b	Higher levels	Plasma	Clinical	(Zhang et al., 2018a)
	MiR-27a, miR-31, miR-1291, miR-139-5p, miR-204, miR-375	Higher levels	GEO database	High throughput	(Wu et al., 2018a)
	miR-1, miR-133a, miR-34a	Lower levels	Myocardium	Experimental: mouse model of MI	(Qipshidze Kelm et al., 2018)

(*Continued*)

TABLE 3 | Continued

Epigenetic modification	Biomarker	Regulation in myocardial infarction	Sample source	Study type	References
	miR-19b, miR-223, miR-483-5p	Higher levels	Plasma	Clinical	(Li et al., 2019)
	miR-17-5p, miR-126-5p, miR-145-3p	Higher levels	Plasma	Clinical	(Xue et al., 2019)
	miR-150	Lower levels	Serum	Clinical	(Lin et al., 2019)
	miR-208b, miR-499	Higher levels	Plasma	Clinical	(Devaux et al., 2012)

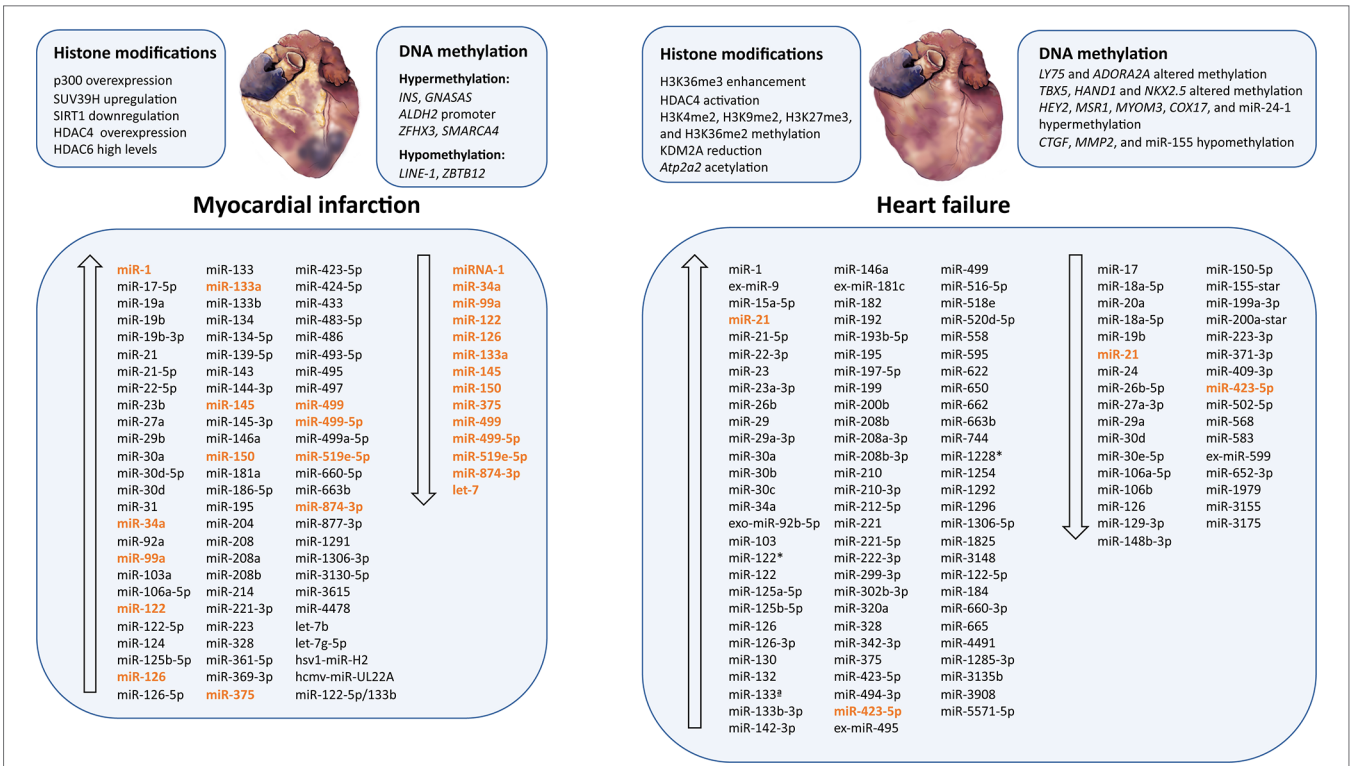


FIGURE 3 | Epigenetic modifications and microRNAs biomarkers dysregulated in myocardial infarction and heart failure. Ascending arrows indicate higher levels or upregulation, and descending arrows denote lower levels or downregulation, both compared to control conditions. Those miRNAs presenting opposite results are shown in orange.

miR-126 may be useful indicators of AMI (Long et al., 2012a). However, when miR-1 was compared to cTnT, the authors found that cTnT was more specific and sensitive than miR-1 (Li et al., 2014a). Experiments performed in a rat model of MI revealed dysregulation of several miRNAs in the myocardium. Specifically, miR-31, miR-208, and miR-214 were upregulated, and miR-126 and miR-499-5p were downregulated in infarcted rats compared to sham-operated animals (Ji et al., 2009; Shi et al., 2010). MiR-499 has been widely analyzed as a possible biomarker of MI. MiR-499 has been reported to be produced almost exclusively in the heart and plasma and is significantly increased in individuals with AMI (Adachi et al., 2010; Devaux et al., 2012). MiR-499 positively correlates with serum creatine kinase-MB (CK-MB) and cTnI increasing their diagnostic accuracy (Chen et al., 2015b; Zhang et al., 2015b). Thus, miR-499 might be a suitable biomarker for MI and a predictor of myocardial ischemia risk (Adachi et al., 2010; Chen et al., 2015b; Zhang et al., 2015b). These results were confirmed in the mouse model of MI, with elevated serum miR-208a levels. However, the expression of miR-499 was significantly reduced in the MI region, whereas miR-208a remained unchanged in the same area. One explanation is that the damaged heart might release miR-499 into the circulation (Xiao et al., 2014). Other authors observed a high correlation between circulating miRNA-208a in STEMI patients and the levels of cTnI and CK-MB mass liberated from the infarcted zone (Białek et al., 2015). Thus, cardiac miR-208 and miR-499 seemed to be better biomarkers for predicting AMI than miR-1 (Liu et al., 2015b; Liu et al., 2018a). Another study analyzed the expression of miR-208a in the myocardium and serum of infarcted rats compared to control groups, as well as the expression of cAMP-PKA to

evaluate the effect of this signaling pathway in the primary stages of MI; they found increased expression of miR-208a and cAMP-PKA. Moreover, the transfection of human myocardial cells with the miR-208a analog significantly increased the amount of cAMP-PKA protein. Thus, higher expression of miR-208a in the infarcted myocardium and serum may play a role in MI by affecting the cAMP-PKA signaling pathway (Feng et al., 2016).

D'Alessandra et al. (2010) investigated plasma levels of miRNAs in acute STEMI patients and infarcted mice and found higher levels of miR-1, miR-133a, miR-133b, and miR-499-5p compared to controls, whereas miR-122 and miR-375 levels were lower only in STEMI patients. Peak miR-1, miR-133a, and miR-133b expression correlated with cTnI levels in time, whereas the time course of miR-499-5p was slower (D'Alessandra et al., 2010). This was later confirmed in an exhaustive meta-analysis of relevant publications (Cheng et al., 2014). Similarly, geriatric patients with acute NSTEMI had greater miR-499-5p levels, exhibiting greater precision in diagnosis than cTnT in patients with mild ST elevation (Olivieri et al., 2013). On the other hand, increased levels of miR-1, miR-133a, miR-208b, and miR-499 in patients with AMI have been demonstrated to not be superior to cTnT (Li et al., 2013). The use of miR-133a as a biomarker in reperfused STEMI has been evaluated and compared to cardiovascular magnetic resonance imaging; high levels of miR-133a correlated with an increased infarct scar size, worse myocardial recovery, and prominent reperfusion injury. Nevertheless, miR-133a did not add further predictive information to cardiovascular magnetic resonance and conventional markers used in clinical practice in high-risk STEMI patients (Eitel et al., 2012). Moreover, the circulating levels of miR-133a were significantly enhanced in AMI patients compared to coronary heart disease and myocardial ischemia patients, presenting a similar trend as plasma cTnI concentration. Remarkably, we found a positive correlation between circulating miR-133a levels and the severity of coronary artery stenosis. Thus, circulating miR-133a may be a suitable tool for AMI diagnosis and predicting the presence and severity of coronary damage in coronary heart disease patients (Wang et al., 2013a). These results were later confirmed (Yuan et al., 2016; Zhu et al., 2018). Nevertheless, in another study analyzing miR-133a and miR-423-5p and their relationship with cardiac biomarkers, such as B-type natriuretic peptide (BNP), C-reactive protein, and cTnI in MI patients, an increase in circulating levels of both miRNAs was observed, but these changes were not associated with LV remodeling or BNP. The authors claimed that miR-133a and miR-423-5p are not useful biomarkers of LV remodeling after MI (Bauters et al., 2013). Another controversial pair of biomarkers is miR-423-5p and miR-30d, which were found to be higher in STEMI patients without a significant correlation with cTnI (Eryılmaz et al., 2016). In addition, the analysis of circulating miR-124a and miR-133 in STEMI and cardiogenic shock patients revealed a significant upregulation of both molecules. A negative correlation was found between miR-133 and MMP-9 levels, and a relationship between miR-124 and soluble ST2 levels, a marker associated with cardiac damage. Surprisingly, this study did not connect any of the miRNAs to the extent of the injury, disease progression, or the prognosis of patient outcomes. In this case, miRNAs would not bring any benefit compared to current markers (Goldbergova et al., 2018). Moreover, elevated circulating miR-1254 was described as predicting adverse LV remodeling in STEMI patients when compared to magnetic resonance imaging. However, the diagnosis and prognosis values of miR-1254 require further research (de Gonzalo-Calvo et al., 2018). Other investigations have described miR-150-3p and miR-486-3p as being upregulated, whereas miR-26a-5p, miR-126-3p, and miR-191-5p were significantly downregulated in STEMI patients (Hsu et al., 2014). In the same manner, circulating miR-19b-3p, miR-134-5p, and miR-186-5p have been reported to be significantly elevated in the initial stages of AMI. The expression of miR-19b-3p and miR-134-5p in the plasma reached a maximum earlier than miR-186-5p. However, all three positively correlated with cTnI and achieved peak expression before cTnI, which was 8 h after admission. Interestingly, the expression of these circulating miRNAs was not altered by heparin and medications for AMI, and the combination of all three miRNAs increased their diagnostic efficacy (Wang et al., 2016b). Moreover, a higher miR-122-5p/133b ratio was found in serum from STEMI patients (Cortez-Dias et al., 2016). The NSTEMI patients presented higher serum levels of miR-4478, soluble leptin receptor, cTnI, CKMB, urea, creatinine, glucose, cholesterol, TG, and ALP but lower levels of ALT compared to healthy individuals (Gholikhani-Darbroud et al., 2017). Moreover, there was an increase in miR-143 expression in monocytes from STEMI patients, whereas miR-1, miR-92a, miR-99a, and miR-223 expression was significantly reduced. Also, monocytic expression of miR-143 positively correlated with high-sensitivity C-reactive protein (hs-CRP), but not cTnT. These findings demonstrated that circulating monocytes could also be suitable biomarkers (Parahuleva et al., 2017).

Interestingly, cell-specific miRNA patterns are able to distinguish STEMI and NSTEMI patients. A correlation was found between miRNA 30d-5p and plasma, platelets, and leukocytes in patients with STEMI and NSTEMI. Furthermore, miR-221-3p and miR-483-5p were associated with plasma and platelets, but only in NSTEMI patients (Ward et al., 2013).

High levels of plasma miR-134 and miR-328 are described as being possible AMI biomarkers, as they correlate with a superior risk of developing HF and mortality. However, the miRNA levels were not superior to high-sensitivity cTnT (hs-cTnT) concentrations (He et al., 2014). In addition, elevated levels of miR-19a, miR-22-5p, miR-27a, miR-30a, miR-30a-5p, miR-30d-5p, miR-31, miR-34a, miR-122-5p, miR-125b-5p, miR-133, miR-133b, miR-139-5p, miR-150, miR-181a, miR-195, miR-204, miR-208, miR-208b, miR-221-3p, miR-375, miR-486, miR-497, miR-499a-5p, miR-663b, miR-1291, and let-7b can be potential biomarkers for AMI, increased risk of mortality, or HF (Devaux et al., 2012; Long et al., 2012b; Devaux et al., 2013; Li et al., 2014b; Lv et al., 2014; Peng et al., 2014; Zhong et al., 2014; Han et al., 2015; Yao et al., 2015; Zhang et al., 2015c; Coskunpinar et al., 2016; Jia et al., 2016; Maciejak et al., 2016; O'Sullivan et al., 2016; Zhu et al., 2016; Liu et al., 2017; Zhang and Xie, 2017; Alavi-Moghaddam et al., 2018; Maciejak et al., 2018; Wu et al., 2018a; Wang et al., 2019b). Other potential biomarkers for AMI are downregulated in patients' plasma, such as miR-99a, miR-122-5p, and miR-874-3p (Yang et al., 2016a; Yan et al., 2017; Wang et al., 2019b). Interestingly, high levels of the combination

of miR-21-5p, miR-361-5p, and miR-519e-5p or the reduction of miR-519e-5p correlates with cTnI concentrations, significantly increasing the diagnostic accuracy in AMI patients (Wang et al., 2014b; Liu et al., 2015a). Similarly, miR-21 and miR-124 have similar diagnostic ability compared to CK, CK-MB, and cTnI (Zhang et al., 2016; Guo et al., 2017).

In an attempt to predict HF and cardiovascular death after AMI, circulating miR-145, the N-terminal fragment of the precursor BNP, myocardial-band CK, and cTnI concentrations were analyzed for short- and long-term clinical outcomes. As a result, the authors concluded that miR-145 was a significant independent predictor of cardiac events, predicting long-term outcomes after AMI (Dong et al., 2015). Later, another group found that miR-145 levels were significantly lower in AMI patients and correlate with increased serum BNP and cTnT and decreased LV ejection fraction (Zhang et al., 2017b).

An miRNA array revealed differences in the miRNA expression patterns in patients with different phases of HF after MI. Specifically, human miR-369-3p, miR-433, miR-493-5p, miR-495, and miR-3615 were overexpressed, whereas miR-877-3p, miR-1306-3p, hsv1-miR-H2, miR-3130-5p, and hcmv-miR-UL22A were underexpressed in these patients. Thus, these circulating miRNAs are novel candidates as biomarkers of MI and HF (Liang et al., 2015).

An important aspect of circulating miRNAs as biomarkers is their temporal release, source, and transportation. Using the ischemia–reperfusion injury model, Deddens et al. (2016) showed that the ischemic myocardium releases extracellular vesicles. They also demonstrated that these extracellular vesicles transported specific miRNAs from the heart and muscle and were quickly detected in plasma. Interestingly, these vesicles had a high miRNAs content and rapid detection compared to traditional injury markers. This makes them a promising tool for the early detection of MI (Deddens et al., 2016). Along the same lines, microparticles and the expression levels of miR-92a were investigated in AMI and stable coronary artery disease patients and compared to cTnI. The number of microparticles and expression levels of miR-92a were higher in AMI patients than in the stable coronary artery disease patients and control groups, with a positive correlation between the levels of microparticles and cTnI. Thus, microparticles containing miR-92a may be suitable for MI diagnosis and possibly regulate dysfunctional endothelial tissue in AMI patients (Zhang et al., 2017c). However, according to Grabmaier et al. (2017), miR-92a seems to not be a good biomarker of adverse ventricular remodeling in post-AMI patients. The authors evaluated circulating miR-1, miR-21, miR-29b, and miR-92a from the SITAGRAMI trial population and found that miR-1, miR-21, and miR-29b expression was higher in AMI patients. The levels of miR-1 and miR-29b in plasma post-AMI correlated with variations in infarct volume, and the levels of miR-29b and changes in LV ejection fraction over time were also associated (Grabmaier et al., 2017).

Investigation of the expression of miR-103a in AMI patients with and without high blood pressure and the effect on endothelial cell function revealed increased levels of miR-103a in all patients but no changes in peripheral blood mononuclear cells. Moreover, miR-103a suppressed the expression of Piezo1 protein, which diminished the capacity to produce capillary tubes and the viability of human umbilical vein endothelial cells (HUVECs). Thus, miR-103a may take part in the development of high blood pressure and the initiation of AMI *via* regulation of Piezo1 expression (Huang et al., 2016a).

In a study based on samples from the HUNT study biobank, Bye et al. (2016) analyzed the utility of circulating miRNAs to predict future fatal AMI in healthy participants. MiR-424-5p and miR-26a-5p were associated exclusively with risk in men and women, respectively, suggesting a gender-specific association. They discovered that the best model for predicting future AMI consisted of miR-106a-5p, miR-424-5p, let-7g-5p, miR-144-3p, and miR-660-5p, and these miRNAs were proposed as a panel to enhance the prediction of AMI risk in healthy individuals (Bye et al., 2016).

Platelet activation is critical for AMI pathogenesis, but the role of platelet miRNAs (pmiRNAs) as biomarkers in AMI and their correlation with indices of platelet activity are unclear. Assessment of pmiR-126 expression in STEMI patients revealed reduced levels and a correlation with plasma cTnI. However, pmiR-126 expression did not correlate well with platelet activity indices, and its potential diagnostic utility is limited (Li et al., 2017b).

MiR-1, miR-133a, and miR-34a induce adverse structural remodeling to impair cardiac contractile function. Increased levels of all three miRNAs have been shown in the hearts of old MI mice compared to young MI mice, and the miR-1 increase was more prolonged and corresponded to LV wall thinning. This suggests that significantly increased levels of miR-1 in the aged post-MI heart could be a biomarker for high-risk prediction (Qipshidze Kelm et al., 2018). In addition, miRNA-21 has been reported to be overexpressed in the serum of ancient patients with AMI and to positively correlate with serum levels of CK-MB and cTnI. *In vitro* experiments with human cardiomyocytes transfected with the miR-21 mimic short hairpin RNA have shown that, following tumor necrosis factor α (TNF-α) induction, apoptosis rates are downregulated. The upregulation of miR-21 expression in the serum of elderly patients with AMI inhibited apoptosis induced by TNF-α in human cardiomyocytes *via* activation of the JNK/p38/caspase-3 signaling pathway (Wang et al., 2017b). Along the same lines, cardiomyocyte apoptosis and hypoxic reduction of cell growth can be promoted by miR-23b overexpression, suggesting that it could be a potential biomarker for STEMI (Zhang et al., 2018a).

A recent study explored the diagnostic use of circulating miRNAs in patients with acute chest pain in the emergency department. They found that higher circulating miR-19b, miR-223, and miR-483-5p levels may be clinically useful for AMI diagnosis in early phases (Li et al., 2019). Similarly, circulating miR-17-5p, miR-126-5p, and miR-145-3p levels are elevated in plasma from AMI patients. Combining these three miRNAs achieves a more precise AMI diagnosis (Xue et al., 2019). Interestingly, next-generation miRNA sequencing from whole blood samples has been useful for identifying new biomarkers of MI (Kanuri et al., 2018).

Heart Failure

Heart failure is a chronic and progressive condition that hampers the ability of the heart to pump enough blood to the body and

fulfill its needs. Heart failure is caused by multiple disorders, such as hypertension, cardiomyopathy, MI, arrhythmias, or valvular diseases, among others (Khatibzadeh et al., 2013). Numerous scientific reports connect HF and epigenetic modifications (**Table 4** and **Figure 3**). High-density epigenome-wide mapping of DNA methylation in the myocardium and blood from dilated cardiomyopathy patients and healthy individuals has been analyzed. This technology has been used to find regions of epigenetic susceptibility and new biomarkers related to HF and heart dysfunction; they recognized different patterns of epigenetic methylation that were preserved through tissues—the CpGs regions identified as novel biomarkers of HF (Meder et al., 2017; Rau and Vondriska, 2017). Differentially methylated DNA regions were also identified in blood leukocytes from HF patients (Li et al., 2017a). Dilated cardiomyopathy is an important cause of HF. Genome-wide cardiac DNA methylation in idiopathic dilated cardiomyopathy patients revealed abnormal DNA methylation, which was related to important variations in the expression of lymphocyte antigen 75 (*LY75*) and adenosine receptor A2A (*ADORA2A*) mRNA (Haas et al., 2013). Similarly, genome-wide maps of DNA methylation and enrichment of histone 3 lysine-36 trimethylation (H3K36me3) in pathological and healthy hearts were analyzed. Differences in DNA methylation were found in promoter CpG islands, genes, intragenic CpG islands, and H3K36me3-rich regions of the genome. The promoters of upregulated genes had altered DNA methylation, but not the promoters of downregulated genes. In particular, an abundance of *DUX4* transcripts was associated with differences in DNA methylation and H3K36me3 enrichment. Although further studies need to be carried out, there is evidence that the expression of genes critical for the development of cardiomyopathies may be controlled by the epigenome (Movassagh et al., 2011). Moreover, in patients with dilated cardiomyopathy, there is an altered methylation pattern in the regulatory regions of cardiac development genes, such as T-box protein 5 (*TBX5*), heart and neural crest derivatives expressed 1 (*HAND1*), and NK2 homeobox 5 (*NKX2.5*) (Jo et al., 2016). Koczor et al. (2013) also studied the differential methylation patterns in patients with dilated cardiomyopathy, which is characterized by congestive HF. Computational analysis detected few differentially methylated gene promoters (*AURKB*, *BTNL9*, *CLDN5*, and *TK1*). This study provides relevant information on DNA methylation and altered expression in dilated cardiomyopathy that would help in treatment (Koczor et al., 2013).

Furthermore, epigenetic modifications have been proposed to play an important role in HF progression in the murine model of pressure overload. The researchers observed a reduction in sarcoplasmic reticulum Ca^{2+}ATPase (*Atp2a2*) levels and a significant induction of β-myosin-heavy chain (*Myh7*) mRNA levels. They also detected H3K4me2, H3K9me2, H3K27me3, and H3K36me2 and a reduction in the lysine-specific demethylase KDM2A after 8 weeks of transverse aortic constriction (Angrisano et al., 2014). *Atp2a2* is a determinant of cardiac function, and its reduced activity is a clear feature of HF. Gorski et al. (2019) investigated the role of lysine acetylation in *Atp2a2* function in HF patients and found that acetylation at lysine 492 is regulated by SIRT1 and HAT p300 and significantly reduced the gene activity (Gorski et al., 2019). All of this knowledge would be fundamental to identifying potential biomarkers and new epigenetic drugs in HF therapy. Interestingly, an association has been reported between epigenetic remodeling in the atrial natriuretic peptide (*ANP*) and *BNP* promoters and reactivation of the fetal gene program in HF. Their reported upregulation in HF patients did not respond to an increase in histone acetylation but HDAC4, which is exported from the nucleus. In contrast, demethylation of H3K9 and dissociation of heterochromatin protein 1 from gene promoters were regulated by HDAC4. Thus, HDAC4 is fundamental to histone methylation in HF caused by increased cardiac load and a potential target for treatment (Hohl et al., 2013). More recently, Glezeva et al. (2019) performed targeted DNA methylation sequencing to detect DNA methylation alterations in coding and ncRNA in cardiac interventricular septal tissue from HF patients. They found hypermethylation in *HEY2*, *MSR1*, *MYOM3*, *COX17*, and miR-24-1 and hypomethylation in *CTGF*, *MMP2*, and miR-155. Therefore, they defended a unique cohort of loci useful as diagnostic and therapeutic targets in HF (Glezeva et al., 2019).

More than 10 years ago, few reports suggested that specific miRNAs are differentially regulated in the failing heart (Divakaran and Mann, 2008; Small and Olson, 2011). Since then, an extensive evidence base has been published in the literature regarding the use of miRNAs as possible biomarkers for HF diagnosis and prognosis. In evaluating whether miRNAs can differentiate clinical HF from healthy individuals and from non-HF dyspnea, miRNA arrays have revealed miR423-5p enrichment in the blood of HF patients (Tijsen et al., 2010). However, criticisms have been raised in this study regarding age differences between groups, reduced sample size, and statistics (Kumarswamy et al., 2010). Moreover, patients with HF of different etiologies presented with different expression levels of circulating miRNAs. Ischemic HF patients were found to have a positive transcoronary gradient for miR-423-5p, miR-423, and miR-34a, but the nonischemic HF group was positive only for miR-21-3p and miR-30a. The transcoronary concentration gradient suggests that the failing heart may selectively release the miRNAs into the coronary circulation. These miRNAs could be useful for discriminating different etiologies of HF (Goldraich et al., 2014; De Rosa et al., 2018).

Circulating miRNAs have been screened in an attempt to identify any that could be used for the prognosis of ischemic HF in post-AMI patients. Knowing that p53 has been involved in HF development in mice (Sano et al., 2007), the authors took great interest in p53-responsive miRNAs. The serum levels of miR-34a, miR-192, and miR-194 were significantly and coordinately upregulated in AMI patients with ischemic HF progression, and all three were p53-responsive. Interestingly, these miRNAs were contained in extracellular vesicles, suggesting that they are circulating regulators of HF. Furthermore, there was a significant correlation between the LV end-diastolic dimension 1 year after AMI and the miR-194 and miR-34a expression levels. Thus, although further investigations are needed, these results suggest the usefulness of miR-34a, miR-192, and miR-194 in predicting the risk of HF progression after AMI (Evans and Mann, 2013; Matsumoto et al., 2013; Klenke et al., 2018).

TABLE 4 | Epigenetic biomarkers in heart failure.

Epigenetic modification	Biomarker	Regulation in heart failure	Sample source	Study type	References
DNA methylation	*LY75* and *ADORA2A*	Aberrant DNA methylation	Left ventricle myocardium and zebrafish	Clinical and experimental: zebrafish	(Haas et al., 2013)
	TBX5, *HAND1*, and *NKX2.5*	Altered DNA methylation	Myocardium	Clinical	(Jo et al., 2016)
	HEY2, *MSR1*, *MYOM3*, *COX17*, miR-24-1, *CTGF*, *MMP2*, miR-155	*HEY2*, *MSR1*, *MYOM3*, *COX17*, and miR-24-1: hypermethylation; *CTGF*, *MMP2*, and miR-155: hypomethylation	Myocardium	Clinical	(Glezeva et al., 2019)
Histone modifications	H3K36me3	H3K36me3 enhancement	Myocardium	Clinical	(Movassagh et al., 2011)
	HDAC4	HDAC4 activation	Myocardium	Clinical and experimental: mouse model of pressure overload	(Hohl et al., 2013)
	H3K4me2, H3K9me2, H3K27me3, H3K36me2, KDM2A	H3K4me2, H3K9me2, H3K27me3, and H3K36me2 methylation and KDM2A reduction	Myocardium	Experimental: mouse model of pressure overload	(Angrisano et al., 2014)
	Atp2a2	*Atp2a2* acetylation	Ventricular myocytes and myocardium	Clinical and experimental: mouse model of pressure overload in *MHC-SIRT1−/−* Tg mice and swine model of MI	(Gorski et al., 2019)
miRNA	miR423-5p	Higher levels	Plasma	Clinical	(Tijsen et al., 2010)
	miR-192	Upregulation	Serum	Clinical	(Matsumoto et al., 2013)
	miR-122*, miR-200b, miR-520d-5p, miR-622, miR-1228*, miR-558	miR-122*, miR-200b, miR-520d-5p, miR-622, miR-1228*: upregulation; miR-558: downregulation	Whole peripheral blood	Clinical	(Vogel et al., 2013)
	miR-103, miR-142-3p, miR-30b, miR-342-3p	Differentially expressed	Plasma	Clinical	(Ellis et al., 2013)
	miR-210, miR-30a	Upregulation	Serum	Clinical	(Zhao et al., 2013)
	miR-210	Higher levels	Plasma, mononuclear cells, and skeletal muscles	Clinical and experimental: Dahl salt-sensitive rats	(Endo et al., 2013)
	miR-1	Higher levels	Plasma	Clinical	(Zhang et al., 2013b)
	miR-423-5p	Positive transcoronary gradients	Transcoronary gradients	Clinical	(Goldraich et al., 2014)
	miR-423-5p	Lower levels	Plasma	Clinical	(Seronde et al., 2015)
	MiR-30c, miR-146a, miR-221, miR-328, miR-375	Downregulation	Serum	Clinical	(Watson et al., 2015)
	miR-21, miR-650, miR-744, miR-516-5p, miR-1292, miR-182, miR-1228, miR-595, miR-663b, miR-1296, miR-1825, miR-299-3p, miR-662 miR-122, miR-3148, miR-518e, miR-129-3p, miR-3155, miR-3175, miR-583, miR-568, miR-30d, miR-200a-star, miR-1979, miR-371-3p, miR-155-star, miR-502-5p	miR-21, miR-650, miR-744, miR-516-5p, miR-1292, miR-182, miR-1228, miR-595, miR-663b, miR-1296, miR-1825, miR-299-3p, miR-662 miR-122, miR-3148, miR-518e: increased; miR-129-3p, miR-3155, miR-3175, miR-583, miR-568, miR-30d, miR-200a-star, miR-1979, miR-371-3p, miR-155-star, miR-502-5p: decreased	Serum	Clinical	(Cakmak et al., 2015)
	miR-1233, miR-671-5p, miR-183-3p, miR-190a, miR-193b-3p, miR-193b-5p, miR-211-5p, miR-494	miR-1233, miR-671-5p: Upregulation; miR-183-3p, miR-190a, miR-193b-3p, miR-193b-5p, miR-211-5p, miR-494: downregulation	Whole blood and plasma	Clinical	(Wong et al., 2015)

(*Continued*)

TABLE 4 | Continued

Epigenetic modification	Biomarker	Regulation in heart failure	Sample source	Study type	References
	miR-1, miR-21	miR-1: downregulation; miR-21: upregulation	Serum	Clinical	(Sygitowicz et al., 2015)
	miR-126	Downregulation	Serum	Clinical	(Wei et al., 2015)
	miR-1, miR-21, miR-23, miR-29, miR-130, miR-195, miR-199	Upregulation	Myocardial biopsy	Clinical	(Lai et al., 2015)
	miR-106a-5p, miR-223-3p, miR-652-3p, miR-199a-3p, miR-18a-5p	Downregulation	Plasma	Clinical	(Vegter et al., 2016)
	miR-148b-3p, miR-409-3p	Downregulation	Serum and left atrial tissue	Clinical	(Chen et al., 2016)
	miR-122-5p, miR-184	Upregulation	H9C2 cells and blood and myocardium	Experimental: *in vitro* and rat model of post-MI HF	(Liu et al., 2016)
	miR-660-3p, miR-665, miR-1285-3p, miR-4491	Upregulation	Plasma and heart	Clinical	(Li et al., 2016)
	miR-18a-5p, miR-26b-5p, miR-27a-3p, miR-30e-5p, miR-106a-5p, miR-199a-3p, miR-652-3p	Lower levels	Plasma	Clinical	(Ovchinnikova et al., 2016)
	miR-19b	Lower levels	Serum and myocardial	Clinical and experimental: *in vitro*	(Beaumont et al., 2017)
	miR-30d	Lower levels	Serum	Clinical	(Xiao et al., 2017)
	miR-195-3p	Higher levels	Plasma	Clinical	(He et al., 2017)
	miR-22-3p	Higher levels	Blood	Clinical	(van Boven et al., 2017)
	miR-150-5p	Downregulation	Blood	Clinical	(Scrutinio et al., 2017)
	miR-133b-3p, miR-208b-3p, miR-125a-5p, miR-125b-5p, miR-126-3p, miR-21-5p, miR-210-3p, miR-29a-3p, miR-320a, miR-494-3p	Upregulation	Blood	Experimental: sheep model of HF	(Wong et al., 2017)
	miR-146a	Upregulation	Exosomal and total plasma	Clinical and experimental: *in vitro*	(Beg et al., 2017)
	miR-9, miR-495, miR-599, miR-181c	ex-miR-9, ex-miR-181c, ex-miR-495: increased; ex-miR-599: decreased	Exosomal and total plasma	Experimental: dogs with myxomatous mitral valve disease, mitral valve prolapse	(Yang et al., 2017b)
	miR-21-5p, miR-23a-3p, miR-222-3p	Higher levels	Plasma	Clinical and experimental: rat model of post-MI HF	(Dubois-Deruy et al., 2017)
	miRNA-21	Higher levels	Serum	Clinical	(Zhang et al., 2017a)
	miR-132	Higher levels	Plasma	Clinical	(Masson et al., 2018)
	miR-1254, miR-1306-5p	Higher levels	Blood	Clinical	(Bayés-Genis et al., 2018)
	miR-423, miR-34a, miR-21-3p, miR-30a	miR-21-3p, miR-30a: Positive transcoronary gradient in non-ischemic HF; miR-423, miR-34a: Negative transcoronary gradient in ischemic HF	Transcoronary gradients	Clinical	(De Rosa et al., 2018)
	miR-3135b, miR-3908, miR-5571-5p	Upregulation	Plasma	Clinical	(Chen et al., 2018a)
	miR-302b-3p	Higher levels	Plasma	Clinical	(Li et al., 2018a)
	exo-miR-92b-5p	increased	Serum	Clinical	(Wu et al., 2018c)
	miR-26b, miR-208b, miR-499	Higher levels	Peripheral blood mononuclear cells	Clinical	(Marketou et al., 2018)
	miR-423-5p, miR-221-5p, miR-212-5p, miR-193b-5p, miR-15a-5p, miR-208a-3p	Upregulation	Plasma, mouse myocardium and NRVMs cells	Clinical and experimental: *in vitro*, and murine model of hypertrophy and HF	(Shah et al., 2018b)
	miR-192	Upregulation	Serum	Clinical	(Klenke et al., 2018)

(*Continued*)

TABLE 4 | Continued

Epigenetic modification	Biomarker	Regulation in heart failure	Sample source	Study type	References
	miR-34a, miR-208b, miR-126, miR-24, miR-29a	miR-34a, miR-208b, miR-126: upregulation; miR-24, miR-29a: downregulation	Serum	Clinical	(Lakhani et al., 2018)
	miR-17, miR-20a, miR-106b	Lower levels	Plasma	Clinical	(Shah et al., 2018a)
	miR-197-5p	Upregulation	Plasma	Clinical	(Liu et al., 2018b)
	miR-133a, miR-221	Higher levels	Plasma	Clinical	(Guo et al., 2018)
	exo-miR-92b-5p	Higher levels	Serum	Clinical	(Wu et al., 2018b; Wu et al., 2018c)

Vogel et al. (2013) assessed the genome-wide miRNA expression profiles in HF patients with reduced ejection fraction (HFrEF). They demonstrated that dysregulated levels of miRNAs, such as miR-122*, miR-200b, miR-520d-5p, miR-622, miR-1228* (upregulated), or miR-558 (downregulated) significantly correlate with disease severity, as indicated by LV ejection fraction (Vogel et al., 2013). Moreover, Ellis et al. (2013) tried to find differences between HF patients and non–HF-related breathlessness, and between HFrEF and HF with preserved ejection fraction (HFpEF); although they found a differential expression of miR-103, miR-142-3p, miR-30b, and miR-342-3p in HF and breathless patients, individually, classical biomarkers such as NT-proBNP and hs-cTnT exhibited greater sensitivity and specificity. However, the combination of miRNAs with NT-proBNP significantly improved prediction performance (Ellis et al., 2013). Similarly, elevated plasma levels of miR-210 were reported in congestive HF patients, although no significant correlation was observed with BNP. However, patients with an improved BNP profile presented with low plasma miR-210 levels. MiR-210 might reflect a mismatch between heart contraction and oxygen demand in the peripheral tissues (Endo et al., 2013). Interestingly, miR-210 and miR-30a expression is upregulated in HF patients, with a tendency toward fetal values (Zhao et al., 2013). Moreover, changes in myocardial miRNA in patients with stable and end-stage HF partially resemble the fetal myocardium. Target mRNA levels negatively correlate with changes in highly expressed miRNAs in HF and fetal hearts. The circulation is dominated by miRNAs, fragments of tRNAs, and small cytoplasmic RNAs. Heart- and muscle-specific circulating miRNAs (myomirs) are also increased in advanced HF, correlating with cTnI levels. These findings support miRNA-based therapies and the use of circulating miRNAs as biomarkers for heart injury (Akat et al., 2014). Cardiac fibroblast–derived miRNAs, such as miR-660-3p, miR-665, miR-1285-3p, and miR-4491, have also been found to be significantly upregulated in heart and plasma during HF, discriminating patients from controls (Li et al., 2016). However, miRNAs in the pericardial fluid are not related to cardiovascular pathologies or clinically assessed stages of HF. MicroRNAs may be paracrine signaling factors that intervene in cardiac cells crosstalk (Kuosmanen et al., 2015).

In another study performed in patients with chronic congestive HF, microarray profiling demonstrated increased expression of miR-21, miR-122, miR-182, miR-299-3p, miR-516-5p, miR-518e, miR-595, miR-650, miR-662, miR-663b, miR-744, miR-1228, miR-1292, miR-1296, miR-1825, and miR-3148 and decreased expression of miR-30d, miR-129-3p, and miR-502-5p, miR-155-star miR-200a-star, miR-371-3p, miR-583, miR-568, miR-1979, miR-3155, and miR-3175. Among these miRNAs, miR-182 seemed to have a better prognostic value than hs-CRP (Cakmak et al., 2015). Furthermore, miR-30c, miR-146a, miR-221, miR-328, and miR-375 had different expression levels in HFrEF and HFpEF. The combination of two or more miRNAs with BNP could significantly improve the discrimination of these pathological conditions compared to BNP alone (Watson et al., 2015). Additional miRNAs have been identified as promising biomarkers to discriminate HF from healthy individuals and to differentiate HFrEF from HFpEF: miR-125a-5p, miR-183-3p, miR-190a, miR-193b-3p, miR-193b-5p, miR-211-5p, miR-494, miR-545-5p, miR-550a-5p, miR-638, miR-671-5p, miR-1233, miR-3135b, miR-3908, and miR-5571-5p. The use of a combination of miRNAs and NT-proBNP increases its discernment capacity (Schulte et al., 2015; Wong et al., 2015; Chen et al., 2018a). Similarly, increased levels of miR-133a and miR-221 can be used as suitable HF diagnostic biomarkers in elderly HF patients, and the combination of NT-proBNP and miR-133a can improve the diagnostic accuracy (Guo et al., 2018). Serum levels of miR-1, miR-21, and miR-208a have also been analyzed in symptomatic HF patients. Expression of miR-1 is reduced in symptomatic HF patients, with decreasing levels correlating with increasing severity. In contrast, miR-21 has been shown to be overexpressed with no relation to HF severity. No circulating miR-208a has been observed in symptomatic HF patients. A negative correlation between miR-1 expression and NT-proBNP has been reported in HF patients, whereas miR-21 and galectin-3 have been positively correlated. Therefore, dysregulated levels of miR-1 and miR-21 may be fundamental for HF progression (Sygitowicz et al., 2015). An inverse correlation between miR-1 levels and ejection fraction has also been reported. Thus, elevated levels of miR-1 may inhibit cardiac function and be a predictor of the onset of HF secondary to AMI (Zhang et al., 2013b).

MiR-126 has also been studied in atrial fibrillation and/or HF patients, with downregulated expression in patients and positive correlation with LV ejection fraction but a negative association with the cardiothoracic ratio and NT-proBNP. Thus, the reduction in miR-126 expression is a potential indicator of severity in atrial fibrillation and HF (Wei et al., 2015). A significant negative correlation has also been found between several miRNAs and classical clinical biomarkers indicative of a worse clinical

outcome in HF patients. MiR-16-5p has been correlated to CRP, miR-106a-5p to creatinine, miR-223-3p to growth differentiation factor 15, miR-652-3p to soluble ST-2, miR-199a-3p to procalcitonin and galectin-3, and miR-18a-5p to procalcitonin (Vegter et al., 2016). Furthermore, an analysis of myocyte and fibroblast-related miRNAs and mRNAs in myocardium samples from HF patients and control individuals revealed that miR-1, miR-21, miR-23, miR-29, miR-130, miR-195, and miR-199 are significantly upregulated in HF patients, whereas miR-30, miR-133, miR-208, and miR-320 do not significantly change. Related mRNAs, such as caspase 3, collagenase I, collagenase III, and transforming growth factor (TGF), are also upregulated in HF patients. MicroRNAs involved in apoptosis, hypertrophy, and fibrosis are upregulated in the myocardium of HF patients and may be suitable biomarkers in the early stages of chronic HF and future therapeutic targets (Lai et al., 2015).

Evaluation of miR-148b-3p and miR-409-3p in mitral regurgitation patients, asymptomatic mitral regurgitation patients, and controls revealed that circulating and tissue miR-148b-3p and circulating miR-409-3p are significantly downregulated in mitral regurgitation patients with HF, and miR-148b-3p is significantly downregulated only in the mitral regurgitation patients without HF. Notably, the mRNAs of target genes of both miRNAs have been shown to be upregulated in HF patients with mitral regurgitation. Thus, circulating miR-148b-3p may be used as a biomarker of HF and miR-409-3p for incident HF in mitral regurgitation patients (Chen et al., 2016).

Specific overexpression of miR-221 in the hearts of transgenic mice has been shown to induce cardiac dysfunction and HF by impairing autophagy. In addition, *in vitro* miR-221 upregulation inhibits autophagic vesicle formation. Thus, autophagy balance and cardiac remodeling are regulated by miR-221 levels through modulation of the p27/CDK2/mTOR axis, and miR-221 might be a therapeutic target in HF (Su et al., 2015). Furthermore, high-throughput sequencing has been used to determine the differential miRNA pattern in a rat model of post-MI HF. Upregulation of miR-122-5p and miR-184 was found in HF rats, describing a proapoptotic role of both miRNAs (Liu et al., 2016). In another study using the same model, the authors identified a significant increase in miR-21-5p, miR-23a-3p, and miR-222-3p and their target *SOD2* in the plasma and myocardium of HF rats. They showed a direct interaction between miR-222-3p and *SOD2*. An inhibition or increase in *SOD2* expression was found when human cardiomyocytes were transfected with miR-222-3p mimic or inhibitor, respectively (Dubois-Deruy et al., 2017).

Myocardial fibrosis–related miRNAs, such as miR-19b, are reduced in the myocardium and serum of HF patients with aortic stenosis. Inhibition of miR-19b in cultured human fibroblasts increases the expression of connective tissue growth factor protein and the enzyme lysyl oxidase (LOX). This could lead to excessive collagen fibril cross-linking and a subsequent increase in LV stiffness in aortic stenosis patients, particularly those with HF. Thus, miR-19b could be a biomarker of alterations in the myocardial collagen network (Beaumont et al., 2017).

Numerous studies have been performed to find miRNAs with a predictive value in HF patients. Increased levels of miR-1, miR-21, miR-21-5p, miR-22-3p, miR-29a-3p, miR30d, miR-125a-5p, miR-125b-5p, miR-126-3p, miR-133b-3p, miR-195-3p, miR-197-5P, miR-208b-3p, miR-210-3p, miR-302b-3p, miR-320a, and miR-494-3p (Zhang et al., 2013b; He et al., 2017; van Boven et al., 2017; Wong et al., 2017; Xiao et al., 2017; Zhang et al., 2017a; Li et al., 2018a; Liu et al., 2018b;) or decreased levels of miR-17, miR-18a-5p, miR-20a, miR-150, miR-26b-5p, miR-27a-3p, miR-30e-5p, miR-106a-5p, miR-106b, miR-150-5p, miR-199a-3p, miR-423-5p, and miR-652-3p (Seronde et al., 2015; Ovchinnikova et al., 2016; Scrutinio et al., 2017; Shah et al., 2018a; Lin et al., 2019) have been described as potential biomarkers in HF patients. These discoveries may serve to develop miRNA-based therapies and to identify new pharmacological targets.

Beg et al. (2017) measured exosomal and total plasma miRNAs separately in HF patients to distinguish between the transfer of biological materials for signaling alteration in distant organs (exosomal) and the level of tissue damage (plasma). They found that the circulating exosomal miR-146a/miR-16 ratio was higher in HF patients, with miR-146a induced in response to inflammation. These results suggest circulating exosomal miR-146a as a biomarker of HF (Beg et al., 2017). Moreover, elevation of exosomal miRNA exo-miR-92b-5p has been suggested as a potential biomarker for the diagnosis of HF (Wu et al., 2018b; Wu et al., 2018c). In a preclinical study in dogs with myxomatous mitral valve disease, dysregulation of exosomal miR-9, miR-495, and miR-599 was observed as the dogs aged. In addition, levels of miR-9, miR-599, miR-181c, and miR-495 changed in myxomatous mitral valve disease. Thus, the exosomal miRNA expression level appears to be more specific to disease states than total plasma miRNA (Yang et al., 2017b). Furthermore, the downregulation of miR-425 and miR-744 in the plasma exosomes has been shown to induce cardiac fibrosis by suppressing TGFβ1 expression (Wang et al., 2018a).

Circulating miR-132 levels increased in chronic HF with disease severity, and lower levels improve risk prediction for HF readmission beyond traditional risk factors, but not for mortality. MiR-132 may be useful for finding strategies that would reduce rehospitalization in HF patients (Masson et al., 2018; Panico and Condorelli, 2018). Moreover, in an exhaustive analysis of two independent cohorts using a strict quality evaluation for miRNA testing, an association was found between high levels of miR-1254 and miR-1306-5p and mortality and HF hospitalization in HF patients. However, these two circulating miRNAs were not shown to improve standard predictors of prognostication, such as age, sex, hemoglobin, renal function, and NT-proBNP (Bayés-Genis et al., 2018).

MiR-26b, miR-208b, and miR-499 expression levels have been assessed in peripheral blood mononuclear cells from hypertensive HFpEF patients to evaluate their association with their exercise capacity. All three miRNAs were expressed at higher levels in the patients group, but miR-208b showed the strongest correlations with cardiopulmonary exercise test parameters, including oxygen uptake, exercise duration, and the minute ventilation–carbon dioxide production relationship (Marketou et al., 2018). In a study performed in patients and a mice model of hypertrophy and HF, miRNAs dysregulation was shown to occur during HF development in animals, with downregulation of target genes. These miRNAs were associated with adverse LV remodeling in

humans, suggesting coordinated regulation of miRNA-mRNA. They also revealed target clusters of genes, such as autophagy, metabolism, and inflammation, implicated in HF mechanisms, (Shah et al., 2018b).

With the intention to establish a biomarkers panel useful for early detection of HF resulting from MI, Lakhani et al. (2018) found significant upregulation of miR-34a, miR-208b, miR-126, TGFβ-1, TNF-α, IL-6, and MMP-9 and reduced miR-24 and miR-29a levels. A positive association between IL-10 and ejection fraction in MI patients also suggested an important role of IL-10 in predicting HF (Lakhani et al., 2018).

Systems biology analyses of LV remodeling after MI allow molecular comprehensions; for example, miRNA modulation may be used as a marker of HF evolution. Two systems biology strategies were used to define an miRNA mark of LV remodeling in MI. They integrated either multiomics data (proteins and ncRNAs) produced from post-MI plasma or proteomic data generated from a rat model of MI. As a result, several miRNAs were associated with LV remodeling: miR-21-5p, miR-23a-3p, miR-222-3p, miR-17-5p, miR-21-5p, miR-26b-5p, miR-222-3p, miR-335-5p, and miR-375. These outcomes support the use of integrative systems biology analyses for the definition of miRNA marks of HF evolution (Charrier et al., 2019).

LIMITATIONS AND PERSPECTIVES OF THE EPIGENETIC BIOMARKERS

Limitations of the current field include the lack of large multicenter studies to provide convincing evidence for clinical applicability. Rather than a single ncRNA, it is likely that there will be patterns of different ncRNAs and other biomarkers (e.g., protein-based) that, together with machine-learning algorithms, will provide more sensitive and specific diagnostic and prognostic approaches to CVDs. Also, several technical challenges must be overcome before CE-marked ncRNA biomarkers will enter the clinical realm. DNA methylation and histone modifications are epigenetic mechanisms that have been reported to be sources of potential biomarkers useful in clinical practice. However, each CVD is regulated by multiple epigenetic pathways, and different CVDs are regulated by the same epigenetic mechanism, most of which are still under study. For example, hypermethylation of H3K79 (Rodriguez-Iturbe, 2006; Duarte et al., 2012) and *ACE2* promoter (Fan et al., 2017) in hypertensive patients has been described. Moreover, H3K4 and H3K9 were also hypermethylated in both mouse models of hypertension (Pojoga et al., 2011) and HF (Angrisano et al., 2014). This makes it difficult to select and implement a set of biomarkers for a particular CVD. Another potential problem is the quality of the samples, especially those obtained from collections in the pathology department. These samples are usually preserved in formaldehyde and paraffin, which highly degrades DNA. The stability, size, and integrity of a sample depend on the duration of fixation and storage (Kristensen et al., 2009). Thus, assessment of the quality of DNA is fundamental. However, the DNA methylation analysis can be performed successfully using polymerase chain reaction (PCR) methods with small amplicons in old samples (Tournier et al., 2012; Wong et al., 2014). In other cases, it is important to carefully adjust the protocol. It is also important to consider that frozen and paraffin-preserved samples may have different results, and they should not be compared without appropriate correction (García-Giménez et al., 2017).

Among the epigenetic biomarkers, miRNAs are the most promising, and numerous studies have been carried out in the last few years. The relatively easy detection and accessibility to samples in fluids, such as blood, urine, or saliva, make them very attractive. However, a few issues should be solved before their implementation in the clinical practice. The main problem is that miRNAs usually target multiple mRNAs from different genes, and one gene can be targeted by several miRNAs. This complex network should be deeply investigated before determining the use of a specific miRNA as a biomarker for the diagnosis or treatment of a particular disease (Akhtar et al., 2016). Regarding sample preparation, it is highly recommended to use plasma instead of whole blood, because if it is hemolyzed, the circulating miRNA content can be altered. Increasing the centrifugation time is also important in order to reduce platelet contamination (de Gonzalo-Calvo et al., 2017; García-Giménez et al., 2017).

Recently, great advances have been made to implement the new technology in the detection of new epigenetic biomarkers. However, a few concerns should be alleviated before their clinical implementation. Studies with big cohorts in different independent laboratories, using the same experimental design, sample preparation, methodology, and disease specifications, are necessary. Small patient cohorts should be considered as pilot studies before the validation of results in bigger sample analysis. The method of detection should be standardized for clinical application, and the clinical trials have to be randomized and prospective. It is also important to compare the new biomarkers with the classical biomarkers in order to validate them and determine their usefulness. The sensitivity and specificity for a certain disease also have to be determined for each biomarker (Engelhardt, 2012; García-Giménez et al., 2017). Regarding the method of DNA methylation detection, the luminometric methylation assay and the methylation analysis of CpG islands in repeatable elements (LINE-1) are widely used. Although there is a certain correlation with the measurements obtained with both methods, the comparison is not recommended, since a consistent bias between the results has been described (Knothe et al., 2016). Interestingly, a large multicenter study comparing DNA methylation assays compatible with routine clinical use has been performed. According to the authors, good agreement was observed between DNA methylation assays, which can be implemented in large-scale validation studies, development of new biomarkers, and clinical diagnostics (BLUEPRINT Consortium, 2016). The most used system to detect miRNAs is quantitative PCR, being the normalization protocol critical. Most laboratories use housekeeping genes or miRNAs as normalizers, changing their expression levels within serums. Another approach employs identical volumes of serum for all samples, generating different amounts of total RNA (Chen et al., 2008; Wang et al., 2009; Rockenbach et al., 2012). Both approaches include spike-in normalization, which consists of adding RNA of known sequence and quantity to calibrate measurements. However, spike-in normalization does not consider internal variation in circulating

miRNA between different individuals. Thus, a combination of both methods should always be performed to guarantee results reliability (van Empel et al., 2012). Polymerase chain reaction technology has to be performed with rigorous controls to avoid artifacts in the amplification step. To overcome this problem, digital PCR based on the amplification of one single molecule per reaction constitutes a valuable option (Hindson et al., 2013). Another attractive alternative for accurate measuring RNAs is the direct nucleic acid sequencing, although it is still expensive when considering large screening analysis (Kozomara and Griffiths-Jones, 2011). Finally, it is also important to understand the processes controlling miRNAs release and stability. The correlation between circulating and tissue miRNAs is not clear, and several studies indicate that miRNA levels in blood are not a reflection of changes in the tissue of origin. The reason is that miRNAs can also be produced by immune cells (Zheng et al., 2018).

CONCLUDING REMARKS

Over the past few years, a great amount of research has focused on epigenetics and its dynamic cross-talk with genetics. Unveiling a personalized epigenetic pattern can provide a large amount of information on epigenetic machinery that could be employed to tailor diagnosis and therapeutic strategies in CVDs. Recent advances in technology and data analysis have made it possible to create detailed epigenetic maps, which may represent a new tool in the clinical practice to discern cardiovascular risk beyond traditional risk determinants. Epigenetic information can also help in predicting individual drug responses. Importantly, epigenetic biomarkers are gaining ground in the scientific community as tools for the diagnosis and prognosis of CVDs. However, discrepancies in specific diagnostic biomarkers make replication of the current results in independent laboratories, with multiple research centers and a big sample size, mandatory. All of this will lead to a standardized clinical application in the near future.

AUTHOR CONTRIBUTIONS

CS-B and AB-G conceived the idea and wrote the manuscript with support from CG-M. CG-M performed the drawings and structure of the figures. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

This work was supported by grants from the Spanish Ministry of Economy and Competitiveness-MINECO (SAF2017-84324-C2-1-R), the Instituto de Salud Carlos III (PIC18/0014, PI18/00256), the Red de Terapia Celular–TerCel (RD16/0011/0006) and the CIBER Cardiovascular (CB16/11/00403) projects, as part of the Plan Nacional de I+D+I, and it was co-funded by ISCIII-Sudirección General de Evaluación y el Fondo Europeo de Desarrollo Regional (FEDER). This work was also funded by the Fundació La MARATÓ de TV3 (201516-10, 201502-20), the Generalitat de Catalunya (SGR2017 00483, SLT002/16/00234), the CERCA Programme/Generalitat de Catalunya, and "la Caixa" Banking Foundation.

ACKNOWLEDGMENTS

We apologize to all authors whose work could not be mentioned because of space limitations or inadvertent omissions. We are greatly grateful to Sonia V Forcales for her comments and discussion on epigenetic regulation.

REFERENCES

Adachi, T., Nakanishi, M., Otsuka, Y., Nishimura, K., Hirokawa, G., Goto, Y., et al. (2010). Plasma microRNA 499 as a biomarker of acute myocardial infarction. *Clin. Chem.* 56, 1183–1185. doi: 10.1373/clinchem.2010.144121

Ai, J., Zhang, R., Li, Y., Pu, J., Lu, Y., Jiao, J., et al. (2010). Circulating microRNA-1 as a potential novel biomarker for acute myocardial infarction. *Biochem. Biophys. Res. Commun.* 391, 73–77. doi: 10.1016/j.bbrc.2009.11.005

Akat, K. M., Moore-McGriff, D., Morozov, P., Brown, M., Gogakos, T., Correa Da Rosa, J., et al. (2014). Comparative RNA-sequencing analysis of myocardial and circulating small RNAs in human heart failure and their utility as biomarkers. *Proc. Natl. Acad. Sci. U. S. A* 111, 11151–11156. doi: 10.1073/pnas.1401724111

Akhtar, M. M., Micolucci, L., Islam, M. S., Olivieri, F., and Procopio, A. D. (2016). Bioinformatic tools for microRNA dissection. *Nucleic Acids Res.* 44, 24–44. doi: 10.1093/nar/gkv1221

Alavi-Moghaddam, M., Chehrazi, M., Alipoor, S. D., Mohammadi, M., Baratloo, A., Mahjoub, M. P., et al. (2018). A preliminary study of microRNA-208b after acute myocardial infarction: impact on 6-month survival. *Dis. Markers* 2018, 2410451–7. doi: 10.1155/2018/2410451

Alikhani-Koopaei, R., Fouladkou, F., Frey, F. J., and Frey, B. M. (2004). Epigenetic regulation of 11 beta-hydroxysteroid dehydrogenase type 2 expression. *J. Clin. Investig.* 114, 1146–1157. doi: 10.1172/JCI21647

Angrisano, T., Schiattarella, G. G., Keller, S., Pironti, G., Florio, E., Magliulo, F., et al. (2014). Epigenetic switch at atp2a2 and myh7 gene promoters in pressure overload-induced heart failure. *PLoS One* 9, e106024. doi: 10.1371/journal.pone.0106024

Authors/Task Force members, Windecker, S., Kolh, P., Alfonso, F., Collet, J.-P., Cremer, J., et al. (2014). 2014 ESC/EACTS guidelines on myocardial revascularization: the task force on myocardial revascularization of the European Society of Cardiology (ESC) and the European Association for Cardio-Thoracic Surgery (EACTS) developed with the special contribution of the European Association of Percutaneous Cardiovascular Interventions (EAPCI). *Eur. Heart J.* 35, 2541–2619. doi: 10.1093/eurheartj/ehu278

Babuin, L., and Jaffe, A. S. (2005). Troponin: the biomarker of choice for the detection of cardiac injury. *CMAJ* 173, 1191–1202. doi: 10.1503/cmaj/051291

Baccarelli, A., Rienstra, M., and Benjamin, E. J. (2010). Cardiovascular epigenetics: basic concepts and results from animal and human studies. *Circ. Cardiovasc. Genet.* 3, 567–573. doi: 10.1161/CIRCGENETICS.110.958744

Baptista, R., Marques, C., Catarino, S., Enguita, F. J., Costa, M. C., Matafome, P., et al. (2018). MicroRNA-424(322) as a new marker of disease progression in pulmonary arterial hypertension and its role in right ventricular hypertrophy by targeting SMURF1. *Cardiovasc. Res.* 114, 53–64. doi: 10.1093/cvr/cvx187

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233. doi: 10.1016/j.cell.2009.01.002

Bauters, C., Kumarswamy, R., Holzmann, A., Bretthauer, J., Anker, S. D., Pinet, F., et al. (2013). Circulating miR-133a and miR-423-5p fail as biomarkers for left ventricular remodeling after myocardial infarction. *Int. J. Cardiol.* 168, 1837–1840. doi: 10.1016/j.ijcard.2012.12.074

Bayés-Genis, A., Lanfear, D. E., de Ronde, M. W. J., Lupón, J., Leenders, J. J., Liu, Z., et al. (2018). Prognostic value of circulating microRNAs on heart failure–related morbidity and mortality in two large diverse cohorts of general heart failure patients. *Eur. J. Heart Fail.* 20, 67–75. doi: 10.1002/ejhf.984

Beaumont, J., López, B., Ravassa, S., Hermida, N., José, G. S., Gallego, I., et al. (2017). MicroRNA-19b is a potential biomarker of increased myocardial collagen cross-linking in patients with aortic stenosis and heart failure. *Sci. Rep.* 7, 40696. doi: 10.1038/srep40696

Beekman, M., Nederstigt, C., Suchiman, H. E. D., Kremer, D., van der Breggen, R., Lakenberg, N., et al. (2010). Genome-wide association study (GWAS)–identified disease risk alleles do not compromise human longevity. *Proc. Natl. Acad. Sci. U. S. A* 107, 18046–18049. doi: 10.1073/pnas.1003540107

Beg, F., Wang, R., Saeed, Z., Devaraj, S., Masoor, K., and Nakshatri, H. (2017). Inflammation-associated microRNA changes in circulating exosomes of heart failure patients. *BMC Res. Notes* 10, 751. doi: 10.1186/s13104-017-3090-y

Bekkering, S., van den Munckhof, I., Nielen, T., Lamfers, E., Dinarello, C., Rutten, J., et al. (2016). Innate immune cell activation and epigenetic remodeling in symptomatic and asymptomatic atherosclerosis in humans *in vivo*. *Atherosclerosis* 254, 228–236. doi: 10.1016/j.atherosclerosis.2016.10.019

Białek, S., Górko, D., Zajkowska, A., Kołtowski, Ł., Grabowski, M., Stachurska, A., et al. (2015). Release kinetics of circulating miRNA-208a in the early phase of myocardial infarction. *Kardiologia Polska* 73, 613–619. doi: 10.5603/KP.a2015.0067

Bildirici, A. E., Arslan, S., Özbilüm Şahin, N., Berkan, Ö., Beton, O., and Yilmaz, M. B. (2018). MicroRNA-221/222 expression in atherosclerotic coronary artery plaque versus internal mammarian artery and in peripheral blood samples. *Biomarkers* 23, 670–675. doi: 10.1080/1354750X.2018.1474260

BLUEPRINT Consortium. (2016). Quantitative comparison of DNA methylation assays for biomarker development and clinical applications. *Nat. Biotechnol.* 34, 726–737. doi: 10.1038/nbt.3605

Bogdarina, I., Welham, S., King, P. J., Burns, S. P., and Clark, A. J. L. (2007). Epigenetic modification of the renin–angiotensin system in the fetal programming of hypertension. *Circ. Res.* 100, 520–526. doi: 10.1161/01.RES.0000258855.60637.58

Bye, A., Røsjø, H., Nauman, J., Silva, G. J. J., Follestad, T., Omland, T., et al. (2016). Circulating microRNAs predict future fatal myocardial infarction in healthy individuals—the HUNT study. *J. Mol. Cell Cardiol.* 97, 162–168. doi: 10.1016/j.yjmcc.2016.05.009

Cakmak, H. A., Coskunpinar, E., Ikitimur, B., Barman, H. A., Karadag, B., Tiryakioglu, N. O., et al. (2015). The prognostic value of circulating microRNAs in heart failure: preliminary results from a genome-wide expression study. *J. Cardiovasc. Med. (Hagerstown)* 16, 431–437. doi: 10.2459/JCM.0000000000000233

Cao, J., and Yan, Q. (2012). Histone ubiquitination and deubiquitination in transcription, DNA damage response, and cancer. *Front. Oncol.* 2, 26. doi: 10.3389/fonc.2012.00026

Charrier, H., Cuvelliez, M., Dubois-Deruy, E., Mulder, P., Richard, V., Bauters, C., et al. (2019). Integrative system biology analyses identify seven microRNAs to predict heart failure. *Noncoding RNA* 5, E22–E30. doi: 10.3390/ncrna5010022

Chelbi, S. T., Mondon, F., Jammes, H., Buffat, C., Mignot, T.-M., Tost, J., et al. (2007). Expressional and epigenetic alterations of placental serine protease inhibitors: SERPINA3 is a potential marker of preeclampsia. *Hypertension* 49, 76–83. doi: 10.1161/01.HYP.0000250831.52876.cb

Chen, F., Yang, J., Li, Y., and Wang, H. (2018a). Circulating microRNAs as novel biomarkers for heart failure. *Hellenic J. Cardiol.* 59, 209–214. doi: 10.1016/j.hjc.2017.10.002

Chen, J., Xu, L., Hu, Q., Yang, S., Zhang, B., and Jiang, H. (2015a). MiR-17-5p as circulating biomarkers for the severity of coronary atherosclerosis in coronary artery disease. *Int. J. Cardiol.* 197, 123–124. doi: 10.1016/j.ijcard.2015.06.037

Chen, M.-C., Chang, T.-H., Chang, J.-P., Huang, H.-D., Ho, W.-C., Lin, Y.-S., et al. (2016). Circulating miR-148b-3p and miR-409-3p as biomarkers for heart failure in patients with mitral regurgitation. *Int. J. Cardiol.* 222, 148–154. doi: 10.1016/j.ijcard.2016.07.179

Chen, S., Chen, R., Zhang, T., Lin, S., Chen, Z., Zhao, B., et al. (2018b). Relationship of cardiovascular disease risk factors and noncoding RNAs with hypertension: a case-control study. *BMC Cardiovasc. Disord.* 18, 58. doi: 10.1186/s12872-018-0795-3

Chen, W., and Li, S. (2017). Circulating microRNA as a novel biomarker for pulmonary arterial hypertension due to congenital heart disease. *Pediatr. Cardiol.* 38, 86–94. doi: 10.1007/s00246-016-1487-3

Chen, X., Ba, Y., Ma, L., Cai, X., Yin, Y., Wang, K., et al. (2008). Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res.* 18, 997–1006. doi: 10.1038/cr.2008.282

Chen, X., Zhang, L., Su, T., Li, H., Huang, Q., Wu, D., et al. (2015b). Kinetics of plasma microRNA-499 expression in acute myocardial infarction. *J. Thorac. Dis.* 7, 890–896. doi: 10.3978/j.issn.2072-1439.2014.11.32

Cheng, C., Wang, Q., You, W., Chen, M., and Xia, J. (2014). MiRNAs as biomarkers of myocardial infarction: a meta-analysis. *PLoS ONE* 9, e88566. doi: 10.1371/journal.pone.0088566

Cho, H.-M., Lee, H.-A., Kim, H. Y., Han, H. S., and Kim, I. K. (2011). Expression of Na+-K+ -2Cl- cotransporter 1 is epigenetically regulated during postnatal development of hypertension. *Am. J. Hypertens.* 24, 1286–1293. doi: 10.1038/ajh.2011.136

Choi, J.-H., Nam, K.-H., Kim, J., Baek, M. W., Park, J.-E., Park, H.-Y., et al. (2005). Trichostatin A exacerbates atherosclerosis in low density lipoprotein receptor-deficient mice. *Arterioscler. Thromb. Vasc. Biol.* 25, 2404–2409. doi: 10.1161/01.ATV.0000184758.07257.88

Choi, S. Y., Yun, J., Lee, O. J., Han, H. S., Yeo, M. K., Lee, M. A., et al. (2013). MicroRNA expression profiles in placenta with severe preeclampsia using a PNA-based microarray. *Placenta* 34, 799–804. doi: 10.1016/j.placenta.2013.06.006

Cortez-Dias, N., Costa, M. C., Carrilho-Ferreira, P., Silva, D., Jorge, C., Calisto, C., et al. (2016). Circulating miR-122-5p/miR-133b ratio is a specific early prognostic biomarker in acute myocardial infarction. *Circ. J.* 80, 2183–2191. doi: 10.1253/circj.CJ-16-0568

Coskunpinar, E., Cakmak, H. A., Kalkan, A. K., Tiryakioglu, N. O., Erturk, M., and Ongen, Z. (2016). Circulating miR-221-3p as a novel marker for early prediction of acute myocardial infarction. *Gene* 591, 90–96. doi: 10.1016/j.gene.2016.06.059

Costantino, S., Libby, P., Kishore, R., Tardif, J.-C., El-Osta, A., and Paneni, F. (2018). Epigenetics and precision medicine in cardiovascular patients: from basic concepts to the clinical arena. *Eur. Heart J.* 39, 4150–4158. doi: 10.1093/eurheartj/ehx568

D'Alessandra, Y., Devanna, P., Limana, F., Straino, S., Di Carlo, A., Brambilla, P. G., et al. (2010). Circulating microRNAs are new and sensitive biomarkers of myocardial infarction. *Eur. Heart J.* 31, 2765–2773. doi: 10.1093/eurheartj/ehq167

Dasgupta, C., Chen, M., Zhang, H., Yang, S., and Zhang, L. (2012). Chronic hypoxia during gestation causes epigenetic repression of the estrogen receptor-α gene in ovine uterine arteries *via* heightened promoter methylation. *Hypertension* 60, 697–704. doi: 10.1161/HYPERTENSIONAHA.112.198242

de Gonzalo-Calvo, D., Cediel, G., Bär, C., Núñez, J., Revuelta-Lopez, E., Gavara, J., et al. (2018). Circulating miR-1254 predicts ventricular remodeling in patients with ST-segment-elevation myocardial infarction: a cardiovascular magnetic resonance study. *Sci. Rep.* 8, 15115. doi: 10.1038/s41598-018-33491-y

de Gonzalo-Calvo, D., Cenarro, A., Civeira, F., and Llorente-Cortés, V. (2016). microRNA expression profile in human coronary smooth muscle cell–derived microparticles is a source of biomarkers. *Clin. Investig. Arterioscler.* 28, 167–177. doi: 10.1016/j.arteri.2016.05.005

de Gonzalo-Calvo, D., Iglesias-Gutiérrez, E., and Llorente-Cortés, V. (2017). Biomarcadores epigenéticos y enfermedad cardiovascular: los microARN circulantes. *Rev. Española Cardiol.* 70, 763–769. doi: 10.1016/j.recesp.2017.02.027

De Rosa, S., Eposito, F., Carella, C., Strangio, A., Ammirati, G., Sabatino, J., et al. (2018). Transcoronary concentration gradients of circulating microRNAs in heart failure. *Eur. J. Heart Fail.* 20, 1000–1010. doi: 10.1002/ejhf.1119

Deddens, J. C., Vrijsen, K. R., Colijn, J. M., Oerlemans, M. I., Metz, C. H. G., van der Vlist, E. J., et al. (2016). Circulating extracellular vesicles contain miRNAs and are released as early biomarkers for cardiac injury. *J. Cardiovasc Transl. Res.* 9, 291–301. doi: 10.1007/s12265-016-9705-1

Devaux, Y., Vausort, M., Goretti, E., Nazarov, P. V., Azuaje, F., Gilson, G., et al. (2012). Use of circulating microRNAs to diagnose acute myocardial infarction. *Clin. Chem.* 58, 559–567. doi: 10.1373/clinchem.2011.173823

Devaux, Y., Vausort, M., McCann, G. P., Zangrando, J., Kelly, D., Razvi, N., et al. (2013). MicroRNA-150: a novel marker of left ventricular remodeling after

acute myocardial infarction. *Circ. Cardiovasc. Genet.* 6, 290–298. doi: 10.1161/CIRCGENETICS.113.000077

Divakaran, V., and Mann, D. L. (2008). The emerging role of microRNAs in cardiac remodeling and heart failure. *Circ. Res.* 103, 1072–1083. doi: 10.1161/CIRCRESAHA.108.183087

Dong, Y.-M., Liu, X.-X., Wei, G.-Q., Da, Y.-N., Cha, L., and Ma, C.-S. (2015). Prediction of long-term outcome after acute myocardial infarction using circulating miR-145. *Scand. J. Clin. Lab. Invest.* 75, 85–91. doi: 10.3109/00365513.2014.981855

Duarte, J. D., Zineh, I., Burkley, B., Gong, Y., Langaee, T. Y., Turner, S. T., et al. (2012). Effects of genetic variation in H3K79 methylation regulatory genes on clinical blood pressure and blood pressure response to hydrochlorothiazide. *J. Transl. Med.* 10, 56. doi: 10.1186/1479-5876-10-56

Dubois-Deruy, E., Cuvelliez, M., Fiedler, J., Charrier, H., Mulder, P., Hebbar, E., et al. (2017). MicroRNAs regulating superoxide dismutase 2 are new circulating biomarkers of heart failure. *Sci. Rep.* 7, 14747. doi: 10.1038/s41598-017-15011-6

Eitel, I., Adams, V., Dieterich, P., Fuernau, G., de Waha, S., Desch, S., et al. (2012). Relation of circulating MicroRNA-133a concentrations with myocardial damage and clinical prognosis in ST-elevation myocardial infarction. *Am. Heart J.* 164, 706–714. doi: 10.1016/j.ahj.2012.08.004

Ek, W. E., Hedman, Å. K., Enroth, S., Morris, A. P., Lindgren, C. M., Mahajan, A., et al. (2016). Genome-wide DNA methylation study identifies genes associated with the cardiovascular biomarker GDF-15. *Hum. Mol. Genet.* 25, 817–827. doi: 10.1093/hmg/ddv511

Ellis, K. L., Cameron, V. A., Troughton, R. W., Frampton, C. M., Ellmers, L. J., and Richards, A. M. (2013). Circulating microRNAs as candidate markers to distinguish heart failure in breathless patients. *Eur. J. Heart Fail.* 15, 1138–1147. doi: 10.1093/eurjhf/hft078

Endo, K., Naito, Y., Ji, X., Nakanishi, M., Noguchi, T., Goto, Y., et al. (2013). MicroRNA 210 as a biomarker for congestive heart failure. *Biol. Pharm. Bull.* 36, 48–54. doi: 10.1248/bpb.b12-00578

Engelhardt, S. (2012). Small RNA biomarkers come of age. *J. Am. Coll. Cardiol.* 60, 300–303. doi: 10.1016/j.jacc.2012.04.018

Eryılmaz, U., Akgullu, C., Beser, N., Yıldız, Ö., Kurt Ömürlü, İ., and Bozdogan, B. (2016). Circulating microRNAs in patients with ST-elevation myocardial infarction. *Anatol J Cardiol* 16, 392–396. doi: 10.5152/AnatolJCardiol.2015.6603

Evans, S., and Mann, D. L. (2013). Circulating p53-responsive microRNAs as predictive biomarkers in heart failure after acute myocardial infarction: the long and arduous road from scientific discovery to clinical utility. *Circ. Res.* 113, 242–244. doi: 10.1161/CIRCRESAHA.113.301951

Fagard, R. H. (2006). Exercise is good for your blood pressure: effects of endurance training and resistance training. *Clin. Exp. Pharmacol. Physiol.* 33, 853–856. doi: 10.1111/j.1440-1681.2006.04453.x

Fan, R., Mao, S.-Q., Gu, T.-L., Zhong, F.-D., Gong, M.-L., Hao, L.-M., et al. (2017). Preliminary analysis of the association between methylation of the ACE2 promoter and essential hypertension. *Mol. Med. Rep.* 15, 3905–3911. doi: 10.3892/mmr.2017.6460

FDA-NIH Biomarker Working Group. (2016). BEST (Biomarkers, EndpointS, and other Tools) Resource.

Feng, G., Yan, Z., Li, C., and Hou, Y. (2016). microRNA-208a in an early stage myocardial infarction rat model and the effect on cAMP-PKA signaling pathway. *Mol. Med. Rep.* 14, 1631–1635. doi: 10.3892/mmr.2016.5402

Findeisen, H. M., Gizard, F., Zhao, Y., Qing, H., Heywood, E. B., Jones, K. L., et al. (2011). Epigenetic regulation of vascular smooth muscle cell proliferation and neointima formation by histone deacetylase inhibition. *Arterioscler Thromb Vasc Biol.* 31, 851–860. doi: 10.1161/ATVBAHA.110.221952

Fiorito, G., Guarrera, S., Valle, C., Ricceri, F., Russo, A., Grioni, S., et al. (2014). B-vitamins intake, DNA-methylation of one carbon metabolism and homocysteine pathway genes and myocardial infarction risk: the EPICOR study. *Nutr. Metab. Cardiovasc. Dis.* 24, 483–488. doi: 10.1016/j.numecd.2013.10.026

Fraga, M. F., Ballestar, E., Paz, M. F., Ropero, S., Setien, F., Ballestar, M. L., et al. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl. Acad. Sci. U. S. A* 102, 10604–10609. doi: 10.1073/pnas.0500398102

Franceschini, N., and Le, T. H. (2014). Genetics of hypertension: discoveries from the bench to human populations. *Am. J. Physiol. Renal Physiol.* 306, F1–F11. doi: 10.1152/ajprenal.00334.2013

Friso, S., Pizzolo, F., Choi, S.-W., Guarini, P., Castagna, A., Ravagnani, V., et al. (2008). Epigenetic control of 11 beta-hydroxysteroid dehydrogenase 2 gene promoter is related to human hypertension. *Atherosclerosis* 199, 323–327. doi: 10.1016/j.atherosclerosis.2007.11.029

Gao, J., Yang, S., Wang, K., Zhong, Q., Ma, A., and Pan, X. (2019). Plasma miR-126 and miR-143 as potential novel biomarkers for cerebral atherosclerosis. *J. Stroke Cerebrovasc. Dis.* 28, 38–43. doi: 10.1016/j.jstrokecerebrovasdis.2018.09.008

García-Giménez, J. L., Mena-Mollá, S., Beltrán-García, J., and Sanchis-Gomar, F. (2017). Challenges in the analysis of epigenetic biomarkers in clinical samples. *Clin. Chem. Lab. Med.* 55, 1474–1477. doi: 10.1515/cclm-2016-1162

Gholikhani-Darbroud, R., Khaki-Khatibi, F., Mansouri, F., Hajahmadipoorrafsanjani, M., and Ghojazadeh, M. (2017). Decreased circulatory microRNA-4478 as a specific biomarker for diagnosing nonn–ST-segment elevation myocardial infarction (NSTEMI) and its association with soluble leptin receptor. *Bratisl. Lek. Listy* 118, 684–690. doi: 10.4149/BLL_2017_129

Glezeva, N., Moran, B., Collier, P., Moravec, C. S., Phelan, D., Donnellan, E., et al. (2019). Targeted DNA methylation profiling of human cardiac tissue reveals novel epigenetic traits and gene deregulation across different heart failure patient subtypes. *Circ. Heart Fail.* 12, e005765. doi: 10.1161/CIRCHEARTFAILURE.118.005765

Goldbergova, M. P., Lipkova, J., Fedorko, J., Sevcikova, J., Parenica, J., Spinar, J., et al. (2018). MicroRNAs in pathophysiology of acute myocardial infarction and cardiogenic shock. *Bratisl. Lek. Listy* 119, 341–347. doi: 10.4149/BLL_2018_064

Goldraich, L. A., Martinelli, N. C., Matte, U., Cohen, C., Andrades, M., Pimentel, M., et al. (2014). Transcoronary gradient of plasma microRNA 423-5p in heart failure: evidence of altered myocardial expression. *Biomarkers* 19, 135–141. doi: 10.3109/1354750X.2013.870605

Gorski, P. A., Jang, S. P., Jeong, D., Lee, A., Lee, P., Oh, J. G., et al. (2019). Role of SIRT1 in modulating acetylation of the sarco-endoplasmic reticulum Ca^{2+}-ATPase in heart failure. *Circ. Res.* 124, e63–e80. doi: 10.1161/CIRCRESAHA.118.313865

Goyal, R., Goyal, D., Leitzke, A., Gheorghe, C. P., and Longo, L. D. (2010). Brain renin–angiotensin system: fetal epigenetic programming by maternal protein restriction during pregnancy. *Reprod. Sci.* 17, 227–238. doi: 10.1177/1933719109351935

Grabmaier, U., Clauss, S., Gross, L., Klier, I., Franz, W. M., Steinbeck, G., et al. (2017). Diagnostic and prognostic value of miR-1 and miR-29b on adverse ventricular remodeling after acute myocardial infarction—the SITAGRAMI-miR analysis. *Int. J. Cardiol.* 244, 30–36. doi: 10.1016/j.ijcard.2017.06.054

Greißel, A., Culmes, M., Burgkart, R., Zimmermann, A., Eckstein, H.-H., Zernecke, A., et al. (2016). Histone acetylation and methylation significantly change with severity of atherosclerosis in human carotid plaques. *Cardiovasc. Pathol.* 25, 79–86. doi: 10.1016/j.carpath.2015.11.001

Guarrera, S., Fiorito, G., Onland-Moret, N. C., Russo, A., Agnoli, C., Allione, A., et al. (2015). Gene-specific DNA methylation profiles and LINE-1 hypomethylation are associated with myocardial infarction risk. *Clin. Epigenet.* 7, 133. doi: 10.1186/s13148-015-0164-3

Gunel, T., Hosseini, M. K., Gumusoglu, E., Kisakesen, H. I., Benian, A., and Aydinli, K. (2017). Expression profiling of maternal plasma and placenta microRNAs in preeclamptic pregnancies by microarray technology. *Placenta* 52, 77–85. doi: 10.1016/j.placenta.2017.02.019

Guo, M., Luo, J., Zhao, J., Shang, D., Lv, Q., and Zang, P. (2018). Combined use of circulating miR-133a and NT-proBNP improves heart failure diagnostic accuracy in elderly patients. *Med. Sci. Monit.* 24, 8840–8848. doi: 10.12659/MSM.911632

Guo, M.-L., Guo, L.-L., and Weng, Y.-Q. (2017). Implication of peripheral blood miRNA-124 in predicting acute myocardial infarction. *Eur. Rev. Med. Pharmacol. Sci.* 21, 1054–1059.

Haas, J., Frese, K. S., Park, Y. J., Keller, A., Vogel, B., Lindroth, A. M., et al. (2013). Alterations in cardiac DNA methylation in human dilated cardiomyopathy. *EMBO Mol. Med.* 5, 413–429. doi: 10.1002/emmm.201201553

Han, L., Zhao, Y., Luo, Q. Q., Liu, X. X., Lu, S. S., and Zou, L. (2017). The significance of miR-145 in the prediction of preeclampsia. *Bratisl. Lek. Listy* 118, 523–528. doi: 10.4149/BLL_2017_101

Han, Z., Zhang, L., Yuan, L., Liu, X., Chen, X., Ye, X., et al. (2015). Change of plasma microRNA-208 level in acute myocardial infarction patients and its clinical significance. *Ann. Transl. Med.* 3, 307. doi: 10.3978/j.issn.2305-5839.2015.10.25

Hao, X.-Z., and Fan, H.-M. (2017). Identification of miRNAs as atherosclerosis biomarkers and functional role of miR-126 in atherosclerosis progression through MAPK signalling pathway. *Eur. Rev. Med. Pharmacol. Sci.* 21, 2725–2733.

He, F., Lv, P., Zhao, X., Wang, X., Ma, X., Meng, W., et al. (2014). Predictive value of circulating miR-328 and miR-134 for acute myocardial infarction. *Mol. Cell Biochem* 394, 137–144. doi: 10.1007/s11010-014-2089-0

He, X., Ji, J., Wang, T., Wang, M.-B., and Chen, X.-L. (2017). Upregulation of circulating miR-195-3p in heart failure. *Cardiology* 138, 107–114. doi: 10.1159/000476029

Hindson, C. M., Chevillet, J. R., Briggs, H. A., Gallichotte, E. N., Ruf, I. K., Hindson, B. J., et al. (2013). Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat. Methods* 10, 1003–1005. doi: 10.1038/nmeth.2633

Hoareau-Aveilla, C., and Meggetto, F. (2017). Crosstalk between microRNA and DNA methylation offers potential biomarkers and targeted therapies in ALK-positive lymphomas. *Cancers (Basel)* 9, E100–E114. doi: 10.3390/cancers9080100

Hohl, M., Wagner, M., Reil, J.-C., Müller, S.-A., Tauchnitz, M., Zimmer, A. M., et al. (2013). HDAC4 controls histone methylation in response to elevated cardiac load. *J. Clin. Invest.* 123, 1359–1370. doi: 10.1172/JCI61084

Hromadnikova, I., Kotlabova, K., Ondrackova, M., Pirkova, P., Kestlerova, A., Novotna, V., et al. (2015). Expression profile of C19MC microRNAs in placental tissue in pregnancy-related complications. *DNA Cell Biol.* 34, 437–457. doi: 10.1089/dna.2014.2687

Hsu, A., Chen, S.-J., Chang, Y.-S., Chen, H.-C., and Chu, P.-H. (2014). Systemic approach to identify serum microRNAs as potential biomarkers for acute myocardial infarction. *Biomed. Res. Int.* 2014, 418628–418613. doi: 10.1155/2014/418628

Huang, L., Li, L., Chen, X., Zhang, H., and Shi, Z. (2016a). MiR-103a targeting Piezo1 is involved in acute myocardial infarction through regulating endothelium function. *Cardiol. J.* 23, 556–562. doi: 10.5603/CJ.a2016.0056

Huang, Y., Chen, J., Zhou, Y., Yu, X., Huang, C., Li, J., et al. (2016b). Circulating miR-30 is related to carotid artery atherosclerosis. *Clin. Exp. Hypertens.* 38, 489–494. doi: 10.3109/10641963.2016.1163370

Huang, Y., Tang, S., Ji-yan, C., Huang, C., Li, J., Cai, A.-P., et al. (2017a). Circulating miR-92a expression level in patients with essential hypertension: a potential marker of atherosclerosis. *J. Hum. Hypertens.* 31, 200–205. doi: 10.1038/jhh.2016.66

Huang, Y.-Q., Huang, C., Chen, J.-Y., Li, J., and Feng, Y.-Q. (2017b). Plasma expression level of miRNA let-7 is positively correlated with carotid intima-media thickness in patients with essential hypertension. *J. Hum. Hypertens.* 31, 843–847. doi: 10.1038/jhh.2017.52

Huang, Y.-Q., Li, J., Huang, C., and Feng, Y.-Q. (2018). Plasma microRNA-29c levels are associated with carotid intima-media thickness and is a potential biomarker for the early detection of atherosclerosis. *Cell. Physiol. Biochem.* 50, 452–459. doi: 10.1159/000494158

Huo, K.-G., Richer, C., Berillo, O., Mahjoub, N., Fraulob-Aquino, J. C., Barhoumi, T., et al. (2019). miR-431-5p knockdown protects against angiotensin II–induced hypertension and vascular injury. *Hypertension* 73, 1007–1017. doi: 10.1161/HYPERTENSIONAHA.119.12619

Iorio, M. V., Piovan, C., and Croce, C. M. (2010). Interplay between microRNAs and the epigenetic machinery: an intricate network. *Biochim. Biophys. Acta* 1799, 694–701. doi: 10.1016/j.bbagrm.2010.05.005

Irmak, M. K., and Sizlan, A. (2006). Essential hypertension seems to result from melatonin-induced epigenetic modifications in area postrema. *Med. Hypotheses* 66, 1000–1007. doi: 10.1016/j.mehy.2005.10.016

Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* 33 Suppl, 245–254. doi: 10.1038/ng1089

Jaffe, A. S., Babuin, L., and Apple, F. S. (2006). Biomarkers in acute cardiac disease: the present and the future. *J. Am. Coll. Cardiol.* 48, 1–11. doi: 10.1016/j.jacc.2006.02.056

Jairajpuri, D. S., Malalla, Z. H., Mahmood, N., and Almawi, W. Y. (2017). Circulating microRNA expression as predictor of preeclampsia and its severity. *Gene* 627, 543–548. doi: 10.1016/j.gene.2017.07.010

Jeong, H. S., Kim, J.-Y., Lee, S. H., Hwang, J., Shin, J. W., Song, K. S., et al. (2017). Synergy of circulating miR-212 with markers for cardiovascular risks to enhance estimation of atherosclerosis presence. *PLoS One* 12, e0177809. doi: 10.1371/journal.pone.0177809

Ji, X., Takahashi, R., Hiura, Y., Hirokawa, G., Fukushima, Y., and Iwai, N. (2009). Plasma miR-208 as a biomarker of myocardial injury. *Clin. Chem.* 55, 1944–1949. doi: 10.1373/clinchem.2009.125310

Jia, K., Shi, P., Han, X., Chen, T., Tang, H., and Wang, J. (2016). Diagnostic value of miR-30d-5p and miR-125b-5p in acute myocardial infarction. *Mol. Med. Rep.* 14, 184–194. doi: 10.3892/mmr.2016.5246

Jiang, D., Sun, M., You, L., Lu, K., Gao, L., Hu, C., et al. (2019). DNA methylation and hydroxymethylation are associated with the degree of coronary atherosclerosis in elderly patients with coronary heart disease. *Life Sci.* 224, 241–248. doi: 10.1016/j.lfs.2019.03.021

Jiang, F., Li, J., Wu, G., Miao, Z., Lu, L., Ren, G., et al. (2015). Upregulation of microRNA-335 and microRNA-584 contributes to the pathogenesis of severe preeclampsia through downregulation of endothelial nitric oxide synthase. *Mol. Med. Rep.* 12, 5383–5390. doi: 10.3892/mmr.2015.4018

Jiang, Y.-Z., Jiménez, J. M., Ou, K., McCormick, M. E., Zhang, L.-D., and Davies, P. F. (2014). Hemodynamic disturbed flow induces differential DNA methylation of endothelial Kruppel-like factor 4 promoter *in vitro* and *in vivo*. *Circ. Res.* 115, 32–43. doi: 10.1161/CIRCRESAHA.115.303883

Jiang, Y.-Z., Manduchi, E., Stoeckert, C. J., and Davies, P. F. (2015). Arterial endothelial methylome: differential DNA methylation in athero-susceptible disturbed flow regions *in vivo*. *BMC Genomics* 16, 506. doi: 10.1186/s12864-015-1656-4

Jin, P., Gu, W., Lai, Y., Zheng, W., Zhou, Q., and Wu, X. (2017). The circulating microRNA-206 level predicts the severity of pulmonary hypertension in patients with left heart diseases. *Cell. Physiol. Biochem.* 41, 2150–2160. doi: 10.1159/000475569

Jo, B.-S., Koh, I.-U., Bae, J.-B., Yu, H.-Y., Jeon, E.-S., Lee, H.-Y., et al. (2016). Methylome analysis reveals alterations in DNA methylation in the regulatory regions of left ventricle development genes in human dilated cardiomyopathy. *Genomics* 108, 84–92. doi: 10.1016/j.ygeno.2016.07.001

Kanuri, S. H., Ipe, J., Kassab, K., Gao, H., Liu, Y., Skaar, T. C., et al. (2018). Next generation MicroRNA sequencing to identify coronary artery disease patients at risk of recurrent myocardial infarction. *Atherosclerosis* 278, 232–239. doi: 10.1016/j.atherosclerosis.2018.09.021

Kato, N., Loh, M., Takeuchi, F., Verweij, N., Wang, X., Zhang, W., et al. (2015). Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. *Nat. Genet.* 47, 1282–1293. doi: 10.1038/ng.3405

Khatibzadeh, S., Farzadfar, F., Oliver, J., Ezzati, M., and Moran, A. (2013). Worldwide risk factors for heart failure: a systematic review and pooled analysis. *Int. J. Cardiol.* 168, 1186–1194. doi: 10.1016/j.ijcard.2012.11.065

Kim, H. J., Kim, S. Y., Lim, J. H., Kwak, D. W., Park, S. Y., and Ryu, H. M. (2015). Quantification and application of potential epigenetic markers in maternal plasma of pregnancies with hypertensive disorders. *IJMS* 16, 29875–29888. doi: 10.3390/ijms161226201

Kim, S. H., Kim, G. J., Umemura, T., Lee, S. G., and Cho, K. J. (2017). Aberrant expression of plasma microRNA-33a in an atherosclerosis-risk group. *Mol. Biol. Rep.* 44, 79–88. doi: 10.1007/s11033-016-4082-z

Klenke, S., Eul, S., Peters, J., Neumann, T., Adamzik, M., and Frey, U. H. (2018). Circulating miR-192 is a prognostic marker in patients with ischemic cardiomyopathy. *Future Cardiol.* 14, 283–289. doi: 10.2217/fca-2017-0108

Knothe, C., Shiratori, H., Resch, E., Ultsch, A., Geisslinger, G., Doehring, A., et al. (2016). Disagreement between two common biomarkers of global DNA methylation. *Clin. Epigenet.* 8, 60–17. doi: 10.1186/s13148-016-0227-0

Koczor, C. A., Lee, E. K., Torres, R. A., Boyd, A., Vega, J. D., Uppal, K., et al. (2013). Detection of differentially methylated gene promoters in failing and nonfailing human left ventricle myocardium using computation analysis. *Physiol. Genomics* 45, 597–605. doi: 10.1152/physiolgenomics.00013.2013

Kontaraki, J. E., Marketou, M. E., Zacharis, E. A., Parthenakis, F. I., and Vardas, P. E. (2014). MicroRNA-9 and microRNA-126 expression levels in patients with essential hypertension: potential markers of target-organ damage. *J. Am. Soc. Hypertens.* 8, 368–375. doi: 10.1016/j.jash.2014.03.324

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693–705. doi: 10.1016/j.cell.2007.02.005

Kozomara, A., and Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* 39, D152–D157. doi: 10.1093/nar/gkq1027

Krishnan, R., Mani, P., Sivakumar, P., Gopinath, V., and Sekar, D. (2017). Expression and methylation of circulating microRNA-510 in essential hypertension. *Hypertens. Res.* 40, 361–363. doi: 10.1038/hr.2016.147

Kristensen, L. S., Wojdacz, T. K., Thestrup, B. B., Wiuf, C., Hager, H., and Hansen, L. L. (2009). Quality assessment of DNA derived from up to 30 years old formalin fixed paraffin embedded (FFPE) tissue for PCR-based methylation analysis using SMART-MSP and MS-HRM. *BMC Cancer* 9, 453. doi: 10.1186/1471-2407-9-453

Kulkarni, A., Chavan-Gautam, P., Mehendale, S., Yadav, H., and Joshi, S. (2011). Global DNA methylation patterns in placenta and its association with maternal hypertension in pre-eclampsia. *DNA Cell Biol.* 30, 79–84. doi: 10.1089/dna.2010.1084

Kumar, A., Kumar, S., Vikram, A., Hoffman, T. A., Naqvi, A., Lewarchik, C. M., et al. (2013). Histone and DNA methylation–mediated epigenetic downregulation of endothelial Kruppel-like factor 2 by low-density lipoprotein cholesterol. *Arteriosclerosis, Thrombosis, and Vascular Biol.* 33, 1936–1942. doi: 10.1161/ATVBAHA.113.301765

Kumarswamy, R., Anker, S. D., and Thum, T. (2010). MicroRNAs as circulating biomarkers for heart failure: questions about MiR-423-5p. *Circ. Res.* 106, e8–author reply e9. doi: 10.1161/CIRCRESAHA.110.220616

Kuosmanen, S. M., Hartikainen, J., Hippeläinen, M., Kokki, H., Levonen, A.-L., and Tavi, P. (2015). MicroRNA profiling of pericardial fluid samples from patients with heart failure. *PLoS One* 10, e0119646. doi: 10.1371/journal.pone.0119646

Laganà, A. S., Vitale, S. G., Sapia, F., Valenti, G., Corrado, F., Padula, F., et al. (2018). miRNA expression for early diagnosis of preeclampsia onset: hope or hype? *J. Matern. Fetal. Neonatal. Med.* 31, 817–821. doi: 10.1080/14767058.2017.1296426

Lai, K.-B., Sanderson, J. E., Izzat, M. B., and Yu, C.-M. (2015). Micro-RNA and mRNA myocardial tissue expression in biopsy specimen from patients with heart failure. *Int. J. Cardiol.* 199, 79–83. doi: 10.1016/j.ijcard.2015.07.043

Lakhani, H. V., Khanal, T., Gabi, A., Yousef, G., Alam, M. B., Sharma, D., et al. (2018). Developing a panel of biomarkers and miRNA in patients with myocardial infarction for early intervention strategies of heart failure in West Virginian population. *PLoS One* 13, e0205329. doi: 10.1371/journal.pone.0205329

Lee, H.-A., Baek, I., Seok, Y. M., Yang, E., Cho, H.-M., Lee, D.-Y., et al. (2010). Promoter hypomethylation upregulates Na^+-K^+-$2Cl^-$ cotransporter 1 in spontaneously hypertensive rats. *Biochem. Biophys. Res. Commun.* 396, 252–257. doi: 10.1016/j.bbrc.2010.04.074

Li, B., Feng, Z.-H., Sun, H., Zhao, Z.-H., Yang, S.-B., and Yang, P. (2017a). The blood genome-wide DNA methylation analysis reveals novel epigenetic changes in human heart failure. *Eur. Rev. Med. Pharmacol. Sci.* 21, 1828–1836.

Li, G., Song, Y., Li, Y.-D., Jie, L.-J., Wu, W.-Y., Li, J.-Z., et al. (2018a). Circulating miRNA-302 family members as potential biomarkers for the diagnosis of acute heart failure. *Biomark. Med* 12, 871–880. doi: 10.2217/bmm-2018-0132

Li, H., Fan, J., Yin, Z., Wang, F., Chen, C., and Wang, D. W. (2016). Identification of cardiac-related circulating microRNA profile in human chronic heart failure. *Oncotarget* 7, 33–45. doi: 10.18632/oncotarget.6631

Li, K., Chen, Z., Qin, Y., and Wei, Y. (2018b). MiR-664a-3p expression in patients with obstructive sleep apnea: a potential marker of atherosclerosis. *Medicine (Baltimore)* 97, e9813. doi: 10.1097/MD.0000000000009813

Li, L., Li, S., Wu, M., Chi, C., Hu, D., Cui, Y., et al. (2019). Early diagnostic value of circulating microRNAs in patients with suspected acute myocardial infarction. *J. Cell Physiol.* 234, 13649–13658. doi: 10.1002/jcp.28045

Li, L.-M., Cai, W.-B., Ye, Q., Liu, J.-M., Li, X., and Liao, X.-X. (2014a). Comparison of plasma microRNA-1 and cardiac troponin T in early diagnosis of patients with acute myocardial infarction. *World. J. Emerg. Med.* 5, 182–186. doi: 10.5847/wjem.j.issn.1920-8642.2014.03.004

Li, S., Guo, L. Z., Kim, M. H., Han, J.-Y., and Serebruany, V. (2017b). Platelet microRNA for predicting acute myocardial infarction. *J. Thromb. Thrombolysis* 44, 556–564. doi: 10.1007/s11239-017-1537-6

Li, T., Cao, H., Zhuang, J., Wan, J., Guan, M., Yu, B., et al. (2011). Identification of miR-130a, miR-27b and miR-210 as serum biomarkers for atherosclerosis obliterans. *Clin. Chim. Acta* 412, 66–70. doi: 10.1016/j.cca.2010.09.029

Li, Y., Du, W., Zhao, R., Hu, J., Li, H., Han, R., et al. (2017c). New insights into epigenetic modifications in heart failure. *Front. Biosci. (Landmark Ed)* 22, 230–247. doi: 10.2741/4483

Li, Y.-Q., Zhang, M.-F., Wen, H.-Y., Hu, C.-L., Liu, R., Wei, H.-Y., et al. (2013). Comparing the diagnostic values of circulating microRNAs and cardiac troponin T in patients with acute myocardial infarction. *Clinics (Sao Paulo)* 68, 75–80. doi: 10.6061/clinics/2013(01)OA12

Li, Z., Lu, J., Luo, Y., Li, S., and Chen, M. (2014b). High association between human circulating microRNA-497 and acute myocardial infarction. *Sci. World J.* 2014, 931845–931847. doi: 10.1155/2014/931845

Liang, J., Bai, S., Su, L., Li, C., Wu, J., Xia, Z., et al. (2015). A subset of circulating microRNAs is expressed differently in patients with myocardial infarction. *Mol. Med. Rep.* 12, 243–247. doi: 10.3892/mmr.2015.3422

Lin, X., Zhang, S., and Huo, Z. (2019). Serum circulating miR-150 is a predictor of post–acute myocardial infarction heart failure. *Int. Heart J.* 60, 280–286. doi: 10.1536/ihj.18-306

Liu, G., Niu, X., Meng, X., and Zhang, Z. (2018a). Sensitive miRNA markers for the detection and management of NSTEMI acute myocardial infarction patients. *J. Thorac. Dis.* 10, 3206–3215. doi: 10.21037/jtd.2018.05.141

Liu, W., Zheng, J., Dong, J., Bai, R., Song, D., Ma, X., et al. (2018b). Association of miR-197-5p, a circulating biomarker for heart failure, with myocardial fibrosis and adverse cardiovascular events among patients with stage C or D heart failure. *Cardiology* 141, 212–225. doi: 10.1159/000493419

Liu, X., Dong, Y., Chen, S., Zhang, G., Zhang, M., Gong, Y., et al. (2015a). Circulating microRNA-146a and microRNA-21 predict left ventricular remodeling after ST-elevation myocardial infarction. *Cardiology* 132, 233–241. doi: 10.1159/000437090

Liu, X., Fan, Z., Zhao, T., Cao, W., Zhang, L., Li, H., et al. (2015b). Plasma miR-1, miR-208, miR-499 as potential predictive biomarkers for acute myocardial infarction: an independent study of Han population. *Exp. Gerontol.* 72, 230–238. doi: 10.1016/j.exger.2015.10.011

Liu, X., Meng, H., Jiang, C., Yang, S., Cui, F., and Yang, P. (2016). Differential microRNA Expression and regulation in the rat model of post-infarction heart failure. *PLoS One* 11, e0160920. doi: 10.1371/journal.pone.0160920

Liu, X., Yuan, L., Chen, F., Zhang, L., Chen, X., Yang, C., et al. (2017). Circulating miR-208b: a potentially sensitive and reliable biomarker for the diagnosis and prognosis of acute myocardial infarction. *Clin. Lab.* 63, 101–109. doi: 10.7754/Clin.Lab.2016.160632

Liu, Y., Liu, P., Yang, C., Cowley, A. W., and Liang, M. (2014). Base-resolution maps of 5-methylcytosine and 5-hydroxymethylcytosine in Dahl S rats: effect of salt and genomic sequence. *Hypertension* 63, 827–838. doi: 10.1161/HYPERTENSIONAHA.113.02637

Long, G., Wang, F., Duan, Q., Chen, F., Yang, S., Gong, W., et al. (2012a). Human circulating microRNA-1 and microRNA-126 as potential novel indicators for acute myocardial infarction. *Int. J. Biol. Sci.* 8, 811–818. doi: 10.7150/ijbs.4439

Long, G., Wang, F., Duan, Q., Yang, S., Chen, F., Gong, W., et al. (2012b). Circulating miR-30a, miR-195 and let-7b associated with acute myocardial infarction. *PLoS One* 7, e50926. doi: 10.1371/journal.pone.0050926

Luque, A., Farwati, A., Krupinski, J., and Aran, J. M. (2018). Association between low levels of serum miR-638 and atherosclerotic plaque vulnerability in patients with high-grade carotid stenosis. *J. Neurosurg.* 131, 1–8. doi: 10.3171/2018.2.JNS171899

Lv, P., Zhou, M., He, J., Meng, W., Ma, X., Dong, S., et al. (2014). Circulating miR-208b and miR-34a are associated with left ventricular remodeling after acute myocardial infarction. *IJMS* 15, 5774–5788. doi: 10.3390/ijms15045774

Ma, S.-C., Zhang, H.-P., Kong, F.-Q., Zhang, H., Yang, C., He, Y.-Y., et al. (2016). Integration of gene expression and DNA methylation profiles provides a molecular subtype for risk assessment in atherosclerosis. *Mol. Med. Rep.* 13, 4791–4799. doi: 10.3892/mmr.2016.5120

Maciejak, A., Kiliszek, M., Opolski, G., Segiet, A., Matlak, K., Dobrzycki, S., et al. (2016). miR-22-5p revealed as a potential biomarker involved in the acute phase of myocardial infarction *via* profiling of circulating microRNAs. *Mol. Med. Rep.* 14, 2867–2875. doi: 10.3892/mmr.2016.5566

Maciejak, A., Kostarska-Srokosz, E., Gierlak, W., Dluzniewski, M., Kuch, M., Marchel, M., et al. (2018). Circulating miR-30a-5p as a prognostic biomarker of left ventricular dysfunction after acute myocardial infarction. *Sci. Rep.* 8, 9883. doi: 10.1038/s41598-018-28118-1

Magenta, A., Sileno, S., D'Agostino, M., Persiani, F., Beji, S., Paolini, A., et al. (2018). Atherosclerotic plaque instability in carotid arteries: miR-200c as a promising biomarker. *Clin. Sci.* 132, 2423–2436. doi: 10.1042/CS20180684

Majchrzak-Celińska, A., Kosicka, K., Paczkowska, J., Główka, F. K., Bręborowicz, G. H., Krzyścin, M., et al. (2017). HSD11B2, RUNX3, and LINE-1 methylation

in placental DNA of hypertensive disorders of pregnancy patients. *Reprod. Sci.* 24, 1520–1531. doi: 10.1177/1933719117692043

Mao, Z., Wu, F., and Shan, Y. (2018). Identification of key genes and miRNAs associated with carotid atherosclerosis based on mRNA-seq data. *Medicine (Baltimore)* 97, e9832. doi: 10.1097/MD.0000000000009832

Marketou, M. E., Kontaraki, J. E., Maragkoudakis, S., Patrianakos, A., Konstantinou, J., Nakou, H., et al. (2018). MicroRNAs in peripheral mononuclear cells as potential biomarkers in hypertensive patients with heart failure with preserved ejection fraction. *Am. J. Hypertens.* 31, 651–657. doi: 10.1093/ajh/hpy035

Masson, S., Batkai, S., Beermann, J., Bär, C., Pfanne, A., Thum, S., et al. (2018). Circulating microRNA-132 levels improve risk prediction for heart failure hospitalization in patients with chronic heart failure. *Eur. J. Heart Fail.* 20, 78–85. doi: 10.1002/ejhf.961

Matsumoto, S., Sakata, Y., Suna, S., Nakatani, D., Usami, M., Hara, M., et al. (2013). Circulating p53-responsive microRNAs are predictive indicators of heart failure after acute myocardial infarction. *Circ. Res.* 113, 322–326. doi: 10.1161/CIRCRESAHA.113.301209

Meder, B., Haas, J., Sedaghat-Hamedani, F., Kayvanpour, E., Frese, K., Lai, A., et al. (2017). Epigenome-wide association study identifies cardiac gene patterning and a novel class of biomarkers for heart failure. *Circulation* 136, 1528–1544. doi: 10.1161/CIRCULATIONAHA.117.027355

Mei, Z., Huang, B., Mo, Y., and Fan, J. (2017). An exploratory study into the role of miR-204-5p in pregnancy-induced hypertension. *Exp Ther Med* 13, 1711–1718. doi: 10.3892/etm.2017.4212

Miyamoto, S., Kawamura, T., Morimoto, T., Ono, K., Wada, H., Kawase, Y., et al. (2006). Histone acetyltransferase activity of p300 is required for the promotion of left ventricular remodeling after myocardial infarction in adult mice *in vivo*. *Circulation* 113, 679–690. doi: 10.1161/CIRCULATIONAHA.105.585182

Moutinho, C., and Esteller, M. (2017). MicroRNAs and epigenetics. *Adv. Cancer Res.* 135, 189–220. doi: 10.1016/bs.acr.2017.06.003

Movassagh, M., Choy, M.-K., Knowles, D. A., Cordeddu, L., Haider, S., Down, T., et al. (2011). Distinct epigenomic features in end-stage failing human hearts. *Circulation* 124, 2411–2422. doi: 10.1161/CIRCULATIONAHA.111.040071

Mu, S., Shimosawa, T., Ogura, S., Wang, H., Uetake, Y., Kawakami-Mori, F., et al. (2011). Epigenetic modulation of the renal β-adrenergic-WNK4 pathway in salt-sensitive hypertension. *Nat. Med.* 17, 573–580. doi: 10.1038/nm.2337

Nagata, S., Marunouchi, T., and Tanonaka, K. (2019). Histone deacetylase inhibitor SAHA treatment prevents the development of heart failure after myocardial infarction *via an* induction of heat-shock proteins in rats. *Biol. Pharm. Bull.* 42, 453–461. doi: 10.1248/bpb.b18-00785

Nakatochi, M., Ichihara, S., Yamamoto, K., Naruse, K., Yokota, S., Asano, H., et al. (2017). Epigenome-wide association of myocardial infarction with DNA methylation sites at loci related to cardiovascular disease. *Clin. Epigenet.* 9, 54. doi: 10.1186/s13148-017-0353-3

Nishiguchi, T., Imanishi, T., and Akasaka, T. (2015). MicroRNAs and cardiovascular diseases. *Biomed. Res. Int.* 2015, 682857. doi: 10.1155/2015/682857

Noack, F., Ribbat-Idel, J., Thorns, C., Chiriac, A., Axt-Fliedner, R., Diedrich, K., et al. (2011). miRNA expression profiling in formalin-fixed and paraffin-embedded placental tissue samples from pregnancies with severe preeclampsia. *J. Perinat. Med.* 39, 267–271. doi: 10.1515/jpm.2011.012

Nuyt, A. M., and Alexander, B. T. (2009). Developmental programming and hypertension. *Curr. Opin. Nephrol. Hypertens.* 18, 144–152. doi: 10.1097/MNH.0b013e328326092c

O'Sullivan, J. F., Neylon, A., McGorrian, C., and Blake, G. J. (2016). miRNA-93-5p and other miRNAs as predictors of coronary artery disease and STEMI. *Int. J. Cardiol.* 224, 310–316. doi: 10.1016/j.ijcard.2016.09.016

Olivieri, F., Antonicelli, R., Lorenzi, M., D'Alessandra, Y., Lazzarini, R., Santini, G., et al. (2013). Diagnostic potential of circulating miR-499-5p in elderly patients with acute non ST-elevation myocardial infarction. *Int. J. Cardiol.* 167, 531–536. doi: 10.1016/j.ijcard.2012.01.075

Ovchinnikova, E. S., Schmitter, D., Vegter, E. L., ter, J. M., Valente, M. A. E., Liu, L. C. Y., et al. (2016). Signature of circulating microRNAs in patients with acute heart failure. *Eur. J. Heart Fail.* 18, 414–423. doi: 10.1002/ejhf.332

Panico, C., and Condorelli, G. (2018). microRNA-132: a new biomarker of heart failure at last? *Eur. J. Heart Fail.* 20, 86–88. doi: 10.1002/ejhf.1044

Parahuleva, M. S., Euler, G., Mardini, A., Parviz, B., Schieffer, B., Schulz, R., et al. (2017). Identification of microRNAs as potential cellular monocytic biomarkers in the early phase of myocardial infarction: a pilot study. *Sci. Rep.* 7, 15974. doi: 10.1038/s41598-017-16263-y

Parthenakis, F., Marketou, M., Kontaraki, J., Patrianakos, A., Nakou, H., Touloupaki, M., et al. (2017). Low levels of microRNA-21 are a marker of reduced arterial stiffness in well-controlled hypertension. *J. Clin. Hypertens. (Greenwich)* 19, 235–240. doi: 10.1111/jch.12900

Peng, L., Chun-guang, Q., Bei-fang, L., Xue-zhi, D., Zi-hao, W., Yun-fu, L., et al. (2014). Clinical impact of circulating miR-133, miR-1291 and miR-663b in plasma of patients with acute myocardial infarction. *Diagn. Pathol.* 9, 89. doi: 10.1186/1746-1596-9-89

Pereira-da-Silva, T., Coutinho Cruz, M., Carrusca, C., Cruz Ferreira, R., Napoleão, P., and Mota Carmo, M. (2018). Circulating microRNA profiles in different arterial territories of stable atherosclerotic disease: a systematic review. *Am J Cardiovasc Dis* 8, 1–13.

Pojoga, L. H., Williams, J. S., Yao, T. M., Kumar, A., Raffetto, J. D., do Nascimento, G. R. A., et al. (2011). Histone demethylase LSD1 deficiency during high-salt diet is associated with enhanced vascular contraction, altered NO-cGMP relaxation pathway, and hypertension. *AJP: Heart Circ. Physiol.* 301, H1862–H1871. doi: 10.1152/ajpheart.00513.2011

Qipshidze Kelm, N., Piell, K. M., Wang, E., and Cole, M. P. (2018). MicroRNAs as predictive biomarkers for myocardial injury in aged mice following myocardial infarction. *J. Cell Physiol.* 233, 5214–5221. doi: 10.1002/jcp.26283

Qiu, X.-K., and Ma, J. (2018). Alteration in microRNA-155 level correspond to severity of coronary heart disease. *Scand. J. Clin. Lab. Invest.* 78, 219–223. doi: 10.1080/00365513.2018.1435904

Rask-Andersen, M., Martinsson, D., Ahsan, M., Enroth, S., Ek, W. E., Gyllensten, U., et al. (2016). Epigenome-wide association study reveals differential DNA methylation in individuals with a history of myocardial infarction. *Hum. Mol. Genet.* 25, 4739–4748. doi: 10.1093/hmg/ddw302

Rau, C. D., and Vondriska, T. M. (2017). DNA methylation and human heart failure: mechanisms or prognostics. *Circulation* 136, 1545–1547. doi: 10.1161/CIRCULATIONAHA.117.029840

Reed, G. W., Rossi, J. E., and Cannon, C. P. (2017). Acute myocardial infarction. *Lancet* 389, 197–210. doi: 10.1016/S0140-6736(16)30677-8

Rezaei, M., Eskandari, F., Mohammadpour-Gharehbagh, A., Harati-Sadegh, M., Teimoori, B., and Salimi, S. (2018). Hypomethylation of the miRNA-34a gene promoter is associated with severe preeclampsia. *Clin. Exp. Hypertens.* 41, 1–5. doi: 10.1080/10641963.2018.1451534

Rivière, G., Lienhard, D., Andrieu, T., Vieau, D., Frey, B. M., and Frey, F. J. (2011). Epigenetic regulation of somatic angiotensin-converting enzyme by DNA methylation and histone acetylation. *Epigenetics* 6, 478–489. doi: 10.4161/epi.6.4.14961

Rockenbach, G., De Melo Neto, A. J., Barcellos, N. T., and Wolff, F. H. (2012). Ethnic differences in viral dominance patterns in patients with hepatitis B virus and hepatitis C virus dual infection. *Hepatology* 55, 1640–author reply 1640. doi: 10.1002/hep.24533

Rodriguez-Iturbe, B. (2006). Arteriolar remodeling in essential hypertension: are connective tissue growth factor and transforming growth factor involved? *Kidney Int.* 69, 1104–1105. doi: 10.1038/sj.ki.5000222

Rossetto, D., Avvakumov, N., and Côté, J. (2012). Histone phosphorylation: a chromatin modification involved in diverse nuclear events. *Epigenetics* 7, 1098–1108. doi: 10.4161/epi.21975

Sabia, C., Picascia, A., Grimaldi, V., Amarelli, C., Maiello, C., and Napoli, C. (2017). The epigenetic promise to improve prognosis of heart failure and heart transplantation. *Transplant. Rev. (Orlando)* 31, 249–256. doi: 10.1016/j.trre.2017.08.004

Sano, M., Minamino, T., Toko, H., Miyauchi, H., Orimo, M., Qin, Y., et al. (2007). p53-Induced inhibition of Hif-1 causes cardiac dysfunction during pressure overload. *Nature* 446, 444–448. doi: 10.1038/nature05602

Sato, F., Tsuchiya, S., Meltzer, S. J., and Shimizu, K. (2011). MicroRNAs and epigenetics. *FEBS J.* 278, 1598–1609. doi: 10.1111/j.1742-4658.2011.08089.x

Schulte, C., Westermann, D., Blankenberg, S., and Zeller, T. (2015). Diagnostic and prognostic value of circulating microRNAs in heart failure with preserved and reduced ejection fraction. *WJC* 7, 843–860. doi: 10.4330/wjc.v7.i12.843

Scrutinio, D., Conserva, F., Passantino, A., Iacoviello, M., Lagioia, R., and Gesualdo, L. (2017). Circulating microRNA-150-5p as a novel biomarker for advanced heart failure: a genome-wide prospective study. *J. Heart Lung Transplant.* 36, 616–624. doi: 10.1016/j.healun.2017.02.008

Sekar, D., Shilpa, B. R., and Das, A. J. (2017). Relevance of microRNA 21 in different types of hypertension. *Curr. Hypertens. Rep.* 19, 57. doi: 10.1007/s11906-017-0752-z

Seronde, M.-F., Vausort, M., Gayat, E., Goretti, E., Ng, L. L., Squire, I. B., et al. (2015). Circulating microRNAs and outcome in patients with acute heart failure. *PLoS One* 10, e0142237. doi: 10.1371/journal.pone.0142237

Shah, R. V., Rong, J., Larson, M. G., Yeri, A., Ziegler, O., Tanriverdi, K., et al. (2018a). Associations of circulating extracellular RNAs with myocardial remodeling and heart failure. *JAMA Cardiol.* 3, 871–876. doi: 10.1001/jamacardio.2018.2371

Shah, R., Ziegler, O., Yeri, A., Liu, X., Murthy, V., Rabideau, D., et al. (2018b). MicroRNAs associated with reverse left ventricular remodeling in humans identify pathways of heart failure progression. *Circ. Heart Fail.* 11, e004278. doi: 10.1161/CIRCHEARTFAILURE.117.004278

Shi, B., Guo, Y., Wang, J., and Gao, W. (2010). Altered expression of microRNAs in the myocardium of rats with acute myocardial infarction. *BMC Cardiovasc. Disord.* 10, 11. doi: 10.1186/1471-2261-10-11

Singh, K., Williams, J., Brown, J., Wang, E. T., Lee, B., Gonzalez, T. L., et al. (2017). Up-regulation of microRNA-202-3p in first trimester placenta of pregnancies destined to develop severe preeclampsia, a pilot study. *Pregnancy Hypertens.* 10, 7–9. doi: 10.1016/j.preghy.2017.04.002

Small, E. M., and Olson, E. N. (2011). Pervasive roles of microRNAs in cardiovascular biology. *Nature* 469, 336–342. doi: 10.1038/nature09783

Smolarek, I., Wyszko, E., Barciszewska, A. M., Nowak, S., Gawronska, I., Jablecka, A., et al. (2010). Global DNA methylation changes in blood of patients with essential hypertension. *Med. Sci. Monit.* 16, CR149–CR155.

Strimbu, K., and Tavel, J. A. (2010). What are biomarkers? *Curr. Opin. HIV AIDS* 5, 463–466. doi: 10.1097/COH.0b013e32833ed177

Su, M., Wang, J., Wang, C., Wang, X., Dong, W., Qiu, W., et al. (2015). MicroRNA-221 inhibits autophagy and promotes heart failure by modulating the p27/CDK2/mTOR axis. *Cell Death Differ.* 22, 986–999. doi: 10.1038/cdd.2014.187

Sun, T., Dong, Y.-H., Du, W., Shi, C.-Y., Wang, K., Tariq, M.-A., et al. (2017). The role of MicroRNAs in myocardial infarction: from molecular mechanism to clinical application. *IJMS* 18, 745. doi: 10.3390/ijms18040745

Sygitowicz, G., Tomaniak, M., Błaszczyk, O., Kołtowski, Ł., Filipiak, K. J., and Sitkiewicz, D. (2015). Circulating microribonucleic acids miR-1, miR-21 and miR-208a in patients with symptomatic heart failure: preliminary results. *Arch. Cardiovasc. Dis.* 108, 634–642. doi: 10.1016/j.acvd.2015.07.003

Talens, R. P., Jukema, J. W., Trompet, S., Kremer, D., Westendorp, R. G. J., Lumey, L. H., et al. (2012). Hypermethylation at loci sensitive to the prenatal environment is associated with increased incidence of myocardial infarction. *Int. J. Epidemiol.* 41, 106–115. doi: 10.1093/ije/dyr153

Thomas, H., Diamond, J., Vieco, A., Chaudhuri, S., Shinnar, E., Cromer, S., et al. (2018). Global atlas of cardiovascular disease 2000–2016: the path to prevention and control. *Glob. Heart* 13, 143–163. doi: 10.1016/j.gheart.2018.09.511

Tijsen, A. J., Creemers, E. E., Moerland, P. D., de Windt, L. J., van der Wal, A. C., Kok, W. E., et al. (2010). MiR423-5p as a circulating biomarker for heart failure. *Circ. Res.* 106, 1035–1039. doi: 10.1161/CIRCRESAHA.110.218297

Tournier, B., Chapusot, C., Courcet, E., Martin, L., Lepage, C., Faivre, J., et al. (2012). Why do results conflict regarding the prognostic value of the methylation status in colon cancers? The role of the preservation method. *BMC Cancer* 12, 12–12. doi: 10.1186/1471-2407-12-12

Ura, B., Feriotto, G., Monasta, L., Bilel, S., Zweyer, M., and Celeghini, C. (2014). Potential role of circulating microRNAs as early markers of preeclampsia. *Taiwan J. Obstet. Gynecol.* 53, 232–234. doi: 10.1016/j.tjog.2014.03.001

Valencia-Morales, M. D. P., Zaina, S., Heyn, H., Carmona, F. J., Varol, N., Sayols, S., et al. (2015). The DNA methylation drift of the atherosclerotic aorta increases with lesion progression. *BMC Med. Genomics* 8, 7. doi: 10.1186/s12920-015-0085-1

van Boven, N., Akkerhuis, K. M., Anroedh, S. S., Rizopoulos, D., Pinto, Y., Battes, L. C., et al. (2017). Serially measured circulating miR-22-3p is a biomarker for adverse clinical outcome in patients with chronic heart failure: the Bio-SHiFT study. *Int. J. Cardiol.* 235, 124–132. doi: 10.1016/j.ijcard.2017.02.078

van Empel, V. P. M., de Windt, L. J., and da Costa Martins, P. A. (2012). Circulating miRNAs: reflecting or affecting cardiovascular disease? *Curr. Hypertens. Rep.* 14, 498–509. doi: 10.1007/s11906-012-0310-7

Vegter, E. L., Ovchinnikova, E. S., van Veldhuisen, D. J., Jaarsma, T., Berezikov, E., van der Meer, P., et al. (2017). Low circulating microRNA levels in heart failure patients are associated with atherosclerotic disease and cardiovascular-related rehospitalizations. *Clin. Res. Cardiol.* 106, 598–609. doi: 10.1007/s00392-017-1096-z

Vegter, E. L., Schmitter, D., Hagemeijer, Y., Ovchinnikova, E. S., van der Harst, P., Teerlink, J. R., et al. (2016). Use of biomarkers to establish potential role and function of circulating microRNAs in acute heart failure. *Int. J. Cardiol.* 224, 231–239. doi: 10.1016/j.ijcard.2016.09.010

Vogel, B., Keller, A., Frese, K. S., Leidinger, P., Sedaghat-Hamedani, F., Kayvanpour, E., et al. (2013). Multivariate miRNA signatures as biomarkers for non-ischaemic systolic heart failure. *Eur. Heart J.* 34, 2812–2822. doi: 10.1093/eurheartj/eht256

Wang, C., Xu, G., Wen, Q., Peng, X., Chen, H., Zhang, J., et al. (2019a). CBS promoter hypermethylation increases the risk of hypertension and stroke. *Clinics (Sao Paulo)* 74, e630. doi: 10.6061/clinics/2019/e630

Wang, F., Demura, M., Cheng, Y., Zhu, A., Karashima, S., Yoneda, T., et al. (2014a). Dynamic CCAAT/enhancer binding protein-associated changes of DNA methylation in the angiotensinogen gene. *Hypertension* 63, 281–288. doi: 10.1161/HYPERTENSIONAHA.113.02303

Wang, F., Long, G., Zhao, C., Li, H., Chaugai, S., Wang, Y., et al. (2013a). Plasma microRNA-133a is a new marker for both acute myocardial infarction and underlying coronary artery stenosis. *J. Transl. Med.* 11, 222. doi: 10.1186/1479-5876-11-222

Wang, F., Long, G., Zhao, C., Li, H., Chaugai, S., Wang, Y., et al. (2014b). Atherosclerosis-related circulating miRNAs as novel and sensitive predictors for acute myocardial infarction. *PLoS One* 9, e105734. doi: 10.1371/journal.pone.0105734

Wang, F., Ma, Y., Wang, H., and Qin, H. (2017a). Reciprocal regulation between microRNAs and epigenetic machinery in colorectal cancer. *Oncol Lett.* 13, 1048–1057. doi: 10.3892/ol.2017.5593

Wang, H., Naghavi, M., Allen, C., Barber, R. M., Bhutta, Z. A., Carter, A., et al. (2016a). Articles Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 388, 1459–1544. doi: 10.1016/S0140-6736(16)31012-1

Wang, K., Zhang, S., Marzolf, B., Troisch, P., Brightman, A., Hu, Z., et al. (2009). Circulating microRNAs, potential biomarkers for drug-induced liver injury. *Proc. Natl. Acad. Sci. U. S. A* 106, 4402–4407. doi: 10.1073/pnas.0813371106

Wang, K.-J., Zhao, X., Liu, Y.-Z., Zeng, Q.-T., Mao, X.-B., Li, S.-N., et al. (2016b). Circulating MiR-19b-3p, MiR-134-5p and MiR-186-5p are promising novel biomarkers for early diagnosis of acute myocardial infarction. *Cell. Physiol. Biochem.* 38, 1015–1029. doi: 10.1159/000443053

Wang, L., Liu, J., Xu, B., Liu, Y.-L., and Liu, Z. (2018a). Reduced exosome miR-425 and miR-744 in the plasma represents the progression of fibrosis and heart failure. *Kaohsiung J. Med. Sci.* 34, 626–633. doi: 10.1016/j.kjms.2018.05.008

Wang, P., Shen, C., Diao, L., Yang, Z., Fan, F., Wang, C., et al. (2015). Aberrant hypermethylation of aldehyde dehydrogenase 2 promoter upstream sequence in rats with experimental myocardial infarction. *Biomed. Res. Int.* 2015, 503692–503613. doi: 10.1155/2015/503692

Wang, X., Falkner, B., Zhu, H., Shi, H., Su, S., Xu, X., et al. (2013b). A genome-wide methylation study on essential hypertension in young African American males. *PLoS One* 8, e53938. doi: 10.1371/journal.pone.0053938

Wang, Y., Chang, W., Zhang, Y., Zhang, L., Ding, H., Qi, H., et al. (2019b). Circulating miR-22-5p and miR-122-5p are promising novel biomarkers for diagnosis of acute myocardial infarction. *J. Cell Physiol.* 234, 4778–4786. doi: 10.1002/jcp.27274

Wang, Y., Yan, L., Zhang, Z., Prado, E., Fu, L., Xu, X., et al. (2018b). Epigenetic regulation and its therapeutic potential in pulmonary hypertension. *Front. Pharmacol.* 9, 241. doi: 10.3389/fphar.2018.00241

Wang, Y.-L., and Yu, W. (2018). Association of circulating microRNA-122 with presence and severity of atherosclerotic lesions. *PeerJ* 6, e5218. doi: 10.7717/peerj.5218

Wang, Z.-H., Sun, X.-Y., Li, C.-L., Sun, Y.-M., Li, J., Wang, L.-F., et al. (2017b). miRNA-21 expression in the serum of elderly patients with acute myocardial infarction. *Med. Sci. Monit.* 23, 5728–5734. doi: 10.12659/MSM.904933

Ward, J. A., Esa, N., Pidikiti, R., Freedman, J. E., Keaney, J. F., Tanriverdi, K., et al. (2013). Circulating cell and plasma microRNA profiles differ between non–ST-segment and ST-segment-elevation myocardial infarction. *Fam. Med. Med. Sci. Res.* 2, 108. doi: 10.4172/2327-4972.1000108

Watson, C. J., Gupta, S. K., O'Connell, E., Thum, S., Glezeva, N., Fendrich, J., et al. (2015). MicroRNA signatures differentiate preserved from reduced ejection fraction heart failure. *Eur. J. Heart Fail.* 17, 405–415. doi: 10.1002/ejhf.244

Wei, L., Zhao, S., Wang, G., Zhang, S., Luo, W., Qin, Z., et al. (2018). SMAD7 methylation as a novel marker in atherosclerosis. *Biochem. Biophys. Res. Commun.* 496, 700–705. doi: 10.1016/j.bbrc.2018.01.121

Wei, X. J., Han, M., Yang, F. Y., Wei, G. C., Liang, Z. G., Yao, H., et al. (2015). Biological significance of miR-126 expression in atrial fibrillation and heart failure. *Braz. J. Med. Biol. Res.* 48, 983–989. doi: 10.1590/1414-431x20154590

Wierda, R. J., Rietveld, I. M., van Eggermond, M. C. J. A., Belien, J. A. M., van Zwet, E. W., Lindeman, J. H. N., et al. (2015). Global histone H3 lysine 27 triple methylation levels are reduced in vessels with advanced atherosclerotic plaques. *Life Sci.* 129, 3–9. doi: 10.1016/j.lfs.2014.10.010

Wissler, R. W. (1991). Update on the pathogenesis of atherosclerosis. *Am. J. Med.* 91, 3S–9S. doi: 10.1016/0002-9343(91)90050-8

Wong, L. L., Armugam, A., Sepramaniam, S., Karolina, D. S., Lim, K. Y., Lim, J. Y., et al. (2015). Circulating microRNAs in heart failure with reduced and preserved left ventricular ejection fraction. *Eur. J. Heart Fail.* 17, 393–404. doi: 10.1002/ejhf.223

Wong, L. L., Rademaker, M. T., Saw, E. L., Lew, K. S., Ellmers, L. J., Charles, C. J., et al. (2017). Identification of novel microRNAs in the sheep heart and their regulation in heart failure. *Sci. Rep.* 7, 8250. doi: 10.1038/s41598-017-08574-x

Wong, S. Q., Li, J., Tan, A. Y.-C., Vedururu, R., Pang, J.-M. B., Do, H., et al. (2014). Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. *BMC Med. Genomics* 7, 23–10. doi: 10.1186/1755-8794-7-23

Wu, K., Zhao, Q., Li, Z., Li, N., Xiao, Q., Li, X., et al. (2018a). Bioinformatic screening for key miRNAs and genes associated with myocardial infarction. *FEBS Open Bio.* 8, 897–913. doi: 10.1002/2211-5463.12423

Wu, T., Chen, Y., Du, Y., Tao, J., Li, W., Zhou, Z., et al. (2018b). Circulating exosomal miR-92b-5p is a promising diagnostic biomarker of heart failure with reduced ejection fraction patients hospitalized for acute heart failure. *J. Thorac. Dis.* 10, 6211–6220. doi: 10.21037/jtd.2018.10.52

Wu, T., Chen, Y., Du, Y., Tao, J., Zhou, Z., and Yang, Z. (2018c). Serum exosomal MiR-92b-5p as a potential biomarker for acute heart failure caused by dilated cardiomyopathy. *Cell. Physiol. Biochem.* 46, 1939–1950. doi: 10.1159/000489383

Xiao, J., Gao, R., Bei, Y., Zhou, Q., Zhou, Y., Zhang, H., et al. (2017). Circulating miR-30d predicts survival in patients with acute heart failure. *Cell. Physiol. Biochem.* 41, 865–874. doi: 10.1159/000459899

Xiao, J., Shen, B., Li, J., Lv, D., Zhao, Y., Wang, F., et al. (2014). Serum microRNA-499 and microRNA-208a as biomarkers of acute myocardial infarction. *Int. J. Clin. Exp. Med.* 7, 136–141.

Xu, M., Li, J., Chen, X., Han, L., Li, L., and Liu, Y. (2019). MTHFD1 promoter hypermethylation increases the risk of hypertension. *Clin. Exp. Hypertens.* 41, 422–427. doi: 10.1080/10641963.2018.1501057

Xu, P., Zhao, Y., Liu, M., Wang, Y., Wang, H., Li, Y.-X., et al. (2014). Variations of microRNAs in human placentas and plasma from preeclamptic pregnancy. *Hypertension* 63, 1276–1284. doi: 10.1161/HYPERTENSIONAHA.113.02647

Xu, S., Pelisek, J., and Jin, Z. G. (2018). Atherosclerosis is an epigenetic disease. *Trends Endocrinol. Metab.* 29, 739–742. doi: 10.1016/j.tem.2018.04.007

Xue, S., Liu, D., Zhu, W., Su, Z., Zhang, L., Zhou, C., et al. (2019). Circulating MiR-17-5p, MiR-126-5p and MiR-145-3p are novel biomarkers for diagnosis of acute myocardial infarction. *Front. Physiol.* 10, 123. doi: 10.3389/fphys.2019.00123

Yamada, Y., Nishida, T., Horibe, H., Oguri, M., Kato, K., and Sawabe, M. (2014). Identification of hypo- and hypermethylated genes related to atherosclerosis by a genome-wide analysis of DNA methylation. *Int. J. Mol. Med.* 33, 1355–1363. doi: 10.3892/ijmm.2014.1692

Yan, Y., Song, X., Li, Z., Zhang, J., Ren, J., Wu, J., et al. (2017). Elevated levels of granzyme B correlated with miR-874-3p downregulation in patients with acute myocardial infarction. *Biomark. Med.* 11, 761–767. doi: 10.2217/bmm-2017-0144

Yang, G., Weng, X., Zhao, Y., Zhang, X., Hu, Y., Dai, X., et al. (2017a). The histone H3K9 methyltransferase SUV39H links SIRT1 repression to myocardial infarction. *Nat. Commun.* 8, 14941. doi: 10.1038/ncomms14941

Yang, Q., Jia, C., Wang, P., Xiong, M., Cui, J., Li, L., et al. (2014). MicroRNA-505 identified from patients with essential hypertension impairs endothelial cell migration and tube formation. *Int. J. Cardiol.* 177, 925–934. doi: 10.1016/j.ijcard.2014.09.204

Yang, Q., Lu, J., Wang, S., Li, H., Ge, Q., and Lu, Z. (2011). Application of next-generation sequencing technology to profile the circulating microRNAs in the serum of preeclampsia versus normal pregnant women. *Clin. Chim. Acta* 412, 2167–2173. doi: 10.1016/j.cca.2011.07.029

Yang, S.-Y., Wang, Y.-Q., Gao, H.-M., Wang, B., and He, Q. (2016a). The clinical value of circulating miR-99a in plasma of patients with acute myocardial infarction. *Eur. Rev. Med. Pharmacol. Sci.* 20, 5193–5197.

Yang, V. K., Loughran, K. A., Meola, D. M., Juhr, C. M., Thane, K. E., Davis, A. M., et al. (2017b). Circulating exosome microRNA associated with heart failure secondary to myxomatous mitral valve disease in a naturally occurring canine model. *J. Extracell Vesicles* 6, 1350088. doi: 10.1080/20013078.2017.1350088

Yang, W., Wang, A., Zhao, C., Li, Q., Pan, Z., Han, X., et al. (2016b). miR-125b enhances IL-8 production in early-onset severe preeclampsia by targeting sphingosine-1-phosphate lyase 1. *PLoS One* 11, e0166940. doi: 10.1371/journal.pone.0166940

Yao, X.-L., Lu, X.-L., Yan, C.-Y., Wan, Q.-L., Cheng, G.-C., and Li, Y.-M. (2015). Circulating miR-122-5p as a potential novel biomarker for diagnosis of acute myocardial infarction. *Int. J. Clin. Exp. Pathol.* 8, 16014–16019.

Yilmaz, S. G., Isbir, S., Kunt, A. T., and Isbir, T. (2018). Circulating microRNAs as novel biomarkers for atherosclerosis. *In Vivo* 32, 561–565. doi: 10.21873/invivo.11276

Yuan, L., Liu, X., Chen, F., Zhang, L., Chen, X., Huang, Q., et al. (2016). Diagnostic and prognostic value of circulating microRNA-133a in patients with acute myocardial infarction. *Clin. Lab.* 62, 1233–1241. doi: 10.7754/Clin.Lab.2015.151023

Zampetaki, A., Zeng, L., Margariti, A., Xiao, Q., Li, H., Zhang, Z., et al. (2010). Histone deacetylase 3 is critical in endothelial survival and atherosclerosis development in response to disturbed flow. *Circulation* 121, 132–142. doi: 10.1161/CIRCULATIONAHA.109.890491

Zhang, C., Li, Q., Ren, N., Li, C., Wang, X., Xie, M., et al. (2015a). Placental miR-106a~363 cluster is dysregulated in preeclamptic placenta. *Placenta* 36, 250–252. doi: 10.1016/j.placenta.2014.11.020

Zhang, D., Yu, Z.-Y., Cruz, P., Kong, Q., Li, S., and Kone, B. C. (2009). Epigenetics and the control of epithelial sodium channel expression in collecting duct. *Kidney Int.* 75, 260–267. doi: 10.1038/ki.2008.475

Zhang, J., Li, Y., and Zhao, Q. (2018a). Circulating miR-23b as a novel biomarker for early risk stratification after ST-elevation myocardial infarction. *Med. Sci. Monit.* 24, 1517–1523. doi: 10.12659/MSM.908060

Zhang, J., Xing, Q., Zhou, X., Li, J., Li, Y., Zhang, L., et al. (2017a). Circulating miRNA-21 is a promising biomarker for heart failure. *Mol. Med. Rep.* 16, 7766–7774. doi: 10.3892/mmr.2017.7575

Zhang, L. X., Du, J., Zhao, Y. T., Wang, J., Zhang, S., Dubielecka, P. M., et al. (2018b). Transgenic overexpression of active HDAC4 in the heart attenuates cardiac function and exacerbates remodeling in infarcted myocardium. *J. Appl. Physiol.* 125, 1968–1978. doi: 10.1152/japplphysiol.00006.2018

Zhang, L., Chen, X., Su, T., Li, H., Huang, Q., Wu, D., et al. (2015b). Circulating miR-499 are novel and sensitive biomarker of acute myocardial infarction. *J. Thorac. Dis.* 7, 303–308. doi: 10.3978/j.issn.2072-1439.2015.02.05

Zhang, L.-N., Liu, P.-P., Wang, L., Yuan, F., Xu, L., Xin, Y., et al. (2013a). Lower ADD1 gene promoter DNA methylation increases the risk of essential hypertension. *PLoS ONE* 8, e63455. doi: 10.1371/journal.pone.0063455

Zhang, M., Cheng, Y.-J., Sara, J. D., Liu, L.-J., Liu, L.-P., Zhao, X., et al. (2017b). Circulating microRNA-145 is associated with acute myocardial infarction and heart failure. *Chin. Med. J.* 130, 51–56. doi: 10.4103/0366-6999.196573

Zhang, R., Lan, C., Pei, H., Duan, G., Huang, L., and Li, L. (2015c). Expression of circulating miR-486 and miR-150 in patients with acute myocardial infarction. *BMC Cardiovasc. Disord.* 15, 51. doi: 10.1186/s12872-015-0042-0

Zhang, R., Niu, H., Ban, T., Xu, L., Li, Y., Wang, N., et al. (2013b). Elevated plasma microRNA-1 predicts heart failure after acute myocardial infarction. *Int. J. Cardiol.* 166, 259–260. doi: 10.1016/j.ijcard.2012.09.108

Zhang, W.-Q., and Xie, B.-Q. (2017). A meta-analysis of the relations between blood microRNA-208b detection and acute myocardial infarction. *Eur. Rev. Med. Pharmacol. Sci.* 21, 848–854.

Zhang, X., Wang, X., Wu, J., Peng, J., Deng, X., Shen, Y., et al. (2018c). The diagnostic values of circulating miRNAs for hypertension and bioinformatics analysis. *Biosci. Rep.* 38, BSR20180525. doi: 10.1042/BSR20180525

Zhang, Y., Cheng, J., Chen, F., Wu, C., Zhang, J., Ren, X., et al. (2017c). Circulating endothelial microparticles and miR-92a in acute myocardial infarction. *Biosci. Rep.* 37, BSR20170047. doi: 10.1042/BSR20170047

Zhang, Y., Liu, Y.-J., Liu, T., Zhang, H., and Yang, S.-J. (2016). Plasma microRNA-21 is a potential diagnostic biomarker of acute myocardial infarction. *Eur. Rev. Med. Pharmacol. Sci.* 20, 323–329.

Zhao, D.-S., Chen, Y., Jiang, H., Lu, J.-P., Zhang, G., Geng, J., et al. (2013). Serum miR-210 and miR-30a expressions tend to revert to fetal levels in Chinese adult patients with chronic heart failure. *Cardiovasc. Pathol.* 22, 444–450. doi: 10.1016/j.carpath.2013.04.001

Zhao, L., Chen, C.-N., Hajji, N., Oliver, E., Cotroneo, E., Wharton, J., et al. (2012). Histone deacetylation inhibition in pulmonary hypertension: therapeutic potential of valproic acid and suberoylanilide hydroxamic acid. *Circulation* 126, 455–467. doi: 10.1161/CIRCULATIONAHA.112.103176

Zheng, B., Xi, Z., Liu, R., Yin, W., Sui, Z., Ren, B., et al. (2018). The function of microRNAs in B-cell development, lymphoma, and their potential in clinical practice. *Front. Immunol.* 9, 936. doi: 10.3389/fimmu.2018.00936

Zhong, J., He, Y., Chen, W., Shui, X., Chen, C., and Lei, W. (2014). Circulating microRNA-19a as a potential novel biomarker for diagnosis of acute myocardial infarction. *IJMS* 15, 20355–20364. doi: 10.3390/ijms151120355

Zhu, J., Yao, K., Wang, Q., Guo, J., Shi, H., Ma, L., et al. (2016). Circulating miR-181a as a potential novel biomarker for diagnosis of acute myocardial infarction. *Cell. Physiol. Biochem.* 40, 1591–1602. doi: 10.1159/000453209

Zhu, L., Liu, F., Xie, H., and Feng, J. (2018). Diagnostic performance of microRNA-133a in acute myocardial infarction: a meta-analysis. *Cardiol. J.* 25, 260–267. doi: 10.5603/CJ.a2017.0126

Zhu, X.-M., Han, T., Sargent, I. L., Yin, G.-W., and Yao, Y.-Q. (2009). Differential expression profile of microRNAs in human placentas from preeclamptic pregnancies vs normal pregnancies. *Am. J. Obstet. Gynecol.* 200, 661.e1–661.e7. doi: 10.1016/j.ajog.2008.12.045

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Soler-Botija, Gálvez-Montón and Bayés-Genís. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

MINI REVIEW
published: 11 October 2019
doi: 10.3389/fgene.2019.00985

Identification of Biomarkers in Neuropsychiatric Disorders Based on Systems Biology and Epigenetics

*Jacob Peedicayil**

Department of Pharmacology and Clinical Pharmacology, Christian Medical College, Vellore, India

Clinically useful biomarkers are available for some neuropsychiatric disorders like fragile X syndrome, Rett syndrome, and Huntington's disease. Despite many decades of research on the pathogenesis of neuropsychiatric disorders like schizophrenia (SZ), bipolar disorder (BD), and major depressive disorder (MDD), the exact pathogenesis of these disorders remains unclear, and there are no clinically useful biomarkers for these disorders. However, there is increasing evidence that abnormal epigenetic mechanisms of gene expression contribute to the pathogenesis of SZ, BD, and MDD. Both systems (or network) biology and epigenetics (a component of systems biology) attempt to make sense of biological systems that are highly dynamic and multi-compartmental. This article suggests that systems biology, emphasizing the epigenetic component of systems biology, could help identify clinically useful biomarkers in neuropsychiatric disorders like SZ, BD, and MDD.

Keywords: biomarkers, epigenetic, network biology, neuropsychiatric disorders, systems biology

OPEN ACCESS

Edited by:
Momiao Xiong,
University of Texas Health Science Center, United States

Reviewed by:
Fu-Ying Tian,
Emory University, United States
Daniel Tarquinio,
Center for Rare Neurological Diseases, United States

***Correspondence:**
Jacob Peedicayil
jpeedi@cmcvellore.ac.in

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *02 May 2019*
Accepted: *17 September 2019*
Published: *11 October 2019*

Citation:
Peedicayil J (2019) Identification of Biomarkers in Neuropsychiatric Disorders Based on Systems Biology and Epigenetics. Front. Genet. 10:985. doi: 10.3389/fgene.2019.00985

INTRODUCTION

A biomarker, a short form for biological marker, has been defined as a feature that is objectively quantified and evaluated as an indicator of normal biological processes, pathological processes, or a pharmacological response to a therapeutic intervention (Biomarkers Definitions Working Group, 2001). In addition, there is another type of biomarker termed physiological biomarkers, which are indicators of the body's physiological functioning, such as heart rate, breathing rate, and the rate and pitch of speech (Adams et al., 2017). Biomarkers have many uses such as in the evaluation of drug effects in preclinical and clinical drug trials, in the diagnosis of patients with a disease, for staging diseases, as indicators of disease prognosis, and for predicting and monitoring clinical response to an intervention (Biomarkers Definitions Working Group, 2001).

Although clinically useful biomarkers are available for several medical disorders, as well as neuropsychiatric disorders like fragile X syndrome (FXS), Huntington's disease (HD), and Rett syndrome (RTT), there are at present none available for neuropsychiatric disorders like schizophrenia (SZ), bipolar disorder (BD), and major depressive disorder (MDD) (Davis et al., 2015; Kruse et al., 2017). The current article discusses the possible use of systems biology in the identification of biomarkers for neuropsychiatric disorders like SZ, BD, and MDD. Since epigenetics, like systems biology, attempts to make sense of biological systems that are highly dynamic and multi-compartmental (Housely et al., 2015), among the different components of systems biology, this article gives emphasis to the epigenetic component of systems biology. Another reason for the emphasis on the epigenetic component of systems biology is that there is increasing evidence that abnormal epigenetic mechanisms of gene expression play a crucial

role in the pathogenesis of neuropsychiatric disorders like SZ, BD, and MDD (Peedicayil and Grayson, 2018a; Peedicayil and Grayson, 2018b).

Types of Neuropsychiatric Disorders

Neuropsychiatric disorders comprise a wide range of disorders. They include neurological/neurosurgical disorders and psychiatric disorders. Although neurological and psychiatric disorders differ from each other in many ways, they are also similar to each other, and in some ways are like two sides of the same coin (Peedicayil et al., 2016a). It has been suggested that neurology and psychiatry are two sub-specialties of neuropsychiatry, which is part of the broader specialty of neurosciences (Peedicayil et al., 2016a).

Reductionism in Neuropsychiatric Disorders

Neuropsychiatric disorders are complex, heterogeneous disorders resulting from the interaction of various factors including genetic, epigenetic, neurobiological, and environmental factors (Lin and Huang, 2016). Can complex biological phenomena like neuropsychiatric symptoms like hallucinations, delusions, disorganized thinking, and mood swings be reduced to specific genes? Noted biologists like Lewontin (1991) and Rose (1995) suggest that psychiatric disorders cannot be reduced to specific genes. Strohman (1997) suggests that epigenetic defects underlying common disorders cannot be identified. He suggests that in future, genetic testing will be restricted to the rare disorders that show Mendelian inheritance. More recently, Drayna (2006) suggests that although simple human behaviors instinctive and crucial to survival and reproduction may be reducible to a set of genes, more generally, human behavior cannot be viewed as a product of a set of genes. Gold (2009) opines that research on the biology of psychiatric disorders is a gamble, like all scientific research. His answer to the question whether reduction is possible in psychiatry is that we will only know after the science has been done.

These workers' ideas appear to contradict those of Francis Crick (1966) who in *Of Molecules and Men* suggests that the ultimate aim of the modern movement of biology is to explain all biology in terms of physics and chemistry. Even Crick's colleague James Watson (2003) felt that the secret of life lies in the sequence of bases in DNA. Watson felt that there is no need to invoke vitalism (the theory that the origin and phenomena of life are determined by a force or principle distinct from purely physical or chemical forces) to explain life, and, instead, life can be explained by physicochemical processes. However, both Watson and Crick have been criticized by others (Lewontin, 1991; Strohman, 1997) for their extreme reductionist views.

It is significant that despite a lot of research spread across about a century, there is no conclusive and unambiguous evidence of consistent changes in biochemical (Kruse et al., 2017), neuropathological (Gandal et al., 2018), and neuroimaging studies (Brugger and Howes, 2017) of neuropsychiatric disorders like SZ, BD, and MDD. Presently, the best way to diagnose whether someone has such a disorder or not is to take a good history and conduct a good mental status, neurological, and physical examination (Kruse et al., 2017).

The Role of Epigenetics in Neuropsychiatric Disorders

A large amount of research on the epigenetics of neuropsychiatric disorders has been conducted over the past few decades. The data gathered so far have shown some interesting disparities (Peedicayil et al., 2016b): the role of epigenetics in the development of neuropsychiatric disorders with a major neurological component like FXS, HD, and RTT has been well characterized. However, in neuropsychiatric disorders with a major psychiatric component like SZ, BD, and MDD, the elucidation of the role of epigenetics in the development of disease is proving to be arduous. The reasons suggested for this disparity could be the following (Peedicayil et al., 2016b): the investigation of the role of epigenetics in neuropsychiatric disorders with a major neurological component started earlier; neuropsychiatric disorders with a greater neurological component are biologically less complex; there is a greater role played by environmental factors in the development of neuropsychiatric disorders with a greater psychiatric component. These three explanations could be related to each other (Peedicayil et al., 2016b).

Difficulties in Identifying Biomarkers in Neuropsychiatric Disorders

There are several difficulties in finding clinically useful biomarkers for many neuropsychiatric disorders. Liu (2016) has elegantly discussed these problems: First, for many neuropsychiatric disorders, we have a limited knowledge of the pathogenesis of the disorder, and the pathogenesis involves genetic, epigenetic, and environmental factors. Second, many neuropsychiatric disorders have subtypes. Hence, it is difficult to obtain specific, stable, and consistent biomarkers for clinical use. The variation in gene expression between cells, tissues, and patient populations makes identification of biomarkers difficult. Third, the use of the techniques, instruments, and machines for measuring disease parameters are complicated. Additionally, brain tissues are difficult to access, and peripheral tissues have to be used as proxies for brain tissues (Lin and Huang, 2016). Moreover, for many disorders like SZ, BD, and MDD, there are no suitable animal models (Lin and Huang, 2016).

There already are molecular tests for diagnosing some neuropsychiatric disorders. Such neuropsychiatric disorders have a greater neurological than a psychiatric component. They include RTT (Eyal et al., 2019), HD (Nance, 2017), and FXS (Wattendorf and Muenke, 2005). It must be noted that the molecular tests for these disorders involve genetic rather than epigenetic testing.

For the past several decades, a lot of research has been conducted to determine the genetic basis of neuropsychiatric disorders like SZ, BD, and MDD. Such research has the potential to throw light on the pathogenesis of these disorders and also identify genetic biomarkers for the disorders. Such research includes genetic linkage studies, genetic association studies, and genome-wide association studies (GWAS). So far, no genetic mutation or polymorphism predisposing to such disorders has been conclusively identified (Ebstein, 2018; Peedicayil and Grayson, 2018a; Peedicayil and Grayson, 2018b). In GWAS,

several associations have been identified (Ebstein, 2018; Peedicayil and Grayson, 2018a; Peedicayil and Grayson, 2018b). However, association does not imply causation (Altman and Krzywinski, 2015). Research on the epigenetic mechanisms underlying neuropsychiatric disorders like SZ, BD, and MDD has led to several findings (Guidotti et al., 2014; Ikegame et al., 2013; Kang et al., 2019; Liu et al., 2018; Mor et al., 2013; Tseng et al., 2014; Ziegler et al., 2016) (**Table 1**). However, these need confirmation and validation. In this context, it has been suggested that it would be a good idea to combine genetic and epigenetic data, as well as other "omic" data in order to distinguish signals from background noise and get a clearer picture about the pathogenesis of these disorders (Califano et al.2012; Feinberg, 2018; Wang et al., 2018).

Systems (Network) Biology and Neuropsychiatric Disorders

It is becoming increasingly clear that a clear biological function usually cannot be attributed to a single molecule. Instead, most biological traits arise from complex interactions between a cell's many constituents like DNA, RNA, and small molecules (Barabasi and Oltvai, 2004). A key challenge for biology in this century is to determine the structure and dynamics of the complex intercellular web of interactions contributing to the structure and functioning of a cell. Many types of interaction webs or networks emerge from a sum of these interactions. None of these networks are independent. Instead, they form a "network of networks" that is responsible for the behavior of a cell. A major challenge of contemporary biology is to theoretically and experimentally map out, understand, and model, in quantifiable terms, the topological (structural) and dynamic properties of the various networks that control the behavior of a cell (Barabasi and Oltvai, 2004).

The new area of systems or network biology could provide a solution for this challenge. Systems biology was pioneered by the noted scientist Leroy Hood using the galactose gene regulatory circuit in the budding yeast *Saccharomyces cerevisiae* (Ideker and Hood, 2019). Systems biology regards biology as an information science, and investigates biological systems as a whole, including their interactions with the environment (Wang et al., 2010). It evolved from the field of systems engineering in which a linked collection of component parts constitute a network whose output the engineer wishes to predict. It refers to a comprehensive quantitative analysis of the manner in which all components of a biological system interact functionally over time (Aderem, 2005). Major developments in technology have taken place since the 1980s. They include automated DNA sequencing, microarray analysis, advances in mass spectrometry, next-generation sequencing, and the internet. The knowledge of the complete sequences of genomes, along with technology allowing the monitoring of the flow of information resulting in specific cell functions, enabled systems biology to develop (Aderem, 2005), a discipline that may change the intellectual and experimental landscape on which we stand (Hiesinger and Hassan, 2005).

All systems can be analyzed by defining their static topology (architecture) and their dynamic (time-dependent) response to perturbation (Loscalzo, 2018). Any system of interacting elements can be schematically represented as a network comprising individual elements (nodes) connected by edges. The nature of the edges reflects the degree of complexity of the system. In simple systems, the nodes are linked linearly with a few feedback or feed-forward loops modulating the system in predictable ways. In complex systems, the nodes are linked in complicated non-linear networks. An important property of complex systems is that simplifying their structures by identifying and characterizing their individual nodes or edges or simple sub-structures need not yield a predictable understanding of a system's behavior. Hence, the system is greater than, or different from, the sum of its individual parts (Loscalzo, 2018).

Systems biology will help us attain a more holistic picture of disease states and could vindicate the reductionist approach to biology (Hiesinger and Hassan, 2005). It will not only facilitate basic biological research but also provide new ways to understand human diseases, identify biomarkers, and develop treatments for diseases (Wang et al., 2015). Moreover, systems biology may help answer questions related to complex organs like the brain, questions which cannot be answered with only the currently available tools of molecular biology and genomics (Villoslada et al., 2009).

TABLE 1 | A partial list of epigenetic changes in some neuropsychiatric disorders.

Sl. No.	Epigenetic Change	Disorder	Tissue
1.	Hypermethylated GAD1, RELN genes	SZ, BD	Peripheral blood cells
2.	Hypermethylated BDNF gene	SZ, BD, MDD	Peripheral blood
3.	Elevated miR-382-5p	SZ	Olfactory epithelium
4.	Several miRNAs	SZ	PBMC
5.	FKBP5 hypermethylation	PTSD	Peripheral blood
6.	Hypomethylated MAOA gene	PD	Peripheral blood
7.	Decreased 5-hmC	MDD	Leukocytes

BDNF, Brain-derived neurotrophic factor; BD, Bipolar disorder; GAD1, Glutamic acid decarboxylase1; hmC, Hydroxymethylcytosine; MDD, Major depressive disorder; miRNA, microRNA; PBMC, Peripheral blood mononuclear cells; PD, Panic disorder; PTSD, Post-traumatic stress disorder; SZ, Schizophrenia. References: 1. Guidotti et al. (2014); 2. Ikegame et al. (2013); 3. Mor et al. (2013); 4. Liu et al. (2018); 5. Kang et al. (2019); 6. Ziegler et al. (2016); 7. Tseng et al. (2014).

Systems Biology and Biomarkers in Neuropsychiatric Disorders

Systems biology could help identify biomarkers for neuropsychiatric disorders (**Figure 1**). As discussed by Lausted et al. (2014), the challenge in identifying biomarkers for complex disorders is to distinguish a small signal from a large amount of noise. The usual approach to blood-based biomarker discovery is to compare molecular profiles of blood samples from normal individuals with those from patients. Inevitably, large numbers of differences are found. However, a lot of these biomarkers is noise (Köhler and Seitz, 2012). A systems approach to biomarkers provides powerful tools for distinguishing signals from noise (Ideker et al., 2011; Lausted et al., 2014). This is because networks provide a distinct and rational framework for describing interactions between genes,

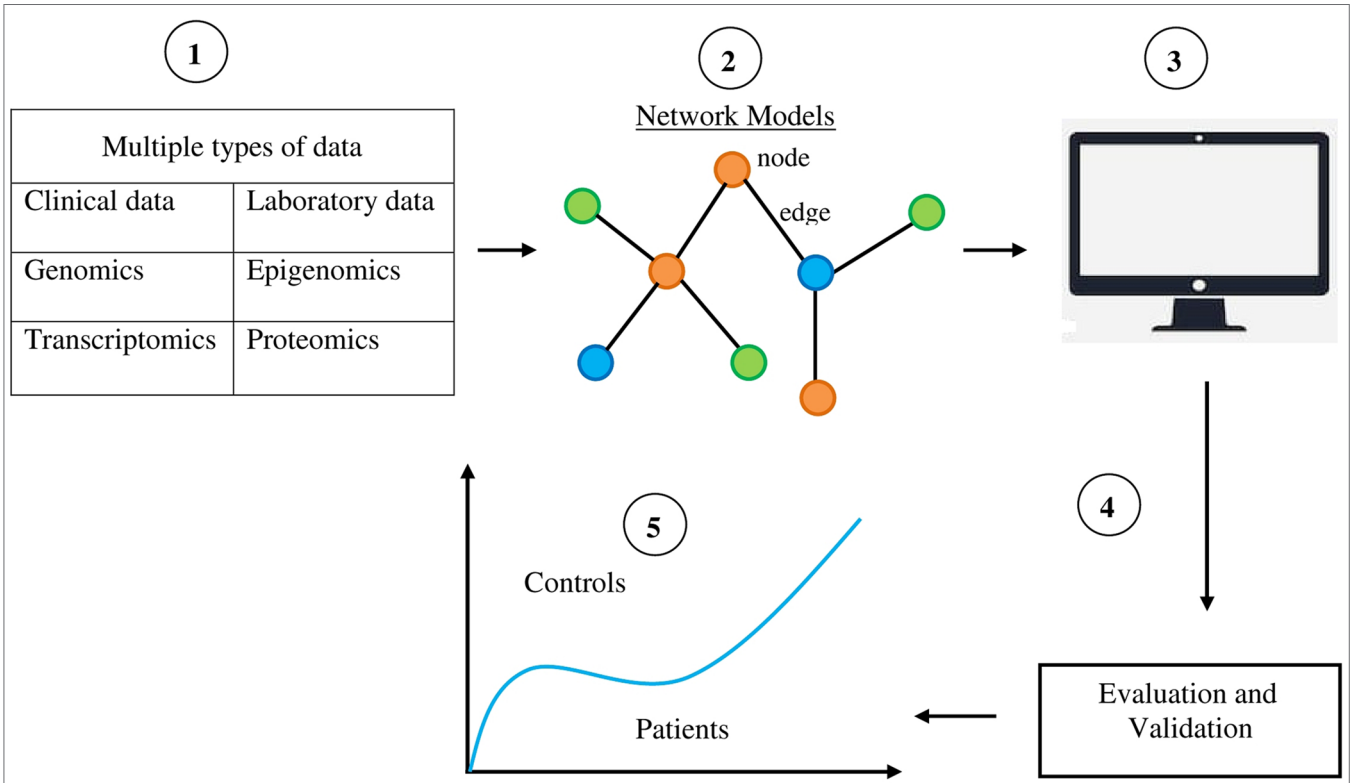


FIGURE 1 | Diagrammatic and simplified representation of the use of systems biology to identify biomarkers for neuropsychiatric disorders. 1) Collection of multiple types of data from patients, thereby providing the resources for the discovery of biomarkers. 2) The data gathered about molecules like DNA, RNA, proteins, and metabolites are organized by a network model. Components of networks like nodes and edges provide the materials for identifying disease biomarkers. 3) Computerized selection is done of specific features and abilities in network components for classifying various phenotypic states. 4) The data obtained are evaluated and validated in order to distinguish signal from noise. 5) Systems biology-based biomarkers are used to distinguish phenotypic states like normal and disease states.

RNA, proteins, and metabolites, and organizing the available data simultaneously (Liu, 2016). Molecules interact as a network in performing their functions. The nodes represent these molecules and the edges represent their physical and functional relationships. The network provides a topological representation of a complex system and the data characterize its specific condition by quantitatively measured values of a large number of molecules. Systems biology uses sophisticated computer software "omics"-based discovery tools and advanced computational techniques to understand the behavior of biological systems and identify diagnostic and prognostic biomarkers for complex disorders (Alawieh et al., 2012). A systems biology biomarker differs from traditional individual biomarkers in that a systems biology biomarker is a sub-network comprising two or more differentially expressed components in control samples *versus* disease samples (Wang et al., 2015).

Lausted et al. (2014) suggests that a systems biology approach for the discovery of biomarkers needs to use the following principles: 1) Blood is the ideal tissue/fluid for assessing biomarkers since it bathes all organs and contains secreted or released proteins from all these organs (however, it must be noted that for neuropsychiatric disorders there is a caveat regarding this principle in that the blood–brain barrier does not permit many molecules from crossing). 2) The diagnostic analyses should be conducted in a longitudinal manner so that changes in disease states can be followed. 3) The analyses should be quantitative. 4) Each patient should be his or her own control. 5) Multiple biomarkers should be measured since testing the status of multiple networks within the organ of interest is advantageous and probably needed. 6) Biomarkers may be of different informational types, like mRNAs, miRNAs, proteins, metabolites, and lipids.

In order to overcome the current limitations of systems biology and boost the efficiency of the systems biology approach for identifying biomarkers in neuropsychiatric disorders, researchers are coming up with innovative ideas and solutions like using neuroimaging techniques to study structural brain changes in patients (Frank et al., 2018), using induced pluripotent stem cell technology to model brain disorders (Schadt et al., 2014), and using endophenotyes (measurable components unseen by the unaided eye along the pathway between disease and distal genotype) of diseases (Gottesman and Hanson, 2005).

There is currently a new initiative called "The Psychiatric Cell Map Initiative" which aims to identify the physical and genetic interaction networks of neuropsychiatric disorders, and then using these data to connect genomic data to neuroscience and finally the clinic (Willsey et al., 2018). The initiative will include geneticists, structural biologists, neurobiologists, systems biologists, and clinicians; use many experimental approaches; and create a collaborative team for

long-term investigation. Its goal is to determine novel molecular and functional interaction data and pathway-level insights with regard to risk genes. The results of this initiative could have several applications, including identification of clinically useful biomarkers (Willsey et al., 2018).

Concluding Remarks

Neuropsychiatric disorders appear to be entirely biological: based on the activities of genetic and epigenetic mechanisms of expression of genes in neurons and other types of cells in different parts of the brain. As James Watson (1963) remarked in his 1962 Nobel banquet speech, the day he and Francis Crick discovered the structure of DNA, "they knew a new world had been opened and that an old world that seemed rather mystical was gone." There is unlikely to be a need to invoke mysticism or vitalism to explain partly or entirely our thoughts and feelings, normal or abnormal. However, due to the inordinate complexity of the brain, it remains to be seen whether neuropsychiatric disorders like SZ, BD, and MDD can be reduced to proteins, amines, or nucleic acids. For several decades, researchers have tried to find proteins and amines as biomarkers for these disorders, with no avail (Kruse et al., 2017). If these disorders could not be reduced to these molecules despite voluminous research, they may not also be reducible to nucleic acids like DNA. Regarding the human brain and mind, "the whole may be greater than the sum of its parts," a phrase attributed to Aristotle in its original form. Peter Medawar (1984) in *The Limits of Science* states that science can solve questions that come under the realm of science, but may not be able to solve questions that come under the realms of religion and philosophy. I feel that the development of neuropsychiatric disorders like SZ, BD, and MDD comes under the realm of science, and not religion and philosophy, and should be solvable by the methods of science. The methods and techniques of systems biology, incorporating epigenetic and other data, may help identify clinically useful biomarkers for neuropsychiatric disorders.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

The author acknowledges Dr Abraham Verghese for his comments on the manuscript.

REFERENCES

Adams, Z., McClure, E. A., Gray, K. M., Danielson, C. K., Treiber, F. A., and Ruggiero, K. J. (2017). Mobile devices for the remote acquisition of physiological and behavioral biomarkers in psychiatric clinical research. *J. Psychiatr. Res.* 85, 1–14. doi: 10.1016/j.jpsychires.2016.10.019

Aderem, A. (2005). Systems biology: its practice and challenges. *Cell* 121, 511–513. doi: 10.1016/j.cell.2005.04.020

Alawieh, A., Zaraket, F. A., Li, J.-L., Mondello, S., Nokkari, A., Razafsha, M., et al. (2012). Systems biology, bioinformatics, and biomarkers in neuropsychiatry. *Front. Neurosci.* 6, 187. doi: 10.3389/fnins.2012.00187

Altman, N., and Krzywinski, M. (2015). Association, causation, and correlation. *Nat. Methods* 12, 899–900. doi: 10.1038/nmeth.3587

Biomarkers Definitions Working Group (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin. Pharmacol. Ther* 69, 89–95. doi: 10.1067/mcp.2001.113989

Barabasi, A.-L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113. doi: 10.1038/nrg1272

Brugger, S. P., and Howes, O. D. (2017). Heterogeneity and homogeneity of regional brain structure in schizophrenia: a meta-analysis. *JAMA Psychiatry* 74, 1104–1111. doi: 10.1001/jamapsychiatry.2017.2663

Califano, A., Butte, A. J., Friend, S., Ideker, T., and Schadt, E. (2012). Leveraging models of cell regulation and GWAS data in integrative network-based association studies. *Nat. Genet* 44, 841–847. doi: 10.1038/ng.2355

Crick, F. (1966). *Of molecules and men*. Seattle: University of Washington Press.

Davis, J., Maes, M., Andreazza, A., McGrath, J. J., Tye, S. J., and Berk, M. (2015). Towards a classification of biomarkers of neuropsychiatric disease: from encompass to compass. *Mol. Psychiatry* 2015 (20), 152–153. doi: 10.1038/mp.2014.139

Drayna, D. (2006). Is our behavior written in our genes? *N. Engl. J. Med.* 354, 7–9. doi: 10.1056/NEJMp058215

Ebstein, R. P. (2018). Has the gloom lifted on genome-wide association studies? *Biol. Psychiatry* 83, 544–545. doi: 10.1016/j.biopsych.2018.02.002

Eyal, O., Berkenstadt, M., Reznik-Wolf, H., Poran, H., Ziv-Baran, T., Greenbaum, L., et al. (2019). Prenatal diagnosis for *de novo* mutations: experience from a tertiary care center over a 10-year period. *Mol. Genet. Genom. Med.* 7, e573. doi: 10.1002/mgg3.573

Feinberg, A. P. (2018). The key role of epigenetics in human disease prevention and mitigation. *New Engl. J. Med.* 378, 1323–1324. doi: 10.1056/NEJMra1402513

Frank, E., Maier, D., Pajula, J., Suvitaival, T., Borgan, F., Butz-Ostendorf, M., et al. (2018). Platform for systems medicine research and diagnostic applications in psychotic disorders—the METSY project. *Eur. Psychiatry* 50, 40–46. doi: 10.1016/j.eurpsy.2017.12.001

Gandal, M. J., Haney, J. R., Parikshak, N. N., Leppa, V., Ramaswami, G., Hartl, C., et al. (2018). Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* 359, 693–697. doi: 10.1126/science.aad6469

Gold, I. (2009). Reduction in psychiatry. *Can. J. Psychiatry* 54, 506–512. doi: 10.1177/070674370905400802

Gottesman, I. I., and Hanson, D. R. (2005). Human development: biological and genetic processes. *Ann. Rev. Psychol.* 56, 263–286. doi: 10.1146/annurev.psych.56.091103.070208

Guidotti, A., Auta, J., Davis, J. M., Dong, E., Gavin, D. P., Grayson, D. R., et al. (2014). Toward the identification of peripheral epigenetic biomarkers of schizophrenia. *J. Neurogenet.* 28, 41–52. doi: 10.3109/01677063.2014.892485

Hiesinger, P. R., and Hassan, B. A. (2005). Genetics in the age of systems biology. *Cell* 123, 1173–1174. doi: 10.1016/j.cell.2005.12.006

Housely, J., Hill, C. S., and Rugg-Gunn, P. J. (2015). Annual meeting of the EpiGeneSys Network of Excellence—advancing epigenetics towards systems biology. *Bioessays* 37, 592–595. doi: 10.1002/bies.201500015

Ideker, T., Dutkowski, J., and Hood, L. (2011). Boosting signal-to-noise in complex biology: prior knowledge is power. *Cell* 144, 860–863. doi: 10.1016/j.cell.2011.03.007

Ideker, T., and Hood, L. (2019). A blueprint for systems biology. *Clin. Chem.* 65, 342–344. doi: 10.1373/clinchem.2018.291062

Ikegame, T., Bundo, M., Murata, Y., Kasai, K., Kato, T., and Iwamoto, K. (2013). DNA methylation of the BDNF gene and its relevance to psychiatric disorders. *J. Hum. Genet.* 58, 434–438. doi: 10.1038/jhg.2013.65

Kang, J. I., Kim, T. Y., Choi, J. H., So, H. S., and Kim, S. J. (2019). Allele-specific DNA methylation level of FKBP5 is associated with post-traumatic stress disorder. *Psychoneuroendocrinology* 103, 1–7. doi: 10.1016/j.psyneuen.2018.12.226

Köhler, K., and Seitz, H. (2012). Validation processes of protein biomarkers in serum—a cross platform comparison. *Sensors (Basel)* 12, 12710–12728. doi: 10.3390/s120912710

Kruse, J. L., James, M. L., and Guze, B. H., (2017). "Medical assessment and laboratory testing in psychiatry," in *Kaplan & Sadock's comprehensive textbook of psychiatry*. Eds. Sadock, B. J., Sadock, V. J., and Ruiz, P. (Philadelphia, PA: McGraw-Hill), 1008–1026.

Lausted, C., Lee, I., Zhou, Y., Qin, S., Sung, J., Price, N. D., et al. (2014). Systems approach to neurodegenerative disease biomarker discovery. *Ann. Rev. Pharmacol. Toxicol.* 54, 457–481. doi: 10.1146/annurev-pharmtox-011613-135928

Lewontin, R. C. (1991). *Biology as ideology: the doctrine of DNA*. New York: HarperCollins Publishers.

Lin, C.-C., and Huang, T.-L., (2016). "Epigenetic biomarkers in neuropsychiatric disorders," in *Neuropsychiatric disorders and epigenetics*. Eds. Yasui, D. H., Peedicayil, J., and Grayson, D. R. (Boston: MA: Elsevier), 35–66. doi: 10.1016/B978-0-12-800226-1.00003-4

Liu, Z.-P. (2016). Identifying network-based biomarkers of complex diseases from high-throughput data. *Biomark. Med.* 10, 633–650. doi: 10.2217/bmm-2015-0035

Liu, S., Zhang, F., Wang, X., Shugart, Y. Y., Zhao, Y., Li, X., et al. (2018). Diagnostic value of blood-derived microRNAs for schizophrenia: results of a meta-analysis and validation. *Sci. Rep* 7, 15328. doi: 10.1038/s41598-017-15751-5

Loscalzo, J. (2018). "Network medicine: systems biology in health and disease," in *Harrison's principles of internal medicine*. Eds. Jameson, J. L., Fauci, A. S., Kasper, D. L., Hauser, S. L., Longo, D. L., and Loscalzo, J. (New York: McGraw-Hill), 3515–3522.

Medawar, P. (1984). *The limits of science*. Oxford: Oxford University Press.

Mor, E., Kano, S.-I., Colantuoni, C., Sawa, A., Navon, R., and Shomron, N. (2013). MicroRNA-382 expression is elevated in the olfactory neuroepithelium of schizophrenia patients. *Neurobiol. Dis.* 55, 1–10. doi: 10.1016/j.nbd.2013.03011

Nance, M. A. (2017). Genetic counseling and testing for Huntington's disease: a historical review. *Am. J. Med. Genet. B* 174, 75–92. doi: 10.1002/ajmg.b.32453

Peedicayil, J., Grayson, D. R., and Yasui, D. H., (2016a). "Introduction to neuropsychiatric disorders and epigenetics," in *Neuropsychiatric disorders and epigenetics*. Eds. Yasui, D. H., Peedicayil, J., and Grayson, D. R. (Boston, MA: Elsevier), 3–8. doi: 10.1016/B978-0-12-800226-1.00001-0

Peedicayil, J., Grayson, D. R., and Yasui, D. H., (2016b). "Neuropsychiatric disorders and epigenetics: summary and outlook," in *Neuropsychiatric disorders and epigenetics*. Eds. YasuiPeedicayil, D. H.J., and Grayson, D. R. (Boston, MA: Elsevier), 401–407. doi: 10.1016/B978-0-12-800226-1.00019-8

Peedicayil, J., and Grayson, D. R. (2018a). An epigenetic basis for an omnigenic model of psychiatric disorders. *J. Theor. Biol.* 443, 52–55. doi: 10.1016/j.jtbi.2018.01.027

Peedicayil, J., and Grayson, D. R. (2018b). Some implications of an epigenetic-based omnigenic model of psychiatric disorders. *J. Theor. Biol.* 452, 81–84. doi: 10.1016/j.jtbi.2018.05.014

Rose, S. (1995). The rise of neurogenetic determinism. *Nature* 373, 380–382. doi: 10.1038/373380a0

Schadt, E. E., Buchanan, S., Brennand, K. J., and Merchant, K. M. (2014). Evolving toward a human-cell based and multiscale approach to drug discovery for CNS disorders. *Front. Pharmacol.* 5, 252. doi: 10.3389/fphar.2014.00252

Strohman, R. C. (1997). Epigenesis and complexity: the coming Kuhnian revolution in biology. *Nat. Biotechnol.* 15, 194–200. doi: 10.1038/nbt0397-194

Tseng, P.-T., Lin, P.-Y., Lee, Y., Hung, C.-F., Lung, F.-W., Chen, C.-S., et al. (2014). Age-associated decrease in global DNA methylation in patients with major depression. *Neuropsychiatr. Dis. Treat.* 10, 2105–2114. doi: 10.2147/NDTS71997

Villoslada, P., Steinman, L., and Baranzini, S. E. (2009). Systems biology and its application to the understanding of neurological diseases. *Ann. Neurol.* 65, 124–139. doi: 10.1002/ana.21634

Wang, K., Lee, I., Carlson, G., Hood, L., and Galas, D. (2010). Systems biology and the discovery of diagnostic biomarkers. *Dis. Markers* 28, 199–207. doi: 10.1155/2010/130861

Wang, R.-S., Maron, B. A., and Loscalzo, J. (2015). Systems medicine: evolution of systems biology from bench to bedside. *WIREs Syst. Biol. Med.* 7, 141–161. doi: 10.1002/wsbm.1297

Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F. C. P., et al. (2018). Comprehensive functional genomic resource and integrative model for the human brain. *Science* 362, eaat8464. doi: 10.1126/science.aat8464

Watson, J. D. (1963). *Les prix Nobel*. Liljestrand, G., editor. Stockholm: Nobel Foundation.

Watson, J. D. (2003). *DNA: The secret of life*. London: Arrow Books.

Wattendorf, D. J., and Muenke, M. (2005). Diagnosis and management of fragile X syndrome. *Am. Fam. Physician* 72, 111–113.

Willsey, A. J., Morris, M. T., Wang, S., Willsey, H. R., Sun, N., Teerikorpi, N., et al. (2018). The Psychiatric Cell Map Initiative: a convergent systems biological approach to illuminating key molecular pathways in neuropsychiatric disorders. *Cell* 174, 505–520. doi: 10.1016/j.cell.2018.06.016

Ziegler, C., Richter, J., Mahr, M., Gajewska, A., Schiele, M. A., Gehrmann, A., et al. (2016). MAOA gene hypomethylation in panic disorder—reversibility of an epigenetic risk pattern by psychotherapy. *Transl. Psychiatry* 6, e773. doi: 10.1038/tp.2016.41

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Peedicayil. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

REVIEW
published: 16 October 2019
doi: 10.3389/fgene.2019.00986

Recent Advances in the Targeting of Epigenetic Regulators in B-Cell Non-Hodgkin Lymphoma

Marcelo L. Ribeiro[1,2], Diana Reyes-Garau[1], Marc Armengol[1], Miranda Fernández-Serrano[1] and Gaël Roué[1]**

[1] Laboratory of Experimental Hematology, Department of Hematology, Vall d'Hebron Institute of Oncology (VHIO), Vall d'Hebron University Hospital, Autonomous University of Barcelona, Barcelona, Spain, [2] Laboratory of Immunopharmacology and Molecular Biology, Sao Francisco University Medical School, Braganca Paulista, São Paulo, Brazil

In the last 10 years, major advances have been made in the diagnosis and development of selective therapies for several blood cancers, including B-cell non-Hodgkin lymphoma (B-NHL), a heterogeneous group of malignancies arising from the mature B lymphocyte compartment. However, most of these entities remain incurable and current treatments are associated with variable efficacy, several adverse events, and frequent relapses. Thus, new diagnostic paradigms and novel therapeutic options are required to improve the prognosis of patients with B-NHL. With the recent deciphering of the mutational landscapes of B-cell disorders by high-throughput sequencing, it came out that different epigenetic deregulations might drive and/or promote B lymphomagenesis. Consistently, over the last decade, numerous epigenetic drugs (or epidrugs) have emerged in the clinical management of B-NHL patients. In this review, we will present an overview of the most relevant epidrugs tested and/or used so far for the treatment of different subtypes of B-NHL, from first-generation epigenetic therapies like histone acetyl transferases (HDACs) or DNA-methyl transferases (DNMTs) inhibitors to new agents showing selectivity for proteins that are mutated, translocated, and/or overexpressed in these diseases, including EZH2, BET, and PRMT. We will dissect the mechanisms of action of these epigenetic inhibitors, as well as the molecular processes underlying their lack of efficacy in refractory patients. This review will also provide a summary of the latest strategies being employed in preclinical and clinical settings, and will point out the most promising lines of investigation in the field.

Keywords: B-cell lymphoma, DNMT, EZH2, HDAC, PRMT inhibitor, BET bromodomain inhibitor (BETi), combination therapy

OPEN ACCESS

Edited by:
Jiucun Wang,
Fudan University, China

Reviewed by:
Naoko Hattori,
National Cancer Center Research Institute (Japan), Japan
Maurizio D'Esposito,
Italian National Research Council (CNR), Italy

****Correspondence:***
Marcelo L. Ribeiro
mlribeiro@vhio.net
Gaël Roué
groue@vhebron.net

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *11 June 2019*
Accepted: *17 September 2019*
Published: *16 October 2019*

Citation:
Ribeiro ML, Reyes-Garau D, Armengol M, Fernández-Serrano M and Roué G (2019) Recent Advances in the Targeting of Epigenetic Regulators in B-Cell Non-Hodgkin Lymphoma. Front. Genet. 10:986. doi: 10.3389/fgene.2019.00986

INTRODUCTION

Characteristics of B-Cell Non-Hodgkin Lymphoma (B-NHL)

At the origin of 4% of all cancers and more than 90% of the cases of lymphoma, B-NHLs comprise a heterogeneous group of lymphoid neoplasms. According to the last World Health Organization hematopoietic and lymphoid tumor classification, more than 40 distinct entities are categorized, according to a combination of morphological, immunophenotypic, genetic, and clinical features, having each entity its own clinical course and requiring specific treatments (**Table 1**)

TABLE 1 | Classification of B-cell non-Hodgkin lymphoma.

Name	Cell of origin	Genetic aberration	Involved genes	Frequency (%)	References
B-ALL	Hematopoietic stem cell or B-cell progenitor	Hyperdiploidy	–	31*	(Yeoh et al., 2002)
		t(12;21)(p13.2;q22.1)	*ETV6/RUNX1*	28*	
		t(1;19)(q23;p13.3)	*TCF3/PBX1*	9*	
		t(9;22)(q34.1;q11.2)	*BCR/ABL1*	5*	
CLL/SLL	Naïve (unmutated IGHV subset) or memory (mutated IGHV subset) B cell	del(13q14.3)	–	54	(Haferlach et al., 2007)
		–	*SF3B1*	21	(Landau et al., 2015)
		Trisomy of 12	–	14	
		del(11q)	*ATM*	12	
LPL	Post-follicular B cell	–	*MYD88*	90	(Hunter et al., 2014)
		del(6q)	–	43	
		–	*CXCR4*	27	
			ARID1A	17	
NMZL	Post-germinal center marginal zone B cell	Gains of 3p	–	24	(Rinaldi et al., 2011)
		Gains of 18q	–	24	
		del(6q23)	*TNFAIP3*	16	
EMZL-MALT	Post-germinal center marginal zone B cell	Trisomy of 3	–	31	(Streubel et al., 2004)
		del(6q23)	*TNFAIP3*	30	(Rinaldi et al., 2011)
		t(11;18)(q21;q21)	*BIRC3/MALT1*	13	
		Trisomy of 18	–	11	
SMZL	Marginal zone B cell with or without antigen exposure	del(7q)	–	26	(Rinaldi et al., 2011)
		Gains of 3q	–	20	(Parry et al., 2015)
		del(17p13) or mutation	*TP53*	16	
		–	*KLF2*	12	
HCL	Late activated memory B cell	–	*BRAF*	100	(Tiacci et al., 2011)
		–	*MAP2K1*	48	(Waterfall et al., 2013)
FL	Germinal center B cell	t(14;18)(q32;q21)	*IGH/BCL2*	89	(Horsman et al., 1995)
		–	*KMT2D*	82	(Okosun et al., 2014)
		–	*CREBBP*	64	
		–	*TNFRSF14*	35	
MCL	Peripheral B cell of the inner mantle zone	t(11;14)(q13;q32)	*IGH/CCND1*	57	(Beà et al., 1999)
		Gain of 3q26	–	49	(Beà et al., 2013)
		del(11q) or mutation	*ATM*	41	
		del(17p13) or mutation	*CCND1*	34	
			TP53	27	
DLBCL-GCB	Peripheral mature B cell of germinal center origin	–	*KMT2D*	46	(Dubois et al., 2016)
		del(1p36) or mutation	*TNFRSF14*	38	(Scholtysik et al., 2015)(Roulland et al., 2018)
		Gains of 2p16	*REL*	30	
		t(18q21)	*BCL2*	28	
DLBCL-ABC	Peripheral mature B cell of germinal center exit or post-germinal center	del(9p21)	*CDKN2A*	47	(Scholtysik et al., 2015)(Dubois et al., 2016)
		del(6q21) or mutation	*PRDM1*	41	
		–	*KMT2D*	41	
		–	*MYD88*	28	
BL	Germinal center B cell	t(8;14)(q24;q32)	*MYC/IGH*	77	(Toujani et al., 2009)
		–Other t(8q24)	*ID3*	58	(Schmitz et al., 2012)
			TCF3	29	
			MYC	8–15	

**Frequency in pediatric cases. B-ALL, B-cell acute lymphocytic leukemia; CLL/SLL, chronic lymphocytic leukemia/small lymphocytic lymphoma; LPL, lymphoplasmacytic lymphoma; NMZL, nodal marginal zone lymphoma; EMZL-MALT, extranodal marginal zone lymphoma of mucosa-associated lymphoid tissue; SMZL, splenic marginal zone lymphoma; HCL, hairy cell leukemia; FL, follicular lymphoma; MCL, mantle cell lymphoma; DLBCL-GCB, diffuse large B-cell lymphoma of germinal center B-cell subtype; DLBCL-ABC, diffuse large B-cell lymphoma of activated B-cell subtype; BL, Burkitt lymphoma.*

(Campo et al., 2011; Scott and Gascoyne, 2014; Swerdlow et al., 2016). Originated from either mature or immature B cells, B-NHLs are characterized by the proliferation of lymphocytes, mainly in lymphoid organs and in extranodal tissues. Their diversity can often be traced to a particular stage of differentiation, from the bone marrow where the normal precursor B cell is originated to secondary lymphoid tissues where B cells undergo multiple rounds of selection before their differentiation into plasma cells or memory B cells. During these processes, the VDJ heavy chain is formed, followed by VJ light-chain gene rearrangement, which allows the pre-B cells to express intracytoplasmic μ-heavy chains. Subsequently, immature immunoglobulin (Ig)-positive B cells are formed. Within the lymph node, and in contact with a determined antigen, naïve B cells can mature into IgM-secreting plasma cells or may proliferate into primary follicles to form germinal center (GC) centroblasts. Upon maturation, they further differentiate into centrocytes, which give place to memory B

cells or plasma cells. Within the GC, somatic hypermutation in the *Ig heavy or light chain variable region (IGHV* or *IGHL)* genes leads to increased antigen affinity.

Although tightly regulated, the B-cell differentiation process and especially the antibody diversification phase can be accompanied by inherited events that may favor lymphomagenesis, such as chromosomal translocations, oncogene activation, and/or inactivating mutations in tumor suppressor genes. Infection by determined viruses, such as the Epstein–Barr virus, has also been involved in lymphomagenesis. The malignant counterparts of the early B-cell differentiation steps account for B lymphoblastic lymphomas, which harbor high similarity with B progenitor cells. On the other hand, mantle cell lymphomas (MCLs) and a subset of chronic lymphocytic leukemia (CLL) with unmutated *IGHV* are thought to derive from naive B cells and pre-GC mature B cells expressing the CD5 surface marker. Other GC-originated lymphomas, including follicular lymphoma (FL), Burkitt's lymphoma (BL), a subset of diffuse large B-cell lymphoma (DLBCL), and Hodgkin's lymphoma (HL), present mutations in *IGHV* gene. Additional entities, including marginal zone lymphoma (MZL), lymphoplasmacytic lymphoma, CLL with somatic *IGHV* mutation, another subset of DLBCL, and multiple myeloma (MM) correspond to post-GC cells. Each lymphoma subtype retains key features of their cell of origin as judged by the similarity of immunophenotype, histological appearance, and gene expression profiles (Seifert et al., 2013) (**Table 1**). The putative normal B-cell counterpart of each B-cell lymphoma is summarized in **Figure 1**.

In the last decade, loads of evidences have suggested an association between the frequent alterations in chromatin state and epigenetic regulators observed in B-NHL patients, and disease formation and progression.

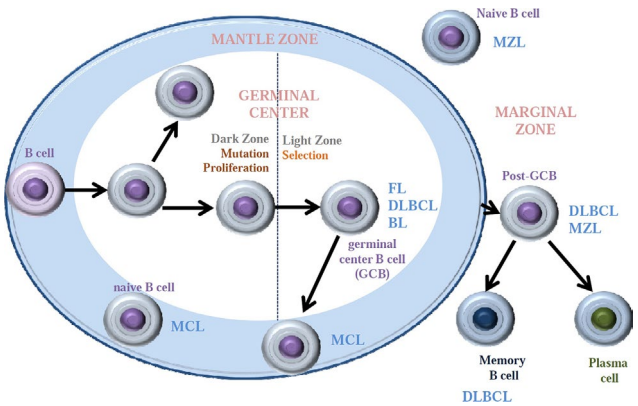


FIGURE 1 | Major B-cell non-Hodgkin lymphoma subtypes arise from different cell of origin within the lymph node. Mantle cell lymphomas (MCL) arise from naive B cells or germinal center (GC) B cells found within the mantle zone. Marginal zone lymphomas initiate from naive B cells or GCB that have entered the marginal zone. GCB are the origin of follicular lymphomas (FL), Burkitt lymphoma (BL), and diffuse large B-cell lymphomas (DLBCL) when still in the germinal center. This last DLBCL appears to also form GCB within the marginal zone or from fully developed memory B cells.

Altered Chromatin-Modifying Enzymes in B-NHL

Contrary to the general belief that only accumulations of DNA mutations might lead directly to the development of tumorigenic processes, it has been progressively reported a growing subset of epigenetic alterations lying at the basis of many malignancies, including those occurring in lymph nodes. Interestingly, in B-cell lymphomas, certain somatic mutations in chromatin-modifying enzymes account for several epigenetic alterations, suggesting that an aberrant epigenetic landscape in B-NHL may be a consequence of genetic alterations associated with a particular lymphoma subtype. For instance, deleterious and/or loss of function mutations in the histone acetyltransferase *CREB binding protein (CREBBP)* or the *E1A binding protein 300 (EP330)* have been reported in about 40% of DLBCL and FL patients as well as in other lymphoma subtypes (Morin et al., 2011; Pasqualucci et al., 2011b). Recurrent point mutations in the histone acetyl transferase (HAT) recruiting gene *myocyte enhancer binding factor 2B (MEF2B)* have been also described in 15% of FL and 13% of DLBCL patients with germinal center B cell (DLBCL-GCB) subtype (Morin et al., 2011). Although no mutations have been reported in the genes coding for histone deacetylases (HDACs), several members of this family like *HDAC1*, *2*, and *6* can be overexpressed in DLBCL, in association with a decrease in the DNA accessibility to the transcription machinery (Marquard et al., 2009).

In addition to mutations in chromatin-regulatory proteins, epigenetic modifications at chromatin level are also commonly observed in B-NHL as a result of profound changes in DNA methylation patterns. Indeed, while hypo- and hyper-DNA methylation status have been linked to the pathogenesis of several cancer subtypes, somatic mutations in epigenetic genes codifying for DNA methylation regulators have been particularly well associated to a repressed chromatin state and to malignant processes in B-NHL (Esteller et al., 2001; Hassler et al., 2013). Among the main reported alterations, activating mutations in *enhancer of zeste homolog 2 (EZH2)*, a histone methyltransferase (HMT) gene, were found in 22% of DLBCL-GCB patients and 7% of FL patients (Morin et al., 2010). Further loss-of-function mutations were observed in the *histone-Lysine N-Methyltransferase 2D (MLL2/KMT2D)* gene in about 90% of FL and 30% of DLBCL patients (Morin et al., 2011; Pasqualucci et al., 2011b; Lohr et al., 2012). Concretely, *MLL2* presents a defective SET domain when mutated by either truncation or frameshift mutations, leading to a reduced H3K4 methylation activity (Shilatifard, 2008; Morin et al., 2011; Pasqualucci et al., 2011b; Lohr et al., 2012).

Hence, B-NHL occurrence as a result of disruption in epigenetic mechanisms has generated a strong rationale to target epigenetic and chromatin regulators for drug discovery attempts. To address these alterations, several Food and Drug Administration (FDA)–approved epigenetic-modulating agents, whose clinical use has been mainly restrained so far to other hematological malignancies (Popovic et al., 2013), are now being made available for their evaluation in B-NHL.

These agents include the HDAC inhibitors romidepsin (FK228, depsipeptide), vorinostat (suberanilohydroxamic acid, SAHA), panobinostat (LBH589), and belinostat (PXD101); the DNA methyltransferase (DNMT) inhibitors (hypomethylating agents, HMAs) azacitidine (5-azacytidine) and decitabine (5-aza-2′-deoxycytidine); and the isocitrate dehydrogenase (IDH) inhibitors enasidenib (AG-221) and ivosidenib (AG-120) (**Table 2**).

TARGETING WRITER EPIGENETIC ENZYMES

DNMT Inhibitors

DNA methylation is responsible for the control of gene expression and for maintaining genomic stability during embryogenesis and tissue differentiation (Meissner, 2010). This process is clonally inherited and preserved in daughter cells, and occurs through the inclusion of a methyl group at cytosine residues in CpG dinucleotides (**Figure 2**). It is carried out by the DNMTs, namely DNMT1 which primarily mediates maintenance methylation during cell division, and DNMT3A and 3B that regulate *de novo* DNA methylation (Belinsky et al., 2003; Hermann et al., 2004). DNA methylation is thought to have a significant role in the regulation of lymphoid compartment, as it has been demonstrated that differential recruitment of DNMT1, DNMT3A, and DNMT3B and consequent specific DNA methylation patterns are determined at early stages during lymphopoiesis and B-cell activation (Shaknovich et al., 2011; Lai et al., 2013).

While on the one hand, DNA methylation is essential for cell homeostasis, on the other hand, disturbance in methylation pattern have been widely described in cancer. Changes in CpG methylation are indeed commonly associated with malignant transformation and tumor progression (Berdasco and Esteller, 2010). In addition, accumulating evidences suggest that aberrant epigenetic regulation, including DNA methylation, exerts an important role in regulating each cancer's hallmarks (Flavahan et al., 2017). Illustrating this relationship in B-NHL, Shaknovich and collaborators demonstrated the relevance of DNA methylation in defining the molecular DLBCL subtypes (Shaknovich et al., 2010). It was further proposed that DNMT1 and DNMT 3B overexpression may play a role in malignant progression of these tumors (Amara et al., 2010) and also in BL neoplasm (Robaina et al., 2015). In line with this, the disruption of DNA methylation pattern is correlated with disease severity and patient survival in DLBCL and FL (De et al., 2013).

Considering that the majority of cancers, including B-NHL, harbor an altered DNA methylation pattern, and also taking into account the reversibility of this alteration, the idea to modulate the methylation machinery to restore a "normal" DNA methylation state has attracted great attention in cancer treatment (Azad et al., 2013). The first two DNA methylation epigenetic compounds (DNMTi) ratified by the FDA and the European Medicines Agency for cancer treatment, azacitidine and decitabine (Jones et al., 2016), were initially described as promising chemotherapeutic agents against myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML), although with moderate efficacy and high toxicity (Li et al., 1970; Vogler et al., 1976). In further trials, low-dose decitabine and azacitidine demonstrated to be effective in these patients, improving both the response and the overall survival (OS), leading to their further approval (**Table 2** and **Figure 3**) (Silverman et al., 2002; Fenaux et al., 2009; Lübbert et al., 2016). In B-NHL patients, two phase I studies using decitabine have been completed so far, but the response to therapy and the effect on DNA methylation were moderate (Stewart et al., 2009; Blum et al., 2010). Currently, azacitidine and decitabine are being evaluated alone or in combination in approximately 10 active clinical trials involving relapsed/refractory R/R B-NHL patients (**Table 3**). Considering the preliminary data of these trials, it seems premature to conclude that DNMTis can be used as monotherapy in B-NHL.

Although the mechanism of action of DNMTi is not well understood, the activity of decitabine and azacitidine is known to involve their incorporation into the DNA of proliferating cells, followed by irreversible inhibition of DNMT1 enzymatic activity and the addressing of this latest to proteasomal degradation

TABLE 2 | FDA-approved epigenetic drugs for hematological malignancies.

Agent	Target	Indication	Year of approval	Current development
Azacitidine	DNMT	MDS CMML AML	2004	
Decitabine	DNMT	MDS AML	2006	Atherosclerosis
Vorinostat	HDAC	CTCL	2006	
Romidepsin	HDAC	CTCL PTCL	2009 2011	HIVAutism
Belinostat	HDAC	PTCL	2014	Ovarian cancer CTCL
Panobinostat	HDAC	MM	2015	CML MDS Breast cancer Prostate cancer

DNMT, DNA methyltransferase; HDAC, histone deacetylase; MDS, myelodysplastic syndrome; CMML, chronic myelomonocytic leukemia; AML, acute myeloid leukemia; CTCL, cutaneous T-cell lymphoma; PTCL, peripheral T-cell lymphoma; MM, multiple myeloma.

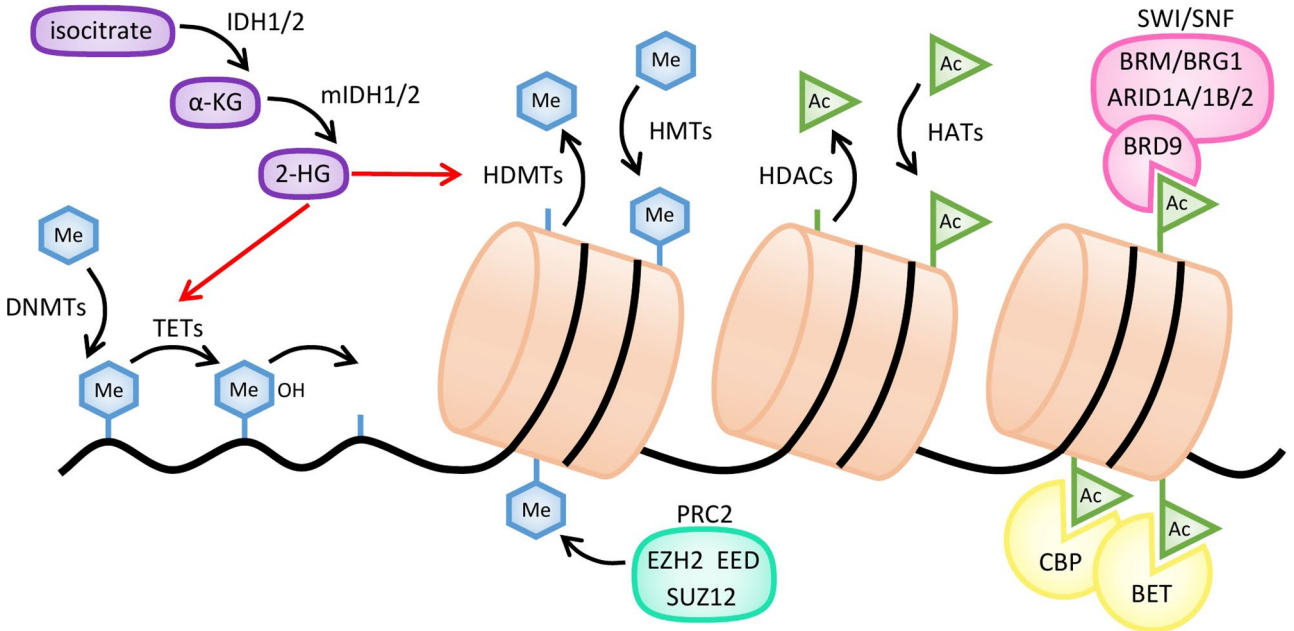


FIGURE 2 | Mechanisms of action of common epigenetic enzymes. Histone methylation is regulated by histone methyltransferases (HMTs), such as the EZH2 subunit of polycomb repressive complex 2 (PRC2), and histone demethylases (HDMTs). DNA methylation is established by DNA methyltransferases (DNMTs) and reversed by several enzymes like TET hydroxymethylases. Demethylation of both histones and DNA is inhibited by 2-hydroxyglutarate (2-HG), produced from α-ketoglutarate (α-KG) by mutant forms of IDH1/2 enzymes (mIDH). Histone acetylation is regulated by histone acetyltransferases (HATs) and histone deacetylases (HDACs). Bromodomain-containing proteins, such as CREB-binding protein (CBP), BET, or the BRD9 subunit of the SWI/SNF complex, bind to acetylated residues of histones.

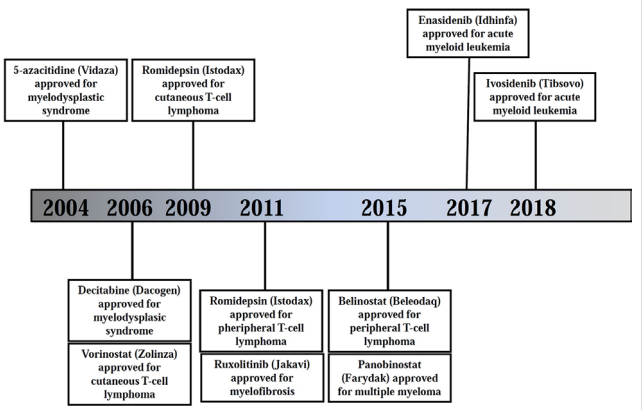


FIGURE 3 | Timeline of FDA approvals of epigenetic-modulating therapies in hematological cancers. Source: https://www.accessdata.fda.gov/scripts/cder/daf/

(Ghoshal et al., 2005; Juttermann et al., 2006). Accordingly, two main molecular effects have been described for DNMTi inhibitors: (1) a global demethylation of gene promoters (mainly tumor suppressor genes) and (2) the activation of immune system and the triggering of an anti-tumor immune response (Groudine et al., 1981; Almstedt et al., 2010; Goodyear et al., 2010; Chiappinelli et al., 2015;). As an illustration, in DLBCL it has been described that decitabine can reverse DNA methylation and restore expression of important cancer-related pathways *in vitro* and *in vivo* (Li et al., 2002; Clozel et al., 2013), although in other studies a less drastic and transient effect was observed (Karpf, 2004; McGarvey et al., 2006; Egger et al., 2007). Furthermore, DNMT inhibition is also linked to the demethylation of gene bodies, leading to oncogene downregulation (Wong et al., 2013; Yang et al., 2014).

Several new DNMTis have been developed in the last decade with potential activity in hematological malignancies. Among them, thioguanine (2-amino-1,7-dihydro-6H-purine-6-thione (6-tG)) has been approved by FDA to treat AML patients (Munshi et al., 2014). Its mechanism of action involves its incorporation into DNA, decrease in DNMT activity and DNA methylation, blockade of DNA and RNA synthesis, and ultimately cell death (Hogarth et al., 2008; Yuan et al., 2011; Flesner et al., 2014). Recently described as an experimental DNMTi, 5-fluoro-2′-deoxycytidine (FdCyd) is currently undergoing a phase I/II clinical trial in combination with other drugs (Kinders et al., 2011; Newman et al., 2015). Its mechanism of action involves the ability to block DNMT-dependent DNA methylation (Jones and Taylor, 1980; Beumer et al., 2008). 5,6-Dihydro-5-azacytidine is a reduced, hydrolytically stable form of 5-azacytidine nucleoside (Beisler et al., 1979). The mechanism of action is very similar to that described for azacytidine, with the advantage of a lower toxicity (Avramis et al., 1989). However, its evaluation in clinical settings revealed a reduced response rate and the rise of significant adverse effects (Samuels et al., 1998). Zebularine is another DNMTi, which has been previously described as tumor-selective

TABLE 3 | Selected examples of epigenetic drugs under clinical evaluation in B-NHL patients as single agents.

Epigenetic drug class	Drug	Diseases	Sponsor	Trial identifier
HDAC inhibitor	Vorinostat	FL, indolent B-NHL, MCL	Merck Sharp & Dohme Corp.	NCT00875056
	Chidamide (Epidaza)	R/R B-NHL	Sun Yat-sen University	NCT03245905
	Abexinostat	R/R FL	Xynomic Pharmaceuticals, Inc.	NCT03934567
	Mocetinostat (MGCD0103)	DLBCL, FL	Memorial Sloan Kettering Cancer Center and MethylGene Inc.	NCT02282358
	Ricolinostat(ACY-1215)	R/R NHL	Acetylon Pharmaceuticals Inc. and Columbia University	NCT02091063
	Panobinostat	CLL	Peter MacCallum Cancer Centre, Australia	NCT01658241
	CDX101	Advanced lymphomas	Cancer Research UK and Oxford University Hospitals	NCT01977638
IDH1 inhibitor	Ivosidenib(AG-120)	Advanced hematologic malignancies with an IDH1 mutation	Agios Pharmaceuticals, Inc.	NCT02074839
BET bromodomaininhibitor	CPI-0610	Progressive lymphomas	Constellation Pharmaceuticals	NCT01949883
	BMS-986158	Lymphomas	Dana-Farber Cancer Institute	NCT03936465
	Molibresib (GSK525762)	NHL	GlaxoSmithKline	NCT01943851
EZH1/2 inhibitor	DS-3201b	R/R B-NHL	Daiichi Sankyo Co., Ltd.	NCT02732275
	CPI-1205	Progressive B-cell lymphomas	Constellation Pharmaceuticals	NCT02395601
	Tazemetostat	R/R B-NHL	Eisai Co., Ltd.	NCT03009344
	Tazemetostat	R/R NHL with EZH2 gene mutations	National Cancer Institute (NCI)	NCT03213665
PRMT inhibitor	JNJ-64619178	R/R B-NHL	Janssen Research & Development, LLC	NCT03573310
	GSK3326595	MDS, AML	GlaxoSmithKline	NCT03614728
	GSK3368715	DLBCL	GlaxoSmithKline	NCT03666988
DNMT inhibitor	Decitabine	R/R DLBCL	Mingzhi Zhang, Zhengzhou University	NCT03579082
EED inhibitor	MAK683	Advanced DLBCL	Novartis Pharmaceuticals	NCT02900651

EED, embryonic ectoderm development. Source: https://clinicaltrials.gov/.

inhibitor of DNMTs (Cheng et al., 2004). Although there are a lot of evidences, both *in vitro* and *in vivo*, indicating the potential of zebularine as a demethylating agent in a wide range of tumors (Agrawal et al., 2018), its poor bioavailability has prevented its introduction into clinical trials (Ben-Kasus et al., 2005). More recently, 4′-thio-2′-deoxycytidine (TdCyd) and its 5-aza analog, 5-aza-TdCyd, have been reported to downregulate DNMT1 and to exhibit anti-tumor activity *in vitro* and in human leukemia and lung cancer xenografts (Thottassery et al., 2014). Among these last molecules, TdCyd has already entered into phase I clinical evaluation (NCT02423057 and NCT03366116). Further molecules were developed with superior anti-tumoral efficacy and included guadecitabine (SGI-110), a second-generation DNMTi that harbors an improved DNA methylation inhibition in solid tumors both *in vitro* and *in vivo* (Chuang et al., 2010; Srivastava et al., 2015). A phase I clinical trial has provided promising results in patients with MDS and AML (Issa et al., 2015). Fluorocyclopentenylcytosine (RX-3117) is a cytidine analog that presents an anti-tumor activity in a large set of tumor cells and *in vivo*. Its mechanism of action is associated with an inhibition of DNMT1 (Choi et al., 2012). This agent is being evaluated in a phase II study with R/R pancreatic or advanced bladder cancer (NCT02030067).

EZH2 Inhibitors

EZH2 constitutes the catalytic subunit of the polycomb repressive complex 2 (PRC2). Its structure is composed by a SET domain, typical in chromatin-associated regulators of gene expression (Xiao et al., 2003). It catalyzes histone H3 lysine 27 tri-methylation (H3K27me3) and the subsequent formation of heterochromatic regions and downregulation of the nearby genes (Bracken and Helin, 2009; Ferrari et al., 2014) (**Figure 2**). In B lymphocytes, EZH2 becomes expressed and inhibited in a cyclic manner. First, in pre-B lymphocytes, induction of EZH2 expression is required for an optimal V(D)J recombination. Later on, during the migration to lymphoid tissues, it is downregulated until the GC reaction occurs, after which it becomes re-expressed to allow the silencing of the anti-proliferative genes *cyclin-dependent kinase inhibitor 2A (CDKN2A)* and *cyclin-dependent kinase inhibitor 1A (CDKN1A1)* and the pro-differentiation genes *interferon regulatory factor 4 (IRF4)* and *PR domain zinc finger protein 1 (PRDM1/BLIMP1)* during the somatic hypermutation and isotype switch processes. Finally, EZH2 becomes repressed when mature B cells leave the GC (Velichutina et al., 2010; Béguelin et al., 2013). Gain-of-function mutations in *EZH2* have been reported in several solid tumors and hematological cancers. The consequence of those mutations in GC lymphocytes is the irreversible silencing of certain cell cycle checkpoint and plasma cell differentiation genes (Béguelin et al., 2013). The main gain-of-function mutation identified in DLBCL and FL patients includes a tyrosine deletion (Y641) at the EZH2 SET domain that increases the levels of H3K27me3, promoting a repressed state of cell differentiation and the repression of tumor suppressor genes (Morin et al., 2010; McCabe et al., 2012a). Similar effects have been described as a consequence of the A677G mutation in EZH2, which has been characterized in multiple human lymphoma cell lines. A change in the substrate preferences accounts for the aberrant H3K27me3 levels observed in cells bearing EZH2 mutant forms. Indeed, wt EZH2 displays preference for less methylated substrates whereas Y641 and A667G mutants prefer either substrates with higher methylation levels or show equal

affinity for all three substrates (H3K73me0, me1, and me2) (McCabe et al., 2012a). Interestingly, these gain-of-function EZH2 variants expressed in GC B-cell lymphomas seem to synergize with *BCL2* deregulation, favoring the progression of these malignancies (Béguelin et al., 2013). On the other hand, overexpression of wt EZH2 has been also reported in B-NHL (Van Kemenade et al., 2001; Visser and Gunster, 2001), with a positive correlation being observed between *EZH2* transcript levels, tumor aggressiveness, and disease prognosis (Abd Al Kader et al., 2013). Taking into account these considerations, it looks reasonable that inhibiting EZH2 activity could result in a potential therapeutic strategy to treat B-NHL.

In this context, many efforts directed to develop highly selective EZH2 inhibitors have been made in the last decade. EZH2 activity was initially targeted by means of the carbocyclic adenosine analog 3-deazaneplanocin A (DZNep), an inhibitor of *S*-adenosylhomocysteine hydrolase. DZNep promotes a global increase in the levels of 5-adensylhomocystein and a further inhibition in the activity of many methyltransferases, including EZH2. Nevertheless, due to its mechanism of action, it resulted to be too unspecific as many other methyltransferases were similarly affected. In 2012, a small chemical compound named El1 with a good capacity to inhibit the Y641 mutant and wt EZH2 form was evaluated for the treatment of DLBCL. This compound was designed as a competitive inhibitor of the EZH2 methyl group donor *S*-adenosyl--methionine (SAM). Unlike DZNep, El1 showed a 10,000-fold selectivity for EZH2 over other HMTs and a 90-fold selectivity over EZH1 methyltransferase. This compound promoted a global decrease in methyl donor availability, leading to a lower global levels of H3K27me3 (Qi et al., 2012). Other subsequent compounds directed specifically against EZH2 are the dual EZH2/1 inhibitors UNC1999, with a potent capacity to suppress H3K27me3 and H3K27me2 levels and to inhibit proliferation of mixed lineage leukemia (MLL)-rearranged cells, and the OR-S1 and OR-S2 inhibitors, which were assessed for the treatment of DLBCL, AML, and MM (Konze et al., 2013; Honma et al., 2017). Later on, EPZ0005687 and GSK126, two selective and SAM-competitive EZH2 inhibitors with a higher inhibitory capacity for the mutant EZH2 form, were developed and tested in DLBCL and FL (McCabe et al., 2012b; Knutson et al., 2014). In 2014, GSK126 entered into phase I clinical trials with B-NHL and MM patients (NCT02082977) (Zeng et al., 2016; Yap et al., 2018), but unfortunately that study had to be discontinued as a consequence of insufficient therapeutic activity, evidencing the need to keep working in the improvement of those inhibitors. Also in 2014, CPI-360 and its more potent and stable analog, CPI-169, were reported to be effective EZH2 inhibitors for the treatment of several B-NHL subtypes (Vaswani et al., 2016). An improved version of these latest, CPI-1205, showed a higher oral bioavailability and was first tested in preclinical studies with xenograft mouse models generated from human B-NHL cell lines and further challenged in phase I trials for the treatment of DLBCL (NCT02395601).

Valemetostat (DS-3201) is another potent wild-type (wt) and mutant EZH1/2 inhibitor that demonstrated a strong anti-proliferative effect against NHL, DLBCL, and T-cell lymphoma (Maruyama et al., 2017). Currently, tazemetostat (EPZ-6438), another SAM competitive inhibitor with a high affinity for the wt and the mutant EZH2 forms, is being evaluated in clinical studies to treat R/R B-NHL and MM patients (NCT03456726) (Knutson et al., 2014; Gulati et al., 2018), reaching an overall response rate of 38% in a phase I clinical trial (Italiano et al., 2018).

Despite first promising results, single-agent treatment with EZH2 inhibitors is in general slightly effective in aggressive lymphomas. Among the possible mechanism(s) of resistance, overactivation of the phosphatidylinositol 3-kinase (PI3K) and mitogen-activated protein kinase (MAPK) pathways has been identified in GSK126-resistant DLBCL cells (Bisserier and Wajapeyee, 2018). Thus, it looks reasonable to prioritize the discovery of new drug combination associating EZH2 inhibitors with other compounds targeting key signaling pathways in order to prevent and/or overcome the occurrence of EZH2i resistance in lymphoid neoplasm with mutated EZH2.

PRMT Inhibitors

A conserved biological mechanism within all eukaryotic organisms, from yeast to higher mammals, is arginine methylation (Migliori et al., 2010). This post-translational modification is mediated by *N*-arginine methyltransferases (PRMTs), which catalyze the transfer of a methyl group, from SAM to the omega nitrogens found in terminus guanidine group of an arginine residue of the side chain. This transfer may occur in one or both nitrogens (Bedford and Clarke, 2009). Among the nine different members of the PRMT family (Schubert et al., 2003), PRMT1 is the major enzyme responsible for arginine methylation followed by PRMT5, according to the observation that PRMT1 and PRMT5 knockout mice die at an early stage during development whereas mice lacking any of the other seven PRMTs are fully viable (Hadjikyriacou et al., 2015). Protein modifications performed by PRMTs are traditionally related to important genetic processes such as DNA repair and gene transcription, among others. More recently, PRMT functions have been linked to carcinogenesis and metastasis, giving to these enzymes the status of potent therapeutic targets in a variety of cancers where they are overexpressed, including colon, breast, prostate, and lung cancers, neuroblastomas, leukemias, and B-cell lymphoma (Yoshimatsu et al., 2011).

Within this family, upregulation of PMRT1 and PRMT5 has been widely associated with hematological malignancies (Greenblatt et al., 2016; Smith et al., 2018). In particular, the expression and function of PMRT5 have been extensively examined during lymphomagenesis, as this enzyme is highly expressed in primary samples and cell lines from different leukemia and lymphoma subtypes, where it promotes the repression of tumor suppressors such as the retinoblastoma proteins. In these models, experimental studies have suggested that PRMT5 upregulation may be caused by overexpression of *MYC* and *NOTCH* oncogenes (Wang et al., 2008). In transformed DLBCL, the *S-methyl-5'-thioadenosine phosphorylase (MTAP)* gene encoding for a critical methionine metabolism enzyme is deleted due to its proximity to the tumor suppressor gene *CDKN2A* (Dreyling et al., 1998), and this phenomenon sensitizes cancer cells to PRMT5 inactivation

(Marjon et al., 2016). A remarkable interplay has also been described between PRMT5 and the *B cell lymphoma 6 (BCL6)* oncogene during the lymphomagenesis in the GC (Lu et al., 2018), suggesting that pharmacological inhibition of arginine methylation could be of special interest in BCL6-driven lymphoma. Regarding PRMT1, an interesting interaction exists between this enzyme and EZH2 in DLBCL-GCB tumors. Indeed, recent works have reported an increase in PRMT1-related histone arginine methylation in DLBCL-GCB cells resistant to EZH2 inhibition, in association with BCL-2 overexpression and modulation of the B-cell receptor (BCR) downstream signaling, supporting the rational association of EZH2 and PRMT1 inhibitors in DLBCL patient samples (Goverdhan et al., 2017).

Among the multiple functional inhibitors that have been developed to target the different members of the family, PRMT1 and PRMT5 small molecule inhibitors have already shown great potential against B-NHL, either alone or upon their combination with other agents. As an illustration, promising results have been obtained with the specific PRMT5 inhibitor EPZ015666 (GSK3235025) when used as single agent in *in vitro* and *in vivo* models of MCL (Chan-Penebre et al., 2015).

TARGETING ERASER EPIGENETIC ENZYMES: HDAC INHIBITORS

By favoring an open chromatin state, histone acetylation allows numerous transcription factors to bind DNA and to activate gene expression. At the same time, acetylated histones increase DNA accessibility to transcriptional activators and counteract the function of transcriptional repressors (McClure et al., 2018). Acetylation of histones and non-histone proteins is regulated through a correct balance between HAT and HDAC activities. Among these enzymes, the most advanced subfamily is human HDACs, which have been classified into four classes according to their sequence homology, activity, and subcellular localization. HDACs 1, 2, 3, and 8 constitute class I. HDAC 4, 5, 6, 7, 9, and 10 belong to class II. Class III includes sirtuin 1 (SIRT1) and sirtuin 7 (SIRT7), two NAD-dependent structurally unrelated protein deacetylases (Minucci and Pelicci, 2006). Finally, class IV is represented by HDAC11. In contrast to class II HDACs which show a heterogeneous expression pattern, class I HDACs are found at particularly high levels in lymphoid cell lines and primary tumors, suggesting a predominant role of these latest in lymphomagenesis. Accordingly, the design of HDAC inhibitors (HDACis) in lymphoid malignancies has been mainly centered on this latest group of enzymes (Gloghini et al., 2009).

Several structurally distinct classes of HDACis have been developed. These molecules can be divided into five chemical groups: hydroxamic acids, cyclic peptides, electrophilic ketones, short-chain fatty acids, and benzamides. Pan-HDACis have the capacity to inhibit almost all HDACs with the exception of class III HDACs and include the hydroxamic acid derivatives vorinostat, givinostat (ITF2357), abexinostat, panobinostat, belinostat, and trichostatin A, the carboxylate sodium butyrate, and the cyclic peptide trapoxin (Bradner et al., 2010; Di Costanzo et al., 2014). Taking into account that HDACs can also modulate the function of several non-histone proteins regulating a number of physiological processes (Lane and Chabner, 2009), and that HDACs can simultaneously exert pro- and anti-leukemic activities (Heideman et al., 2013; Santoro et al., 2013), blocking individual HDACs with isotype-selective inhibitors specific for one or two classes of HDACs might represent a strategy of choice for the treatment of lymphoid tumors. In line with this, the isotype-selective HDACis include the benzamides entinostat (MS-275, SNDX-275) and mocetinostat (MGCD0103) (Fournel et al., 2008; Vannini et al., 2004), the hydroxamic acid derivative rocilinostat (ACY-1215) (Santo et al., 2012), and the cyclic peptide romidepsin, which show preference for HDAC1-6-8, HDAC6, and HDAC1-2, respectively (Lemoine and Younes, 2010). Several HDACis like vorinostat, mocetinostat, and entinostat can be administered orally; conversely, other agents like romidepsin are given intravenously (Batlevi et al., 2016; Mann et al., 2007; Younes et al., 2011; Holkova et al., 2017). By inhibiting the catalytic activity of their target HDAC(s), these compounds impair the formation of HDAC–substrate complexes, thus altering the transcriptomic pattern of the malignant cells as well as the activity of non-histone proteins, ultimately leading to growth arrest, differentiation, and induction of apoptosis (Qiu et al., 2000). Of importance, when compared to their malignant counterparts, healthy tissues are generally unaffected by HDACis (Mai et al., 2005).

A number of preclinical studies have highlighted a role for HDACi therapy in a range of B-cell lymphoma, including DLBCL, HL, and BL, either alone or in combination with other epidrugs such as HMAs, with small molecule agents or with standard chemotherapeutics (Buglio et al., 2008; Kretzner et al., 2011; Kewitz et al., 2012; Ageberg et al., 2013; Klein et al., 2013; Rozati et al., 2016; Garrido Castro et al., 2018). Among these studies, the weak HDACi valproic acid was shown to overcome DLBCL cell resistance to the standard R-CHOP (rituximab, cyclophosphamide, doxorubicin, vincristine, prednisone) chemotherapeutic regimen (Ageberg et al., 2013). In preclinical models of DLBCL and MCL, panobinostat, belinostat, depsipeptide, and vorinostat were shown to evoke tumor growth arrest, differentiation, and/or apoptosis *in vitro* and/or *in vivo*, mediated by the accumulation of DNA damage upon PARP trapping (Valdez et al., 2018), G1 cell cycle arrest consequent to an increase in the expression of the cyclin-dependent kinase inhibitor p21, acetylation of histone H3 (Xue et al., 2016), or transcriptional activation of the BCL-2 family proapoptotic members BIM, BMF, and NOXA (Kalac et al., 2011; Xargay-Torrent et al., 2011).

Based on these preclinical studies, several HDACis have entered clinical trials under different modalities (monotherapies or in combination). Many of these trials have been conducted in DLBCLs, FLs, and HLs using HDACis, either alone or in combinatorial therapies (Watanabe et al., 2010; Stathis et al., 2011; Younes et al., 2012; Oki et al., 2013; Ogura et al., 2014; Chen et al., 2015; Morschhauser et al., 2015) (**Table 2** and **Figure 3**). As monotherapy, HDACis have shown a wide range of response in lymphoma patients, varying from complete

remissions (CRs) to no response. In the absence of biomarkers for prediction of clinical outcome, the molecular mechanisms of resistance are poorly understood. Vorinostat was the first proved in relapsed B-NHL patients, including FL, MZL, and MCL. In a phase II study including relapsed FL, non-FL indolent NHL and MCL patients, oral vorinostat showed low levels as a single agent, with the exception of FL, in which an overall response rate (ORR) of 47–49% (referring to the proportion of patients with tumor size reduction of a predefined amount and for a minimum time period) and a CR rate of 23% was observed (Kirschbaum et al., 2011; Ogura et al., 2014). This agent was also tolerated, but displayed limited activity in another phase II trial against R/R DLBCL, with only 1/18 patients presenting complete response (Crump et al., 2008).

With the pan-HDACis abexinostat and quisinostat, or the class-specific mocetinostat and entinostat, the response rates were quite variable (from 12% to 56%), and mostly dependent on the drug and on the lymphoma subtype. The most robust responses were obtained with abexinostat in FL patients (56% ORR). This latest drug showed a unique pharmacokinetic profile and an optimized oral dosing schedule that allowed for a superior anti-tumoral activity. In a recent phase II study with patients with R/R B-NHL or CLL, among the evaluable patients the ORR was 28%, with highest responses observed in FL patients (ORR 56%) and DLBCL (ORR 31%) (Ribrag et al., 2017). A phase II clinical trial with mocetinostat in patients with R/R DLBCL and FL showed promising results (Batlevi et al., 2017), whereas for entinostat only one B-NHL patient has been included in phase II trial; therefore, no conclusion can be made on its efficacy in this subgroup of patients (Kummar et al., 2007).

Similar to DNMTis, the effectiveness of the first-generation HDACis carries significant toxicity and is limited to hematopoietic malignancies, which makes them challenging to combine (Suraweera et al., 2018). It is believed that part of this toxicity may be related to the capacity of HDACis to alter directly the function of many non-histone proteins. Toxicity may also be due to widespread activity across HDAC isoforms; therefore, the focus of second-generation HDACi discovery was to enhance the discrimination over HDAC family members (Galli et al., 2010; Knipstein and Gore, 2011; Younes et al., 2011; Santo et al., 2012; Evens et al., 2016). In this context, targeting HDAC6 was associated to the upregulation of CD20 and consequent enhanced efficacy of anti-CD20 monoclonal antibody therapy (Bobrowicz et al., 2017). Also, tucidinostat (CS055/chidamide), the first oral subtype-selective HDACi, was approved for the treatment of refractory/relapsed PTCL by the China Food and Drug Administration. This compound inhibits HDAC1, HDAC2, HDAC3, and HDAC10, and has entered a phase II clinical trial as single-agent treatment for patients with R/R B-NHL (NCT03245905) based on preliminary evidences of clinical activity in DLBCL (Yang et al., 2018).

Another approach to maximize efficacy with manageable toxicity consists in developing dual inhibitors. In this field, CUDC-907, a novel first-in-class oral small molecule inhibitor of both HDAC (class I and II) and PI3K (class Iα, β, and δ), has demonstrated excellent levels of activity (55% ORR) and tolerability in DLBCL patients in a phase IA clinical trial (Younes et al., 2016). In a second phase IB trial, the drug has been tested in patients with R/R DLBCL and showed a response rate of 37%, with a higher effect in MYC-altered *versus* MYC-unaltered patients (Oki et al., 2017). As a result of these encouraging initial data, this agent is currently being evaluated in a phase II study including DLBCL patients, and also in a phase I trial involving pediatric patients with lymphomas (NCT02674750 and NCT02909777).

TARGETING READER EPIGENETIC ENZYMES

BET Inhibitors

Among the post-translational modifiers with ability to orchestrate chromatin organization, bromodomain (BD)-containing proteins are readers of Ac-K residues at the N-terminal histone tails. They act as scaffolds that enable histone attachment to the chromatin and form active multi-protein transcription complexes, thereby modulating chromatin dynamics and ultimately diversifying gene expression (Filippakopoulos et al., 2012; Chaidos et al., 2015; Smith and Zhou, 2016). This family of proteins contains 46 members, comprising nuclear proteins with HAT or HMT activity, chromatin remodelers, helicases, transcription co-activators, and mediators or scaffold proteins. They are subdivided into eight subfamilies (I to VIII), based on their structure and sequence similarities. Subfamily II is the most studied one and includes the bromodomain-containing proteins mBRDT, BRD2, BRD3, and BRD4 (Padmanabhan et al., 2016). Besides the presence of two bromodomains (BD1 and BD2) that allow acetylated chromatin recognition, these proteins harbor an extra-terminal domain, which is responsible for protein–protein interactions. This bromodomain and extra-terminal (BET) subfamily has thus the capacity to act as protein adaptors facilitating the recruitment of chromatin remodelers and transcription factors for further initiation and elongation of transcription (Delmore et al., 2011; Chaidos et al., 2015; Padmanabhan et al., 2016). Several reports have highlighted the importance of the BET proteins action over DNA enhancers for the regulation of certain oncogenes expression (Lovén et al., 2013). Altogether, these studies make BET proteins attractive therapeutic targets in cancer.

As interfering with this family of proteins may serve as a strategy to address transcription irrespective of the presence of epigenetic mutations, BET proteins inhibitors have been a significant area of focus in the last decade, in cancer but also in inflammation, fibrosis, and heart diseases (Vakoc, 2015). Drug developmental studies have paid special attention to the Ac-K binding sites in the bromodomains, as these deep hydrophobic pockets with conserved asparagine and/or aspartate residues make BET proteins highly druggable (Cox et al., 2016). Indeed, the most common drug targeting approach in this family has been the development of small molecules that could block the lysine-binding pocket and disrupt the interactions between BDs and the Ac-K on chromatins (Smith and Zhou, 2016).

In 2005, a first bromodomain inhibitor developed by the Zhou laboratory, namely NP1, has the ability to target the BD of the P300/CBP-associated factor transcriptional coactivator (Zeng

et al., 2005). This step was followed by the discovery in 2006 of MS7972, a weakly binding fragment specific for CREBBP-BD, hindering its binding to acetylated p53 (Sachchidanand et al., 2006). Among BET proteins, the first target considered to be druggable was BRD4, as a pioneering RNAi base unveiled its critical role in the maintenance of AML. In this study, authors found that BRD4-dependent transcriptional activity could be efficiently targeted by the pan-BET thieno-triazolo-1,4-diazepine (+)-JQ1 (Filippakopoulos et al., 2010; Zuber et al., 2011). This class of diazepine-based small molecule inhibitors, which also includes the benzodiazepine I-BET151 (GSK1210151A) (Dawson et al., 2011) and I-BET762 (GSK525762) (Mirguet et al., 2013) (NCT01943851), utilizes the methyltriazolo-diazepine ring system as the acetyl-mimetic. Further studies demonstrated that inhibition of BRD4 by (+)-JQ1 unveiled the MYC downregulation and, consequently, a genome-wide inhibition of its target genes (Filippakopoulos et al., 2010; Delmore et al., 2011). These results underlined significant preclinical activity of this inhibitor in MYC-driven B-NHL, including the aggressive, so-called "double hit" lymphoma (DHL), characterized by simultaneous oncogenic activation of *MYC* and/or *BCL2/BCL6* (Johnson-Farley et al., 2015). Accordingly, (+)-JQ1 could increase survival of mice xenografted with MYC-driven lymphoma, including those ones bearing either TP53 deletions or intrinsic resistant to the topoisomerase II inhibitor etoposide (Hogg et al., 2016).

These promising results from (+)-JQ1 encouraged the development of BET inhibitors with similar chemical structure, including the BRD4 inhibitor CPI203 characterized by a higher bioavailability profile in mice (Normant et al., 2012; King et al., 2013). This agent displayed remarkable efficacy in different preclinical models of B-NHL, either as single agent or in combination with the BCL-2 antagonist venetoclax in DHLs (Esteve-Arenys et al., 2018), in DLBCL-ABC (Ceribelli et al., 2014) and in both ABC and GCB subtypes of DLBCL in combination with blockade of the CXCR4 chemokine receptor (Recasens-zorzo et al., 2018). In these studies, BRD4i activity was mainly related to the blockade of MYC transcriptional program. This is of special interest, as despite its central role in multiple hematological malignancies, including various subtypes of B-NHL, direct targeting of MYC was considered impossible until the demonstration that BET inhibition could regulate MYC activity in varied contexts, thanks to alleviation of BRD4 occupancy on MYC super-enhancers. Importantly, beside MYC, different anti-apoptotic proteins like BCL-2 and MCL-1 are also downregulated, either by direct transcription repression or as a downstream consequence of BRD4 antagonism (Vakoc, 2015). Unlike the expected general effects of BET inhibition in the elongation of transcription of several genes, changes in the expression of only a small subset of genes was observed in cultures and/or animals receiving this therapy, suggesting that bromodomain inhibitors might be suitable modulators of certain disease-associated genes. As an illustration, high levels of BRD4 co-localize in CLL cells with super-enhancer sites of genes and microRNAs belonging to the BCR-mediated signaling pathway with possible tumor-initiating activity, including *miR-21*, *miR-15*, *TCL1*, *IL21R*, and *IL4R*. Accordingly, in a mouse model of CLL, exposure to the BET inhibitor PLX51107 promoted an expression downmodulation of several tumor-associated genes, followed by consistent reduction in tumor burden (Ozer et al., 2018).

According to these promising results, in the last years a number of clinical leads have entered into trials for the treatment of hematological patients. Nevertheless, several side effects have been reported including some bone marrow and gastrointestinal toxicity that has forced to dose discontinuation or reduction. Nowadays, 18 BET inhibitors are being assessed in clinical trials either as single agents or in combination with other compounds (**Table 4**). While the data from various solid tumor trials look mitigated, several BETis including birabresib (OTX015, MK-8628), molibresib (GSK525762), RO6870810/TEN-010, and mivebresib (ABBV-075) have demonstrated remarkable clinical efficacy in myeloproliferative disorders, while other small molecule inhibitors such as PFI-1, BI-894999, FT-1101, INCB-54329, and CPI0610, a pharmacological derivative of CPI203, are currently undergoing human clinical trials in these patients (**Table 3**). Among these different molecules, molibresib has demonstrated an 18.5% ORR in various subtypes of NHLs including a CR in a DLBCL case (Dickinson et al., 2018). CPI0610 has also been evaluated in a phase I clinical trial (NCT01949883) in 64 R/R FL, DLBCL, or HL patients, showing leading to a complete remission in one FL case and in four DLBCL patients (Blum et al., 2018). In addition, the compound INCB057643 is currently being tested in a third phase I trial involving lymphoma patients, including some FL and DLBCL cases. In this evaluation trial, a CR has been achieved in one FL case whereas in two other patients, the disease has been stabilized (Forero-Torres et al., 2017). In the dose-escalation, open-label, phase I study with OTX015,

TABLE 4 | Drug combinations with non-approved epigenetic agents in B-NHL.

Epigenetic drug class	Drug	Agent used in combination	Trial identifier
HDAC inhibitor	CUDC-907	Rituximab, venetoclax, and bendamustine	NCT01742988
	Entinostat	Isotretinoin	NCT00098891
		Molibresib	NCT03925428
	Mocetinostat	Azacitidine	NCT00543582
EZH2 inhibitor	Tazemetostat	Fluconazole, omeprazole, and repaglinide	NCT03028103
		Atezolizumab and obinutuzumab	NCT02220842
		Prednisolone	NCT01897571
	PF 06821497	SOC	NCT03460977
BET inhibitor	Molibresib	Entinostat	NCT03925428
	RO6870810	Venetoclax and rituximab	NCT03255096
INCB057643	Gemcitabine, paclitaxel, rucaparib, abiraterone, ruxolitinib, and azacitidine		NCT02711137
	FT-1101	Azacitidine	NCT02543879

Source: https://clinicaltrials.gov/.

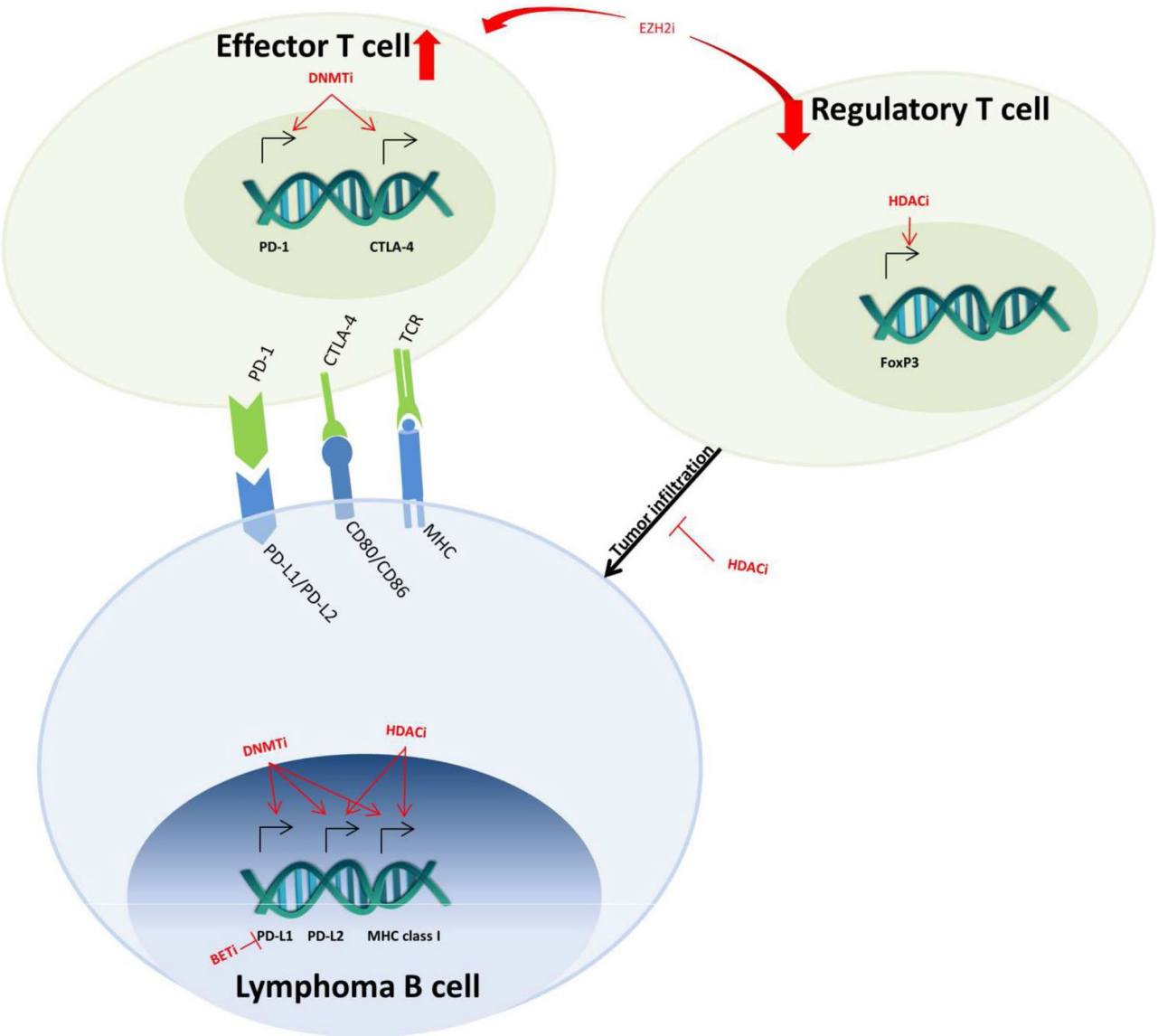


FIGURE 4 | Epigenetic-targeted effects on immuno-oncology mechanisms. In lymphoma B cells, DNMT, HDAC, and BET inhibitors (DNMTi, HDACi, BETi) regulate the expression of MHC class I and PD-1 ligands (PD-L1 and PD-L2). In effector T cells, DNMTi also upregulates the expression of PD-1 and CTLA-4, which leads to T-cell exhaustion. The effects of HDACi on FoxP3 decrease the infiltration of regulatory T cells into the tumor. Lastly, EZH2i decreases the regulatory and increases effector T-cell population in the microenvironment.

a 47% complete remission was reported in 17 DLBCL cases; however, objective responses were seen in only three DLBCL patients and clinical activity in other six B-NHL patients (NCT01713582) (Amorim et al., 2016). More recently, the BETis molibresib, CC-90010, and INCB054329 are being challenged in clinical trials including various hematological malignancies (NCT02431260, NCT01943851, and NCT03220347), but no data have been released so far.

Although at the moment most of the tested compounds aimed at inhibiting BET bromodomains are pan-BET inhibitors, many efforts are being focused in targeting BET proteins in a more specific and novel way. These new approaches include ABBV-744 (which targets bromodomain-containing protein II) (Sheppard et al., 2018), the bivalent BET inhibitors AZD5153 and MT1 (a JQ1-derived BETi) (Rhyasen et al., 2016; Tanaka et al., 2016), and the so-called BET-PROTACs (QCA570, dBET6, BETd-260, and ARV-771) that drive BET proteins to their degradation by proteolysis-targeted chimera (Raina et al., 2016; Winter et al., 2017; Qin et al., 2018). These molecules have shown both to promote apoptosis in MCL-derived cells resistant to the first-in-class Bruton's kinase (BTK) inhibitor ibrutinib as well as to increase survival compared to OTX015-treated MCL xenografts (Sun et al., 2018). Although promising results have been reported for this new generation of BET-targeting agents in preclinical

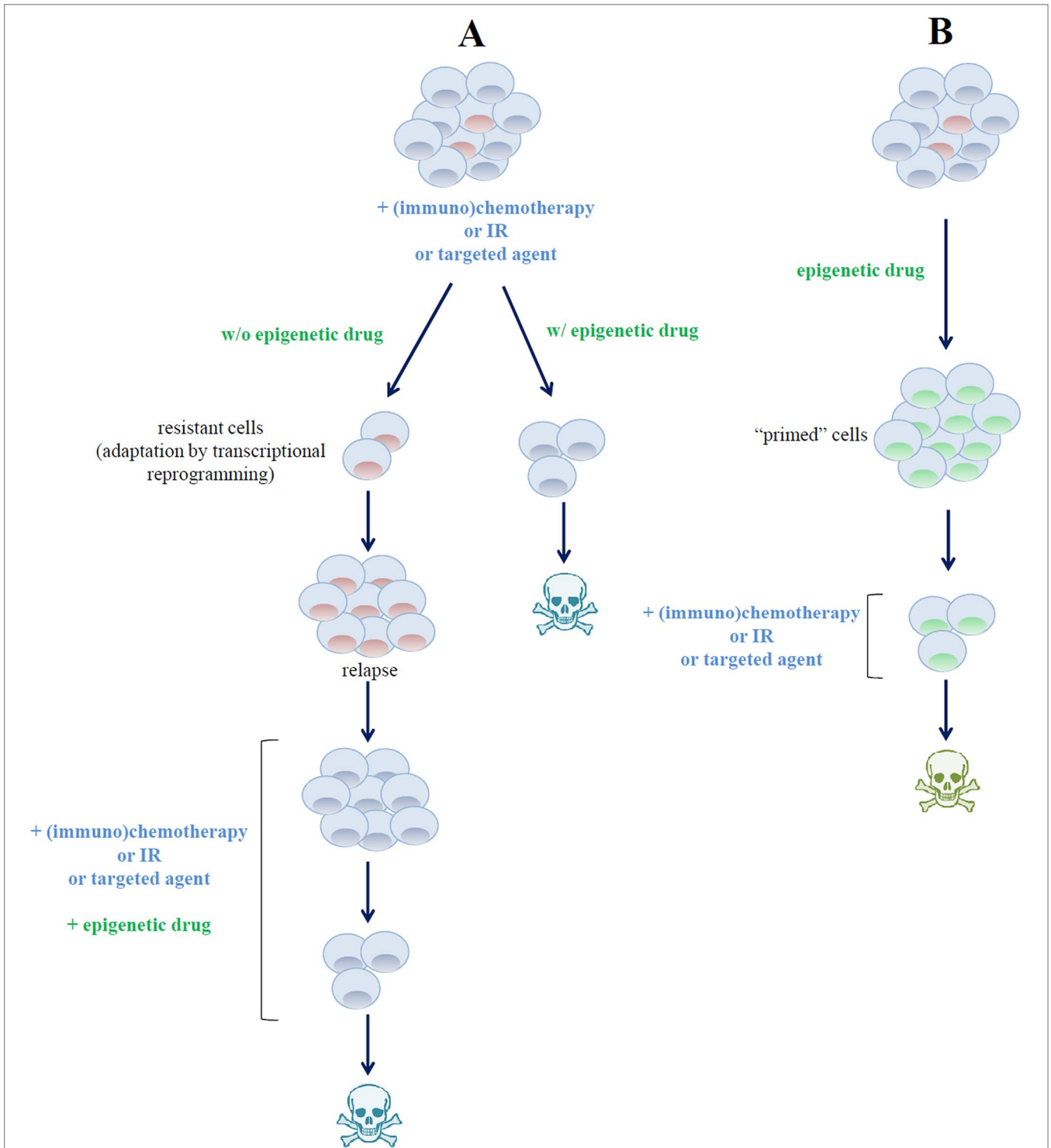


FIGURE 5 | Different strategies for combining drugs targeting epigenetic regulators in B-cell lymphoma. **(A)** Epigenetic drugs can be used to overcome (left axis) or to circumvent (right axis) malignant B-cell resistance to targeted agents or to standard chemotherapeutic regimens. **(B)** Pre-exposure of tumor cells to epigenetic drugs induces profound modifications of cell transcriptional profile, thus priming them to the cytotoxic effect of chemotherapeutic and targeted agents.

studies, their therapeutic window when moving to clinical trials has still to be evaluated.

Non-BET Bromodomain-Containing Proteins: the Histone Acetyltransferase CREB-Binding Protein (CBP)

As previously mentioned, chromatin modifications can regulate several important features of cell function. Among these modifications, histone lysine acetylation is generally associated with activation of gene expression (Shahbazian and Grunstein, 2007). HAT enzymes can deposit acetyl marks on histones and modify chromatin structure. Such marks are also recognized by bromodomains, thus adding a second level of regulation of the transcription process (Kouzarides, 2007). The transcriptional co-activators CBP/p300 are highly homologous, multifunctional proteins that encode a single bromodomain each and possess HAT activity (Chen and Li, 2011; Delvecchio et al., 2013). CBP/p300 act as transcriptional co-factors, involved in the regulation of several biological processes (Dancy and Cole, 2015). Animal studies have shown that CBP and p300 are required for the generation and activity of normal hematopoietic stem cells as well as for adult hematopoietic stem cell maintenance and function (Chan et al., 2011; Rebel et al., 2002). Consequently, CBP ablation has a direct impact on the quiescence, apoptosis, and self-renewal of adult hematopoietic stem cells (Chan et al., 2011) and CBP/p300 have a tumor suppressor role in mice models (Kung et al., 2000; Kang-Decker et al., 2004; Chan et al., 2011). This role of CBP and p300 as tumor suppressors has been also observed in B-NHL, where its inactivating mutation is a common event in FL and DLBCL, providing a rationale for employing drugs with the capacity to modulate acetylation and deacetylation processes in these tumors (Cerchietti et al., 2010; Mullighan et al., 2011; Pasqualucci et al., 2011a).

CHROMATIN REMODELERS: SWI/SNF AND BRG1 AND ARID1

The SWItch/Sucrose Non-Fermentable (SWI/SNF) complex was initially discovered in yeast. It is composed by polypeptides associated with a subset of proteins codified by the SWI1, SWI2, SNF2, SWI3, SWI5, and SWI6 genes (Pazin and Kadonaga, 1997). This complex regulates gene transcription by altering DNA–nucleosome interactions at expenses of ATP consumption, thus facilitating or impeding the accession of the transcription machinery at concrete genomic regions (Workman and Kingston, 2002). Several studies have reported its capacity to repair nucleotide excisions and DNA double-strand breaks by homologous recombination (Chai et al., 2005). The mammalian analog of the SWI/SNF complex (mSWI/SNF) is the BRG1-Associated Factors (BAF) complex. It comprised approximately 11 subunits encoded by 19 distinct genes assembled in different combinations according to its specific molecular mechanism of action, and in a concrete genomic region. Two of the BAF components are the human Brahma (hBRM, also SMARCA2) and the Brahma-related gene 1 (BRG1, also SMARCA4). These proteins are ATPase subunits (Khavari et al., 1993) and either one or the other constitute the core component of the BAF complex. They contain BDs within their structure that recognize and contact acetyl groups present in histone proteins (Wang et al., 1996). Although they share similarities in their domain composition, they interact with different families of transcription factors what confers to them specific functions in the BAF complex (Kadam and Emerson, 2003).

BRG1 has been reported to be the most frequently mutated protein of the BAF complex in cancer. Classically, it has been described as a tumor suppressor gene as inactivating mutations of its protein have been found in numerous solid tumors like breast, lung, gastric, bladder, colon, ovarian cancers, and melanomas (Atlas et al., 2012; Hodis et al., 2012; Jelinic et al., 2014), but also in determined B-NHL subtype like DLBCL and MCL (Cuadros et al., 2017). Concretely, these loss-of-function mutations lead to the upregulation of the pro-survival gene *BCL2L1* in MCL, conferring to this malignancy primary resistance to treatment or eventually relapse after dual exposure to ibutrinib and venetoclax (Agarwal et al., 2019). Other studies described BRG1 as a potent oncogene, since its function was required for AML progression in mice, through its binding to *MYC* enhancer region and consequent aberrant expression of this second oncogene (Shi et al., 2013; Buscarlet et al., 2014).

Beside BRG1, several BRG-/BRM-associated factors (BAF subunits) participate in tumoral progression. Two of these subunits, namely the AT-Rich Interaction Domain 1A (ARID1A/BAF250A) and its homologous ARID1B/BAF250B, contain domains capable of recognizing and binding to AT-enriched genomic regions and C terminus region, stimulating the activation of transcription in a glucocorticoid receptor-dependent manner. The presence of each of them in the complex is mutually exclusive, suggesting specific roles at concrete genomic regions (Wang et al., 2004).

Mutations that truncate the ARID1A sequence and promote its degradation have been widely characterized in endometrial carcinoma (Kandoth et al., 2013), colon cancer (Atlas et al., 2012), stomach cancer (Wang et al., 2011), bladder cancer (Gui et al., 2011), neuroblastoma (Sausen et al., 2013), and pancreatic or hepatocellular carcinoma (Biankin et al., 2012; Fujimoto et al., 2012), evidencing the role of this protein in preventing tumoral progression. Similar to the mutations reported for ARID1A, *truncating mutations have also been identified for ARID1B* although in a lesser frequency and *most of them associated with neurodevelopmental disorders (*Santen et al., 2012*) or neuroblastomas (*Lee et al., 2017*).* ARID1B knockdown has been reported to destabilize the SWI/SNF complex and inhibit cell proliferation in both ARID1A-mutant cancer cell lines and primary tumor cells, suggesting that this protein could constitute an interesting therapeutic target for the treatment of ARID1A-mutant tumors (Helming et al., 2014).

INDIRECT INHIBITION OF EPIGENETIC DYSREGULATION BY IDH INHIBITORS

The enzyme isocitrate dehydrogenase (IDH) catalyzes the conversion of isocitrate into α-ketoglutarate (α-KG) by oxidative decarboxylation using $NADP^+$ as a cofactor. The IDH1 isomer is located in the cytosol and the peroxisomes, whereas IDH2 is

found in the mitochondria. IDH enzymes play an important role in the tricarboxylic (TCA) or Krebs' cycle, but are also related with other cellular functions such as the regulation of redox balance (Dang et al., 2016; Dang and Su, 2017). Mutations in *IDH* genes are most commonly found in the R132 codon of *IDH1* and the R172 and R140 codons of *IDH2*, which correspond to evolutionarily conserved residues in the enzyme active site which is critical for substrate binding. Mutant forms of IDH have much lower catalytic activity and are associated with metabolic alterations. More importantly, mutant IDH enzymes gain neomorphic activity as they convert α-KG into 2-hydroxyglutarate (2-HG). Under homeostatic conditions, 2-HG is only produced by errors in catalysis and it is maintained at low levels due to the action of 2-HG-hydroxigenases (2-HGHD). Unlike in bacteria and plants, 2-HG has no known physiological function in mammals (Dang and Su, 2017). 2-HG is structurally similar to α-KG and acts as a competitive inhibitor, blocking the activity of α-KG-dependent dioxygenases. This group of enzymes includes the TET family of hydroxylases, which participate in DNA demethylation, and the JMJ domain-containing histone demethylases (Dang and Su, 2017). The consequent aberrant hypermethylation of both DNA and histones has been associated to a blockade in differentiation in hematopoietic cells (Figueroa et al., 2010; Losman et al., 2013), hepatocytes (Saha et al., 2014), and mesenchymal stem cells (Jin et al., 2015), among other cell types.

Homozygous missense mutations in both *IDH1* or *IDH2* have been described in several cancer types, including glioma, cholangiocarcinoma, and hematological tumors, such as AML and MDS (Dang et al., 2016). Although infrequent, mutations have also been found in lymphoid malignancies like angioimmunoblastic T-cell lymphomas (Cairns et al., 2012) and acute lymphocytic leukemia, both in pediatric (Andersson et al., 2011; Tang et al., 2012) and adult cases (Kang et al., 2009; Abbas et al., 2010; Zhang et al., 2012). Dysregulation of the IDH pathway has also been reported in CLL, as leukemic B cells from these patients show overexpression of IDH1 and lower levels of IDH2 when compared to healthy B cells (Van Damme et al., 2016).

Two IDH inhibitors have been recently approved by the FDA for the treatment of R/R AML in adults. Enasidenib (AG-221) targets IDH2 with R172S, R172K, and R140Q mutations, whereas ivosidenib (AG-120) targets IDH1 with susceptible mutations, such as R132H and R132C (Han et al., 2019). Other non-approved IDH inhibitors are currently in clinical trials involving patients with advanced hematological cancers. Among these molecules, AG-881 is a pan-inhibitor of both IDH1 and IDH2 that can penetrate the blood–brain barrier, while IDH305, FT-2102, and BAY-1436032 are IDH1-specific inhibitors (Dang et al., 2016; Montalban-Bravo and DiNardo, 2018). At the preclinical level, the pharmacological IDH2 inhibitor AGI-6780 displayed synergistic cytotoxicity in MCL and BL cell lines in combination with the proteasome inhibitor carfilzomib, mediated by the blockade of tricarboxylic acid cycle and the decrease in ATP levels, as a consequence of enhanced IDH2 enzymatic inhibition (Bergaggio et al., 2019). Thus, although activating mutations of IDH genes are rare in B-NHL, there may be some room to evaluate, alone or in combination with standard chemotherapy, some of the molecules exhibiting clinical activity in non-lymphoid patients.

COMBINATION INVOLVING EPIGENETIC-TARGETING APPROACHES

Concomitant Targeting of Different Epigenetic Modulators

In recent years, thanks to the many works directed to characterize and get a better understanding of the human epigenome, it came out that more than 50% of the human cancers account for aberrant changes in chromatin organization at certain genomic regions, as a consequence of mutations in enzymes involved in the regulation of chromatin structure (You and Jones, 2012; The Cancer Genome Atlas Research Network, 2013). Changes in the activity of these chromatin modifiers can lead not only to the initiation of a tumor formation process but also to its progression, metastasis, development of drug resistances, and further relapse and/or escape from immune surveillance (Jones et al., 2016). Therapeutic modulation of such alterations can be achieved with chemical compounds that broadly affect the structure of the DNA such as DNMTis, histone HDACis, or BETis (**Figure 4**). While single-agent clinical trials with these compounds have been conducted with some success in MDS or R/R AML patients receiving azacitidine (Scott, 2016; Schuh et al., 2017) or in R/R FL, MZL, and MCL patients treated with vorinostat (Kirschbaum et al., 2011; Ogura et al., 2014), the association of these agents with other compounds has also been tested. As an example, the combinatorial treatment with vorinostat and the sirtuin inhibitor niacinamide was evaluated in R/R NHL and HL cases (NCT00691210) (Amengual et al., 2013), but it achieved a modest efficiency with an ORR below 50% (Olsen et al., 2007). Other examples include the combination of panobinostat with decitabine which displayed synergistic caspase-dependent cell death in DLBCL cells (Kalac et al., 2011) or the combination of romidepsin with the antimetabolite pralatrexate for the treatment of relapsed PTCL (Amengual et al., 2018).

A different therapeutic approach consists in targeting specifically certain chromatin regulatory proteins to achieve a more restricted effect in the transcription of a concrete subset of genes. Promising examples are the inhibition of the DOT1-like (DOT1L) histone H3K79 methyltransferase with pinometostat (EPZ-5676) in adults with MLL/KMT2A-driven leukemia (NCT02141828) (Stein et al., 2018) or inhibition of histone H3K4 and K9 demethylation by the lysine-specific demethylase 1 (LSD1) inhibitor seclidemstat, currently being assessed in clinical trials to treat refractory Ewing sarcomas (NCT03600649).

Combinations with chemical compounds that broadly affect an epigenetic mark and a specific inhibitor of a chromatin-modifying enzyme, such as the EZH2 inhibitor GSK126 and romidepsin, have also been assessed in preclinical studies with DLBCL-GCB cell lines, leading to synergistic tumor growth inhibition effects in mice (Lue et al., 2019). Another example of the strategies currently

evaluated in clinical studies is the concomitant treatment of drug-resistant MM with panobinostat and bortezomib (NCT01083602) (Richardson et al., 2016).

Finally, and in concordance with the concept that acquired resistance to chemotherapy is tightly linked to changes in chromatin structure, many efforts have been made in identifying combinational strategies associating different types of cytotoxic drugs to small molecule regulators of chromatin modifiers. As an example, the dinitroazetidine derivative RRx-001 administered in combination with radiation, chemotherapy, or immunotherapies promotes the generation of reactive oxygen and nitrogen species, leading to the oxidation of the cysteines present at the catalytic sites of DNMTs and HDACs. This phenomenon entrains the inhibition in DNMT and HDAC enzymatic activities, with subsequent alterations in the chromatin structure. The therapeutic benefits of this compound have been assessed in phase II clinical trials both as a radio- and chemo-sensitizer, as well as a way to prone tumor response to conventional therapies (NCT02215512, NCT02452970, NCT02096341, NCT02871843) (Oronsky et al., 2017; Zhao et al., 2017).

Combination of Epigenetic Drugs With Other Classes of Anti-Tumoral Drugs

The use of epigenetic agents combined with other anti-tumoral drugs may represent the future of epigenetic-targeted therapies (**Figure 5**). The rationale of such combinations would be, on the one hand, to benefit from the transcriptional effects of targeting epigenome. Indeed, growing evidences are showing that epigenetic therapy, using DNMTi or HDACi, in combination with conventional therapy or immunotherapy, might be an up-and-coming step toward the development of new and efficient cancer treatment strategies (Brahmer et al., 2012; Sharma and Allison 2015; Topalian et al., 2015; Issa et al., 2017). Accordingly, the acquired capacity of tumors to resist chemotherapy is related with changes in the cancer cell's epigenome, which might affect directly the cell cycle and/or some key apoptosis regulators (Fodale et al., 2011).

In a phase I study, Clozel and collaborators proposed a new approach to beaten chemotherapy resistance in DLBCL patients. The authors demonstrated a high rate of complete remission when a 5-day exposure to azacitidine followed by treatment with R-CHOP was employed. Mechanistically, the treatment leads to the demethylation of the chemoresistance-associated gene SMAD1 and subsequent chemosensitization (Clozel et al., 2013). Based on these results, an ongoing phase I study using azacitidine combined with R-CHOP in therapy-naive DLBCL, grade 3B FL, or transformed FL patients is showing promising preliminary results (NCT02343536). Finally, the safety and tolerability of adding oral azacitidine to R-ICE therapy is being evaluated in R/R DLBCL patients (NCT03450343).

Regarding HDACi, *in vitro* studies have demonstrated that this class of agents can synergize with chemotherapy. Globally, these trials have had mixed heterogeneous results. Among the potently successful studies, in indolent B-NHL the vorinostat/rituximab combination exhibited a nice activity with an acceptable safety profile and durable responses (Chen et al., 2015). Ageberg and collaborators also showed that valproic acid sensitizes to CHOP and enhances the CHOP ability to induce apoptosis in DLBCL cell lines (Ageberg et al., 2013). Subsequently, it has been shown in a small set of DLBCL patients that the administration of valproate before R-CHOP treatment upregulated the CD20 levels and increased the efficacy of anti-CD20-based therapy (Damm et al., 2015). Recently, the VALFRID phase I trial (NCT01622439) showed that valproate when added to standard R-CHOP therapy is secure, tolerable, and increases OS in DLBCL patients (Drott et al., 2018). The efficacy of vorinostat combined with cyclophosphamide, etoposide, and prednisone (R-CVEP) was evaluated in aged patients with R/R DLBCL (NCT00667615); however, the R-CVEP association did not reach the criteria for cohort expansion (Straus et al., 2015). Similarly, the combination of vorinostat with R-CHOP was evaluated in the SWOG S0806 phase I/II trial (NCT00972478) without success in DLBCL patients (Persky et al., 2018). Panobinostat was tested in combination with conventional therapy and although the data from the clinical trial NCT01238692 suggested that as a single agent this drug induces a durable response in a subset of R/R DLBCL patients, its combination with rituximab did not improve the response rate (Assouline et al., 2016). Similarly, Barnes and collaborators observed that this combination was effective in a minority of DLBCL patients heavily pretreated (NTC01282476) (Barnes et al., 2018). The combination with immunomodulatory drug (IMiD) lenalidomide was assessed in a phase I/II clinical trial in patients with R/R HL (NCT01460940); however, the combination was not advantageous over single-agent treatment and raised relevant concerns regarding the toxicity (Maly et al., 2017). Finally, preclinical data have shown that belinostat exhibits synergistic cytotoxic activity in DLBCL cell lines when associated to the microtubule-interfering drug vincristine, mediated by the prevention of cell polyploidy (Havas et al., 2016).

Regarding EZH2 inhibitors, combinatorial treatments with tazemetostat and the anti-programmed death-ligand 1 (PDL1) antibody atezolizumab (NCT02220842), prednisone alone, or combined with other components of CHOP regimen are currently being evaluated in patients with refractory DLBCL (NCT02889523) (Gulati et al., 2018). Moreover, combinations with EZH2 inhibitors and inhibitors of the BCR signaling cascade such as ibrutinib, the spleen tyrosine kinase (SYK) inhibitor tamatinib, the mammalian target of rapamycin (mTOR) inhibitor everolimus, or MAPK inhibitor have also been challenged in pre-clinical models of DLBCL (Brach et al., 2017; Lue et al., 2017). Other therapeutic strategies currently assessed in pre-clinical studies for the treatment of MM consist in combining the inhibitor tazemetostat with IMiDs such as lenalidomide or pomalidomide (Dang et al., 2016), glucocorticoid receptor agonists (dexamethasone or prednisolone), proteasome inhibitors (bortezomib or ixazomib) (Drew et al., 2017), or HDACis (Issa et al., 2017).

Finally, in combination with the CDK4/6 inhibitor palbocilib, the BETi JQ1 has shown synergistic activity in MCL *in vitro* and *in vivo* (Sun et al., 2015). Another member of the CDK family, CDK9, is a core component of the assembly of the positive transcription elongation factor complex (P-TEFb), which is recruiting by BRD4.

In relation with this, the BETi BI-894999 shows profound synergy with CDK9 inhibitors alvocidib and LDC000067 in both *in vitro* and *in vivo* models of hematological malignancies (Doroshow et al., 2017). Among other promising combinations, CPI203 combined with the proteasome inhibitor bortezomib or with lenalidomide was particularly efficient in aggressive bortezomib-resistant MCL tumors (Moros et al., 2014), and GS-5829 synergistically interacted with venetoclax or with BCR-interfering agents in preclinical models of DLBCL, MCL, and/or CLL (Bates et al., 2016; Kim et al., 2017).

CONCLUSIONS

Besides the well-known genomic changes, several epigenetic modifications that result in an altered chromatin state and alterations in the DNA methylation status have been described in lymphoma cells. In general, these alterations favor the malignant transformation and/or tumor progression. Among the mechanisms that may apply to several lymphoma entities, epigenetic activation of suppressors of lineage fidelity leads to downregulation of lineage-specific genes, while additional silencing of essential transcription factors through H3K27 trimethylation avoids the restoration of the cell type characteristic expression program. Therefore, there is undoubtedly an important clinical role for epigenetic drugs across the spectrum of lymphoid malignancies, including B-NHL.

In the last decade, the progresses in the awareness of epigenetic changes in lymphoma cells have paved the way for targeted therapy alternatives employing epigenetic drugs. Treatment approaches such as HDAC inhibition or DNMT blockade have shown remarkable activity in specific subsets of lymphoma patients who remained unresponsive to or relapsed after standard therapy. These drugs have already been added into routine use for patients with a particular lymphoma/leukemia subtype and are the most broadly studied now. However, the identification of biomarkers of clinical sensitivity/resistance to these agents is still needed in order to better identify those lymphoma patients suitable for treatment with these drugs, and for the design of rationally based targeted combination therapies. Although several epigenetic drugs can be successfully combined with standard chemotherapy, allowing to decrease the chemotherapy doses and to limit toxicities and adverse effects, co-administration of two epigenetic modulators like DNA hypomethylating agents and HDAC inhibitors, for example, can also show synergistic molecular effects, resulting in increased antitumor activity.

In the light of the large number of drugs currently in clinical development in B-NHL patients, selection of the most relevant targeted therapies will be extremely important to move the field ahead. Epigenetic drugs with more specific targets, such as EZH2 inhibitors or BRD4 inhibitors, but also the newer epigenetic agents like PRMT5 and IDH inhibitors, are also of great interest, as demonstrated by a particularly rapid translation from bench to bedside within the past 5 years.

Despite these considerable advances in epigenetic drug therapy in B-cell lymphoma, there is still some way to go before reaching a complete overview of the complex landscape of the epigenetic modifications occurring during the lymphomagenesis, and much work is still to be done to improve the rationale use of epigenetic drugs in lymphoma patients. According to the promising reports from several trials involving the newest agents and the most innovative drug combinations in B-NHL patients with relapse disease, it seems that we are entering a very exciting era for the field of epigenetics in lymphoma.

AUTHOR CONTRIBUTIONS

MR, DR, MA, MF and GR made a substantial contribution to all aspects of the preparation of this manuscript.

FUNDING

The authors received financial support from Fondo de Investigación Sanitaria PI15/00102 and PI18/01383, European Regional Development Fund (ERDF) "Una manera de hacer Europa" (to GR). The authors received fundings from TG Therapeutics and Celgene Corp to support researches unrelated to the present work. Funders were involved neither in the design, nor in the writing of this review.

REFERENCES

Abbas, S., Lugthart, S., Kavelaars, F. G., Schelen, A., Koenders, J. E., Zeilemaker, A., et al. (2010). Acquired mutations in the genes encoding IDH1 and IDH2 both are recurrent aberrations in acute myeloid leukemia: prevalence and prognostic value. *Blood* 116, 2122–2126. doi: 10.1182/blood-2009-11-250878

Abd Al Kader, L., Oka, T., Takata, K., Sun, X., Sato, H., Murakami, I., et al. (2013). In aggressive variants of non-Hodgkin lymphomas, Ezh2 is strongly expressed and polycomb repressive complex PRC1.4 dominates over PRC1.2. *Virchows Arch.* 463, 697–711. doi: 10.1007/s00428-013-1428-y

Agarwal, R., Chan, Y. C., Tam, C. S., Hunter, T., Vassiliadis, D., Teh, C. E., et al. (2019). Dynamic molecular monitoring reveals that SWI–SNF mutations mediate resistance to ibrutinib plus venetoclax in mantle cell lymphoma. *Nat. Med.* 25, 119–129. doi: 10.1038/s41591-018-0243-z

Ageberg, M., Rydström, K., Relander, T., and Drott, K. (2013). The histone deacetylase inhibitor valproic acid sensitizes diffuse large B-cell lymphoma cell lines to CHOP-induced cell death. *Am. J. Transl. Res.* 5, 170–183.

Agrawal, K., Das, V., Vyas, P., and Hajdúch, M. (2018). Nucleosidic DNA demethylating epigenetic drugs —A comprehensive review from discovery to clinic. *Pharmacol. Ther.* 188, 45–89. doi: 10.1016/j.pharmthera.2018.02.006

Almstedt, M., Blagitko-Dorfs, N., Duque-Afonso, J., Karbach, J., Pfeifer, D., Jäger, E., et al. (2010). The DNA demethylating agent 5-aza-2'2'-deoxycytidine induces expression of NY-ESO-1 and other cancer/testis antigens in myeloid leukemia cells. *Leuk. Res.* 34, 899–905. doi: 10.1016/j.leukres.2010.02.004

Amara, K., Ziadi, S., Hachana, M., Soltani, N., Korbi, S., and Trimeche, M. (2010). DNA methyltransferase DNMT3b protein overexpression as a prognostic factor in patients with diffuse large B-cell lymphomas. *Cancer Sci.* 101, 1722–1730. doi: 10.1111/j.1349-7006.2010.01569.x

Amengual, J. E., Clark-Garvey, S., Kalac, M., Scotto, L., Marchi, E., Neylon, E., et al. (2013). Sirtuin and pan-class I/II deacetylase (DAC) inhibition is synergistic in preclinical models and clinical studies of lymphoma. *Blood.* 122, 2104–2113. doi: 10.1182/blood-2013-02-485441

Amengual, J. E., Lichtenstein, R., Lue, J., Sawas, A., Deng, C., Lichtenstein, E., et al. (2018). A phase 1 study of romidepsin and pralatrexate reveals marked activity in relapsed and refractory T-cell lymphoma. *Blood.* 131, 397–407. doi: 10.1182/blood-2017-09-806737

Amorim, S., Stathis, A., Gleeson, M., Iyengar, S., Magarotto, V., Leleu, X., et al. (2016). Bromodomain inhibitor OTX015 in patients with lymphoma or multiple myeloma: a dose-escalation, open-label, pharmacokinetic, phase 1 study. *Lancet Haematol.* 3, e196–e204. doi: 10.1016/S2352-3026(16)00021-1

Andersson, A. K., Miller, D. W., Lynch, J. A., Lemoff, A. S., Cai, Z., Pounds, S. B., et al. (2011). IDH1 and IDH2 mutations in pediatric acute leukemia. *Leukemia* 25, 1570–1577. doi: 10.1038/leu.2011.133

Assouline, S. E., Nielsen, T. H., Yu, S., Alcaide, M., Chong, L., MacDonald, D., et al. (2016). Phase 2 study of panobinostat with or without rituximab in relapsed diffuse large B-cell lymphoma. *Blood*. 128, 185–194. doi: 10.1182/blood-2016-02-699520

Atlas, T. C. G. N., Muzny, D., Bainbridge, M., Chang, K., Dinh, H., Drummond, J., et al. (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487, 330–337. doi: 10.1038/nature11252

Avramis, V. I., Powell, W. C., and Mecum, R. A. (1989). Cellular metabolism of 5,6-dihydro-5-azacytidine and its incorporation into DNA and RNA of human lymphoid cells CEM/O and CEM/dCk(-). (–). *Cancer Chemother. Pharmacol.* 24, 155–160. doi: 10.1007/BF00300235

Azad, N., Zahnow, C. A., Rudin, C. M., and Baylin, S. B. (2013). The future of epigenetic therapy in solid tumours - —Lessons from the past. *Nat. Rev. Clin. Oncol.* 10, 256–266. doi: 10.1038/nrclinonc.2013.42

Barnes, J. A., Redd, R., Fisher, D. C., Hochberg, E. P., Takvorian, T., Neuberg, D., et al. (2018). Panobinostat in combination with rituximab in heavily pretreated diffuse large B-cell lymphoma: Results of a phase II study. *Hematol. Oncol.* 36, 633–637. doi: 10.1002/hon.2515

Bates, J., Kusam, S., Tannheimer, S., Clarke, A., Kenney, T., Breckenridge, D., et al. (2016). Combination of the BET inhibitor GS-5829 and a BCL2 inhibitor resulted in broader activity in DLBCL and MCL cell lines. *Blood* 128, 5104.

Batlevi, C. L., Crump, M., Andreadis, C., Rizzieri, D., Assouline, S. E., Fox, S., et al. (2017). A phase 2 study of mocetinostat, a histone deacetylase inhibitor, in relapsed or refractory lymphoma. *Br. J. Haematol.* 178, 434–441. doi: 10.1111/bjh.14698

Batlevi, C. L., Kasamon, Y., Bociek, R. G., Lee, P., Gore, L., Copeland, A., et al. (2016). ENGAGE- 501: Phase II study of entinostat (SNDX-275) in relapsed and refractory Hodgkin lymphoma. *Haematologica* 101, 968–975. doi: 10.3324/haematol.2016.142406

Beà, S., Ribas, M., Hernandez, J. M., Bosch, F., Pinyol, M., Hernández, L., et al. (1999). Increased number of chromosomal imbalances and high-level DNA amplifications in mantle cell lymphoma are associated with blastoid variants. *Blood* 93, 4365–4374.

Beà, S., Valdés-Mas, R., Navarro, A., Salaverria, I., Martín-Garcia, D., Jares, P., et al. (2013). Landscape of somatic mutations and clonal evolution in mantle cell lymphoma. *Proc. Natl. Acad. Sci. U. S. A.* 110,18250–18255. doi: 10.1073/pnas.1314608110

Bedford, M. T., and Clarke, S. G. (2009). Protein arginine methylation in mammals: Who, what, and why. *Mol. Cell* 33, 1–13. doi: 10.1016/j.molcel.2008.12.013

Béguelin, W., Popovic, R., Teater, M., Jiang, Y., Bunting, K. L., Rosen, M., et al. (2013). EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. *Cancer Cell* 23, 677–692. doi: 10.1016/j.ccr.2013.04.011

Beisler, J. A., Abbasi, M. M., and Driscoll, J. S. (1979). Synthesis and antitumor activity of 5-azacytosine arabinoside. *J. Med. Chem.* 22, 1230–1234. doi: 10.1021/jm00196a015

Belinsky, S. A., Klinge, D. M., Stidley, C. A., Issa, J. P., Herman, J. G., March, T. H., et al. (2003). Inhibition of DNA methylation and histone deacetylation prevents murine lung cancer. *Cancer Res.* 63, 7089–7093.

Ben-Kasus, T., Ben-Zvi, Z., Marquez, V. E., Kelley, J. A., and Agbaria, R. (2005). Metabolic activation of zebularine, a novel DNA methylation inhibitor, in human bladder carcinoma cells. *Biochem. Pharmacol.* 70, 121–133. doi: 10.1016/j.bcp.2005.04.010

Berdasco, M., and Esteller, M. (2010). Aberrant epigenetic landscape in cancer: How cellular identity goes awry. *Dev. Cell.* 19, 698–711. doi: 10.1016/j.devcel.2010.10.005

Bergaggio, E., Riganti, C., Garaffo, G., Vitale, N., Mereu, E., Bandini, C., et al. (2019). IDH2 inhibition enhances proteasome inhibitor responsiveness in hematological malignancies. *Blood* 133, 156 LP–15 167. doi: 10.1182/blood-2018-05-850826

Beumer, J. H., Parise, R. A., Newman, E. M., Doroshow, J. H., Synold, T. W., Lenz, H. J., et al. (2008). Concentrations of the DNA methyltransferase inhibitor 5-fluoro-2′- deoxycytidine (FdCyd) and its cytotoxic metabolites in plasma of patients treated with FdCyd and tetrahydrouridine (THU). *Cancer Chemother. Pharmacol.* 62, 363–368. doi: 10.1007/s00280-007-0603-8

Biankin, A. V., Waddell, N., Kassahn, K. S., Gingras, M. C., Muthuswamy, L. B., Johns, A. L., et al. (2012). Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature*. 491, 399–405. doi: 10.1038/nature11547

Bisserier, M., and Wajapeyee, N. (2018). Mechanisms of resistance to ezh2 inhibitors in diffuse large B-cell lymphomas. *Blood*. 131, 2125–2137. doi: 10.1182/blood-2017-08-804344

Blum, K. A., Abramson, J., Maris, M., Flinn, I., Goy, A., Mertz, J., et al. (2018). 41OA phase I study of CPI-0610, a bromodomain and extra terminal protein (BET) inhibitor in patients with relapsed or refractory lymphoma. *Ann. Oncol.* 29, A410. doi: 10.1093/annonc/mdy048

Blum, K. A., Liu, Z., Lucas, D. M., Chen, P., Xie, Z., Baiocchi, R., et al. (2010). Phase I trial of low dose decitabine targeting DNA hypermethylation in patients with chronic lymphocytic leukaemia and non-Hodgkin lymphoma: dose-limiting myelosuppression without evidence of DNA hypomethylation. *Br. J. Haematol.* 29, 410. doi: 10.1111/j.1365-2141.2010.08213.x

Bobrowicz, M., Dwojak, M., Pyrzynska, B., Stachura, J., Muchowicz, A., Berthel, E., et al. (2017). HDAC6 inhibition upregulates CD20 levels and increases the efficacy of anti-CD20 monoclonal antibodies. *Blood* 130, 1628–1638. doi: 10.1182/blood-2016-08-736066

Brach, D., Johnston-blackwellBlackwell, D., Drew, A., Lingaraj, T., Motwani, V., Warholic, N. M., et al. (2017). EZH2 inhibition by tazemetostat results in altered dependency on B-cell activation signaling in DLBCL 16. *Mol. Cancer Ther.* 2586-2597. doi: 10.1158/1535-7163.MCT-16-0840

Bracken, A. P., and Helin, K. (2009). Polycomb group proteins: Navigators of lineage pathways led astray in cancer. *Nat. Rev. Cancer* 9, 773–784. doi: 10.1038/nrc2736

Bradner, J. E., West, N., Grachan, M. L., Greenberg, E. F., Haggarty, S. J., Warnow, T., et al. (2010). Chemical phylogenetics of histone deacetylases. *Nat. Chem. Biol.* 6, 238–243. doi: 10.1038/nchembio.313

Brahmer, J. R., Tykodi, S. S., Chow, L. Q. M., Hwu, W.-J., Topalian, S. L., Hwu, P., et al. (2012). Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N. Engl. J. Med.* 366, 2455–2465. doi: 10.1056/NEJMoa1200694

Buglio, D., Georgakis, G. V., Hanabuchi, S., Arima, K., Khaskhely, N. M., Liu, Y.-J., et al. (2008). Vorinostat inhibits STAT6-mediated TH2 cytokine and TARC production and induces cell death in Hodgkin lymphoma cell lines. *Blood* 112, 1424–1433. doi: 10.1182/blood-2008-01-133769

Buscarlet, M., Krasteva, V., Ho, L., Simon, C., Hébert, J., Wilhelm, B., et al. (2014). Essential role of BRG, the ATPase subunit of BAF chromatin remodeling complexes, in leukemia maintenance. *Blood*. 123, 1720–1728. doi: 10.1182/blood-2013-02-483495

Cairns, R. A., Iqbal, J., Lemonnier, F., Kucuk, C., De Leval, L., Jais, J.-P., et al. (2012). IDH2 mutations are frequent in angioimmunoblastic T-cell lymphoma. *Blood* 119, 1901–1903. doi: 10.1182/blood-2011-11-391748

Campo, E., Swerdlow, S. H., Harris, N. L., Pileri, S., Stein, H., and Jaffe, E. S. (2011). The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. *Blood* 117, 5019–5032. doi: 10.1182/blood-2011-01-293050

Cerchietti, L. C., Hatzi, K., Caldas-Lopes, E., Yang, S. N., Figueroa, M. E., Morin, R. D., et al. (2010). BCL6 repression of EP300 in human diffuse large B cell lymphoma cells provides a basis for rational combinatorial therapy. *J. Clin. Invest.* 120, 4569–4582. doi: 10.1172/JCI42869

Ceribelli, M., Kelly, P. N., Shaffer, A. L., Wright, G. W., Xiao, W., Yang, Y., et al. (2014). Blockade of oncogenic IκB kinase activity in diffuse large B-cell lymphoma by bromodomain and extraterminal domain protein inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 111, 11365–11370. doi: 10.1073/pnas.1411701111

Chai, B., Huang, J., Cairns, B. R., and Laurent, B. C. (2005). Distinct roles for the RSC and Swi/Snf ATP-dependent chromatin remodelers in DNA double-strand break repair. *Genes Dev.* 19, 1656–1661. doi: 10.1101/gad.1273105

Chaidos, A., Caputo, V., and Karadimitris, A. (2015). Inhibition of bromodomain and extra-terminal proteins (BET) as a potential therapeutic approach in haematological malignancies: Emerging preclinical and clinical evidence. *Ther. Adv. Hematol.* 6, 128–141. doi: 10.1177/2040620715576662

Chan-Penebre, E., Kuplast, K. G., Majer, C. R., Boriack-Sjodin, P. A., Wigle, T. J., Johnston, L. D., et al. (2015). A selective inhibitor of PRMT5 with in vivo and in vitro potency in MCL models. *Nat. Chem. Biol.* 11, 432–437. doi: 10.1038/nchembio.1810

Chan, W.-I., Hannah, R. L., Dawson, M. A., Pridans, C., Foster, D., Joshi, A., et al. (2011). The transcriptional coactivator Cbp regulates self-renewal and

differentiation in adult hematopoietic stem cells. *Mol. Cell. Biol.* 31, 5046–5060. doi: 10.1128/MCB.05830-11

Chen, J., and Li, Q. (2011). Life and death of transcriptional co-activator p300. *Epigenetics*. 6, 957–961. doi: 10.4161/epi.6.8.16065

Chen, R., Frankel, P., Popplewell, L., Siddiqi, T., Ruel, N., Rotter, A., et al. (2015). A phase II study of vorinostat and rituximab for treatment of newly diagnosed and relapsed/refractory indolent non-Hodgkin lymphoma. *Haematologica* 100, 357–362. doi: 10.3324/haematol.2014.117473

Cheng, J. C., Yoo, C. B., Weisenberger, D. J., Chuang, J., Wozniak, C., Liang, G., et al. (2004). Preferential response of cancer cells to zebularine. *Cancer Cell*. 6, 151–158. doi: 10.1016/j.ccr.2004.06.023

Chiappinelli, K. B., Strissel, P. L., Desrichard, A., Li, H., Henke, C., Akman, B., et al. (2015). Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell*. 162, 974–986. doi: 10.1016/j.cell.2015.07.011

Choi, W. J., Chung, H. J., Chandra, G., Alexander, V., Zhao, L. X., Lee, H. W., et al. (2012). Fluorocyclopentenyl-cytosine with broad spectrum and potent antitumor activity. *J. Med. Chem.* 55, 4521–4525. doi: 10.1021/jm3004009

Chuang, J. C., Warner, S. L., Vollmer, D., Vankayalapati, H., Redkar, S., Bearss, D. J., et al. (2010). S110, a 5-aza-2'2'-deoxycytidine-containing dinucleotide, is an effective DNA methylation inhibitor in vivo and can reduce tumor growth. *Mol. Cancer Ther.* 9, 1443–1450. doi: 10.1158/1535-7163.MCT-09-1048

Clozel, T., Yang, S. N., Elstrom, R. L., Tam, W., Martin, P., Kormaksson, M., et al. (2013). Mechanism-based epigenetic chemosensitization therapy of diffuse large B-cell lymphoma. *Cancer Discov.* 3, 1002–1019. doi: 10.1158/2159-8290.CD-13-0117

Cox, O. B., Krojer, T., Collins, P., Monteiro, O., Talon, R., Bradley, A., et al. (2016). A poised fragment library enables rapid synthetic expansion yielding the first reported inhibitors of PHIP(2), an atypical bromodomain. *Chem. Sci.* 7, 2322–2330. doi: 10.1039/C5SC03115J

Crump, M., Coiffier, B., Jacobsen, E. D., Sun, L., Ricker, J. L., Xie, H., et al. (2008). Phase II trial of oral vorinostat (suberoylanilide hydroxamic acid) in relapsed diffuse large-B-cell lymphoma. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* 19, 964–969. doi: 10.1093/annonc/mdn031

Cuadros, M., Sánchez-Martín, V., Herrera, A., Baliñas, C., Martín-Padrón, J., Boyero, L., et al. (2017). BRG1 regulation by miR-155 in human leukemia and lymphoma cell lines. *Clin. Transl. Oncol.* 19, 1010–1017. doi: 10.1007/s12094-017-1633-2

Damm, J. K., Gordon, S., Ehinger, M., Jerkeman, M., Gullberg, U., Hultquist, A., et al. (2015). Pharmacologically relevant doses of valproate upregulate CD20 expression in three diffuse large B-cell lymphoma patients in vivo. *Exp. Hematol. Oncol.* 4, eColection. doi: 10.1186/2162-3619-4-4

Dancy, B. M., and Cole, P. A. (2015). Protein lysine acetylation by p300/CBP. *Chem. Rev*. 115, 2419–2452. doi: 10.1021/cr500452k

Dang, L., and Su, S.-S. M. (2017). Isocitrate dehydrogenase mutation and (R)-2-hydroxyglutarate: from basic discovery to therapeutics development. *Annu. Rev. Biochem.* 86, 305–331. doi: 10.1146/annurev-biochem-061516-044732

Dang, L., Yen, K. E., and Attar, E. C. (2016). IDH mutations in cancer and progress toward development of targeted therapeutics. *Ann. Oncol.* 27, 599–608. doi: 10.1093/annonc/mdw013

Dawson, M. A., Prinjha, R. K., Dittmann, A., Giotopoulos, G., Bantscheff, M., Chan, W. I., et al. (2011). Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature*. 478, 529–533. doi: 10.1038/nature10509

De, S., Shaknovich, R., Riester, M., Elemento, O., Geng, H., Kormaksson, M., et al. (2013). Aberration in DNA methylation in B-cell lymphomas has a complex origin and increases with disease severity. *PLoS Genet*. 9, e1003137. doi: 10.1371/journal.pgen.1003137

Delmore, J. E., Issa, G. C., Lemieux, M. E., Rahl, P. B., Shi, J., Jacobs, H. M., et al. (2011). BET bromodomain inhibition as as a therapeutic strategy to target c-Myc. *Cell* 146, 904–917. doi: 10.1016/j.cell.2011.08.017

Delvecchio, M., Gaucher, J., Aguilar-Gurrieri, C., Ortega, E., and Panne, D. (2013). Structure of the p300 catalytic core and implications for chromatin targeting and HAT regulation. *Nat. Struct. Mol. Biol.* 20, 1040–1046. doi: 10.1038/nsmb.2642

Di Costanzo, A., Del Gaudio, N., Migliaccio, A., and Altucci, L. (2014). Epigenetic drugs against cancer: an evolving landscape. *Arch. Toxicol.* 88, 1651–1668. doi: 10.1007/s00204-014-1315-6

Dickinson, M., Kamdar, M., Huntly, B. J., Larrea, C. F., De, Cordoba, R., Mateos, M.-V., et al. (2018). A phase I study of molibresib (GSK525762), a selective bromodomain (BRD) and extra terminal protein (BET) inhibitor: results from part 1 of a phase I/II open label single agent study in subjects with non-Hodgkin's lymphoma (NHL). *Blood* 132, 1682–1682. doi: 10.1182/BLOOD-2018-99-117089

Doroshow, D. B., Eder, J. P., and LoRusso, P. M. (2017). BET inhibitors: a novel epigenetic approach. *Ann. Oncol.* 28, 1776–1787. doi: 10.1093/annonc/mdx157

Drew, A. E., Motwani, V., Campbell, J. E., Tang, C., Smith, J. J., Chesworth, R., et al. (2017). Abstract 5060: activity of the EZH2 inhibitor tazemetostat as a monotherapy and in combination with multiple myeloma therapies in preclinical models. *Cancer Res.* 77, n5060. doi: 10.1158/1538-7445.AM2017-5060

Dreyling, M. H., Roulston, D., Bohlander, S. K., Vardiman, J., and Olopade, O. I. (1998). Codeletion of CDKN2 and MTAP genes in a subset of non-Hodgkin's lymphoma may be associated with histologic transformation from low-grade to diffuse large-cell lymphoma. *Genes Chromosomes Cancer* 22, 72–78. doi: 10.1002/(SICI)1098-2264(199805)22:1<72::AID-GCC10>3.3.CO;2-G

Drott, K., Hagberg, H., Papworth, K., Relander, T., and Jerkeman, M. (2018). Valproate in combination with rituximab and CHOP as first-line therapy in diffuse large B-cell lymphoma (VALFRID). *Blood Adv.* 26, 1386–1392. doi: 10.1182/bloodadvances.2018019240

Dubois, S., Viailly, P., Mareschal, S., Bohers, E., Bertrand, P., Ruminy, P., et al. (2016). Next-generation sequencing in diffuse large B-cell lymphoma highlights molecular divergence and therapeutic opportunities : a LYSA study *Clin. Cancer Res.* 22, 2919–2929. doi: 10.1158/1078-0432.CCR-15-2305

Egger, G., Aparicio, A. M., Escobar, S. G., and Jones, P. A. (2007). Inhibition of histone deacetylation does not block resilencing of p16 after 5-aza-2'-deoxycytidine treatment. *Cancer Res.* 67, 346–356. doi: 10.1158/0008-5472.CAN-06-2845

Esteller, M., Corn, P. G., Baylin, S. B., and Herman, J. G. (2001). A gene hypermethylation profile of human cancer. *Cancer Res.* 61, 3225–3229.

Esteve-Arenys, A., Valero, J. G., Chamorro-Jorganes, A., Gonzalez, D., Rodriguez, V., Dlouhy, I., et al. (2018). The BET bromodomain inhibitor CPI203 overcomes resistance to ABT-199 (venetoclax) by downregulation of BFL-1/A1 in in vitro and in vivo models of MYC+/BCL2+ double hit lymphoma. *Oncogene*. 37, 1830–1844. doi: 10.1038/s41388-017-0111-1

Evens, A. M., Balasubramanian, S., Vose, J. M., Harb, W., Gordon, L. I., Langdon, R., et al. (2016). A phase I/II multicenter, open-label study of the oral histone deacetylase inhibitor abexinostat in relapsed/refractory lymphoma. *Clin. Cancer Res.* 22, 1059–1066. doi: 10.1158/1078-0432.CCR-15-0624

Fenaux, P., Mufti, G. J., Hellstrom-Lindberg, E., Santini, V., Finelli, C., Giagounidis, A., et al. (2009). Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: A randomised, open-label, phase III study. *Lancet Oncol.* 10, 223–232. doi: 10.1016/S1470-2045(09)70003-8

Ferrari, K. J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., et al. (2014). Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. *Mol. Cell*. 53, 49–62. doi: 10.1016/j.molcel.2013.10.030

Figueroa, M. E., Abdel-Wahab, O., Lu, C., Ward, P. S., Patel, J., Shih, A., et al. (2010). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* 18, 553–567. doi: 10.1016/j.ccr.2010.11.015

Filippakopoulos, P., Picaud, S., Mangos, M., Keates, T., Lambert, J. P., Barsyte-Lovejoy, D., et al. (2012). Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell*. 149, 214–231. doi: 10.1016/j.cell.2012.02.013

Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W. B., Fedorov, O., et al. (2010). Selective inhibition of BET bromodomains. *Nature*. 468, 1067–1073. doi: 10.1038/nature09504

Flavahan, W. A., Gaskell, E., and Bernstein, B. E. (2017). Epigenetic plasticity and the hallmarks of cancer. *Science*. 357, p.eaal2380. doi: 10.1126/science.aal2380

Flesner, B. K., Kumar, S. R., and Bryan, J. N. (2014). 6-Thioguanine and zebularine down-regulate DNMT1 and globally demethylate canine malignant lymphoid cells. *BMC Vet. Res.* 10, 290. doi: 10.1186/s12917-014-0290-8

Fodale, V., Pierobon, M., Liotta, L., and Petricoin, E. (2011). Mechanism of cell adaptation: When and how do cancer cells develop chemoresistance? *Cancer J.* 17, 89–95. doi: 10.1097/PPO.0b013e318212dd3d

Forero-Torres, A., Rosen, S., Smith, D. C., Lesser, G., Peguero, J., Gupta, S., et al. (2017). Preliminary results from an ongoing phase 1/2 Study of INCB057643, a bromodomain and extraterminal (BET) protein inhibitor, in patients (pts) with advanced malignancies. *Blood* 130, 4048.

Fournel, M., Bonfils, C., Hou, Y., Yan, P. T., Trachy-Bourget, M.-C., Kalita, A., et al. (2008). MGCD0103, a novel isotype-selective histone deacetylase inhibitor, has

broad spectrum antitumor activity in vitro and in vivo. *Mol. Cancer Ther.* 7, 759–768. doi: 10.1158/1535-7163.MCT-07-2026

Fujimoto, A., Totoki, Y., Abe, T., Boroevich, K. A., Hosoda, F., Nguyen, H. H., et al. (2012). Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat. Genet.* 44, 760–764. doi: 10.1038/ng.2291

Galli, M., Salmoiraghi, S., Golay, J., Gozzini, A., Crippa, C., Pescosta, N., et al. (2010). A phase II multiple dose clinical trial of histone deacetylase inhibitor ITF2357 in patients with relapsed or progressive multiple myeloma. *Ann. Hematol.* 89, 185–190. doi: 10.1007/s00277-009-0793-8

Garrido Castro, P., van Roon, E. H. J., Pinhanços, S. S., Trentin, L., Schneider, P., Kerstjens, M., et al. (2018). The HDAC inhibitor panobinostat (LBH589) exerts in vivo anti-leukaemic activity against MLL-rearranged acute lymphoblastic leukaemia and involves the RNF20/RNF40/WAC-H2B ubiquitination axis. *Leukemia* 32, 323–331. doi: 10.1038/leu.2017.216

Ghoshal, K., Datta, J., Majumder, S., Bai, S., Kutay, H., Motiwala, T., et al. (2005). 5-Aza-deoxycytidine induces selective degradation of DNA methyltransferase 1 by a proteasomal pathway that requires the KEN box, bromo-adjacent homology domain, and nuclear localization signal. *Mol. Cell. Biol.* 25, 4727–4741. doi: 10.1128/MCB.25.11.4727-4741.2005

Gloghini, A., Buglio, D., Khaskhely, N. M., Georgakis, G., Orlowski, R. Z., Neelapu, S. S., et al. (2009). Expression of histone deacetylases in lymphoma: implication for the development of selective inhibitors. *Br. J. Haematol.* 147, 515–525. doi: 10.1111/j.1365-2141.2009.07887.x

Goodyear, O., Agathanggelou, A., Novitzky-Basso, I., Siddique, S., McSkeane, T., Ryan, G., et al. (2010). Induction of a CD8+ T-cell response to the MAGE cancer testis antigen by combined treatment with azacitidine and sodium valproate in patients with acute myeloid leukemia and myelodysplasia. *Blood*. 116, 1908–1918. doi: 10.1182/blood-2009-11-249474

Goverdhan, A., Lee, H.-H., Havranek, O., Davis, R. E., and Hung, M.-C., (2017). "Abstract 13: PRMT1 as a therapeutic target in diffuse large B-cell lymphoma," in *Epigenetics* (American Association for Cancer Research), 13–13. doi: 10.1158/1557-3265.HEMMAL17-13

Greenblatt, S. M., Liu, F., and Nimer, S. D. (2016). Arginine methyltransferases in normal and malignant hematopoiesis. *Exp. Hematol.* 44, 435–441. doi: 10.1016/j.exphem.2016.03.009

Groudine, M., Eisenman, R., and Weintraub, H. (1981). Chromatin structure of endogenous retroviral genes and activation by an inhibitor of DNA methylation. *Nature*. 292, 311–317. doi: 10.1038/292311a0

Gui, Y., Guo, G., Huang, Y., Hu, X., Tang, A., Gao, S., et al. (2011). Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat. Genet.* 44, 17–19. doi: 10.1038/ng.907

Gulati, N., Béguelin, W., and Giulino-Roth, L. (2018). Enhancer of zeste homolog 2 (EZH2) inhibitors. *Leuk. Lymphoma* 59, 1574–1585. doi: 10.1080/10428194.2018.1430795

Hadjikyriacou, A., Yang, Y., Espejo, A., Bedford, M. T., and Clarke, S. G. (2015). Unique features of human protein arginine methyltransferase 9 (PRMT9) and its substrate RNA splicing factor SF3B2. *J. Biol. Chem.* 290, 16723–16743. doi: 10.1074/jbc.M115.659433

Haferlach, C., Dicker, F., Schnittger, S., Kern, W., and Haferlach, T. (2007). Comprehensive genetic characterization of CLL: a study on 506 cases analysed with chromosome banding analysis, interphase FISH, IgV(H) status and immunophenotyping. *Leukemia* 21, 2442–2451. doi: 10.1038/sj.leu.2404935

Han, M., Jia, L., Lv, W., Wang, L., and Cui, W. (2019). Epigenetic enzyme mutations: role in tumorigenesis and molecular inhibitors. *Front. Oncol.* 9, 1–9. doi: 10.3389/fonc.2019.00194

Hassler, M. R., Schiefer, A.-I., and Egger, G. (2013). Combating the epigenome: epigenetic drugs against non-Hodgkin's lymphoma. *Epigenomics* 5, 397–415. doi: 10.2217/epi.13.39

Havas, A. P., Rodrigues, K. B., Bhakta, A., Demirjian, J. A., Hahn, S., Tran, J., et al. (2016). Belinostat and vincristine demonstrate mutually synergistic cytotoxicity associated with mitotic arrest and inhibition of polyploidy in a preclinical model of aggressive diffuse large B cell lymphoma. *Cancer Biol. Ther.* 17, 1240–1252. doi: 10.1080/15384047.2016.1250046

Heideman, M. R., Wilting, R. H., Yanover, E., Velds, A., de Jong, J., Kerkhoven, R. M., et al. (2013). Dosage-dependent tumor suppression by histone deacetylases 1 and 2 through regulation of c-Myc collaborating genes and p53 function. *Blood* 121, 2038–2050. doi: 10.1182/blood-2012-08-450916

Helming, K. C., Wang, X., Wilson, B. G., Vazquez, F., Haswell, J. R., Manchester, H. E., et al. (2014). ARID1B is a specific vulnerability in ARID1A-mutant cancers. *Nat. Med.* 20, 251–254. doi: 10.1038/nm.3480

Hermann, A., Goyal, R., and Jeltsch, A. (2004). The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J. Biol. Chem.* 279, 48350–48359. doi: 10.1074/jbc.M403427200

Herviou, L., Kassambara, A., Boireau, S., Robert, N., Requirand, G., Vincent, L., et al. (2016). Targeting EZH2 in multiple myeloma could be promising for a subgroup of MM patients in combination with IMiDs. *Blood*. 128, 311.

Hodis, E., Watson, I. R., Kryukov, G. V., Arold, S. T., Imielinski, M., Theurillat, J. P., et al. (2012). A landscape of driver mutations in melanoma. *Cell*. 150, 251–266. doi: 10.1016/j.cell.2012.06.024

Hogarth, L. A., Redfern, C. P. F., Teodoridis, J. M., Hall, A. G., Anderson, H., Case, M. C., et al. (2008). The effect of thiopurine drugs on DNA methylation in relation to TPMT expression. *Biochem. Pharmacol.* 76, 1024–1035. doi: 10.1016/j.bcp.2008.07.026

Hogg, S. J., Newbold, A., Vervoort, S. J., Cluse, L. A., Martin, B. P., Gregory, G. P., et al. (2016). BET inhibition induces apoptosis in aggressive B-cell lymphoma via epigenetic regulation of BCL-2 family members. *Mol. Cancer Ther.* 15, 2030–2041. doi: 10.1158/1535-7163.MCT-15-0924

Holkova, B., Yazbeck, V., Kmieciak, M., Bose, P., Ma, S., Kimball, A., et al. (2017). A phase 1 study of bortezomib and romidepsin in patients with chronic lymphocytic leukemia/small lymphocytic lymphoma, indolent B-cell lymphoma, peripheral T-cell lymphoma, or cutaneous T-cell lymphoma. *Leuk. Lymphoma* 58, 1349–1357. doi: 10.1080/10428194.2016.1276287

Honma, D., Kanno, O., Watanabe, J., Kinoshita, J., Hirasawa, M., Nosaka, E., et al. (2017). Novel orally bioavailable EZH1/2 dual inhibitors with greater antitumor efficacy than an EZH2 selective inhibitor. *Cancer Sci.* 108, 2069–2078. doi: 10.1111/cas.13326

Horsman, D. E., Gascoyne, R. D., Coupland, R. W., Coldman, A. J., and Adomat, S. A. (1995). Comparison of cytogenetic analysis, Southern analysis, and polymerase chain reaction for the detection of t(14; 18) in follicular lymphoma. *Am. J. Clin. Pathol.* 103, 472–478. doi: 10.1093/ajcp/103.4.472

Hunter, Z. R., Xu, L., Yang, G., Zhou, Y., Liu, X., Cao, Y., et al. (2014). The genomic landscape of Waldenström macroglobulinemia is characterized by highly recurring MYD88 and WHIM-like CXCR4 mutations, and small somatic deletions associated with B-cell lymphomagenesis. *Blood* 123, 1637–1646. doi: 10.1182/blood-2013-09-525808

Issa, J. P. J., Roboz, G., Rizzieri, D., Jabbour, E., Stock, W., O'Connell, C., et al. (2015). Safety and tolerability of guadecitabine (SGI-110) in patients with myelodysplastic syndrome and acute myeloid leukaemia: A multicentre, randomised, dose-escalation phase 1 study. *Lancet Oncol.* 16, 1099–1110. doi: 10.1016/S1470-2045(15)00038-8

Issa, M. E., Takhsha, F. S., Chirumamilla, C. S., Perez-Novo, C., Vanden Berghe, W., and Cuendet, M. (2017). Epigenetic strategies to reverse drug resistance in heterogeneous multiple myeloma. *Clin. Epigenetics*. 9, 17. doi: 10.1186/s13148-017-0319-5

Italiano, A., Soria, J., Toulmonde, M., Michot, J., Lucchesi, C., Varga, A., et al. (2018). Articles Tazemetostat , an EZH2 inhibitor , in relapsed or refractory B-cell non-Hodgkin lymphoma and advanced solid tumours : a first-in-human, open-label , phase 1 study 2045, *Lancet Oncol.* 1–11. doi: 10.1016/S1470-2045(18)30145-1

Jelinic, P., Mueller, J. J., Olvera, N., Dao, F., Scott, S. N., Shah, R., et al. (2014). Recurrent SMARCA4 mutations in small cell carcinoma of the ovary. *Nat. Genet.* 46, 424–426. doi: 10.1038/ng.2922

Jin, Y., Elalaf, H., Watanabe, M., Tamaki, S., Hineno, S., Matsunaga, K., et al. (2015). Mutant IDH1 dysregulates the differentiation of mesenchymal stem cells in association with gene-specific histone modifications to cartilage- and bone-related genes. *PLoS One* 10, 1–15. doi: 10.1371/journal.pone.0131998

Johnson-Farley, N., Veliz, J., Bhagavathi, S., and Bertino, J. R. (2015). ABT-199, a BH3 mimetic that specifically targets Bcl-2, enhances the antitumor activity of chemotherapy, bortezomib and JQ1 in ""double hit"" lymphoma cells. *Leuk. Lymphoma* 56, 2146–2152. doi: 10.3109/10428194.2014.981172

Jones, P. A., Issa, J. P. J., and Baylin, S. (2016). Targeting the cancer epigenome for therapy. *Nat. Rev. Genet.* 17, 630–641. doi: 10.1038/nrg.2016.93

Jones, P. A., and Taylor, S. M. (1980). Cellular differentiation, cytidine analogs and DNA methylation. *Cell*. 20, 85–93. doi: 10.1016/0092-8674(80)90237-8

Juttermann, R., Li, E., and Jaenisch, R. (2006). Toxicity of 5-aza-2'2′-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation. *Proc. Natl. Acad. Sci.* 91, 11797–11801. doi: 10.1073/pnas.91.25.11797

Kadam, S., and Emerson, B. M. (2003). Transcriptional specificity of human SWI/SNF BRG1 and BRM chromatin remodeling complexes. *Mol. Cell.* 1, 177–189. doi: 10.1016/S1097-2765(03)00034-0

Kalac, M., Scotto, L., Marchi, E., Amengual, J., Seshan, V. E., Bhagat, G., et al. (2011). HDAC inhibitors and decitabine are highly synergistic and associated with unique gene-expression and epigenetic profiles in models of DLBCL. *Blood* 118, 5506–5516. doi: 10.1182/blood-2011-02-336891

Kandoth, C., Schultz, N., Cherniack, A. D., Akbani, R., Liu, Y., Shen, H., et al. (2013). Integrated genomic characterization of endometrial carcinoma. *Nature*. 497, 67–73. doi: 10.1038/nature12113

Kang-Decker, N., Tong, C., Boussouar, F., Baker, D. J., Xu, W., Leontovich, A. A., et al. (2004). Loss of CBP causes T cell lymphomagenesis in synergy with p27 Kip1 insufficiency. *Cancer Cell*. 5, 177–189. doi: 10.1016/S1535-6108(04)00022-4

Kang, M. R., Kim, M. S., Oh, J. E., Kim, Y. R., Song, S. Y., Seo, S., et al. (2009). Mutational analysis of IDH1 codon 132 in glioblastomas and other common cancers. *Int. J. Cancer* 125, 353–355. doi: 10.1002/ijc.24379

Karpf, A. R. (2004). Limited gene activation in tumor and normal epithelial cells treated with the DNA methyltransferase Inhibitor 5-aza-2'2′-deoxycytidine. *Mol. Pharmacol.* 65, 18–27. doi: 10.1124/mol.65.1.18

Kewitz, S., Bernig, T., and Staege, M. S. (2012). Histone deacetylase inhibition restores cisplatin sensitivity of Hodgkin's lymphoma cells. *Leuk. Res.* 36, 773–778. doi: 10.1016/j.leukres.2012.02.021

Khavari, P. A., Peterson, C. L., Tamkun, J. W., Mendel, D. B., and Crabtree, G. R. (1993). BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription. *Nature*. 366, 170–174. doi: 10.1038/366170a0

Kim, E., Ten Hacken, E., Clarke, A., Keating, M. J., Wierda, W. G., Ferrajoli, A., et al. (2017). The clinical BET inhibitor, GS-5829, is active against chronic lymphocytic leukemia as single agent and in combination with B-cell receptor signaling inhibitors. *Blood* 130, 3844.

Kinders R. J., Wang, L., Kummar, K. S., Balasubramanian, P., Zhu, W., Parchment, R. E., et al. (2011). Investigation of 5-fluorodeoxycytidine with tetrahydrouracil as a demethylation regimen in solid tumors. *Mol. Cancer Ther.* 10, A106. doi: 10.1158/1535-7163.TARG-11-A106

King, B., Trimarchi, T., Reavie, L., Xu, L., Mullenders, J., Ntziachristos, P., et al. (2013). The ubiquitin ligase FBXW7 modulates leukemia-initiating cell activity by regulating MYC stability. *Cell* 153, 1552–1566. doi: 10.1016/j.cell.2013.05.041

Kirschbaum, M., Frankel, P., Popplewell, L., Zain, J., Delioukina, M., Pullarkat, V., et al. (2011). Phase II study of vorinostat for treatment of relapsed or refractory indolent non-Hodgkin's lymphoma and mantle cell lymphoma. *J. Clin. Oncol.* 29, 1198–1203. doi: 10.1200/JCO.2010.32.1398

Klein, J. M., Henke, A., Sauer, M., Bessler, M., Reiners, K. S., Engert, A., et al. (2013). The histone deacetylase inhibitor LBH589 (panobinostat) modulates the crosstalk of lymphocytes with Hodgkin lymphoma cell lines. *PLoS One* 8, e79502. doi: 10.1371/journal.pone.0079502

Knipstein, J., and Gore, L. (2011). Entinostat for treatment of solid tumors and hematologic malignancies. *Expert Opin. Investig. Drugs* 20, 1455–1467. doi: 10.1517/13543784.2011.613822

Knutson, S. K., Kawano, S., Minoshima, Y., Warholic, N. M., Huang, K.-C., Xiao, Y., et al. (2014). Selective inhibition of EZH2 by EPZ-6438 leads to potent antitumor activity in EZH2-mutant non-Hodgkin lymphoma. *Mol. Cancer Ther.* 13, 842–854. doi: 10.1158/1535-7163.MCT-13-0773

Konze, K. D., Ma, A., Li, F., Barsyte-Lovejoy, D., Parton, T., MacNevin, C. J., et al. (2013). An orally bioavailable chemical probe of the lysine methyltransferases EZH2 and EZH1. *ACS Chem. Biol.* 8, 1324–1334. doi: 10.1021/cb400133j

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*. 128, 693–705. doi: 10.1016/j.cell.2007.02.005

Kretzner, L., Scuto, A., Dino, P. M., Kowolik, C. M., Wu, J., Ventura, P., et al. (2011). Combining histone deacetylase inhibitor vorinostat with aurora kinase inhibitors enhances lymphoma cell killing with repression of c-Myc, hTERT, and microRNA levels. *Cancer Res.* 71, 3912–3920. doi: 10.1158/0008-5472.CAN-10-2259

Kummar, S., Gutierrez, M., Gardner, E. R., Donovan, E., Hwang, K., Chung, E. J., et al. (2007). Phase I trial of MS-275, a histone deacetylase inhibitor, administered weekly in refractory solid tumors and lymphoid malignancies. *Clin. Cancer Res.* 13, 5411–5417. doi: 10.1158/1078-0432.CCR-07-0791

Kung, A. L., Rebel, V. I., Bronson, R. T., Ch'ng, L. E., Sieff, C. A., Livingston, D. M., et al. (2000). Gene dose-dependent control of hematopoiesis and hematologic tumor suppression by CBP. *Genes Dev.* 14, 172–177.

Lai, A. Y., Mav, D., Shah, R., Grimm, S. A., Phadke, D., Hatzi, K., et al. (2013). DNA methylation profiling in human B cells reveals immune regulatory elements and epigenetic plasticity at Alu elements during B-cell activation. *Genome Res.* 12, 2030–2041. doi: 10.1101/gr.155473.113

Landau, D. A., Tausch, E., Taylor-Weiner, A. N., Stewart, C., Reiter, J. G., Bahlo, J., et al. (2015). Mutations driving CLL and their evolution in progression and relapse. *Nature* 526, 525–530. doi: 10.1038/nature15395

Lane, A. A., and Chabner, B. A. (2009). Histone deacetylase inhibitors in cancer therapy. *J. Clin. Oncol.* 27, 5459–5468. doi: 10.1200/JCO.2009.22.1291

Lee, S. H., Kim, J.-S., Zheng, S., Huse, J. T., Bae, J. S., Lee, J. W., et al. (2017). ARID1B alterations identify aggressive tumors in neuroblastoma. *Oncotarget*. 8, 45943–45950. doi: 10.18632/oncotarget.17500

Lemoine, M., and Younes, A. (2010). Histone deacetylase inhibitors in the treatment of lymphoma. *Discov. Med.* 10, 462–470.

Li, L. H., Olin, E. J., Buskirk, H. H., and Reineke, L. M. (1970). Cytotoxicity and mode of action of 5-azacytidine on L1210 leukemia. *Cancer Res.* 30, 2760–2769.

Li, Y., Nagai, H., Ohno, T., Yuge, M., Hatano, S., Ito, E., et al. (2002). Aberrant DNA methylation of p57 KIP2 gene in the promoter region in lymphoid malignancies of B-cell phenotype. *Blood*. 100, 2572–2577. doi: 10.1182/blood-2001-11-0026

Lohr, J. G., Stojanov, P., Lawrence, M. S., Auclair, D., Chapuy, B., Sougnez, C., et al. (2012). Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc. Natl. Acad. Sci. U. S. A.* 109, 3879–3884. doi: 10.1073/pnas.1121343109

Losman, J.-A., Looper, R., Koivunen, P., Lee, S., Schneider, R. K., McMahon, C., et al. (2013). (R)-2-hydroxyglutarate is sufficient to promote leukemogenesis and its effects are reversible. *Science (80-.).* 339, 1621–1625. doi: 10.1126/science.1231677

Lovén, J., Hoke, H. A., Lin, C. Y., Lau, A., Orlando, D. A., Vakoc, C. R., et al. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*. 153, 320–334. doi: 10.1016/j.cell.2013.03.036

Lu, X., Fernando, T. M., Lossos, C., Yusufova, N., Liu, F., Fontán, L., et al. (2018). PRMT5 interacts with the BCL6 oncoprotein and is required for germinal center formation and lymphoma cell survival. *Blood* 132, 2026–2039. doi: 10.1182/blood-2018-02-831438

Lübbert, M., Suciu, S., Hagemeijer, A., Rüter, B., Platzbecker, U., Giagounidis, A., et al. (2016). Decitabine improves progression-free survival in older high-risk MDS patients with multiple autosomal monosomies: results of a subgroup analysis of the randomized phase III study 06011 of the EORTC Leukemia cooperative group and german MDS study group. *Ann. Hematol.* 95, 191–199. doi: 10.1007/s00277-015-2547-0

Lue, J. K., Prabhu, S. A., Liu, Y., Gonzalez, Y., Verma, A., Mundi, P. S., et al. (2019). Precision targeting with EZH2 and HDAC inhibitors in epigenetically dysregulated lymphomas. *Clin. Cancer Res.* 25, 5271–5283. doi: 10.1158/1078-0432.CCR-18-3989

Lue, J. K., Prabhu, S. A., Liu, Y., Verma, A., Elemento, O., and Amengual, J. E. (2017). Dual inhibition of EZH2 and HDAC is synergistic in EZH2 Dysregulated Lymphomas. *Hematol. Oncol.* 35, 254–255. doi: 10.1002/hon.2438_120

Mai, A., Massa, S., Rotili, D., Cerbara, I., Valente, S., Pezzi, R., et al. (2005). Histone deacetylation in epigenetics: an attractive target for anticancer therapy. *Med. Res. Rev.* 25, 261–309. doi: 10.1002/med.20024

Maly, J. J., Christian, B. A., Zhu, X., Wei, L., Sexton, J. L., Jaglowski, S. M., et al. (2017). A phase I/II trial of panobinostat in combination with lenalidomide in patients with relapsed or refractory Hodgkin lymphoma. *Clin. Lymphoma Myeloma Leuk.* 17, 347–353. doi: 10.1016/j.clml.2017.05.008

Mann, B. S., Johnson, J. R., Cohen, M. H., Justice, R., and Pazdur, R. (2007). FDA approval summary: Vorinostat for treatment of advanced primary cutaneous T-cell lymphoma. *Oncologist* 12, 1247–1252. doi: 10.1634/theoncologist.12-10-1247

Marjon, K., Cameron, M. J., Quang, P., Clasquin, M. F., Mandley, E., Kunii, K., et al. (2016). MTAP deletions in cancer create vulnerability to targeting of the MAT2A/PRMT5/RIOK1 axis. *Cell Rep.* 15, 574–587. doi: 10.1016/j.celrep.2016.03.043

Marquard, L., Poulsen, C. B., Gjerdrum, L. M., de Nully Brown, P., Christensen, I. J., Jensen, P. B., et al. (2009). Histone deacetylase 1, 2, 6 and acetylated

histone H4 in B- and T-cell lymphomas. *Histopathology* 54, 688–698. doi: 10.1111/j.1365-2559.2009.03290.x

Maruyama, D., Tobinai, K., Makita, S., Ishida, T., Kusumoto, S., Ishitsuka, K., et al. (2017). First-in-human study of the EZH1/2 dual inhibitor DS-3201b in patients with relapsed or refractory non-Hodgkin lymphomas — Preliminary results. *Blood*. 130, 4070.

McCabe, M. T., Graves, A. P., Ganji, G., Diaz, E., Halsey, W. S., Jiang, Y., et al. (2012a). Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). *Proc. Natl. Acad. Sci. U. S. A*. 109, 2989–2994. doi: 10.1073/pnas.1116418109

McCabe, M. T., Ott, H. M., Ganji, G., Korenchuk, S., Thompson, C., Van Aller, G. S., et al. (2012b). EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature*. 492, 108–112. doi: 10.1038/nature11606

McClure, J. J., Li, X., and Chou, C. J. (2018). Advances and challenges of HDAC inhibitors in cancer therapeutics. *Adv. Cancer Res.* 183–211. doi: 10.1016/bs.acr.2018.02.006

McGarvey, K. M., Fahrner, J. A., Greene, E., Martens, J., Jenuwein, T., and Baylin, S. B. (2006). Silenced tumor suppressor genes reactivated by DNA demethylation do not return to a fully euchromatic chromatin state. *Cancer Res.* 66, 3541–3549. doi: 10.1158/0008-5472.CAN-05-2481

Meissner, A. (2010). Epigenetic modifications in pluripotent and differentiated cells. *Nat. Biotechnol.* 28, 1079–1088. doi: 10.1038/nbt.1684

Migliori, V., Phalke, S., Bezzi, M., and Guccione, E. (2010). Arginine/lysine-methyl/methyl switches: biochemical role of histone arginine methylation in transcriptional regulation. *Epigenomics* 2, 119–137. doi: 10.2217/epi.09.39

Minucci, S., and Pelicci, P. G. (2006). Histone deacetylase inhibitors and the promise of epigenetic (and more) treatments for cancer. *Nat. Rev. Cancer* 6, 38–51. doi: 10.1038/nrc1779

Mirguet, O., Gosmini, R., Toum, J., Clément, C. A., Barnathan, M., Brusq, J.-M., et al. (2013). Discovery of epigenetic regulator I-BET762: lead optimization to afford a clinical candidate inhibitor of the BET bromodomains. *J. Med. Chem.* 56, 7501–7515. doi: 10.1021/jm401088k

Montalban-Bravo, G., and DiNardo, C. D. (2018). The role of IDH mutations in acute myeloid leukemia. *Futur. Oncol.* 14, 979–993. doi: 10.2217/fon-2017-0523

Morin, R. D., Johnson, N. A., Severson, T. M., Mungall, A. J., An, J., Goya, R., et al. (2010). Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat. Genet.* 42, 181–185. doi: 10.1038/ng.518

Morin, R. D., Mendez-Lago, M., Mungall, A. J., Goya, R., Mungall, K. L., Corbett, R. D., et al. (2011). Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* 476, 298–303. doi: 10.1038/nature10351

Moros, A., Rodríguez, V., Saborit-Villarroya, I., Montraveta, A., Balsas, P., Sandy, P., et al. (2014). Synergistic antitumor activity of lenalidomide with the BET bromodomain inhibitor CPI203 in bortezomib-resistant mantle cell lymphoma. *Leukemia* 28, 2049–2059. doi: 10.1038/leu.2014.106

Morschhauser, F., Terriou, L., Coiffier, B., Bachy, E., Varga, A., Kloos, I., et al. (2015). Phase 1 study of the oral histone deacetylase inhibitor abexinostat in patients with Hodgkin lymphoma, non-Hodgkin lymphoma, or chronic lymphocytic leukaemia. *Invest. New Drugs* 33, 423–431. doi: 10.1007/s10637-015-0206-x

Mullighan, C. G., Zhang, J., Kasper, L. H., Lerach, S., Payne-Turner, D., Phillips, L. A., et al. (2011). CREBBP mutations in relapsed acute lymphoblastic leukaemia. *Nature*. 471, 235–239. doi: 10.1038/nature09727

Munshi, P. N., Lubin, M., and Bertino, J. R. (2014). 6-Thioguanine: a drug with unrealized potential for cancer therapy. *Oncologist*. 19, 760–765. doi: 10.1634/theoncologist.2014-0178

Newman, E. M, Morgan, R. J., Kummar, S., Beumer, J. H., Blanchard, M. S., Ruel, C., et al. (2015). A phase I, pharmacokinetic, and pharmacodynamic evaluation of the DNA methyltransferase inhibitor 5-fluoro-2'2'-deoxycytidine, administered with tetrahydrouridine. *Cancer Chemother. Pharmacol.* 75, 537–547.

Normant, E., Cummings, R., Bellon, S., Bailey, C., Albrecht, B., Hewitt, M., et al. (2012). "Abstract LB-237: in vitro and in vivo characterization of CPI-267203, a potent inhibitor of bromodomain-containing proteins," in *Experimental and molecular therapeutics* (American Association for Cancer Research), LB–237-LB-237. doi: 10.1158/1538-7445.AM2012-LB-237

Ogura, M., Ando, K., Suzuki, T., Ishizawa, K., Oh, S. Y., Itoh, K., et al. (2014). A multicentre phase II study of vorinostat in patients with relapsed or refractory indolent B-cell non-Hodgkin lymphoma and mantle cell lymphoma. *Br. J. Haematol.* 165, 768–776. doi: 10.1111/bjh.12819

Oki, Y., Buglio, D., Fanale, M., Fayad, L., Copeland, A., Romaguera, J., et al. (2013). Phase I study of panobinostat plus everolimus in patients with relapsed or refractory lymphoma. *Clin. Cancer Res.* 19, 6882–6890. doi: 10.1158/1078-0432.CCR-13-1906

Oki, Y., Kelly, K. R., Flinn, I., Patel, M. R., Gharavi, R., Ma, A., et al. (2017). CUDC-907 in relapsed/refractory diffuse large B-cell lymphoma, including patients with MYC-alterations: Results from an expanded phase I trial. *Haematologica* 102, 1923–1930. doi: 10.3324/haematol.2017.172882

Okosun, J., Bödör, C., Wang, J., Araf, S., Yang, C.-Y., Pan, C., et al. (2014). Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma. *Nat Genet*. 46, 176–181. doi: 10.1038/ng.2856

Olsen, E. A., Kim, Y. H., Kuzel, T. M., Pacheco, T. R., Foss, F. M., Parker, S., et al. (2007). Phase IIB multicenter trial of vorinostat in patients with persistent, progressive, or treatment refractory cutaneous T-cell lymphoma. *J. Clin. Oncol.* 25, 3109–3115. doi: 10.1200/JCO.2006.10.2434

Oronsky, B., Paulmurugan, R., Foygel, K., Scicinski, J., Knox, S. J., Peehl, D., et al. (2017). RRx-001: a systemically non-toxic M2-to-M1 macrophage stimulating and prosensitizing agent in phase II clinical trials. *Expert Opin. Investig. Drugs*. 26, 109–119. doi: 10.1080/13543784.2017.1268600

Ozer, H. G., El-Gamal, D., Powell, B., Hing, Z. A., Blachly, J. S., Harrington, B., et al. (2018). BRD4 profiling identifies critical chronic lymphocytic leukemia oncogenic circuits and reveals sensitivity to PLX51107, a novel structurally distinct BET inhibitor. *Cancer Discov.* 8, 458–477. doi: 10.1158/2159-8290.CD-17-0902

Padmanabhan, B., Mathur, S., Manjula, R., and Tripathi, S. (2016). Bromodomain and extra-terminal (BET) family proteins: new therapeutic targets in major diseases. *J. Biosci.* 41, 295–311. doi: 10.1007/s12038-016-9600-6

Parry, M., Rose-Zerilli, M. J. J., Ljungström, V., Gibson, J., Wang, J., Walewska, R., et al. (2015). Genetics and prognostication in splenic marginal zone lymphoma: revelations from deep sequencing. *Clin. Cancer Res.* 21, 4174–4183. doi: 10.1158/1078-0432.CCR-14-2759

Pasqualucci, L., Dominguez-Sola, D., Chiarenza, A., Fabbri, G., Grunn, A., Trifonov, V., et al. (2011a). Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature*. 471, 189–195. doi: 10.1038/nature09730

Pasqualucci, L., Trifonov, V., Fabbri, G., Ma, J., Rossi, D., Chiarenza, A., et al. (2011b). Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat. Genet.* 43, 830–837. doi: 10.1038/ng.892

Pazin, M. J., and Kadonaga, J. T. (1997). SWI2/SNF2 and related proteins: ATP-driven motors that disrupt protein- –DNA interactions? *Cell*. 88, 737–740. doi: 10.1016/S0092-8674(00)81918-2

Persky, D. O., Li, H., Rimsza, L. M., Barr, P. M., Popplewell, L. L., Bane, C. L., et al. (2018). A phase I/II trial of vorinostat (SAHA) in combination with rituximab-CHOP in patients with newly diagnosed advanced stage diffuse large B-cell lymphoma (DLBCL): SWOG S0806. *Am. J. Hematol.* 93, 486–493. doi: 10.1002/ajh.25010

Popovic, R., Shah, M. Y., and Licht, J. D. (2013). Epigenetic therapy of hematological malignancies: Where are we now? *Ther. Adv. Hematol.* 4, 81–91. doi: 10.1177/2040620712466864

Qi, W., Chan, H., Teng, L., Li, L., Chuai, S., Zhang, R., et al. (2012). Selective inhibition of Ezh2 by a small molecule inhibitor blocks tumor cells proliferation. *Proc. Natl. Acad. Sci.* 109, 21360–21365. doi: 10.1073/pnas.1210371110

Qin, C., Hu, Y., Zhou, B., Fernandez-Salas, E., Yang, C.-Y., Liu, L., et al. (2018). Discovery of QCA570 as an exceptionally potent and efficacious proteolysis targeting chimera (PROTAC) degrader of the bromodomain and extra-terminal (BET) proteins capable of inducing complete and durable tumor regression. *J. Med. Chem.* 61, 6685–6704. doi: 10.1021/acs.jmedchem.8b00506

Qiu, L., Burgess, A., Fairlie, D. P., Leonard, H., Parsons, P. G., and Gabrielli, B. G. (2000). Histone deacetylase inhibitors trigger a G2 checkpoint in normal cells that is defective in tumor cells. *Mol. Biol. Cell* 11, 2069–2083. doi: 10.1091/mbc.11.6.2069

Raina, K., Lu, J., Qian, Y., Altieri, M., Gordon, D., Rossi, A. M. K., et al.. (2016). PROTAC-induced BET protein degradation as a therapy for castration-resistant prostate cancer. *Proc. Natl. Acad. Sci. U. S. A.* 113, 7124–7129. doi: 10.1073/pnas.1521738113

Rebel, V. I., Kung, A. L., Tanner, E. A., Yang, H., Bronson, R. T., and Livingston, D. M. (2002). Distinct roles for CREB-binding protein and p300

in hematopoietic stem cell self-renewal. *Proc. Natl. Acad. Sci.* 99, 14789–14794. doi: 10.1073/pnas.232568499

Recasens-zorzo, C., Cardesa-salzmann, T., Petazzi, P., Ros-blanco, L., Esteve-arenys, A., Clot, G., et al. (2018). Pharmacological modulation of CXCR4 cooperates with BET bromodomain inhibition in diffuse large B-cell lymphoma. *Haematologica* 104, 778–788. doi: 10.3324/haematol.2017.180505

Rhyasen, G. W., Hattersley, M. M., Yao, Y., Dulak, A., Wang, W., Petteruti, P., et al. (2016). AZD5153: a novel bivalent BET bromodomain inhibitor highly active against hematologic malignancies. *Mol. Cancer Ther.* 15, 2563–2574. doi: 10.1158/1535-7163.MCT-16-0141

Ribrag, V., Kim, W. S., Bouabdallah, R., Lim, S. T., Coiffier, B., Illes, A., et al. (2017). Safety and efficacy of abexinostat, a pan-histone deacetylase inhibitor, in non-Hodgkin lymphoma and chronic lymphocytic leukemia: Results of a phase II study. *Haematologica* 102, 903–909. doi: 10.3324/haematol.2016.154377

Richardson, P. G., Hungria, V. T. M., Yoon, S. S., Beksac, M., Dimopoulos, M. A., Elghandour, A., et al. (2016). Panobinostat plus bortezomib and dexamethasone in previously treated multiple myeloma: Outcomes by prior treatment. *Blood*. 15, 2563–2574. doi: 10.1182/blood-2015-09-665018

Rinaldi, A., Mian, M., Chigrinova, E., Arcaini, L., Bhagat, G., Novak, U., et al. (2011). Genome-wide DNA profiling of marginal zone lymphomas identifies subtype-specific lesions with an impact on the clinical outcome. *Blood* 117, 1595–1604. doi: 10.1182/blood-2010-01-264275

Robaina, M. C., Mazzoccoli, L., Arruda, V. O., Reis, F. R., de, S., Apa, A. G., et al. (2015). Deregulation of DNMT1, DNMT3B and miR-29s in Burkitt lymphoma suggests novel contribution for disease pathogenesis. *Exp. Mol. Pathol.* 98, 200–207. doi: 10.1016/j.yexmp.2015.03.006

Roulland, S., Kasbekar, M., Young, R. M., Shaffer, A. L., Hodson, D. J., Xiao, W., et al. (2018). Genetics and pathogenesis of diffuse large B-cell lymphoma. *N. Engl. Med.* 378, 1396–1407. doi: 10.1056/NEJMoa1801445

Rozati, S., Cheng, P. F., Widmer, D. S., Fujii, K., Levesque, M. P., and Dummer, R. (2016). Romidepsin and azacitidine synergize in their epigenetic modulatory effects to induce apoptosis in CTCL. *Clin. Cancer Res.* 22, 2020–2031. doi: 10.1158/1078-0432.CCR-15-1435

Sachchidanand, R-S, L., Yan, S., Mutjaba, S., Liu, W. J., Zeng, L., et al. (2006). Target structure-based discovery of small molecules that block human p53 and CREB binding protein association. *Chem. Biol.* 13, 81–90. doi: 10.1016/j.chembiol.2005.10.014

Saha, S. K., Parachoniak, C. A., Ghanta, K. S., Fitamant, J., Ross, K. N., Najem, M. S., et al. (2014). Mutant IDH inhibits HNF-4α to block hepatocyte differentiation and promote biliary cancer. *Nature* 513, 110–152. doi: 10.1038/nature13441

Samuels,B.L.,Herndon,J.E.,Harmon,D.C.,Carey,R.,Aisner,J.,Corson,J.M.,etal.(1998). Dihydro-5-azacytidine and cisplatin in the treatment of malignant mesothelioma: a phase II study by the Cancer and Leukemia Group B. *Cancer.* 82, 1578–1584. doi: 10.1002/(SICI)1097-0142(19980415)82:8<1578::AID-CNCR21>3.0.CO;2-0

Santen, G. W. E., Aten, E., Sun, Y., Almomani, R., Gilissen, C., Nielsen, M., et al. (2012). Mutations in SWI/SNF chromatin remodeling complex gene ARID1B cause Coffin-Siris syndrome. *Nat. Genet*. 44, 379–380. doi: 10.1038/ng.2217

Santo, L., Hideshima, T., Kung, A. L., Tseng, J.-C., Tamang, D., Yang, M., et al. (2012). Preclinical activity, pharmacodynamic, and pharmacokinetic properties of a selective HDAC6 inhibitor, ACY-1215, in combination with bortezomib in multiple myeloma. *Blood* 119, 2579–2589. doi: 10.1182/blood-2011-10-387365

Santoro, F., Botrugno, O. A., Dal Zuffo, R., Pallavicini, I., Matthews, G. M., Cluse, L., et al. (2013). A dual role for Hdac1: oncosuppressor in tumorigenesis, oncogene in tumor maintenance. *Blood* 121, 3459–3468. doi: 10.1182/blood-2012-10-461988

Sausen, M., Leary, R. J., Jones, S., Wu, J., Reynolds, C. P., Liu, X., et al. (2013). Integrated genomic analyses identify ARID1A and ARID1B alterations in the childhood cancer neuroblastoma. *Nat. Genet*. 45, 12–17. doi: 10.1038/ng.2493

Schmitz, R., Young, R. M., Ceribelli, M., Jhavar, S., Xiao, W., Zhang, M., et al. (2012). Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. *Nature* 490, 116–120. doi: 10.1038/nature11378

Scholtysik, R., Kreuz, M., Hummel, M., Rosolowski, M., Szczepanowski, M., Klapper, W., et al. (2015). Characterization of genomic imbalances in diffuse large B-cell lymphoma by detailed SNP-chip analysis. *Int. J. Cancer* 136, 1033–1042. doi: 10.1002/ijc.29072

Schubert, H. L., Blumenthal, R. M., and Cheng, X. (2003). Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem. Sci.* 28, 329–335. doi: 10.1016/S0968-0004(03)00090-2

Schuh, A. C., Döhner, H., Pleyer, L., Seymour, J. F., Fenaux, P., and Dombret, H. (2017). Azacitidine in adult patients with acute myeloid leukemia. *Crit. Rev. Oncol. Hematol.* 116, 159–177. doi: 10.1016/j.critrevonc.2017.05.010

Scott, D. W., and Gascoyne, R. D. (2014). The tumour microenvironment in B cell lymphomas. *Nat. Rev. Cancer* 14, 517–534. doi: 10.1038/nrc3774

Scott, L. J. (2016). Azacitidine: A review in myelodysplastic syndromes and acute myeloid leukaemia. *Drugs.* 14, 517–534. doi: 10.1007/s40265-016-0585-0

Seifert, M., Scholtysik, R., and Küppers, R., (2013). "Origin and pathogenesis of B cell lymphomas," in *Methods Mol. Biol. (Clifton, N.J.)*, 1–25. doi: 10.1007/978-1-62703-269-8_1

Shahbazian, M. D., and Grunstein, M. (2007). Functions of site-specific histone acetylation and deacetylation. *Annu. Rev. Biochem.* 76, 75–100. doi: 10.1146/annurev.biochem.76.052705.162114

Shaknovich, R., Cerchietti, L., Tsikitas, L., Kormaksson, M., De, S., Figueroa, M. E., et al. (2011). DNA methyltransferase 1 and DNA methylation patterning contribute to germinal center B-cell differentiation. *Blood.* 118, 3559–3569. doi: 10.1182/blood-2011-06-357996

Shaknovich, R., Geng, H., Johnson, N. A., Tsikitas, L., Cerchietti, L., Greally, J. M., et al. (2010). DNA methylation signatures define molecular subtypes of diffuse large B-cell lymphoma. *Blood.* 116, e81–89. doi: 10.1182/blood-2010-05-285320

Sharma, P., and Allison, J. P. (2015). Immune checkpoint targeting in cancer therapy: Toward combination strategies with curative potential. *Cell.* 161, 205–214. doi: 10.1016/j.cell.2015.03.030

Sheppard, G. S., Wang, L., Fidanze, S. D., Hasvold, L. A., Liu, D., Pratt, J. K., et al. (2018). "Abstract 931: Discovery of ABBV-744, a first-in-class highly BDII-selective BET bromodomain inhibitor," in *Cancer chemistry* (American Association for Cancer Research), 931–931. doi: 10.1158/1538-7445.AM2018-931

Shi, J., Whyte, W. A., Zepeda-Mendoza, C. J., Milazzo, J. P., Shen, C., Roe, J. S., et al. (2013). Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. *Genes Dev.* 27, 2648–2662. doi: 10.1101/gad.232710.113

Shilatifard, A. (2008). Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. *Curr. Opin. Cell Biol.* 20, 341–348. doi: 10.1016/j.ceb.2008.03.019

Silverman, L. R., Demakos, E. P., Peterson, B. L., Kornblith, A. B., Holland, J. C., Odchimar-Reissig, R., et al. (2002). Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: A study of the cancer and leukemia group B. *J. Clin. Oncol.* 20, 2429–2440. doi: 10.1200/JCO.2002.04.117

Smith, E., Zhou, W., Shindiapina, P., Sif, S., Li, C., and Baiocchi, R. A. (2018). Recent advances in targeting protein arginine methyltransferase enzymes in cancer therapy. *Expert Opin. Ther. Targets* 22, 527–545. doi: 10.1080/14728222.2018.1474203

Smith, S. G., and Zhou, M.-M. (2016). The bromodomain: A new target in emerging epigenetic medicine. *ACS Chem. Biol.* 11, 598–608. doi: 10.1021/acschembio.5b00831

Srivastava, P., Paluch, B. E., Matsuzaki, J., James, S. R., Collamat-Lai, G., Taverna, P., et al. (2015). Immunomodulatory action of the DNA methyltransferase inhibitor SGI-110 in epithelial ovarian cancer cells and xenografts. *Epigenetics*. 10, 237–246. doi: 10.1080/15592294.2015.1017198

Stathis, A., Hotte, S. J., Chen, E. X., Hirte, H. W., Oza, A. M., Moretto, P., et al. (2011). Phase I study of decitabine in combination with vorinostat in patients with advanced solid tumors and non-Hodgkin's lymphomas. *Clin. Cancer Res.* 17, 1582–1590. doi: 10.1158/1078-0432.CCR-10-1893

Stein, E. M., Garcia-Manero, G., Rizzieri, D. A., Tibes, R., Berdeja, J. G., Savona, M. R., et al. (2018). The DOT1L inhibitor pinometostat reduces H3K79 methylation and has modest clinical activity in adult acute leukemia. *Blood*. 131, 2661–2669. doi: 10.1182/blood-2017-12-818948

Stewart, D. J., Issa, J. P., Kurzrock, R., Nunez, M. I., Jelinek, J., Hong, D., et al. (2009). Decitabine effect on tumor global DNA methylation and other parameters in a phase I trial in refractory solid tumors and lymphomas. *Clin. Cancer Res.* 15, 3881–3888. doi: 10.1158/1078-0432.CCR-08-2196

Straus, D. J., Hamlin, P. A., Matasar, M. J., Lia Palomba, M., Drullinsky, P. R., Zelenetz, A. D., et al.. (2015). Phase I/II trial of vorinostat with rituximab, cyclophosphamide, etoposide and prednisone as palliative treatment for elderly patients with relapsed or refractory diffuse large B-cell lymphoma not eligible for autologous stem cell transplantation. *Br. J. Haematol.* 168, 663–670. doi: 10.1111/bjh.13195

Streubel, B., Simonitsch-Klupp, I., Müllauer, L., Lamprecht, A., Huber, D., Siebert, R., et al. (2004). Variable frequencies of MALT lymphoma-associated genetic aberrations in MALT lymphomas of different sites. *Leukemia* 18, 1722–1726. doi: 10.1038/sj.leu.2403501

Sun, B., Fiskus, W., Qian, Y., Rajapakshe, K., Raina, K., Coleman, K. G., et al. (2018). BET protein proteolysis targeting chimera (PROTAC) exerts potent lethal activity against mantle cell lymphoma cells. *Leukemia* 32, 343–352. doi: 10.1038/leu.2017.207

Sun, B., Shah, B., Fiskus, W., Qi, J., Rajapakshe, K., Coarfa, C., et al. (2015). Synergistic activity of BET protein antagonist-based combinations in mantle cell lymphoma cells sensitive or resistant to ibrutinib. *Blood* 126, 1565–1574. doi: 10.1182/blood-2015-04-639542

Suraweera, A., O'Byrne, K. J., and Richard, D. J. (2018). Combination therapy with histone deacetylase inhibitors (HDACi) for the Treatment of cancer: achieving the full therapeutic potential of HDACi. *Front. Oncol.* 8, 92. doi: 10.3389/fonc.2018.00092

Swerdlow, S. H., Campo, E., Pileri, S. A., Harris, N. L., Stein, H., Siebert, R., et al. (2016). The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* 127, 2375–2390. doi: 10.1182/blood-2016-01-643569

Tanaka, M., Roberts, J. M., Seo, H.-S., Souza, A., Paulk, J., Scott, T. G., et al. (2016). Design and characterization of bivalent BET inhibitors. *Nat. Chem. Biol.* 12, 1089–1096. doi: 10.1038/nchembio.2209

Tang, J.-Y., Chang, C.-C., Lin, P.-C., and Chang, J.-G. (2012). Isocitrate dehydrogenase mutation hot spots in acute lymphoblastic leukemia and oral cancer. *Kaohsiung J. Med. Sci.* 28, 138–144. doi: 10.1016/j.kjms.2011.10.023

The Cancer Genome Atlas Research Network (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* 45, 1113–1120. doi: 10.1038/ng.2764

Thottassery, J. V., Sambandam, V., Allan, P. W., Maddry, J. A., Maxuitenko, Y. Y., Tiwari, K., et al. (2014). Novel DNA methyltransferase-1 (DNMT1) depleting anticancer nucleosides, 4′-thio-2′-deoxycytidine and 5-aza-4′-thio-2′-deoxycytidine. *Cancer Chemother. Pharmacol.* 74, 291–302. doi: 10.1007/s00280-014-2503-z

Tiacci, E., Trifonov, V., Schiavoni, G., Holmes, A., Kern, W., Martelli, M. P., et al. (2011). BRAF mutations in hairy-cell leukemia. *N. Engl. J. Med.* 364, 2305–2315. doi: 10.1056/NEJMoa1014209

Topalian, S. L., Drake, C. G., and Pardoll, D. M. (2015). Immune checkpoint blockade: a common denominator approach to cancer therapy. *Cancer Cell.* 27, 450–461. doi: 10.1016/j.ccell.2015.03.001

Toujani, S., Dessen, P., Ithzar, N., Danglot, G., Richon, C., Vassetzky, Y., et al. (2009). High resolution genome-wide analysis of chromosomal alterations in Burkitt's lymphoma. *PLoS One* 4, e7089. doi: 10.1371/journal.pone.0007089

Vakoc, C. R. (2015). "BET bromodomain inhibition as a therapeutic approach in hematological malignancies," in *Histone recognition*. Ed. Zhou, M. M. (Heidelberg, Germany: Springer), 185–198. doi: 10.1007/978-3-319-18102-8_9

Valdez, B. C., Li, Y., Murray, D., Liu, Y., Nieto, Y., Champlin, R. E., et al. (2018). Combination of a hypomethylating agent and inhibitors of PARP and HDAC traps PARP1 and DNMT1 to chromatin, acetylates DNA repair proteins, down-regulates NuRD and induces apoptosis in human leukemia and lymphoma cells. *Oncotarget* 9, 3908–3921. doi: 10.18632/oncotarget.23386

Van Damme, M., Crompot, E., Meuleman, N., Maerevoet, M., Mineur, P., Bron, D., et al. (2016). Characterization of TET and IDH gene expression in chronic lymphocytic leukemia: comparison with normal B cells and prognostic significance. *Clin. Epigenetics* 8, 1–11. doi: 10.1186/s13148-016-0298-y

Van Kemenade, F. J., Raaphorst, F. M., Blokzijl, T., Fieret, E., Hamer, K. M., Satijn, D. P. E., et al. (2001). Coexpression of BMI-1 and EZH2 polycomb-group proteins is associated with cycling cells and degree of malignancy in B-cell non-Hodgkin lymphoma. *Blood.* 97, 3896–3901. doi: 10.1182/blood.V97.12.3896

Vannini, A., Volpari, C., Filocamo, G., Casavola, E. C., Brunetti, M., Renzoni, D., et al. (2004). Crystal structure of a eukaryotic zinc-dependent histone deacetylase, human HDAC8, complexed with a hydroxamic acid inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* 101, 15064–15069. doi: 10.1073/pnas.0404603101

Vaswani, R. G., Gehling, V. S., Dakin, L. A., Cook, A. S., Nasveschuk, C. G., Duplessis, M., et al. (2016). Identification of (R)-N-((4-Methoxy-6-methyl-2-oxo-1,2-dihydropyridin-3-yl)methyl)-2-methyl-1-(1-(1-(2,2,2-trifluoroethyl) piperidin-4-yl)ethyl)-1H-indole-3-carboxamide (CPI-1205), a potent and selective inhibitor of histone methyltransferase EZH2, Suitabl. *J. Med. Chem.* 59, 9928–9941. doi: 10.1021/acs.jmedchem.6b01315

Velichutina, I., Shaknovich, R., Geng, H., Johnson, N. A., Gascoyne, R. D., Melnick, A. M., et al. (2010). EZH2-mediated epigenetic silencing in germinal center B cells contributes to proliferation and lymphomagenesis. *Blood.* 116, 5247–5255. doi: 10.1182/blood-2010-04-280149

Visser, H. P. J., and Gunster, M. J. (2001). "The Polycomb group protein EZH2 is upregulated in proliferating, cultured human mantle cell lymphoma - Visser - 2001," in *Br. J. of Haematol.* Mar; 112 (4), 950–958. (Wiley Online Library. Br. J). doi: 10.1046/j.1365-2141.2001.02641.x

Vogler, W. R., Miller, D. S., and Keller, J. W. (1976). 5 Azacytidine (NSC 102816): A new drug for the treatment of myeloblastic leukemia. *Blood.* 48, 331–337.

Wang, K., Kan, J., Yuen, S. T., Shi, S. T., Chu, K. M., Law, S., et al. (2011). Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat. Genet.* 43, 1219–1223. doi: 10.1038/ng.982

Wang, L., Pal, S., and Sif, S. (2008). Protein arginine methyltransferase 5 suppresses the transcription of the RB family of tumor suppressors in leukemia and lymphoma cells. *Mol. Cell. Biol.* 28, 6262–6277. doi: 10.1128/MCB.00923-08

Wang, W., Côté, J., Xue, Y., Zhou, S., Khavari, P. A., Biggar, S. R., et al. (1996). Purification and biochemical heterogeneity of the mammalian SWI—SNF complex. *EMBO J.* 15, 5370–5382. doi: 10.1002/j.1460-2075.1996.tb00921.x

Wang, X., Nagl, N. G., Wilsker, D., Van Scoy, M., Pacchione, S., Yaciuk, P., et al. (2004). Two related ARID family proteins are alternative subunits of human SWI/SNF complexes. *Biochem. J.* 383, 319–325. doi: 10.1042/BJ20040524

Watanabe, T., Kato, H., Kobayashi, Y., Yamasaki, S., Morita-Hoshi, Y., Yokoyama, H., et al. (2010). Potential efficacy of the oral histone deacetylase inhibitor vorinostat in a phase I trial in follicular and mantle cell lymphoma. *Cancer Sci.* 101, 196–200. doi: 10.1111/j.1349-7006.2009.01360.x

Waterfall, J. J., Arons, E., Walker, R. L., Pineda, M., Roth, L., Killian, J. K., et al. (2013). High prevalence of MAP2K1 mutations in variant and IGHV4-34-expressing hairy-cell leukemias. *Nat. Genet.* 46, 8–10. doi: 10.1038/ng.2828

Winter, G. E., Mayer, A., Buckley, D. L., Erb, M. A., Roderick, J. E., Vittori, S., et al. (2017). BET bromodomain proteins function as master transcription elongation factors independent of CDK9 recruitment. *Mol. Cell* 67, 5–18.e19. doi: 10.1016/j.molcel.2017.06.004

Wong, Y. F., Jakt, L. M., and Nishikawa, S. I. (2013). Prolonged treatment with DNMT inhibitors Induces distinct effects in promoters and gene-bodies. *PLoS One*. 8, e71099. doi: 10.1371/journal.pone.0071099

Workman, J. L., and Kingston, R. E. (2002). Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu. Rev. Biochem.* 67, 545–579. doi: 10.1146/annurev.biochem.67.1.545

Xargay-Torrent, S., López-Guerra, M., Saborit-Villarroya, I., Rosich, L., Campo, E., Roué, G., et al. (2011). Vorinostat-induced apoptosis in mantle cell lymphoma is mediated by acetylation of proapoptotic BH3-only gene promoters. *Clin. Cancer Res.* 17, 3956–3968. doi: 10.1158/1078-0432.CCR-10-3412

Xiao, B., Wilson, J. R., and Gamblin, S. J. (2003). SET domains and histone methylation. *Curr. Opin. Struct. Biol.* 13, 699–705. doi: 10.1016/j.sbi.2003.10.003

Xue, K., Gu, J. J., Zhang, Q., Mavis, C., Hernandez-Ilizaliturri, F. J., Czuczman, M. S., et al. (2016). Vorinostat, a histone deacetylase (HDAC) inhibitor, promotes cell cycle arrest and re-sensitizes rituximab- and chemo-resistant lymphoma cells to chemotherapy agents. *J. Cancer Res. Clin. Oncol.* 142, 379–387. doi: 10.1007/s00432-015-2026-y

Yang, D., Zhang, W.-P., Yang, J.-M., He, M.-X., Cheng, C., and Chen, J. (2018). Secondary skin involvement in gastric diffuse large B-cell lymphoma treated with chidamide: A case report. *Medicine (Baltimore)* 97, e13093. doi: 10.1097/MD.0000000000013093

Yang, X., Han, H., DeCarvalho, D. D., Lay, F. D., Jones, P. A., and Liang, G. (2014). Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell.* 26, 577–590. doi: 10.1016/j.ccr.2014.07.028

Yap, T. A., Johnson, P. W. M., Winter, J., Leonard, J., Giulino-Roth, L., Horner, T., et al.. (2018). A phase I, open-label study of GSK2816126, an enhancer of zeste homolog 2 (EZH2) inhibitor, in patients with relapsed/refractory diffuse large B-cell lymphoma (DLBCL), transformed follicular lymphoma (tFL), other non-Hodgkin's lymphomas (NHL), multiple my. *J. Clin. Oncol.* 34, TPS2595. doi: 10.1200/jco.2016.34.15_suppl.tps2595

Yeoh, E.-J., Ross, M. E., Shurtleff, S. A., Williams, W. K., Patel, D., Mahfouz, R., et al.. (2002). Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 1, 133–143. doi: 10.1016/S1535-6108(02)00032-6

Yoshimatsu, M., Toyokawa, G., Hayami, S., Unoki, M., Tsunoda, T., Field, H. I., et al. (2011). Dysregulation of PRMT1 and PRMT6, type I arginine methyltransferases, is involved in various types of human cancers. *Int. J. Cancer* 128, 562–573. doi: 10.1002/ijc.25366

You, J. S., and Jones, P. A. (2012). Cancer genetics and epigenetics: Two sides of the same coin? *Cancer Cell*. 22, 9–20. doi: 10.1016/j.ccr.2012.06.008

Younes, A., Berdeja, J. G., Patel, M. R., Flinn, I., Gerecitano, J. F., Neelapu, S. S., et al. (2016). Safety, tolerability, and preliminary activity of CUDC-907, a first-in-class, oral, dual inhibitor of HDAC and PI3K, in patients with relapsed or refractory lymphoma or multiple myeloma: An open-label, dose-escalation, phase 1 trial. *Lancet Oncol.* 17, 622–631. doi: 10.1016/S1470-2045(15)00584-7

Younes, A., Oki, Y., Bociek, R. G., Kuruvilla, J., Fanale, M., Neelapu, S., et al. (2011). Mocetinostat for relapsed classical Hodgkin's lymphoma: An open-label, single-arm, phase 2 trial. *Lancet Oncol.* 12, 1222–1228. doi: 10.1016/S1470-2045(11)70265-0

Younes, A., Sureda, A., Ben-Yehuda, D., Zinzani, P. L., Ong, T.-C., Prince, H. M., et al. (2012). Panobinostat in patients with relapsed/refractory Hodgkin's lymphoma after autologous stem-cell transplantation: Results of a phase II study. *J. Clin. Oncol.* 30, 2197–2203. doi: 10.1200/JCO.2011.38.1350

Yuan, B., Zhang, J., Wang, H., Xiong, L., Cai, Q., Wang, T., et al. (2011). 6-Thioguanine reactivates epigenetically silenced genes in acute lymphoblastic leukemia cells by facilitating proteasome-mediated degradation of DNMT. *Cancer Res.* 71, 1904–1911. doi: 10.1158/0008-5472.CAN-10-3430

Zeng, D., Liu, M., and Pan, J. (2016). Blocking EZH2 methylation transferase activity by GSK126 decreases stem cell-like myeloma cells. *Oncotarget*. 8, 3396–3411. doi: 10.18632/oncotarget.13773

Zeng, L., Li, J., Muller, M., Yan, S., Mujtaba, S., Pan, C., et al. (2005). Selective small molecules blocking HIV-1 Tat and coactivator PCAF association. *J. Am. Chem. Soc.* 127, 2376–2377. doi: 10.1021/ja044885g

Zhang, Y., Wei, H., Tang, K., Lin, D., Zhang, C., Mi, Y., et al. (2012). Mutation analysis of isocitrate dehydrogenase in acute lymphoblastic leukemia. *Genet. Test. Mol. Biomarkers* 16, 991–995. doi: 10.1089/gtmb.2011.0323

Zhao, H., Ning, S., Nolley, R., Scicinski, J., Oronsky, B., Knox, S. J., et al. (2017). The immunomodulatory anticancer agent, RRx-001, induces an interferon response through epigenetic induction of viral mimicry. *Clin. Epigenetics*. 9, 4. doi: 10.1186/s13148-017-0312-z

Zuber, J., Shi, J., Wang, E., Rappaport, A. R., Herrmann, H., Sison, E. A., et al. (2011). RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature*. 478, 524–528. doi: 10.1038/nature10334

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Ribeiro, Reyes-Garau, Armengol, Fernández-Serrano and Roué. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 18 October 2019
doi: 10.3389/fgene.2019.01001

Association of Sperm Methylation at *LINE-1*, Four Candidate Genes, and Nicotine/Alcohol Exposure With the Risk of Infertility

Wenjing Zhang[1,2†], Min Li[1†], Feng Sun[3†], Xuting Xu[4], Zhaofeng Zhang[1], Junwei Liu[1], Xiaowei Sun[1], Aiping Zhang[5], Yupei Shen[1], Jianhua Xu[1], Maohua Miao[1], Bin Wu[1], Yao Yuan[1], Xianliang Huang[6], Huijuan Shi[1]* and Jing Du[1]**

[1] NHC Key Laboratory of Reproduction Regulation (Shanghai Institute of Planned Parenthood Research), Fudan University, Shanghai, China, [2] Reproductive Medical Center, Changhai Hospital of Shanghai, Shanghai, China, [3] Department of Obstetrics and Gynecology, International Peace Maternity and Child Health Hospital, Shanghai Jiao Tong University, Shanghai, China, [4] Huzhou Key Laboratory of Molecular Medicine, Huzhou Central Hospital, Zhejiang, China, [5] Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders, Ministry of Education, Shanghai Jiao Tong University, Shanghai, China, [6] Shanghai Institute of Planned Parenthood Research Hospital, Shanghai, China

OPEN ACCESS

Edited by:
Yun Liu,
Fudan University, China

Reviewed by:
Osman A. El-Maarri,
University of Bonn, Germany
Na Zhu,
Columbia University, United States
Peng Chen,
Jilin University, China

***Correspondence:**
Xianliang Huang
hxl7400641@163.com
Huijuan Shi
shihuijuan2011@163.com
Jing Du
dujing42@126.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *21 January 2019*
Accepted: *20 September 2019*
Published: *18 October 2019*

Citation:
Zhang W, Li M, Sun F, Xu X, Zhang Z, Liu J, Sun X, Zhang A, Shen Y, Xu J, Miao M, Wu B, Yuan Y, Huang X, Shi H and Du J (2019) Association of Sperm Methylation at LINE-1, Four Candidate Genes, and Nicotine/Alcohol Exposure With the Risk of Infertility. Front. Genet. 10:1001. doi: 10.3389/fgene.2019.01001

In this study, we examined whether smoking and drinking affect sperm quality and the DNA methylation of the repetitive element *LINE-1*, *MEST*, *P16*, *H19*, and *GNAS* in sperm. Semen samples were obtained from 143 male residents in a minority-inhabited district of Guizhou province in southwest China. Quantitative DNA methylation analysis of the samples was performed using MassARRAY EpiTYPER assays. Sperm motility was significantly lower in both the nicotine-exposed (P = 0.0064) and the nicotine- and alcohol-exposed (P = 0.0008) groups. Follicle-stimulating hormone (FSH) levels were higher in the nicotine-exposed group (P = 0.0026). The repetitive element *LINE-1* was hypermethylated in the three exposed groups, while *P16* was hypomethylated in the alcohol and both the alcohol and nicotine exposure groups. Our results also show that alcohol and nicotine exposure altered sperm cell quality, which may be related to the methylation levels of *MEST* and *GNAS*. In addition, *MEST*, *GNAS*, and the repetitive element *LINE1* methylation was significantly associated with the concentration of sperm as well as FSH and luteinizing hormone levels.

Keywords: DNA methylation, nicotine/alcohol exposure, male infertility, imprint gene, sperm

INTRODUCTION

Infertility is a major public health concern that affects 10%–20% of married couples attempting to conceive, and male infertility is the only or a common factor (Sharlip et al., 2002; Dada et al., 2008). Recent epidemiological studies have provided plenty of pieces of evidence that environmental exposure, lifestyle, and DNA methylation are closely related to male infertility (Kobayashi et al., 2017; Laqqan et al., 2017; Nasri et al., 2017; Santi et al., 2017; Donkin and Barres, 2018; Siddeek et al., 2018). Our studies before had shown associations of aberrant DNA methylation of several genes in spermatozoa with male infertility, but the study before chose only an asthenozoospermia patient, and other environmental elements such as smoking and drinking were not examined (Xu et al., 2016). It has been shown that long-term alcohol consumption and tobacco use have adverse effects

on fertility (Besingi and Johansson, 2014; Hamad et al., 2018). Tobacco exposure is associated with impaired fecundability (Chavarro et al., 2009), slightly lower sperm viability, and reduced ejaculate volume, and spermatocyte apoptosis and disruption of the seminiferous tubules were observed (Kunzle et al., 2003; Gunes et al., 2018). Although cigarette smoking and alcohol consumption have been shown to affect DNA methylation patterns of human sperm, be related to semen quality, and have effects on endocrine control of reproductive and sexual functions, their effects on semen parameters is controversial (Gaur et al., 2010; Hansen et al., 2012; Povey et al., 2012; Chang et al., 2017; Laqqan et al., 2017; Alkhaled et al., 2018; Al Khaled et al., 2018).

Several studies have shown that smoking and drinking have similar effects on oxidative stress and DNA methylation in males of reproductive age and animal models, as has been observed for other chemicals, such as cadmium and bisphenol A (BPA) (Besingi and Johansson, 2014; Miao et al., 2014; Pierron et al., 2014; Jenkins et al., 2017; McCarthy et al., 2018). In imprinted genes, the methylation level present at pro-meiosis should be maintained throughout the male gamete development (Yaman and Grandjean, 2006; Marques et al., 2011). Nicotine exposure may alter the methylation of imprinted and non-imprinted genes in sperm that are associated with oligozoospermia and asthenozoospermia (Dai et al., 2017; Dong et al., 2017; Laqqan et al., 2017). Studies have also shown that chronic paternal alcohol exposure induced behavioral abnormalities in offspring due to alterations in the methylation of imprinted genes in sperm (Kim et al., 2014; Liang et al., 2014; Chastain and Sarkar, 2017).

Mesoderm-specific transcript (*MEST*) and *GNAS* are two maternally imprinted genes that are expressed from the paternal allele. The germ-line differentially methylated regions (DMRs) in *MEST* and *GNAS* exhibit differences in methylation levels between sperm and egg. During spermatogenesis, sperm genomic imprinting (especially in germ-line DMRs) is vulnerable to environmental factors (Marques et al., 2008). In addition, alcohol exposure could cause hypomethylation of *H19* in the sperm of offspring and reduce the mean sperm concentration (Marques et al., 2011), and the methylation levels of *H19* are related to sperm parameters, sperm chromatin, and DNA integrity (Montjean et al., 2015; Darbandi et al., 2018).

The promoter of long interspersed nucleotide element (*LINE-1*), which is used as a surrogate for global methylation, is enriched with methylated CpG dinucleotides and is usually silenced in normal tissues (Rangwala et al., 2009). However, the methylation levels of the *LINE-1* DMRs were lower in BPA-exposed spermatozoa and asthenozoospermia (Miao et al., 2014; Xu et al., 2016). In addition, the P16 protein may inhibit mitosis in spermatogonia and is related to a loss of testicular function (Xin-Chang et al., 2002; Jeong et al., 2017), and we showed that increased methylation defects in the *P16* DMR may be associated with low sperm motility (Xu et al., 2016). The imprint and non-imprint methylation marks at these DMRs are established during gametogenesis and affected by environmental exposures (Li et al., 2016). *P16* methylation is strongly associated with smoking in different pathological conditions, including lung cancer and cervical cancer (Han et al., 2016; Han et al., 2017; Wang et al., 2017). However, the relationship between these genes, tobacco/alcohol exposure, and male infertility has not yet been elucidated.

To investigate the methylation modifications that occur under exposure to alcohol and nicotine, we performed a cohort study of the methylation at the repetitive element *LINE-1* and four genes (*MEST*, *P16*, *H19*, and *GNAS*) in 143 subjects. The aim of this study was to assess whether the DNA methylation of these five genes is associated with the risk of male infertility under tobacco/alcohol exposure.

METHODS

Subjects and Clinical Data

This study included 143 male residents from a minority-inhabited district in Sandu county of Guizhou province. All participants were interviewed by trained Chinese-speaking researchers and were asked about their demography, disease history, and lifestyle factors, including tobacco use and alcohol consumption. The standards for smokers and drinkers were as we described before (Liang et al., 2017; Yang et al., 2017). We defined smoking as consuming at least one cigarette per day for more than 6 months and drinking as consuming an alcohol beverage (beer, wine, and liquor) at least once a week for more than 6 months (Witkiewitz et al., 2017; Pang et al., 2018). As per the standards of the World Health Organization, semen samples were collected after 2 days of abstinence. Sperm counts and motility were assessed by a computer-aided sperm analysis system (Cyto-S; Alpha Innotech Corp., San Leandro, CA, USA) at 37°C. The remainder of the semen sample was stored at −80°C until further examination and DNA extraction. Our study was approved by the Ethics Committee of Shanghai Institute of Planned Parenthood Research, and the local approval of Guizhou province was not required. The individuals included in this study gave written informed consent before participating. All procedures were carried out in accordance with the approved guidelines and local regulations.

TABLE 1 | Primer sequence of five genes.

Target gene	F	R	CpG	Position
MEST	GGGTTTAGAGGTATAAGAAAGAGGG	TTTCTAAAAACAACCAAACCCCTAC	1–17	chr7:130130648–130131063
P16	GTGGGTTTTAGTTTGTAGTTAAGGG	ATTATCTCCTCCTCCTCCTAACCTAA	1–35	chr9:21994026–21994435
H16	GAGATTTGAGGTGAATTTTAGGGA	CAAAACAAAATCCCCACAACC	1–20	chr11:2019635–2019930
LINE1	GGTGATTTTTGTATTTTTATTTGAGGT	CAAAAACAAACAAACCTCCTTAAACT	1–28	chr16:33760542–33761007
GNAS	GTTTTAGAGTTTTAGGGAAGGGGAG	ATCCCAAACTAACCAACTAAACCTC	1–19	chr20:57415713–57416072

DNA Methylation Assay

DNA was extracted from the semen samples by using the QIAamp DNA Mini Kit (Qiagen, Valencia, CA, USA) stored at −80°C. Bisulfite conversion of DNA was carried out using the Epitect Bisulfite Kit (Qiagen). Quantitative analysis of DNA methylation was performed using MassARRAY EpiTYPER assays (Sequenom, San Diego, CA, USA) according to a published protocol (He et al., 2013). Primers used in this study were designed using Methprimer (http://epidesigner.com; **Table 1**). CpG units that yielded data in >90% of the samples passed the initial quality control assessment. Epigenetic changes at DMRs are important in controlling the levels of gene expression. Thus methylation was measured at 76 CpG dinucleotides in the DMRs at the repetitive element *LINE-1* and four genes (*LINE-1*: 20 CpG sites; *MEST*: 7q32, 130486175–130506297, 12 CpG sites; *P16*: 9p21, 21967752–21995043, 18 CpG sites; *H19*: 11p15.5, 142575532–142578146, 14 CpG sites; and *GNAS*: 20q13.3, 58839740–58911195, 12 CpG sites; **Table 1**).

Statistical Analysis

All data were analyzed with peak picking spectra interpretation tools to generate the ratios of methyl CpG/total CpG. EpiTyper software (Sequenom, San Diego, CA, USA) was used to quantify the methylated fraction of all CpG units. Statistics 18.0 software (SPSS, Inc., Somers, NY, USA) was used to perform all the statistical analyses in this study. Pearson's correlation coefficient test and analysis of variance followed by Dunnett's *post hoc t* test were used to compare the categorical variables and the differences in the mean values of continuous variables between the two groups. All tests were two-tailed. Principal component analysis (PCA) was performed to identify underlying factors. Kaiser–Meyer–Olkin (KMO) value and Bartlett's test of sphericity were checked to confirm that the data were suitable for factor analysis. The criterion of eigenvalue >1.0 was applied to determine the number of factors retained. Items were included in the factor on which they loaded highest (minimum accepted 0.4).

RESULTS

The results of the sperm motility assessment and the levels of follicle-stimulating hormone (FSH), luteinizing hormone (LH), and testosterone (T) are shown in **Table 2**. Sperm motility was significantly lower in the nicotine-exposed (P = 0.0064) and the nicotine- and alcohol-exposed (P = 0.0008) groups than in the control group. FSH levels were higher in the nicotine-exposed group (P = 0.0026). Methylation of each CpG site and adjusted linear regression of alcohol and nicotine exposure are shown in **Suppl 1** and **Suppl 2**.

As shown in **Figure 1**, the average methylation levels of the repetitive element *LINE-1* and four assessed genes in sperm from 143 minority male residents were compared. In general, the methylation levels in the repetitive element *LINE-1* were higher in the three exposed groups (P < 0.001, P = 0.017, and P < 0.001, respectively) than in the control group, whereas methylation levels were lower in *P16* in the nicotine-exposed group and in the nicotine- and alcohol-exposed group after correction of multiple testing (P < 0.001, **Figure 1**). Individual CpG sites within the same gene showed similar trends in methylation level. Compared to the controls, the methylation levels of nine CpG sites in the repetitive element *LINE-1* (sites 2, 4.5.6, 7, 9, 14, 19, 20, 23, 25.26, and 27) were higher in the alcohol-exposed group, while the methylation levels of three CpG sites in the repetitive element *LINE-1* (sites 7, 8, and 9) were higher in the nicotine-exposed group, and the levels of 14 CpG sites were significantly higher in the nicotine- and alcohol-exposed group (**Figure 2**).

Among the imprinted genes, we found that the methylation levels of two CpG sites in the *GNAS* DMR (sites 1 and 3) were significantly lower in the alcohol-exposed and the nicotine- and alcohol-exposed groups, and only one site (site 11) in the *MEST* was lower in the nicotine-exposed group. However, the methylation levels of most CpGs in *H19* did not show obvious differences among the three groups (**Figure 2**).

Our results showed that the methylation levels of eight CpGs in *P16* in the nicotine-exposed and nicotine- and

TABLE 2 | Characteristics of all subjects.

Items mean (SD)	Neither nicotine nor alcohol exposed	Alcohol exposed only	Nicotine exposed only	Both nicotine and alcohol exposed	*P1*	*P2*	*P3*
N	48	16	16	63			
Age, years	30.13 (5.52)	36.56 (7.29)	39.60 (4.53)	38.68 (6.31)	**0.0004**	**<0.001**	**<0.001**
Marriage age, years,	24.35 (4.11)	21.07 (2.99)	27.92 (5.71)	24.69 (5.69)	**0.0064**	**0.0242**	0.7351
Body mass index	23.10 (4.02)	22.01 (2.05)	22.48 (2.71)	22.52 (2.42)	0.3186	0.6952	0.3651
Sexual absence, days	7.02 (5.70)	4.56 (4.29)	7.00 (7.48)	5.56 (8.47)	0.1196	0.9909	0.3039
Semen volume, ml,	2.96 (1.21)	2.50 (1.64)	2.90 (1.57)	2.52(1.39)	0.2366	0.8802	0.0869
Sperm concentration, $\times 10^6$/ml	65.47 (45.30)	65.40 (55.79)	55.31 (42.42)	58.23(47.94)	0.9958	0.4445	0.4213
Motility (moving forward), %	49.13 (12.20)	42.64 (17.97)	37.03 (20.34)	37.86(19.86)	0.1091	**0.0064**	**0.0008**
Vitality, %	68.50 (15.39)	63.00 (15.09)	62.38 (18.06)	67.30(16.20)	0.3733	0.3408	0.7583
LH, IU/L	5.32 (2.36)	4.25 (2.28)	6.02 (3.05)	4.94(3.97)	0.1181	0.3552	0.5603
T, IU/L	5.34 (1.96)	4.51 (2.47)	5.75 (1.81)	5.00(2.60)	0.1744	0.4735	0.4574
FSH, IU/L	5.05 (2.39)	4.16 (2.36)	7.87 (4.50)	5.69(5.74)	0.2003	**0.0026**	0.4734

P1: neither nicotine nor alcohol exposed vs alcohol exposed only; P2: neither nicotine- nor alcohol exposed vs nicotine exposed only; P3: neither nicotine nor alcohol exposed vs both nicotine and alcohol exposed. The bold represent p <0.05.

BMI, body mass index; FSH, Follicle-stimulating hormone, LH, luteinizing hormone; T, testosterone.

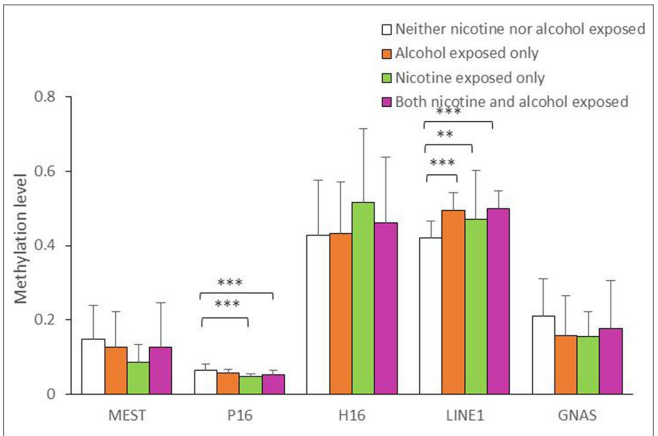


FIGURE 1 | Average methylation level of the five genes in all the subjects. The comparison of average methylation levels at *MEST, P16, H19, LINE1*, and *GNAS* in human sperms of 143 minority male residents, respectively. Data are means ± SD. Statistically significant differences are represented with asterisks: **P < 0.01, ***P < 0.001. Not significant, P > 0.05.

Correlation tests for gene modulation levels and phenotypic indices showed that the average methylation levels of *MEST* and *GNAS* were inversely correlated with sperm concentration [r = −0.522 (P = 0.038) and r = −0.557 (P = 0.025), respectively; **Figure 3A**] in the alcohol-exposed group. The average methylation levels of *MEST* and *GNAS* were positively correlated with LH levels [r = 0.344 (P = 0.012) and r = 0.365 (P = 0.006), respectively], and the methylation of the repetitive element *LINE1* was positively correlated with FSH level (r = 0.436, P = 0.001; **Figure 3B**) in the nicotine- and alcohol-exposed group. However, no association between gene methylation and the phenotypic indices was observed in the nicotine-exposed group (P > 0.05, **Table 3**).The multivariate correlation pattern between the variables was investigated using PCA. The KMO measure was 0.669, indicating that sufficient correlation existed between these variables to proceed with factor analysis. Five components were extracted by factor analysis using PCA (**Suppl 4**). Variables located near each other such as MEST, GNAS, FSH, and LH were strongly correlated (**Figure 4**).

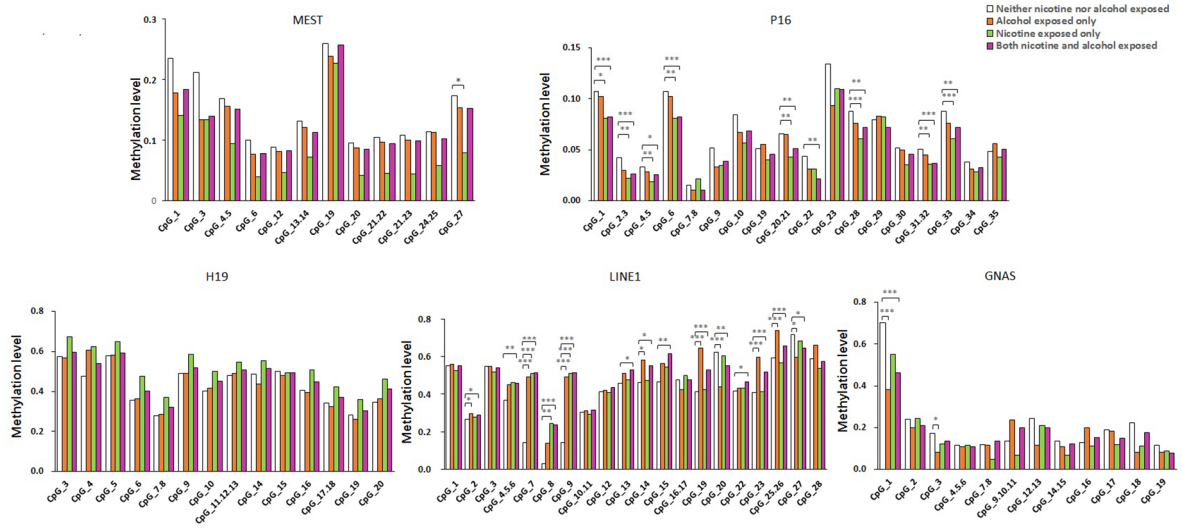


FIGURE 2 | Methylation level of the five genes in all the subjects. The comparison of site-specific methylation levels at *MEST, P16, H19, LINE1*, and *GNAS* DMR of the neither nicotine- nor alcohol-exposed controls (N = 48) with alcohol-exposed only cases (N = 16), nicotine-exposed only cases (N = 16), both nicotine- and alcohol-exposed cases (N = 63), respectively. The data are the geometric mean ratios of methylation levels at each CpG site. *P < 0.05, **P < 0.01, ***P < 0.001. Not significant, P > 0.05

alcohol-exposed groups were lower than in the controls. In contrast, there were only one CpG sites with lower methylation levels in the alcohol-exposed group when compared to the control group (**Figure 2**).

The data were checked for normal distributions using the Shapiro–Wilk test. The correlations in methylation between loci were analyzed using the Pearson's test. The methylation between *MEST* and *P16*, and *MEST* and *GNAS* were positively correlated, respectively (**Suppl 3**). The correlations between the average methylation levels and seminological parameters or hormones were analyzed using the Pearson's (normal distributions) or Spearman's correlation (abnormal distributions), respectively.

DISCUSSION

In this study, we observed that alcohol and nicotine exposure altered sperm cell quality, which may be related to the methylation levels of *MEST* and *GNAS*. The methylation levels of *MEST*, *GNAS*, and the repetitive element *LINE1* were significantly associated with sperm concentration and FSH and LH levels.

Recent studies have shown that tobacco use and alcohol consumption may increase the risk of global aberrant DNA methylation (Hamid et al., 2009; Semmler et al., 2015). Aberrant methylation of *LINE-1* has been reported in many recent studies on aging (Cho et al., 2015) and cancer

(Pattamadilok et al., 2008; Suzuki et al., 2013; Suter et al., 2004), and *LINE-1* methylation status could reflect the influence of environmental conditions or lifestyle habits on the genome (Schernhammer et al., 2010). High tobacco use might lead to a high risk of *LINE-1* hypermethylation-related cancers in men (Andreotti et al., 2014; Karami et al., 2015). To the best of our knowledge, our study showed, for the first time, that hypermethylation of the *LINE-1* gene in sperm was associated with alcohol and nicotine exposure. However, the association between *LINE-1* methylation level and alcohol and tobacco exposure in sperm needs to be further explored in future studies.

In our study, compared to the control group, hypomethylation of *P16* was observed in the tobacco-exposed group. In addition, we found that *GNAS* methylation was decreased in the alcohol-exposed group, and *P16* methylation was decreased in the nicotine- and alcohol-exposed group. Previous studies have suggested that chronic paternal alcohol exposure might contribute to mental deficits in offspring *via* abnormal methylation of imprinted genes (such as *H19* and *Peg3*) in sperm (Liang et al., 2014) and that methylation levels could be easily modified by air pollution, heavy metals, and other environmental factors, both *in vivo* and *in vitro* (Baccarelli and Bollati, 2009; Haggarty, 2013). An association between nicotine/alcohol exposure and methylation has been demonstrated in pregnant women (Barua and Junaid, 2015; Lee et al., 2015). In this study, we studied the influence of nicotine/alcohol exposure in male residents from Guizhou province, a population with a low fertility rate. The relationship between nicotine/alcohol exposure and the methylation of these five genes in male residents of Guizhou has not been previously reported. Abnormal DNA methylation in spermatozoa seems to be involved in environmental factor-induced transgenerational disruptive spermatogenesis (Anway et al., 2005; Anway et al., 2006). The impact of environmental factors on the epigenetic phenotype might affect offspring through abnormal spermatozoa methylation (Anway and Skinner, 2006). Some evidence has suggested that abnormal DNA methylation of imprinted genes may be associated with spermatogenesis failure (Boissonnas et al., 2010), and an observable decrease in the concentration of sperm was reported in patients with *H19* hypomethylation (Stouder et al., 2011). In addition, aberrant *MEST* DNA methylation has been shown to be significantly associated with increased FSH levels (Klaver et al., 2013). However, we found that alcohol exposure altered sperm cell quality and was related to the hypomethylation of *MEST* and *GNAS*. *MEST* and *GNAS* methylation levels were significantly associated with increased LH levels, and *LINE1* methylation was significantly associated with increased FSH levels. However, further studies are needed to explore the mechanisms underlying the association between chronic nicotine and alcohol exposure and aberration methylation of *MEST*, *GNAS*, and *LINE1* with sperm quality and abnormal FSH and LH levels.

Our study has several limitations. First, our study had a small sample size, which restricts the generalizability of our results. Therefore, our results must be verified in larger cohorts using different techniques. In addition, further studies are needed

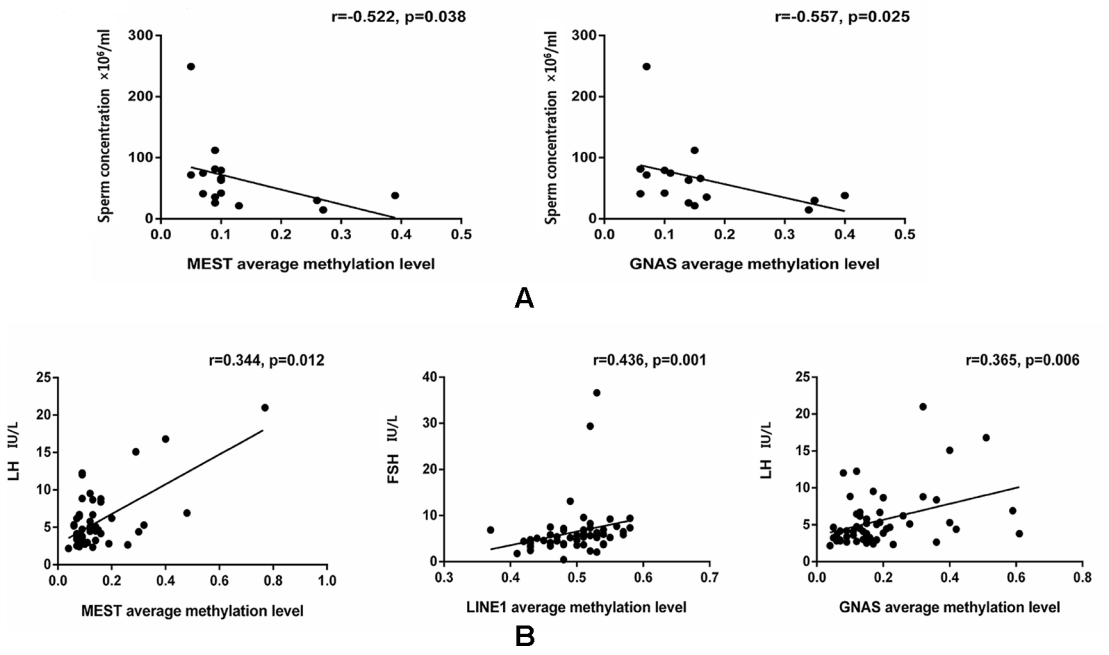


FIGURE 3 | Comparison of the gene methylation levels and semen quality/hormone level. **(A)** The significant correlations between the average methylation levels, seminological parameters and hormones in the alcohol exposed only group. **(B)** The significant correlations between the average methylation levels, seminological parameters and hormones in the both nicotine and alcohol exposed group. The dots are the intersection points between the average methylation levels, seminological parameters and hormones.

TABLE 3 | Comparison of 5 gene methylation levels and semen quality/hormone level in all subjects.

	Sperm motility (% moving forward)		Concentration (106 spermatozoa/ml)		Sperm vitality (% alive)		FSH		LH		T	
	r	P	r	P	r	P	r	P	r	P	r	P
Neither nicotine nor alcohol Exposed												
MEST	–0.033	0.828	0.317	0.030	0.121	0.525	0.022	0.886	0.051	0.735	0.025	0.867
P16	–0.024	0.872	0.020	0.894	–0.119	0.524	0.211	0.154	–0.181	0.223	–0.006	0.970
H19	–0.276	0.061	–0.147	0.323	–0.373	0.039	0.263	0.078	0.177	0.240	0.027	0.857
LINE1	–0.101	0.496	–0.224	0.125	–0.171	0.357	0.071	0.635	–0.059	0.695	–0.040	0.789
GNAS	–0.171	0.246	–0.248	0.090	–0.199	0.283	0.081	0.586	0.120	0.420	0.072	0.628
Alcohol exposed only												
MEST	0.082	0.763	–0.522	**0.038**	0.024	0.955	–0.143	0.626	0.262	0.366	–0.389	0.169
P16	–0.114	0.674	–0.291	0.274	0.582	0.130	0.026	0.930	0.333	0.245	0.019	0.949
H19	–0.284	0.286	–0.350	0.184	0.306	0.461	–0.266	0.359	0.207	0.478	–0.089	0.761
LINE1	0.169	0.531	0.016	0.953	0.518	0.188	–0.082	0.780	0.113	0.701	–0.448	0.108
GNAS	–0.054	0.843	–0.557	**0.025**	–0.005	0.986	–0.207	0.478	0.176	0.547	0.044	0.881
Nicotine exposed only												
MEST	0.041	0.884	–0.474	0.075	–0.333	0.347	0.088	0.745	–0.044	0.871	–0.362	0.169
P16	–0.354	0.196	–0.104	0.712	0.034	0.925	–0.108	0.692	–0.327	0.216	0.219	0.415
H19	–0.134	0.633	–0.427	0.113	0.604	0.064	–0.076	0.778	–0.071	0.795	–0.062	0.820
LINE1	0.056	0.849	0.041	0.890	–0.240	0.534	0.046	0.869	0.175	0.532	–0.122	0.664
GNAS	0.382	0.160	0.062	0.825	–0.544	0.104	0.209	0.438	0.482	0.058	–0.358	0.173
Both nicotine and alcohol exposed												
MEST	–0.062	0.648	–0.196	0.144	–0.023	0.889	0.241	0.082	0.344	**0.012**	0.016	0.910
P16	0.064	0.631	0.072	0.589	–0.050	0.761	0.097	0.481	0.019	0.892	–0.074	0.593
H19	–0.140	0.295	–0.103	0.441	–0.065	0.685	–0.059	0.673	0.024	0.864	0.184	0.186
LINE1	–0.037	0.780	–0.077	0.561	–0.067	0.676	0.436	**0.001**	0.173	0.206	0.168	0.221
GNAS	–0.187	0.159	–0.218	0.101	–0.087	0.593	0.194	0.156	0.365	**0.006**	0.066	0.630

The bold represent p<0.05.

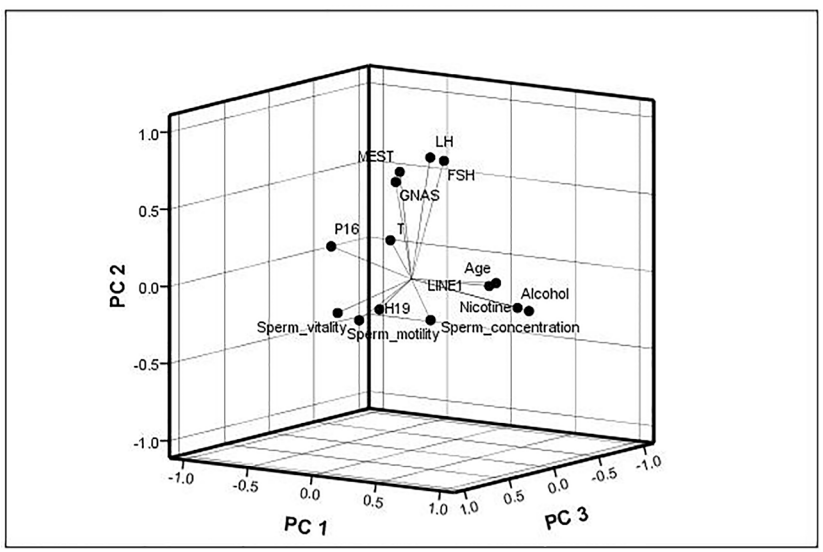


FIGURE 4 | Multivariate correlation patterns. The Kaiser–Meyer–Olkin measure was 0.669, indicating that sufficient correlation existed between these variables to proceed with factor analysis. Three components were extracted by factor analysis using principal component analysis. Factor 1 labeled as "PC1" contains nicotine exposed, alcohol exposed, age, LINE1, P16, and sperm motility and had loadings of 0.852, 0.829, 0.678, 0.578, –0.516, and –0.403, respectively. This factor explained 20.0% of the total variance. Factor 2 labeled as "PC2" contains MEST, GNAS, follicle-stimulating hormone (FSH), and luteinizing hormone (LH) and had loadings of 0.672, 0.629, 0.775, and 0.789, respectively, explaining 17.5% of the total variance. Factor 3 labeled as "PC3" contains sperm vitality, H19, and sperm concentration and had loadings of 0.746, and 0.632, and –0.516, respectively, explaining 9.3% of the total variance.

to explore how changes in methylation due to smoking and drinking affect fertility.

In conclusion, our results show that the different methylation levels of four genes, *MEST*, *P16*, *LINE-1*, and *GNAS*, alter the sperm cells of patients who consume alcohol and use nicotine. Both smoking and drinking impair sperm/semen quality and hormone levels. Thus, methylation of *MEST*, *GNAS*, and *LINE1* may be associated with sperm concentration and FSH and LH levels.

ETHICS STATEMENT

Our study was approved by the Ethics Committee of Shanghai Institute of Planned Parenthood Research (SIPPR) and the local approval of Guizhou province was not required.

AUTHOR CONTRIBUTIONS

Designed and coordinated the study: WZ, AZ, HS, JD. Performed the experiments: FS, ZZ, YS, JX, ML. Analyzed the data: XX, JL, XS, MM, JD. Contributed reagents/materials/analysis tools: YY, FS, BW. Wrote the paper: WZ, FS, JD. Helped in critical revision of the manuscript for important intellectual content: DJ, XH, ML.

ACKNOWLEDGMENTS

We thank Chunli Zhong, Yuan Yang, Kanglian Chen, and Jiang Zhu for help in collecting samples from the Population and Family Planning Institute of Guizhou Province. We would like to acknowledge research support from the National Nature Science Foundation of China (Nos. 81571503, 81270744, and 81771655) and the Major State Basic Research Development Program of China (973 Program, 2014CB943104).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01001/full#supplementary-material

REFERENCES

Al Khaled, Y., Tierling, S., Laqqan, M., Lo Porto, C., and Hammadeh, M. E. (2018). Cigarette smoking induces only marginal changes in sperm DNA methylation levels of patients undergoing intracytoplasmic sperm injection treatment. *Andrologia* 50. doi: 10.1111/and.12818

Alkhaled, Y., Laqqan, M., Tierling, S., Lo Porto, C., Amor, H., and Hammadeh, M. E. (2018). Impact of cigarette-smoking on sperm DNA methylation and its effect on sperm parameters. *Andrologia*. doi: 10.1111/and.12950

Andreotti, G., Karami, S., Pfeiffer, R. M., Hurwitz, L., Liao, L. M., Weinstein, S. J., et al. (2014). LINE1 methylation levels associated with increased bladder cancer risk in pre-diagnostic blood DNA among US (PLCO) and European (ATBC) cohort study participants. *Epigenetics* 9, 404–415. doi: 10.4161/epi.27386

Anway, M. D., and Skinner, M. K. (2006). Epigenetic transgenerational actions of endocrine disruptors. *Endocrinology* 147, S43–S49. doi: 10.1210/en.2006-0640

Anway, M. D., Cupp, A. S., Uzumcu, M., and Skinner, M. K. (2005). Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science* 308, 1466–1469. doi: 10.1126/science.1108190

Anway, M. D., Memon, M. A., Uzumcu, M., and Skinner, M. K. (2006). Transgenerational effect of the endocrine disruptor vinclozolin on male spermatogenesis. *J. Androl.* 27, 868–879. doi: 10.2164/jandrol.106.000349

Baccarelli, A., and Bollati, V. (2009). Epigenetics and environmental chemicals. *Curr. Opin. Pediatr.* 21, 243–251. doi: 10.1097/MOP.0b013e32832925cc

Barua, S., and Junaid, M. A. (2015). Lifestyle, pregnancy and epigenetic effects. *Epigenomics* 7, 85–102. doi: 10.2217/epi.14.71

Besingi, W., and Johansson, A. (2014). Smoke-related DNA methylation changes in the etiology of human disease. *Hum Mol. Genet.* 23, 2290–2297. doi: 10.1093/hmg/ddt621

Boissonnas, C. C., Abdalaoui, H. E., Haelewyn, V., Fauque, P., Dupont, J. M., Gut, I., et al. (2010). Specific epigenetic alterations of IGF2-H19 locus in spermatozoa from infertile men. *Eur. J. Hum. Genet.* 18, 73–80. doi: 10.1038/ejhg.2009.117

Chang, R. C., Skiles, W. M., Chronister, S. S., Wang, H., Sutton, G. I., Bedi, Y. S., et al. (2017). DNA methylation-independent growth restriction and altered developmental programming in a mouse model of preconception male alcohol exposure. *Epigenetics* 12, 841–853. doi: 10.1080/15592294.2017.1363952

Chastain, L. G., and Sarkar, D. K. (2017). Alcohol effects on the epigenome in the germline: role in the inheritance of alcohol-related pathology. *Alcohol* 60, 53–66. doi: 10.1016/j.alcohol.2016.12.007

Chavarro, J. E., Rich-Edwards, J. W., Rosner, B. A., and Willett, W. C. (2009). Caffeinated and alcoholic beverage intake in relation to ovulatory disorder infertility. *Epidemiology* 20, 374–381. doi: 10.1097/EDE.0b013e31819d68cc

Cho, Y. H., Woo, H. D., Jang, Y., Porter, V., Christensen, S., Hamilton, R. F., Jr., et al. (2015). The Association of LINE-1 Hypomethylation with Age and Centromere Positive Micronuclei in Human Lymphocytes. *PLoS One* 10, e0133909. doi: 10.1371/journal.pone.0133909

Dada, R., Kumar, R., Shamsi, M. B., Tanwar, M., Pathak, D., Venkatesh, S., et al. (2008). Genetic screening in couples experiencing recurrent assisted procreation failure. *Indian J. Biochem. Biophys.* 45, 116–120.

Dai, J., Wang, Z., Xu, W., Zhang, M., Zhu, Z., Zhao, X., et al. (2017). Paternal nicotine exposure defines different behavior in subsequent generation via hyper-methylation of mmu-miR-15b. *Sci. Rep.* 7, 7286. doi: 10.1038/s41598-017-07920-3

Darbandi, M., Darbandi, S., Agarwal, A., Baskaran, S., Dutta, S., Sengupta, P., et al. (2018). Reactive oxygen species-induced alterations in H19-Igf2 methylation patterns, seminal plasma metabolites, and semen quality. *J. Assist. Reprod. Genet.* doi: 10.1007/s10815-018-1350-y

Dong, H., Wang, Y., Zou, Z., Chen, L., Shen, C., Xu, S., et al. (2017). Abnormal methylation of imprinted genes and cigarette smoking: assessment of Their association with the risk of male infertility. *Reprod. Sci.* 24, 114–123. doi: 10.1177/1933719116650755

Donkin, I., and Barres, R. (2018). Sperm epigenetics and influence of environmental factors. *Mol. Metabol.* 14, 1–11. doi: 10.1016/j.molmet.2018.02.006

Gaur, D. S., Talekar, M. S., and Pathak, V. P. (2010). Alcohol intake and cigarette smoking: impact of two major lifestyle factors on male fertility. *Indian J Pathol Microbiol* 53, 35–40. doi: 10.4103/0377-4929.59180

Gunes, S., Metin Mahmutoglu, A., Arslan, M. A., and Henkel, R. (2018). Smoking-induced genetic and epigenetic alterations in infertile men. *Andrologia* 50, e13124. doi: 10.1111/and.13124

Haggarty, P. (2013). Epigenetic consequences of a changing human diet. *Proc. Nutr. Soc.* 72, 363–371. doi: 10.1017/S0029665113003376

Hamad, M. F., Dayyih, W. A. A., Laqqan, M., AlKhaled, Y., Montenarh, M., and Hammadeh, M. E. (2018). The status of global DNA methylation in the spermatozoa of smokers and non-smokers. *Reprod. Biomed. Online* 37, 581–589. doi: 10.1016/j.rbmo.2018.08.016

Hamid, A., Wani, N. A., and Kaur, J. (2009). New perspectives on folate transport in relation to alcoholism-induced folate malabsorption–association with epigenome stability and cancer development. *FEBS J.* 276, 2175–2191. doi: 10.1111/j.1742-4658.2009.06959.x

Han, J. C., Xu, F., Chen, N., Qi, G. B., Wei, Y. J., Li, H. B., et al. (2016). Promoter methylations of RASSF1A and p16 is associated with clinicopathological features in lung cancers. *J. Cancer Res. Ther.* 12, 340–349. doi: 10.4103/0973-1482.154926

Han, Y. D., Wang, X. B., Cui, N. H., Zhang, S., Wang, C., and Zheng, F. (2017). Associations of P16INK4a promoter hypermethylation with squamous intraepithelial lesion, cervical cancer and their clinicopathological features: a meta-analysis. *Oncotarget* 8, 1871–1883. doi: 10.18632/oncotarget.12202

Hansen, M. L., Thulstrup, A. M., Bonde, J. P., Olsen, J., Hakonsen, L. B., and Ramlau-Hansen, C. H. (2012). Does last week's alcohol intake affect semen quality or reproductive hormones? A cross-sectional study among healthy young Danish men. *Reprod. Toxicol.* 34, 457–462. doi: 10.1016/j.reprotox.2012.06.004

He, J., Zhang, A., Fang, M., Fang, R., Ge, J., Jiang, Y., et al. (2013). Methylation levels at IGF2 and GNAS DMRs in infants born to preeclamptic pregnancies. *BMC Genomics* 14, 472. doi: 10.1186/1471-2164-14-472

Jenkins, T. G., James, E. R., Alonso, D. F., Hoidal, J. R., Murphy, P. J., Hotaling, J. M., et al. (2017). Cigarette smoking significantly alters sperm DNA methylation patterns. *Andrology* 5, 1089–1099. doi: 10.1111/andr.12416

Jeong, M. G., Song, H., Shin, J. H., Jeong, H., Kim, H. K., and Hwang, E. S. (2017). Transcriptional coactivator with PDZ-binding motif is required to sustain testicular function on aging. *Aging Cell* 16, 1035–1042. doi: 10.1111/acel.12631

Karami, S., Andreotti, G., Liao, L. M., Pfeiffer, R. M., Weinstein, S. J., Purdue, M. P., et al. (2015). LINE1 methylation levels in pre-diagnostic leukocyte DNA and future renal cell carcinoma risk. *Epigenetics* 10, 282–292. doi: 10.1080/15592294.2015.1006505

Kim, P., Choi, C. S., Park, J. H., Joo, S. H., Kim, S. Y., Ko, H. M., et al. (2014). Chronic exposure to ethanol of male mice before mating produces attention deficit hyperactivity disorder-like phenotype along with epigenetic dysregulation of dopamine transporter expression in mouse offspring. *J. Neurosci. Res.* 92, 658–670. doi: 10.1002/jnr.23275

Klaver, R., Tuttelmann, F., Bleiziffer, A., Haaf, T., Kliesch, S., and Gromoll, J. (2013). DNA methylation in spermatozoa as a prospective marker in andrology. *Andrology* 1, 731–740. doi: 10.1111/j.2047-2927.2013.00118.x

Kobayashi, N., Miyauchi, N., Tatsuta, N., Kitamura, A., Okae, H., Hiura, H., et al. (2017). Factors associated with aberrant imprint methylation and oligozoospermia. *Sci. Rep.* 7, 42336. doi: 10.1038/srep42336

Kunzle, R., Mueller, M. D., Hanggi, W., Birkhauser, M. H., Drescher, H., and Bersinger, N. A. (2003). Semen quality of male smokers and nonsmokers in infertile couples. *Fertil. Steril.* 79, 287–291. doi: 10.1016/S0015-0282(02)04664-2

Laqqan, M., Tierling, S., Alkhaled, Y., LoPorto, C., and Hammadeh, M. E. (2017). Alterations in sperm DNA methylation patterns of oligospermic males. *Reprod. Biol.* 17, 396–400. doi: 10.1016/j.repbio.2017.10.007

Laqqan, M., Tierling, S., Alkhaled, Y., Porto, C. L., Solomayer, E. F., and Hammadeh, M. E. (2017). Aberrant DNA methylation patterns of human spermatozoa in current smoker males. *Reprod. Toxicol.* 71, 126–133. doi: 10.1016/j.reprotox.2017.05.010

Lee, B. Y., Park, S. Y., Ryu, H. M., Shin, C. Y., Ko, K. N., Han, J. Y., et al. (2015). Changes in the methylation status of DAT, SERT, and MeCP2 gene promoters in the blood cell in families exposed to alcohol during the periconceptional period. *Alcohol. Clin. Exp. Res.* 39, 239–250. doi: 10.1111/acer.12635

Li, Y., Xie, C., Murphy, S. K., Skaar, D., Nye, M., Vidal, A. C., et al. (2016). Lead exposure during early human development and DNA methylation of imprinted gene regulatory elements in adulthood. *Environ. Health Perspect.* 124, 666–673. doi: 10.1289/ehp.1408577

Liang, F., Diao, L., Liu, J., Jiang, N., Zhang, J., Wang, H., et al. (2014). Paternal ethanol exposure and behavioral abnormities in offspring: associated alterations in imprinted gene methylation. *Neuropharmacology* 81, 126–133. doi: 10.1016/j.neuropharm.2014.01.025

Liang, H., Xu, W., Chen, J., Shi, H., Zhu, J., Liu, X., et al. (2017). The Association between exposure to environmental bisphenol A and gonadotropic hormone levels among men. *PLoS One* 12, e0169217. doi: 10.1371/journal.pone.0169217

Marques, C. J., Costa, P., Vaz, B., Carvalho, F., Fernandes, S., Barros, A., et al. (2008). Abnormal methylation of imprinted genes in human sperm is associated with oligozoospermia. *Mol. Hum. Reprod.* 14, 67–74. doi: 10.1093/molehr/gam093

Marques, C. J., Joao Pinho, M., Carvalho, F., Bieche, I., Barros, A., and Sousa, M. (2011). DNA methylation imprinting marks and DNA methyltransferase expression in human spermatogenic cell stages. *Epigenetics* 6, 1354–1361. doi: 10.4161/epi.6.11.17993

McCarthy, D. M., Morgan, T. J., Jr., Lowe, S. E., Williamson, M. J., Spencer, T. J., Biederman, J., et al. (2018). Nicotine exposure of male mice produces behavioral impairment in multiple generations of descendants. *PLoS Biol.* 16, e2006497. doi: 10.1371/journal.pbio.2006497

Miao, M., Zhou, X., Li, Y., Zhang, O., Zhou, Z., Li, T., et al. (2014). LINE-1 hypomethylation in spermatozoa is associated with Bisphenol A exposure. *Andrology* 2, 138–144. doi: 10.1111/j.2047-2927.2013.00166.x

Montjean, D., Zini, A., Ravel, C., Belloc, S., Dalleac, A., Copin, H., et al. (2015). Sperm global DNA methylation level: association with semen parameters and genome integrity. *Andrology* 3, 235–40. doi: 10.1111/andr.12001

Nasri, F., Gharesi-Fard, B., Namavar Jahromi, B., Farazi-Fard, M. A., Banaei, M., Davari, M., et al. (2017). Sperm DNA methylation of H19 imprinted gene and male infertility. *Andrologia* 49. doi: 10.1111/and.12766

Pang, Y., Holmes, M. V., Guo, Y., Yang, L., Bian, Z., Chen, Y., et al. (2018). Smoking, alcohol, and diet in relation to risk of pancreatic cancer in China: a prospective study of 0.5 million people. *Cancer Med.* 7, 229–239. doi: 10.1002/cam4.1261

Pattamadilok, J., Huapai, N., Rattanatanyong, P., Vasurattana, A., Triratanachat, S., Tresukosol, D., et al. (2008). LINE-1 hypomethylation level as a potential prognostic factor for epithelial ovarian cancer. *Int. J. Gynecol. Cancer* 18, 711–717. doi: 10.1111/j.1525-1438.2007.01117.x

Pierron, F., Baillon, L., Sow, M., Gotreau, S., and Gonzalez, P. (2014). Effect of low-dose cadmium exposure on DNA methylation in the endangered European eel. *Environ. Sci. Technol. Lett.* 48, 797–803. doi: 10.1021/es4048347

Povey, A. C., Clyma, J. A., McNamee, R., Moore, H. D., Baillie, H., Pacey, A. A., et al. (2012). Modifiable and non-modifiable risk factors for poor semen quality: a case-referent study. *Hum. Reprod.* 27, 2799–2806. doi: 10.1093/humrep/des183

Rangwala, S. H., Zhang, L., and Kazazian, Jr, H. H. (2009). Many LINE1 elements contribute to the transcriptome of human somatic cells. *Genome Biol.* 10, R100. doi: 10.1186/gb-2009-10-9-r100

Santi, D., De Vincentis, S., Magnani, E., and Spaggiari, G. (2017). Impairment of sperm DNA methylation in male infertility: a meta-analytic study. *Andrology* 5, 695–703. doi: 10.1111/andr.12379

Schernhammer, E. S., Giovannucci, E., Kawasaki, T., Rosner, B., Fuchs, C. S., and Ogino, S. (2010). Dietary folate, alcohol and B vitamins in relation to LINE-1 hypomethylation in colon cancer. *Gut* 59, 794–799. doi: 10.1136/gut.2009.183707

Semmler, A., Heese, P., Stoffel-Wagner, B., Muschler, M., Heberlein, A., Bigler, L., et al. (2015). Alcohol abuse and cigarette smoking are associated with global DNA hypermethylation: results from the German Investigation on Neurobiology in Alcoholism (GINA). *Alcohol* 49, 97–101. doi: 10.1016/j.alcohol.2015.01.004

Sharlip, I. D., Jarow, J. P., Belker, A. M., Lipshultz, L. I., Sigman, M., Thomas, A. J., et al. (2002). Best practice policies for male infertility. *Fertil Steril* 77, 873–882. doi: 10.1016/S0015-0282(02)03105-9

Siddeek, B., Mauduit, C., Simeoni, U., and Benahmed, M. (2018). Sperm epigenome as a marker of environmental exposure and lifestyle, at the origin of diseases inheritance. *Mutat Res* 778, 38–44. doi: 10.1016/j.mrrev.2018.09.001

Stouder, C., Somm, E., and Paoloni-Giacobino, A. (2011). Prenatal exposure to ethanol: a specific effect on the H19 gene in sperm. *Reprod. Toxicol.* 31, 507–512. doi: 10.1016/j.reprotox.2011.02.009

Suter, C. M., Martin, D. I., and Ward, R. L. (2004). Hypomethylation of L1 retrotransposons in colorectal cancer and adjacent normal tissue. *Int J Colorectal Dis* 19, 95–101. doi: 10.1007/s00384-003-0539-3

Suzuki, M., Shiraishi, K., Eguchi, A., Ikeda, K., Mori, T., Yoshimoto, K., et al. (2013). Aberrant methylation of LINE-1, SLIT2, MAL and IGFBP7 in non-small cell lung cancer. *Oncol. Rep.* 29, 1308–1314. doi: 10.3892/or.2013.2266

Wang, B. H., Li, Y. Y., Han, J. Z., Zhou, L. Y., Lv, Y. Q., Zhang, H. L., et al. (2017). Gene methylation as a powerful biomarker for detection and screening of non-small cell lung cancer in blood. *Oncotarget* 8, 31692–31704. doi: 10.18632/oncotarget.15919

Witkiewitz, K., Hallgren, K. A., Kranzler, H. R., Mann, K. F., Hasin, D. S., Falk, D. E., et al. (2017). Clinical validation of reduced alcohol consumption after treatment for alcohol dependence using the World Health Organization risk drinking levels. *Alcohol. Clin. Exp. Res.* 41, 179–186. doi: 10.1111/acer.13272

Xin-Chang, Z., Peng, W., Zhao-Yuan, H., Xiao-Bin, H., Ru-Jin, Z., and Yi-Xun, L. (2002). Expression of P16(INK4a) in testis of rhesus monkey during heat

stress and testosterone undecanoate induced azoospermia or oligozoospermia. *Contraception* 65, 251–255. doi: 10.1016/S0010-7824(01)00305-5

Xu, J., Zhang, A., Zhang, Z., Wang, P., Qian, Y., He, L., et al. (2016). DNA methylation levels of imprinted and nonimprinted genes DMRs associated with defective human spermatozoa. *Andrologia* 48, 939–947. doi: 10.1111/and.12535

Yaman, R., and Grandjean, V. (2006). Timing of entry of meiosis depends on a mark generated by DNA methyltransferase 3a in testis. *Mol. Reprod. Dev.* 73, 390–397. doi: 10.1002/mrd.20430

Yang, F., Li, L., Chen, J. P., Liu, X. Q., Zhong, C. L., Yang, Y., et al. (2017). Couple's infertility in relation to male smoking in a Chinese rural area. *Asian J. Androl.* 19, 311–315 doi: 10.4103/1008-682X.168685.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Zhang, Li, Sun, Xu, Zhang, Liu, Sun, Zhang, Shen, Xu, Miao, Wu, Yuan, Huang, Shi and Du. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

REVIEW
published: 30 October 2019
doi: 10.3389/fgene.2019.01044

Perspectives on miRNAs as Epigenetic Markers in Osteoporosis and Bone Fracture Risk: A Step Forward in Personalized Diagnosis

*Michela Bottani[1], Giuseppe Banfi[1,2] and Giovanni Lombardi[1,3]**

[1] IRCCS Istituto Ortopedico Galeazzi, Laboratory of Experimental Biochemistry & Moelcular Biology, Milano, Italy, [2] Vita-Salute San Raffaele University, Milano, Italy, [3] Department of Physiology & Pharmacology, Gdańsk University of Physical Education & Sport, Gdańsk, Poland

Aging is associated with an increased incidence of age-related bone diseases. Current diagnostics (e.g., conventional radiology, biochemical markers), because limited in specificity and sensitivity, can distinguish between healthy or osteoporotic subjects but they are unable to discriminate among different underlying causes that lead to the same bone pathological condition (e.g., bone fracture risk). Among recent, more sensitive biomarkers, miRNAs — the non-coding RNAs involved in the epigenetic regulation of gene expression, have emerged as fundamental post-transcriptional modulators of bone development and homeostasis. Each identified miRNA carries out a specific role in osteoblast and osteoclast differentiation and functional pathways (osteomiRs). miRNAs bound to proteins or encapsulated in exosomes and/or microvesicles are released into the bloodstream and biological fluids where they can be detected and measured by highly sensitive and specific methods (e.g., quantitative PCR, next-generation sequencing). As such, miRNAs provide a prompt and easily accessible tool to determine the subject-specific epigenetic environment of a specific condition. Their use as biomarkers opens new frontiers in personalized medicine. While miRNAs circulating levels are lower than those found in the tissue/cell source, their quantification in biological fluids may be strategic in the diagnosis of diseases that affect tissues, such as bone, in which biopsy may be especially challenging. For a biomarker to be valuable in clinical practice and support medical decisions, it must be (easily) measurable, validated by independent studies, and strongly and significantly associated with a disease outcome. Currently, miRNAs analysis does not completely satisfy these criteria, however. Starting from *in vitro* and *in vivo* observations describing their biological role in bone cell development and metabolism, this review describes the potential use of bone-associated circulating miRNAs as biomarkers for determining predisposition, onset, and development of osteoporosis and bone fracture risk. Moreover, the review focuses on their clinical relevance and discusses the pre-analytical, analytical, and post-analytical issues in their measurement, which still limits their routine application. Taken together, research and clinical findings may be helpful for creating miRNA-based diagnostic tools in the diagnosis and treatment of bone diseases.

Keywords: biomarkers, circulating miRNAs, miRNA signature, extra-analytical variability, sensitivity and specificity, osteopenia/osteoporosis, fracture risk

OPEN ACCESS

Edited by:
Nejat Dalay,
Istanbul University, Turkey

Reviewed by:
Ling-Qing Yuan,
Central South University, China
Daniele Bellavia,
Rizzoli Orthopaedic Institute (IRCCS), Italy

***Correspondence:**
Giovanni Lombardi
giovanni.lombardigrupposandonato.it;
giovanni.lombardi@awf.gda.pl

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *19 April 2019*
Accepted: *30 September 2019*
Published: *30 October 2019*

Citation:
Bottani M, Banfi G and Lombardi G (2019) Perspectives on miRNAs as Epigenetic Markers in Osteoporosis and Bone Fracture Risk: A Step Forward in Personalized Diagnosis. Front. Genet. 10:1044. doi: 10.3389/fgene.2019.01044

INTRODUCTION

Biogenesis of miRNAs and Their Biological Role

MicroRNAs (miRNAs) are short, single-stranded non-coding RNAs (18–22 nucleotides in length) that inhibit gene expression. Lee et al. (1993) discovered in *Caenorhabditis elegans* — a short, single-stranded non-coding RNA (lin-4) that downregulated lin-14 gene expression through a direct antisense RNA–RNA interaction. Since then, miRNAs have been discovered in all living kingdoms (Lagos-Quintana et al., 2001; Reinhart et al., 2002; Cerutti and Casas-Mollano, 2006; Dang et al., 2011; Bloch et al., 2017) and in viruses, as well (Grundhoff and Sullivan, 2011). Among the databases that record the ever growing number of miRNAs being discovered, miRBase (www.mirbase.org) is a comprehensive and constantly updated miRNAs database that provides universal nomenclature, information about sequence, predicted target genes, and additional annotations (Griffiths-Jones et al., 2006). Currently, it contains 38,589 entries, more than 1,900 of which are human.

Though widely discussed, miRNAs biogenesis is not yet fully understood. Briefly, miRNAs are transcribed by RNA polymerase II (Pol II) from encoding sequences (miRNA genes) located within non-coding DNA sequences, introns or untranslated regions (UTR) of protein-coding genes (Ha and Kim, 2014; Hammond, 2015). miRNA genes can be found in clusters within a chromosomal locus; they are transcribed as polycistronic primary transcripts and subsequently processed as single miRNA precursors. miRNAs within the same cluster are thought to target related mRNAs (Lee et al., 2002; Wang et al., 2016). Furthermore, the same miRNA encoding genes can be duplicated in different loci: the derived mature miRNAs (grouped within a miRNA family) have an identical seed region and share the same mRNA targets (Bartel, 2009). A long primary transcript (pri-miRNA) is processed in the nucleus by the RNase III DROSHA-DGCR8 cofactor complex that removes the stem loop-flanking structure generating the ~60 nt hairpin pre-miRNA.

After its exportation into the cytosol in a process mediated by exportin 5 (EXP5), RNase III DICER cleaves the loop to generate a double stranded (ds) miRNA. One miRNA strand, the passenger strand, is incorporated into the RNA-induced silencing complex (RISC) as a mature miRNA, while the other, the star strand, is degraded. Both strands in some miRNAs are bioactive and each strand is loaded into a RISC. The RISC protein argonaute-2 (AGO-2) is responsible for targeting a specific mRNA based on the complementarity of a 7-nt miRNA sequence ("seed region," position 2-to-7). The ds miRNA–mRNA complex induces degradation of the target mRNA, inhibition of its translation, and consequent modulation of the downstream cellular processes. Other DICER- or DROSHA-independent non-canonical miRNA biogenesis pathways exist (Ha and Kim, 2014; Hammond, 2015). Finally, miRNAs expression undergoes multilevel regulation: epigenetically in DNA methylation and histone modifications (e.g., histone acetylation) (Saito et al., 2006; Scott et al., 2006; Lujambio et al., 2008; Lujambio and Esteller, 2009) and through the regulation of proteins involved in miRNAs maturation (Davis-Dusenbery and Hata, 2010). Beside their more known inhibitory function, there are evidence suggesting that at least some miRNAs can induce gene expression under specific conditions. In this process, miRNA-associated ribonucleoproteins (miRNPs) play a key role as reviewed in (Valinezhad Orang et al., 2014).

One of the first demonstrations of the key role of miRNAs was the embryonic lethality of the DICER-1- and DGCR8-double knockout (KO) in mice (Bernstein et al., 2003; Wang et al., 2007). Conditional inactivation of DICER in mice embryonic stem (ES) impaired proliferation and differentiation and compromised miRNA biogenesis (Suh et al., 2004; Murchison et al., 2005). Several miRNAs display a cell- or tissue-specific expression profile, while others are more widely expressed (Ludwig et al., 2016). Since they are also present in human biological fluids (Weber et al., 2010), their abundance and stability in human serum and plasma prompted the idea for their potential use as biomarkers (Chen et al., 2008).

Figure 1 illustrates the canonical miRNA biogenetic pathway and notions about their nomenclature.

Aim

Based on the potentialities of miRNAs as biomarkers, research efforts have been spent in studying and defining the relationships between their altered expression and human disease, particularly bone diseases (Bellavia et al., 2019; Hadjiargyrou and Komatsu, 2019; Van Meurs et al., 2019). The search term "miRNA" on PubMed retrieves 83,067 records, 53,240 (64%) of which were published in the last 5 years.

Different from previous reviews, the aim of this paper is to comprehensively review the available data about the potential next use, or even the actual use, of circulating miRNAs as biological indexes for osteoporosis and bone fracture risk. We gleaned information from each article that claimed miRNAs diagnostic, prognostic, and/or predictive properties, including information about the pre-analytical phase, quantification platforms, and normalization methods used. Several articles also reported the sensitivity and specificity parameters in evaluating the clinical potential of a specific miRNA as a biomarker to assess the presence of disease and, at the same time, the absence of the disease in healthy individuals. Since sensitivity and specificity are inversely correlated, they can be plotted on a receiver operating characteristic (ROC) curve as 1-specificity vs. sensitivity (Hajian-Tilaki, 2013).

miRNA can be found in human biofluids and in blood as free (mainly protein-associated) and exosome-/microvesicle-/LDL-associated miRNAs. These two distinct subsets are believed to exert different functions: the free fraction is somehow passively released from cells during normal recycling of the subcellular components, whereas the encapsulated fraction is actively released and finely packaged together with other components with specific functions addressed to other target tissues. In these terms, free-miRNAs can be considered classical biomarkers, while encapsulated miRNAs more likely act as endocrine-like factors (Bayraktar et al., 2017). This review will discuss bone tissue and bone-associated free-circulating miRNAs in relation to osteoporosis and the related risk of bone fracture. In addition, the review will systematically

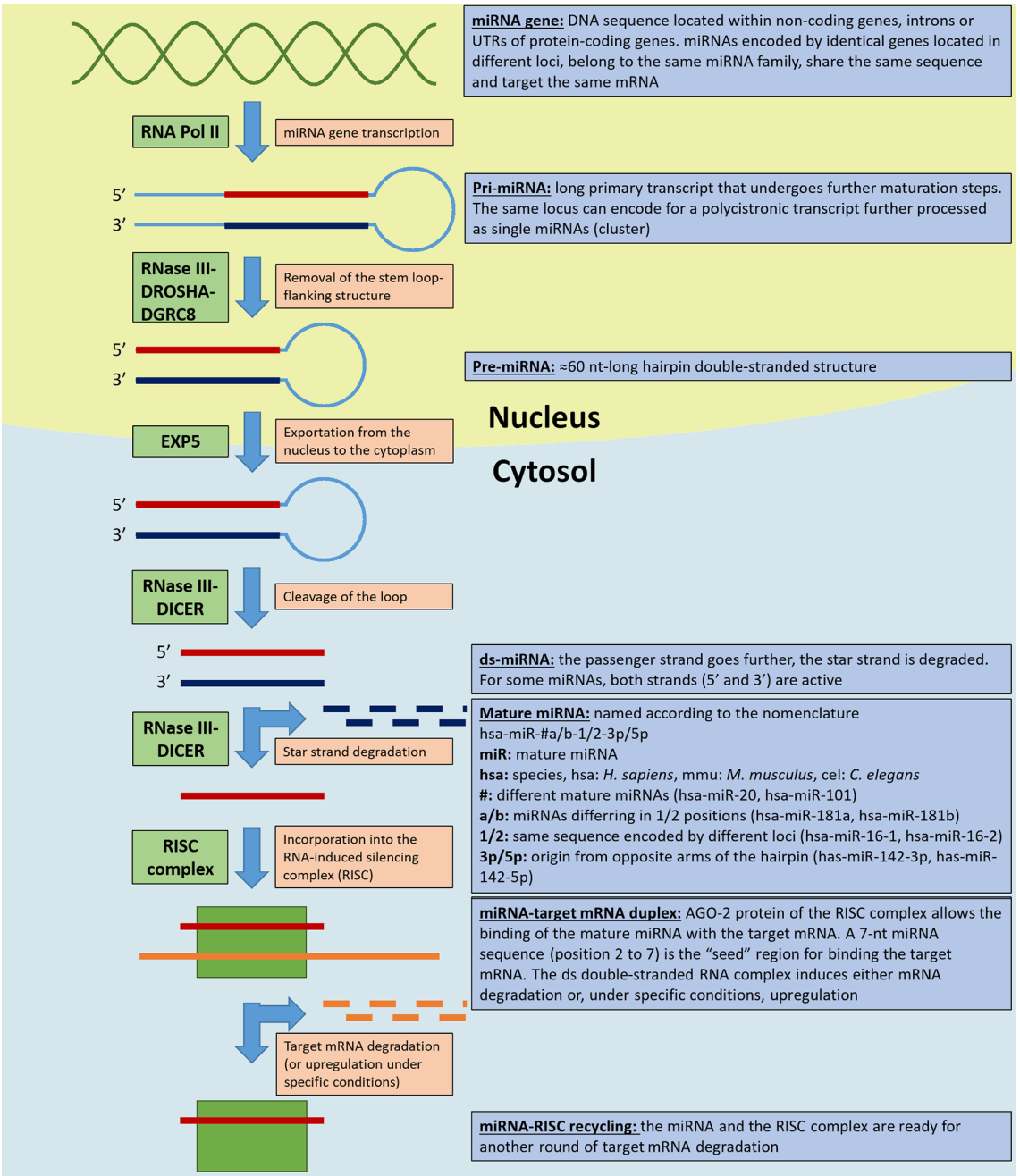


FIGURE 1 | miRNA biogenesis and nomenclature. The figure illustrates the key steps in miRNAs biogenetic canonical pathways. The light orange boxes indicate the step, the green boxes the key enzyme/enzyme complexes involved in the process, and the light blue boxes the miRNAs and miRNAs precursor nomenclature and specifications (according to Griffiths-Jones et al., 2006). RNA Pol II, RNA polymerase II; EXP5, exportin 5; RISC, RNA-induced silencing complex; AGO-2, argonaute-2 protein.

describe the *in vivo–in vitro* evidence for the role, the pathways, and the putative target genes of these miRNAs.

MIRNAS AS BIOMARKERS

Borrowing from Morrow and de Lemos (2007), the three essential features of a novel cardiovascular biomarker for clinical use are: measurability in a certain clinical setting; validation by multiple studies; and direct impact on medical decision making and patient management.

The measurability criterion requires an accurate and reproducible analytical method that can provide reliable measures rapidly and at reasonable cost. Furthermore, pre-analytical issues (conditions of measurement and sample handling, type, and stability) must be known and solved beforehand in order to control for variables in

the biomarker's measurability/detectability. The validation criterion requires a strong and consistent association between the outcome/disease of interest and the biomarker level based on evidence from multiple clinical studies. Moreover, in order to directly impact medical decision making, a novel biomarker must perform better than existing tests and the associated risk might be modified by a specific therapy (Morrow and de Lemos, 2007). These criteria are still burdened by several issues regarding the pre-analytical, analytical, and post-analytical phases in miRNAs.

miRNAs as Biomarkers: Strengths

These limitations notwithstanding, the use of circulating (or also tissue) miRNAs as biomarkers is nearly ready for implementation in clinical practice. Interest in these molecules arises from the fact that, as epigenetic regulators of gene expression, they act as modulators rather than effectors of a specific biological function. As such, they provide a prompt and easily accessible tool to determine the epigenetic environment of a specific condition. And as subject-specific epigenetic determinants of a condition, they can be considered a personalized signature for tailor-made diagnosis and/or treatment. Circulating miRNAs are easily detectable in biofluids such as (but not only) plasma, serum, and urine, which are minimal/non-invasive sources of biomarkers with broad applicability in clinical research and repositories (Weber et al., 2010; Hackl et al., 2016). Although circulating miRNAs levels are lower than those found in tissues and cells (Jarry et al., 2014), this feature is advantageous, especially in diseases affecting tissues such as bone in which biopsy may be problematic (Hackl et al., 2016). Furthermore, circulating miRNAs can be detected with reliable methods based on polymerase chain reaction (PCR); reverse transcription quantitative PCR (RT-qPCR) is the most widely used owing to its high sensitivity, specificity, and reproducibility (Bustin and Nolan, 2004). Another important advantage of miRNAs as biomarkers is their stability in biofluids due to their encapsulation in extracellular vesicles (ectosomes or exosomes) and in high-density lipoproteins (HDL) and their association with proteins (Argonaute2 or nucleophosmin); miRNAs packaging is correlated with the way they are taken up by target cells (Arroyo et al., 2011; Chen et al., 2012; Li et al., 2012). miRNAs concentration in plasma, as evaluated by qPCR, is highly variable. El-Hefnawy et al. (2004) detected miRNAs concentration in the range of 1–10 μg/L, while Weber et al. (2010) reported a median concentration of 308 μg/L. Differences among healthy humans are physiological and any variation in blood processing conditions can affect circulating miRNA levels (Mitchell et al., 2008; Kroh et al., 2010; Cheng et al., 2013a).

miRNAs as Biomarkers: Weaknesses

Pre-Analytical Issues in miRNA Evaluation

In the pre-analytical phase, two sets of variables can affect miRNAs evaluation: patient-related and sampling-related factors.

Patient-related factors: lifestyle habits and diseases

Among patient-related factors, lifestyle habits and diseases affect circulating miRNA levels. Studies have shown that cigarette smoking (Takahashi et al., 2013), physical activity (Baggish et al., 2011; Faraldi et al., 2019), diet (Witwer, 2012), vitamin D levels (Bellavia et al., 2016; Bellavia et al., 2019), and head-down tilt (HDT) bed rest (Ling et al., 2017) can modify the level of a specific miRNA in circulation, whereas gender does not seem to significantly contribute to total variability (Chen et al., 2008). Also, miRNA levels are affected by circadian rhythm (Shende et al., 2011).

The total amount of circulating miRNAs is reduced in chronic kidney disease patients (Neal et al., 2011), while its correlation with liver disease is unknown (Hackl et al., 2016). As a consequence, any clinical study validating a panel of circulating miRNAs as biomarkers must follow pre-analytical protocols with strict criteria for sample collection (preferentially in the morning) and for patient inclusion and exclusion (type of diet, glomerular filtration rate, and fasting time before sample collection) to minimize the effect of variables on the validation process (Hackl et al., 2016).

Sampling-related factors: source/matrix, sample collection, and handling

A key step in the validation of a novel biomarker is selection of the correct matrix (Livesey et al., 2008; Kavsak and Hammett-Stabler, 2014). Serum and plasma miRNAs evaluated in the same blood sample are stable, and measurements in healthy individuals are reproducible, consistent, and linkable (Chen et al., 2008; Mitchell et al., 2008). In blood sample collection and handling, phlebotomy is the chief source of variability and contamination with non-circulating miRNAs (Kroh et al., 2010; Cheng et al., 2013a). In detail, miRNA quantification can be affected by the type of collection tube and anticoagulant coating, in addition to blood cell count, needle gauge (Kroh et al., 2010), and hemolysis (Kirschner et al., 2011). Since the total amount of miRNAs contained in cells is considerably higher than in circulation, quantification of circulating miRNAs can be affected by the signal coming from non-circulating miRNA contamination (e.g., the skin contaminant within the needle). In addition, miRNAs can be released by activated platelets or by hemolytic erythrocytes (Kirschner et al., 2011; Willeit et al., 2013). Another often unconsidered source of variability is tourniquet application, together with clenching the fist and maintaining it closed, that can alter blood levels of electrolytes, muscle enzymes, free hemoglobin, water, and low-molecular-weight molecules. Also at the needle insert site the concentration of some blood analytes may be increased (Lima-Oliveira et al., 2013; Lima-Oliveira et al., 2016). For the collection of plasma samples, it is important to use the right anticoagulant: heparin, potassium ethylendiaminotetraacetate (K2/K3 EDTA), sodium fluoride/potassium oxalate (NaF/KOx), or sodium citrate. Heparin (Garcia et al., 2002; Boeckel et al., 2013) and sodium citrate are not recommended for RT-qPCR-based miRNA quantification because they alter the activity of the enzymes used in PCR-based assays (Hackl et al., 2016). Conversely, EDTA is considered the right choice for PCR-based miRNA evaluation because it is easily removed from the PCR mastermix (Zampetaki and Mayr, 2012). Alternatively,

NaF/KOx may be used when EDTA is not available, although it can increase the miRNA detection rate (Kim et al., 2012). Centrifugation speed and length to separate plasma can affect miRNAs detection in EDTA-plasma possibly due to platelet-derived miRNAs (Cheng et al., 2013a), while miRNAs evaluation in serum samples is less sensitive to this process (Hackl et al., 2016). miRNAs in blood samples are stable up to 24 h at room temperature (Mitchell et al., 2008) due to their association with proteins or extracellular vesicles. This is important in clinical routine, especially when unexpected delays prolong turnaround time. Interestingly, miRNAs are reported to be stable also in extreme conditions (e.g., low and high pH) or after repeated freezing/thawing cycles (Chen et al., 2008). The ongoing discovery of novel miRNAs, together with the limited number of stability tests, calls for the need of standardized protocols in sample collection and handling in order to minimize pre-analytical sources of error (Cheng et al., 2013a). Samples can be stored for decades at low temperatures (i.e., < −70°C), which facilitates the retrieval of reliable data in retrospective studies (Zampetaki and Mayr, 2012).

Analytical and Post-Analytical Issues in miRNA Evaluation

In their study comparing 12 commercially available platforms for evaluating miRNA expression levels (7 PCR-based, 3 microarrays, and 2 next generation sequencing [NGS] technologies), Mestdagh et al. (2014) observed marked differences between the platforms. Because different technologies are often used during the validation process, platform choice will affect a method's reproducibility and specificity. For any platform combination, the average validation rate for deregulated miRNA expression is 54.6%, indicating that screening studies and validation studies on different platforms and/or technologies must be performed. Sensitivity is more technology-correlated, with qPCR platforms showing the best score and, as a consequence, higher accuracy and more reliable results. These observations suggest that analytical protocols and platforms must be the same for the discovery and the validation of a biomarker and that further efforts are required to aid in the migration to a final commercial platform (Hackl et al., 2016).

The major post-analytical issues in miRNAs evaluation are data normalization and choice of the right reference gene. Presently, there is no consensus on either issue. The amount of miRNAs in a biofluid is expressed in relative rather than absolute terms by volume unit. This makes it hard to compare results across different labs or across different studies performed in the same lab (Nelson et al., 2008; Hackl et al., 2016). The most common normalization methods for miRNAs expression of RT-qPCR data (reviewed in Faraldi et al., 2018) are based on: exogenous synthetic oligonucleotides; endogenous reference genes; and the average of all the miRNA expressed. The right choice of normalization strategy is crucial to reduce analytical variability and to obtain reliable and reproducible results. Exogenous reference genes are non-human synthetic oligonucleotides usually added to the analyzed biological sample to monitor the efficiency and quality of RNA processing.

In miRNAs quantification, the normalization strategies adopted for RT-qPCR data calculation are based on the use of a single reference gene (i.e., cel-miR-238, cel-miR-39, cel-miR-54) (Ho et al., 2010; Wang et al., 2015; Yang et al., 2017) or on the average of multiple reference exogenous oligonucleotides (Mitchell et al., 2008; Sourvinou et al., 2013). These normalization methods have an important limitation, however: unlike endogenous miRNAs, exogenous oligonucleotides are not affected by pre-analytical variables, consequently, they reduce the analytical but not the pre-analytical variability. The use of one or more endogenous reference genes satisfies this criterion because the genes are affected by the same pre-analytical variables as the same analytical procedures of the target miRNA(s); therefore, this is the most suitable normalization strategy for miRNAs data from RT-qPCR-based quantification techniques (Faraldi et al., 2018).

In human samples, the most commonly used endogenous reference gene is has-miR-16 (Faraldi et al., 2018), but several studies have shown very variable expression between cases and controls and the effect of hemolysis on its levels in blood samples (Hu et al., 2012; Liu et al., 2012; Kirschner et al., 2013). Also for endogenous sequences, the normalization method based on the use of multiple reference genes, identified with the aid of informatics tools, is thought to reduce post-analytical variability (Vandesompele et al., 2002; Andersen et al., 2004). With this procedure, however, the miRNAs set as reference cannot be used later in the analysis as targets (Faraldi et al., 2018). Finally, for large amounts of data or in the absence of an *a priori* reference gene, a commonly applied strategy is to calculate the average expression of all the evaluated endogenous miRNA (Mestdagh et al., 2009). Based on these considerations, it is of key importance to standardize the normalization method by determining the most stable reference gene(s) in each experimental setting (Faraldi et al., 2018). Recently, we demonstrated large differences in results obtained by applying different normalization strategies to RT-qPCR data from a panel of 179 circulating miRNAs. Based on analysis of the between-assay coefficients of variation (CV) and of the CV distribution frequencies, we defined the normalization of a specific miRNA (hsa-miR-320d) as the best strategy in that specific setting (Faraldi et al., 2019).

Specific guidelines to standardize pre-analytical, analytical, and post-analytical variables are desirable in order to obtain reliable and comparable miRNA expression data and to accelerate the definitive clinical implementation of miRNAs-based tests.

MIRNAS AS BIOMARKERS FOR BONE DISEASES

While the multiple roles exerted by tissue and exome/microvesicle-associated miRNAs in bone pathophysiology have been identified and validated, the clinical usefulness of circulating miRNAs in skeletal and muscle-skeletal diseases has not yet been established. This is because studies so far have been designed with a mechanistic purpose in mind and not for identifying circulating miRNAs with diagnostic/prognostic abilities for bone fracture risk or treatment response (Hackl et al., 2016). The potential role of circulating miRNAs as biomarkers for the early identification of altered bone metabolism ranks

high on the clinical research agenda, given the aging population and the growing incidence of age-associated diseases (e.g., metabolic bone diseases and osteoporosis) and the related risk of bone fracture. Reliable diagnostic tools that can prognosticate a subject-specific risk of disease onset or, if already overt, a subject-specific risk of progression and response to therapy are currently lacking. Furthermore, the natural history of age-associated bone diseases is, as never before, tied to a plethora of subject-specific variables. miRNAs and their circulating fraction hold promise: as epigenetic modifiers of gene expression they act much more upstream of the expression process than classical protein markers. This means that changes in their expression, which are likely to be mirrored by changes in their circulating levels, are effective far before their translation into metabolic and structural changes (Materozzi et al., 2018).

Circulating miRNAs and Postmenopausal Osteoporosis

Osteoporosis (OP), one of the most prevalent bone diseases, is characterized by impaired bone strength and quality that increase the risk of bone fracture (NIH, 2001). Currently, dual energy X-ray absorptiometry (DXA) is the diagnostic gold standard, while bone turnover markers are useful in framing the metabolic activity of bone cells [e.g., C-terminal cross-link (CTx), N-terminal pro-peptide of type I collagen (PINP), parathyroid hormone (PTH), bone alkaline phosphatase (BAP), osteocalcin, and tartrate-resistant acid phosphatase 5b (TRAP5b), pyridonline/deoxypyridinoline] and in evaluating the effectiveness of anti-resorptive therapies (Lombardi et al., 2012; Vasikaran and Chubb, 2016). Although valuable, these diagnostic tools have several practical flaws that partially limit their utility: on the one hand, radiological methods can reveal only already established bony architectural modifications, which take several weeks or months to become detectable, and on the other, bone turnover markers are not fully specific for either bone or the metabolic process they are associated with (i.e., formation or resorption) (Lombardi et al., 2012).

Despite limitations in pre-analytical, analytical, and post-analytical standardization, miRNAs still have enormous potential in this setting. Indeed, based on their role as highly sensitive fine-tuners of biological processes, when assayed in combination with conventional diagnostics, they may give a more detailed clinical framing and a prompt measure of response to therapy (Faraldi et al., 2018; Sansoni et al., 2018). This is particularly desirable in complex syndromic conditions, such as OP, in which the prognosis (i.e., bone fracture) depends not only upon the bony metabolic status but also on the whole-body metabolism. Circulating miRNAs can much better describe such a complex network. The still limited information about the role of miRNAs in OP is derived from different types of human samples [serum, circulating monocytes or bone marrow-derived mesenchymal stem cells (BM-MSCs), and bone tissue] obtained from patients of different ethnic groups with low bone mineral density (BMD) or bone fractures and compared with healthy controls or osteoarthritis (OA) patients. Furthermore, differences in quantification platforms and normalization processes make it very hard to compare the study data.

Early evidence that OP correlates with altered expression of circulating miRNAs stems from a microarray analysis of 365 miRNAs in human circulating monocytes collected from postmenopausal Caucasian women with either low or high BMD. Of the 365 miRNAs screened by RT-qPCR analysis, only miR-133a was found significantly upregulated in the low-BMD subjects compared with their normal BMD counterparts (Wang et al., 2012). Using the same experimental protocol, the same authors found another marginally expressed miRNA associated with low BMD: miR-422a (Cao et al., 2014). Supporting the hypothesis for their tissue-specificity, subsequent analysis of miR-133a and miR-422a expression in isolated circulating B cells derived from the same subjects disclosed no difference between the two groups (Wang et al., 2012; Cao et al., 2014). Based on these results, the authors speculated that these two miRNAs might be monocyte-specific biomarkers for postmenopausal OP. Mature miR-133a is transcribed from two different loci (18q11.2 and 20q13.33). It was previously described as an inhibitor of osteoblast differentiation by directly targeting RUNX2 in murine pre-myogenic C2C12 and pre-osteoblastic MC3T3-E1 cells (Li et al., 2008; Zhang et al., 2011b). The miR-422a expression level in osteoblast-like cells was described to be decreased after treatment with peptide-15, a factor that increases bone development (Palmieri et al., 2008). Since monocytes are osteoclast precursors, a bioinformatics analysis has highlighted three osteoclast-related potential target genes for miR-133a (CXCL11, CXCR3, and SLC39A1) and five for miR-422a (CBL, CD226, IGF1, PAG1, TOB2) (Wang et al., 2012; Cao et al., 2014). The latter studies, however, suffered from several limitations: limited sample size (10 subjects per group); no evidence of a correlation between miR-133a or miR-422a and target gene expression; and no information about the stem-loop arm of origin of these miRNAs.

In another study, Chen et al. (2014a) evaluated the expression profile of 721 human miRNAs in CD14+ mononuclear cells from peripheral blood (PBMCs) collected from postmenopausal OP women. They found seven differentially expressed miRNAs compared with the non-OP group: four (miR-218, miR-503, miR-305, and miR-618) were downregulated and three (miR-107, miR-133a, and miR-411) were upregulated. Also, miR-133a was confirmed as upregulated in circulating monocytes from postmenopausal OP women (Wang et al., 2012); however, only miR-503, the most deregulated one, was validated by RT-qPCR, and its anti-osteoclastogenic effects were investigated *in vivo* and *in vitro*. Overexpression of miR-503, after pre-miR-503 transfection in OP-derived CD14+, drastically inhibited M-CSF/RANKL-induced osteoclastogenesis, while its suppression by antagomiR-503 promoted osteoclast differentiation. The authors identified and validated RANK mRNA as a target for miR-503. Furthermore, in ovariectomized (OVX) mice, antagomiR-503 increased RANK protein expression, and promoted bone loss and resorption, whereas agomiR-503 prevented bone loss and resorption (Chen et al., 2014a). Because miR-503 downregulation has a key role in postmenopausal OP onset, it may be a target for new therapeutic strategies for OP.

Using a different approach, a study evaluated the miRNA profile differences in human *bone marrow*-derived mesenchymal stromal cells (BM-MCSs) from OP patients and non-OP controls. In this case, 1,040 miRNAs were screened using a microarray in BM-MCSs collected from healthy premenopausal women (control group, n = 5) and postmenopausal OP women (n = 5) (Yang et al., 2013). Following RT-qPCR validation, miR-21 was found downregulated in the OP women, as confirmed in the MSCs from OVX mice. Further experiments revealed that Spry1 negatively regulates fibroblast growth factor (FGF) and extracellular signal-regulated kinase–mitogen-activated protein kinase (ERK-MAPK) signaling pathways and that it is directly targeted by miR-21. As a consequence, the TNFα-mediated inhibition of miR-21 may impair bone formation, as observed in OP induced by estrogen deficiency. This mRNA seems to be a main regulator of osteoblastic differentiation of MSCs and in postmenopausal OP onset (Yang et al., 2013). Moreover, osteoclast precursors express miR-21, which is upregulated during TNF-α/RANKL-induced osteoclastogenesis (Sugatani et al., 2011; Kagiya and Nakamura, 2013). miR-21 expression is upregulated by the osteoclastogenesis transcription factor c-Fos that binds the miR-21 promoter (Kagiya and Nakamura, 2013) which, in turn, downregulates c-Fos inhibitor-programmed cell death 4 (PDCD4). This positive c-Fos/miR-21/PDCD4 feedback loop regulates and promotes RANKL-induced osteoclastogenesis (Sugatani et al., 2011). In addition, miR-21 is involved in estrogen-induced osteoclasts apoptosis: estrogens inhibit miR-21 expression by inducing Fas-ligand (FasL), another miR-21 target, which in turn inhibits osteoclastogenesis and promotes osteoclast apoptosis (Garcia Palacios et al., 2005; Sugatani and Hruska, 2013).

More recent studies have been focused on whole blood, serum or plasma miRNA profiling in patients with or without OP. Circulating levels of miR-133a, miR-146a, and miR-21 have been assayed by RT-qPCR in plasma samples of Chinese postmenopausal women, grouped as normal, osteopenic or OP. miR-21 was downregulated while miR133a was upregulated in the OP and osteopenic women compared with the controls and both correlated with BMD; miR-146a was unchanged (Li et al., 2014). miR-21 was found downregulated in the BM-MCSs of postmenopausal OP women (Yang et al., 2013), while the monocyte expression of miR-133a was associated with low BMD values (Wang et al., 2012). A study investigated the discriminatory potential between OP and osteopenia of six miRNAs (miR-130b-3p, miR-151a-3p, miR-151b, miR-194-5p, and miR-590-5p) which were found upregulated in OP. Of these six, miR-194-5p was the most upregulated and its expression negatively correlated with BMD. The association between miR-194-5p circulating levels and BMD was later confirmed in a wider cohort of Chinese postmenopausal women with normal, osteopenia, and OP ranges of BMD. The study also reported that miR-194-5p may influence the TGF-β and Wnt signaling pathways, thus acting as a critical factor in the pathophysiology of postmenopausal OP (Meng et al., 2015).

The overexpression of miR-194-5p in mice BM-MSCs was correlated with osteogenesis by targeting both COUP-TFII (chicken ovalbumin upstream promoter-transcription factor II) (Jeong et al., 2014) and STAT1 (signal transducer and activator of transcription 1) (Li et al., 2015b). In parallel, among other 851 miRNAs, miR-27a was validated as the most downregulated one in the serum of postmenopausal OP women compared with their healthy counterparts (You et al., 2016). The MSCs collected from these OP patients displayed an increased adipogenic potential at the expense of osteoblast formation. During osteogenesis, miR-27a is upregulated in MSCs, whereas the opposite occurs during adipogenesis; and indeed, miR-27a silencing in mice impairs bone formation. Myocyte enhancer factor 2c (Mef2c), a transcription factor involved in developmental processes, has been identified and validated as a miR-27a target gene (You et al., 2016). Consistent with previous observations (Lin et al., 2009; Wang and Xu, 2010; Pan et al., 2014), miR-27a expression, is essential for osteoblastic differentiation of MSCs and its downregulation *in vivo* has been associated with bone loss. Bedene et al. (2016) identified, among other nine miRNAs, miR-148a-3p as a potential biomarker for postmenopausal OP based on its significantly higher levels in the plasma samples from OP subjects compared with controls. In CD14+ PBMCs, the RANKL-induced osteoclast differentiation promotes miR-148a expression dependent on the repression of V-maf musculoaponeurotic fibrosarcoma oncogene homolog B (MAFB), a transcription factor whose expression inhibits osteoclastogenesis (Cheng et al., 2013b). miR-148-3p has been found upregulated also in CD14+ PBMCs of patients with systemic lupus erythematous (SLE) in which it was correlated with reduced BMD. Furthermore, treatment of OVX mice with antagomiR-148a slowed bone resorption and increased bone mass (Cheng et al., 2013b). The expression levels of the nine miRNAs assayed by Bedene et al. (2016) revealed that plasma miR-126-3p is also positively associated with BMD at the distal forearm and that miR-423-5p plasma levels are negatively correlated with the 10-year probability of bone fracture in OP.

Using a different approach, Chen et al. (2016) screened a wide range of miRNAs in serum samples from OP mice in order to identify the most stable reference gene (miR-25-3p) for use in data normalization in humans. Fifteen of the screened miRNAs found differentially expressed in the OP mice were then investigated in serum samples from postmenopausal women (7 osteopenic, 10 OP, and 19 healthy women). miR-30b-5p was significantly lower in both the osteopenia and OP samples, while miR-103-3p, miR-142-3p, and miR-328-3p were significantly lower in the OP group only compared with the healthy subjects. The role of miR-103-3p and miR-30b-5p in bone physiology has been validated in *in vitro* studies of osteogenesis: miR-30b-5p expression, whose target is Runx2, decreases during late-stage osteoblast differentiation (Eguchi et al., 2013), while miR-103-3p inhibits osteoblasts differentiation and proliferation by directly targeting Runx2 (Zuo et al., 2015) and Cav1.2 (Sun et al., 2015), respectively. Despite the limited sample size, the serum levels of these four miRNAs in OP patients were positively correlated with BMD. The ROC analysis revealed their diagnostic potential for OP based on the following AUC–sensitivity–specificity values: 0.800–80%–72.2% (miR-103-3p), 0.789–70%–79.0% (miR-142-3p), 0.793–70.6%–79.0% (miR-30b-5p), and 0.874–80%–100% (miR-328-3p) (Chen et al., 2016).

In a study series, circulating monocytes from 12 postmenopausal Mexican-Mestizo women, divided in normal (control group) and OP groups were assayed using a microarray platform for the expression profile of 2,578 miRNAs. The results showed that the three most upregulated miRNAs in the OP group were miR-1270, miR-548x-3p, and miR-8084, while the three most downregulated were miR-6124, miR-6165, and miR-6824-5p. Among the upregulated miRNAs, only miR-1270 was further validated. Based on bioinformatics analysis, nine genes have been identified as possible targets of miR-1270, and RT-qPCR finally validated the interferon regulatory factor-8 (IRF8) gene, an inhibitor of osteoclastogenesis (Zhao et al., 2009; Jimenez-Ortega et al., 2017; Saito et al., 2017), which was significantly downregulated in the OP group. The same research team discovered another monocytic miRNA, miR-708-5p, as a potential biomarker for postmenopausal OP. Next generation sequencing (NGS) of the 46 miRNAs found differentially regulated in the two groups revealed that miR-708-5p and miR-3161 were the two most upregulated in the OP group, whereas miR-4422 and miR-939-3p were the two most downregulated. These four miRNAs were then assayed using RT-qPCR, but only miR-708-5p was validated as it was found significantly upregulated in OP patients compared with controls. Bioinformatics analysis of miR-708-5p disclosed ten potential targets involved in osteoclastogenesis, only five of which (AKT1, AKT2, PARP1, FKBP5, and MP2K3) were effectively downregulated in the OP subjects compared with controls (De-La-Cruz-Montoya et al., 2018). The major limitations besides the small sample size in these two studies were the use of different quantification platforms (microarray and NGS) in preliminary screening of differential miRNA expression and the use of two different normalization strategies for RT-qPCR data analysis. These limitations make it difficult to correlate the data. In any case, miRNA-708-5p and miR-1270 may be suitable biomarkers for postmenopausal OP but require an independent validation study with a larger sample using the same protocol for data quantification and analysis.

The last paper published by this research group is the most complete work to date. The potential of miRNAs as biomarkers for OP was evaluated in serum samples (Ramirez-Salazar et al., 2018). The study was divided in two experimental parts: in the discovery stage, 40 postmenopausal Mexican-Mestizo women (grouped into OP subjects and healthy controls) were recruited, while the validation stage comprised Mexican-Mestizo women with OP, osteopenia, and bone fractures, plus healthy postmenopausal Mexican-Mestizo women. In the discovery stage, microarray analysis of 754 serum miRNAs identified seven miRNAs (miR-1227-3p, miR-139-5p, miR-140-3p, miR-17-5p, miR-197-3p, miR-23b-3p, and miR-885-5p) in which the levels were significantly higher in the OP than in the healthy subjects. Only the three most upregulated (miR-140-3p, miR-23b-3p, and miR-885-5p) were used in the validation stage. The study confirmed by RT-qPCR the higher serum levels of miR-140-3p and miR-23b-3p in the groups with osteopenia, OP or bone fracture, and higher levels of miR-885-5p in the osteopenia group than in healthy subjects. ROC analysis for miR-140-3p and miR-23b-3p, in which their ability to discriminate between OP and healthy women was evaluated, demonstrated that the two miRNAs might be good candidates as biomarkers for BMD loss: AUC of 0.84, 0.96, and 0.92 for miR-140-3p in the osteopenia, OP, and bone fracture group, respectively, compared with the healthy controls, and AUC of 0.73, 0.69, and 0.88, respectively, for miR-23b-3p. Furthermore, miR-140-3p and miR-23b-3p were significantly correlated with BMD in each cohort. Target genes databases predicted AKT1, AKT2, AKT3, BMP2, FOXO3, GSK3B, IL6R, PRKACB, RUNX2, and WNT5B as bone-related genes potentially targeted by miR-140-3p and miR-23b-3p. Other potential osteogenic related target genes have been validated *in vitro* and *in vivo*: SMAD3 (Liu et al., 2016) and RUNX2 (Deng et al., 2017) for miR-23b-3p, and BMP2 (Hwang et al., 2014) for miR-140-5p. The study underlined the importance of miR-140-3p and miR-23b-3p as biomarkers of bone loss and risk of fracture, despite the small sample size especially of the control group.

Table 1 presents information about circulating miRNAs associated with OP.

miRNAs, Bone Fragility, and Bone Fracture Risk in Postmenopausal Women

Bone fragility and fractures are the clinically relevant consequences of OP and have a negative impact on quality of life. Considering the objective limit of bone biopsy in healthy individuals, studies have compared the miRNA expression profile of OP bone with osteoarthritis (OA) samples as control. Thirteen of 760 miRNAs assayed by microarray cards were found differentially expressed in bone specimens from the femur heads of eight women with OP hip fracture compared to the femur heads from eight women with severe hip OA but without OP hip fracture, in seven of which the miRNAs were overexpressed in OP bones. In the following replication stage, the results showed that miR-518f was overexpressed and miR-187 downregulated in OP compared with OA bone (Garmilla-Ezquerra et al., 2015). Finally, the expression profile of 1,932 miRNAs was compared between fresh femoral neck trabecular bone from postmenopausal women with OP hip fracture and from postmenopausal women with OA non-OP hip fracture (control group). Following validation, only two (miR-320a and miR-483-5p) of the 82 miRNAs differently expressed between the two groups were significantly overexpressed in the OP vs. the OA samples (De-Ugarte et al., 2015). miRNA-320a targets RUNX2 and β-catenin (Yu et al., 2011; Sun et al., 2012), while miRNA-483-5p downregulates IGF2 expression in OP-derived human osteoblast cultures (De-Ugarte et al., 2015).

To identify circulating miRNAs as biomarkers for OP fracture, Seeliger et al. (2014) assayed a panel of 83 serum miRNAs in OP and non-OP patients with either femoral neck or pertrochanteric fracture. Eleven miRNAs (miR-100-5p, miR-122a-5p, miR-124-3p, miR-125b-5p, miR-148a-3p, miR-21-5p, miR-223-3p, miR-23-3p, miR-24-3p, miR-25-3p, and miR-27a-3p) were found at significantly higher levels in the OP sera. Together with miR-93 and miR-637, these miRNAs were subsequently validated in another set of serum samples: nine miRNAs (miR-100, miR-122a, miR-124a, miR-125b, miR-148a, miR-21, miR-23a, miR-24, and miR-93) were significantly higher in the OP sera than in the controls and they were proposed as markers to differentiate OP from non-OP bone fracture. Interestingly, miR-21 was previously

TABLE 1 | miRNAs related to postmenopausal OP.

Study	Study design	Biomarker source	Sample handling	Quantification platform	Evaluated miRNA	Normalization strategy	Validated miRNA biomarker	Potential target gene	AUC-Sensitivity(%)-Specificity(%)	Limits
(Wang et al., 2012)	20 PM Caucasian women (age 57-68 years): 10 with low BMD (hip/spine Z-score < -0.84); 10 with high BMD (hip/spine Z-score > 0.84)	Circulating monocytes	Monocytes separated by density gradients in UNI-SEP tubes (sodium metrizoate 9.6% and polysucrose 5.6% with 1.077 g/ml density), and isolated using a negative isolation kit	Screening: TaqMan Human MicroRNA Array v1.0 Validation: TaqMan RT-qPCR	Screening: 365 miRNAs tested Validation: miR-133a and miR-382	RNU48	↑miR-133a in low vs. high BMD group	CXCR3, CXCL11, and SLC39A1 (identified for miR-133a using miRDB and TargetScan database but not validated)	/	Small sample size; no significant correlation between the expression level of miR-133a and the potential target genes; no information about the stem-loop arm of miRNA origin; no ROC analysis
(Cao et al., 2014)	21 PM Caucasian women (age 57-68 years): 10 with low BMD (hip/spine Z-score < -0.84); 10 with high BMD (hip/spine Z-score > 0.84)	Circulating monocytes	Monocytes separated by density gradients in UNI-SEP tubes (sodium metrizoate 9.6% and polysucrose 5.6% with 1.077 g/ml density), and isolated using a negative isolation kit	Screening: TaqMan Human MicroRNA Array v1.0 Validation: TaqMan RT-qPCR	Screening: 365 miRNAs tested Validation: miR-27b, miR-422a, miR-151, and miR-152	RNU48	↑miR-422a in low vs. high BMD group	CD226, CBL, IGF1, TOB2, and PAG1 (identified for miR-422a using TargetScan database but not validated)	/	Small sample size; no significant correlation between miR-422a and the evaluated target genes; no information about the stem-loop arm of miRNA origin; no ROC analysis
(Chen et al., 2014a)	31 Chinese PM women with OP and 30 healthy women (age 50-59 years).	PBMCs CD14+	Ficoll-Paque separation step and CD14 antibody-coated magnetic cell sorting MicroBeads used for buffy coat PBMCs isolation and $CD14^+$ purification, respectively	Screening: MicroRNA microarray by LC Sciences Validation: SYBR Green RT-qPCR	Screening: 721 miRNAs tested Validation: miR-503	snRNU6	↓miR-503 in OP group vs. non-OP group	RANK (validated as miR-503 target gene)	/	Small sample size; no information about the stem-loop arm of miRNA origin; no ROC analysis
(Yang et al., 2013)	5 OP PM women (age 53-63 years) and 5 premenopausal women (age 39-45 years)	BM-MCSs	Percoll density gradient centrifugation methodology obtaining BM-MCSs from the BM	Screening: LC Sciences microarray platform Validation: RT-qPCR	Screening: 1040 miRNAs tested Validation: miR-21	snRNU6	↓miR-21 in PM OP group vs. non-OP group	SPRY1 (identified for miR-21 using Target Scan 6.0 and Pic Tar databases and validated by *in vitro* experiments)	/	Small sample size; no information about the stem-loop arm of miRNA origin; no ROC analysis
(Li et al., 2014)	40 PM Chinese women with normal, 40 with OP, and 40 with osteopenia range BMD (age 46-69 years)	Cell-free plasma	Plasma obtained from fasting blood samples and stored in liquid nitrogen	miRCURY LNA RT-qPCR	miR-21, miR-133a, and miR-146a	miR-16	↓ miR-21 and ↑ miR-133a in OP and osteopenia groups vs. control	/	/	Small sample size; no information about the used anticoagulant; small sample size; arbitrary decision of the reference gene; no evaluation of the target genes; no information about the stem-loop arm of miRNA origin; no ROC analysis

(*Continued*)

TABLE 1 | Continued

Study	Study design	Biomarker source	Sample handling	Quantification platform	Evaluated miRNA	Normalization strategy	Validated miRNA biomarker	Potential target gene	AUC-Sensitivity(%)-Specificity(%)	Limits
(Meng et al., 2015)	Discovery cohort: 25 PM women with OP and 23 PM Chinese women with osteopenia (age 59-70 years) Validation cohort: 24 PM Chinese women with normal, 32 with OP and 30 with osteopenia range BMD (age 59-70 years)	Whole blood	Blood samples lysed using RBC lysis solution and centrifuged for 10 min at 450g	Discovery: Agilent Human miRNA microarray followed by SYBR Green RT-qPCR Validation: SYBR Green RT-qPCR	Discovery cohort: comprehensive miRNA expression analysis (Microarray); miR-130b-3p, miR-151a-3p, miR-151b, miR-194-5p, miR-590-5p, and miR-660-5p (RT-qPCR) Validation cohort: miR-194-5p	snRNU6	↑ miR-130b-3p, miR-151a-3p, miR-151b, miR-194-5p, and miR-590-5p in OP vs. osteopenia (Discovery cohort)↑ miR-194-5p in OP and osteopenia vs. control (Validation cohort)	/	/	Small sample size; no evaluation of the target genes; no ROC analysis
(You et al., 2016)	155 PM Chinese women with PM OP (n = 81, age 51-62 years) or healthy (n = 74, age 40-46 years)	Cell-free serum and BM-MSCs	/	Screening: Agilent Human miRNA Microarray Validation: TaqMan RT-qPCR	Screening: 851 miRNAs Validation: miR-27a	snRNU6	↓ miR-27a in OP vs. control	Mef2c (predicted for miR-27a using TargetScan and PicTar database and validated by *in vitro* studies)	/	The mean age of OP and healthy women is significantly different; no information about the stem-loop arm of miRNA origin; no ROC analysis
(Bedene et al., 2016)	74 PM women (age 55-65 years): 57 controls and 17 OP based on femoral neck/lumbar spine/ total hip T-score ≤–2.5 SD	Cell-free plasma	Blood samples collected in EDTA tubes, centrifuged at 2800 rpm and 4°C for 10 min, then further centrifuged at 9600g and 4°C for 15 min. Plasma samples stored at -80°C	SYBR Green RT-qPCR	miR-7d-5p, miR-7e-5p, miR-30 d-5p, miR-30e-5p, miR-126-3p, miR-148a-3p, miR-199a-3p, miR-423-5p, and miR-574-5p	Combination of let-7a-5p and miR-16-5p as identified by Normfinder	↑ miR-148a-3p in OP vs. control	/	/	Small sample size of the control group; no evaluation of the target genes; no ROC analysis
(Chen et al., 2016)	36 PM women: 19 HC, 7 osteopenic, 10 OP	Cell-free serum	Serum obtained by centrifuging blood samples in two steps: for 10 min at 2000g and 4°C and for 20 min at 12000g and 4°C. Serum stored at -80°C	SYBR RT-qPCR	miR-30a-5p, miR-30e-5p, miR-425-5p, miR-142-3p, miR-191a-3p, miR-215, miR-29b-3p, miR-30b-5p, miR-26a-5p, miR-345-5p, miR-361-5p, miR-185-5p, and miR-103-3p	NormFinder and GeNorm identified miR-25-3p as the most stable reference gene in mice models of OP	↓ miR-30b-5p in both osteopenic and OP vs. HC↓ miR-103-3p, miR-328-3p, and miR-142-3p OP vs. HC	/	0.793 (miR-30b-5p) for both OP and osteopenia vs. HC0.793-70.6-79.0 (miR-30b-5p), 0.800-80-72.2 (miR-103-3p), 0.789-70-79.0 (miR-142-3p) and 0.874-80-100 (miR-328-3p) for OP vs. HC	Different number of subjects recruited in the 3 groups; the reference gene for humans was identified in mice models; no evaluation of the target genes

(*Continued*)

TABLE 1 | Continued

Study	Study design	Biomarker source	Sample handling	Quantification platform	Evaluated miRNA	Normalization strategy	Validated miRNA biomarker	Potential target gene	AUC-Sensitivity(%)-Specificity(%)	Limits
(Jimenez-Ortega et al., 2017)	PM Mexican-Mestizo women: 6 with normal (control group) and 6 with OP hip BMD (age 63-85 years)	PBMCs	Histopaque-1077 kit used for obtaining PCMCs by density gradients. CD14+ obtained by density gradient centrifugation for 30 min at 400g and RT and magnetic bead isolation. negative isolation kit EasySep Human Monocyte Enrichment used for naive monocyte isolation	Screening: Affymetrix GeneChip Human U133 Plus 2.0 Array Validation: TaqMan RT-qPCR	Screening: 2.578 miRNAs tested Validation: miR-1270, miR-548x-3p, and miR-8084	Screening: quantile normalization Validation: RNU44	↑miR-1270 in OP group vs. control group	IRF8 (identified for miR-1270 using PITA v5.0, microRNA.org, miRWalk v2.0, miRDB, and TargetScan Human v7.0 database and validated in study)	/	Small sample size; no ROC analysis
(De-La-Cruz-Montoya et al., 2018)	PM Mexican-Mestizo women: 7 with normal (control group) and 7 with OP hip BMD (age 63-85 years)	Human PBMCs	Blood collected in CPT tubes and PBMCs obtained. CD14+ cells enriched by negative selection (EasySep kit)	Screening: Illumia NextSeq 500 Validation: TaqMan RT-qPCR	Validation: miR-708-5p, miR-3161, miR-939-3p, and miR-4422	Validation: RNU44 and RNU48	↑miR-708-5p in osteoporosis group vs. control group	AKT1, AKT2, FKBP5, PARP1, and MP2K3 (identified for miR-708-5p using miRTarBase and MiRNet and validated in study)	/	Small sample size; no ROC analysis
(Ramirez-Salazar et al., 2018)	Discovery cohort: 40 PM Mexican-Mestizo women: 20 with normal (controls) and 20 with OP hip BMD (age 63-85 years) Validation cohort: 22 normal, 26 OP, 28 osteopenia, 21 with hip fracture BMD	Cell-free serum	Serum obtained within 1h of collection and stored at -80°C	Discovery stage: TaqMan Array Human MicroRNA A+B Cards Set v3.0 Validation stage: TaqMan RT-qPCR	Screening: 754 miRNAs tested Validation: miR-23b-3p miR-140-3p, and miR-885-5p	snRNU6	↑ miR-23b-3p and miR-140-3p in OP, osteopenia and bone fracture group vs. control↑ miR-885-5p in osteopenia vs. control	AKT1, AKT2, AKT3, IL6R, BMP2, GSK3B, FOXO3, PRKACB, WNT5B, and RUNX2 (identified for miR-23b-3p and miR-140-3p using miRWalk v3 database)	0.84 (miR-140-3p) for osteopenia, 0.96 (miR-140-3p) for OP, and 0.92 (miR-140-3p) for fracture vs. HC0.73 (miR-23b-3p) for osteopenia, 0.69 (miR-23b-3p) for OP, and 0.88 (miR-23b-3p) for fracture vs. HC0.69 (miR-885-5p) for osteopenia vs. HC	No validation of the identified target genes

AKT1, AKT serine/threonine kinase 1; AKT2, AKT serine/threonine kinase 2; AKT3, AKT serine/threonine kinase 3; BMD, bone mineral density; BM-MCSs, bone marrow mesenchymal stem cells; BMP2, bone morphogenic protein 2; CBL, casitas B-lineage lymphoma proto oncogene; CD226, cluster of differentiation 226; CXCL11, chemokine (C-X-C motif) ligand 11; CXCR3, chemokine (C-X-C motif) receptor 3; FKBP5, FK506 binding protein 5; FOXO3, forkhead box O3; FZD3, frizzled-3; GSK3B, glycogen synthase kinase 3 beta; HC, healthy controls; IGF1, insulin-like growth factor 1; IL6R, interleukin 6 receptor; IRF8, interferon regulatory factor-8; Mef2c, myocyte enhancer factor 2 c; MP2K3, mitogen-activated protein kinase kinase 3; OP, osteoporosis; OSX, osterix; PAG1, phosphoprotein associated with glycosphingolipid microdomains 1; PARP1, poly(ADP-ribose) polymerase 1; PBMCs, peripheral blood mononuclear cells; PM, postmenopausal; PRKACB, protein kinase cAMP-activated catalytic subunit beta; RANK, receptor activator of nuclear factor κ B; RANKL, receptor activator of nuclear factor k B ligand; RT, room temperature; RT-qPCR, real-time quantitative polymerase chain reaction; RUNX2, runt-related transcription factor 2; SLC39A1, solute carrier family (zinc transporter), member 1; SPRY1, protein sprouty homolog 1; TOB2, transducer of ERBB2, 2; WNT5B, Wnt family member 5B.

found downregulated in both the BM-MCSs and the plasma of OP patients (Yang et al., 2013; Li et al., 2014); these opposite results could be ascribed to the different experimental protocols used, which identified miRNAs that regulate osteoclast/osteoblast differentiation and activity, as previously demonstrated. miR-21 is highly expressed in osteoclast precursors and it is upregulated in the course of TNF-α/RANKL-induced osteoclastogenesis (Fujita et al., 2008; Kagiya and Nakamura, 2013); it stimulates osteoclastogenesis by overcoming PDCD4-mediated c-Fos inhibition (Fujita et al., 2008; Sugatani et al., 2011), while its expression is inhibited by estrogens (Garcia Palacios et al., 2005; Sugatani and Hruska, 2013). miR-23 and miR-24 belong to the miR-23a~27a~24-2 cluster and act as negative regulators of osteoblast differentiation by targeting SATB2 that cooperates with RUNX2 to induce osteogenesis, while miR-23a also inhibits RUNX2 (Hassan et al., 2010). miR-93 inhibits osteoblast mineralization by targeting OSX (Yang et al., 2012). miR-100 negatively regulates BMPR2, a key osteogenic factor for MSCs (Zeng et al., 2012). The overexpression of miR-125b is associated with impaired osteoblast differentiation and proliferation through the modulation of OSX expression (Mizuno et al., 2008; Chen et al., 2014b). miR-124 is progressively downregulated during RANKL-induced osteoclastogenesis and its overexpression affects the maturation of osteoclast precursors *via* suppression of the key osteoclastogenic factor NFATc1, and their migration *via* inhibition of RhoA/Rac1 (Lee et al., 2013).

Following the identification of nine miRNAs whose circulating levels were higher in OP patients than in controls, Seelinger et al. evaluated their expression in the bone tissues: miR-100, miR-125b, miR-21, miR-23a, miR-24, and miR-25 were upregulated also in the OP bone samples. They defined the potential diagnostic value of these miRNAs by means of ROC curve analysis. All the identified serum miRNAs showed significant AUC, sensitivity and specificity in discriminating OP from non-OP subjects: 0.69–62.9%–61.7% (miR-100), 0.77–74.1%–72.1% (miR-122a), 0.69–61.4%–61.0% (miR-124a), 0.76–76.4%–75.0% (miR-125b), 0.61–62.5%–62.3% (miR-148a), 0.63–61.3%–61.7% (miR-21), 0.63–57.4%–56.7% (miR-23a), 0.63–60.3%–60.4% (miR-24), and 0.68–69.0%–68.3% (miR-93). Consequently, the five miRNAs identified in both tissue and serum samples can be used as biomarkers for OP and related hip fractures (Seeliger et al., 2014).

Another study attempted to search for potential miRNAs marking for OP bone fractures. In the discovery stage, Caucasian women with either OP sub-capital hip fracture (n = 8) or severe hip OA (control group, n = 5), which required arthroplasty, were recruited (Panach et al., 2015). The serum levels of 179 miRNAs were analyzed by RT-qPCR. Among the 42 differently regulated miRNAs, six (miR-122-5p, miR-125b-5p, miR-143-3p, miR-21-5p, miR-210, and miR-34a-5p) were selected for the replication stage. miR-122-5p, miR-125b-5p, and miR-21-5p were significantly higher in the OP bone fracture group than the controls. miR-125b-5p and miR-21-5p have been correlated with bone metabolic indexes (Fujita et al., 2008; Mizuno et al., 2008; Sugatani and Hruska, 2013), and the upregulation of miR-21 was consistent with previous observations (Seeliger et al., 2014). ROC analysis of the diagnostic value of the serum miRNAs revealed that miR-122-5p, miR-125b-5p, and miR-21-5p consistently discriminated between the OP patients with fractures (n = 15) and the controls (n = 12) (AUC 0.87 for miR-122-5p, 0.76 for miR-125-5p, and 0.87 for miR-21-5p) (Panach et al., 2015). Using a similar protocol, Weilner et al. (2015) found three other miRNAs potentially correlated with OP fractures in postmenopausal women (n = 7 in the discovery stage, n = 12 in the validation stage) (miR-22-3p, miR-328-3p, and let-7g-5p) and that the levels were significantly lower in the serum of the cases (n = 7 in the discovery stage, n = 11 in the validation stage). Previous *in vitro* experiments demonstrated that let-7 promotes osteoblastogenesis in MSCs *in vitro*, while it induces bone formation *in vivo*. These effects are mediated by the repression of high-mobility group AT-hook 2 (HMGA2) (Wei et al., 2014). *In vitro* experiments on human unrestricted somatic stem cells (USSC) showed that miR-22-3p is upregulated during osteogenic differentiation and that its potential target is CDK6 (Trompeter et al., 2013). Finally, CD44 is a potential target of miRNA-328-3p in macrophages and it is also expressed in osteocytes (Ishimoto et al., 2014). *In vitro* experiments on MSCs collected from two OP patients with bone fracture confirmed the let-7g-5p-mediated effect and miR-22-3p downregulation, and correlated miR-328-3p repression with reduced ALP activity during osteogenic formation (Weilner et al., 2015).

Recent studies have investigated whether single or combined miRNAs discriminate bone fractures in conditions associated with bone fragility. Kocijan et al. (2016) performed a case-control study to identify serum miRNAs correlated with trauma fractures in postmenopausal OP. Three (miR-152-3p, miR-320a, and miR-335-5p) of the 187 tested miRNAs selected based on previously published studies were significantly higher, whereas sixteen (let-7b-5p, miR-140-5p, miR-16-5p, miR-186-5p, miR-19a-3p, miR-19b-3p, miR-215-5p, miR-29b-3p, miR-30e-5p, miR-324-3p, miR-365a-3p, miR-378a-5p, miR-532-5p, miR-550a-3p, miR-7-5p, and miR-93-5p) were significantly lower in postmenopausal women with bone fracture (n = 10) than in the controls without bone fracture (n = 11). ROC analysis showed that miR-140-5p, miR-152-3p, miR-19a-3p, miR-19b-3p, miR-30e-5p, miR-324-3p, miR-335-5p, and miR-550a-3p had a higher discriminating power between individuals with bone fracture and healthy individuals (AUC> 0.9) than BMD or bone turnover markers. miR-335-3p has been reported to promote osteogenic differentiation by binding and downregulating dickkopf-related protein 1 (DKK1), a soluble antagonist of the Wnt signaling pathway (Zhang et al., 2011a). miR-30e has been reported to be downregulated during osteoblastic differentiation of MSC, and its target has been identified in low-density lipoprotein receptor-related protein 6 (LRP6), a known critical factor in Wnt signaling (Wang et al., 2013b). miR-140-5p inhibits osteoblastic differentiation of hMSCs by repressing bone morphogenic protein 2 (BMP2) (Hwang et al., 2014). miR-29 family members (miR-29a-3p, miR-29b-3p, and miR-29c-3p) are upregulated during osteoclastogenesis, while their KO results in altered recruitment and migration of osteoclast precursors without any effect on osteoclast activity (Franceschetti et al., 2013). In addition, six targets (Cdc42, srGAP2, GPR85, NFIA, CD93, and CTR) of the miR-29 family are involved in cytoskeletal

organization, recruitment of osteoclast precursors, and osteoclast function (Franceschetti et al., 2013). However, results for miR-29 family roles are conflicting. The administration of pre-miR-29a in rats limited the bone loss induced by glucocorticoids, while miR-29b expression was downregulated during the differentiation of CD14+ PBMCs into osteoclasts (Rossi et al., 2013; Wang et al., 2013a). These effects are probably associated with the miR-29 family action on Wnt signaling and on osteoblast activity promotion (Wang et al., 2013a). In another study, miR-29b resulted upregulated in RAW264.7 cells treated with TNF-α and RANKL to induce osteoclastogenesis (Kagiya and Nakamura, 2013). Furthermore, miR-29b has been found to promote osteogenesis and to regulate extracellular matrix proteins expression by targeting the expression of HDAC4, TGF3, ACVR2A, CTNNBIP1, DUSP2 and COL1A1, COL5A3, COL4A2, respectively (Li et al., 2009).

Recent studies have discovered other circulating miRNAs associated with OP and OP bone fracture. Chen et al. (2017) tried to find other potential serum and tissue miRNAs in Chinese OP women with hip fractures. Five of the 95 detected miRNAs were significantly upregulated in the OP patients (n = 30) compared with the healthy non-OP controls (n = 30): miR-125b, miR-30, miR-4665-3p, miR-5914, and miR-96. Only miR-125b, miR-30, and miR-5914 were subsequently validated by RT-qPCR. These three miRNAs were also found upregulated in OP bone samples compared with controls. In both cases, miR-125b was the most upregulated, and ROC analysis confirmed its diagnostic potential in postmenopausal OP (AUC 0.898) in accordance with three previous studies (Seeliger et al., 2014; Panach et al., 2015; Kelch et al., 2017).

Yavropoulou et al. (2017) investigated the expression level of fourteen serum miRNAs, previously associated with OP and OP bone fractures in the sera from postmenopausal women with low bone mass and either with (n = 35) or without (n = 35) vertebral fractures. Compared with the controls, miR-124-3p and miR-2861 were higher, whereas miR-21-5p, miR-23a-3p, and miR-29a-3p were lower in the two OP groups compared with the non-OP controls. Furthermore, in the patients with low bone mass, the levels of miR-21-5p were lowest in the patients with vertebral fractures. Together with their above- described role, miR-124-3p, miR-21-5p, miR-23a-3p, miR-2861, and miR-29a-3p are known to positively regulate osteoblast differentiation by targeting HDAC5, a transcriptional factor that affects bone formation mediated by Runx2 (Hu et al., 2011). ROC analysis showed that the associated AUC of miR-21-5p was 0.66, with 66% sensitivity and 71% specificity (Yavropoulou et al., 2017). These results contrasted with those from previous studies that found an association between miR-21-5p and miR-23-3p upregulation with bone fractures in OP (Seeliger et al., 2014; Panach et al., 2015; Kelch et al., 2017). Wang et al. (2018) identified eight out of ten miRNAs in sera and bone tissue samples from OP patients with bone fracture. miR-100, miR-122a, miR-125b, miR-24-3p, and miR-27a-3p levels were higher in serum and upregulated in the bone samples of OP patients (n = 45) than in the non-OP subjects (n = 15), while miR-128 was upregulated only in the OP bone samples. Conversely, miR-145 expression was increased only in the OP serum compared with non-OP, while miR-144-3p was downregulated in the OP serum and the bone samples. Since miR-144-3p has not been associated with OP, the authors further investigated its role in osteoclastogenesis. miR-144 was found to affect osteoclast differentiation by targeting RANK, as well as proliferation and apoptosis.

Recently, Li et al. (2018) conducted a study to validate serum miR-133a as a biomarker for postmenopausal OP with bone fracture. miR-133a upregulation in circulating monocytes and in serum has been associated with postmenopausal OP (Wang et al., 2012; Li et al., 2014). The study reported that serum miR-133a was significantly higher in the postmenopausal OP women with hip fracture than in the healthy controls, and that it negatively correlated with BMD at the lumbar spine. *In vitro*, miR-133a expression was significantly upregulated during RANKL/M-CSF-induced osteoclastogenesis in RAW264.7 and THP-1 cells and its overexpression upregulated NFATc1, c-Fos, and TRAP protein expression (Li et al., 2018). Previous studies have also demonstrated that miR-133a overexpression in the osteoblast cell line MC3T3 suppressed osteoclastogenesis by directly targeting RUNX2 (Zhang et al., 2011b). *In vivo*, miR-133a KO in OVX rats altered the circulating levels of osteoclastogenesis-related factors and prevented bone loss (Li et al., 2018). Taken together, these findings support the diagnostic potential for miR-133a in postmenopausal OP and related bone fracture and highlight the potential of miR-133a as a clinical therapeutic target for postmenopausal OP.

Table 2 summarizes information about circulating miRNAs associated with bone fracture risk in OP.

miRNAs, Fracture Risk, and Physical Activity

Physical activity (PA) is a therapeutic strategy to reduce bone fracture risk, improve bone metabolic status and, eventually, to increase bone mass during childhood, adolescence, and early adulthood or to limit the age-associated decrease in peak bone mass in older age (Xu et al., 2016). PA affects miRNAs expression in tissues and organs, the circulating miRNAs profile reflects this situation as a consequence (Lombardi et al., 2016a). The literature on PA-dependent modifications of osteoporosis- or fracture risk-associated miRNAs is scarce (Lombardi et al., 2016a). The suboptimal understanding of these mechanisms stems from failure to appreciate the complex network of interactions accompanying the metabolic response of bone to PA. This multilevel relationship contemplates: direct effects of PA on bone; whole-body metabolic effects of PA on bone; specific effects of PA on tissues (e.g., skeletal muscle, adipose tissue, immune system, nervous system) besides the release of mediators from bone (e.g., myokines, adipokines, cytokines, and neurotransmitters) that affect bone both directly and indirectly; and PA-dependent release of mediators by bone (osteokines) that affect the expression of bone-acting mediators released by other tissues (Lombardi et al., 2016b; Lombardi, 2019). Recently, we demonstrated that seven from a panel of ten fracture risk-associated miRNAs (miR-100, miR-122-5p, miR-125-5p, miR148a-3p, miR-23a-3p, miR-24-3p, and miR-93-5p) responded to a protocol of PA (8-week repeated sprint training in young healthy males) in a more sensitive way than standard bone metabolism markers, metabolic hormones, and cytokines (Sansoni et al., 2018).

TABLE 2 | miRNAs related to bone fracture risk in postmenopausal OP.

Study	Study design	Biomarker source	Sample handling	Quantification platform	Evaluated miRNA	Normalization strategy	Reported miRNA biomarker	Potential target gene	AUC-Sensitivity (%)-Specificity (%)	Limits
(Garmilla-Ezquerra et al., 2015)	Discovery cohort: 8 women with OP hip fracture, 8 women with severe hip OA without OP fractures (control group) Replication cohort: 19 women with OP hip fracture, 19 women with severe hip OA without OP fractures (control group)	Bone specimens	Trabecular bone cylinders obtained from central part of femoral head using a trephine. Fragments cut into small pieces, washed with PBS, snap-frozen in liquid nitrogen, and stored at -70°C	Discovery stage: TaqMan array human miRNA A + B cards v3 Replication stage: TaqMan RT-qPCR	Discovery stage: 760 miRNAs tested Replication stage: miR-187, miR-193a-3p, miR-214, miR-518f, miR-636, and miR-210	NormFinder and GeNorm programs identified miR-222 and let-7b as most stable normalizators.	↑ miR-518f in OP fractures group vs. control group↑ miR-187 in control group vs. OP fractures group	IGFBP1, DKK1, WISP1, CTNNBIP1 (identified for miR-518f using microRNA.org, mirbase.org, and targetscan.org prediction algorithms but not validated by *in vitro* experiments)	/	Small sample size; OA patients as control group; no validation of the identified target genes; no information about the stem-loop arm of miRNA origin; no ROC analysis.
(De-Ugarte et al., 2015)	Discovery cohort: 6 PM OP women and 6 PM OA women (control group) both with femoral neck fracture Replication cohort: 7 PM OP women and 6 PM OA women (control group) both with femoral neck fracture	Fresh bone specimens	Bone fragments from femoral neck transcervical region reduced to small pieces, washed three times with PBS, and stored at -80°C	Discovery stage: miRCURY LNA™ microRNA Array performed by Exiqon Services Replication stage: RT-qPCR performed by Exiqon Services	Discovery stage: 1932 miRNAs tested Replication stage: miR-675-5p, miR-30c-1-3p, miR-483-5p, miR-542-5p, miR-142-3p, miR-223-3p, miR-32-3p, and miR-320a	Discovery stage: Lowess (Locally Weighted Scatterplot Smoothing) global regression algorithm. Replication stage: average of miR-let-7e-5p expression in each sample	↑ miR-320a and miR-483-5p in OP fractures vs. control group	ARPP-19, BMP3 and 6, BMPR1A, CAMTA1, DNER, ESRRG, IGF1, IGF1R, IL6R, JAK2, PPARGC1A, LEPR, MAPK1, MCL, NR3C1, PDGFD, PTGER3, RARG, RXRA, SGK, SP1, SRF, TFR1 (identified for miR-320a using PicTar, TargetScan Human, miRDB, MiRanda, DIANA-TarBase, and miRTarBase database)SRF and MAPK3 (identified for miR-483-5p using mirTArBase)	/	Small sample size; OA patients as control group; no validation of the identified target genes; no ROC analysis
(Seeliger et al., 2014)	Discovery cohort: 10 OP (7 women and 3 men) and 10 non-OP (10 women) as control group, both with femoral neck or pertrochanteric fracture Replication cohort: 30 OP women and 30 non-OP women (control group), both with femoral neck or pertrochanteric fracture	Discovery stage: cell-free serum Replication stage: cell-free serum and bone tissue	/	Screening: human Serum & Plasma miRNA PCR Array MIHS-106Z Validation: SYBR RT-qPCR	Screening: 83 miRNAs tested Validation: miR-21-5p, miR-23-3p, miR-24-3p, miR-25-3p, miR-27a-3p, miR-93, miR-100-5p, miR-122a-5p, miR-124-3p, miR-125b-5p, miR-148a-3p, miR-223-3p, and miR-637	Average of SNORD96a and snRNU6	↑ miR-21, miR-23a, miR-24, miR-93, miR-100, miR-122a, miR-124a, miR-125b, and miR-148a in OP fracture serum vs. controls↑ miR-21, miR-23a, miR-24, miR-25, miR-100, and miR-125b in bone tissue from OP fracture patients vs. control	PDCD4, cFos (miR-21); RUNX2 (miR-23a/miR-24-2/miR-27a complex); OSX (miR-93); BMPR2 (miR-100); VCAN (miR-124a); RANKL (miR-148a) (identified from previous papers but not validated in this paper)	0.63-61.3-61.7 (miR-21), 0.63-57.4-56.7 (miR-23a), 0.63-60.3-60.4 (miR-24), 0.68-69.0-68.3 (miR-93), 0.69-62.9-61.7 (miR-100), 0.77-74.1-72.1 (miR-122a), 0.69-61.4-61.0 (miR-124a), 0.76-76.4-75.0 (miR-125b), 0.61-62.5-62.3 (miR-148a) for OP fracture vs. non-OP	Small sample size; no validation of the target genes.

(*Continued*)

TABLE 2 | Continued

Study	Study design	Biomarker source	Sample handling	Quantification platform	Evaluated miRNA	Normalization strategy	Reported miRNA biomarker	Potential target gene	AUC-Sensitivity (%)-Specificity (%)	Limits
(Panach et al., 2015)	Discovery stage: 8 Caucasian women with OP subcapital hip fracture and 5 with severe OA of hip requiring surgery (control group) Replication stage: 15 Caucasian women with OP subcapital hip fracture and 12 with severe OA of hip requiring surgery (control group)	Cell-free serum	Serum samples obtained from fasting blood stored at -80°C	Discovery stage: miRCURY LNA Universal RT microRNA PCR, Serum/Plasma Focus microRNA PCR Panel Replication stage: Exiqon LNA RT-qPCR	Screening: 179 miRNAs tested Validation: miR-143-3p, miR-122-5p, miR-125b-5p, miR-210, miR-21-5p, and miR-34a-5p	GeNorm identified miR-93-5p	↑ miR-122-5p, miR-125b-5p, and miR-21-5p in OP fracture vs. control group	/	0.87 (miR-122-5p), 0.76 (miR-125-5p), and 0.87 (miR-21-5p) for OP fracture vs. control group	Small sample size, OA patients as control group; no evaluation of the target genes
(Weilner et al., 2015)	Discovery stage: 7 PM Caucasian women with femoral neck OP fracture and 7 PM women without femoral fracture (control group) Replication stage: 12 PM Caucasian women with femoral neck OP fracture and 11 PM women without femoral fracture (control group)	Cell-free serum	Serum obtained from blood samples centrifugied at RT and 2000g for 15 min, after incubation at RT for 30 min, and stored at -80°C	Screening: Exiqon serum/plasma focus panels Validation: RT-qPCR	Screening: 175 miRNAs tested Validation: miR-10a-5p, miR-10b-5p, miR-22-3p, miR 133b, miR-328-3p, and let-7g-5p	Normalization of Cp-values based on average Cp of the detected miRNAs	↓ miR-22-3p, miR-328-3p, and let-7g-5p in OP fracture serum vs. control group	/	/	Small sample size; no evaluation of the target genes; the mean age of patients recruited for the discovery and validation study was significantly different (71 years and 80 years, respectively); no ROC analysis
(Kocijan et al., 2016)	10 women with PM OP low trauma fracture and 11 healthy PM women without low-trauma fracture	Cell-free serum	Fasting blood samples immediately centrifuged and serum stored a -80°C	SYBR Green RT-qPCR	187 miRNAs tested	Global mean	↑ miR-152-3p, miR-335-5p, miR-320a and↓ let-7b-5p, miR-7-5p, miR-16-5p, miR-19a-3p, miR-19b-3p, miR-29b-3p, miR-30e-5p, miR-93-5p, miR-140-5p, miR-215-5p, miR-186-5p, miR-324-3p, miR-365a-3p, miR-378a-5p, miR-532-5p, and miR-550a-3p in fractured group vs. control group	/	0.962 (miR-152-3p), 0.959 (miR-30e-5p), 0.950 (miR-324-3p), 0.947(miR-140-5p), 0.944 (miR-19b-3p), 0.939 (miR-335-5p), 0.929 (miR-19a-3p), 0.909 (miR-550a-3p), 0.898 (miR-186-5p), 0.898 (miR-532-5p), 0.872 (miR-378a-5p), 0.870 (miR-320a), 0.879 (miR-93-5p), 0.857 (miR-16-5p), 0.853 (miR-215-5p), 0.852 (let-7b-5p), 0.824 (miR-7-5p), 0.838 (miR-29b-3p), and 0.809 (miR-365a-3p) for fracture group vs. control group	Small sample size; no evaluation of the target genes; arbitrary choice of the screened miRNAs

(*Continued*)

TABLE 2 | Continued

Study	Study design	Biomarker source	Sample handling	Quantification platform	Evaluated miRNA	Normalization strategy	Reported miRNA biomarker	Potential target gene	AUC-Sensitivity (%)-Specificity (%)	Limits
(Chen et al., 2017)	30 PM Chinese women with OP and 30 PM Chinese women without OP (control group) both with hip fracture	Cell-free serum and bone tissues	Blood samples allowed to clot, centrifuged at 1500g, then serum isolated and stored.	Screening: Microarray Validation: TaqMan RT-qPCR	Validation: miR-30, miR- 96, miR-125b, miR-4665-3p, and miR-5914	snRNU6	↑ miR-125b, miR-30 and miR-5914 in serum and bone tissues from OP fracture vs. control group	/	0.699 (miR-5914), 0.757 (miR-30), and 0.898 (miR-125b) for OP fracture vs. controls	Small sample size; no target genes evaluation; no information about the stem-loop arm of miRNA origin
(Yavropoulou et al., 2017)	35 PM women with low bone mass without vertebral fractures, 35 with low bone mass and vertebral fractures, 30 HC	Cell-free serum	Blood samples collected in clot activator tubes, placed at RT for 10-60 min, centrifuged for 10 min at 1900g and 4°C. Serum samples centrifuged again for 10 min at 16000g and 4°C and frozen at -80°C	SYBR Green RT-qPCR	14 miRNAs selected based on the existing literature: miR-21-5p, miR-23a-3p, miR-24-2-5p, miR-26a-5p, miR-29a, miR-33a-5p, miR-124-3p, miR-133a, miR-135b-5p, miR-214-3p, miR-218-5p, miR-335-3p, miR-422, and miR-2861	Panel of SNORD95, SNORD96A, and snRNU6-2	↑ miR-124-3p, miR-2861, and ↓ miR-21-5p, miR-23a-3p, miR-29a-3p in OP vs. controls↓miR-21-5p in OP with vertebral fracture vs. OP without vertebral fracture	SPRY1, BMP3, DKK2, and SMAD7 (miR-21-5p); SATB2 and RUNX2 (miR-23a-3p); SATB2 and CALB1 (miR-24-2-5p); EPHA5, COL10A1, and COL19A1 (miR-26a-5p); DUSP2, COL3A1, COL5A3, and PTHLH (miR-29a); DKK2, WIF1, and OSTF1 (miR-33a-5p); HDAC5, NFATC1, and NFATC2, (miR-124-3p); ACVR1B, FOXO1, SIRT1, and SMAD5 (miR-135b-5p); ATP2A3, CTNNB1, and VDR (miR-214-3p); COL1A1, SFRP2, SOST, and EPHA5 (miR-218-5p); DKK1 and SPARC (miR-335-3p); HDAC5 (miR-2861)(Identified using miRBase, DIANA TOOLS, PicTar, miRDB, TargetScanHuman, miRGator, and microRNA database)	0.66-66-71 (miR-21-5p) for OP with vertebral fracture vs. OP without vertebral fracture	Small sample size; no validation of the identified target genes
(Wang et al., 2018)	45 OP patients, 15 non-OP (control group) both with femoral fracture	Cell-free serum and bone tissues	/	RT-qPCR	miR-7-5p, miR-24-3p, miR-27a-3p, miR-100, miR-125b, miR-128, miR-145-5p, miR-211-5p, miR-144-3p, and miR-122a	snRNU6	↑ miR-24-3p, 27a-3p, miR-100, miR-125b, miR-122a, miR-145, and ↓ miR-144-3p in serum from OP fracture vs. non-OP fracture↑ miR-24-3p, 27a-3p, miR-100, miR-125b, miR-128, miR-122a, and ↓ miR-144-3p in bone tissues form OP fracture vs. non-OP fracture	RANK (identified for miR-144-3p using TargetScan online software and validated by *in vitro* study)	/	Small sample size of the non-OP group; no ROC analysis

(*Continued*)

TABLE 2 | Continued

Study	Study design	Biomarker source	Sample handling	Quantification platform	Evaluated miRNA	Normalization strategy	Reported miRNA biomarker	Potential target gene	AUC-Sensitivity (%)-Specificity (%)	Limits
(Li et al., 2018)	10 PM Chinese OP women with hip fracture and 10 HC	Cell-free serum	Blood samples allowed to clot then centrifuged at 1500g to obtain serum	TaqMan RT-qPCR	miR-133a	snRNU6	↑miR-133a in OP with fractures vs. HC	c-Fos, NFATc1, and TRAP for miR-133a identified by *in vitro* experiments	/	Small sample size; no ROC analysis

ACVR1B, activin A receptor type 1B; ALPL, alkaline phosphatase; ANKH, ANKH inorganic pyrophosphate transport regulator; AR, androgen receptor; ARPP-19, cAMP-regulated phosphoprotein 19; ATF4, activating transcription factor 4; ATP2A3, sarcoplasmic/endoplasmic reticulum calcium ATPase 3; BMP2K, BMP2 inducible kinase; BMP3, bone morphogenetic protein 3; BMP6, bone morphogenetic protein 6; BMPR1A, bone morphogenetic protein receptor 1A; BMPR2, bone morphogenetic protein receptor type 2; CALB1, calbindin 1; CAMTA1, calmodulin binding transcription activator 1; CNR1, cannabinoid receptor 1; CNR2, cannabinoid receptor 2; COL10A1, collagen type X alpha 1 chain; COL19A1, collagen type XIX alpha 1 chain; COL1A1, collagen type I alpha 1 chain; COL3A1, collagen type III alpha 1 chain; COL5A3, collagen type V alpha 3 chain; CTNNB1, catenin beta 1, CTNNBIP1, catenin-interacting protein 1; DKK1, Dickkopf WNT signaling pathway inhibitor 1; DKK2, Dickkopf WNT signaling pathway inhibitor 2; DNER, delta and notch-like epidermal growth factor-related receptor; DUSP2, dual specificity phosphatase 2; EPHA5, EPH receptor A5; ESR1, estrogen receptor 1; ESRRG, estrogen related receptor gamma; FOXO1, Forkhead box O1, FSHB, follicle stimulating hormone subunit beta; HC, healthy controls; HDAC5, histone deacetylase 5; IGF1, insulin-like growth factor; IGF1R, insulin-like growth factor 1 receptor; IGFBP1, insulin-like growth factor binding protein 1; IL6R, interleukin 6 receptor, JAK2,Janus kinase 2; LEPR, leptin receptor; LRP6, LDL receptor related protein 6; MAPK1, mitogen-activated protein kinase 1; MAPK3, mitogen-activated protein kinase 3; MCL, myeloid cell leukemia; MSCs, mesenchymal stem cells; NFATC1, nuclear factor of activated T cells 1; NFATC2, nuclear factor of activated T cells 2; NR3C1, nuclear receptor subfamily 3 group C member 1; OA, osteoarthritis; OP, osteoporosis; OSTF1, osteoclast stimulating factor 1; OSX, osterix; PDCD4, programmed cell death 4; PDGFD, platelet-derived growth factor D; PM, postmenopausal; PPARGC1A, peroxisome proliferator-activated receptor gamma coactivator 1-alpha; PTGER3, prostaglandin E receptor 3; PTHLH, parathyroid hormone like hormone; RANKL, receptor activator of nuclear factor k B ligand; RARG, retinoic acid receptor gamma; RT, room temperature; RT-qPCR, real-time quantitative polymerase chain reaction; RUNX2, runt-related transcription factor 2; RXRA, retinoid X receptor alpha; SATB2, SATB homeobox 2; SFRP2, secreted frizzled related protein 2; SGK, serine/threonine protein-kinase; SIRT1, sirtuin 1; SMAD5, SMAD family member 5; SMAD7, SMAD family member 7; SOST, sclerostin; SPARC, secreted protein acidic and cysteine rich; SPRY1, protein sprouty homolog 1; SRF, serum response factor; T2DM, Type 2 diabetes mellitus; TFR1, transferrin receptor protein 1; TRAP, triiodothyronine receptor auxiliary protein; TSC22D3, TSC22 domain family member 3; VCAN, versican; VDR, vitamin D receptor; WIF1, WNT inhibitory factor 1;WISP1, WNT1-inducible-signaling pathway protein 1.

miRNAs in Other Types of OP and Related Fracture Risk

Considering senile OP, a study investigated the role of a specific miRNA (miR-125b) in osteoblast differentiation (Chen et al., 2014b). miR-125b was selected due to its crucial involvement in the epigenetic regulation of proliferation/differentiation of cell lineages (Liu et al., 2011). miR-125b expression levels in BM-MSCs was found upregulated in small mixed gender populations of senile Chinese OP patients (n = 4, 3 women and 1 man) compared with subjects with normal BMD (control group, n = 5, 2 women and 3 men). miR-125b upregulation was associated with impaired BM-MSCs proliferation and osteogenic differentiation and, consistent with these observations, the antagonism of miR-125b in non-OP BM-MSCs promoted proliferation, osteoblast differentiation, and mineralization. In these cells, miR-125b also targeted Osterix (OSX), a key transcription factor for osteogenic differentiation (Chen et al., 2014b). Weilner et al. (2016) found that the presence of miR-31 in circulating microvesicles derived from senescent endothelial cells negatively impacted on the osteogenic differentiation capacity of adipose tissue-derived MSCs. Circulating miR-31 levels were higher in the plasma samples from elderly healthy donors than in young healthy controls, as well as in the plasma from OP patients compared with healthy age-matched controls. miR-31 directly inhibits osteoblast formation by targeting Frizzled-3 (FZD3). Also SATB2, Osx, and RUNX2 have been validated as targets of miR-31 (Baglio et al., 2013; Deng et al., 2014; Xie et al., 2014). This miRNA is involved in osteoclastogenesis: its expression has been found strongly upregulated during RANKL-induced osteoclast differentiation and its inhibition by specific antagomirs results in impaired osteoclast differentiation, actin ring formation, and bone resorption (Mizoguchi et al., 2013). These alterations depend upon the overexpression of the miR-31 target gene RhoA, a GTPase involved in the transduction of extracellular signals to the cytoskeleton (Mizoguchi et al., 2013). This study showed, for the first time, that the miRNA content from senescent cells-derived microvesicles might correlate with the impairment of bone formation and that miR-31 can be used as a biomarker for age-associated diseases such as OP (Weilner et al., 2016). Nonetheless, a larger cohort is needed to confirm these data.

Studies have attempted to correlate circulating and tissue-altered miRNAs expression with the risk of bone fracture in senile OP patients. In bone tissue samples from elderly Chinese patients with bone fracture, miRNA quantification by RT-PCR revealed that miR-214 expression correlated positively with age and negatively with bone formation marker levels (osteocalcin and alkaline phosphatases) (Wang et al., 2013c). The major limitations of the study were: small sample size, unclear comparison between aged and control groups, and missing information about the screened miRNAs and data normalization. In murine pre-osteoblast MC3T3-E1 cells, miR-214 negatively affected osteoblast activity and matrix mineralization by targeting activating transcription factor 4 (ATF4); these features were restored by antagomiR-214 and further accentuated by agomiR-214. Furthermore, miR-214

inhibition improved the bone phenotype in OVX and hind limb-unloaded mice, whereas osteoblast activity was limited and bone mass reduced in miR-214 transgenic mice (Wang et al., 2013c). In 2017, the nine serum miRNAs associated with OP found by Seeliger et al. (2014) were validated also in serum, bone specimens, and cultured osteoblasts and osteoclasts from another cohort of OP (n = 14, 7 women and 7 men) and OA patients (n = 14, 7 women and 7 men) with hip fractures (Kelch et al., 2017). The expression levels of miR-100-5p, miR-122-5p, miR-124-3p, miR-125b-5p, and miR-148a-3p, miR-21-5p, miR-23a-3p, miR-24-3p, and miR-93-5p were assayed by RT-qPCR. The results showed that circulating miR-100-5p, miR-122-5p, miR-124-3p, miR-148a-3p, miR-21-5p, miR-23a-3p, miR-24-3p, and miR-93-5p were significantly upregulated in the OP women and men compared with the controls, but miR-93-5p failed to discriminate between OP and non-OP male patients. Furthermore, miR-125b-5p expression was gender-related. In the OP bone samples, miR-100-5p, miR-125b-5p, miR-21-5p, miR-24-3p, and miR-93-5p were significantly upregulated in the OP patients compared with the controls and correlated with BMD. In particular, miR-21-5p expression values discriminated between osteopenia and OP. miR-100-5p, miR-125b-5p, miR-21-5p, miR-23a-3p, miR-24-3p, and miR-93-5p were upregulated in OP osteoblasts, while miR-100-5p, miR-122-5p, miR-124-3p, miR-125b-5p, miR-148a-3p, miR-21-5p, and miR-93-5p were upregulated in OP osteoclasts. Among these miRNAs, miR-122-5p was previously identified as being upregulated in serum samples from OP patients with bone fracture (Panach et al., 2015). The role of the other miRNAs and their potential target genes have been described above. These results identify miRNAs with high potential as biomarkers for OP, as well as targets for OP therapeutic treatment (Kelch et al., 2017). Recent studies have investigated whether single or combined miRNAs discriminate bone fractures in conditions associated with bone fragility. Interestingly, the nineteen serum miRNAs found altered in postmenopausal women by Kocijan et al. (2016), as previously described, were found altered also in serum samples from trauma fractures in idiopathic OP (premenopausal women, n = 10, and men, n = 16) compared to their controls (n = 28, 12 premenopausal women and 16 men) without bone fracture. Also in these cases, ROC analysis revealed that miR-140-5p, miR-152-3p, miR-19a-3p, miR-19b-3p, miR-30e-5p, miR-324-3p, miR-335-5p, and miR-550a-3p had a higher discriminating power between bone fracture and controls (AUC> 0.9) than BMD or bone turnover markers. Mandourah et al. (2018) recruited 139 subjects and divided them into 5 groups: healthy controls, osteopenic subjects with or without bone fractures, and OP patients with or without bone fractures. Fifteen of the 370 miRNAs screened in the pooled sera were differently regulated in the females with OP and the healthy females, and twenty-five were up or downregulated in the OP females compared with the osteopenic females. Following RT-qPCR validation, miR-122-5p and miR-4516 levels differed between the healthy subjects and the osteopenic/OP patients. Moreover, serum miR-122-5p and miR-4516 levels were lower in the OP patients than the healthy controls and osteopenic patients. miR-4516 was also found to be downregulated in the OP patients with bone fracture and associated with BMD. ROC analysis revealed that only miR-4516 had an acceptable diagnostic value for OP: AUC 0.727, 71% sensitivity, and 62% specificity. Furthermore, the diagnostic value of these two miRNAs increased when combined (AUC 0.752). Overall, these findings indicate that miR-122-5p and miR-4516 downregulation in patient samples may be associated with OP progression. However, miR-122-5p has been found upregulated in the sera of OP patients with hip fracture (Panach et al., 2015).

In order to discriminate between type 2 diabetes (T2DM)- and OP-associated bone fracture, serum levels of 375 miRNAs were evaluated using a low-density qPCR array. Forty-eight miRNAs were differentially expressed between T2DM patients with bone fracture and healthy controls, and 23 miRNAs differentially expressed between OP with bone fracture and healthy controls. Eighteen of these showed the same regulation pattern in the T2DM and the OP patients. Considering the top ten ranking miRNAs (i.e., four-miRNA model signatures with AUC values >0.9 for identifying the T2DM or OP fragility fracture groups), the most abundant miRNAs were miR-382-3p, miR-550a-5p, and miR-96-5p for the T2DM group and miR-188-3p, miR-382-3p, miR-942 for the OP group. miR-382-3p was downregulated in both groups with bone fracture compared with the controls; miR-550a-5p and miR-96-5p were significantly upregulated in the T2DM patients with bone fractures, while miR-188-3p and miR-942 were downregulated, although without reaching statistical significance, in OP bone fractures compared with the controls: these last two miRNAs are associated with bone metabolism (Heilmeier et al., 2016). miR-188 is recognized as a main modulator of the BM-MSCs age-associated osteogenesis-to-adipogenesis shift by targeting histone deacetylase 9 (HDAC9) and the RPTOR-independent companion of mTOR complex 2 (RICTOR). In particular, miR-188 suppression induces osteoblast differentiation and bone formation (Li et al., 2015a). By targeting the heparin-binding EGF-like growth factor (HB-EGF), miR-96 is able to promote osteoblast differentiation (Yang et al., 2014). Analyzing the *in vitro* effects of miR-188-3p, miR-382-3p, and miR-550a-5p on cell proliferation, osteogenesis, and adipogenesis, the authors demonstrated that miR-382-3p and miR-550a-5p enhance and inhibit, respectively, osteogenic differentiation and both affect adipogenesis, whereas miR-188-3p does not impair it. Thus, miR-382-3p and miR-550a-5p have been identified as potential circulating biomarkers for T2DM-associated bone disease, and miR-188-3p and miR-382-2p for bone fractures in OP (Heilmeier et al., 2016).

Table 3 presents information about circulating miRNAs associated with other types of OP and related fracture risk.

Conclusions

The growing body of evidence for the fundamental modulatory role exerted by miRNAs in biological functions, along with aberrant expression in disease onset, underline their potential as biomarkers for the onset and progression of disease. Based on current evidence, age-related bone diseases, especially in OP and OP fractures, may be correlated with altered levels of circulating and tissue miRNA. In

TABLE 3 | miRNAs associated with other types of OP and related fracture risk.

Study	Study design	Biomarker source	Sample handling	Quantification platform	Evaluated miRNA	Normalization strategy	Validated miRNA biomarker	Potential target gene	AUC-Sensitivity(%)-Specificity(%)	Limits
(Chen et al., 2014b)	4 Chinese OP patients (3 women and 1 man, age 76-88 years) and 5 Chinese subjects with normal BMD (2 women and 3 men, age 19-44 years)	BM-MCSs	Bone marrow aspirated from iliac crest and used for BM-MCSs isolation	SYBR Green RT-qPCR	miR-125b	snRNU6	↑miR-125b in OP group vs. non-OP group	OSX (Identified using TargetScan and PicTar database, and validated by *in vitro* experiments) and RUNX2 for miR-125b	/	Small sample size; no information about the stem-loop arm of miRNA origin; no ROC analysis
(Weilner et al., 2016)	14 men (mean age ~53 years) with idiopathic osteoporosis and 11 age-matched HC	Cell-free plasma and plasma microvesicles	Filtration and differential centrifugation methodologies for microvescicle purification	TaqMan RT-qPCR	miR-31	snRNU6	↑miR-31 in OP group vs. HC	FZD3 (validated by *in vitro* experiments for miR-31)	/	Small sample size; no information about the stem-loop arm of miRNA origin; study mainly focused on miRNA evaluation by *in vitro* studies; no ROC analysis
(Wang et al., 2013c)	40 Chinese patients with fracture (age 60-90 years) and 9 Chinese HC (control group)	Bone specimens	Femurs collected during surgery	RT-PCR	Not specified	Not specified in this paper	↑miR-214a in older individuals	ATF4 (identified for miR-214a using miRBase and validated by *in vitro-in vivo* experiments)	/	Small sample size of the HC group; confusing information about the comparisons done among groups; evaluated miRNAs and data normalization not explained in this paper; no information about the stem-loop arm of miRNA origin; no ROC analysis
(Kelch et al., 2017)	28 patients with hip fracture: 7 men + 7 women with OP and 7 men + 7 women with AO (control group)	Cell-free serum and bone tissue	Blood collected 2 h post-fracture (OP) or pre-operation (non-OP) into S-Monovette polypropylene tubes, placed for 30 min at RT upright, centrifuged for 10 min at 1900g, serum stored at -80°CFemoral head samples collected during surgery (within 8 h after fracture in OP group). Cylindrical bone samples obtained from middle of each femoral head, cut into small pieces with Luer forceps, rinsed with D-PBS, collected in TRI-Reagent, snap frozen in liquid nitrogen, and mechanically ground. The bone powder collected with TRI-Reagent and stored at -80°C	miScript SYBR Green RT-qPCR	miR-21-5p, miR-23a-3p, miR-24-3p, miR-93-5p, miR-100-5p, miR-122-5p, miR-124-3p, miR-125b-5p, and miR-148a-3p	SNORD96a	↑ miR-21-5p, miR-23a-3p, miR-24-3p, miR-93-5p, miR-100-5p, miR-122-5p, miR-124-3p, and miR-148a-3p in OP serum vs. control↑ miR-21-5p, miR-24-3p, miR-93-5p, miR-100-5p and miR-125b-5p in OP tissues vs. control↑ miR-21-5p, miR-23a-3p, miR-24-3p, miR-93-5p, miR-100-5p, and miR-125b-5p in OP osteoblasts vs. control↑ miR-21-5p, miR-93-5p, miR-100-5p, miR-122-5p, miR-124-3p, miR-125b-5p, and miR-148a-3p in OP osteoclasts vs. control	/	/	no evaluation of the target genes; no ROC analysis

(*Continued*)

TABLE 3 | Continued

Study	Study design	Biomarker source	Sample handling	Quantification platform	Evaluated miRNA	Normalization strategy	Validated miRNA biomarker	Potential target gene	AUC-Sensitivity(%)-Specificity(%)	Limits
(Kocijan et al., 2016)	Patients with idiopathic (16 men and 10 premenopausal women); HC without low-trauma fracture (16 men and 12 premenopausal women).	Cell-free serum	Fasting blood samples immediately centrifuged and serum stored a -80°C	SYBR Green RT-qPCR	187 miRNAs tested	Global mean	↑ miR-152-3p, miR-335-5p, miR-320a and↓ let-7b-5p, miR-7-5p, miR-16-5p, miR-19a-3p, miR-19b-3p, miR-29b-3p, miR-30e-5p, miR-93-5p, miR-140-5p, miR-215-5p, miR-186-5p, miR-324-3p, miR-365a-3p, miR-378a-5p, miR-532-5p, and miR-550a-3p in fractured groups vs. their control groups	/	0.962 (miR-152-3p), 0.959 (miR-30e-5p), 0.950 (miR-324-3p), 0.947(miR-140-5p), 0.944 (miR-19b-3p), 0.939 (miR-335-5p), 0.929 (miR-19a-3p), 0.909 (miR-550a-3p), 0.898 (miR-186-5p), 0.898 (miR-532-5p), 0.872 (miR-378a-5p), 0.870 (miR-320a), 0.879 (miR-93-5p), 0.857 (miR-16-5p), 0.853 (miR-215-5p), 0.852 (let-7b-5p), 0.824 (miR-7-5p), 0.838 (miR-29b-3p), and 0.809 (miR-365a-3p) for fracture groups vs. control groups	No evaluation of the target genes; arbitrary choice of the screened miRNAs
(Mandourah et al., 2018)	12 (1 male/11 females) non-OP controls, 61 (9 males/52 females) osteopenia without fracture, 15 (2 males/13 females) osteopenia with fracture, 33 (6 males/27 females) OP without fracture, and 18 (2 males/16 females) OP with fracture	Cell-free serum and plasma	Serum/plasma samples obtained by centrifuging at 2500g and RT for 30 min. Supernatants further centrifuged at 14000g and 4°C for 30 min. Samples stored at -80°C	Screening: Human Serum and Plasma miRNA PCR arrays Validation: miScript SYBR Green RT-qPCR	Screening: 370 miRNAs tested Validation: 40 miRNAs tested	SNORD96A and RNU6-6P	↓ miR-122-5p and miR-4516 in OP vs. non-OP and osteopenia patients	BMP2K, FSHB, IGF1R, VDR, SPARC, TSC22D3 and RUNX2 (miR-122-5p and miR-4516); ANKH, ALPL, CNR2, CD44, LRP6, and ESR1 (miR-122-5p); AR and CNR1 (miR-4516) (identified using miRWalk2.0 database but not validated in the study)	0.727-71-62 (miR-4516) and 0.752 (miR-122-5p+miR-4516) for OP	Small sample size; confusing information about the comparisons done among groups; no validation of the identified target genes
(Heilmeier et al., 2016)	80 PM women; two study arms with two groups each:T2DM arm composed of T2DM women with (n = 20) and without (n = 20) fragility fractures since T2DM onsetOP arm composed of healthy non-T2DM PM women with OP fragility fracture (n = 20), and control group of fracture-free PM women (n = 20).	Cell-free serum	Fasting blood placed for 40 min upright and centrifuged for 15 min at 2000g.	SYBR Green Low-density qPCR platform	375 miRNAs tested	Cq values computed using second derivative maximum method provided with LC480 II software.	Most abundant miRNAs among the top 10 four-miRNAs models:↓ miR-382-3p in T2DM and OP with fragility fracture vs. respective controls↑ miR-550a-5p and miR-96-5p in T2DM fragility fracture group vs. controls↓ miR-188-3p and miR-942 in OP fracture group vs. controls	/	10 candidate four-miRNA models displayed AUC values (0.922 -0.965) for identifying fracture status in T2DM.10 candidate four-miRNA models displayed AUC values (0.972 -0.991) for identifying fracture status in OP group.	No evaluation of the target genes; arbitrary choice of the screened miRNAs

ALPL, alkaline phosphatase; ANKH, ANKH inorganic pyrophosphate transport regulator; AR, androgen receptor; ATF4, activating transcription factor 4; BMD, bone mineral density; BM-MCSs, bone marrow mesenchymal stem cells; BMP2K, BMP2 inducible kinase; CNR2, cannabinoid receptor 2; ESR1, estrogen receptor 1; FSHB, follicle stimulating hormone subunit beta;FZD3, frizzled-3; HC, healthy controls; IGF1R, insulin-like growth factor 1 receptor;OA, osteoarthritis; OP, osteoporosis; OSX, osterix; RT-qPCR, real-time quantitative polymerase chain reaction; RUNX2, runt-related transcription factor 2; RUNX2, runt-related transcription factor 2; SPARC, secreted protein acidic and cysteine rich; T2DM, type 2 diabetes mellitus; TSC22D3, TSC22 domain family member 3; VDR, vitamin D receptor; WIF1, WNT inhibitory factor 1; WISP1, WNT1-inducible-signaling pathway protein 1.

addition, the essential regulatory role exerted by miRNAs in bone homeostasis, as revealed by *in vitro* and *in vivo* studies, underscores their huge potential as biomarkers for diagnosis, prognosis, and personalized treatment of age-associated bone-related disease. Unfortunately, clinical studies for identifying circulating miRNAs as markers for bone diseases have employed various different experimental protocols, making it difficult to compare the results obtained from different labs and even from the same lab in some cases. Furthermore, the great majority of the published studies, here reviewed, are featured by limited (and sometimes statistically unjustifiably too limited) sample sizes. For these reasons, more effort must be spent in standardizing the pre-analytical, analytical, and post-analytical stage of miRNAs discovery and validation to obtain valuable biomarkers for clinical practice and to improve the significance by validating, at least the most promising biomarkers, on wide and real life-adherent populations.

AUTHOR CONTRIBUTIONS

MB: Drafting the work, final approval. GB: Conception of the work, critical revision, final approval. GL: Conception of the work, drafting the work, critical revision, final approval.

FUNDING

This study was funded by the Italian Ministry of Health (Ricerca Corrente).

REFERENCES

Andersen, C. L., Jensen, J. L., and Orntoft, T. F. (2004). Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Re.s* 64 (15), 5245–5250. doi: 10.1158/0008-5472.CAN-04-0496

Arroyo, J. D., Chevillet, J. R., Kroh, E. M., Ruf, I. K., Pritchard, C. C., Gibson, D. F., et al. (2011). Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc. Natl. Acad. Sci. U. S. A.* 108 (12), 5003–5008. doi: 10.1073/pnas.1019055108

Baggish, A. L., Hale, A., Weiner, R. B., Lewis, G. D., Systrom, D., Wang, F., et al. (2011). Dynamic regulation of circulating microRNA during acute exhaustive exercise and sustained aerobic exercise training. *J. Physiol.* 589 (Pt 16), 3983–3994. doi: 10.1113/jphysiol.2011.213363

Baglio, S. R., Devescovi, V., Granchi, D., and Baldini, N. (2013). MicroRNA expression profiling of human bone marrow mesenchymal stem cells during osteogenic differentiation reveals Osterix regulation by miR-31. *Gene* 527 (1), 321–331. doi: 10.1016/j.gene.2013.06.021

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136 (2), 215–233. doi: 10.1016/j.cell.2009.01.002

Bayraktar, R., Van Roosbroeck, K., and Calin, G. A. (2017). Cell-to-cell communication: microRNAs as hormones. *Mol. Oncol.* 11 (12), 1673–1686. doi: 10.1002/1878-0261.12144

Bedene, A., Mencej Bedrač, S., Ješe, L., Marc, J., Vrtačnik, P., Preželj, J., et al. (2016). MiR-148a the epigenetic regulator of bone homeostasis is increased in plasma of osteoporotic postmenopausal women. *Wien. Klin. Wochenschr*. 128 (Suppl 7), 519–526. doi: 10.1007/s00508-016-1141-3

Bellavia, D., Costa, V., De Luca, A., Maglio, M., Pagani, S., et al. (2016). Vitamin D level between calcium-phosphorus homeostasis and immune system: new perspective in osteoporosis. *Curr. Osteoporos Rep.* doi: 10.1007/s11914-016-0331-2

Bellavia, D., De Luca, A., Carina, V., Costa, V., Raimondi, L., Salamanna, F., et al. (2019). Deregulated miRNAs in bone health: Epigenetic roles in osteoporosis. *Bone* 122, 52–75. doi: 10.1016/j.bone.2019.02.013

Bernstein, E., Kim, S. Y., Carmell, M. A., Murchison, E. P., Alcorn, H., Li, M. Z., et al. (2003). Dicer is essential for mouse development. *Nat Genet.* 35 (3), 215–217. doi: 10.1038/ng1253

Bloch, S., Wegrzyn, A., Wegrzyn, G., and Nejman-Falenczyk, B. (2017). Small and Smaller-sRNAs and MicroRNAs in the Regulation of Toxin Gene Expression in Prokaryotic Cells: a mini-review. *Toxins. (Basel)* 9 (6). doi: 10.3390/toxins9060181

Boeckel, J. N., Thome, C. E., Leistner, D., Zeiher, A. M., Fichtlscherer, S., and Dimmeler, S. (2013). Heparin selectively affects the quantification of microRNAs in human blood samples. *Clin. Chem.* 59 (7), 1125–1127. doi: 10.1373/clinchem.2012.199505

Bustin, S. A., and Nolan, T. (2004). Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *J. Biomol. Tech.* 15 (3), 155–166.

Cao, Z., Moore, B. T., Wang, Y., Peng, X. H., Lappe, J. M., Recker, R. R., et al. (2014). MiR-422a as a potential cellular microRNA biomarker for postmenopausal osteoporosis. *PLoS One* 9 (5), e97098. doi: 10.1371/journal.pone.0097098

Cerutti, H., and Casas-Mollano, J. A. (2006). On the origin and functions of RNA-mediated silencing: from protists to man. *Curr. Genet.* 50 (2), 81–99. doi: 10.1007/s00294-006-0078-x

Chen, C., Cheng, P., Xie, H., Zhou, H. D., Wu, X. P., Liao, E. Y., et al. (2014a). MiR-503 regulates osteoclastogenesis via targeting RANK. *J. Bone Miner. Res.* 29 (2), 338–347. doi: 10.1002/jbmr.2032

Chen, H., Jiang, H., Can, D., Xu, H., Zhang, K., and Guo, S. (2017). Evaluation of MicroRNA 125b as a potential biomarker for postmenopausal osteoporosis. *Trop. J. Pharm. Res.* 16 (3), 641–647. doi: 10.4314/tjpr.v16i3.20

Chen, J., Li, K., Pang, Q., Yang, C., Zhang, H., Wu, F., et al. (2016). Identification of suitable reference gene and biomarkers of serum miRNAs for osteoporosis. *Sci Rep.* 6, 36347. doi: 10.1038/srep36347

Chen, S., Yang, L., Jie, Q., Lin, Y. S., Meng, G. L., Fan, J. Z., et al. (2014b). MicroRNA125b suppresses the proliferation and osteogenic differentiation of human bone marrowderived mesenchymal stem cells. *Mol. Med. Rep.* 9 (5), 1820–1826. doi: 10.3892/mmr.2014.2024

Chen, X., Ba, Y., Ma, L., Cai, X., Yin, Y., Wang, K., et al. (2008). Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res.* 18 (10), 997–1006. doi: 10.1038/cr.2008.282

Chen, X., Liang, H., Zhang, J., Zen, K., and Zhang, C. Y. (2012). Secreted microRNAs: a new form of intercellular communication. *Trends Cell Biol.* 22 (3), 125–132. doi: 10.1016/j.tcb.2011.12.001

Cheng, H. H., Yi, H. S., Kim, Y., Kroh, E. M., Chien, J. W., Eaton, K. D., et al. (2013a). Plasma processing conditions substantially influence circulating microRNA biomarker levels. *PLoS One* 8 (6), e64795. doi: 10.1371/journal.pone.0064795

Cheng, P., Chen, C., He, H. B., Hu, R., Zhou, H. D., Xie, H., et al. (2013b). miR-148a regulates osteoclastogenesis by targeting V-maf musculoaponeurotic fibrosarcoma oncogene homolog B. *J. Bone Miner. Res.* 28 (5), 1180–1190. doi: 10.1002/jbmr.1845

Dang, Y., Yang, Q., Xue, Z., and Liu, Y. (2011). RNA interference in fungi: pathways, functions, and applications. *Eukaryot Cell* 10 (9), 1148–1155. doi: 10.1128/EC.05109-11

Davis-Dusenbery, B. N., and Hata, A. (2010). Mechanisms of control of microRNA biogenesis. *J. Biochem.* 148 (4), 381–392. doi: 10.1093/jb/mvq096

De-La-Cruz-Montoya, A. H., Ramirez-Salazar, E. G., Martinez-Aguilar, M. M., Gonzalez-de-la-Rosa, P. M., Quiterio, M., Abreu-Goodger, C., et al. (2018). Identification of miR-708-5p in peripheral blood monocytes: Potential marker for postmenopausal osteoporosis in Mexican-Mestizo population. *Exp. Biol. Med. (Maywood)* 243 (13), 1027–1036. doi: 10.1177/1535370218806828

Deng, L., Hu, G., Jin, L., Wang, C., and Niu, H. (2017). Involvement of microRNA-23b in TNF-alpha-reduced BMSC osteogenic differentiation via targeting runx2. *J. Bone Miner. Metab*. 36 (6), 648–660. doi: 10.1007/s00774-017-0886-8

Deng, Y., Bi, X., Zhou, H., You, Z., Wang, Y., Gu, P., et al. (2014). Repair of critical-sized bone defects with anti-miR-31-expressing bone marrow stromal stem cells and poly(glycerol sebacate) scaffolds. *Eur. Cell Mater* 27, 13–24. doi: 10.22203/eCM.v027a02

De-Ugarte, L., Yoskovitz, G., Balcells, S., Guerri-Fernandez, R., Martinez-Diaz, S., Mellibovsky, L., et al. (2015). MiRNA profiling of whole trabecular bone: identification of osteoporosis-related changes in MiRNAs in human hip bones. *BMC Med. Genomics* 8, 75. doi: 10.1186/s12920-015-0149-2

Eguchi, T., Watanabe, K., Hara, E. S., Ono, M., Kuboki, T., and Calderwood, S. K. (2013). OstemiR: a novel panel of microRNA biomarkers in osteoblastic and osteocytic differentiation from mesencymal stem cells. *PLoS One* 8 (3), e58796. doi: 10.1371/journal.pone.0058796

El-Hefnawy, T., Raja, S., Kelly, L., Bigbee, W. L., Kirkwood, J. M., Luketich, J. D., et al. (2004). Characterization of amplifiable, circulating RNA in plasma and its potential as a tool for cancer diagnostics. *Clin. Chem.* 50 (3), 564–573. doi: 10.1373/clinchem.2003.028506

Faraldi, M., Gomarasca, M., Banfi, G., and Lombardi, G. (2018). Free Circulating miRNAs measurement in clinical settings: the still unsolved issue of the normalization. *Adv. Clin. Chem.* 87, 113–139. doi: 10.1016/bs.acc.2018.07.003

Faraldi, M., Gomarasca, M., Sansoni, V., Perego, S., Banfi, G., and Lombardi, G. (2019). Normalization strategies differently affect circulating miRNA profile associated with the training status. *Sci Rep.* 9 (1), 1584. doi: 10.1038/s41598-019-38505-x

Franceschetti, T., Kessler, C. B., Lee, S. K., and Delany, A. M. (2013). miR-29 promotes murine osteoclastogenesis by regulating osteoclast commitment and migration. *J. Biol. Chem.* 288 (46), 33347–33360. doi: 10.1074/jbc.M113.484568

Fujita, S., Ito, T., Mizutani, T., Minoguchi, S., Yamamichi, N., Sakurai, K., et al. (2008). miR-21 Gene expression triggered by AP-1 is sustained through a double-negative feedback mechanism. *J. Mol. Biol.* 378 (3), 492–504. doi: 10.1016/j.jmb.2008.03.015

Garcia Palacios, V., Robinson, L. J., Borysenko, C. W., Lehmann, T., Kalla, S. E., and Blair, H. C. (2005). Negative regulation of RANKL-induced osteoclastic differentiation in RAW264.7 Cells by estrogen and phytoestrogens. *J. Biol. Chem.* 280 (14), 13720–13727. doi: 10.1074/jbc.M410995200

Garcia, M. E., Blanco, J. L., Caballero, J., and Gargallo-Viola, D. (2002). Anticoagulants interfere with PCR used to diagnose invasive aspergillosis. *J. Clin. Microbiol.* 40 (4), 1567–1568. doi: 10.1128/JCM.40.4.1567-1568.2002

Garmilla-Ezquerra, P., Sanudo, C., Delgado-Calle, J., Perez-Nunez, M. I., Sumillera, M., and Riancho, J. A. (2015). Analysis of the bone microRNome in osteoporotic fractures. *Calcif. Tissue Int.* 96 (1), 30–37. doi: 10.1007/s00223-014-9935-7

Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A., and Enright, A. J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* 34 (Database issue), D140–144. doi: 10.1093/nar/gkj112 (Database issue).

Grundhoff, A., and Sullivan, C. S. (2011). Virus-encoded microRNAs. *Virology* 411 (2), 325–343. doi: 10.1016/j.virol.2011.01.002

Ha, M., and Kim, V. N. (2014). Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* 15 (8), 509–524. doi: 10.1038/nrm3838

Hackl, M., Heilmeier, U., Weilner, S., and Grillari, J. (2016). Circulating microRNAs as novel biomarkers for bone diseases - complex signatures for multifactorial diseases? *Mol. Cell Endocrinol.* 432, 83–95. doi: 10.1016/j.mce.2015.10.015

Hadjiargyrou, M., and Komatsu, D. E. (2019). The therapeutic potential of MicroRNAs as orthobiologics for skeletal fractures. *J. Bone Miner. Res.* 34 (5), 797–809. doi: 10.1002/jbmr.3708

Hajian-Tilaki, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J. Int. Med.* 4 (2), 627–635.

Hammond, S. M. (2015). An overview of microRNAs. *Adv. Drug Delivery Rev.* 87, 3–14. doi: 10.1016/j.addr.2015.05.001

Hassan, M. Q., Gordon, J. A., Beloti, M. M., Croce, C. M., van Wijnen, A. J., Stein, J. L., et al. (2010). A network connecting Runx2, SATB2, and the miR-23a~27a~24-2 cluster regulates the osteoblast differentiation program. *Proc. Natl. Acad. Sci. U. S. A.* 107 (46), 19879–19884. doi: 10.1073/pnas.1007698107

Heilmeier, U., Hackl, M., Skalicky, S., Weilner, S., Schroeder, F., Vierlinger, K., et al. (2016). Serum miRNA signatures are indicative of skeletal fractures in postmenopausal women with and without type 2 diabetes and influence osteogenic and adipogenic differentiation of adipose tissue-derived mesenchymal stem cells in vitro. *J. Bone Miner. Res.* 31 (12), 2173–2192. doi: 10.1002/jbmr.2897

Ho, A. S., Huang, X., Cao, H., Christman-Skieller, C., Bennewith, K., Le, Q. T., et al. (2010). Circulating miR-210 as a novel hypoxia marker in pancreatic cancer. *Transl. Oncol.* 3 (2), 109–113. doi: 10.1593/tlo.09256

Hu, R., Liu, W., Li, H., Yang, L., Chen, C., Xia, Z. Y., et al. (2011). A Runx2/miR-3960/miR-2861 regulatory feedback loop during mouse osteoblast differentiation. *J. Biol. Chem.* 286 (14), 12328–12339. doi: 10.1074/jbc.M110.176099

Hu, Z., Dong, J., Wang, L. E., Ma, H., Liu, J., Zhao, Y., et al. (2012). Serum microRNA profiling and breast cancer risk: the use of miR-484/191 as endogenous controls. *Carcinogenesis* 33 (4), 828–834. doi: 10.1093/carcin/bgs030

Hwang, S., Park, S. K., Lee, H. Y., Kim, S. W., Lee, J. S., Choi, E. K., et al. (2014). miR-140-5p suppresses BMP2-mediated osteogenesis in undifferentiated human mesenchymal stem cells. *FEBS Lett.* 588 (17), 2957–2963. doi: 10.1016/j.febslet.2014.05.048

Ishimoto, T., Sugihara, H., Watanabe, M., Sawayama, H., Iwatsuki, M., Baba, Y., et al. (2014). Macrophage-derived reactive oxygen species suppress miR-328 targeting CD44 in cancer cells and promote redox adaptation. *Carcinogenesis* 35 (5), 1003–1011. doi: 10.1093/carcin/bgt402

Jarry, J., Schadendorf, D., Greenwood, C., Spatz, A., and van Kempen, L. C. (2014). The validity of circulating microRNAs in oncology: five years of challenges and contradictions. *Mol. Oncol.* 8 (4), 819–829. doi: 10.1016/j.molonc.2014.02.009

Jeong, B. C., Kang, I. H., Hwang, Y. C., Kim, S. H., and Koh, J. T. (2014). MicroRNA-194 reciprocally stimulates osteogenesis and inhibits adipogenesis via regulating COUP-TFII expression. *Cell Death Dis.* 5, e1532. doi: 10.1038/cddis.2014.485

Jimenez-Ortega, R. F., Ramirez-Salazar, E. G., Parra-Torres, A. Y., Munoz-Montero, S. A., Rangel-Escareno, C., Salido-Guadarrama, I., et al. (2017). Identification of microRNAs in human circulating monocytes of postmenopausal osteoporotic Mexican-Mestizo women: A pilot study. *Exp. Ther. Med.* 14 (6), 5464–5472. doi: 10.3892/etm.2017.5260

Kagiya, T., and Nakamura, S. (2013). Expression profiling of microRNAs in RAW264.7 cells treated with a combination of tumor necrosis factor alpha and RANKL during osteoclast differentiation. *J. Periodontal Res.* 48 (3), 373–385. doi: 10.1111/jre.12017

Kavsak, P. A., and Hammett-Stabler, C. A. (2014). Clinical biochemistry year in review - the clinical "good", the analytical "bad", and the "ugly" laboratory practices. *Clin. Biochem.* 47 (18), 255–256. doi: 10.1016/j.clinbiochem.2014.11.015

Kelch, S., Balmayor, E. R., Seeliger, C., Vester, H., Kirschke, J. S., and van Griensven, M. (2017). miRNAs in bone tissue correlate to bone mineral density and circulating miRNAs are gender independent in osteoporotic patients. *Sci Rep.* 7 (1), 15861. doi: 10.1038/s41598-017-16113-x

Kim, D. J., Linnstaedt, S., Palma, J., Park, J. C., Ntrivalas, E., Kwak-Kim, J. Y., et al. (2012). Plasma components affect accuracy of circulating cancer-related microRNA quantitation. *J. Mol. Diagn.* 14 (1), 71–80. doi: 10.1016/j.jmoldx.2011.09.002

Kirschner, M. B., Edelman, J. J., Kao, S. C., Vallely, M. P., van Zandwijk, N., and Reid, G. (2013). The impact of hemolysis on cell-free microRNA biomarkers. *Front. Genet.* 4, 94. doi: 10.3389/fgene.2013.00094

Kirschner, M. B., Kao, S. C., Edelman, J. J., Armstrong, N. J., Vallely, M. P., van Zandwijk, N., et al. (2011). Haemolysis during sample preparation alters microRNA content of plasma. *PLoS One* 6 (9), e24145. doi: 10.1371/journal.pone.0024145

Kocijan, R., Muschitz, C., Geiger, E., Skalicky, S., Baierl, A., Dormann, R., et al. (2016). Circulating microRNA Signatures in patients with idiopathic and postmenopausal osteoporosis and fragility fractures. *J. Clin. Endocrinol. Metab.* 101 (11), 4125–4134. doi: 10.1210/jc.2016-2365

Kroh, E. M., Parkin, R. K., Mitchell, P. S., and Tewari, M. (2010). Analysis of circulating microRNA biomarkers in plasma and serum using quantitative reverse transcription-PCR (qRT-PCR). *Methods* 50 (4), 298–301. doi: 10.1016/j.ymeth.2010.01.032

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. *Science* 294 (5543), 853–858. doi: 10.1126/science.1064921

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75 (5), 843–854. doi: 10.1016/0092-8674(93)90529-Y

Lee, Y., Jeon, K., Lee, J. T., Kim, S., and Kim, V. N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J.* 21 (17), 4663–4670. doi: 10.1093/emboj/cdf476

Lee, Y., Kim, H. J., Park, C. K., Kim, Y. G., Lee, H. J., Kim, J. Y., et al. (2013). MicroRNA-124 regulates osteoclast differentiation. *Bone* 56 (2), 383–389. doi: 10.1016/j.bone.2013.07.007

Li, C. J., Cheng, P., Liang, M. K., Chen, Y. S., Lu, Q., Wang, J. Y., et al. (2015a). MicroRNA-188 regulates age-related switch between osteoblast and adipocyte differentiation. *J. Clin. Invest.* 125 (4), 1509–1522. doi: 10.1172/JCI77716

Li, H., Wang, Z., Fu, Q., and Zhang, J. (2014). Plasma miRNA levels correlate with sensitivity to bone mineral density in postmenopausal osteoporosis patients. *Biomarkers* 19 (7), 553–556. doi: 10.3109/1354750X.2014.935957

Li, J., He, X., Wei, W., and Zhou, X. (2015b). MicroRNA-194 promotes osteoblast differentiation via downregulating STAT1. *Biochem. Biophys. Res. Commun.* 460 (2), 482–488. doi: 10.1016/j.bbrc.2015.03.059

Li, L., Zhu, D., Huang, L., Zhang, J., Bian, Z., Chen, X., et al. (2012). Argonaute 2 complexes selectively protect the circulating microRNAs in cell-secreted microvesicles. *PLoS One* 7 (10), e46957. doi: 10.1371/journal.pone.0046957

Li, Z., Hassan, M. Q., Jafferji, M., Aqeilan, R. I., Garzon, R., Croce, C. M., et al. (2009). Biological functions of miR-29b contribute to positive regulation of osteoblast differentiation. *J. Biol. Chem.* 284 (23), 15676–15684. doi: 10.1074/jbc.M809787200

Li, Z., Hassan, M. Q., Volinia, S., van Wijnen, A. J., Stein, J. L., Croce, C. M., et al. (2008). A microRNA signature for a BMP2-induced osteoblast lineage commitment program. *Proc. Natl. Acad. Sci. U. S. A.* 105 (37), 13906–13911. doi: 10.1073/pnas.0804438105

Li, Z., Zhang, W., and Huang, Y. (2018). MiRNA-133a is involved in the regulation of postmenopausal osteoporosis through promoting osteoclast differentiation. *Acta Biochim. Biophys. Sin. (Shanghai)* 50 (3), 273–280. doi: 10.1093/abbs/gmy006

Lima-Oliveira, G., Guidi, G. C., Salvagno, G. L., Brocco, G., Danese, E., and Lippi, G. (2016). Estimation of the imprecision on clinical chemistry testing due to fist clenching and maintenance during venipuncture. *Clin. Biochem.* 49 (18), 1364–1367. doi: 10.1016/j.clinbiochem.2016.07.007

Lima-Oliveira, G., Lippi, G., Salvagno, G. L., Montagnana, M., Picheth, G., and Guidi, G. C. (2013). The effective reduction of tourniquet application time after minor modification of the CLSI H03-A6 blood collection procedure. *Biochem. Med. (Zagreb)* 23 (3), 308–315. doi: 10.11613/BM.2013.037

Lin, Q., Gao, Z., Alarcon, R. M., Ye, J., and Yun, Z. (2009). A role of miR-27 in the regulation of adipogenesis. *FEBS J.* 276 (8), 2348–2358. doi: 10.1111/j.1742-4658.2009.06967.x

Ling, S., Zhong, G., Sun, W., Liang, F., Wu, F., et al. (2017). Circulating microRNAs correlated with bone loss induced by 45 days of bed rest. *Front. Physiol.* 8, 69. doi: 10.3389/fphys.2017.00069

Liu, H., Hao, W., Wang, X., and Su, H. (2016). miR-23b targets Smad 3 and ameliorates the LPS-inhibited osteogenic differentiation in preosteoblast MC3T3-E1 cells. *J. Toxicol. Sci.* 41 (2), 185–193. doi: 10.2131/jts.41.185

Liu, J., Gao, J., Du, Y., Li, Z., Ren, Y., Gu, J., et al. (2012). Combination of plasma microRNAs with serum CA19-9 for early detection of pancreatic cancer. *Int. J. Cancer* 131 (3), 683–691. doi: 10.1002/ijc.26422

Liu, L. H., Li, H., Li, J. P., Zhong, H., Zhang, H. C., Chen, J., et al. (2011). miR-125b suppresses the proliferation and migration of osteosarcoma cells through down-regulation of STAT3. *Biochem. Biophys. Res. Commun.* 416 (1-2), 31–38. doi: 10.1016/j.bbrc.2011.10.117

Livesey, J. H., Ellis, M. J., and Evans, M. J. (2008). Pre-analytical requirements. *Clin. Biochem. Rev.* 29 Suppl 1, S11–S15.

Lombardi, G. (2019). Exercise-dependent modulation of bone metabolism and bone endocrine function: new findings and therapeutic perspectives. *J. Sci. Sport. Exercise* 1 (1), 20–28. doi: 10.1007/s42978-019-0010-y

Lombardi, G., Lanteri, P., Colombini, A., and Banfi, G. (2012). Blood biochemical markers of bone turnover: pre-analytical and technical aspects of sample collection and handling. *Clin. Chem. Lab. Med.* 50 (5), 771–789. doi: 10.1515/cclm-2011-0614

Lombardi, G., Perego, S., Sansoni, V., and Banfi, G. (2016a). Circulating miRNA as fine regulators of the physiological responses to physical activity: Pre-analytical warnings for a novel class of biomarkers. *Clin. Biochem.* 49 (18), 1331–1339. doi: 10.1016/j.clinbiochem.2016.09.017

Lombardi, G., Sanchis-Gomar, F., Perego, S., Sansoni, V., and Banfi, G. (2016b). Implications of exercise-induced adipo-myokines in bone metabolism. *Endocrine* 54 (2), 284–305. doi: 10.1007/s12020-015-0834-0

Ludwig, N., Leidinger, P., Becker, K., Backes, C., Fehlmann, T., Pallasch, C., et al. (2016). Distribution of miRNA expression across human tissues. *Nucleic Acids Res.* 44 (8), 3865–3877. doi: 10.1093/nar/gkw116

Lujambio, A., and Esteller, M. (2009). How epigenetics can explain human metastasis: a new role for microRNAs. *Cell Cycle* 8 (3), 377–382. doi: 10.4161/cc.8.3.7526

Lujambio, A., Calin, G. A., Villanueva, A., Ropero, S., Sanchez-Cespedes, M., Blanco, D., et al. (2008). A microRNA DNA methylation signature for human cancer metastasis. *Proc. Natl. Acad. Sci. U. S. A.* 105 (36), 13556–13561. doi: 10.1073/pnas.0803055105

Mandourah, A. Y., Ranganath, L., Barraclough, R., Vinjamuri, S., Hof, R. V., Hamill, S., et al. (2018). Circulating microRNAs as potential diagnostic biomarkers for osteoporosis. *Sci Rep.* 8 (1), 8421. doi: 10.1038/s41598-018-26525-y

Materozzi, M., Merlotti, D., Gennari, L., and Bianciardi, S. (2018). The potential role of mirnas as new biomarkers for osteoporosis. *Int. J. Endocrinol.* 2018, 2342860. doi: 10.1155/2018/2342860

Meng, J., Zhang, D., Pan, N., Sun, N., Wang, Q., Fan, J., et al. (2015). Identification of miR-194-5p as a potential biomarker for postmenopausal osteoporosis. *PeerJ* 3, e971. doi: 10.7717/peerj.971

Mestdagh, P., Hartmann, N., Baeriswyl, L., Andreasen, D., Bernard, N., Chen, C., et al. (2014). Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study. *Nat. Methods* 11 (8), 809–815. doi: 10.1038/nmeth.3014

Mestdagh, P., Van Vlierberghe, P., De Weer, A., Muth, D., Westermann, F., Speleman, F., et al. (2009). A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biol.* 10 (6), R64. doi: 10.1186/gb-2009-10-6-r64

Mitchell, P. S., Parkin, R. K., Kroh, E. M., Fritz, B. R., Wyman, S. K., Pogosova-Agadjanyan, E. L., et al. (2008). Circulating microRNAs as stable blood-based markers for cancer detection. *Proc. Natl. Acad. Sci. U. S. A.* 105 (30), 10513–10518. doi: 10.1073/pnas.0804549105

Mizoguchi, F., Murakami, Y., Saito, T., Miyasaka, N., and Kohsaka, H. (2013). miR-31 controls osteoclast formation and bone resorption by targeting RhoA. *Arthritis Res. Ther.* 15 (5), R102. doi: 10.1186/ar4282

Mizuno, Y., Yagi, K., Tokuzawa, Y., Kanesaki-Yatsuka, Y., Suda, T., Katagiri, T., et al. (2008). miR-125b inhibits osteoblastic differentiation by down-regulation of cell proliferation. *Biochem. Biophys. Res. Commun.* 368 (2), 267–272. doi: 10.1016/j.bbrc.2008.01.073

Morrow, D. A., and de Lemos, J. A. (2007). Benchmarks for the assessment of novel cardiovascular biomarkers. *Circulation* 115 (8), 949–952. doi: 10.1161/CIRCULATIONAHA.106.683110

Murchison, E. P., Partridge, J. F., Tam, O. H., Cheloufi, S., and Hannon, G. J. (2005). Characterization of Dicer-deficient murine embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 102 (34), 12135–12140. doi: 10.1073/pnas.0505479102

Neal, C. S., Michael, M. Z., Pimlott, L. K., Yong, T. Y., Li, J. Y., and Gleadle, J. M. (2011). Circulating microRNA expression is reduced in chronic kidney disease. *Nephrol. Dial. Transplant* 26 (11), 3794–3802. doi: 10.1093/ndt/gfr485

Nelson, P. T., Wang, W. X., Wilfred, B. R., and Tang, G. (2008). Technical variables in high-throughput miRNA expression profiling: much work remains to be done. *Biochim. Biophys. Acta.* 1779 (11), 758–765. doi: 10.1016/j.bbagrm.2008.03.012

NIH Consensus Development Panel on Osteoporosis Prevention, Diagnosis, and Therapy. (2001). Osteoporosis prevention, diagnosis, and therapy. *JAMA* 285 (6), 785–795. doi: 10.1001/jama.285.6785

Palmieri, A., Pezzetti, F., Brunelli, G., Martinelli, M., Lo Muzio, L., Scarano, A., et al. (2008). Peptide-15 changes miRNA expression in osteoblast-like cells. *Implant Dent.* 17 (1), 100–108. doi: 10.1097/ID.0b013e318166d182

Pan, W., Wang, H., Jianwei, R., and Ye, Z. (2014). MicroRNA-27a promotes proliferation, migration and invasion by targeting MAP2K4 in human osteosarcoma cells. *Cell Physiol. Biochem.* 33 (2), 402–412. doi: 10.1159/000356679

Panach, L., Mifsut, D., Tarin, J. J., Cano, A., and Garcia-Perez, M. A. (2015). Serum circulating MicroRNAs as biomarkers of osteoporotic fracture. *Calcif. Tissue Int.* 97 (5), 495–505. doi: 10.1007/s00223-015-0036-z

Ramirez-Salazar, E. G., Carrillo-Patino, S., Hidalgo-Bravo, A., Rivera-Paredez, B., Quiterio, M., Ramirez-Palacios, P., et al. (2018). Serum miRNAs miR-140-3p and miR-23b-3p as potential biomarkers for osteoporosis and osteoporotic fracture in postmenopausal Mexican-Mestizo women. *Gene* 679, 19–27. doi: 10.1016/j.gene.2018.08.074

Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B., and Bartel, D. P. (2002). MicroRNAs in plants. *Genes Dev.* 16 (13), 1616–1626. doi: 10.1101/gad.1004402

Rossi, M., Pitari, M. R., Amodio, N., Di Martino, M. T., Conforti, F., Leone, E., et al. (2013). miR-29b negatively regulates human osteoclastic cell differentiation and function: implications for the treatment of multiple myeloma-related bone disease. *J. Cell Physiol.* 228 (7), 1506–1515. doi: 10.1002/jcp.24306

Saito, E., Suzuki, D., Kurotaki, D., Mochizuki, A., Manome, Y., Suzawa, T., et al. (2017). Down-regulation of Irf8 by Lyz2-cre/loxP accelerates osteoclast differentiation in vitro. *Cytotechnology* 69 (3), 443–450. doi: 10.1007/s10616-016-0013-z

Saito, Y., Liang, G., Egger, G., Friedman, J. M., Chuang, J. C., Coetzee, G. A., et al. (2006). Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. *Cancer Cell* 9 (6), 435–443. doi: 10.1016/j.ccr.2006.04.020

Sansoni, V., Perego, S., Vernillo, G., Barbuti, A., Merati, G., La Torre, A., et al. (2018). Effects of repeated sprints training on fracture risk-associated miRNA. *Oncotarget* 9 (26), 18029–18040. doi: 10.18632/oncotarget.24707

Scott, G. K., Mattie, M. D., Berger, C. E., Benz, S. C., and Benz, C. C. (2006). Rapid alteration of microRNA levels by histone deacetylase inhibition. *Cancer Re.s* 66 (3), 1277–1281. doi: 10.1158/0008-5472.CAN-05-3632

Seeliger, C., Karpinski, K., Haug, A. T., Vester, H., Schmitt, A., Bauer, J. S., et al. (2014). Five freely circulating miRNAs and bone tissue miRNAs are associated with osteoporotic fractures. *J. Bone Miner. Res.* 29 (8), 1718–1728. doi: 10.1002/jbmr.2175

Shende, V. R., Goldrick, M. M., Ramani, S., and Earnest, D. J. (2011). Expression and rhythmic modulation of circulating microRNAs targeting the clock gene Bmal1 in mice. *PLoS One* 6 (7), e22586. doi: 10.1371/journal.pone.0022586

Sourvinou, I. S., Markou, A., and Lianidou, E. S. (2013). Quantification of circulating miRNAs in plasma: effect of preanalytical and analytical parameters on their isolation and stability. *J. Mol. Diagn.* 15 (6), 827–834. doi: 10.1016/j.jmoldx.2013.07.005

Sugatani, T., and Hruska, K. A. (2013). Down-regulation of miR-21 biogenesis by estrogen action contributes to osteoclastic apoptosis. *J. Cell Biochem.* 114 (6), 1217–1222. doi: 10.1002/jcb.24471

Sugatani, T., Vacher, J., and Hruska, K. A. (2011). A microRNA expression signature of osteoclastogenesis. *Blood* 117 (13), 3648–3657. doi: 10.1182/blood-2010-10-311415

Suh, M. R., Lee, Y., Kim, J. Y., Kim, S. K., Moon, S. H., Lee, J. Y., et al. (2004). Human embryonic stem cells express a unique set of microRNAs. *Dev. Biol.* 270 (2), 488–498. doi: 10.1016/j.ydbio.2004.02.019

Sun, J. Y., Huang, Y., Li, J. P., Zhang, X., Wang, L., Meng, Y. L., et al. (2012). MicroRNA-320a suppresses human colon cancer cell proliferation by directly targeting beta-catenin. *Biochem. Biophys. Res. Commun.* 420 (4), 787–792. doi: 10.1016/j.bbrc.2012.03.075

Sun, Z., Cao, X., Hu, Z., Zhang, L., Wang, H., Zhou, H., et al. (2015). MiR-103 inhibits osteoblast proliferation mainly through suppressing Cav1.2 expression in simulated microgravity. *Bone* 76, 121–128. doi: 10.1016/j.bone.2015.04.006

Takahashi, K., Yokota, S., Tatsumi, N., Fukami, T., Yokoi, T., and Nakajima, M. (2013). Cigarette smoking substantially alters plasma microRNA profiles in healthy subjects. *Toxicol. Appl. Pharmaco.l* 272 (1), 154–160. doi: 10.1016/j.taap.2013.05.018

Trompeter, H. I., Dreesen, J., Hermann, E., Iwaniuk, K. M., Hafner, M., Renwick, N., et al. (2013). MicroRNAs miR-26a, miR-26b, and miR-29b accelerate osteogenic differentiation of unrestricted somatic stem cells from human cord blood. *BMC Genomics* 14, 111. doi: 10.1186/1471-2164-14-111

Valinezhad Orang, A., Safaralizadeh, R., and Kazemzadeh-Bavili, M. (2014). Mechanisms of miRNA-mediated gene regulation from common downregulation to mRNA-specific upregulation. *Int. J. Genomics*. 2014, 970607. doi: 10.1155/2014/970607

Van Meurs, J. B., Boer, C. G., Lopez-Delgado, L., and Riancho, J. A. (2019). Role of epigenomics in bone and cartilage disease. *J. Bone Miner. Res.* 34 (2), 215–230. doi: 10.1002/jbmr.3662

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., et al. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3 (7), 0034. doi: 10.1186/gb-2002-3-7-research0034

Vasikaran, S. D., and Chubb, S. A. (2016). The use of biochemical markers of bone turnover in the clinical management of primary and secondary osteoporosis. *Endocrine* 52 (2), 222–225. doi: 10.1007/s12020-016-0900-2

Wang, C., He, H., Wang, L., Jiang, Y., and Xu, Y. (2018). Reduced miR-144-3p expression in serum and bone mediates osteoporosis pathogenesis by targeting RANK. *Biochem. Cell Biol.* 96 (5), 627–635. doi: 10.1139/bcb-2017-0243

Wang, F. S., Chuang, P. C., Lin, C. L., Chen, M. W., Ke, H. J., Chang, Y. H., et al. (2013a). MicroRNA-29a protects against glucocorticoid-induced bone loss and fragility in rats by orchestrating bone acquisition and resorption. *Arthritis Rheum.* 65 (6), 1530–1540. doi: 10.1002/art.37948

Wang, J., Guan, X., Guo, F., Zhou, J., Chang, A., Sun, B., et al. (2013b). miR-30e reciprocally regulates the differentiation of adipocytes and osteoblasts by directly targeting low-density lipoprotein receptor-related protein 6. *Cell Death Dis.* 4, e845. doi: 10.1038/cddis.2013.356

Wang, P., Yang, D., Zhang, H., Wei, X., Ma, T., Cheng, Z., et al. (2015). Early detection of lung cancer in serum by a panel of MicroRNA biomarkers. *Clin. Lung Cancer* 16 (4), 313–319 e311. doi: 10.1016/j.cllc.2014.12.006

Wang, T., and Xu, Z. (2010). miR-27 promotes osteoblast differentiation by modulating Wnt signaling. *Biochem. Biophys. Res. Commun.* 402 (2), 186–189. doi: 10.1016/j.bbrc.2010.08.031

Wang, X., Guo, B., Li, Q., Peng, J., Yang, Z., Wang, A., et al. (2013c). miR-214 targets ATF4 to inhibit bone formation. *Nat. Med.* 19 (1), 93–100. doi: 10.1038/nm.3026

Wang, Y., Li, L., Moore, B. T., Peng, X. H., Fang, X., Lappe, J. M., et al. (2012). MiR-133a in human circulating monocytes: a potential biomarker associated with postmenopausal osteoporosis. *PLoS One* 7 (4), e34641. doi: 10.1371/journal.pone.0034641

Wang, Y., Luo, J., Zhang, H., and Lu, J. (2016). microRNAs in the same clusters evolve to coordinately regulate functionally related genes. *Mol. Biol. Evol.* 33 (9), 2232–2247. doi: 10.1093/molbev/msw089

Wang, Y., Medvid, R., Melton, C., Jaenisch, R., and Blelloch, R. (2007). DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat Genet.* 39 (3), 380–385. doi: 10.1038/ng1969

Weber, J. A., Baxter, D. H., Zhang, S., Huang, D. Y., Huang, K. H., Lee, M. J., et al. (2010). The microRNA spectrum in 12 body fluids. *Clin. Chem.* 56 (11), 1733–1741. doi: 10.1373/clinchem.2010.147405

Wei, J., Li, H., Wang, S., Li, T., Fan, J., Liang, X., et al. (2014). let-7 enhances osteogenesis and bone formation while repressing adipogenesis of human stromal/mesenchymal stem cells by regulating HMGA2. *Stem Cells Dev.* 23 (13), 1452–1463. doi: 10.1089/scd.2013.0600

Weilner, S., Schraml, E., Wieser, M., Messner, P., Schneider, K., Wassermann, K., et al. (2016). Secreted microvesicular miR-31 inhibits osteogenic differentiation of mesenchymal stem cells. *Aging Cell* 15 (4), 744–754. doi: 10.1111/acel.12484

Weilner, S., Skalicky, S., Salzer, B., Keider, V., Wagner, M., Hildner, F., et al. (2015). Differentially circulating miRNAs after recent osteoporotic fractures can influence osteogenic differentiation. *Bone* 79, 43–51. doi: 10.1016/j.bone.2015.05.027

Willeit, P., Zampetaki, A., Dudek, K., Kaudewitz, D., King, A., Kirkby, N. S., et al. (2013). Circulating microRNAs as novel biomarkers for platelet activation. *Circuit Res.* 112 (4), 595–600. doi: 10.1161/CIRCRESAHA.111.300539

Witwer, K. W. (2012). XenomiRs and miRNA homeostasis in health and disease: evidence that diet and dietary miRNAs directly and indirectly influence circulating miRNA profiles. *RNA Biol.* 9 (9), 1147–1154. doi: 10.4161/rna.21619

Xie, Q., Wang, Z., Bi, X., Zhou, H., Wang, Y., Gu, P., et al. (2014). Effects of miR-31 on the osteogenesis of human mesenchymal stem cells. *Biochem. Biophys. Res. Commun.* 446 (1), 98–104. doi: 10.1016/j.bbrc.2014.02.058

Xu, J., Lombardi, G., Jiao, W., and Banfi, G. (2016). Effects of exercise on bone status in female subjects, from young girls to postmenopausal women: an overview of systematic reviews and meta-analyses. *Sports Med.* 46 (8), 1165–1182. doi: 10.1007/s40279-016-0494-0

Yang, L., Cheng, P., Chen, C., He, H. B., Xie, G. Q., Zhou, H. D., et al. (2012). miR-93/Sp7 function loop mediates osteoblast mineralization. *J. Bone Miner. Res.* 27 (7), 1598–1606. doi: 10.1002/jbmr.1621

Yang, M., Pan, Y., and Zhou, Y. (2014). miR-96 promotes osteogenic differentiation by suppressing HBEGF-EGFR signaling in osteoblastic cells. *FEBS Lett.* 588 (24), 4761–4768. doi: 10.1016/j.febslet.2014.11.008

Yang, N., Wang, G., Hu, C., Shi, Y., Liao, L., Shi, S., et al. (2013). Tumor necrosis factor alpha suppresses the mesenchymal stem cell osteogenesis promoter

miR-21 in estrogen deficiency-induced osteoporosis. *J. Bone Miner. Res.* 28 (3), 559–573. doi: 10.1002/jbmr.1798

Yang, R., Fu, Y., Zeng, Y., Xiang, M., Yin, Y., Li, L., et al. (2017). Serum miR-20a is a promising biomarker for gastric cancer. *Biomed Rep.* 6 (4), 429–434. doi: 10.3892/br.2017.862

Yavropoulou, M. P., Anastasilakis, A. D., Makras, P., Tsalikakis, D. G., Grammatiki, M., and Yovos, J. G. (2017). Expression of microRNAs that regulate bone turnover in the serum of postmenopausal women with low bone mass and vertebral fractures. *Eur. J. Endocrinol.* 176 (2), 169–176. doi: 10.1530/EJE-16-0583

You, L., Pan, L., Chen, L., Gu, W., and Chen, J. (2016). MiR-27a is essential for the shift from osteogenic differentiation to adipogenic differentiation of mesenchymal stem cells in postmenopausal osteoporosis. *Cell Physiol. Biochem.* 39 (1), 253–265. doi: 10.1159/000445621

Yu, F., Cui, Y., Zhou, X., Zhang, X., and Han, J. (2011). Osteogenic differentiation of human ligament fibroblasts induced by conditioned medium of osteoclast-like cells. *Biosci. Trends* 5 (2), 46–51. doi: 10.5582/bst.2011.v5.2.46

Zampetaki, A., and Mayr, M. (2012). Analytical challenges and technical limitations in assessing circulating miRNAs. *Thromb. Haemos.t* 108 (4), 592–598. doi: 10.1160/TH12-02-0097

Zeng, Y., Qu, X., Li, H., Huang, S., Wang, S., Xu, Q., et al. (2012). MicroRNA-100 regulates osteogenic differentiation of human adipose-derived mesenchymal stem cells by targeting BMPR2. *FEBS Lett.* 586 (16), 2375–2381. doi: 10.1016/j.febslet.2012.05.049

Zhang, J., Tu, Q., Bonewald, L. F., He, X., Stein, G., Lian, J., et al. (2011a). Effects of miR-335-5p in modulating osteogenic differentiation by specifically downregulating Wnt antagonist DKK1. *J. Bone Miner. Res.* 26 (8), 1953–1963. doi: 10.1002/jbmr.377

Zhang, Y., Xie, R. L., Croce, C. M., Stein, J. L., Lian, J. B., van Wijnen, A. J., et al. (2011b). A program of microRNAs controls osteogenic lineage progression by targeting transcription factor Runx2. *Proc. Natl. Acad. Sci. U. S. A.* 108 (24), 9863–9868. doi: 10.1073/pnas.1018493108

Zhao, B., Takami, M., Yamada, A., Wang, X., Koga, T., Hu, X., et al. (2009). Interferon regulatory factor-8 regulates bone metabolism by suppressing osteoclastogenesis. *Nat. Med.* 15 (9), 1066–1071. doi: 10.1038/nm.2007

Zuo, B., Zhu, J., Li, J., Wang, C., Zhao, X., Cai, G., et al. (2015). microRNA-103a functions as a mechanosensitive microRNA to inhibit bone formation through targeting Runx2. *J. Bone Miner. Res.* 30 (2), 330–345. doi: 10.1002/jbmr.2352

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Bottani, Banfi and Lombardi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

REVIEW
published: 31 October 2019
doi: 10.3389/fgene.2019.01017

From Genetics to Epigenetics, Roles of Epigenetics in Inflammatory Bowel Disease

Zhen Zeng[1,2], *Arjudeb Mukherjee*[3] *and Hu Zhang*[1,2]*

[1] Department of Gastroenterology, West China Hospital, Sichuan University, Chengdu, China, [2] Center for Inflammatory Bowel Disease, West China Hospital, Sichuan University, Chengdu, China, [3] West China School of Medicine, Sichuan University, Chengdu, China

Inflammatory bowel disease (IBD) is a destructive, recurrent, and heterogeneous disease. Its detailed pathogenesis is still unclear, although available evidence supports that IBD is caused by a complex interplay between genetic predispositions, environmental factors, and aberrant immune responses. Recent breakthroughs with regard to its genetics have offered valuable insights into the sophisticated genetic basis, but the identified genetic factors only explain a small part of overall disease variance. It is becoming increasingly apparent that epigenetic factors can mediate the interaction between genetics and environment, and play a fundamental role in the pathogenesis of IBD. This review outlines recent genetic and epigenetic discoveries in IBD, with a focus on the roles of epigenetics in disease susceptibility, activity, behavior and colorectal cancer (CRC), and their potential translational applications.

Keywords: epigenetic modifications, inflammatory bowel disease, disease susceptibility, disease activity, disease behaviour, colorectal cancer, therapeutic translation

OPEN ACCESS

Edited by:
Jiucun Wang,
Fudan University, China

Reviewed by:
Jingying Zhou,
The Chinese University of Hong Kong, China
Nasun Hah,
Salk Institute for Biological Studies, United States

***Correspondence:**
Hu Zhang
zhanghu@scu.edu.cn

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *12 May 2019*
Accepted: *24 September 2019*
Published: *31 October 2019*

Citation:
Zeng Z, Mukherjee A and Zhang H (2019) From Genetics to Epigenetics, Roles of Epigenetics in Inflammatory Bowel Disease. Front. Genet. 10:1017. doi: 10.3389/fgene.2019.01017

INTRODUCTION

It is widely acknowledged that IBD is an extremely complicated disease with an unclear pathogenesis. Crohn's disease (CD) and ulcerative colitis (UC) are the most common subtypes of IBD. It predominantly affects the gastrointestinal tract (GI), and results in repeated abdominal pain, diarrhea, bloody purulent stool, and weight loss, which substantially reduces the quality of life and increases the economic burden of IBD patients (Kaser et al., 2010). Characterized by chronic inflammation and inappropriate immune responses, IBD may develop into stenosis disease, fistula phenotypes or even CRC, posing a serious management challenge. Despite many years of research, the exact pathogenesis has not been completely elucidated. Current data indicate that IBD could be accounted as the result of the complex interplay between genetic predispositions, environmental factors, and aberrant immune responses (Kaser et al., 2010; Zhang et al., 2018). Although recent technological advances have enormously facilitated the genetic research in IBD, the identified genetic factors can only explain a small proportion of overall disease variance (Ventham et al., 2013). Moreover, the great differences in disease manifestations between young and old patients cannot be explained merely by different genotypes; environmental factors should also be given due importance due to the finding that environmental changes could shape pathological gene expression through epigenetic mechanisms (Aleksandrova et al., 2017). Besides, the rapidly growing incidence and steadily increasing prevalence of IBD further impelled us to uncover the role of the genome-environment interaction in the occurrence and development of IBD (Kaplan and Ng, 2017). Epigenetic mechanisms such as DNA methylation, non-coding RNAs, histone

modification, and the positioning of nucleosomes significantly contribute to the interplay between genome and environment (Ventham et al., 2013; Karatzas et al., 2014). Available evidence also supports the critical roles of epigenetic modifications in the disease susceptibility, activity, behavior, and CRC of IBD, which has provided valuable insights into the molecular basis of IBD. Moreover, it is well known that the diagnosis, differential diagnosis, disease surveillance, and treatment of IBD are difficult, and until now, there wasn't a single solution to offer an accurate diagnosis and monitoring of IBD, and completely cure IBD on its own merits (Zhang et al., 2018). In the era of precision medicine, precision diagnosis and treatment have become an increasingly important issue in clinical practice (Li, 2018; Weissman, 2018). So, defining roles of epigenetics in IBD provides new avenues for the development of disease prediction, therapy, and monitoring. In this review, we introduce the recent genetic and epigenetic discoveries in IBD, primarily focusing on the roles of epigenetics in disease susceptibility, activity, behavior, CRC, and the potential translational applications.

ACHIEVEMENTS OF GENETIC RESEARCH IN IBD

Early family and twin studies have demonstrated that genetic factors play a fundamental role in disease susceptibility of IBD. The prevalence of disease (CD or UC) among relatives of IBD patients was significantly higher than that in controls. It should be emphasized that consistent trends were noticeable. The relatives of CD patients were at higher risk of developing CD, and those of UC patients were more likely to be subjected to UC than CD (Satsangi et al., 1994). Twin studies not only suggested that the twin concordance rates were much higher in CD than in UC, but also claimed that twins with IBD represented great consistency in clinical characteristics (Satsangi et al., 1994; Halfvarson et al., 2003). Later linkage analyses and association studies further identified many susceptibility loci (*IBD1-9*) of IBD. Nucleotide binding oligomerization domain containing 2 (*NOD2*, also known as *CARD15*) gene located in the IBD1 locus was firstly demonstrated to be a risk allele of CD, and three rare SNPs (R702W, G908R and 1007fs) were the most studied (Ahmad et al., 2001; Zhang et al., 2018). It is noteworthy that, Helbig et al. (2012) found cigarette smoking to be a possible modulator of the *NOD2* mRNA expression and function, and therefore *NOD2*-smoking interaction (gene–environment interaction) might confer an increased risk to CD. Technological innovations such as Genome-wide association study (GWAS), whole exome sequencing (WES), and fine-mapping have dramatically facilitated genetic research in IBD, identifying more than 240 susceptibility loci of IBD, including TNF superfamily member 15 (*TNFSF15*), interleukin 23 receptor (*IL23R*), autophagy related 16 like 1 (*ATG16L1*), immunity related GTPase M (*IRGM*), PR/SET domain 1 (*PRDM1*), and nuclear dot protein 52 kDa (*NDP52*, also known as *CALCOCO2*) (Ellinghaus et al., 2013; Liu et al., 2015; de Lange et al., 2017). Among these risk loci, some are shared by both CD and UC, while others are specific to one subtype (CD or UC). These data indicate that genetics plays a role in the pathogenesis of both CD and UC. However, it was quite disappointing to discover that the heritability conferred by genetic predisposition is smaller than expected (also known as missing heritability). Available data indicate that the portion of heritability explained by genetic variants was only 13.1% in CD, and 8.2% in UC (Liu et al., 2015). Therefore, understanding the role of other factors such as epigenetic modifications is a vital step in uncovering the sophisticated pathogenesis of IBD.

EPIGENETIC MODIFICATIONS IN IBD

Epigenetic modifications are defined as changes to gene structure and heritable phenotype that cannot be explained by altered DNA sequences. The classic epigenetic mechanisms include DNA methylation, histone modification, non-coding RNAs, and nucleosome positioning. In contrast, some new modifications such as RNA methylation are on the horizon (Ventham et al., 2013; Huang et al., 2019). Among these modifications, DNA methylation and non-coding RNAs are most extensively studied in IBD research.

DNA methylation is one of the chemical modifications of DNA. It is referred to the covalent addition of a methyl group to cytosines, which mostly occurs at cytosine phosphate guanine (CpG) dinucleotides, resulting in 5-methylcytosine formation (Jeltsch et al., 2018; Li et al., 2019). CpG dinucleotides occur in human genome with a low frequency of 1%, and present with nonrandom distribution (Portela and Esteller, 2010). Regions relatively clustered with CpG dinucleotides are named as CpG islands (CGIs) that range from 200bp to 5kb in length, preserve in 1–2% of the genome, and show a decreased transcriptional activity (Tang and Ho, 2007). Several studies have demonstrated aberrant changes of DNA methylation in IBD patients (Tahara et al., 2009a; Cooke et al., 2012; Kang et al., 2016; McDermott et al., 2016). Alterations in the methylation status of IBD-associated genes considerably change the transcriptional activity and expression levels of genes, thereby shaping the disease risk and progression. It is noteworthy that some DNA profiles are claimed to be common to both CD and UC, while others are demonstrated to be specific for CD or UC, which create novel and powerful motivations for disease classification and therapy. In addition, some aberrant methylated genes were initially found to be involved in IBD, and were not identified as IBD risk genes before. In this regard, it would cast new insights into the intricate pathogenesis of IBD. Non-coding RNAs are a group of RNA molecules that are not translated into proteins, including small interfering RNA (siRNA), microRNA (miRNA), PIWI-interacting RNA (piRNA), long non-coding RNA (lncRNA) and others (Gutschner and Diederichs, 2012). Numerous cellular processes such as translation, RNA splicing, gene and chromosome structure modulation, as well as DNA replication and genome defense are correlated with these non-coding RNAs (Winter et al., 2009; Gutschner and Diederichs, 2012; Dong et al., 2018). Current data indicate that non-coding RNAs, especially miRNAs, generally act in 3′ untranslated regions (3′ UTRs) and 5′ UTRs of genes, regulating gene expression at both transcriptional and post-transcriptional levels, and modifying the IBD-correlated mechanisms such as T-cell differentiation, IL23/Th17 signaling pathways, and autophagy; as a result, affecting the disease onset and progression (Wilusz et al., 2009; Winter et al., 2009;

Gutschner and Diederichs, 2012; Kalla et al., 2015; Dong et al., 2018). In accordance with findings of DNA methylation, some non-coding RNAs are also differentially expressed between CD and UC. In this respect, miRNAs can serve as potential biomarkers to provide supplementary information for more precise diagnosis and management of IBD. Collectively, it is an important emerging area as epigenetic modifications play a key regulatory role in gene replication, gene expression, and chromatin remodeling. However, despite rapid progresses being made in the field, other epigenetic patterns such as histone modification and nucleosome positioning are less studied in IBD. Moreover, the functions and precise mechanisms for most of epigenetic modifications are not completely understood. Therefore, it is definitely a pressing need to devote more efforts to annotate the functions and mechanisms of epigenetic changes in IBD. Applying basic research results into reliable biomarkers and therapeutic strategies is also becoming increasingly necessary.

ROLES OF EPIGENETICS IN IBD

Epigenetic modifications are involved in numerous diseases including cancers, neurodevelopmental disorders, cardiovascular diseases, and autoimmune diseases (rheumatoid arthritis, psoriasis, and IBD). Established roles of epigenetic modifications in the pathogenesis of these diseases suggest novel targets for disease therapy. Furthermore, significant associations between epigenetic modifications and disease susceptibility, activity and behavior indicate a potential ability to diagnose and manage disease. In this paper, we introduce the roles of epigenetic modifications in IBD, with a focus on DNA methylation and miRNA profiles (**Tables 1** and **2**).

Estimation of Disease Susceptibility

It is well established that traditional diagnosis and differential diagnosis of IBD are based on comprehensive analysis of clinical characteristics, laboratory parameters, endoscopy, imaging features, and histologic examinations. Other emerging surrogates such as genetic, serological, histologic, and fecal markers have also showed an important potential in disease diagnosis and classification. Although with these methods, some patients are still diagnosed with "IBD-unclassified" or "indeterminate colitis" (Satsangi et al., 2006). Therefore, identification of more diagnostic markers for IBD is of paramount importance. Epigenetic modifications such as DNA methylation and miRNAs are attractive biomarkers for diagnosis at a molecular level. A large number of studies have demonstrated the strength of sensitivity, specificity, and accuracy in the diagnosis of IBD.

Cooke et al. (2012) convincingly claimed that IBD cases displayed different mucosal methylation changes (*THRAP2*, *FANCC*, *GBGT1*, *DOK2* and *TNFSF4*) in comparison to healthy controls. Besides, they also found a significant difference in methylation landscape between CD and UC patients. For example, CD patients showed hypermethylated *GBGT1*, *IGFBP4*, *FAM10A4* and hypomethylated *IFITM1* when compared with UC patients, which provides a possibility for discriminating IBD from controls, and CD from UC. Subsequently, Adams et al. (2014) suggested that CD patients displayed different circulating leukocyte methylation profiles in comparison to healthy controls. They identified 65 probes and 19 differentially methylated regions (DMRs) in pediatric patients with CD, and developed models for each possible combination of two probes to discriminate CD and healthy controls with AUCs

TABLE 1 | Roles of DNA methylation in IBD.

Methylated markers	Methylation status (↑/↓)	Roles
Estimation of disease susceptibility		
THRAP2, FANCC, GBGT1, WDR8 and ITGB2	↑	CD vs. healthy controls
DOK2, TNFSF4 and VMP1	↓	CD vs. healthy controls
THRAP2, FANCC, GBGT1, WDR8, CARD9 and CDH1	↑	UC vs. healthy controls
ICAM3, DOK2, TNFSF4 and VMP1	↓	UC vs. healthy controls
GBGT1, IGFBP4 and FAM10A4	↑	CD vs. UC
IFITM1	↓	CD vs. UC
Assessment of disease activity		
CDH1, GDNF, SLIT2, MDR1, FMR1, GXYLT2 and RARB	↑	Active UC vs. quiescent UC
FOXA2, ROR1, NOTCH3, CDH17, PAD14, TNFSF8, EPHX1, HOXV2 and FRK	↓	Active UC vs. quiescent UC
SLIT2	↑	Active CD vs. quiescent CD, correlates with endoscopic and histological activity
Evaluation of disease behavior		
PAR2	↑	Total colitis phenotypes, steroid-dependent and refractory phenotypes of UC, and stricturing phenotypes
MDR1	↑	Total colitis phenotypes, younger onset of disease, and chronic continuous type of UC
CDH1, CDH13 and GDNF	↑	Long-standing disease course of UC
miR-1247 and CDX1	↑	Refractory UC and severe Mayo endoscopic score
RPS6KA2	↑	Stricturing/penetrating phenotypes of CD, and extensive disease of UC
Cancer surveillance		
RUNX3, MINT1, TGFB2, SLIT2, HS3ST2, TMEFF2, ITGA4, TFPI2, FOXE1, SYNE1 APC, CDH13, MGMT and MLH1	↑	Discriminate UC–CRC from controls
COX-2	↓	Discriminates UC–CRC from controls
miR-137	↑	Discriminates dysplasia and UC-CRC from controls
BMP3, vimentin, EYA4 and NDRG4	↑	Discriminate neoplasia and CRC from controls

TABLE 2 | Roles of miRNAs in IBD.

miRNAs	Expression levels (↑/↓)	Roles
Estimation of disease susceptibility		
miRs-3180-3p, miRplus-E1035 and miRplus-F1159	↑	Active UC vs. active CD and healthy controls
miR-20b, miR-98, miR-125b-1* and let-7e*	↑	Active UC vs. inactive UC, active CD, inactive CD, and healthy controls
miRs-103-2*, miR-362-3p, miR-532-3p, miR-20b, miR-98, miR-125b-1* and let-7e*	↑	UC vs. healthy controls
miR-340* and miR-484	↑	CD vs. healthy controls
Assessment of disease activity		
miR-16, miR-21, miR-24, miR-126, miR-203, miR-28-5p, miR-151-5p, miR-199a-5p, miR-340*, miRplus-E1271 and miR-595	↑	Active UC vs. quiescent UC
miR-200b and miR-124	↓	Active UC vs. quiescent UC
miR-199a-5p, miR-362-3p, miR-532-3p, miRplus-E1271, miR-877 and miR-595	↑	Active CD vs. quiescent CD
miRplus-F1065	↓	Active CD vs. quiescent CD
Evaluation of disease behavior		
miR-23b, miR-106 and miR-191	↑	Colonic CD
miR-19b and miR-629	↓	Colonic CD
miR-16, miR-21, miR-223 and miR-594	↑	Ileal CD
miR-29a, miR-29b, miR-29c, miR-19a-3p and miR-19b-3p	↓	Stricturing phenotypes of CD
miR-31-5p	↑	Stricturing and/or penetrating phenotypes
miR-196b-5p and miR-149-5p	↓	Stricturing and/or penetrating phenotypes
Cancer surveillance		
miR-31 and miR-224	↑	Discriminate dysplasia and CRC from controls, as well as IBD-associated CRC from sporadic CRC
miR-143 and miR-145	↓	Correlate with neoplastic progression of IBD
miR-21	↑	Discriminates dysplasia and CRC from controls
miR-155	↑	Correlates with neoplastic progression of IBD
miR-26b	↑	Discriminates UC–CRC from controls
miR-15b, miR-17, miR-26b and miR-145	↑	Discriminate CRC from controls

ranging from 0.79 to 0.98 (mean value of 0.93). However, no direct comparison between CD and UC has been made in their study. It is worth noting that most methylation changes occurred in proximity to GWAS risk loci. These results accord with a similar finding by Cooke et al. (Cooke et al., 2012). They demonstrated that many identified GWAS risk genes (*CARD9*, *CDH1*, *ICAM3* etc.) presented different methylation status between IBD patients (CD and UC) and healthy controls, suggesting a possibility of mechanistic interactions between the epigenetic and genetic signals. Existing data exhibited that referred SNPs could be located in CGIs, disrupt CpG sites, and therefore interfere CGI methylation states (Cooke et al., 2012). Meanwhile, methylation alterations in or in proximity to the transcription start site and the promoter region of susceptibility genes also exert great influence on gene transcription (Adams et al., 2014). This indicated that genetic risk loci might mediate effects on disease susceptibility through DNA methylation. In 2016, an epigenome-wide association study (EWAS) of 240 newly-diagnosed adult patients with IBD (CD and UC) and 190 controls successfully identified four DMRs (*VMP1*, *ITGB2*, *WDR8* and *CDC42BPB*) in CD versus controls, and two DMRs (*VMP1* and *WDR8*) in UC in comparison with controls, which paralleled the genomic findings that CD and UC not only have their own specific susceptibility loci, but also share overlapping risk loci to some extent. Furthermore, Ventham et al. (2016) also created a diagnostic model of 19 methylation probes that could distinguish CD from UC with a favorable sensitivity of 1 and acceptable accuracy of 0.719. Another 30-probe panel could differentiate IBD patients from controls with a sensitivity, a specificity, and an AUC of 0.812, 0.847, and 0.898, respectively (Ventham et al., 2016). Recently, a British research team has revealed distinct gut segment-specific DNA methylation patterns of intestinal epithelial cells (IECs) between pediatric IBD patients and healthy controls. Their data indicated that disease-specific DNA methylation profiles of IECs (ascending colon) could accurately separate IBD patients from healthy controls with a sensitivity of 75% and a specificity of 100%. Moreover, another ileal methylation signatures were capable of distinguishing CD from UC with a precision of 77% and an AUC of 0.92 (sensitivity of 57%, specificity of 100%) (Howell et al., 2018). Such a high degree of diagnostic value suggests its potential utility in clinical settings. Successful application of DNA methylation markers in cancer detection and surveillance has paved new ways for IBD research. Compared to genetic biomarkers, DNA methylation incorporates cumulative or specific environmental experience (such as smoking and diet) and the influence of age. Besides, current methylation detection encompasses panels of multiple methylation markers rather than a single marker, showing its superiority in sensitivity and specificity (Laird, 2003). Furthermore, DNA methylation biomarkers are stable in the bloodstream, tissues and even in stool, making it convenient to be preserved and detected (Johnson et al., 2016). Moreover, methylation assays for individual DNA methylation surrogate tend to be universal, which is similar to genetic markers (Laird, 2003). However, there are still some factors limiting the routine clinical application. Firstly, as is well known, DNA methylation signatures are cell-specific. Different sampling sites may exhibit a marked

difference in DNA methylation profiles due to different types of cells located in these sites (Cooke et al., 2012). Secondly, substantial (45%) overlap of differentially methylated positions (DMPs) between UC and CD might bring additional hurdles with regard to discriminating between them (McDermott et al., 2016). Thirdly, limitations of technologies applied in DNA methylation analyses significantly restrict clinical translation. The bisulphite-based approaches are still the leading methods used in this field. High quality samples and DNA sequence bias are important and serious challenges for a long time. Although whole genome bisulfite sequencing (WGBS) has displayed advantages in sample requirement, high coverage, and less DNA sequence bias, additional efforts are still in pressing need in order to resolve difficulties in PCR polymerase and bisulfite conversion (Raine et al., 2017). Fourthly, expensive testing costs need to be taken into consideration, which can add to the financial burden and thus, decrease patient acceptance. DNA methylation markers are indeed a powerful and promising tool to make a diagnosis of IBD. However, more studies are warranted prior to their clinical application.

MicroRNAs (miRNAs) are a group of non-coding RNAs with a length of about 22 nucleotides, mediating RNA silencing and gene expression regulation at a post-transcriptional level (Fisher, 2015). Accumulated evidence has showed its critical contribution to disease onset and progression of IBD, which supports possibilities of exploring roles of miRNA markers in diagnosis and differential diagnosis. miRNA expression patterns significantly differ between IBD patients and healthy controls, CD patients and UC patients, as well as between patients in remission and those in active states. Wu et al. (Wu et al., 2011) identified a panel of three peripheral blood miRNAs (miRs-3180-3p, miRplus-E1035 and miRplus-F1159) that were differentially expressed in active UC patients and healthy controls, and they also could distinguish active CD patients from UC patients. In the same study, specific miRNA expression panels of CD and UC have also been reported. Patients with UC displayed higher levels of miRs-103-2*, miR-362-3p, and miR-532-3p compared with healthy controls, irrespective of whether they were in remission or in active status. However, CD patients always displayed increased levels of miR-340* in peripheral blood. A further study has identified four specific miRNA surrogates (miR-20b, miR-98, miR-125b-1*, and let-7e*) in colonic mucosa of UC patients and claimed that they were differentially up-regulated by more than 5-fold in active UC in comparison to inactive UC, active CD, inactive CD, and healthy controls, driving its continuous development in IBD discrimination (Coskun et al., 2013). Zahm et al. (2011) tested the diagnostic ability of 11 serum miRNA markers in pediatric patients with CD, and found that these miRNA surrogates could accurately differentiate CD patients from controls with sensitivities higher than 80%. Among these identified miRNAs, miR-484 outstripped other miRNAs and promising markers, including C-reactive protein (CRP), anti-*Saccharomyces cerevisiae* antibody (ASCA) IgG, erythrocyte sedimentation rate (ESR) and albumin, with an AUC of 0.917, a sensitivity of 82.61%, and a specificity of 84.38%, respectively. However, the discriminative power of these CD-associated miRNAs in distinguishing CD from UC, CD from irritable bowel syndrome (IBS), and CD from celiac disease is unknown. More studies are warranted to elucidate the discriminative capacity with regard to these differential diagnoses. Even though peripheral blood and colon mucosa miRNA markers play a pivotal role in disease diagnosis, limitations including invasiveness, inflexibility, and time consumption make them unacceptable for patients. Saliva miRNA markers might overcome these shortcomings and provide additional diagnostic information. Different saliva miRNA expression signatures between IBD cases and healthy controls may help physicians in disease diagnosis and classification (Schaefer et al., 2015). In order to improve diagnostic accuracy, extended panels may be more helpful. A study of 76 IBD (CD and UC) patients and 38 healthy controls has established classification models comprising of various miRNAs (miR-34b-3p, miR-377-3p, miR-484, miR-574-5p etc.), which could discriminate IBD from healthy controls, and CD from UC, with increased AUCs of 0.89 to 0.98, and low classification error rates of 3.3% and 3.1%, respectively (Chamaillard et al., 2015). More importantly, some studies have observed a considerable overlap of miRNA signatures between IBD and other immune diseases (systemic lupus erythematosus, rheumatoid arthritis, asthma etc.), paralleling the genetic overlap between IBD and other immune diseases, which suggested some shared pathways among them; thereby offering a possibility of knowledge innovation in diagnosis and targeted treatment of IBD (Lees et al., 2011; Wu et al., 2011; Clark et al., 2012). In addition, it is important to note that clear differences of miRNA expression signatures have also been observed in different studies, that is to say, increased levels of miRNAs that were identified in one study otherwise showed a decreased expression in another study, or altered miRNAs couldn't be validated in other studies, which made it somewhat difficult for physicians to make an accurate diagnosis. In addition to different miRNA microarray platforms and sample sizes, other influencing factors such as different sample resources (colon tissues, peripheral blood, stool, saliva etc.) and inconsistent fold change criteria, as well as different therapeutic regimens, disease states (active or quiescent), and disease duration may also account for it (Coskun et al., 2013; Kalla et al., 2015; Schaefer et al., 2015). Thus, these reported miRNA markers are needed to be validated in large-scale, independent, clinically well-matched cohorts.

As for histone modifications and nucleosome positioning, definite evidence is still lacking for the contributions in diagnosis and differential diagnosis of IBD. Available evidence demonstrated complex networks between DNA methylation, miRNAs, histone modifications and nucleosome positioning. (Wang et al., 2013). So, determined DNA methylation or miRNA markers may affect disease susceptibility through histone modifications or nucleosome positioning at some levels. Therefore, further studies are warranted to clarify the detailed interactions, functional pathways and transcription regulation amongst these epigenetic modifications.

Diagnosis and differential diagnosis of IBD are definitely a major clinical challenge. Collection of additional evidence might help achieve a higher diagnostic accuracy of IBD. Emerging molecular markers such as DNA methylation and miRNA markers, along with other surrogates such as *NOD2*, ASCA, antineutrophil cytoplasmic antibody (ANCA), fecal calprotectin (FC) and fecal lactoferrin (FL), have exhibited certain advantages over other classic surrogates with regard to the sensitivity, specificity and accuracy (Zhang et al., 2018). A pooled analysis of different-class markers ensures a more precise diagnosis,

but cost–effectiveness ratio should also be taken into account. Although most of these emerging molecular markers have not been recommended in any guidelines, and are not usually generalized in routine diagnosis, they indeed provide some useful diagnostic information for doctors. Considering that some results were obtained in small sample studies, verification in larger, well-designed, and prospective studies has become increasingly important.

Assessment of Disease Activity

The natural course of IBD is characterized by relapse-remission. A population-based study from Copenhagen has delineated that approximately 18% patients could experience an indolent course, with 57% undergoing moderate activity (no less than two relapses within the first five years, but less than every year), and 25% having aggressive disease (disease relapses every year) during the first 5 years after diagnosis of CD (Jess et al., 2007). The corresponding percentages of UC of indolent, moderate, and aggressive disease course were 13%, 74% and 13%, respectively (Jess et al., 2007). IBD patients with earlier recurrence are at higher risk of relapsing during following years than those with later relapse (Magro et al., 2017). In routine clinical work, patients with relapse are recommended to get microbiological examination of stool, serological tests such as ESR and CRP, and even sigmoidoscopy or colonoscopy, aiming to exclude specific infections and assess disease activity. However, classic markers are not always parallel to disease activity. Some patients with mild or moderate disease activity may display normal serological parameters (Magro et al., 2017). Additionally, other diseases such as infectious enteritis and intestinal tuberculosis can also result in abnormal levels of ESR and CRP, making them unspecific for IBD (Zhang et al., 2018). Even though endoscopy together with histological analysis is recognized as the gold standard for the assessment of disease activity, it is unreasonable to prescribe endoscopy for patients once the disease flares. In recent years, novel epigenetic markers are claimed to be independently correlated with disease activity, and be of practical significance in the assessment of disease activity.

Saito et al. (2011) analyzed colonic methylation levels of UC patients and found that inflamed mucosa exhibited markedly higher methylation status of cadherin 1 (*CDH1*) and glial cell derived neurotrophic factor (*GDNF*) loci compared with quiescent mucosa. Recently, Barnicle et al. (2017). compared the DNA methylation patterns in inflamed and non-inflamed tissues of UC patients, and successfully found four differentially methylated and expressed genes (*ROR1, GXYLT2, RARB,* and *FOXA2*) that were involved in the regulation of Wnt signaling and cell development. A further study of 38 IBD patients (29 UC and 9 CD) revealed a significant correlation between slit guidance ligand 2 (*SLIT2*) methylation and endoscopic and histological activity (Lobatón, 2014). It should be pointed out that altered methylation status was also correlated with changed endoscopic activity in the longitudinal study. *SLIT2* methylation status tended to be elevated in patients who shifted from remission to active states. In addition, a large-scale systematic review of 16 studies further identified 25 differentially methylated inflammatory genes between UC patients and controls (Gould et al., 2016). Among these genes, methylation status of multidrug resistance 1 (*MDR1*), fragile X mental retardation 1 (*FMR1*), *CDH1* and *GDNF* gene was elevated, while methylation status of *NOTCH3*, *CDH17*, *PAD14, TNFSF8, EPHX1, HOXV2, FRK* etc. was decreased in inflamed mucosa in comparison with quiescent mucosa, indicating that histologic methylation profiles can serve as valuable surrogates to evaluate the disease activity of IBD. Associations between serum methylation signatures and disease activity have also been corroborated by several other studies (Gould et al., 2016). However, a recent genome-wide DNA methylation study has drawn a contrary conclusion that peripheral blood DNA methylation was not significantly different between active and inactive disease states (McDermott et al., 2016). Considering great heterogeneity of disease locations, disease duration, disease behaviors, degrees of disease activity, and drug use might affect the epigenetic changes, large scale, well-matched, prospective studies are needed to further verify the relationships between them. It must be stressed that most methylated loci have been confirmed to be IBD susceptibility loci by GWAS, while some methylated loci were firstly identified in these epigenetic studies, and were demonstrated to be involved previously unknown signaling pathways. In this sense, this offered a possibility of unveiling new pathogenic mechanisms of IBD and developing new targets for treatment (Saito et al., 2011; Lin et al., 2012; McDermott et al., 2016). Given that blood collection is more accessible and less invasive than biopsy, some studies compared the DNA methylation changes between peripheral blood and intestinal tissues, and suggested that methylation profiles in peripheral blood could reflect DNA methylation patterns in intestinal tissues (Gould et al., 2016; McDermott et al., 2016). Thus, identification of serum methylation signatures may be more acceptable in the assessment of disease activity. However, there is still a lack of studies directly comparing the diagnostic accuracy of methylation markers with other classic and emerging markers, additional efforts should be made to fill this gap.

miRNAs were firstly reported to be of value in the evaluation of disease activity of IBD in 2008 (Wu et al., 2008). Expression levels of miR-16, miR-21, miR-24, miR-126 and miR-203 were increased in active UC tissues in comparison with quiescent UC tissues. In contrast, miR-200b displayed a lower expression concentration in active UC tissues than in inactive ones (Wu et al., 2008). Among these differentially expressed miRNAs, miR-21 showed the highest fold change of 3.7 between active and inactive disease states. It should be emphasized that no difference has been found in the expression levels of the active UC-associated miRNAs between CD patients and controls. A later study also confirmed that peripheral blood miRNAs could distinguish active IBD from quiescent IBD (Wu et al., 2011). Their data demonstrated that active CD patients displayed an increased expression level of miR-199a-5p, miR-362-3p, miR-532-3p and miRplus-E1271 as well as a decreased level of miRplus-F1065, compared with CD patients in remission. Similarly, as for UC patients, miR-28-5p, miR-151-5p, miR-199a-5p, miR-340* and miRplus-E1271 were elevated in active ones but not in inactive ones. Moreover, miRs-3180-3p, miRplus-E1035 and miRplus-F1159 were demonstrated to be differentially expressed in the active UC patients vs active CD

patients, which supported the hypothesis that the two subtypes of IBD were implicated in different pathogenic mechanisms. Additional serum or tissue miRNA markers such as miR-124, miR-877, miR-595 etc. are also claimed to be instrumental in discriminating active IBD from inactive IBD (Iborra et al., 2013; Koukos et al., 2013; Krissansen et al., 2015). In 2016, a spearman correlation analysis indicated that circulating miR-223 was not only correlated with ESR and hs-CRP, but also correlated with clinical activity index including Crohn's Disease Activity Index (CDAI), Simplified Endoscopic Score for Crohn's Disease (SES-CD), Mayo score, and Ulcerative Colitis Endoscopic Index of Severity (UCEIS) (Wang et al., 2016). However, little is known about the definite predictive values (sensitivity, specificity and AUC) of these miRNAs in detection of disease activity in IBD. Moreover, there is still a lack of evidence about comparative advantages of miRNA markers when compared with other accurate markers such as serum calprotectin (SC), FC and FL. Another important issue that should be stressed is that whether serum expression profiles of IBD-associated miRNAs can reflect miRNA expression patterns in intestinal tissues. Contrary results have been found in some studies (Archanioti et al., 2011; Iborra et al., 2013; Zhang et al., 2018). So, larger comparison studies of paired serum and mucosal tissues are warranted. This further merits additional investigation to see if the combined analysis of serum and histologic miRNA profiles will ensure a more accurate assessment of disease activity.

Evaluation of Disease Behavior

IBD is a heterogeneous entity with distinct disease locations, age of onset, phenotypes, and severity. A majority of patients experience great changes of disease behaviors throughout the disease course. For example, some CD patients with inflammatory phenotypes may convert into stricturing or penetrating phenotypes, and UC patients manifesting proctitis will develop into extensive colitis as the disease progresses. Some convincing evidence suggests that early age onset, extensive disease, the presence of perianal disease, and stricturing or penetrating subtypes are risk factors of progressive course and poor prognosis (Gomollon et al., 2017; Magro et al., 2017). Screening patients with a less favorable course in the early stage of disease is highly recommended. So, it is of paramount importance to identify markers that can help physicians evaluate disease behavior in clinical practice.

Tahara et al. (2009b) firstly demonstrated that protease-activated receptor 2 (*PAR2*) methylation status was independently associated with various clinical disease behaviors in a study of 84 UC patients. Their data indicated that methylation levels of *PAR2* tended to be higher in patients with total colitis in comparison to those with rectal colitis, and increased methylation levels were also correlated with steroid-dependent and steroid-refractory phenotypes. In the same year, Christerson et al. (2009) suggested that PAR2 activation could potentiate intestinal myofibroblast proliferation and stricture formation in patients with CD. Considering that *PAR2* is widely implicated in the regulation of inflammatory responses, cell growth, and stricture formation in IBD, PAR2 methylation markers may serve as a valuable tool in the assessment of disease behavior (Christerson et al., 2009; Tahara et al., 2009b). In the same year, Tahara's research team further identified the putative roles of *MDR1* methylation signatures in UC patients. They suggested that increased methylation levels of *MDR1* gene were not only associated with total colitis phenotypes, but also correlated with younger onset of disease (≤20 years) and chronic continuous types (Tahara et al., 2009a). Available evidence has demonstrated a close association between *MDR1* dysfunction and impaired intestinal epithelial barrier in UC. Moreover, those patients who had progressive disease course were more likely to present severely damaged intestinal epithelial barrier (Schwab et al., 2003; Tahara et al., 2009a). From this point, *MDR1* methylation surrogates may be of important value and significance in evaluation of disease course. Further evidence has indicated that *CDH1*, *CDH13* and *GDNF* methylation occurred more frequently in UC patients with long-standing disease course, and higher methylation status of miR-1247 and caudal type homeobox 1 (*CDX1*) could serve as a predictor of refractory UC and severe Mayo endoscopic score (Saito et al., 2011; Schneider-Stock et al., 2014; Gould et al., 2016). In addition, hypomethylation of ribosomal protein S6 kinase A2 (*RPS6KA2*) has also been identified as a diagnostic aid in the prediction of complicated disease behavior (stricturing/penetrating disease) of CD and extensive disease of UC (Ventham et al., 2016). *RPS6KA2* is a ribosomal kinase that is responsible for the modulation of cell growth, motility and proliferation, as well as the regulation of PI3K/Akt/mTor pathway and autophagy. The latter has been proven to be one of the most important pathogenesis of CD in recent years. Previous studies have declared that gene expression is characterized by region-specificity in intestine (Bates et al., 2002). Given that DNA methylation can regulate gene expression at a post-transcriptional level, a significant difference of DNA methylation status in different segments of intestine may help explain the underlying molecular basis. Moreover, close associations between methylation status and certain disease behaviors highlight the exciting potential of using methylation markers in the assessment and prediction of disease behaviors. However, comparative studies are still in need to assess the exact predictive value in IBD, and extended panels of different molecular markers are also required to improve the accuracy of prediction.

The fact that the expression of miRNAs in intestine is region-specific, provides a basis for studying the specific miRNA expression patterns in IBD patients with different disease locations. Wu et al. (2010) have successfully identified three specifically upregulated miRNAs (miR-23b, miR-106 and miR-191) and two down-regulated miRNAs (miR-19b and miR-629) in tissues from colonic CD, and four miRNAs (miR-16, miR-21, miR-223, and miR-594) with increased expression in tissues from ileal CD, offering a possibility of using miRNA biomarkers to discriminate different subtypes of CD. Moreover, a British study highly suggested that the expression levels of miR-29 family were in correlation with stricturing phenotypes in CD patients (Nijhuis et al., 2014). They conducted a comparative study of mucosa overlying a stricture and paired non-stricturing samples in CD patients, and claimed that expression levels of miR-29a, miR-29b and miR-29c were significantly down-regulated in mucosa overlying a stricture compared with the other. Similarly,

in the serum, there is also a great reduction of expression levels of miR-29a in patients with stricturing phenotypes in comparison to those manifesting inflammatory phenotypes. This data was in accordance with previous findings that a decreased level of miR-29 family is a hallmark of cardiac, hepatic, pulmonary, and renal fibrosis, suggesting its significant contribution in tissue fibrosis (Nijhuis et al., 2014; Lewis et al., 2015b). A later study of 106 patients with CD further suggested that reduced serum expression levels of miR-19-3p (miR-19a-3p and miR-19b-3p) were independently associated with stricturing phenotypes (Lewis et al., 2015a). More importantly, further evidence showed that decreased serum miR-19-3p levels antedated the development of stricture, and remained low in patients with resected strictures. In addition, Lewis et al. (2015a) compared the predictive value of miR-19-3p, disease duration and ileal disease in discriminating stricturing from non-stricturing subtypes, and found that disease duration outperformed the other indicators with an AUC of 0.76, followed by miR-19-3p (AUC = 0.67) and ileal disease (AUC = 0.58). Combined analysis of the three predictors would make the classification efficiency increase markedly, with an AUC of 0.81. Additionally, some other miRNA markers such as miR-31-5p, miR-196b-5p, miR-149-5p etc. have also been confirmed in association with stricturing and/or penetrating phenotypes (Peck et al., 2015). Distinct expression patterns of miRNAs are of high value as a diagnostic and predictive tool in classifying different disease behaviors of IBD patients. Current diagnostic modalities displayed a limited value in discriminating inflammatory stricture from fibrotic stricture, while miR-29 family showed a great potential in identifying stricturing subtypes secondary to fibrosis (Nijhuis et al., 2014). Exploration of additional miRNA markers capable of classifying inflammatory and fibrotic stricture is in an unmet need, as this could guide clinicians in implementing individualized treatment (drug therapy, endoscopic balloon dilation or surgical intervention). In addition, establishing a standardized miRNA processing protocol is in dire need, considering that different RNA isolation methods and miRNA microarray platforms greatly influence the experimental results (Lewis et al., 2015a; Lewis et al., 2015b). Furthermore, functional significance and targeted sites of miRNA markers also deserve in-depth investigation in order to unveil the comprehensive molecular basis of IBD and develop miRNA-based therapeutics.

In the era of precision medicine, physicians are advised to perform risk stratification firstly according to the clinical characteristics, endoscopic findings, and imaging features, as well as molecular markers, and then select the most suitable treatment for individual patient based on risk stratification. Epigenetic patterns indeed provide some important clues for disease risk. Based on the risk analysis, patients can be divided into two groups including high risk group and low risk group, and the two different groups are supposed to receive different treatment regimens. The European Crohn's and Colitis Organization (ECCO) consensus recommends that patients with poor prognosis and progressive disease course better receive early and progressive therapy (immunomodulator or biological agents) and if possible, a combined treatment of immunosuppressant and biological agents. For patients with mild course, an accelerated step-up approach is recommended, which markedly decreases the unnecessary expenses and the risk of severe adverse events (Gomollon et al., 2017; Harbord et al., 2017). However, epigenetic markers have not been included in any guidelines for IBD treatment, suggesting many areas need to be improved. In addition, even we can choose de-escalation or escalation therapy according to risk stratification. The challenge remains to select the most suitable drugs for each individual amongst a variety of drugs, given that different patients show significantly different drug metabolism rates and response rates to therapy. Genetic markers such as thiopurine S-methyltransferase (*TPMT*), nucleoside diphosphate-linked moiety X-type motif 15 (*NUDT15*), and inosine triphosphate pyrophosphatase (*ITPA*) variant loci that implicate drug metabolism have been shown to be of great value in predicting therapeutic efficacy and adverse drug reactions of thiopurines (Lucafo et al., 2018b). With the wide use of biologics (infliximab, adalimumab, vedolizumab and ustekinumab) in clinic, emerging genetic markers (*IL23R, TNFAIP3 and TNFRSF1A*) and other serologic, histologic, and fecal surrogates (CRP, ANCA, membrane-bound TNF, TNF-α, FC etc.) represent as exciting indicators for the prediction of response rates to biologics (Zhang et al., 2018). As for epigenetic profiles, available data has shown that miR-499 was associated with steroid dependence, and a high level of lncRNA growth arrest-specific 5 (GAS5) was claimed to be correlated with poor steroid response (Okubo et al., 2011; Lucafo et al., 2018a). Serum let-7d and let-7e have been found to be candidate biomarkers for the prediction of treatment response to infliximab in CD patients (Fujioka et al., 2014). Moreover, DNA methylation patterns in IECs of pediatric IBD patients were also linked with the requirement of biologics and time to third treatment escalation (Howell et al., 2018). However, definite predictive values of these epigenetic markers are still absent, which limits the clinical application to some extent. Exploring the sensitivity, specificity, predictive accuracy in other prospective and independent cohorts is of utmost importance. Considering that there are still only a limited number of studies demonstrating the roles of epigenetics in the assessment and prediction of therapeutic response, and the selection of therapeutic methods, especially in the field of the immunosuppressant and biologics, additional studies are needed to replicate these findings and find more accurate epigenetic biomarkers.

Cancer Surveillance

IBD is a kind of long-lasting inflammatory disease with an increased risk of developing CRC, especially for patients with UC. Recent studies have shown that the cumulative risk of CRC is approximately 1.6% during fourteen-year follow-up, and UC increases the risk of CRC 2.4-fold in comparison with the normal population (Jess et al., 2012). Additionally, CRC risk increases over time as it is 8% at 20 years and 18% at 30 years after UC diagnosis (Eaden et al., 2001). Even though CRC in IBD merely accounts for a small portion (1–2%) of CRC cases in the general population, it contributes to 15% of all-causes mortality of IBD patients (Breynaert et al., 2008). Therefore, early detection and close surveillance of CRC in IBD patients are of paramount importance. Previous studies have demonstrated positive correlations between CRC and

young age at diagnosis, long disease duration, extensive colitis, male, primary sclerosing cholangitis and a family history of CRC (Jess et al., 2012; Azuara et al., 2013; Luo and Zhang, 2017; Zhen et al., 2018). Serum carcinoembryonic antigen (CEA) testing and fecal occult blood testing (FOB) are the most frequently used noninvasive means of detecting CRC (Ma et al., 2019). However, these detecting means have been claimed to be less efficient with unfavorable sensitivity and specificity. Exploring robust biomarkers has become an urgent need. Emerging biomarkers such as DNA methylation and miRNAs have showed their great potential in detection and surveillance of CRC.

It is well known that DNA methylation modifications occur early in neoplasia and can work as promising early-detection indicators of carcinogenesis. In 2010, Garrity-Park et al. (2010) assessed the methylation status of ten potential genes in intestinal biopsies, and revealed significant associations between runt related transcription factor 3 (*RUNX3*), *MINT1* (also known as *APBA1*) and *COX-2* methylation and UC–CRC (OR=12.6, 9.0 and 0.2, respectively). It is noteworthy that the concurrent presence of *RUNX3*/*MINT1* methylation and COX-2 unmethylation could substantially increase the possibility of UC-CRC (OR = 61.2 and 17.6, respectively). Two years later, Azuara et al. (2013) reported that the methylation status of transforming growth factor beta 2 (*TGFB2*), *SLIT2*, heparan sulfate-glucosamine 3-sulfotransferase 2 (*HS3ST2*), and transmembrane protein with EGF like and two follistatin like domains 2 (*TMEFF2*) in colorectal biopsies could be potential surrogates for an early diagnosis of colorectal dysplasia or CRC in high-risk patients with IBD. Methylation markers of *ITGA4*, *TFPI2*, *FOXE1*, *SYNE1*, *APC*, *CDH13, MGMT* and *MLH1* have also proven to be high-performance screening tools for estimating individual risk for CRC or colorectal neoplasia in IBD patients (Papadia et al., 2014; Gerecke et al., 2015; Scarpa et al., 2016). A recent study by Scarpa et al. (2016) clearly identified that any two or more methylated genes (*APC*, *CDH13*, *MGMT*, *MLH1* and *RUNX3*) in the non-neoplastic mucosa could predict CRC with a sensitivity of 57.1% and a specificity of 93.1%. Such a high specificity made these methylation markers to be an ideal rule-in test to detect CRC. In addition to DNA methylation markers, miRNA methylation patterns are also helpful in detection of CRC. A large study of 238 UC patients showed that methylation of miR-137 could distinguish UC patients with dysplasia or cancer from those without neoplasia with an AUC of 0.77, and miR-1, miR-9, miR-124, miR-137 and miR-34B/C work together could accurately quantify the risk for CRC, dysplasia and neoplasia with good AUC (Toiyama et al., 2017). Considering that low-grade dysplasia (LGD) is more closely associated with UC than with CRC, and LGD does not always progress to CRC, Garrity-Park et al. (2016) extended the scope of research to UC patients with LGD, and demonstrated critical roles of *MINT1* and *RUNX3* in the progression from LGD to CRC. In the same study, researchers also established a predictive model that comprised demographic, clinical, genetic, and epigenetic indicators for detection of synchronous neoplasm, which performed better than any other traditional and experimental model with an AUC of 0.92, a sensitivity of 82.8%, a specificity of 91.2%, a PPV of 95.1% and a NPV of 72.1% (Garrity-Park et al., 2016). In addition to histological methylation markers, methylation modifications in stool are also receiving attention. Kisiel et al. (2013) tested the exfoliated DNA markers in 50 IBD patients. Fecal BMP3, vimentin, EYA4, and NDRG4 methylation markers could accurately compartmentalize CRC from controls with an AUC of 0.97, 0.97, 0.95 and 0.85, respectively. At 89% specificity, methylation BMP3 in combination with methylation NDRG4 could diagnose 100% (9/9) of CRC and 80% (8/10) of dysplasia. A later study further confirmed the predictive ability of methylated BMP3 to detect colorectal neoplasia even in small IBD lesions (Johnson et al., 2016). All the above data clearly highlight the exciting potential of methylation markers in CRC detection and surveillance. Although colonoscopy with biopsy has been proven to be the gold standard for diagnosis and monitoring of CRC or colorectal neoplasia, it is a costly, time-consuming and invasive method. Moreover, its interpretation is subject to high interobserver variability. Methylation markers do provide adjuvant and valuable messages for adjustment of surveillance interval, and formulation of an individualized treatment plan in IBD patients at different risk. Stool and saliva DNA testing, as appealing non-invasive tests, improve the patient compliance in disease monitoring. However, sample size in some studies was quite small, which limited its argumentative strength and diagnostic efficacy. Moreover, some studies neglected the influence of intestinal inflammation and neoplasia on the levels of DNA, which consequently affected the levels of methylation DNA (Johnson et al., 2016). Additionally, the morbidity of CRC exhibited great ethnic differences. Larger studies of different races are also required. It is important to stress that IBD-associated and sporadic CRC patients showed a great difference in clinical features, histopathologic characteristics, and epigenetic changes (Garrity-Park et al., 2016). Many methylation markers including *SEPT9, TWIST1*, *TAC1*, *IGFBP3*, *EYA4* and *SST* have been claimed to be useful in the diagnosis and surveillance of sporadic CRC, while little is known about their roles in carcinogenesis of IBD (Kisiel et al., 2013; Ma et al., 2019). Therefore, prospective studies are desperately warranted to corroborate effects of those markers in IBD-associated CRC.

Insights from miRNA research have led to salient changes in our knowledge of biological processes of CRC in IBD patients. Aberrant expression profiles of miRNAs have been claimed to be associated with IBD-associated CRC. In 2011, a preliminary study identified significant differences of miRNA expression patterns between IBD-dysplasia tissues and inflamed colonic tissues, with 22 miRNAs increased and 10 miRNAs decreased in dysplastic tissues (Olaru et al., 2011). They surprisingly found miR-31 represented a stepwise increase in the progression from normal to chronic inflammation to neoplasia, with the highest levels in CRC, which indicated its potential for an early detection of dysplasia or CRC. In addition, a marked difference of miR-31 between IBD-associated CRC and sporadic CRC made it a favorable biomarker in discriminating between them. A later study also demonstrated the successive increase of miR-224 levels at each stage of IBD progression, and its excellent performance in distinguishing IBD-cancers from non-cancers (Olaru et al., 2013). Subsequent lines of evidence indicated that miR-143, miR-145, miR-21 and miR-155 were ancillary biomarkers in

the diagnosis and surveillance of IBD-associated carcinogenesis (Pekow et al., 2012; Ludwig et al., 2013; Wan et al., 2016). Relying on a single marker to detect CRC is not appropriate, establishing panels embodying different-class markers may further improve diagnostic accuracy. Benderska et al. (2015) have proven that a combined evaluation of ki-67 and miR-26b expression profiles could accurately detect 93% UC-associated colonic carcinoma. Its application in classifying different stages of CRC has also been confirmed. Recently, a Chinese research team developed a blood-based diagnostic model comprising of five circulating miRNA markers (miR-15b, miR-17, miR-21, miR-26b, and miR-145) and CEA, which could correctly diagnose CRC with an AUC of 0.85, followed by CEA of 0.793, and five-miRNA panel of 0.681 (Pan et al., 2017). However, due to the small sample size of this study, the feasibility of this diagnostic model has to be extensively studied in a larger cohort. miRNA surrogates are detectable, stable and quantifiable, with a high diagnostic and surveillance performance in discriminating CRC from controls. In this regard, miRNA biomarkers are of high clinical significance. However, many miRNA markers are not specific to CRC. Aberrant expression patterns identified in CRC are also present in other diseases. Additionally, different miRNA microarray platforms and cell types are also needed to be considered. Although the development of CRC diagnosis and monitoring is progressing at a fast pace, detection and surveillance of CRC remains challenging. Identifying more reliable markers, and establishing more robust diagnostic and surveillance models are becoming increasingly necessary. Elaborating on the roles of miRNAs in the pathogenesis and prognosis of CRC could further enhance our understanding of CRC, ultimately improve the survival quality and prognosis of patients.

FUNCTIONAL STUDY AND THERAPEUTIC TRANSLATION

IBD is a multifactorial disease derived from dysregulated immune responses in genetically susceptible individuals. Aberrant immunoregulation, impaired intestinal epithelial barrier, and abnormal autophagy significantly contribute to the complicated pathogenesis of IBD. Substantial evidence has demonstrated the widespread impacts of epigenetic patterns on IBD-associated signal pathways and functional changes, which facilitates a better understanding of the interactions between genetic and environmental factors, and provides an impetus for translational research on epigenetics-based therapeutics for patients with IBD. In this part, the functional impacts of epigenetic changes in the most extensively investigated pathways of IBD, and the roles of epigenetics in therapeutic translation will be discussed (**Table 3**).

T-cell differentiation and activation, antigen processing (recognition, presentation and binding), and cytokine production are the most studied fields of immunoregulation in IBD (CD and UC). PAR2 activation displayed pro-inflammatory and anti-inflammatory effects on colon, by promoting the production of T-helper cell type 1 (Th1) cytokines (TNF-α, IL-1 and IFN-γ), and the release of calcitonin gene related peptide (CGRP) respectively (Fiorucci et al., 2001; Cenac et al., 2002). Higher methylation levels of PAR2 are associated with severe phenotypes of UC (Tahara et al., 2009b), implying that accumulated inflammation and immune dysfunction derived from PAR2 methylation might result in severe disease behaviors of UC. Besides, PAR2 is also up-regulated by TNF-α (one of the most important mediator in CD and UC), and implicated in the activation of cytosolic phospholipase A2 (cPLA2) and proliferation of intestinal myofibroblast in CD patients, thereby playing a vital role in stricture formation of CD (Christerson et al., 2009). *RUNX3* is a tumor-suppressor gene that is implicated

TABLE 3 | Functional study of epigenetic modifications in IBD.

Epigenetic modifications	Functions
Immunoregulation	
PAR2	Regulates the production of inflammatory cytokines, and the proliferation of intestinal myofibroblast
RUNX3	Regulates T-cell development and TNF-β signaling pathways
TRAF6, IL12B, HLA-DOB, IL16, IGHG1 and THY1	Implicate in lymphocyte development, antigen processing, and cytokine responses
miR-155	Regulates the differentiation of T helper cells and the expression of proinflammatory cytokine, inhibits the expression FOXO3a and the NF-κB signaling pathway
miR-21	Mediates Th2 cell differentiation, modulates T-cell-mediated immune responses, involves in PTEN/PI3K/Akt signaling pathways, and disrupt intestinal epithelial barrier
miR-301a, miR-20b, miR-10a, miR-18a, miR-210, miR-223, miR-155, miR-26a and miR-21	Implicate IL23/Th17 pathways
miR-146a	Modulates Treg cells, dendritic cells and NK cells, and signaling pathways related to NOD2 and TLRs
miR-192, miR-20, miR-143, miR-150, miR-122, miR-29, miR-132, miR-495, miR-512 and miR-671	Implicate NOD2 signaling pathways
miR-146a, miR-144, miR-155, miR-132 and let-7	Implicate TLR signaling pathways
miR-124, let-7, miR-125, miR-26 and miR-101	Implicates STAT3 signaling pathway
Intestinal epithelial barrier	
CDH1	Encodes e-cadherin and mediates adherens junctions
MDR1	Involves in transmembrane transport and functional maintenance of intestinal epithelium
miR-21	Damages tight junctions and increases the permeability
miR-200b	Prevents intestinal inflammation, and protects tight junction and paracellular permeability
miR-122a	Increases the levels of zonulin and weakens the intestinal barrier
Autophagy	
miR-142-3P, miR-106b and miR-93	Decrease the ATG16L1-mediated autophagic activity
miR-196	Decreases the IRGM-mediated autophagic activity

in the pathophysiology of IBD and CRC. One of the IBD (CD and UC) susceptibility loci is located in the chromosomal region 1p36 where *RUNX3* resides (Brenner et al., 2004). *RUNX3* plays a certain role in T-cell development and TNF-β signaling pathways that are associated with the pathogenesis of both CD and UC. Studies have showed that *RUNX3* knockout mice presented over-responsiveness to antigens, over stimulation of T-cells, and spontaneous IBD (Brenner et al., 2004; Garrity-Park et al., 2010). Thus, it seems likely that RUNX3 methylation may contribute to the excessive inflammatory responses in both CD and UC. Moreover, UC–CRC cases presented much higher methylation levels of *RUNX3* than UC controls, indicating that *RUNX3* agonists might play an anti-inflammatory and anti-cancer role in clinical settings (Garrity-Park et al., 2010). Methylation modifications in other genes (*TRAF6*, *IL12B*, *HLA-DOB*, *IL16*, *IGHG1* and *THY1*) were also claimed to be either involved in T-cell or B-cell development, or implicated in antigen processing and cytokine responses, which provided a basis for drug discovery in the future (Gould et al., 2016). In addition to DNA methylation, miRNAs are also implicated in several immunoregulation processes related to IBD. Overexpression of miR-155 mediates a bias towards Th1 differentiation, while loss of miR-155 is prone to Th2 differentiation (Kalla et al., 2015). Previous evidence has suggested that CD was associated with Th1 and Th17 cytokine patterns, whereas UC was thought to be correlated with Th2-mediated inflammation (Brand, 2009). Up-regulated miR-155 exerts a pro-inflammatory effect by inhibiting the expression of Forkhead box O3 (FOXO3a) and therefore promotes the expression of inflammatory cytokines, and IBD-associated NF-κB signaling pathway. However, a deficiency in miR-155 shows a protective effect on experimental colitis by diminishing the expression of proinflammatory cytokine (TNF-α, IL-6, IL-12, IL-17, and IFN-γ), weakening the activation of T-cells, and repressing the Th1-mediated immune responses (Wan et al., 2016). miR-21 is overexpressed in patients with IBD. It mainly mediates UC-associated pathophysiological processes, including Th2 cell differentiation, T-cell-mediated immune responses, PTEN/PI3K/Akt signaling pathway, and the disruption of intestinal epithelial barrier (Kalla et al., 2015; Moein et al., 2019). miR-21 knockout mice with experimental dextran sulfate sodium (DSS) colitis showed an improved survival rate and less inflammation and injury in tissues when compared with wild type mice (Shi et al., 2013). Taking this into consideration, miR-21 inhibition may be a promising therapeutic target for UC patients. In addition, miR-21 also plays a central role in IL23/Th17 axis. IL23/Th17 signaling pathway has been reported to contribute greatly to the pathogenesis of CD. GWAS have identified several susceptibility genes of CD (*IL23R*, *IL12B*, *JAK2*, *STAT3*, *CCR6* and *TNFSF15*) that were involved in IL23/Th17 signaling pathway. Th17 is a novel kind of proinflammatory cell, and is implicated in the intestinal inflammation of CD by promoting the production of proinflammatory cytokines (IL17A, IL17F, IL21, IL22 and IL26) and chemokines (CCL20) (Brand, 2009). Other miRNAs implicated in IL23/Th17 pathways include miR-301a, miR-20b, miR-10a, miR-18a, miR-210, miR-223, miR-155, miR-26a and miR-21 (He et al., 2016; Moein et al., 2019). Recent studies have identified a direct and positive regulatory effect of miR-301a on the differentiation of Th17 cells and the production of proinflammatory cytokines through down regulation of Smad Nuclear Interacting Protein 1 (SNIP1) (He et al., 2016). In this respect, blockers of miR-301a may be a promising therapeutic intervention for CD patients. miR-146a involves the modulation of Treg cells, dendritic cells and NK cells, and signaling pathways related to NOD2 and Toll-like receptors (TLRs) (Kalla et al., 2015; Moein et al., 2019). NOD2 and TLRs are most integral parts in the pathogenesis of IBD, especially for CD. NOD2 can recognize the bacteria-derived muramyl dipeptide (MDP), and activate the NF-κB and caspase3 signaling pathways, and then, produces proinflammatory cytokines and regulates the innate and adaptive immunity of intestine (Kullberg et al., 2008). Moreover, it is also involved in the maintenance of the mucosal antibacterial barrier by regulating the expression of alpha-defensin and beta-defensin (Wehkamp et al., 2004; Voss et al., 2006). Thus, NOD2 variant/deficiency is a certain contributor to the development of CD. Existing data revealed that miR-192 and miR-20 showed inhibitory effects on the expression of *NOD2*, while miR-143 and miR-150 influenced the *NOD2* by targeting the important mediators of *NOD2* signaling pathway. miR-122, miR-29, miR-132, miR-495, miR-512 and miR-671 are other miRNAs associated with the *NOD2* signal and IBD pathogenesis. It's noteworthy that miR-122 designed for Hepatitis C infections is the first miRNA-based therapies in human clinical trials, which hold a great promise for future clinical research in other diseases such as IBD (Janssen et al., 2013). Additionally, an agent targeting miR-29 was also undergoing phase II clinical trials, with the aim of preventing tissue fibrosis. With regard to TLRs, TLR4 is largely activated by the lipopolysaccharide (LPS)-LPS-binding protein (LBP)-CD14 complex, and then triggers the NF-κB signaling pathway and promotes the production of proinflammatory cytokines (Chow et al., 1999). Besides, it is also proposed that TLR4-mediated signals can be modulated by NOD2, and NOD2 mutations can damage the cross-tolerance between NOD2 and TLR4, thus increasing the risk of CD (Kullberg et al., 2008). Available evidence indicated that miR-146a targets TLR4 signaling pathways and plays an anti-inflammatory role in CD, while miR-144 targets TLR2 and serves as a pro-inflammatory marker (Kalla et al., 2015; Moein et al., 2019). Other miRNAs associated with TLR signaling pathways include miR-155, miR-132 and let-7 (Koukos et al., 2013; Moein et al., 2019). Signal transducer and activator of transcription 3 (STAT3) signaling pathway is another vital transduction pathway, which is responsible for prolonging the survival of pathogenic T cells, and exacerbating inflammatory responses, therefore contributing to the pathogenesis of both CD and UC (Sugimoto, 2008). Koukos et al. (2013) have indicated that miR-124, let-7, miR-125, miR-26, and miR-101 could decrease STAT3 phosphorylation, and thereby suppress the inflammatory responses in UC patients. Amongst these miRNAs, miR-124 outperformed others, and showed a decreased level in active states in comparison to quiescent states of UC patients. Collectively, epigenetic patterns show a widespread influence on immunological functions associated with IBD, which provides some new druggable receptors for novel therapeutics. Some miRNA agonists and antagonists have

been developed and successfully applied in mouse models of colitis. For example, treatment with miR-155 antagonists alleviates the inflammatory responses in DSS-induced colitis mouse model (Lu et al., 2017). He et al. (2016) devised miR-301a antisense oligonucleotide and administrated it in trinitrobenzene sulphonic acid (TNBS)-induced mouse colitis model. As a result, a notable decrease in IL-17A cells and pro-inflammatory cytokines has been noticed in the inflamed tissues. Remarkable results gained in animal studies provide a strong driving force for translational studies and for developing novel epigenetics-based therapeutics for patients with IBD.

The impairment of intestinal epithelial barrier is one of the most critical pathogenic factors for IBD, especially for UC. Accumulated evidence has revealed that intestinal epithelial barrier has an established effect on defending against pathogenic microorganism invasion and colonization, preventing toxin translocation, and maintaining immune balance (Latiano et al., 2008; Consortium et al., 2009). IBD patients and even individuals at high risk of developing IBD could present impaired cell-cell junction and increased intestinal permeability (Wolters et al., 2011). Several genes including *CDH1*, *LAMB1*, *HNF4A* and *MYO9B* that are involved in the maintenance of epithelial barrier function have been claimed to be risk genes of UC (Latiano et al., 2008; Wolters et al., 2011). *CDH1* gene is located within the IBD1 locus, and encodes e-cadherin and mediates adherens junctions of colonic epithelia. Its decreased expression level and increased methylation status have been found in active UC and CRC tissues, suggesting a possibility of using CDH1 methylation marker to classify active disease from inactive disease, and CRC from healthy controls (Saito et al., 2011; Cooke et al., 2012). Similar to *CDH1*, *MDR1* gene also encompasses susceptibility loci of UC. It is involved in transmembrane transport and functional maintenance of intestinal epithelium (Tahara et al., 2009a). Mice lacking *MDR1a* gene spontaneously suffered from UC-like intestinal inflammation (Panwala et al., 1998; Ho et al., 2005). And the expression levels of *MDR1* in DSS-induced colitis mouse model and UC patients were reduced in comparison to healthy controls (Ho et al., 2005). Higher methylation levels of *MDR1* in inflammatory tissues relative to normal tissues of UC patients further supported the protective effects of *MDR1* in intestinal epithelium (Tahara et al., 2009a). In addition to methylation profiles, different kinds of miRNAs also showed their protective or destructive function in intestinal barrier. miR-21 damages tight junctions and increases the permeability of intestine through targeting RhoB and PTEN/PI3K/Akt pathways (Yang et al., 2013; Moein et al., 2019). It also regulates the malignant phenotypes of CRC by reducing the phosphatase and tensin homolog (PTEN), indicating a possibility of evaluating the CRC transformation and progression by it. Whereas, miR-200b exerts a protective effect on intestinal inflammation, tight junction, and paracellular permeability by down regulating the expression of IL-8 secondary to the activation of TNF-α, and inhibiting the destabilization of claudin 1 and zonula occludens-1 (ZO-1) (Shen et al., 2017). miR-122a weakens the intestinal barrier by targeting the EGFR pathways and increases the levels of zonulin, thereby increasing intestinal permeability, promoting pathogen invasion, and aggravating intestinal inflammation. Additionally, miR-191a, miR-93, miR-150, miR-675 and miR-874 also can affect functions of intestinal epithelial barrier (Moein et al., 2019). Altogether, diverse epigenetic modifications exert facilitating or damaging effects on intestinal epithelial barrier, which proves a novel avenue for IBD treatment. Producing antagomirs or miRNA mimics that are involved in regulation of intestinal epithelial barrier may be fruitful in future. Unfortunately, there is still no ongoing trial targeting these miRNAs for IBD. Instead, a trail targeting miR-122, miR-196 and miR-34 for glioblastoma multiforme and metastatic breast cancer is in the preclinical phase. Thus, continuous efforts are required to achieve translational research.

Successfully unveiling the contribution of autophagy to the pathogenesis of IBD has been a milestone achievement in the field of IBD research. Autophagy is dynamic cellular recycling process that is responsible for the degradation of abnormal cytoplasmic component (Kim and Lee, 2014). Recent studies have claimed that autophagy greatly affected the pathogenesis of IBD (especially of CD) by modulating the process of pathogen clearance, antimicrobial peptide secretion, inflammatory response, antigen presentation, and the endoplasmic reticulum (ER) stress response (Hooper et al., 2017; Iida et al., 2017). *ATG16L1*, *NOD2* and *IRGM* are the most investigated autophagy-related genes in CD. The interplay between autophagy-related genes and different miRNA offers deep insights into pathophysiological mechanisms of CD. miR-142-3P, miR-106b and miR-93 are claimed to target *ATG16L1*, while miR-196 is involved in IRGM-mediated autophagy. miR-142-3p directly reduces the mRNA and protein levels of ATG16L1, thereby decreasing starvation-induced and L18-MDP-induced autophagic activity (Zhai et al., 2014). A hallmark study revealed that miR-106b was increased while ATG16L1 was decreased in intestinal tissues of active CD patients in comparison to controls. miR-106b and miR-93 were claimed to target *ATG16L1* mRNA, thereby inhibiting the expression levels of ATG16L1 and damaging autophagy-mediated bacteria eradication. Antagonists for miR-106b and miR-93 facilitated the formation of autophagosomes, thus, alleviating intestinal inflammation (Lu et al., 2014). As for miR-196, several studies have seen an increase of it in patients with CD (Zhang et al., 2018). Overexpressed miR-196 can down regulate the protective variant (c.313C) in *IRGM*, thereby causing a disturbance in the regulation of *IRGM*. As a result, the expression levels of *IGRM* and efficacy of autophagy are diminished, and the growth of CD-associated intracellular bacteria (Adherent Invasive *Escherichia coli*, AIEC) is out of control, leading to an increased risk of developing CD (Brest et al., 2011). On the basis of this, miRNA-based regulation in IRGM-dependent autophagy may play a certain role in CD. On the other hand, it may open up a new research direction in autophagy and drug development of CD. Many approved drugs including corticosteroids, aminosalicylates, thiopurines, cyclosporin, tacrolimus and anti-TNF biologics exert their therapeutic effects by modulating signaling pathways that are often directly or indirectly associated with autophagy, but drugs targeting miRNA are still lacking (Hooper et al., 2017). Developing miRNA-based pharmacotherapy that specifically targets autophagy represents a promising therapeutic option for CD patients. However, the cell-type-specific feature of autophagy

makes it difficult to do autophagy-targeted drug discovery (Hooper et al., 2017). Further research is needed to resolve this difficulty.

Although histone alterations have been less studied in IBD, some studies still suggest its potential roles in disease. Acetylation of H4 was upregulated in inflamed tissues and Peyer's patches of CD patients and DSS-induced colitis models, highlighting its pro-inflammatory effects in colon (Tsaprouni et al., 2011). Treated with histone deacetylase (HDAC) inhibitors, mice consequently showed an apparent attenuation in intestinal inflammation. It's important to note that HDAC inhibitors have multiple targets including some other non-histone targets (TLR4, β-defensin 2, STAT3, P53 etc.), and are involved in a variety of IBD-associated signaling pathways such as NF-κB and Foxp3 transduction pathways (Tsaprouni et al., 2011; Ventham et al., 2013). In addition, tight links between lncRNA signatures and IBD-related inflammatory responses have also been described in several studies (Padua et al., 2016). Indeed, histone alterations and lncRNAs are important contributors for IBD activity, but associations with disease susceptibility, behaviors and prognosis are yet to be elucidated in the near future. Although some drugs targeting HDAC are used in clinical trials, most are designed for hematological malignancies and solid tumors. Therefore, annotation of the therapeutic utility of histone alterations and lncRNAs in IBD is also in dire need.

Dramatic success in development and application of biologics to IBD has brought IBD therapy into a new horizon. However, primary non-responders and secondary non-responders to biologics have still remained. Adverse reactions and high economic burden of existing biologic agents are real challenges in IBD treatment, highlighting the need of exploring new therapeutic strategies with good efficacy and less side effects for IBD patients. In-depth understanding of roles of epigenetic alterations in IBD susceptibility, activity, behaviors, and CRC provides a powerful driving force for the development of epigenetics-based therapeutics. Whereas, the process of therapeutic translation is in slow progress. Drug development as a whole is also being faced with numerous challenges. Firstly, currently used DNA methylating/demethylating agents show poor efficiency as a therapeutic modality due to the poor chemical stability, low specificity, and strong secondary effects (Gros et al. 2012). Azacitidine and decitabine are the two drugs approved by the US Food and Drug Administration (FDA) for myelodysplastic syndrome and acute myeloid leukemia, with common side effects such as hepatotoxicity and nephrotoxicity (Issa and Kantarjian, 2009). Constructing highly efficient and selective DNA methylation-based therapeutics is required. Secondly, since gut microbiota can regulate histone acetylation and methylation patterns of intestine, and epigenetic changes are cell/tissue-specific and time-dependent, identifying the biological impacts of gut microbiota on epigenetic patterns, and the etiological contributions of epigenetic modifications to gastrointestinal disorders remain difficult (Aleksandrova et al., 2017). Thirdly, delivery technologies for miRNA modulators to specific cell types and tissues, and off-target effects of miRNA-based therapeutics pose a major challenge for researchers. Fourthly, definite miRNA targets, exact mechanisms of action, and functional impacts of miRNAs should also be taken into account. In addition, more efforts are needed to annotate the long-term effects and pharmacokinetics, pharmacodynamics and pharmacogenetics of miRNA mimics or antagomirs *in vivo* (van Rooij and Kauppinen, 2014). Overcoming these difficulties at the earliest is of paramount importance.

CONCLUSIONS

IBD is an extremely complicated disease and poses a big challenge for physicians with regard to diagnosis and management of patients. In the era of precision medicine, we advocate that diagnosis, treatment and surveillance of diseases must be based on individual genetic markers, phenotypic characteristics, and psychosocial features (Chow et al., 2018). Substantial progress has been made in the genetic study of IBD, with numerous IBD-associated susceptibility loci identified. However, the identified genetic factors can explain only a small portion of overall disease variance, highlighting the need of uncovering the role of other factors such as epigenetic modifications in the occurrence and development of IBD. Epigenetic changes can mediate the interaction between genetics and environment, providing some critical information related to IBD pathogenesis. Recent years have seen a substantial advancement in epigenetics of IBD, particularly with relation to DNA methylation and miRNAs. Significant associations between epigenetic modifications and disease susceptibility, activity, behavior, and IBD-associated CRC have been shown in numerous studies, providing in-depth insights into the molecular basis of IBD, and additional diagnostic and monitoring tools for IBD patients. Several DNA methylation/miRNA-based panels for diagnosis and differential diagnosis, disease activity assessment, disease behavior evaluation, and CRC detection and surveillance have been developed, with good sensitivity, specificity and accuracy. Epigenetic markers are also candidate indicators for the selection of therapeutic methods and the prediction of therapeutic response. Functional studies have showed the significant impacts of epigenetic changes on the IBD-related immunoregulation, maintenance of intestinal epithelial barrier, and modulation of autophagy, notably in the most extensively investigated filed such as T-cell differentiation, IL23/Th17 and STAT3 signaling pathways, and intestinal permeability, which further enhance our knowledge of the biological processes of IBD. Based on the crucial contributions to IBD, pharmacological modulation of epigenetic patterns provides possibilities of therapeutic translation for the future clinical applications. However, current clinical trials or preclinical trials are focused on cancer treatment and obtain some preliminary achievements, providing a glimpse of translational potential of IBD-associated epigenetic modifications. Epigenetic research of IBD is in its infancy, and there are still some challenges to address. More endeavors are needed to compare the performance of epigenetic surrogates with classical and emerging markers, and to establish more robust diagnostic and monitoring panels comprising of different-class of markers. Continuous

efforts should also be made to construct highly efficient and selective therapeutics, identify targets and functional impacts of epigenetic modifications, improve delivery technologies for miRNAs, and elucidate biological effects of gut microbiota on epigenetic patterns. Moreover, considering that histone modifications and nucleosome positioning and other non-coding RNAs such as siRNA, piRNA and lncRNA are less studied in the field of IBD, further efforts should be made to identify the roles of these epigenetic changes in the pathogenesis of IBD. Therefore, it can be concluded that epigenetics plays a critical role in the pathogenesis of IBD, and holds a promise for disease diagnosis and surveillance, as well as for risk prediction and therapeutic innovation.

AUTHOR CONTRIBUTIONS

ZZ and HZ outlined the overall manuscript, and ZZ drafted the manuscript; HZ supervised the preparation of the draft and edited it. AM helped write, proofread, and edit the final manuscript.

FUNDING

This Work was Supported by the National Natural Science Foundation of China [No. 81570502] and by the 1.3.5 Project for Disciplines of Excellence, West China Hospital, Sichuan University [Grant Number: ZYJC18037].

REFERENCES

Adams, A. T., Kennedy, N. A., Hansen, R., Ventham, N. T., O'Leary, K. R., Drummond, H. E., et al. (2014). Two-stage genome-wide methylation profiling in childhood-onset crohn's disease implicates epigenetic alterations at the vmp1/mir21 and hla locissss. *Inflamm. Bowel Dis.* 20 (10), 1784–1793. doi: 10.1097/MIB.0000000000000179

Ahmad, T., Satsangi, J., McGovern, D., Bunce, M., and Jewell, D. P. (2001). Review article: the genetics of inflammatory bowel disease. *Aliment. Pharmacol. Ther.* 15 (6), 731–748. doi: 10.1046/j.1365-2036.2001.00981.x

Aleksandrova, K., Romero-Mosquera, B., and Hernandez, V. (2017). Diet, gut microbiome and epigenetics: emerging links with inflammatory bowel diseases and prospects for management and prevention. *Nutrients* 9 (9), E962. doi: 10.3390/nu9090962

Archanioti, P., Gazouli, M., Theodoropoulos, G., Vaiopoulou, A., and Nikiteas, N. (2011). Micro-RNAs as regulators and possible diagnostic bio-markers in inflammatory bowel disease. *J. Crohns Colitis* 5 (6), 520–524. doi: 10.1016/j.crohns.2011.05.007

Azuara, D., Rodriguez-Moranta, F., de Oca, J., Sanjuan, X., Guardiola, J., Lobaton, T., et al. (2013). Novel methylation panel for the early detection of neoplasia in high-risk ulcerative colitis and crohn's colitis patients. *Inflamm. Bowel Dis.* 19 (1), 165–173. doi: 10.1002/ibd.22994

Barnicle, A., Seoighe, C., Greally, J. M., Golden, A., and Egan, L. J. (2017). Inflammation-associated DNA methylation patterns in epithelium of ulcerative colitis. *Epigenetics* 12 (8), 591–606. doi: 10.1080/15592294.2017.1334023

Bates, M. D., Erwin, C. R., Sanford, L. P., Wiginton, D., Bezerra, J. A., Schatzman, L. C., et al. (2002). Novel genes and functional relationships in the adult mouse gastrointestinal tract identified by microarray analysis. *Gastroenterology* 122 (5), 1467–1482. doi: 10.1053/gast.2002.32975

Benderska, N., Dittrich, A. L., Knaup, S., Rau, T. T., Neufert, C., Wach, S., et al. (2015). miRNA-26b Overexpression in ulcerative colitis-associated carcinogenesis. *Inflamm. Bowel Dis.* 21 (9), 2039–2051. doi: 10.1097/MIB.0000000000000453

Brand, S. (2009). Crohn's disease: Th1, Th17 or both? The change of a paradigm: new immunological and genetic insights implicate Th17 cells in the pathogenesis of Crohn's disease. *Gut* 58 (8), 1152–1167. doi: 10.1136/gut.2008.163667

Brenner, O., Levanon, D., Negreanu, V., Golubkov, O., Fainaru, O., Woolf, E., et al. (2004). Loss of Runx3 function in leukocytes is associated with spontaneously developed colitis and gastric mucosal hyperplasia. *PNAS* 101 (45), 16016–16021. doi: 10.1073/pnas.0407180101

Brest, P., Lapaquette, P., Souidi, M., Lebrigand, K., Cesaro, A., Vouret-Craviari, V., et al. (2011). A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. *Nat. Genet.* 43 (3), 242–245. doi: 10.1038/ng.762

Breynaert, C., Vermeire, S., Rutgeerts, P., and Van Assche, G. (2008). Dysplasia and colorectal cancer in inflammatory bowel disease: a result of inflammation or an intrinsic risk? *Acta Gastroenterol. Belg.* 71 (4), 367–372.

Cenac, N., Coelho, A. M., Nguyen, C., Compton, S., Andrade-Gordon, P., MacNaughton, W. K., et al. (2002). Induction of intestinal inflammation in mouse by activation of proteinase-activated receptor-2. *Am. J. Pathol.* 161 (5), 1903–1915. doi: 10.1016/S0002-9440(10)64466-5

Chamaillard, M., Hübenthal, M., Hemmrich-Stanisak, G., Degenhardt, F., Szymczak, S., Du, Z., et al. (2015). Sparse modeling reveals mirna signatures for diagnostics of inflammatory bowel disease. *Plos One* 10 (10), e0140155. doi: 10.1371/journal.pone.0140155

Chow, J. C., Young, D. W., Golenbock, D. T., Christ, W. J., and Gusovsky, F. (1999). Toll-like receptor-4 mediates lipopolysaccharide-induced signal transduction. *J. Biol. Chem.* 274 (16), 10689–10692. doi: 10.1074/jbc.274.16.10689

Chow, N., Gallo, L., and Busse, J. W. (2018). Evidence-based medicine and precision medicine: complementary approaches to clinical decision-making. *Precis. Clin. Med.* 1 (2), 60–64. doi: 10.1093/pcmedi/pby009

Christerson, U., Keita, A. V., Soderholm, J. D., and Gustafson-Svard, C. (2009). Potential role of protease-activated receptor-2-stimulated activation of cytosolic phospholipase A(2) in intestinal myofibroblast proliferation: Implications for stricture formation in Crohn's disease. *J. Crohns Colitis* 3 (1), 15–24. doi: 10.1016/j.crohns.2008.10.002

Clark, P. M., Dawany, N., Dampier, W., Byers, S. W., Pestell, R. G., and Tozeren, A. (2012). Bioinformatics analysis reveals transcriptome and microRNA signatures and drug repositioning targets for IBD and other autoimmune diseases. *Inflamm. Bowel Dis.* 18 (12), 2315–2333. doi: 10.1002/ibd.22958

Consortium, U. I. G., Barrett, J. C., Lee, J. C., Lees, C. W., Prescott, N. J., Anderson, C. A., et al. (2009). Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. *Nat. Genet.* 41 (12), 1330–1334. doi: 10.1038/ng.483

Cooke, J., Zhang, H., Greger, L., Silva, A. L., Massey, D., Dawson, C., et al. (2012). Mucosal genome-wide methylation changes in inflammatory bowel disease. *Inflamm. Bowel Dis.* 18 (11), 2128–2137. doi: 10.1002/ibd.22942

Coskun, M., Bjerrum, J. T., Seidelin, J. B., Troelsen, J. T., Olsen, J., and Nielsen, O. H. (2013). miR-20b, miR-98, miR-125b-1*, and let-7e* as new potential diagnostic biomarkers in ulcerative colitis. *World J. Gastroenterol.* 19 (27), 4289–4299. doi: 10.3748/wjg.v19.i27.4289

de Lange, K. M., Moutsianas, L., Lee, J. C., Lamb, C. A., Luo, Y., Kennedy, N. A., et al. (2017). Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* 49 (2), 256–261. doi: 10.1038/ng.3760

Dong, Y., Xu, S., Liu, J., Ponnusamy, M., Zhao, Y., Zhang, Y., et al. (2018). Non-coding RNA-linked epigenetic regulation in cardiac hypertrophy. *Int. J. Biol. Sci.* 14 (9), 1133–1141. doi: 10.7150/ijbs.26215

Eaden, J. A., Abrams, K. R., and Mayberry, J. F. (2001). The risk of colorectal cancer in ulcerative colitis: a meta-analysis. *Gut* 48 (4), 526–535. doi: 10.1136/gut.48.4.526

Ellinghaus, D., Zhang, H., Zeissig, S., Lipinski, S., Till, A., Jiang, T., et al. (2013). Association between variants of prdm1 and ndp52 and crohn's disease, based on exome sequencing and functional studies. *Gastroenterology* 145 (2), 339–347. doi: 10.1053/j.gastro.2013.04.040

Fiorucci, S., Mencarelli, A., Palazzetti, B., Distrutti, E., Vergnolle, N., Hollenberg, M. D., et al. (2001). Proteinase-activated receptor 2 is an anti-inflammatory signal for colonic lamina propria lymphocytes in a mouse model of colitis. *PNAS* 98 (24), 13936–13941. doi: 10.1073/pnas.241377298

Fisher, K. (2015). MicroRNA in inflammatory bowel disease: Translational research and clinical implication. *World J. Gastroenterol.* 21 (43), 12274–12282. doi: 10.3748/wjg.v21.i43.12274

Fujioka, S., Nakamichi, I., Esaki, M., Asano, K., Matsumoto, T., and Kitazono, T. (2014). Serum microRNA levels in patients with Crohn's disease during induction therapy by infliximab. *J. Gastroenterol. Hepatol.* 29 (6), 1207–1214. doi: 10.1111/jgh.12523

Garrity-Park, M. M., Loftus, E. V., Bryant, S. C., and Smyrk, T. C. (2016). A biomarker panel to detect synchronous neoplasm in non-neoplastic surveillance biopsies from patients with ulcerative colitis. *Inflamm. Bowel Dis.* 22 (7), 1568–1574. doi: 10.1097/MIB.0000000000000789

Garrity-Park, M. M., Loftus, E. V., Jr, Sandborn, W. J., Bryant, S. C., and Smyrk, T. C. (2010). Methylation status of genes in non-neoplastic mucosa from patients with ulcerative colitis-associated colorectal cancer. *Am. J. Gastroenterol* 105 (7), 1610–1619. doi: 10.1038/ajg.2010.22

Gerecke, C., Scholtka, B., Löwenstein, Y., Fait, I., Gottschalk, U., Rogoll, D., et al. (2015). Hypermethylation of ITGA4, TFPI2 and VIMENTIN promoters is increased in inflamed colon tissue: putative risk markers for colitis-associated cancer. *J. Cancer Res. Clin. Oncol.* 141 (12), 2097–2107. doi: 10.1007/s00432-015-1972-8

Gomollon, F., Dignass, A., Annese, V., Tilg, H., Van Assche, G., Lindsay, J. O., et al. (2017). 3rd European Evidence-based consensus on the diagnosis and management of crohn's disease 2016: part 1: diagnosis and medical management. *J. Crohns Colitis* 11 (1), 3–25. doi: 10.1093/ecco-jcc/jjw168

Gould, N. J. D., Davidson, K. L. M., Nwokolo, C. U., and Arasaradnam, R. P. (2016). A systematic review of the role of DNA methylation on inflammatory genes in ulcerative colitis. *Epigenomics* 8 (5), 667–684. doi: 10.2217/epi-2016-0006

Gros, C., Fahy, J., Halby, L., Dufau, I., Erdmann, A., Gregoire, J. M., et al. (2012). DNA methylation inhibitors in cancer: recent and future approaches. *Biochimie* 94 (11), 2280–2296. doi: 10.1016/j.biochi.2012.07.025

Gutschner, T., and Diederichs, S. (2012). The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biol.* 9 (6), 703–719. doi: 10.4161/rna.20481

Halfvarson, J., Bodin, L., Tysk, C., Lindberg, E., and Jarnerot, G. (2003). Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics. *Gastroenterology* 124 (7), 1767–1773. doi: 10.1016/S0016-5085(03)00385-8

Harbord, M., Eliakim, R., Bettenworth, D., Karmiris, K., Katsanos, K., Kopylov, U., et al. (2017). Third european evidence-based consensus on diagnosis and management of ulcerative colitis. part 2: current management. *J. Crohns Colitis* 11 (7), 769–784. doi: 10.1093/ecco-jcc/jjx009

He, C., Shi, Y., Wu, R., Sun, M., Fang, L., Wu, W., et al. (2016). miR-301a promotes intestinal mucosal inflammation through induction of IL-17A and TNF-alpha in IBD. *Gut* 65 (12), 1938–1950. doi: 10.1136/gutjnl-2015-309389

Helbig, K. L., Nothnagel, M., Hampe, J., Balschun, T., Nikolaus, S., Schreiber, S., et al. (2012). A case-only study of gene-environment interaction between genetic susceptibility variants in NOD2 and cigarette smoking in Crohn's disease aetiology. *BMC Med. Genet.* 13, 14. doi: 10.1186/1471-2350-13-14

Ho, G. T., Gaya, D. R., and Satsangi, J. (2005). Multidrug resistance (MDR1) gene in inflammatory bowel disease: a key player? *Inflamm. Bowel Dis.* 11 (11), 1013–1019. doi: 10.1097/01.MIB.0000186488.53493.c3

Hooper, K. M., Barlow, P. G., Stevens, C., and Henderson, P. (2017). Inflammatory bowel disease drugs: a focus on autophagy. *J. Crohns Colitis* 11 (1), 118–127. doi: 10.1093/ecco-jcc/jjw127

Howell, K. J., Kraiczy, J., Nayak, K. M., Gasparetto, M., Ross, A., Lee, C., et al. (2018). DNA methylation and transcription patterns in intestinal epithelial cells from pediatric patients with inflammatory bowel diseases differentiate disease subtypes and associate with outcome. *Gastroenterology* 154 (3), 585–598. doi: 10.1053/j.gastro.2017.10.007

Huang, H., Weng, H., Zhou, K., Wu, T., Zhao, B. S., Sun, M., et al. (2019). Histone H3 trimethylation at lysine 36 guides m(6)A RNA modification co-transcriptionally. *Nature* 567 (7748), 414–419. doi: 10.1038/s41586-019-1016-7

Iborra, M., Bernuzzi, F., Correale, C., Vetrano, S., Fiorino, G., Beltrán, B., et al. (2013). Identification of serum and tissue micro-RNA expression profiles in different stages of inflammatory bowel disease. *Clin.Exp. Immunol.* 173 (2), 250–258. doi: 10.1111/cei.12104

Iida, T., Onodera, K., and Nakase, H. (2017). Role of autophagy in the pathogenesis of inflammatory bowel disease. *World J Gastroenterol.* 23 (11), 1944–1953. doi: 10.3748/wjg.v23.i11.1944

Issa, J. P., and Kantarjian, H. M. (2009). Targeting DNA methylation. *Clin. Cancer Res.* 15 (12), 3938–3946. doi: 10.1158/1078-0432.CCR-08-2783

Janssen, H. L., Reesink, H. W., Lawitz, E. J., Zeuzem, S., Rodriguez-Torres, M., Patel, K., et al. (2013). Treatment of HCV infection by targeting microRNA. *N. Engl. J. Med.* 368 (18), 1685–1694. doi: 10.1056/NEJMoa1209026

Jeltsch, A., Broche, J., and Bashtrykov, P. (2018). Molecular processes connecting dna methylation patterns with dna methyltransferases and histone modifications in mammalian genomes. *Genes (Basel)* 9 (11), E566. doi: 10.3390/genes9110566

Jess, T., Riis, L., Vind, I., Winther, K. V., Borg, S., Binder, V., et al. (2007). Changes in clinical characteristics, course, and prognosis of inflammatory bowel disease during the last 5 decades: a population-based study from Copenhagen, Denmark. *Inflamm. Bowel Dis.* 13 (4), 481–489. doi: 10.1002/ibd.20036

Jess, T., Rungoe, C., and Peyrin-Biroulet, L. (2012). Risk of colorectal cancer in patients with ulcerative colitis: a meta-analysis of population-based cohort studies. *Clin. Gastroenterol. Hepatol.* 10 (6), 639–645. doi: 10.1016/j.cgh.2012.01.010

Johnson, D. H., Taylor, W. R., Aboelsoud, M. M., Foote, P. H., Yab, T. C., Cao, X., et al. (2016). DNA methylation and mutation of small colonic neoplasms in ulcerative colitis and crohn's colitis: implications for surveillance. *Inflamm. Bowel Dis.* 22 (7), 1559–1567. doi: 10.1097/MIB.0000000000000795

Kalla, R., Ventham, N. T., Kennedy, N. A., Quintana, J. F., Nimmo, E. R., Buck, A. H., et al. (2015). MicroRNAs: new players in IBD. *Gut* 64 (3), 504–513. doi: 10.1136/gutjnl-2014-307891

Kang, K., Bae, J. H., Han, K., Kim, E. S., Kim, T. O., and Yi, J. M. (2016). A genome-wide methylation approach identifies a new hypermethylated gene panel in ulcerative colitis. *Int. J. Mol. Sci.* 17 (8), E1291. doi: 10.3390/ijms17081291

Kaplan, G. G., and Ng, S. C. (2017). Understanding and preventing the global increase of inflammatory bowel disease. *Gastroenterology* 152 (2), 313–321. e312. doi: 10.1053/j.gastro.2016.10.020

Karatzas, P. S., Gazouli, M., Safioleas, M., and Mantzaris, G. J. (2014). DNA methylation changes in inflammatory bowel disease. *Ann. Gastroenterol.* 27 (2), 125–132.

Kaser, A., Zeissig, S., and Blumberg, R. S. (2010). Inflammatory bowel disease. *Annu. Rev. Immunol.* 28, 573–621. doi: 10.1146/annurev-immunol-030409-101225

Kim, K. H., and Lee, M. S. (2014). Autophagy–a key player in cellular and body metabolism. *Nat. Rev. Endocrinol.* 10 (6), 322–337. doi: 10.1038/nrendo.2014.35

Kisiel, J. B., Yab, T. C., Nazer Hussain, F. T., Taylor, W. R., Garrity-Park, M. M., Sandborn, W. J., et al. (2013). Stool DNA testing for the detection of colorectal neoplasia in patients with inflammatory bowel disease. *Aliment. Pharmacol. Ther.* 37 (5), 546–554. doi: 10.1111/apt.12218

Koukos, G., Polytarchou, C., Kaplan, J. L., Morley-Fletcher, A., Gras-Miralles, B., Kokkotou, E., et al. (2013). MicroRNA-124 regulates STAT3 expression and is down-regulated in colon tissues of pediatric patients with ulcerative colitis. *Gastroenterology* 145 (4), 842-852 e842. doi: 10.1053/j.gastro.2013.07.001

Krissansen, G. W., Yang, Y., McQueen, F. M., Leung, E., Peek, D., Chan, Y. C., et al. (2015). Overexpression of miR-595 and miR-1246 in the sera of patients with active forms of inflammatory bowel disease. *Inflamm. Bowel Dis.* 21 (3), 520–530. doi: 10.1097/MIB.0000000000000285

Kullberg, B. J., Ferwerda, G., de Jong, D. J., Drenth, J. P., Joosten, L. A., Van der Meer, J. W., et al. (2008). Crohn's disease patients homozygous for the 3020insC NOD2 mutation have a defective NOD2/TLR4 cross-tolerance to intestinal stimuli. *Immunology* 123 (4), 600–605. doi: 10.1111/j.1365-2567.2007.02735.x

Laird, P. W. (2003). The power and the promise of DNA methylation markers. *Nat. Rev. Cancer* 3 (4), 253–266. doi: 10.1038/nrc1045

Latiano, A., Palmieri, O., Valvano, M. R., D'Inca, R., Caprilli, R., Cucchiara, S., et al. (2008). The association of MYO9B gene in Italian patients with inflammatory bowel diseases. *Aliment. Pharmacol. Ther.* 27 (3), 241–248. doi: 10.1111/j.1365-2036.2007.03551.x

Lees, C. W., Barrett, J. C., Parkes, M., and Satsangi, J. (2011). New IBD genetics: common pathways with other diseases. *Gut* 60 (12), 1739–1753. doi: 10.1136/gut.2009.199679

Lewis, A., Mehta, S., Hanna, L. N., Rogalski, L. A., Jeffery, R., Nijhuis, A., et al. (2015a). Low serum levels of microrna-19 are associated with a stricturing crohn's disease phenotype. *Inflamm. Bowel Dis.* 21 (8), 1926–1934. doi: 10.1097/MIB.0000000000000443

Lewis, A., Nijhuis, A., Mehta, S., Kumagai, T., Feakins, R., Lindsay, J. O., et al. (2015b). Intestinal fibrosis in Crohn's disease: role of microRNAs as fibrogenic modulators, serum biomarkers, and therapeutic targets. *Inflamm. Bowel Dis.* 21 (5), 1141–1150. doi: 10.1097/MIB.0000000000000298

Li, L., Fu, K., Zhou, W., and Snyder, M. (2019). Applying circulating tumor DNA methylation in the diagnosis of lung cancer. *Precis. Clin. Med.* 2 (1), 45–56. doi: 10.1093/pcmedi/pbz003

Li, W. (2018). Precision medicine: to cure and relieve more. *Precis. Clin. Med.* 1 (1), 3–4. doi: 10.1093/pcmedi/pby002

Lin, Z., Hegarty, J. P., Yu, W., Cappel, J. A., Chen, X., Faber, P. W., et al. (2012). Identification of disease-associated DNA methylation in B cells from Crohn's disease and ulcerative colitis patients. *Dig. Dis. Sci.* 57 (12), 3145–3153. doi: 10.1007/s10620-012-2288-z

Liu, J. Z., van Sommeren, S., Huang, H., Ng, S. C., Alberts, R., Takahashi, A., et al. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* 47 (9), 979–986. doi: 10.1038/ng.3359

Lobatón, T. (2014). Relationship between methylation and colonic inflammation in inflammatory bowel disease. *World J. Gastroenterol.* 20 (30), 10591–10598. doi: 10.3748/wjg.v20.i30.10591

Lu, C., Chen, J., Xu, H. G., Zhou, X., He, Q., Li, Y. L., et al. (2014). MIR106B and MIR93 prevent removal of bacteria from epithelial cells by disrupting ATG16L1-mediated autophagy. *Gastroenterology* 146 (1), 188–199. doi: 10.1053/j.gastro.2013.09.006

Lu, Z. J., Wu, J. J., Jiang, W. L., Xiao, J. H., Tao, K. Z., Ma, L., et al. (2017). MicroRNA-155 promotes the pathogenesis of experimental colitis by repressing SHIP-1 expression. *World J. Gastroenterol.* 23 (6), 976–985. doi: 10.3748/wjg.v23.i6.976

Lucafo, M., Di Silvestre, A., Romano, M., Avian, A., Antonelli, R., Martelossi, S., et al. (2018a). Role of the long non-coding rna growth arrest-specific 5 in glucocorticoid response in children with inflammatory bowel disease. *Basic Clin. Pharmacol. Toxicol.* 122 (1), 87–93. doi: 10.1111/bcpt.12851

Lucafo, M., Franca, R., Selvestrel, D., Curci, D., Pugnetti, L., Decorti, G., et al. (2018b). Pharmacogenetics of treatments for inflammatory bowel disease. *Expert Opin. Drug Metab. Toxicol.* 14 (12), 1209–1223. doi: 10.1080/17425255.2018.1551876

Ludwig, K., Fassan, M., Mescoli, C., Pizzi, M., Balistreri, M., Albertoni, L., et al. (2013). PDCD4/miR-21 dysregulation in inflammatory bowel disease-associated carcinogenesis. *Virchows Arch.* 462 (1), 57–63. doi: 10.1007/s00428-012-1345-5

Luo, C., and Zhang, H. (2017). The role of proinflammatory pathways in the pathogenesis of colitis-associated colorectal cancer. *Mediators Inflamm.* 2017, 5126048. doi: 10.1155/2017/5126048

Ma, Z., Williams, M., Cheng, Y. Y., and Leung, W. K. (2019). Roles of methylated dna biomarkers in patients with colorectal cancer. *Dis. Markers* 2019, 2673543. doi: 10.1155/2019/2673543

Magro, F., Gionchetti, P., Eliakim, R., Ardizzone, S., Armuzzi, A., Barreiro-de Acosta, M., et al. (2017). Third european evidence-based consensus on diagnosis and management of ulcerative colitis. part 1: definitions, diagnosis, extra-intestinal manifestations, pregnancy, cancer surveillance, surgery, and ileo-anal pouch disorders. *J. Crohns Colitis* 11 (6), 649–670. doi: 10.1093/ecco-jcc/jjx008

McDermott, E., Ryan, E. J., Tosetto, M., Gibson, D., Burrage, J., Keegan, D., et al. (2016). DNA methylation profiling in inflammatory bowel disease provides new insights into disease pathogenesis. *J. Crohns Colitis* 10 (1), 77–86. doi: 10.1093/ecco-jcc/jjv176

Moein, S., Vaghari-Tabari, M., Qujeq, D., Majidinia, M., Nabavi, S. M., and Yousefi, B. (2019). MiRNAs and inflammatory bowel disease: An interesting new story. *J. Cell Physiol.* 234 (4), 3277–3293. doi: 10.1002/jcp.27173

Nijhuis, A., Biancheri, P., Lewis, A., Bishop, C. L., Giuffrida, P., Chan, C., et al. (2014). In Crohn's disease fibrosis-reduced expression of the miR-29 family enhances collagen expression in intestinal fibroblasts. *Clin. Sci. (Lond)* 127 (5), 341–350. doi: 10.1042/CS20140048

Okubo, M., Tahara, T., Shibata, T., Yamashita, H., Nakamura, M., Yoshioka, D., et al. (2011). Association study of common genetic variants in pre-microRNAs in patients with ulcerative colitis. *J. Clin. Immunol.* 31 (1), 69–73. doi: 10.1007/s10875-010-9461-y

Olaru, A. V., Selaru, F. M., Mori, Y., Vazquez, C., David, S., Paun, B., et al. (2011). Dynamic changes in the expression of MicroRNA-31 during inflammatory bowel disease-associated neoplastic transformation. *Inflamm. Bowel Dis.* 17 (1), 221–231. doi: 10.1002/ibd.21359

Olaru, A. V., Yamanaka, S., Vazquez, C., Mori, Y., Cheng, Y., Abraham, J. M., et al. (2013). MicroRNA-224 negatively regulates p21 expression during late neoplastic progression in inflammatory bowel disease. *Inflamm. Bowel Dis.* 19 (3), 471–480. doi: 10.1097/MIB.0b013e31827e78eb

Padua, D., Mahurkar-Joshi, S., Law, I. K., Polytarchou, C., Vu, J. P., Pisegna, J. R., et al. (2016). A long noncoding RNA signature for ulcerative colitis identifies IFNG-AS1 as an enhancer of inflammation. *Am. J. Physiol. Gastrointest. Liver Physiol.* 311 (3), G446–G457. doi: 10.1152/ajpgi.00212.2016

Pan, C., Yan, X. B., Li, H., Huang, L. S., Yin, M. M., Yang, Y. Z., et al. (2017). Systematic literature review and clinical validation of circulating microRNAs as diagnostic biomarkers for colorectal cancer. *Oncotarget* 8 (40), 68317–68328. doi: 10.18632/oncotarget.19344

Panwala, C. M., Jones, J. C., and Viney, J. L. (1998). A novel model of inflammatory bowel disease: mice deficient for the multiple drug resistance gene, mdr1a, spontaneously develop colitis. *J. Immunol.* 161 (10), 5733–5744.

Papadia, C., Louwagie, J., Del Rio, P., Grooteclaes, M., Coruzzi, A., Montana, C., et al. (2014). FOXE1 and SYNE1 genes hypermethylation panel as promising biomarker in colitis-associated colorectal neoplasia. *Inflamm. Bowel Dis.* 20 (2), 271–277. doi: 10.1097/01.MIB.0000435443.07237.ed

Peck, B. C., Weiser, M., Lee, S. E., Gipson, G. R., Iyer, V. B., Sartor, R. B., et al. (2015). MicroRNAs classify different disease behavior phenotypes of crohn's disease and may have prognostic utility. *Inflamm. Bowel Dis.* 21 (9), 2178–2187. doi: 10.1097/MIB.0000000000000478

Pekow, J. R., Dougherty, U., Mustafi, R., Zhu, H., Kocherginsky, M., Rubin, D. T., et al. (2012). miR-143 and miR-145 are downregulated in ulcerative colitis: Putative regulators of inflammation and protooncogenes. *Inflamm. Bowel Dis.* 18 (1), 94–100. doi: 10.1002/ibd.21742

Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. *Nat. Biotechnol.* 28 (10), 1057–1068. doi: 10.1038/nbt.1685

Raine, A., Manlig, E., Wahlberg, P., Syvanen, A. C., and Nordlund, J. (2017). SPlinted Ligation Adapter Tagging (SPLAT), a novel library preparation method for whole genome bisulphite sequencing. *Nucleic Acids Res.* 45 (6), e36. doi: 10.1093/nar/gkw1110

Saito, S., Kato, J., Hiraoka, S., Horii, J., Suzuki, H., Higashi, R., et al. (2011). DNA methylation of colon mucosa in ulcerative colitis patients: Correlation with inflammatory status. *Inflamm. Bowel Dis.* 17 (9), 1955–1965. doi: 10.1002/ibd.21573

Satsangi, J., Jewell, D. P., Rosenberg, W. M. C., and Bell, J. I. (1994). Genetics of inflammatory bowel-disease. *Gut* 35 (5), 696–700. doi: 10.1136/gut.35.5.696

Satsangi, J., Silverberg, M. S., Vermeire, S., and Colombel, J. F. (2006). The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut* 55 (6), 749–753. doi: 10.1136/gut.2005.082909

Scarpa, M., Scarpa, M., Castagliuolo, I., Erroi, F., Kotsafti, A., Basato, S., et al. (2016). Aberrant gene methylation in non-neoplastic mucosa as a predictive marker of ulcerative colitis-associated CRC. *Oncotarget* 7 (9), 10322–10331. doi: 10.18632/oncotarget.7188

Schaefer, J. S., Attumi, T., Opekun, A. R., Abraham, B., Hou, J., Shelby, H., et al. (2015). MicroRNA signatures differentiate Crohn's disease from ulcerative colitis. *BMC Immunol.* 16, 5. doi: 10.1186/s12865-015-0069-0

Schneider-Stock, R., Tahara, T., Shibata, T., Okubo, M., Ishizuka, T., Nakamura, M., et al. (2014). DNA methylation status of epithelial-mesenchymal transition (emt) - related genes is associated with severe clinical phenotypes in ulcerative colitis (uc). *PLoS ONE* 9 (10), e107947. doi: 10.1371/journal.pone.0107947

Schwab, M., Schaeffeler, E., Marx, C., Fromm, M. F., Kaskas, B., Metzler, J., et al. (2003). Association between the C3435T MDR1 gene polymorphism and susceptibility for ulcerative colitis. *Gastroenterology* 124 (1), 26–33. doi: 10.1053/gast.2003.50010

Shen, Y., Zhou, M., Yan, J., Gong, Z., Xiao, Y., Zhang, C., et al. (2017). miR-200b inhibits TNF-alpha-induced IL-8 secretion and tight junction disruption of intestinal epithelial cells in vitro. *Am. J. Physiol. Gastrointest. Liver Physiol.* 312 (2), G123–g132. doi: 10.1152/ajpgi.00316.2016

Shi, C., Liang, Y., Yang, J., Xia, Y., Chen, H., Han, H., et al. (2013). MicroRNA-21 knockout improve the survival rate in DSS induced fatal colitis through protecting against inflammation and tissue injury. *PLoS One* 8 (6), e66814. doi: 10.1371/journal.pone.0066814

Sugimoto, K. (2008). Role of STAT3 in inflammatory bowel disease. *World J Gastroenterol.* 14 (33), 5110–5114. doi: 10.3748/wjg.14.5110

Tahara, T., Shibata, T., Nakamura, M., Yamashita, H., Yoshioka, D., Okubo, M., et al. (2009a). Effect of MDR1 gene promoter methylation in patients with ulcerative colitis. *Int. J. Mol. Med.* 23 (4), 521-527. doi: 10.3892/ijmm_00000160

Tahara, T., Shibata, T., Nakamura, M., Yamashita, H., Yoshioka, D., Okubo, M., et al. (2009b). Promoter methylation of protease-activated receptor (PAR2) is associated with severe clinical phenotypes of ulcerative colitis (UC). *Clin. Exp. Med.* 9 (2), 125–130. doi: 10.1007/s10238-008-0025-x

Tang, W. Y., and Ho, S. M. (2007). Epigenetic reprogramming and imprinting in origins of disease. *Rev. Endocr. Metab. Disord.* 8 (2), 173–182. doi: 10.1007/s11154-007-9042-4

Toiyama, Y., Okugawa, Y., Tanaka, K., Araki, T., Uchida, K., Hishida, A., et al. (2017). A panel of methylated microrna biomarkers for identifying high-risk patients with ulcerative colitis-associated colorectal cancer. *Gastroenterology* 153 (6), 1634–1646.e1638. doi: 10.1053/j.gastro.2017.08.037

Tsaprouni, L. G., Ito, K., Powell, J. J., Adcock, I. M., and Punchard, N. (2011). Differential patterns of histone acetylation in inflammatory bowel diseases. *J. Inflamm. (Lond)* 8 (1), 1. doi: 10.1186/1476-9255-8-1

van Rooij, E., and Kauppinen, S. (2014). Development of microRNA therapeutics is coming of age. *EMBO Mol. Med.* 6 (7), 851–864. doi: 10.15252/emmm.201100899

Ventham, N. T., Kennedy, N. A., Adams, A. T., Kalla, R., Heath, S., O'Leary, K. R., et al. (2016). Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease. *Nat. Commun.* 7, 13507. doi: 10.1038/ncomms13507

Ventham, N. T., Kennedy, N. A., Nimmo, E. R., and Satsangi, J. (2013). Beyond gene discovery in inflammatory bowel disease: the emerging role of epigenetics. *Gastroenterology* 145 (2), 293–308. doi: 10.1053/j.gastro.2013.05.050

Voss, E., Wehkamp, J., Wehkamp, K., Stange, E. F., Schroder, J. M., and Harder, J. (2006). NOD2/CARD15 mediates induction of the antimicrobial peptide human beta-defensin-2. *J. Biol. Chem.* 281 (4), 2005–2011. doi: 10.1074/jbc.M511044200

Wan, J., Xia, L., Xu, W., and Lu, N. (2016). Expression and function of mir-155 in diseases of the gastrointestinal tract. *Int. J. Mol. Sci.* 17 (5), E709. doi: 10.3390/ijms17050709

Wang, H., Zhang, S., Yu, Q., Yang, G., Guo, J., Li, M., et al. (2016). Circulating microrna223 is a new biomarker for inflammatory bowel disease. *Medicine* 95 (5), e2703. doi: 10.1097/MD.0000000000002703

Wang, Z., Yao, H., Lin, S., Zhu, X., Shen, Z., Lu, G., et al. (2013). Transcriptional and epigenetic regulation of human microRNAs. *Cancer Lett.* 331 (1), 1–10. doi: 10.1016/j.canlet.2012.12.006

Wehkamp, J., Harder, J., Weichenthal, M., Schwab, M., Schaffeler, E., Schlee, M., et al. (2004). NOD2 (CARD15) mutations in Crohn's disease are associated with diminished mucosal alpha-defensin expression. *Gut* 53 (11), 1658–1664. doi: 10.1136/gut.2003.032805

Weissman, S. M. (2018). Personalized medicine: a new horizon for medical therapy. *Precis. Clin. Med.* 1 (1), 1–2. doi: 10.1093/pcmedi/pby001

Wilusz, J. E., Sunwoo, H., and Spector, D. L. (2009). Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev.* 23 (13), 1494–1504. doi: 10.1101/gad.1800909

Winter, J., Jung, S., Keller, S., Gregory, R. I., and Diederichs, S. (2009). Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat. Cell Biol.* 11 (3), 228–234. doi: 10.1038/ncb0309-228

Wolters, V. M., Xu, W., Zhao, X., Walters, T. D., Griffiths, A. M., Silverberg, M. S., et al. (2011). Replication of genetic variation in the MYO9B gene in Crohn's disease. *Hum. Immunol.* 72 (7), 592–597. doi: 10.1016/j.humimm.2011.03.025

Wu, F., Guo, N. J., Tian, H. Y., Marohn, M., Gearhart, S., Bayless, T. M., et al. (2011). Peripheral blood micrornas distinguish active ulcerative colitis and crohn's disease. *Inflamm. Bowel Dis.* 17 (1), 241–250. doi: 10.1002/ibd.21450

Wu, F., Zhang, S., Dassopoulos, T., Harris, M. L., Bayless, T. M., Meltzer, S. J., et al. (2010). Identification of microRNAs associated with ileal and colonic Crohn's disease. *Inflamm. Bowel Dis.* 16 (10), 1729–1738. doi: 10.1002/ibd.21267

Wu, F., Zikusoka, M., Trindade, A., Dassopoulos, T., Harris, M. L., Bayless, T. M., et al. (2008). MicroRNAs are differentially expressed in ulcerative colitis and alter expression of macrophage inflammatory peptide-2 alpha. *Gastroenterology* 135 (5), 1624-1635 e1624. doi: 10.1053/j.gastro.2008.07.068

Yang, Y., Ma, Y., Shi, C., Chen, H., Zhang, H., Chen, N., et al. (2013). Overexpression of miR-21 in patients with ulcerative colitis impairs intestinal epithelial barrier function through targeting the Rho GTPase RhoB. *Biochem. Biophys. Res. Commun.* 434 (4), 746–752. doi: 10.1016/j.bbrc.2013.03.122

Zahm, A. M., Thayu, M., Hand, N. J., Horner, A., Leonard, M. B., and Friedman, J. R. (2011). Circulating microRNA is a biomarker of pediatric Crohn disease. *J. Pediatr. Gastroenterol. Nutr.* 53 (1), 26–33. doi: 10.1097/MPG.0b013e31822200cc

Zhai, Z., Wu, F., Dong, F., Chuang, A. Y., Messer, J. S., Boone, D. L., et al. (2014). Human autophagy gene ATG16L1 is post-transcriptionally regulated by MIR142-3p. *Autophagy* 10 (3), 468–479. doi: 10.4161/auto.27553

Zhang, H., Zeng, Z., Mukherjee, A., and Shen, B. (2018). Molecular diagnosis and classification of inflammatory bowel disease. *Expert Rev. Mol. Diagn.* 18 (10), 867–886. doi: 10.1080/14737159.2018.1516549

Zhen, Y., Luo, C., and Zhang, H. (2018). Early detection of ulcerative colitis-associated colorectal cancer. *Gastroenterol. Rep. (Oxf)* 6 (2), 83–92. doi: 10.1093/gastro/goy010

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Zeng, Mukherjee and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 07 November 2019
doi: 10.3389/fgene.2019.00909

Mutations in *KIAA1109*, *CACNA1C*, *BSN*, *AKAP13*, *CELSR2*, and *HELZ2* Are Associated With the Prognosis in Endometrial Cancer

*Zhiwei Qiao, Ying Jiang, Ling Wang, Lei Wang, Jing Jiang and Jingru Zhang **

The Department of Gynaecology, Liaoning Cancer Hospital & Institute, Cancer Hospital of China Medical University, Shengyang, China

Endometrial cancer (EC) is one of the most common gynecologic malignancies. Emerging studies had demonstrated the mutations in genes could serve as diagnostic or prognostic markers for human cancers. In this study, we screened mutated genes in EC and found that the mutations in *KIAA1109, CACNA1C, BSN, AKAP13, CELSR2,* and *HELZ2* were correlated to the overall survival time in patients with EC. Bioinformatics analysis showed *KIAA1109* was involved in regulating NIK/NF-kappaB signaling, *CACNA1C* was found to regulate cell migration and proliferation, *BSN* was found to regulate Wnt signaling pathway, *CELSR2* was involved in regulating cell–cell adhesion, nuclear import, and protein folding, *HELZ2* was found to regulate multiple immune related biological processes, and *AKAP13* was involved in regulating translation, mRNA nonsense-mediated decay, rRNA processing, translational initiation, and mRNA splicing *via* spliceosome. The findings provided a novel therapeutic strategy in patients with EC.

Keywords: endometrial cancer, bioinformatics analyses, mutation, overall survival time, biomarkers

OPEN ACCESS

Edited by:
Rui Henrique, Portuguese Oncology Institute, Portugal

Reviewed by:
Noritaka Yamaguchi, Chiba University, Japan
Jaime Prat, Autonomous University of Barcelona, Spain

****Correspondence:***
Jingru Zhang
yi85317870@163.com

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *12 May 2019*
Accepted: *28 August 2019*
Published: *07 November 2019*

Citation:
Qiao Z, Jiang Y, Wang L, Wang L, Jiang J and Zhang J (2019) Mutations in KIAA1109, CACNA1C, BSN, AKAP13, CELSR2, and HELZ2 Are Associated With the Prognosis in Endometrial Cancer. Front. Genet. 10:909. doi: 10.3389/fgene.2019.00909

INTRODUCTION

Endometrial cancer (EC) is one of the most common gynecologic malignancies (Attarha et al., 2011). Despite the prognosis of the early stage EC is good with a 5-year survival rate of 69–88% (Gottwald et al., 2010). However, the prognosis of metastatic EC remained very poor, with a median survival of 7–12 months. Therefore, there is an urgent need to identify novel biomarkers for the prognosis of EC. Moreover, the mechanisms underlying the progression of EC remained largely unclear.

With the development of next-generation sequencing, multiple EC related mutations were identified. Emerging studies had demonstrated the mutations in genes could serve as diagnostic or prognostic markers for human cancers. For example, McConechy et al. identified a series of mutations in *PTEN, CTNNB1, PIK3CA, ARID1A, ARID5B*, and *KRAS* were associated with EC (Mcconechy et al., 2012). The mutations in FGFR2 were associated with poor outcomes in endometrioid endometrial cancer (Jeske et al., 2017). The genetic alterations in CTCF could promote EC cell survival and alter cell polarity (Marshall et al., 2017). Jing et al. found that MUC16 mutations could improve patients' prognosis by enhancing the infiltration of cytotoxic T lymphocytes in the EC microenvironment (Jing and Jing, 2014).

The present study identified prognosis related gene mutations in EC by analyzing TCGA databases (Collins, 2007). The mutations in 6 genes were correlated to the overall survival time in patients with EC. Bioinformatics analysis was used to predict the potential functions of these

genes. The purpose of this study was to evaluate the impact of somatic tumor mutation on recurrence-free survival in this patient population.

MATERIALS AND METHODS

Data Mining With cBioPortal and TCGA Database

In this study, we identified the gene mutations in EC using TCGA database (https://portal.gdc.cancer.gov/). All searches were performed according to cBioPortal's online instructions (http://www.cbioportal.org/index.do) (Jianjiong et al., 2013). The survival analysis related to gene mutations was performed on the TCGA database (https://portal.gdc.cancer.gov/).

Co-Expression Network Analysis

In this study, the Pearson correlation coefficient was calculated according to the expression value between lncRNA–mRNA pair using cBioPortal's online instructions (http://www.cbioportal.org/index.do). The top 500 co-expressing genes were selected as potential targets of mutated genes in EC.

Bioinformatics Analysis

GO and KEGG pathway enrichment analysis were performed to determine the biological significance of DEGs, using the Database for Annotation, Visualization, and Integrated Discovery (Dennis et al., 2003) (DAVID; version 6.8; http://david.ncifcrf.gov/).

Patients' Prognostic Analyses

Survival curves were depicted using the Kaplan-Meier method and compared with log-rank test. Cox proportional hazards regression analysis was used for univariate and multivariate analyses to explore the association of clinical features, gene mutational status, and patients' prognosis. All the prognostic analyses were conducted by survival R package.

Statistical Analysis

The two groups were compared using Student's t-test. Overall survival time analyses were estimated using the Kaplan-Meier product-limit estimator, and then a log-rank test was conducted to compare wildtype and mutation status. Overall survival was measured from the date of surgery to the date of last contact or death. Patients alive were censored at the date of last contact or clinic visit. Stata v14.2 (College Station, TX) was used to conduct statistical analysis.

RESULTS

Screening of Mutated Genes in Endometrial Cancer

The present study analyzed TCGA database to identify mutated genes in EC. As shown in **Figure 1**, the top 50 mutated genes in EC included *TTN, MUC4, MUC16, PIK3CA, KMT2C, KMT2D, SYNE1, FLG, SYNE2, EP300, OBSCN, ADGRV1, RYR2, LRP1B, USH2A, MUC17, NEB, MDN1, MUC5B, CSMD1, PCLO, HUWE1, FBXW7, DMD, NSD1, NAV3, DNAH8, DST, PLEC, AHNAK2, LRP2, MKI67, DNAH2, TENM1, DNAH10, PRKDC, FAT1, TP53, HMCN1, ZFHX4, DNAH6, UBR4, NOTCH1, CREBBP, NIPBL, EYS, AHNAK, CSMD3, XIRP2,* and *MACF1*. Among these genes, *TTN*, *MUC4*, and *PIK3CA* are the most frequently mutated genes. The mutation rates in *TTN*, *MUC4*, and *PIK3CA* from the TCGA provisional data sets were 43.25% (125/289), 31.83% (92/289), and 29.41% (85/289), respectively.

The Somatic Mutations of *KIAA1109, CACNA1C, BSN, AKAP13, CELSR2,* and *HELZ2* Were Correlated to Overall Survival Time in Patients With EC

Next, we screened somatic mutations associated with overall survival time in patients with EC. As shown in **Figure 2**, Log-rank test showed that mutations in *KIAA1109, CACNA1C, BSN, AKAP13,* and *HELZ2* were significantly associated with the longer overall survival time in EC patients, however, mutations in CELSR2 were significantly associated with the shorter overall survival time in EC patients.

Mutation Profiles in *KIAA1109*, *CACNA1C*, *BSN*, *AKAP13*, *CELSR2*, and *HELZ2* in EC

The mutation rates in *KIAA1109, CACNA1C, BSN, AKAP13, CELSR2,* and *HELZ2* from the TCGA provisional data sets were 6.92% (20/289), 7.27% (21/289), 7.96% (23/289), 7.61% (22/289), 6.92% (20/289), and 7.27% (21/289), respectively in **Figure 3**. A, majority of mutations identified were missense and nonsense resulting in amino acid, changes and a truncation of these proteins. However, there was no evidence of a mutational hotspot in *KIAA1109, CACNA1C, BSN, AKAP13, CELSR2,* and *HELZ2* in EC patients (**Figure 4**).

The Effect of Mutations on mRNA Expressions of *KIAA1109, CACNA1C, BSN, AKAP13, CELSR2,* and *HELZ2* in EC Patients

Furthermore, we detected the effect of mutations in *KIAA1109, CACNA1C, BSN, AKAP13, CELSR2,* and *HELZ2* on mRNA expression based on the RNA-Seq data. As shown in **Figure 5**, we found the mutations in *CACNA1C*, *BSN*, *CELSR2*, and *HELZ2* did not result in a significant alteration of their mRNA levels. However, we found that the mRNA levels in *KIAA1109* and *AKAP13* mutated EC samples were lower than that in *KIAA1109* and *AKAP13* wild type EC samples.

Bioinformatics Analysis of *KIAA1109*, *CACNA1C*, *BSN*, *AKAP13*, *CELSR2*, and *HELZ2* in EC Patients

Furthermore, we performed bioinformatics analysis to reveal the potential functions of *KIAA1109, CACNA1C, BSN, AKAP13, CELSR2,* and *HELZ2* using their co-expressing mRNAs in EC patients. The present study selected the top 500 correlated genes as the potential targets of *KIAA1109, CACNA1C, BSN,*

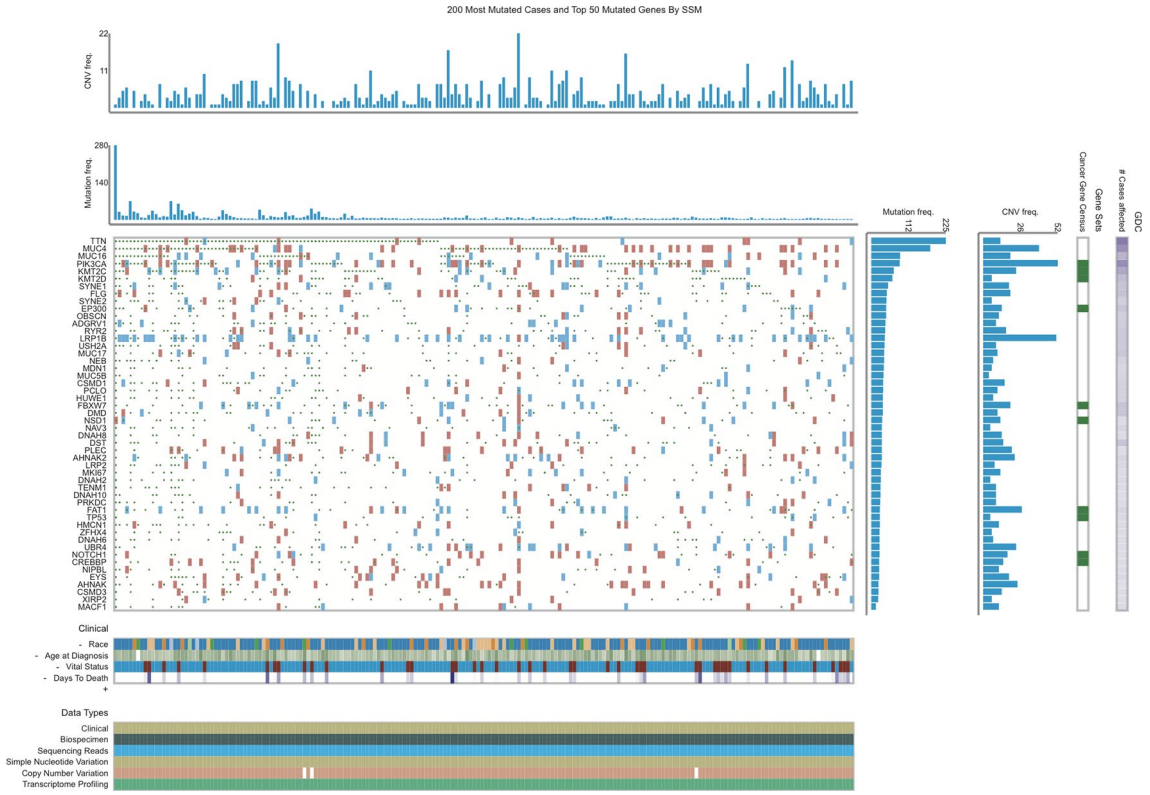


FIGURE 1 | Identification of mutated genes in EC using TCGA database.

AKAP13, CELSR2, and *HELZ2*. Bioinformatics analysis showed *KIAA1109* was involved in regulating rRNA processing, translation, transcription, NIK/NF-kappaB signaling, and histone acetylation. The results were shown in **Figure 6**. *CACNA1C* was involved in regulating collagen fibril organization, cell-matrix adhesion, cellular response to amino acid stimulus, cell adhesion, and negative regulation of cell proliferation. *BSN* was involved in regulating epidermis development, cilium movement, smoothened signaling pathway, Wnt signaling pathway, planar cell polarity pathway, and cilium morphogenesis. *AKAP13* was involved in regulating translation, mRNA nonsense-mediated decay, rRNA processing, translational initiation, and mRNA splicing *via* spliceosome. *CELSR2* was involved in regulating cell–cell adhesion, keratinocyte differentiation, spliceosomal snRNP assembly, nuclear import, and protein folding. *HELZ2* was involved in regulating type I interferon signaling pathway, innate immune response, immune response, inflammatory response, and T cell activation.

DISCUSSION

Endometrial cancer (EC) is one of the most common gynecologic malignancies. However, the mechanisms underlying EC progression remained unclear. Previous studies had showed the mutations in several genes were related to EC. For example, *MUC16* mutations improve EC prognosis through enhancing the infiltration of cytotoxic T lymphocytes. *PTEN* and *PIK3CA* mutations played crucial roles in grade 3 EC (Jing and Jing, 2014). The present study screened mutated genes in EC. Our results showed *TTN*, *MUC4*, and *PIK3CA* were the most frequently mutated genes in the EC, which was consistent with previous studies. Moreover, we identified the mutations in 6 genes were associated with the prognosis of EC. The results showed that mutations in *KIAA1109, CACNA1C, BSN, AKAP13,* and *HELZ2* were significantly associated with the longer overall survival time in EC patients. However, mutations in *CELSR2* were significantly associated with the shorter overall survival time in EC patients. These results suggested the important roles of these genes in the progression and prognosis of EC.

KIAA1109, located on the chromosome 4, was reported to be associated with susceptibility to celiac disease. Of note, 2 recent studies indicated *KIAA1109* was associated with the prognosis of human cancers. For example, Qing et al. reported mutations in *KIAA1109, DNAH5* and *KCNH7* were associated with poor survival of Chinese esophageal squamous cell carcinoma patients (Tao et al., 2017). Tindall et al. found genetic variation of *KIAA1109* might be associated with prostate cancer susceptibility in men with a family history of the disease (Tindall et al., 2010). *CACNA1C* gene encodes an alpha-1 subunit of a voltage-dependent calcium channel (Fayi et al., 2016). The mutations in *CACNA1C* were observed in various types of human diseases, such as ventricular fibrillation, and schizophrenia (Charles et al., 2007). Previous studies showed *CACNA1C* was

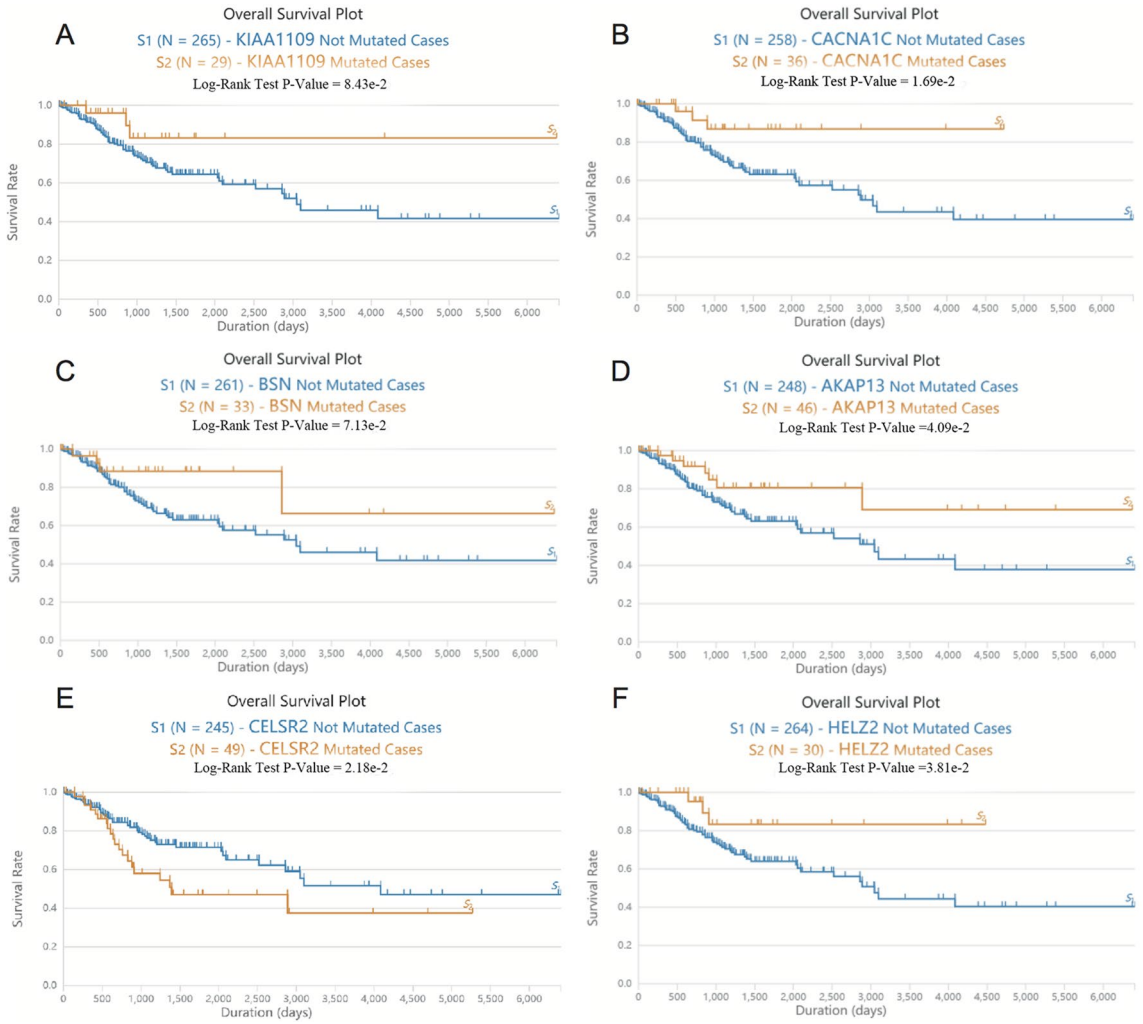


FIGURE 2 | The somatic mutations of *KIAA1109, CACNA1C, BSN, AKAP13, CELSR2*, and *HELZ2* were correlated to overall survival time in patients with EC. **(A–F)** Log-rank test showed that mutations in *KIAA1109* **(A)**, *CACNA1C* **(B)**, *BSN* **(C)**, *AKAP13* **(D)**, *CELSR2* **(E)** and *HELZ2* **(F)** were associated with the overall survival time in EC patients.



FIGURE 3 | Mutation profiles of *KIAA1109*, *CACNA1C*, *BSN*, *AKAP13*, *CELSR2* and *HELZ2* found in EC.

down-regulated in multiple human cancers (Fastje et al., 2009), including brain tumors, kidney cancers and lung cancers, suggested its regulatory roles in cancer progression. BSN encoded a scaffolding protein involved in organizing the presynaptic cytoskeleton. BSN has been demonstrated to have chemo-preventive, antiproliferative, antifungal, and anti-carcinogenic activities. In addition, *BSN* has been reported to induce G1 phase arrest through increase of p21 and p27. In PCa, *BSN* was involved in regulating cell apoptosis in cancer cells (Xu et al., 2016). The dysregulation and mutation of *AKAP13* were found to be associated with the progression of colorectal cancer and breast cancer. Bentin et al. showed *AKAP13* is essential for the phosphorylation of ERαS305 (Toaldo et al., 2015), which leads to tamoxifen resistance in breast cancer. *HELZ2* encoded b a nuclear transcriptional co-activator for peroxisome proliferator activated receptor alpha (Jakobsson et al., 2010). However, its roles in human cancers remained largely unclear. *CELSR2* was found to be dysregulated in breast cancer (Jiang et al., 2018). However, the potential functions of *CELSR2* in EC remained unknown.

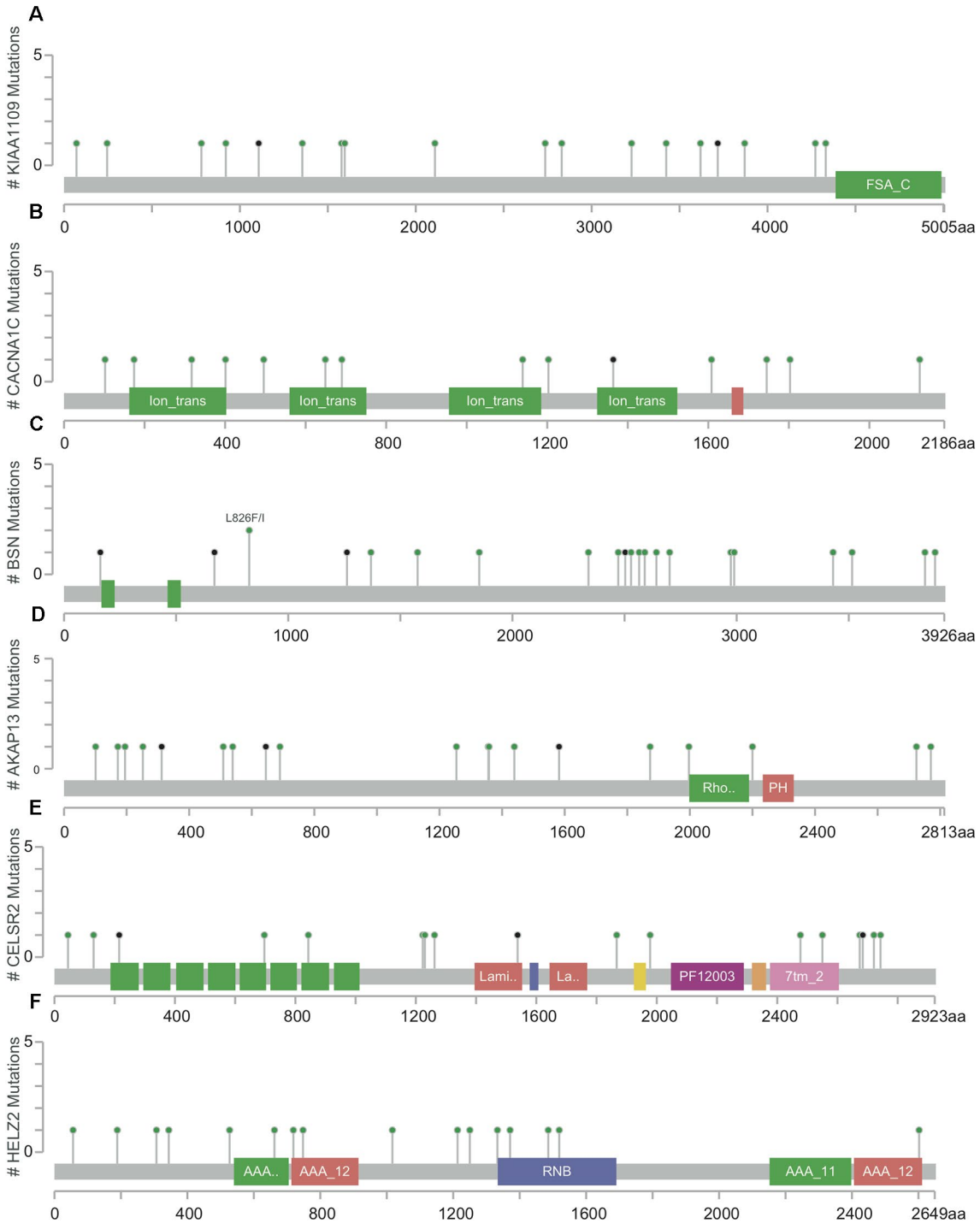


FIGURE 4 | Detailed mutation maps of *KIAA1109* **(A)**, *CACNA1C* **(B)**, *BSN* **(C)**, *AKAP13* **(D)**, *CELSR2* **(E)** and *HELZ2* **(F)** found in patients with EC. Each dot above the protein molecule represents a mutation, which spreads across the entire encoded protein of these genes.

In the present study, we performed co-expression analysis to reveal the potential roles of these mutated genes in EC. The results showed *KIAA1109* was involved in regulating NIK/NF-kappaB signaling. Of note, NF-kappaB signaling had been demonstrated to be a key regulator in cancers. Suppressing of NF-kappaB signaling could inhibit cell growth and invasion in multiple cancers. For example, NF-κB suppresses apoptosis and promotes the proliferation of bladder cancer cells. A recent study

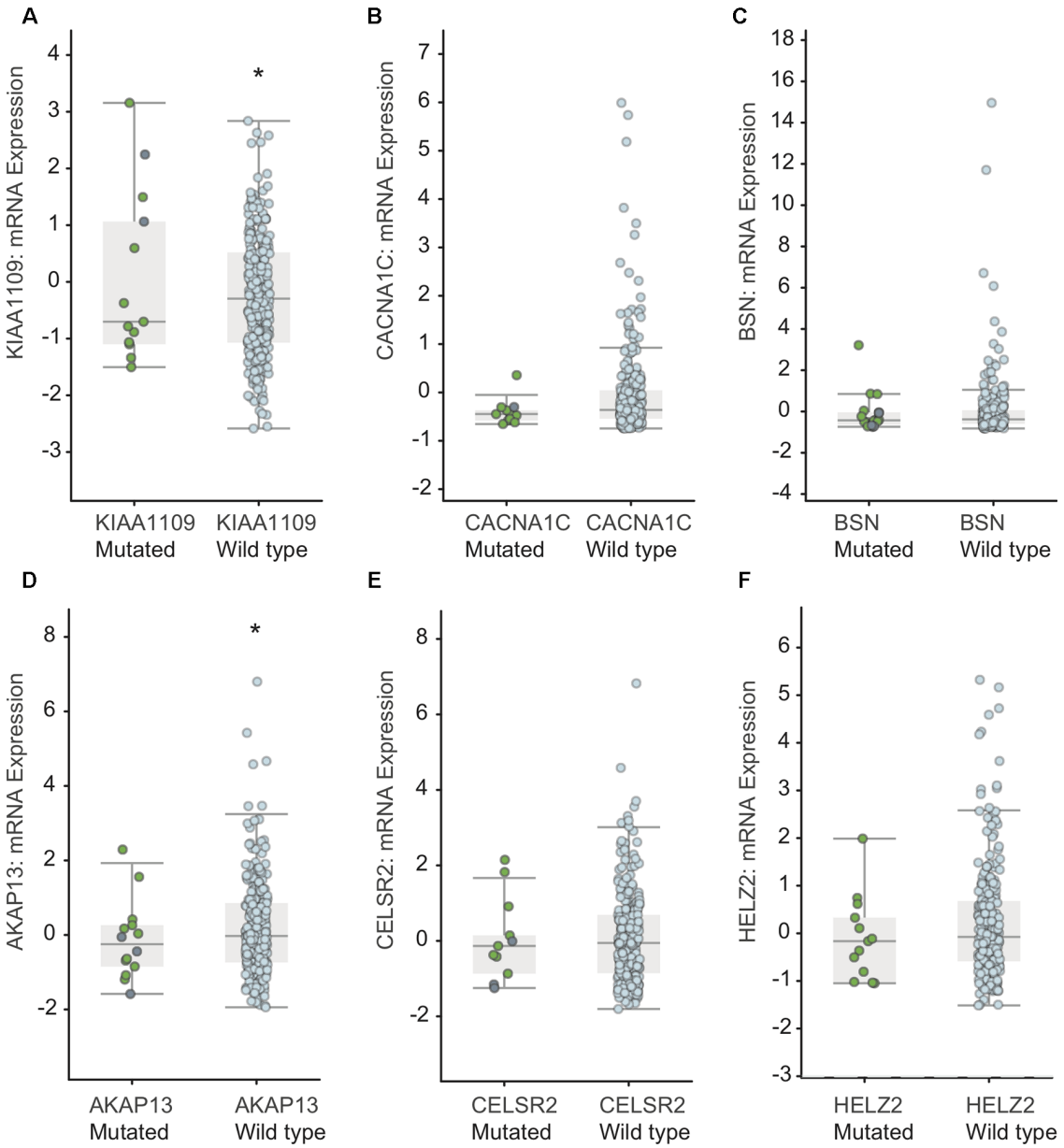


FIGURE 5 | Association between mutations in *KIAA1109* **(A)**, *CACNA1C* **(B)**, *BSN* **(C)**, *AKAP13* **(D)**, *CELSR2* **(E)** and *HELZ2* **(F)** and mRNA levels in EC samples. ** means p value < 0.05 between the two groups.*

showed liposomal curcumin targeting EC through the NF-κB Pathway. Bioinformatics analysis revealed *CACNA1C* played important roles in regulation of EC metastasis and proliferation. *BSN* was found to regulate Wnt signaling pathway. Mounting evidence has confirmed the activation of Wnt/β-catenin signaling was associated with multiple cancers, including EC. *AKAP13* was predicted as a RNA processing regulator. *CELSR2* was involved in regulating cell–cell adhesion, keratinocyte differentiation, spliceosomal snRNP assembly, nuclear import, and protein folding. *HELZ2* was involved in regulating type I interferon signaling pathway, innate immune response, immune response, inflammatory response, and T cell activation. These results suggested these mutated genes played important roles in EC tumorigenesis and progression.

Despite that bioinformatics analyses were conducted to predict the potential functions of these mutated genes in EC, several limitations still existed in this study. First, the mutated sites of these genes should be further validated in EC clinical samples using Sanger sequencing. Second, the molecular function of these key mutated genes in EC remained unclear. Therefore, gain or loss of function assays should be further conducted to investigate their important roles in EC.

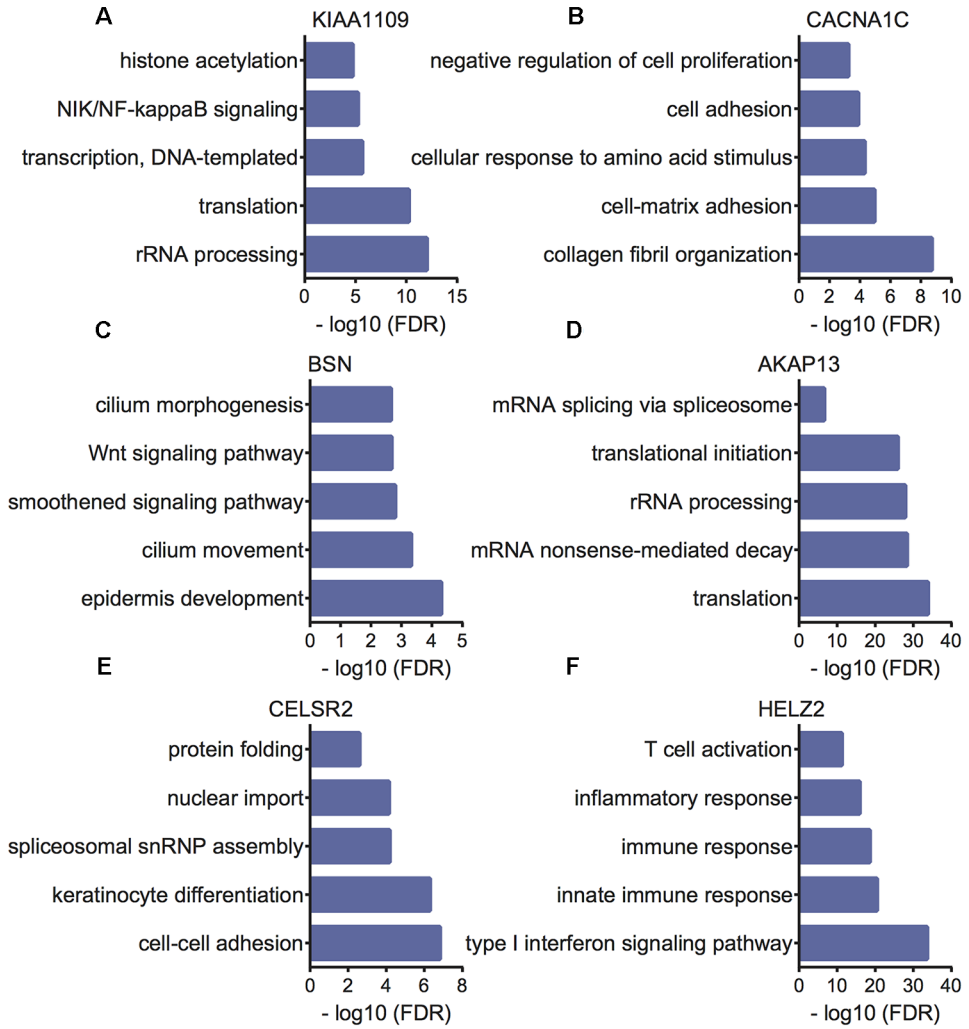


FIGURE 6 | Bioinformatics analysis of *KIAA1109* **(A)**, *CACNA1C* **(B)**, *BSN* **(C)**, *AKAP13* **(D)**, *CELSR2* **(E)** and *HELZ2* **(F)** in EC samples.

In conclusion, we screened mutated genes in EC and found that the mutations in *KIAA1109, CACNA1C, BSN, AKAP13, CELSR2*, and *HELZ2* correlated with the overall survival time in patients with EC. Bioinformatics analysis showed *KIAA1109* was involved in regulating NIK/NF-kappaB signaling, *CACNA1C* was found to regulate cell migration and proliferation, *BSN* was found to regulate Wnt signaling pathway, *CELSR2* was involved in regulating cell-cell adhesion, nuclear import, and protein folding, and *HELZ2* was found to regulate multiple immune related biological processes. The findings provided a novel therapeutic strategy in patients with EC.

DATA AVAILABILITY

All datasets analysed in this study can be found in the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/). We screened the gene mutations in EC. We have downloaded these data from the database and the top 500 mutated genes in EC were listed in **Supplementary Table 1**.

AUTHOR CONTRIBUTIONS

JZ designed experiments; ZQ and YJ analyzed the data. All authors wrote and approved the manuscript.

FUNDING

This work is supported by the Natural Science Foundation of Liaoning Province, China (Grant no. 20170540570).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00909/full#supplementary-material

SUPPLEMENTARY TABLE 1 | The top 500 mutated genes in EC.

REFERENCES

Attarha, S., Mints, M., Andersson, S., and Souchelnytskyi, S., (2011). Endometrial cancer and application of proteomics. *Exp. Oncol.* 33, 174–177.

Charles, A., Pollevick, G. D., Cordeiro, J. M., Oscar, C., Sanguinetti, M. C., Yoshiyasu, A., et al. (2007). Loss-of-function mutations in the cardiac calcium channel underlie a new clinical entity characterized by ST-segment elevation, short QT intervals, and sudden cardiac death. *Circ.* 115, 442–449. doi: 10.1161/CIRCULATIONAHA.106.668392

Collins, A. (2007). The Cancer Genome Atlas (TCGA) pilot project. *Cancer Res.* 67.

Dennis, G., Jr., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, C. H., et al. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 4 (9) R60.

Fastje, C. D., Le, K., Sun, N. N., Wong, S. S., Sheppard, P. R., and Witte, M. L. (2009). Prenatal exposure of mice to tungstate is associated with decreased transcriptome-expression of the putative tumor suppressor gene, DMBT1: implications for childhood leukemia. *Land Contam. Reclam.* 17, 169–178. doi: 10.2462/09670513.931

Fayi, N., Xiaoli, W., Panpan, Z., Hao, Y., Wenhua, Z., Yaling, Z., et al. (2016). Genetic analysis of SNPs in *CACNA1C* and ANK3 gene with schizophrenia: a comprehensive meta-analysis. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 168, 637–648. doi: 10.1002/ajmg.b.32348

Gottwald, L., Pluta, P., Piekarski, J., Spych, M., Hendzel, K., Topczewskatylinska, K., et al. (2010). Long-term survival of endometrioid endometrial cancer patients. *Arch. Med. Sci.* 6, 937–944. doi: 10.5114/aoms.2010.19305

Jakobsson, L., Franco, C. A., Bentley, K., Collins, R. T., Ponsioen, B., Aspalter, I. M., et al. (2010). Endothelial cells dynamically compete for the tip cell position during angiogenic sprouting. *Nat. Cell Biol.* 12, 943–953. doi: 10.1038/ncb2103

Jeske, Y. W., Ali, S., Byron, S. A., Gao, F., Mannel, R. S., Ghebre, R. G., et al. (2017). FGFR2 mutations are associated with poor outcomes in endometrioid endometrial cancer: an NRG oncology/gynecologic oncology group study. *Gynecol. Oncol.* 145, 366–373. doi: 10.1016/j.ygyno.2017.02.031

Jiang, L., Zhang, X., Xiang, C., Geradts, J., Wei, Q., Liang, Y., et al. (2018). Differential cellular localization of CELSR2 and ING4 and correlations with hormone receptor status in breast cancer. *Histol. Histopathol.* 33, 11979. doi: 10.14670/HH-11-979

Jianjiong, G., Bülent Arman, A., Ugur, D., Gideon, D., Benjamin, G., Onur, S. S., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* 6, pl1–pl1. doi: 10.1126/scisignal.2004088

Jing, H., and Jing, S. (2014). MUC16 mutations improve patients' prognosis by enhancing the infiltration and antitumor immunity of cytotoxic T lymphocytes in the endometrial cancer microenvironment. *Oncoimmunology*.

Marshall, A. D., Bailey, C. G., Champ, K., Vellozzi, M., O'Young, P., Metierre, C., et al. (2017). CTCF genetic alterations in endometrial carcinoma are pro-tumorigenic. *Oncogene* 36, 4100–4110. doi: 10.1038/onc.2017.25

Mcconechy, M. K., Jiarui, D., Cheang, M. C. U., Wiegand, K. C., Janine, S., Tone, A. A., et al. (2012). Use of mutation profiles to refine the classification of endometrial carcinomas. *J. Pathol.* 228, 20–30. doi: 10.1002/path.4056

Tao, Q., Zhu, S., Suo, C., Zhang, L., Zheng, Y., and Shi, L. (2017). Somatic mutations in ZFHX4 gene are associated with poor overall survival of Chinese esophageal squamous cell carcinoma patients. *Sci. Rep.* 7, 4951. doi: 10.1038/s41598-017-04221-7

Tindall, E. A., Hoang, H. N., Southey, M. C., English, D. R., Hopper, J. L., Giles, G. G., et al. (2010). The 4q27 locus and prostate cancer risk. *BMC Cancer* 10, 69. doi: 10.1186/1471-2407-10-69

Toaldo, C. B., Alexi, X., Beelen, K., Kok, M., Hauptmann, M., Jansen, M., et al. (2015). Protein Kinase A-induced tamoxifen resistance is mediated by anchoring protein AKAP13. *BMC Cancer* 15, 1–12. doi: 10.1186/s12885-015-1591-4

Xu, L., Wang, Z., He, S. Y., Zhang, S. F., Luo, H. J., Zhou, K., et al. (2016). Bax-interacting factor-1 inhibits cell proliferation and promotes apoptosis in prostate cancer cells. *Oncol. Rep.* 36, 3513. doi: 10.3892/or.2016.5172

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Qiao, Jiang, Wang, Wang, Jiang and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

MINI REVIEW
published: 13 November 2019
doi: 10.3389/fgene.2019.01134

Alteration of the DNA Methylation Signature of Renal Erythropoietin-Producing Cells Governs the Sensitivity to Drugs Targeting the Hypoxia-Response Pathway in Kidney Disease Progression

Koji Sato[1,2], *Naonori Kumagai*[3] *and Norio Suzuki*[1]*

[1] Division of Oxygen Biology, United Centers for Advanced Research and Translational Medicine, Tohoku University Graduate School of Medicine, Sendai, Japan, [2] Division of Nephrology, Endocrinology, and Vascular Medicine, Tohoku University Graduate School of Medicine, Sendai, Japan, [3] Department of Pediatrics, School of Medicine, Fujita Health University, Toyoake, Japan

OPEN ACCESS

Edited by:
Yun Liu,
Fudan University,
China

Reviewed by:
Johannes Schödel,
University of Erlangen Nuremberg,
Germany
Yi Fang,
Fudan University, China

***Correspondence:**
Norio Suzuki
sunorio@med.tohoku.ac.jp

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: 29 July 2019
Accepted: 18 October 2019
Published: 13 November 2019

Citation:
Sato K, Kumagai N and Suzuki N (2019) Alteration of the DNA Methylation Signature of Renal Erythropoietin-Producing Cells Governs the Sensitivity to Drugs Targeting the Hypoxia-Response Pathway in Kidney Disease Progression. Front. Genet. 10:1134. doi: 10.3389/fgene.2019.01134

Chronic kidney disease (CKD) affects more than 10% of the population worldwide and burdens citizens with heavy medical expenses in many countries. Because a vital erythroid growth factor, erythropoietin (EPO), is secreted from renal interstitial fibroblasts [renal EPO-producing (REP) cells], anemia arises as a major complication of CKD. We determined that hypoxia-inducible factor 2α (HIF2α), which is inactivated by HIF-prolyl hydroxylase domain-containing proteins (PHDs) in an oxygen-dependent manner, tightly regulates EPO production in REP cells at the gene transcription level to maintain oxygen homeostasis. HIF2α-mediated disassembly of the nucleosome in the *EPO* gene is also involved in hypoxia-inducible EPO production. In renal anemia patients, anemic and pathological hypoxia is ineffective toward EPO induction due to the inappropriate over-activation of PHDs in REP cells transformed into myofibroblasts (MF-REP cells) due to kidney damage. Accordingly, PHD inhibitory compounds are being developed for the treatment of renal anemia. However, our studies have demonstrated that the promoter regions of the genes encoding EPO and HIF2α are highly methylated in MF-REP cells, and the expression of these genes is epigenetically silenced with CKD progression. This finding notably indicates that the efficacy of PHD inhibitors depends on the CKD stage of each patient. In addition, a strategy for harvesting renal cells, including REP cells from the urine of patients, is proposed to identify plausible biomarkers for CKD and to develop personalized precision medicine against CKD by a non-invasive strategy.

Keywords: chronic kidney disease, DNA methylation, fibrosis, hypoxia, renal anemia, urine exfoliated cells

RENAL ANEMIA

Currently, over 10% of the population worldwide suffers from chronic kidney disease (CKD), which is characterized by kidney dysfunction and/or proteinuria that persists for more than 3 months (Levery et al., 2005). A gradual decline in kidney function results in sclerotic lesions, cardiovascular disease, and mortality (Imai et al., 2009; Hill et al., 2016). While the etiologies of CKD are diverse, ranging from lifestyle-related diseases to autoimmune disorders, CKD progression is commonly accompanied by kidney fibrosis, in which myofibroblasts emerge and proliferate in the renal tubular interstitium (Quaggin and Kapus, 2011). Because kidneys are the major organs producing erythroid growth factor erythropoietin (EPO) in adult mammals (Suzuki, 2015; Hirano and Suzuki, 2019), erythropoiesis is often impaired in CKD patients (Nangaku and Eckardt, 2006). The liver supportively produces EPO under anemic conditions, but hepatic EPO production cannot adequately compensate for renal EPO production in renal anemia patients. In fact, mice lacking renal *EPO* gene expression exhibit severe anemia, although *EPO*-gene expression is induced in their hepatocytes (Yamazaki et al., 2013; Hirano et al., 2017).

Because EPO is required for erythropoiesis, gene-modified mouse lines lacking EPO production exhibit embryonic lethality due to severe anemia (Wu et al., 1995; Yamazaki et al., 2013). Since red blood cells are essential for oxygen delivery to every organ, renal anemia severely decreases the quality of life (QOL) of CKD patients. To maintain oxygen homeostasis, EPO production in the kidney is dramatically enhanced under hypoxic/anemic conditions (Suzuki and Yamamoto, 2016). As CKD progresses, renal EPO production becomes impaired, and renal anemia then develops (Nangaku and Eckardt, 2006; Souma et al., 2015). Intriguingly, recent studies have shown that proper treatment of renal anemia is associated with the prognosis of CKD patients and that the plasma EPO concentration tightly correlates with kidney function and fibrosis (Inomata et al., 1997; Singh et al., 2006; Pfeffer et al., 2009). Thus, plasma EPO is expected to be a plausible biomarker to estimate the CKD grade (Tsubakihara et al., 2015).

For treatment of renal anemia, recombinant human EPO reagents have been used as erythropoiesis-stimulating agents (ESAs) for more than 30 years, and these reagents have dramatically improved the QOL of CKD patients (Jones et al., 2004). However, the invasiveness of subcutaneous ESA injections and the formulation costs of ESAs are problems that need to be solved (Schiller et al., 2008). Additionally, ESAs are frequently ineffective for patients suffering from chronic inflammation because EPO-dependent erythropoiesis is strongly suppressed by high serum concentrations of inflammatory cytokines and hepcidin, which negatively regulates iron usage for hemoglobin (Ganz, 2003; Smrzova et al., 2005; Suzuki et al., 2016; Petrulienė et al., 2017).

RENAL ERYTHROPOIETIN-PRODUCING CELLS

Using genetically modified mouse lines, we and others demonstrated that the ability to produce EPO is present in most fibroblasts that are positive for CD73 and platelet-derived growth factor receptor β (PDGFRβ) in the interstitium spreading from the cortico-medullary boundary to the renal cortex (**Figures 1A**, **B**; Maxwell et al., 1993; Pan et al., 2011; Yamazaki et al., 2013). The cells that produce EPO in response to a hypoxic microenvironment are known as REP (renal EPO-producing) cells (Suzuki et al., 2007; Obara et al., 2008). REP cells are fundamentally quiescent in terms of the cell cycle, and EPO production in the majority of REP cells is absent in healthy mice (Souma et al., 2013; Yamazaki et al., 2013). Under hypoxic/anemic conditions, the percentage of "ON-REP cells," in which EPO production is ongoing, in the total REP cell population is increased. However, only up to 10% of REP cells are ON-REP cells, even under very severe chronic anemia conditions, suggesting that most REP cells are reservoirs (referred to as OFF-REP cells) in preparation for much more severe conditions that require high amounts of EPO (**Figures 1C**, **D**; Yamazaki et al., 2013; Souma et al., 2015). Thus, the total amount of EPO secretion from a kidney is correlated with the ratio of ON-REP cells to total REP cells, rather than the extent of EPO-production levels in each cell (Eckardt et al., 1993; Obara et al., 2008; Suzuki, 2015). Additionally, these data indicate that small numbers of ON-REP cells are sufficient for recovery from anemia because EPO-production levels in each ON-REP cell are very high.

Whereas the origins of myofibroblasts coming into existence in the fibrotic kidneys of CKD patients are controversial and considered various (LeBleu et al., 2013), we and others have demonstrated that resident interstitial fibroblasts, including REP cells in healthy kidneys, are transformed into myofibroblasts under pathological conditions (**Figures 1C**, **D**; Humphreys et al., 2010; Asada et al., 2011; Souma et al., 2013). Importantly, REP cells gain proliferative activity and lose EPO-production ability after transformation (Souma et al., 2013). Thus, REP cells are closely related to the two major pathologies of CKD: renal anemia and fibrosis. Therefore, investigations of REP cells and myofibroblast-transformed REP (MF-REP) cells hold the key to elucidating the molecular pathology of CKD.

Various studies have proposed that the transformation of REP cells into MF-REP cells is promoted by the SMAD and NFκB transcription factors, which are activated by transforming growth factor beta (TGFβ) and tumor necrosis factor alpha, respectively (Wynn and Ramalingam, 2012; Souma et al., 2015). Additionally, DNA methylation in the *EPO*-gene promoter is thought to be involved in the loss of EPO-production ability in MF-REP cells (Chang et al., 2016). To further elucidate the molecular pathology of CKD by characterizing MF-REP cells, we recently established a myofibroblast cell line derived from mouse REP cells, and the cell line was referred to as Replic (REP cell-lineage immortalized and cultivable) cells (Sato et al., 2019). The genomic region of the *EPO*-gene promoter is highly methylated in Replic cells, and cell-autonomous TGFβ signaling supports their myofibroblast properties, which include the expression of genes for α smooth muscle actin, fibronectin, and collagens, among others.

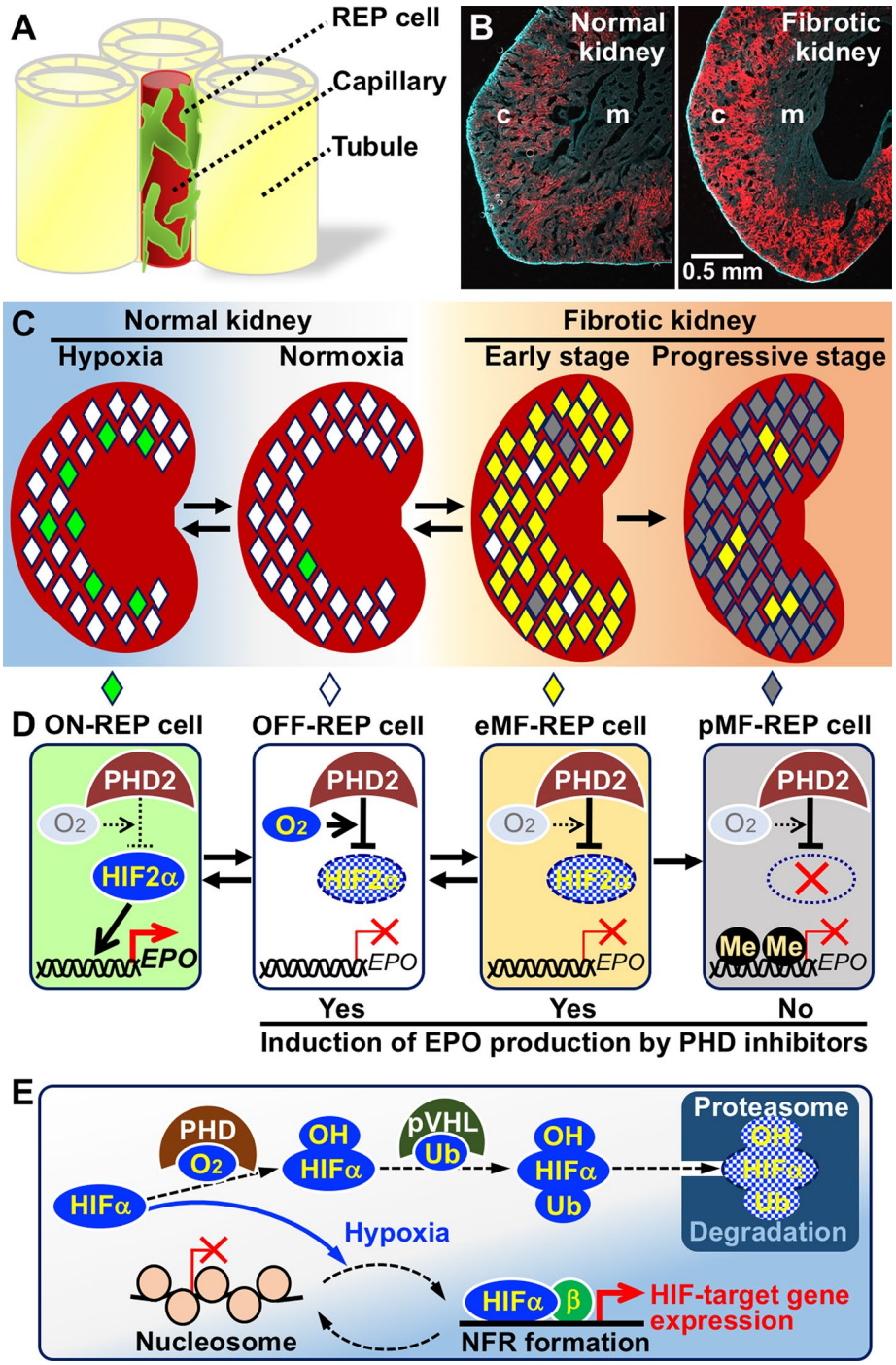


FIGURE 1 | Mechanisms of hypoxia-inducible erythropoietin (EPO) production in renal EPO-producing (REP) cells and failure of EPO production in fibrotic kidney. **(A)** A schema of REP cell localization in the interstitia between renal tubules. REP cells directly associate with capillaries (Souma et al., 2016). **(B)** REP cells (red) distributed to the outer medulla (m) and cortex (c) of a normal healthy kidney (left) are expanded in a fibrotic kidney (right) of a genetically modified mouse line specifically expressing tdTomato fluorescence in REP cells (Yamazaki et al., 2013). **(C)** Distributions of ON-REP (green), OFF-REP (white), early myofibroblast (eMF)-REP (yellow), and progressive MF (pMF)-REP (gray) in normal kidneys and fibrotic kidneys. Note that a small fraction of REP cells produce EPO even under hypoxic conditions (left). **(D)** *EPO*-gene regulation by the PHD2-HIF2α pathway in REP cells and MF-REP cells. In eMF-REP cells (reversibly transformed REP cells), PHD2 over-activation results in inactivation of *EPO*-gene transcription. Therefore, PHD inhibitors may induce EPO production. Because the genes for EPO and HIF2α are epigenetically inactivated due to DNA methylation (Me) in pMF-REP cells (irreversibly transformed REP cells), PHD inhibitors are ineffective. **(E)** Molecular mechanism of hypoxia-inducible transcriptional regulation. HIFα proteins are always synthesized and degraded by the ubiquitin (Ub)-proteasome pathway *via* PHD-mediated hydroxylation (OH) in oxygen-replete cells. In hypoxic cells, PHD is inactivated, and HIFα proteins are stabilized. In some HIF-target gene promoters, HIFα/β complexes mediate the disassembly of nucleosome structures to form nucleosome-free regions under hypoxic conditions.

EPO-GENE REGULATION IN REP CELLS

EPO production in REP cells is strictly regulated at the gene transcription level, and transcription is likely regulated by an ON/OFF mechanism in each cell (Obara et al., 2008). Gene expression data of separately isolated ON- and OFF-REP cells indicated that hypoxia-inducible genes are highly expressed in ON-REP cells compared to OFF-REP cells, suggesting that there is a hypoxic threshold to activate *EPO*-gene expression in REP cells and that the oxygen levels of the microenvironments around ON-REP cells are below the threshold (Yamazaki et al., 2013). The expression levels of almost all hypoxia-inducible genes, including genes related to angiogenesis, glycolysis, and cell survival, are commonly regulated by hypoxia-inducible transcription factors (HIFs) (**Figure 1E**; Wang and Semenza, 1993; Lendahl et al., 2009; Suzuki et al., 2017).

HIFs consist of two subunits, namely, HIFα and HIFβ (also known as ARNT), and they bind to specific DNA sequences ($^{A}_{G}$CGTG) in the regulatory regions of their target genes (Semenza et al., 1991; Lendahl et al., 2009; Haase, 2013). Under normal air conditions (normoxia), specific prolyl residues of HIFα are hydroxylated with HIF-specific prolyl hydroxylase domain proteins (PHDs) by means of intracellular oxygen, and hydroxylated HIFα proteins are degraded by the ubiquitin-proteasome system (**Figure 1E**; Lendahl et al., 2009). In cells with insufficient oxygen for PHD-mediated HIFα hydroxylation, HIFα proteins avoid degradation and activate transcription of their target genes. There are three isoforms encoded by the different genes for the PHD and HIFα proteins, respectively. Among the isoforms, PHD2 and HIF2α primarily control *EPO*-gene expression in a hypoxia-inducible manner in REP cells (**Figure 1D**; Castrop and Kurtz, 2010; Souma et al., 2016). Therefore, dysfunction of the PHD2-HIF2α-*EPO* axis in REP cells is considered the molecular cause of renal anemia. Notably, polycythaemia-related polymorphisms are found in the genes for PHD2 and HIF2α but not in those for the other isoforms, and these polymorphisms are predicted to lead to HIF2α stabilization followed by *EPO*-gene induction without hypoxic stimuli (Bento et al., 2014).

Due to the difficulty of isolating sufficient levels of REP cells for molecular biology analyses, hepatocytes and genetically modified mice have been used for studies on *EPO*-gene regulation. With transgenic mouse strategies, the murine *Epo*-gene regulatory region for REP-cell-specific and hypoxia-inducible expression was determined to be approximately 10 kb upstream from the transcription start site of the *Epo* gene (Hirano et al., 2017). We also discovered that histones located in the *EPO*-gene promoter are always acetylated regardless of hypoxic EPO induction and that histones are dissociated from the nucleosome structure in the *EPO*-gene promoter of hepatocytes under hypoxic conditions through HIF2α activation (Suzuki et al., 2011; Tojo et al., 2015). Nucleosome disassembly results in the formation of a nucleosome-free region (NFR) that has an open chromatin structure for the direct association between transcription factors and promoters and allows the induction of *EPO*-gene transcription (**Figure 1E**; Suzuki et al., 2017; Suzuki et al., 2018a).

STEPWISE MECHANISMS OF *EPO*-GENE SILENCING IN MF-REP CELLS

Since mice lacking PHD2 expression in REP cells are resistant to renal EPO deficiency caused by kidney injury, inappropriate over-activation of PHD2 is considered responsible for *EPO*-gene inactivation in MF-REP cells (**Figure 1D**; Souma et al., 2016). Although the oxygen affinities of PHDs are ordinarily very low compared to those of other oxygen-dependent enzymes, including collagen hydroxylases and epigenetic regulators (see below; Hancock et al., 2017; Chakraborty et al., 2019), unknown mechanisms are speculated to allow PHDs to use oxygen in MF-REP cells even under pathological hypoxic conditions. Indeed, PHD inhibitory compounds are being developed as medicines for renal anemia treatment, and clinical trials of these compounds are showing anticipated effects (**Figure 1D**; Akizawa et al., 2019).

In addition to PHD over-activation, DNA methylation in the *EPO* promoter is involved in *EPO*-gene silencing in MF-REP cells (Chang et al., 2016; Sato et al., 2019). Because hyper-methylation of gene promoter regions blunts gene transcription by tightly compacting the chromatin structure and blocking associations with transcription factor complexes (Jones, 2012; Schübeler, 2015), PHD inhibitors are predicted to be ineffective in cells in which the *EPO* promoter is highly methylated (**Figure 1D**). Consistent with this hypothesis, in Replic cells, neither PHD inhibitors nor HIF2α overexpression activated the *Epo* gene, which is highly methylated (Sato et al., 2019). Thus, the transformation of REP cells into myofibroblasts is divided into at least two consecutive stages: the early MF-REP (eMF-REP) cell stage with over-activation of PHD and the progressive MF-REP (pMF-REP) cell stage with hyper-methylation of the *EPO* promoter. PHD inhibitors are theoretically effective at inducing EPO production in the former cell type but ineffective in the latter cell type, which likely corresponds to Replic cells. Intriguingly, transformation of REP cells is reversible in the early stages of kidney injury (**Figure 1D**; Souma et al., 2013).

We recently discovered that the gene promoter for HIF2α is also highly methylated and that both the mRNA and protein of HIF2α are undetectable in pMF-REP cells, even under hypoxic conditions (Sato et al., 2019). This finding indicates that DNA methylation in specific gene promoters is one of the causes of EPO deficiency in CKD. DNA methylation is mediated by 3 DNA methyltransferases (DNMTs): DNMT1, DNMT3A, and DNMT3B. DNMT1 is essential for the maintenance of DNA methylation patterns beyond mitosis to inherit epigenetic memory (Jeltsch, 2006), while *de novo* DNA methylation is mediated by DNMT3A and DNMT3B (Hsieh, 1999). This transformation enhances the expression of mRNAs for DNMT1 and DNMT3B by TGFβ signaling (Souma et al., 2013), suggesting that these DNMTs are involved in the loss of EPO-production ability in MF-REP cells. In fact, 5-aza-2'-deoxycytidine (5-aza), an inhibitor of DNMT1, restores EPO production in primary-cultured mouse MF-REP cells by reducing DNA methylation in the *Epo*-gene promoter (Chang et al., 2016). In contrast, DNA methylation in the gene promoters for EPO and HIF2α in Replic

cells was resistant to 5-aza treatment, whereas the other genomic regions tested were sensitive (Sato et al., 2019). This discrepancy in 5-aza efficacy between the primary-cultured MF-REP cells and Replic cells, which may represent eMF-REP and pMF-REP cells, respectively, is explained by differences in the activity of *de novo* DNA methylation because expression of *de novo* DNMTs (DNMT3A and DNMT3B) is induced by TGFβ signaling (Cardenas et al., 2014), which is autonomously promoted in Replic cells.

EFFICACY OF PHD INHIBITORS IN EPO-INDUCTION IS RELATED TO THE TRANSFORMATION STAGE OF MYOFIBROBLAST-TRANSFORMED REP CELLS

As an alternative to ESAs, PHD inhibitors are a promising group of next-generation medicines for renal anemia treatment because they are orally administrable small compounds (Martin et al., 2017). The first PHD inhibitor, roxadustat, was launched in 2018 in China, where there are more than 100 million CKD patients (Zhang et al., 2012). However, it is concerned that PHD inhibitors cause unexpected side effects through their widespread activation of HIF-target genes in addition to *EPO*, and these genes include genes that are related to energy metabolism, angiogenesis, and cell survival (Wang and Semenza, 1993). Although obvious adverse events, such as tumor malignancy, have not been observed in clinical trials thus far (Akizawa et al., 2019), further long-term observation is necessary to confirm both the beneficial and unfavorable side effects of PHD inhibitors.

PHDs catalyze oxygenation reactions of the specific prolyl residues of HIFαs to produce hydroxylated HIFαs using oxygen, iron, ascorbate, and α-ketoglutarate. These substrates are also used by a variety of α-ketoglutarate-dependent dioxygenases, including important epigenetic regulators, TET (ten-eleven translocation) family DNA demethylases and KDM (histone lysine demethylase) family histone demethylases (Itoh et al., 2013; Kohli and Zhang, 2013). Since PHDs show the lowest affinity for oxygen among these dioxygenases, PHDs are the first dioxygenases inactivated by hypoxia and can thus sensitively detect hypoxia in cells. On the other hand, the other dioxygenases are less susceptible to hypoxia than PHDs. Notably, very recent studies have shown that some KDMs are as sensitive to hypoxia as PHDs, and further studies are expected to unveil mechanisms involving direct sensing of hypoxia by epigenetic regulators in addition to PHDs (Batie et al., 2019; Chakraborty et al., 2019). Since the current PHD inhibitors commonly block the specific association of α-ketoglutarate with PHDs, other α-ketoglutarate-dependent dioxygenases are unresponsive to these compounds.

In summary, PHD inhibitors are considered to be effective in eMF-REP cells but not in pMF-REP cells with methylation-based silencing of the genes for HIF2α and EPO (**Figure 1D**). Our preliminary experiments using mouse models have shown that EPO production is induced by PHD inhibitors in undamaged or slightly damaged REP Cells of fibrotic kidneys through HIF2α accumulation but not in severely damaged areas. In contrast, PHD inhibitors activate EPO production in almost all the REP cells of healthy kidneys within 6 H after peritoneal injection of the drug (Suzuki et al., 2018b). Clinical trials have demonstrated that PHD inhibitors induce erythropoiesis in nephric patients suffering from any CKD stage and end-stage renal disease, but anephric patients barely respond to PHD inhibitors with regard to EPO induction (Bernhardt et al., 2010). Taken together, these results suggest that EPO produced by a small number of REP cells in the kidney is sufficient to induce erythropoiesis in renal anemia patients. Indeed, as mentioned above, ON-REP cells constitute less than 10% of the REP cells in mouse models of severe chronic anemia (**Figure 1C**; Yamazaki et al., 2013). These observations also suggest that the efficacy of PHD inhibitors differs among renal anemia patients and that the population of eMF-REP and healthy REP Cells in each patient defines their responsiveness to PHD inhibitors.

PERSPECTIVES: NON-INVASIVE STRATEGIES FOR PERSONALIZED PRECISION MEDICINE FOR CHRONIC KIDNEY DISEASE

Here, we summarize the epigenetic and molecular mechanisms of *EPO*-gene silencing in CKD patients and propose the stepwise transformation of REP cells into eMF-REP and pMF-REP cells in injured kidneys (**Figures 1C**, **D**). We also suggest that PHD-inhibitor responsiveness varies in patients and is dependent on the degree of REP cell transformation, which fundamentally correlates with the degree of kidney fibrosis in CKD. Thus, diagnosing the degree of kidney fibrosis is expected to inform us not only about CKD conditions/prognoses but also about PHD-inhibitor responsiveness of CKD patients. Currently, an invasive biopsy is widely adopted for the diagnosis of the complicated pathology of CKD (Mise et al., 2014). However, non-invasive biomarkers for the progression of CKD are being explored. For example, urine concentrations of N-acetyl-β-D-glucosaminidase (Bazzi et al., 2002) can be used. However, the quantitative relationship of the biomarkers to the degree of kidney fibrosis should be investigated in detail.

We propose that urine exfoliated cells can be used for the diagnosis and prediction of CKD. Urine contains several types of kidney cells, including tubular epithelial cells and podocytes, which are living and proliferative in *ex vivo* culture (Dörrenhaus et al., 2000; Kumagai et al., 2000; Vogelmann et al., 2003; Oliveira Arcolino et al., 2015). Therefore, these cells have been utilized as the experimental source of human renal epithelial cells and investigated as biomarkers for the early detection of bladder cancer (Rahmoune et al., 2005; Shimizu et al., 2013). Additionally, urine from CKD patients contains more cultivable exfoliated cells than urine from healthy individuals, which is advantageous for diagnosis (Detrisac et al., 1983). Importantly, our preliminary RT-PCR experiments detected the expression of mRNAs for EPO, HIF2α, and CD73

in cultured cells from the urine of patients with kidney disease, indicating that the exfoliated cell cultures contain REP cells and/or MF-REP cells. In addition, REP cells and MF-REP cells can be purified from the mixtures of exfoliated cells with cell surface expression of CD73 or PDGFRβ using cell sorters (Armulik et al., 2011; Pan et al., 2011).

With small numbers of urine exfoliated cells, high-sensitivity PCR-based techniques are expected to detect HIF2α mRNA expression and *EPO*-gene methylation. NFRs are also detectable with PCR, as we have identified NFRs in hypoxia-inducible gene promoters (Tojo et al., 2015; Suzuki et al., 2018a). Taking advantage of living cells, drug sensitivity may be directly investigated in urine exfoliated cells. Although further studies are needed, exfoliated cells in urine would provide novel diagnostic strategies to distinguish pMF-REP and eMF-REP for the prediction of PHD-inhibitor responsiveness, as well as plausible biomarkers for kidney fibrosis and CKD prognosis.

AUTHOR CONTRIBUTIONS

NS conceived the idea. KS, NK, and NS wrote the manuscript and created the figures.

FUNDING

This work was supported in part by Takeda Life Science Foundation and SENSHIN Medical Research Foundation (for NS). The funders have no role in this study.

ACKNOWLEDGMENTS

We thank Atsuko Konuma, Riona Asai, Taku Nakai, Rio Sasaki, Koichiro Kato and Masayuki Yamamoto (Tohoku University) for technical help and scientific comments. It has been a pleasure conducting this study with Tohoku University Advanced Research Center for Innovations in Next-Generation Medicine (INGEM).

REFERENCES

Akizawa, T., Nangaku, M., Yamaguchi, T., Arai, M., Koretomo, R., and Matsui, A. (2019). A placebo-controlled, randomized trial of enarodustat in patients with chronic kidney disease followed by long-term trial. *Am. J. Nephrol.* 49, 165–174. doi: 10.1159/000496929

Armulik, A., Genové, G., and Betsholtz, C. (2011). Pericytes: developmental, physiological, and pathological perspectives, problems, and promises. *Dev. Cell.* 21, 193–215. doi: 10.1016/j.devcel.2011.07.001

Asada, N., Takase, M., Nakamura, J., Oguchi, A., Asada, M., and Suzuki, N. (2011). Dysfunction of fibroblasts of external origin underlies renal fibrosis and renal anemia in mice. *J. Clin. Invest.* 121, 3981–3990. doi: 10.1172/JCI57301

Batie, M., Frost, J., Frost, M., Wilson, J. W., Schofield, P., and Rocha, S. (2019). Hypoxia induces rapid changes to histone methylation and reprograms chromatin. *Science* 363, 1222–1226. doi: 10.1126/science.aau5870

Bazzi, C., Petrini, C., Rizza, V., Arrigo, G., Napodano, P., and Paparella, M. (2002). Urinary N-acetyl-beta-glucosaminidase excretion is a marker of tubular cell dysfunction and a predictor of outcome in primary glomerulonephritis. *Nephrol. Dial Transplant* 17, 1890–1896. doi: 10.1093/ndt/17.11.1890

Bento, C., Percy, M. J., Gardie, B., Maia, T. M., van Wijk, R., and Perrotta, S. (2014). Genetic basis of congenital erythrocytosis: mutation update and online databases. *Hum. Mutat.* 35, 15–26. doi: 10.1002/humu.22448

Bernhardt, W. M., Wiesener, M. S., Scigalla, P., Chou, J., Schmieder, R. E., and Günzler, V. (2010). Inhibition of prolyl hydroxylases increases erythropoietin production in ESRD. *J. Am. Soc. Nephrol.* 21, 2151–2156. doi: 10.1681/ASN.2010010116

Cardenas, H., Vieth, E., Lee, J., Segar, M., Liu, Y., and Nephew, K. P. (2014). TGF-β induces global changes in DNA methylation during the epithelial-to-mesenchymal transition in ovarian cancer cells. *Epigenetics* 9, 1461–1472. doi: 10.4161/15592294.2014.971608

Castrop, H., and Kurtz, A. (2010). Functional evidence confirmed by histological localization: overlapping expression of erythropoietin and HIF-2alpha in interstitial fibroblasts of the renal cortex. *Kidney Int.* 77, 269–271. doi: 10.1038/ki.2009.470

Chakraborty, A. A., Laukka, T., Myllykoski, M., Ringel, A. E., Booker, M. A., and Tolstorukov, M. Y. (2019). Histone demethylase KDM6A directly senses oxygen to control chromatin and cell fate. *Science* 363, 1217–1222. doi: 10.1126/science.aaw1026

Chang, Y. T., Yang, C. C., Pan, S. Y., Chou, Y. H., Chang, F. C., and Lai, C. F. (2016). DNA methyltransferase inhibition restores erythropoietin production in fibrotic murine kidneys. *J. Clin. Invest.* 126, 721–731. doi: 10.1172/JCI82819

Detrisac, C. J., Mayfield, R. K., Colwell, J. A., Garvin, A. J., and Sens, D. A. (1983). In vitro culture of cells exfoliated in the urine by patients with diabetes mellitus. *J. Clin. Invest.* 71, 170–173. doi: 10.1172/JCI110747

Dörrenhaus, A., Müller, J. I., Golka, K., Jedrusik, P., Schulze, H., and Föllmann, W. (2000). Cultures of exfoliated epithelial cells from different locations of the human urinary tract and the renal tubular system. *Arch. Toxicol.* 74, 618–626. doi: 10.1007/s002040000173

Eckardt, K. U., Koury, S. T., Tan, C. C., Schuster, S. J., Kaissling, B., and Ratcliffe, P. J. (1993). Distribution of erythropoietin producing cells in rat kidneys during hypoxic hypoxia. *Kidney Int.* 43, 815–823. doi: 10.1038/ki.1993.115

Ganz, T. (2003). Hepcidin, a key regulator of iron metabolism and mediator of anemia of inflammation. *Blood* 102, 783–788. doi: 10.1182/blood-2003-03-0672

Haase, V. H. (2013). Regulation of erythropoiesis by hypoxia-inducible factors. *Blood Rev.* 27, 41–53. doi: 10.1016/j.blre.2012.12.003

Hancock, R. L., Masson, N., Dunne, K., Flashman, E., and Kawamura, A. (2017). The activity of JmjC histone lysine demethylase KDM4A is highly sensitive to oxygen concentrations. *ACS Chem. Biol.* 12, 1011–1019. doi: 10.1021/acschembio.6b00958

Hill, N. R., Fatoba, S. T., Oke, J. L., Hirst, J. A., O'Callaghan, C. A., and Lasserson, D. S. (2016). Global prevalence of chronic kidney disease - A systematic review and meta-analysis. *PloS One* 11, e0158765. doi: 10.1371/journal.pone.0158765

Hirano, I., and Suzuki, N. (2019). The neural crest as the first production site of the erythroid growth factor erythropoietin. *Front. Cell Dev. Biol.* 7, 105. doi: 10.3389/fcell.2019.00105

Hirano, I., Suzuki, N., Yamazaki, S., Sekine, H., Minegishi, N., and Shimizu, R. (2017). Renal anemia model mouse established by transgenic rescue with an erythropoietin gene lacking kidney-specific regulatory elements. *Mol. Cell Biol.* 37, e00451–e00416. doi: 10.1128/MCB.00451-16

Hsieh, C. L. (1999). In vivo activity of murine de novo methyltransferases, Dnmt3a and Dnmt3b. *Mol. Cell Biol.* 19, 8211–8218. doi: 10.1128/MCB.19.12.8211

Humphreys, B. D., Lin, S. L., Kobayashi, A., Hudson, T. E., Nowlin, B. T., and Bonventre, J. V. (2010). Fate tracing reveals the pericyte and not epithelial origin of myofibroblasts in kidney fibrosis. *Am. J. Pathol.* 176, 85–97. doi: 10.2353/ajpath.2010.090517

Imai, E., Horio, M., Watanabe, T., Iseki, K., Yamagata, K., and Hara, S. (2009). Prevalence of chronic kidney disease in the Japanese general population. *Clin. Exp. Nephrol.* 13, 621–630. doi: 10.1007/s10157-009-0199-x

Inomata, S., Itoh, M., Imai, H., and Sato, T. (1997). Serum levels of erythropoietin as a novel marker reflecting the severity of diabetic nephropathy. *Nephron* 75, 426–430. doi: 10.1159/000189580

Itoh, Y., Suzuki, T., and Miyata, N. (2013). Small-molecular modulators of cancer-associated epigenetic mechanisms. *Mol. Biosyst.* 9, 873–896. doi: 10.1039/c3mb25410k

Jeltsch, A. (2006). Molecular enzymology of mammalian DNA methyltransferases. *Curr. Top. Microbiol. Immunol.* 301, 203–225. doi: 10.1007/3-540-31390-7_7

Jones, M., Ibels, L., Schenkel, B., and Zagari, M. (2004). Impact of epoetin alfa on clinical end points in patients with chronic renal failure: a meta-analysis. *Kidney Int.* 65, 757–767. doi: 10.1111/j.1523-1755.2004.00450.x

Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* 13, 484–492. doi: 10.1038/nrg3230

Kohli, R. M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* 502, 472–479. doi: 10.1038/nature12750

Kumagai, N., Inoue, C. N., Kondo, Y., and Iinuma, K. (2000). Mitogenic action of lysophosphatidic acid in proximal tubular epithelial cells obtained from voided human urine. *Clin. Sci. (Lond)* 99, 561–567. doi: 10.1042/cs0990561

LeBleu, V. S., Taduri, G., O'Connell, J., Teng, Y., Cooke, V. G., and Woda, C. (2013). Origin and function of myofibroblasts in kidney fibrosis. *Nat. Med.* 19, 1047–1053. doi: 10.1038/nm.3218

Lendahl, U., Lee, K. L., Yang, H., and Poellinger, L. (2009). Generating specificity and diversity in the transcriptional response to hypoxia. *Nat. Rev. Genet.* 10, 821–832. doi: 10.1038/nrg2665

Levery, A. S., Eckardt, K. U., Tsukamoto, Y., Levin, A., Coresh, J., and Rossert, J. (2005). Definition and classification of chronic kidney disease: a position statement from kidney disease: improving global outcomes (KDIGO). *Kidney Int.* 67, 2089–2100. doi: 10.1111/j.1523-1755.2005.00365.x

Martin, E. R., Smith, M. T., Maroni, B. J., Zuraw, Q. C., and deGoma, E. M. (2017). Clinical trial of vadadustat in patients with anemia secondary to stage 3 or 4 chronic kidney disease. *Am. J. Nephrol.* 45, 380–388. doi: 10.1159/000464476

Maxwell, P. H., Osmond, M. K., Pugh, C. W., Heryet, A., Nicholls, L. G., and Tan, C. C. (1993). Identification of the renal erythropoietin-producing cells using transgenic mice. *Kidney Int.* 44, 1149–1162. doi: 10.1038/ki.1993.362

Mise, K., Hoshino, J., Ubara, Y., Sumida, K., Hiramatsu, R., and Hasegawa, E. (2014). Renal prognosis a long time after renal biopsy on patients with diabetic nephropathy. *Nephrol. Dial Transplant* 29, 109–118. doi: 10.1093/ndt/gft349

Nangaku, M., and Eckardt, K. U. (2006). Pathogenesis of renal anemia. *Semin. Nephrol.* 26, 261–268. doi: 10.1016/j.semnephrol.2006.06.001

Obara, N., Suzuki, N., Kim, K., Nagasawa, T., Imagawa, S., and Yamamoto, M. (2008). Repression via the GATA box is essential for tissue-specific erythropoietin gene expression. *Blood* 111, 5223–5232. doi: 10.1182/blood-2007-10-115857

Oliveira Arcolino, F., Tort Piella, A., Papadimitriou, E., Bussolati, B., Antonie, D. J., and Murray, P. (2015). Human urine as a noninvasive sourse of kidney cells. *Stem Cell Int.* 2015, 362562. doi: 10.1155/2015/362562

Pan, X., Suzuki, N., Hirano, I., Yamazaki, S., Minegishi, N., and Yamamoto, M. (2011). Isolation and characterization of renal erythropoietin-producing cells from genetically produced anemia mice. *PloS One* 6, e25839. doi: 10.1371/journal.pone.0025839

Petrulienė, K., Žiginskienė, E., Kuzminskis, V., Nedzelskienė, I., and Bumblytė, I. A. (2017). Hepcidin serum levels and resistance to recombinant human erythropoietin therapy in hemodialysis patients. *Medicina (Kaunas).* 53, 90–100. doi: 10.1016/j.medici.2017.03.001

Pfeffer, M. A., Burdmann, E. A., Chen, C. Y., Cooper, M. E., de Zeeuw, D., and Eckardt, K. U. (2009). A trial of darbepoetin alfa in type 2 diabetes and chronic kidney disease. *N Engl. J. Med.* 361, 2019–2032. doi: 10.1056/NEJMoa0907845

Quaggin, S. E., and Kapus, A. (2011). Scar wars: mapping the fate of epithelial-mesenchymal-myofibroblast transition. *Kidney Int.* 80, 41–50. doi: 10.1038/ki.2011.77

Rahmoune, H., Thompson, P. W., Ward, J. M., Smith, C. D., Hong, G., and Brown, J. (2005). Glucose transporters in human renal proximal tubular cells isolated from the urine of patients with non-insulin-dependent diabetes. *Diabetes* 54, 3427–3434. doi: 10.2337/diabetes.54.12.3427

Sato, K., Hirano, I., Sekine, H., Miyauchi, K., Nakai, T., and Kato, K. (2019). An immortalized cell line derived from renal erythropoietin-producing (REP) cells demonstrates their potential to transform into myofibroblasts. *Sci. Rep.* 9, 11254. doi: 10.1038/s41598-019-47766-5

Schiller, B., Doss, S., DE Cock, E., Del Aguila, M. A., and Nissenson, A. R. (2008). Costs of managing anemia with erythropoiesis-stimulating agents during hemodialysis: a time and motion study. *Hemodial Int.* 12, 441–449. doi: 10.1111/j.1542-4758.2008.00308.x

Schübeler, D. (2015). Function and information content of DNA methylation. *Nature* 517, 321–326. doi: 10.1038/nature14192

Semenza, G. L., Koury, S. T., Nejfelt, M. K., Gearhart, J. D., and Antonarakis, S. E. (1991). Cell-type-specific and hypoxia-inducible expression of the human erythropoietin gene in transgenic mice. *Proc. Natl. Acad. Sci. U. S. A.* 88, 8725–8729. doi: 10.1073/pnas.88.19.8725

Shimizu, T., Suzuki, H., Nojima, M., Kitamura, H., Yamamoto, E., and Maruyama, R. (2013). Methylation of a panel of microRNA genes is a novel biomarker for detection of bladder cancer. *Eur. Urol.* 63, 1091–1100. doi: 10.1016/j.eururo.2012.11.030

Singh, A. K., Szczech, L., Tang, K. L., Barnhart, H., Sapp, S., and Wolfson, M. (2006). Correction of anemia with epoetin alfa in chronic kidney disease. *N Engl. J. Med.* 355, 2085–2098. doi: 10.1056/NEJMoa065485

Smrzova, J., Balla, J., and Bárány, P. (2005). Inflammation and resistance to erythropoiesis-stimulating agents–what do we know and what needs to be clarified?. *Nephrol. Dial Transplant* 20 Suppl 8, viii2–viii7. doi: 10.1093/ndt/gfh1109

Souma, T., Nezu, M., Nakano, D., Yamazaki, S., Hirano, I., and Sekine, H. (2016). Erythropoietin synthesis in renal myofibroblasts is restored by activation of hypoxia signaling. *J. Am. Soc. Nephrol.* 27, 428–438. doi: 10.1681/ASN.2014121184

Souma, T., Suzuki, N., and Yamamoto, M. (2015). Renal erythropoietin-producing cells in health and disease. *Front. Physiol.* 6, 167. doi: 10.3389/fphys.2015.00167

Souma, T., Yamazaki, S., Moriguchi, T., Suzuki, N., Hirano, I., and Pan, X. (2013). Plasticity of renal erythropoietin-producing cells governs fibrosis. *J. Am. Soc. Nephrol.* 24, 1599–1616. doi: 10.1681/ASN.2013010030

Suzuki, N. (2015). Erythropoietin gene expression: developmental-stage specificity, cell-type specificity, and hypoxia inducibility. *Tohoku J. Exp. Med.* 235, 233–240. doi: 10.1620/tjem.235.233

Suzuki, N., Gradin, K., Poellinger, L., and Yamamoto, M. (2017). Regulation of hypoxia-inducible gene expression after HIF activation. *Exp. Cell Res.* 356, 182–186. doi: 10.1016/j.yexcr.2017.03.013

Suzuki, N., Obara, N., Pan, X., Watanabe, M., Jishage, K., and Minegishi, N. (2011). Specific contribution of the erythropoietin gene 3' enhancer to hepatic erythropoiesis after late embryonic stages. *Mol. Cell Biol.* 31, 3896–3905. doi: 10.1128/MCB.05463-11

Suzuki, N., Obara, N., and Yamamoto., M. (2007). Use of gene-manipulated mice in the study of erythropoietin gene expression. *Methods Enzymol.* 435, 157–177. doi: 10.1016/S0076-6879(07)35009-X

Suzuki, N., Sasaki, Y., Kato, K., Yamazaki, S., Kurasawa, M., Yorozu, K., et al. (2016). Efficacy estimation of erythropoiesis-stimulating agents using erythropoietin-deficient anemic mice. *Haematologica* 101, e356–e360. doi: 10.3324/haematol.2015.140814

Suzuki, N., Matsuo-Tezuka, Y., Sasaki, Y., Sato, K., Miyauchi, K., and Kato, K. (2018b). Iron attenuates erythropoietin production by decreasing hypoxia-inducible transcription factor 2α concentrations in renal interstitial fibroblasts. *Kidney Int.* 94, 900–911. doi: 10.1016/j.kint.2018.06.028

Suzuki, N., Vojnovic, N., Lee, K. L., Yang, H., Gradin, K., and Poellinger, L. (2018a). HIF-dependent and reversible nucleosome disassembly in hypoxia-inducible gene promoters. *Exp. Cell Res.* 366, 181–191. doi: 10.1016/j.yexcr.2018.03.020

Suzuki, N., and Yamamoto, M. (2016). Roles of renal erythropoietin-producing (REP) cells in the maintenance of systemic oxygen homeostasis. *Pflugers Arch.* 468, 3–12. doi: 10.1007/s00424-015-1740-2

Tojo, Y., Sekine, H., Hirano, I., Pan, X., Souma, T., and Tsujita, T. (2015). Hypoxia signaling cascade for erythropoietin production in hepatocytes. *Mol. Cell Biol.* 35, 2658–2572. doi: 10.1128/MCB.00161-15

Tsubakihara, Y., Akizawa, T., Iwasaki, M., and Shimazaki, R. (2015). High hemoglobin levels maintained by an erythropoiesis-stimulating agent improve renal survival in patients with severe renal impairment. *Ther. Apher Dial* 19, 457–465. doi: 10.1111/1744-9987.12308

Vogelmann, S. U., Nelson, W. J., Myers, B. D., and Lemley, K. V. (2003). Urinary excretion of viable podocytes in health and renal disease. *Am. J. Physiol. Renal Physiol.* 285, F40–F48. doi: 10.1152/ajprenal.00404.2002

Wang, G. L., and Semenza, G. L. (1993). General involvement of hypoxia-inducible factor 1 in transcriptional response to hypoxia. *Proc. Natl. Acad. Sci. U. S. A.* 90, 4304–4308. doi: 10.1073/pnas.90.9.4304

Wu, H., Liu, X., Jaenisch, R., and Lodish, H. F. (1995). Generation of committed erythroid BFU-E and CFU-E progenitors does not require erythropoietin or the erythropoietin receptor. *Cell.* 83, 59–67. doi: 10.1016/0092-8674(95)90234-1

Wynn, T. A., and Ramalingam, T. R. (2012). Mechanisms of fibrosis: therapeutic translation for fibrotic disease. *Nat. Med.* 18, 1028–1040. doi: 10.1038/nm.2807

Yamazaki, S., Souma, T., Hirano, I., Pan, X., Minegishi, N., and Suzuki, N. (2013). A mouse model of adult-onset anaemia due to erythropoietin deficiency. *Nat. Commun.* 4, 1950. doi: 10.1038/ncomms2950

Zhang, L., Wang, F., Wang, L., Wang, W., Liu, B., and Liu, J. (2012). Prevalence of chronic kidney disease in China: a cross-sectional survey. *Lancet* 379, 815–822. doi: 10.1016/S0140-6736(12)60033-6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Sato, Kumagai and Suzuki. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

REVIEW
published: 14 November 2019
doi: 10.3389/fgene.2019.01150

DNA Methylation Cancer Biomarkers: Translation to the Clinic

*Warwick J. Locke[1,2], Dominic Guanzon[1,2], Chenkai Ma[1], Yi Jin Liew[1,2], Konsta R. Duesing[1], Kim Y.C. Fung[1,2] and Jason P. Ross[1,2]**

[1] Molecular Diagnostics Solutions, CSIRO Health and Biosecurity, North Ryde, NSW, Australia, [2] Probing Biosystems Future Science Platform, CSIRO Health and Biosecurity, Canberra, ACT, Australia

Carcinogenesis is accompanied by widespread DNA methylation changes within the cell. These changes are characterized by a globally hypomethylated genome with focal hypermethylation of numerous 5'-cytosine-phosphate-guanine-3' (CpG) islands, often spanning gene promoters and first exons. Many of these epigenetic changes occur early in tumorigenesis and are highly pervasive across a tumor type. This allows DNA methylation cancer biomarkers to be suitable for early detection and also to have utility across a range of areas relevant to cancer detection and treatment. Such tests are also simple in construction, as only one or a few loci need to be targeted for good test coverage. These properties make cancer-associated DNA methylation changes very attractive for development of cancer biomarker tests with substantive clinical utility. Across the patient journey from initial detection, to treatment and then monitoring, there are several points where DNA methylation assays can inform clinical practice. Assays on surgically removed tumor tissue are useful to determine indicators of treatment resistance, prognostication of outcome, or to molecularly characterize, classify, and determine the tissue of origin of a tumor. Cancer-associated DNA methylation changes can also be detected with accuracy in the cell-free DNA present in blood, stool, urine, and other biosamples. Such tests hold great promise for the development of simple, economical, and highly specific cancer detection tests suitable for population-wide screening, with several successfully translated examples already. The ability of circulating tumor DNA liquid biopsy assays to monitor cancer *in situ* also allows for the ability to monitor response to therapy, to detect minimal residual disease and as an early biomarker for cancer recurrence. This review will summarize existing DNA methylation cancer biomarkers used in clinical practice across the application domains above, discuss what makes a suitable DNA methylation cancer biomarker, and identify barriers to translation. We discuss technical factors such as the analytical performance and product-market fit, factors that contribute to successful downstream investment, including geography, and how this impacts intellectual property, regulatory hurdles, and the future of the marketplace and healthcare system.

Keywords: DNA methylation, diagnostic, translation, cancer, epigenetics, liquid biopsy

OPEN ACCESS

Edited by:
Jiucun Wang,
Fudan University,
China

Reviewed by:
Carmen Jeronimo,
Portuguese Oncology Institute,
Portugal
Jorg Tost,
Commissariat à l'Energie Atomique et aux Energies Alternatives,
France

***Correspondence:**
Jason P. Ross
jason.ross@csiro.au

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *11 June 2019*
Accepted: *22 October 2019*
Published: *14 November 2019*

Citation:
Locke WJ, Guanzon D, Ma C, Liew YJ, Duesing KR, Fung KYC and Ross JP (2019) DNA Methylation Cancer Biomarkers: Translation to the Clinic. Front. Genet. 10:1150. doi: 10.3389/fgene.2019.01150

INTRODUCTION

Cancer is defined by extensive genetic changes and associated dysregulation in gene function and activity (Nakagawa and Fujita, 2018). However, cancer is not an exclusively genetic disease and its progression is dependent on a host of additional biological processes such as immune activity, the tissue microenvironment, and epigenetics (Hanahan and Weinberg, 2011). Epigenetics is a second layer of information encoded onto the genome that guides genomic function and activity. Epigenetics acts through two mechanisms: (1) modifications to chromosomal proteins that alter the 3D conformation of the genome and/or protein-DNA interactions and (2) chemical modification of the DNA strand itself (Kondo, 2009). Change in the 3D structure of DNA is enacted *via* post-translational modifications of the histone proteins at the center of the simplest DNA structure, the nucleosome. Histone modifications can lead to either tightly packed and inactive conformations or open and accessible DNA (termed heterochromatin and euchromatin respectively). The best characterized chemical modification of DNA is the methylation of cytosine to 5-methylcytosine (5mC) that occurs almost exclusively in the context of a cytosine base linked by the DNA phosphate-backbone to guanosine, termed a CpG site. DNA methylation is considered a "soft" and potentially reversible change to the genome that can define or adapt to tumor biology and is functionally equivalent to genetic changes like mutation or deletion (Kulis and Esteller, 2010).

Epigenetic changes are considered to be among the earliest and most comprehensive genomic aberrations occurring during carcinogenesis (Alvarez et al., 2011) and reviewed in (Feinberg et al., 2006). These changes can be broadly characterized as focal hypermethylation and global hypomethylation (Ross et al., 2010). Each mechanism has their own role to play in defining carcinogenesis. Hypomethylation occurs predominantly at repetitive regions and has been demonstrated to be a carcinogenic process in its own right (Gaudet et al., 2003). Hypomethylation also promotes genomic instability, causing missegregation of chromosomes during cell division (Prada et al., 2012) and the unwanted activation of transposable elements within the genome, leading to further genetic damage (Daskalos et al., 2009). Hypermethylation can drive the silencing of key tumor suppressors (Belinsky et al., 1998) or regulatory regions within the genome leading to dysregulation of cell growth or altered response to cancer therapies (Stone et al., 2015). Such epigenetic mechanisms can synergize with known driver mutations to facilitate cancer development or evolution (Tao et al., 2019). Despite the varied and complex nature of changes to the epigenetic landscape, many cancers exhibit a high degree of concordance across tissues, or within the tissue of origin (Zhang and Huang, 2017; Yang et al., 2017b; Hoadley et al., 2018). The robust and common nature of DNA methylation aberrations in cancer and the stability of cell-free DNA in body fluids are attractive properties for diagnostic development. The widespread nature of epigenetic change across the genome can also facilitate increases in sensitivity and specificity by utilizing multiple target loci in a single assay. When combined with the informative nature of these changes regarding cancer biology, DNA methylation-based biomarkers have great potential to transform the treatment and observation of cancer and other diseases.

The value of epigenetic changes as candidate biomarkers is reflected in the scientific literature with thousands of studies published to date that associate DNA methylation with clinical parameters. However, there is a paucity of markers that have been successfully translated into clinical practice (**Figure 1**). Historically, this has in part been due to limitations of technology to assess epigenetic information at a large scale or in a cost-effective manner. Recent improvements in DNA sequencing and other molecular technologies have helped overcome these initial barriers. However, translation is still a slow and costly process. In this review, we will discuss the current state of the DNA methylation biomarker landscape, the current barriers to translation (be they scientific or regulatory), and what the future may look like for this emerging field of diagnostics.

DESIGNING AN EFFECTIVE ASSAY

Clinical Utility

Traditional diagnostic approaches based on clinical pathology utilize patient biopsied cancerous tissue. Histological analysis of tumor specimens has long been the gold standard for tumor subtyping and diagnosis. Modern epigenetic methods may also make use of such samples, allowing for novel molecular diagnostics to be run in parallel to traditional techniques. DNA methylation analysis does not require any special handling of tumor specimens and can also be applied with similar efficiency to fresh frozen and formalin fixed paraffin embedded tissue. Indeed, early market offerings in the DNA methylation oncology diagnostic space were based upon detecting hypermethylated DNA using fresh tumor biopsies or fixed tissue blocks in glioblastoma, prostate, and colorectal cancer (CRC) (e.g. *MGMT*, *GSTP1*, and *MLH1* based assays) (Esteller et al., 1998; Herman et al., 1998; Esteller et al., 2000).

DNA methylation analysis is not limited to tissue specimens and can be readily extended to almost any bodily fluid (typically termed a "liquid biopsy"). Various bodily fluids contain a host of informative molecules linked to tumorigenesis, growth, immune/cancer interactions, and cell death, circulating tumor cells (CTC) and microvesicles such as exosomes (Wang et al., 2017). These molecules are easily assayed using non- or minimally-invasive techniques and are of extremely high value where tumor tissue, from surgery or biopsy, is not available. Circulating tumor DNA (ctDNA), which is the cancer-originating component of cell free DNA (cfDNA), can provide a window into a tumors mutational and epigenetic profile and has a range of benefits over a traditional tissue biopsy approach (Gai and Sun, 2019).

Simple tissue biopsies only sample a subpopulation of all cell types and with intra-tumoral heterogeneity/clonality could provide a misleading image of the true cellular makeup of the tumor. Recent studies indicate that ctDNA may better capture this natural variation by facilitating sampling of a broader proportion of tumor cells (Dagogo-Jack and Shaw, 2018). This is due to unbiased nature of ctDNA, in that all cell types are likely to make some contribution to the total DNA population. While

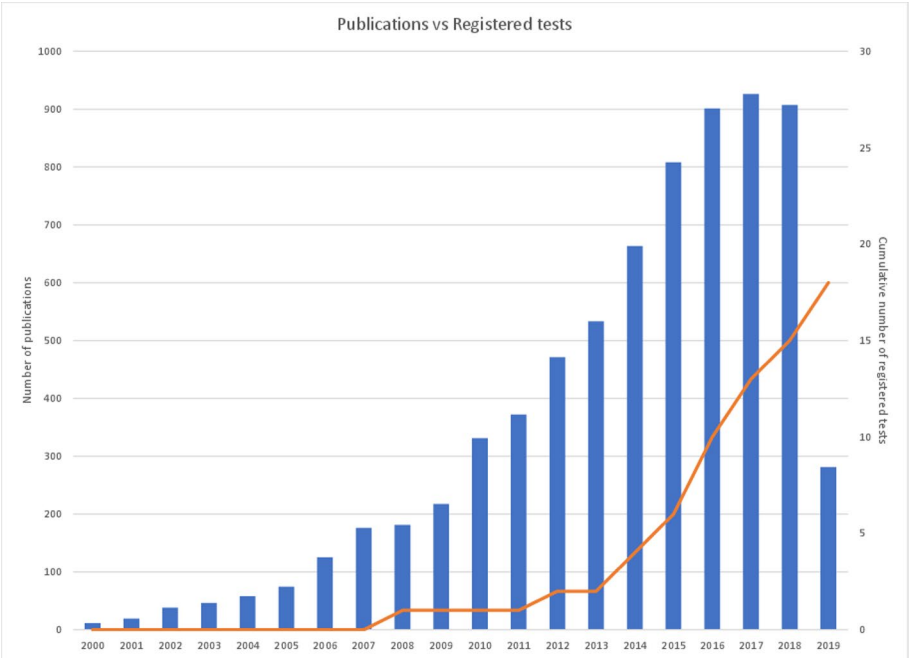


FIGURE 1 | Cancer epigenetic biomarker publications per annum versus cumulative registered DNA-methylated based IVDs. The figure demonstrates the number of cancer epigenetic biomarker academic publications per annum over the last 20 years (left axis) in comparison with the cumulative number of registered cancer epigenetic diagnostic tests available on the market (right axis). A PubMed search utilizing the term 'epigenetic biomarkers cancer' was used to determine the number of publications per year and the number of registered tests is referenced in **Table 1**.

many registered translated tests [i.e. those with Food and Drug Administration (FDA) pre-market approval (PMA), a European CE mark (CE-IVD) or registered as a lab-developed test (LDT) through the Centers for Medicare and Medicaid Services [CMS]] utilize ctDNA as their target (**Table 1**), ctDNA is not a cure-all for current diagnostic shortfalls. Circulating DNA from rare tumor sub-populations may only be present in ctDNA at vanishingly small levels, making it difficult to detect by even the most sensitive methods. Such limitations should be considered when designing new assays or assessing diagnostic results from liquid biopsy. Despite this, with the right biomarker it is possible to design a simple liquid biopsy that can detect cancer or tumor characteristics with excellent sensitivity. Additionally, such an assay may be run serially with minimal impact on patients, even where biopsy is impossible or impractical, such as during advanced metastatic disease. Overall, from an efficacy and translation standpoint, liquid biopsy is an extremely attractive strategy with the capacity to greatly transform disease diagnosis and management in the near future.

There are at least six broad diagnostic areas in which a DNA methylation cancer liquid biopsy test may be combined with traditional screening and medical imaging for better patient outcomes:

1. Primary diagnosis: determine individuals potentially presenting with cancer and who should be followed up by traditional screening exams.
2. Triage: after indeterminate results from imaging or biopsy, a further test to decide invasive and/or non-invasive follow-up.
3. Choice of therapy: diagnostics which influence the choice of treatment. This can include prognostic markers, which grade tumors as likely to be treatable with first-line therapy, or companion diagnostics, in which a test result is linked with efficacy of a particular treatment (may also be based on direct tumor biopsy).
4. Response to therapy and treatment failure: the measurement of ctDNA tumor load in the blood to monitor the initial response to therapy and to detect a later rise in ctDNA load consistent with subsequent resistance to therapy. Testing can be serial to detect trend.
5. Residual disease monitoring: for determining minimal residual disease after surgery, adjuvant chemotherapy or radiotherapy and identify patients at increased risk of disease recurrence.
6. Recurrence: early detection of recurrence to give more opportunity for treatment with curative intent.

Most existing tests for cancer screening, diagnosis, or monitoring are protein immunoassays or imaging. For example, many countries have adopted the prostate specific antigen (PSA) test as a population screen for prostate cancer; and the fecal occult blood test (FOBT) or improved fecal immunochemical test (FIT) for population screening of CRC. Although inexpensive and widely used, none of the screening or recurrence tests have the ideal performance characteristics for their respective cancer type, providing opportunity for development of alternate tests, such as DNA methylation tests, to better inform clinical management.

TABLE 1 | Current registered liquid biopsy tests in the marketplace.

Test name	Tissue	Biomarker(s)	Biosample	Population	Intended clinical use	Manufacturer/ Distributor	Approval
AssureMDx™	Bladder	*TWIST1*, *ONECUT2*, *OTX1* (+ *FGFR3*, *TERT*, *HRAS* mutations)	Voided urine	Patients diagnosed with hematuria	Detection of bladder cancer to avoid cystoscopy	MdxHealth	2017*
Bladder CARE™	Bladder	*SOX1*, *IRAK3* and methylated LINE1	Voided urine	Patients with a history of bladder cancer, smokers, and specific occupations, not currently included in a bladder cancer screening program	Detection of bladder cancer	Pangea	2019*
Bladder EpiCheck®	Bladder	Score over 15 methylation markers	Voided urine	Monitoring for tumor recurrence in patients previously diagnosed with bladder cancer	Surveillance of non-muscle-invasive bladder cancer (NMIBC)	Nucleix	2017†
therascreen® PITX2 RGQ	Breast	*PITX2*	Formalin fixed paraffin-embedded (FFPE) tumor tissue taken from primary lesions	Lymph node-positive, ER+, HER2– high-risk breast cancer patients treated with anthracycline chemotherapy	Predict response to anthracycline-based chemotherapy	Qiagen	2018†
IvyGene®	Breast, colon, liver, lung	Score	Blood, 40 ml	Direct to consumer	Detection of cancer	Laboratory for Advanced Medicine	2018*
GynTect®	Cervical	*ASTN1*, *DLX1*, *ITGA4*, *RXFP3*, *SOX17*, *ZNF671*	Cervical smear in STM medium	Women who are HPV-positive with abnormal cytology findings (Pap III, Pap IIID)	Triage of unclear cervical cancer screening tests	Oncgnostics	2019†
QIAsure	Cervical	*FAM19A4*, *hsa-mir124-2*	Cervical scrape, vaginal sample	Women who are high-risk HPV positive or have ASC-US cytology	Triage of unclear cervical cancer screening tests	Qiagen	2016†
Cologuard®	Colorectal	*NDRG4*, *BMP3* (+ *KRAS* mutation, occult hemoglobin)	Stool	Patients, 50 years and older, at average risk who are typical candidates for CRC screening	Detection of colorectal cancer (CRC)	Exact Sciences	2014‡
ColoSure™	Colorectal	*VIM*	Stool	Patients unwilling or unable to undergo a more invasive exam	Detection of CRC	LabCorp	2008*
COLVERA™	Colorectal	*IKZF1*, *BCAT1*	Plasma, 3.9 ml	Detect both residual disease and recurrent disease in CRC patients	Detection of residual disease post-surgical resection, for surveillance of recurrent CRC after primary treatment	Clinical Genomics	2016*
Epi proColon®	Colorectal	*SEPT9*	Plasma, 3.5 ml	Patients, 50 years or older, with average risk for CRC, who decline other CRC screening	Detection of CRC	Epigenomics	2016†,‡
Human MGMT Gene Methylation Detection	Glioblastoma	*MGMT*	Tumor biopsy	Glioblastoma patients	Predict response to alkylating agent chemotherapy such as Temozolomide	Xiamen SpacegenCo	2016†
PredictMDx™	Glioblastoma	*MGMT*	Tumor biopsy	Glioblastoma patients	Predict response to alkylating agent chemotherapy such as Temozolomide	LabCorp	2012*
therascreen® MGMT Pyro®	Glioblastoma	*MGMT*	Blood ctDNA or FFPE Tumor biopsy	Glioblastoma patients	Predict response to alkylating agent chemotherapy such as Temozolomide	Qiagen	2015†

(*Continued*)

TABLE 1 | Continued

Test name	Tissue	Biomarker(s)	Biosample	Population	Intended clinical use	Manufacturer/ Distributor	Approval
HCCBloodTest	Liver	*SEPT9*	Plasma, 3.5 ml	Patients with cirrhosis	Detection of hepatocellular carcinoma	Epigenomics	2019†
Epi proLung®	Lung	*SHOX2*, *PTGER4*	Plasma, 3.5 ml	Increased risk patients defined by life history, presentation with symptoms, radiological findings in the lung	Detection of lung cancer in patients at increased risk for the disease	Epigenomics	2017†
ConfirmMDx	Prostate	*GSTP1*, *APC*, *RASSF1*	Prostate biopsy	Men with established risk factors	Detection of occult prostate cancer on previously biopsied, histopathologically negative tissue	MdxHealth	2012*
EPICUP™	Unknown Origin	Human Methylation450 BeadChip	Fresh frozen or FFPE tumor biopsy	Patients with cancer of unknown primary (CUP) origin	Predict cancer tissue of origin to enable direction of tumor type-specific therapy	Ferrer	2015†

**CLIA LDT, †CE-IVD, ‡FDA PMA*
HPV, human papillomavirus; ASC-US, atypical squamous cells of undetermined significance.

Performance Characteristics

The six diagnostic areas above are best addressed with a blend of liquid biopsy technologies. For population screening (primary diagnosis and triage) the diagnostic must be inexpensive, non-invasive, reliable, and have high specificity to reduce false positive results and unnecessary follow-up procedures. With residual disease monitoring and recurrence, the diagnostic test should exhibit high sensitivity. Ideally, response to therapy and treatment failure diagnostic tests should be rapid and inexpensive. The response area is well suited for point-of-care devices which allow immediate decision-making about treatment efficacy and allow inexpensive serial testing to quickly flag the onset of resistance to current therapy. Choice of therapy diagnostics are tailored to fit clinical decision-making around treatment and can also be designed as companion diagnostics developed in conjunction with a partnered therapeutic intervention.

Going forward, the general nature of DNA methylation ctDNA diagnostics and their economic and high-throughput nature suggests these markers will continue to have a growing role in the six broad areas outlined above. While somatic mutation screening surveys of ctDNA using next-generation sequencing (NGS) and the examination of CTCs are more expensive, they offer unique insight around treatment options and the development of resistance as these approaches can reveal "druggable mutations". The economics of the therapy response market support these expensive tests; often this precision oncology information informs whether the prescription of expensive chemotherapy drugs will be efficacious.

Combining With Other Modes of Detection

The performance of liquid biopsy ctDNA somatic mutation tests is reduced in earlier stage tumors, likely due to far lower levels of ctDNA in the blood (Bettegowda et al., 2014). Reduced sensitivity for earlier stage tumors is also observed with DNA methylation-based liquid biopsy tests of ctDNA, even though the methylation changes are apparent in early stage tumor sections (Church et al., 2014; Pedersen et al., 2015a). For example, *BCAT1* and *IKZF1* are hypermethylated in 97.8% and 86.8% of CRC tumor biopsies, respectively; yet, the ability to detect ctDNA using the same assay vary by staging, tumor size, location, and lymphatic invasion (Pedersen et al., 2015a; Symonds et al., 2016; Jedi et al., 2018; Symonds et al., 2018). Early stage tumors are not highly vascularized and have little central necrosis, which may explain the low ctDNA concentration in the blood. To raise the likelihood of detecting these rare ctDNA fragments, many biomarkers can be screened at the same time (Elazezy and Joosse, 2018), however this raises the test complexity and price. The diagnostic power to detect tumors can be increased by combining multi-analyte modes of detection into a single test, such as the CancerSEEK test which combines sequencing ctDNA with detection of serum protein biomarkers (Cohen et al., 2017; Cohen et al., 2018). Another option is to examine ctDNA in alternate clinical specimens, e.g. in urine to diagnose bladder or prostate cancer, sputum for lung cancer, cerebrospinal fluid for glioma, and stool for CRC.

In the instance of CRC, an early stage tumor that sheds little ctDNA into the bloodstream may cause bleeding into the bowel and the FOBT or FIT will detect the hemoglobin resulting from this bleeding (Symonds et al., 2016). Conversely, late stage cancers might be more readily detected *via* the blood than stool (Ahlquist et al., 2012) and there is some evidence that people perceive a DNA-based stool test as preferable over FOBT (Schroy and Heeren, 2005) and a blood-based ctDNA test over a stool based test (Osborne et al., 2012; Adler et al., 2014). A comparison of the sensitivity of FIT and DNA-based tests to detect advanced precancerous lesions, early- and late-stage cancer is presented on **Table 2**. For CRC, a three-protein ELISA panel has been developed that has higher sensitivity and specificity for early stage I-II disease than the FOBT (Fung et al., 2015). This work has been translated into a company (https://www.rhythmbio.com/).

Liquid biopsy tests and traditional medical imaging can also be combined as they offer complementary means to detect

TABLE 2 | Comparison of commercially available assays for CRC.

	OC-SENSOR®	ColoGuard®	Epi proColon®	Epi proColon®	IKZF1/BCAT1
Assay	Fecal immunochemical test (FIT); 100 μg Hb/g	*KRAS* mutations, methylated *NDRG4*, *BMP3* and hemoglobin	Methylated *SEPT9*	Methylated *SEPT9*	Methylated *BCAT1* and/or *IKZF1*
Biosample	Stool	Stool	Blood	Blood	Blood
Study cohort size	9989 (65 tumors)	9989 (65 tumors)	1544 (44 tumors)#	1510 (53 tumors)#	2101 (85 tumors)
Specificity	94.9%	86.6%	80.0%	91.5%	93.8%
Sensitivity	73.8%	92.3%	68.2%	48.2%	65.9%
Advanced precancerous lesions*	23.8%	42.4%	21.6%	11.2%	6.2%
Stage I	65.5%	89.7%	41.1%	35.0%	37.9%
Stage II	76.2%	100.0%	83.3%	63.0%	69.0%
Stage III	90.0%	90.0%	80.0%	46.0%	72.5%
Stage IV	75.0%	75.0%	100.0%	77.4%	93.8%
Data reference	(Imperiale et al., 2014)	(Imperiale et al., 2014)	(Potter et al., 2014)	(Church et al., 2014)	(Pedersen et al., 2015b)

**Defined as advanced adenomas and sessile serrated polyps measuring 1 cm or more*
#Standardized estimates

and monitor cancer. While ctDNA tests are less expensive than imaging, and can characterize a tumor and potentially earlier detection of recurrence (Ulrich and Paweletz, 2018), these tests do not routinely identify the location of the tumor. The ability for medical imaging to identify the location of tumor(s) is particularly important pre-surgery and in metastatic disease. In the context of lung cancer screening, liquid biopsy tests have the advantage that they do not present high-risk populations (typically older smokers) with a lung radiation dose. In the primary diagnosis setting, ctDNA tests can be used as triage diagnostics after a scan, when a low-dose CT scan or mammography, for example, reveals an indeterminate mass. With sufficient specificity and sensitivity, ctDNA tests may replace riskier biopsy procedures.

Product-Market Fit

With the increasing rise in chronic illness, aging populations, and climbing national healthcare expenditure, governments are increasingly looking at costs per relevant clinical outcome (Anderson and Frogner, 2008; Anderson et al., 2014). Fee-for-service payment models which reward volume will be replaced by quality metrics which value health outcomes achieved per dollar spent (Conway, 2009). Furthermore, the rise in precision therapies in oncology is creating the opportunity for more tailored treatments. As such, global healthcare in the 21st century is characterized by evidence-based medicine, patient-centered care, and cost effectiveness (Bae, 2015). In determining the market value of an *in vitro* diagnostic (IVD), technology investors and healthcare payers need to be provided with the appropriate evidence. It follows that the perceived value of an IVD is proportional to the quality of the evidence.

A well-tested case around product-market fit is useful for defining the clinical gap, who might be willing to order the IVD, who are the payers and if the proposed technological solution is a match for the identified marketplace. Factors to consider are price, assay time and performance metrics like sensitivity and specificity as well as logistics and the potential to meet market demand and expectations. For example, in a centralized lab model, one needs to consider how the analyte(s) are transported and the sample conditions required, and for assays with large potential markets, such as the primary diagnosis of common cancers, how the IVD be simplified, sequenced, and automated to scale to potentially huge volumes of tests per year. The 2010 review on the development of Epi proColon (Payne, 2010) provides an informative narrative on the development and translation of epigenetic diagnostics. From experience, Payne emphasizes that the platform and degree of test automation must be considered early in development and that the test should be robust to detect very low numbers of target molecules in a high background of non-target DNA.

Diagnostics should not just have technical or classification merit but must directly inform clinical decision making in a timely manner. Assays which define prognostic risk or estimate survival can assist in clinical decision-making regarding prescription of a more aggressive protocol or second-line therapy in poor prognosis cases, or in cases of likely predicted recurrence of metastatic disease, an increase in patient surveillance. While the latest molecular technologies can offer benefits to IVD performance metrics, IVDs depending on new technologies can be expensive to implement, automate, and regulate. More expensive IVDs are potentially a better fit for clinical decisions with large financial costs or health risks, such as a decision to administer a second-line therapy or to undertake a significant surgical procedure.

The utility of a primary diagnosis IVD should not just be considered in terms of the number of additional cancers detected over standard care, but also the costs and risks, both for the patient and to the healthcare system, for reporting false positive results. The costs and risks for each tumor type are contextualized by the incidence rate and available follow-up procedures. The true positive rate, known as positive predictive value (PPV), can be increased by targeting the clinical translation to higher risk sub-populations, such as smokers for lung cancer and BRCA mutation carriers for ovarian cancer, but even then, issues remain (Pearce et al., 2015). The problem of unnecessary procedures and patient psychological harm is very real in screening programs. For example, using low-dose CT for lung cancer screening, results from The National Lung Screening Trial revealed that

80.5% of cancer-free participants experienced unnecessary follow-up imaging studies, with 2.2% of participants having an invasive bronchoscopy procedure and 1.3% unnecessary surgery. Another screening study using cancer antigen 125 (CA-125) found that for each ovarian and peritoneal cancer detected by screening, an additional two women had false-positive surgery with a surgical complication rate of 3.1% (Jacobs et al., 2016).

Product-market fit needs to be considered early in the diagnostic development process. The identification of the prospective markets, clinically relevant patient group(s) and what clinical decisions happen after a positive test result should inform considerations around price, required turnaround time and minimal sensitivity and specificity metrics. The market size informs scale considerations and the biosample collection procedure, the need for ambient or cold chain transport logistics. All these parameters collectively inform the design of the diagnostic assay.

Pre-Analytic Conditions

Before the collection of clinical samples, the pre-analytic conditions for how tumor biopsies, blood, or other biosamples will be prepared and stored for later analysis require consideration, including quality control for sample integrity (e.g. cell lysis or nucleic acid degradation). It is commonplace for tumor tissue sections to be stored in a fixative. By necessity, diagnostic tests utilizing tissue sections need to be robust to analyze potentially heavily degraded DNA in formalin fixed paraffin-embedded (FFPE) samples. As the liquid biopsy diagnostics area matures toward increased clinical translation, there is a strong focus on controlling for pre-analytical variables. Guidelines are now coalescing around the optimal preanalytical conditions for analyzing cfDNA (Meddeb et al., 2019) and two large consortia have formed to standardize pre-analytical steps and downstream protocols. The CANCER-ID European Public-Private-Partnership (www.cancer-id.eu) commenced at the start of 2015 and has 36 partners from 13 countries with aims to establish standard protocols for clinical validation of blood-based biomarkers. The USA-based Blood Profiling Atlas in Cancer (BloodPAC; www.bloodpac.org) consortium formed in 2016 is aggregating, harmonizing, and making freely available data from CTC, ctDNA, protein and exosome assays, and the associated clinical data and biosample collection protocols.

Suitable Tissue and Analytes

The stability of epigenetic marks on DNA means there a few limitations on possible analytes with almost all tissues useful for designing DNA methylation-based diagnostics.

Blood

Blood represents a rich source of information on tumor biology and is usually the tissue of choice for ctDNA studies. DNA methylation can be assayed easily using existing methods. There is potential for other epigenetic data to be determined from ctDNA, such as nucleosome positioning and gene activity. Using sequencing approaches, ctDNA fragment ends can be used to estimate genomic activity (Snyder et al., 2016) and predict gene expression (Ulz et al., 2016) without biopsying the tumor itself. While this is a very early area of research these findings open the window to detailed assessments of intra-tumoral biology without access to tumor tissue and without dependence on just one epigenetic mark (i.e. DNA methylation).

The use of ctDNA in clinical settings does have a set of known caveats, in particular, low yields of DNA and the level of contaminating DNA from other cells. The bulk of cfDNA found in blood derives from nucleated blood cells, with a proportion from vascular endothelial cells and liver (Moss et al., 2018). Special consideration must be taken in handling blood samples in the clinical setting, as white blood cell lysis can produce large quantities of fragmented DNA. Typically, ctDNA represents only a very small fraction of total cfDNA, so inappropriate handling of blood samples may result in near complete loss of measurable signal. This risk can be abrogated through the use of careful blood processing techniques or specialized cfDNA collection tubes which stabilize white blood cells (Meddeb et al., 2019). Examples include PAXgene® Blood ccfDNA Tube (Qiagen), Cell-Free DNA Collection Tube (Roche), cf-DNA/cf-RNA Preservative Tube (Norgen Biotek), and Cell-Free DNA BCT® (Streck).

Urine

Sources of cfDNA in urine can be categorized into three sources: pre-renal that can be mostly attributed to blood cells (from the systemic circulation), renal, and post-renal from the bladder urothelium. The median relative contributions of these three tissues are around 52%, 32%, and 5%, respectively. These values do vary largely across patient urine samples, but the ranked order is consistent (Cheng et al., 2017). Compared to blood cfDNA testing, urine cfDNA has two advantages; firstly, it is far easier and cheaper to obtain urine than blood, making urine an ideal biofluid in resource-limited settings (Lawn et al., 2012). Secondly, urine is thought to be a more sensitive alternative for early detection or monitoring recurrence of cancers in the genitourinary tract (Lin et al., 2017). Presently, none of the registered cancer IVDs are based purely on urinary cfDNA. One major reason is because the workflow in preserving urine cfDNA has yet to be standardized. The activity of DNase I in urine, relative to serum, is around 100-fold higher (Bryzgunova and Laktionov, 2015); as such, the half-life of urine cfDNA at body temperature is around 2.6–5.1 h (Cheng et al., 2017).

For clinical purposes, methods that stabilize urine cfDNA and prevent the lysis of nucleated cells are imperative to ease end-user collection. Some products addressing this unmet need have entered the market. These preservatives tend to be colored liquids (to provide visual indication for their addition), or as a dried coating lining the collection container. Examples include Urine Preservation (Norgen Biotek), Cell-Free DNA Urine Preserve (Streck), Quick-DNA Urine Kit (Zymo Research), and NextCollect™ (Novogene). For research purposes, there are kits designed specifically for urine cfDNA, but across the kits, the extracted DNA displays significantly different yields and size profiles (Diefenbach et al., 2018; Streleckiene et al., 2018).

As isolating urine cfDNA remains a technically challenging problem, current biomarker discovery efforts are mostly based on the cellular fraction of the collected urine. Compared to

blood, practical use of urine markers in detecting or monitoring cancer is limited. One contributing factor is that cut-offs or thresholds derived from clinical studies tend to be specific to the study despite a focus on the same marker; the lack of standardized methodology also leads to different definitions of optimality. Binary thresholds resulting from differing definitions are problematic, more so for patients close to the cut-off point (Lotan et al., 2010). Currently marketed tests address this limitation by relying on a panel of biomarkers (Gallioli et al., 2019), or constrain themselves to recurrence monitoring.

Stool

Analysis of fecal material is useful for a range of bowel conditions, e.g. efficiency of digestion, leaky gut syndrome, inflammatory bowel disease, dysbiosis, acute infections, and CRC (Siddiqui et al., 2017). Stool testing for CRC is widely used and robust collection regimes are well established with home-based collection kits routinely used. Test kits have a stabilization agent as this is critical for maximizing the performance of fecal DNA-based tests (Olson et al., 2005; Nechvatal et al., 2008) and stool contains polymerase chain reaction (PCR) inhibitors, which need to be removed (Flekna et al., 2007). The fraction of human epithelial cell origin DNA in stool is small compared to total bacterial DNA, so a PCR diagnostic assay must also be robust to this background (Nechvatal et al., 2008).

Airway

Studies have demonstrated that methylated DNA can be detected within respiratory derived biological samples, specifically sputum (Hulbert et al., 2017), bronchoalveolar lavage (Um et al., 2017), nasal washing/brushing (Yang et al., 2017a; Nino et al., 2018), and exhaled breath condensate (EBC) (Xiao et al., 2014). Not surprisingly, the majority of the literature has focused on the role of this methylated DNA in lung associated pathologies such as asthma, cystic fibrosis, and lung cancer (Konstantinidi et al., 2015).

After a radiological procedure highlights an indeterminate lung mass, a reasonable first step in the investigation is the cytological analysis of sputum to detect lung cancer associated cells. This has a clinical sensitivity of 66%. Further follow-up tests with higher sensitivity are likely required, such as the biopsy of suspected lung nodules (90% sensitivity), but this is a highly invasive and risky procedure, with a 15% chance of collapsing a lung (pneumothorax) (Rivera et al., 2013). Detection of methylated ctDNA is presenting as a viable alternative to cytology of sputum. In a large cohort of lung cancer patients, it was demonstrated that measuring the methylation pattern of eight genes had a lung cancer prediction accuracy of 82%-86%, and a negative predictive value (NPV) from 88% to 94% to rule out cancer (Leng et al., 2017). Another study demonstrated that using the methylation status of genes *TAC1*, *HOXA17*, and *SOX17* in sputum had a sensitivity of 93% to detect lung cancer (Hulbert et al., 2017). These studies show that DNA methylation detection in sputum has greater sensitivity than sputum cytology. However, a major problem is that there is no standardization of sputum acquisition and handling so pre-analytical variables remain a major challenge to translation (Rivera et al., 2013).

Bronchoalveolar lavage is a process where bronchoscopy is used to locate the lung lesion, which is subsequently washed (lavage) with 10–20 ml of isotonic saline and collected for analysis. For easily visible and accessible central lesions, forcep biopsies of the lesions are performed (74% sensitivity) followed by bronchoalveolar lavage (48% sensitivity). However, the sensitivity is lower for peripheral lung lesions as they are difficult to locate and visualize, with a sensitivity of 57% and 43% for transbronchial biopsies and bronchoalveolar lavage, respectively (Rivera et al., 2013). While cytology analysis is typically performed on bronchoalveolar lavage (Carvalho et al., 2017) there is also opportunity to use the lavage fluid for liquid biopsy. Methylated *SHOX2* and *RASSF1A* gene promoters were detected in lavage fluid from 322 patients with a sensitivity of 81% for lung cancer detection (Zhang et al., 2017). Other studies have also revealed high sensitivity for lung cancer detection using bronchoalveolar lavage fluid, with 75% and 78% for *PCDHGA12* (Jeong et al., 2018) and *SHOX2* (Dietrich et al., 2012) methylated DNA, respectively. However, these studies use different methodologies to process the lavage samples, so cannot be directly compared.

The collection of EBC is a novel non-invasive measurement method for lung cancer detection. A portable FDA approved device exists for EBC (RTube™ by Respiratory Research, Inc.), however, there are several caveats with using EBC collection for diagnostic purposes. These include the dilution of analytes in the breath condensate and the contamination with DNA from ambient air, saliva, and the nasal epithelium (Horvath et al., 2017; Koc et al., 2019). Furthermore, normalizing for varying levels of condensation arising from different collection methods is a well-known issue (Horvath et al., 2017). While there are several challenges to overcome in developing an IVD, the non-invasive nature of EBC compared to bronchoalveolar lavage makes EBC an attractive biological sample.

Technology

Overview

Bisulfite treatment is the gold standard method for mapping methylated cytosines in DNA and was developed by Australian scientists from the CSIRO and Kanematsu Laboratories in Sydney (Frommer et al., 1992; Clark et al., 1994). With this method, sodium bisulfite is used to convert cytosine residues to uracil residues in single-stranded DNA, under conditions whereby 5-methylcytosine (5mC) remains non-reactive. The 5-hydroxymethylcytosine (5-hmC) epigenetic mark, which is mostly confined to embryonic stem cells and to an extent brain and liver, is indistinguishable from 5mC using bisulfite conversion (Huang et al., 2010). Alternatives to bisulfite treatment are to use enzymes sensitive (or specific) to DNA methylation within their cleavage site or affinity capture using a binding protein or antibody. Bisulfite-treatment can be coupled with multiplexed probe-based detection. Methods which selectively determine the presence of methylated DNA are a good fit for liquid biopsy applications, whereas methods estimating the fraction of DNA methylated at a CpG site (often called the beta-value), are better suited for examining tissue. A brief description and classification of commonly used methods is presented on **Table 3**.

TABLE 3 | Summarised methods for the detection of DNA methylation in liquid biopsy.

Method name	Class*	Sub-class	Bisulfite-based	Description	Citation
Whole Genome Bisulfite Sequencing (WGBS)	GW	High-throughput sequencing	Y	Various approaches for the tagging and sequencing of bisulfite converted DNA. Adapter tagging can be done before or after conversion. Different approaches may introduce biases.	(Lister et al., 2009)
Bisulfite Sanger Sequencing (Bis-Seq)	TGT	Bisulfite conversion specific amplification	Y	Bisulfite-treated genomic DNA subjected to amplification with conversion-specific PCR primers. Primers are unbiased and contain no CpGs	(Frommer et al., 1992)
Nested PCR	TGT	Bisulfite conversion specific amplification	Y	Bisulfite-treated genomic DNA subjected to amplification with conversion-specific PCR primers, followed by secondary PCR with primers targets within the first PCR fragment to enhance specificity or sensitivity.	(Herman et al., 1996)
Methylation-specific PCR	TGT	Bisulfite conversion specific amplification	Y	Bisulfite-treated DNA subjected to amplification. PCR primers intentionally biased by including multiple CpGs in binding sites.	(Herman et al., 1996)
MethylLight	TGT	Fluorescence probe PCR	Y	Conversion or Methylation specific PCR with the addition of a TaqMan probe. Methylation specificity obtained by including CpGs in the primers, probe or both.	(Eads et al., 2000)
Quantitative Allele-specific Real-time Target and Signal amplification (QuARTS)	TGT	Fluorescence probe PCR	Y	Bisulfite-treated genomic DNA subjected to PCR with probes targeting alternate methylation states. Probe fluorescence activated by an additional oligo binding immediately upstream.	(Zou et al., 2012)
HeavyMethyl	TGT	PCR with blocker	Y	Competitive inhibition of PCR using primers combined with a blocker oligo that target alternate methylation states.	(Cottrell et al., 2004)
Cold-PCR	TGT	Preferential denaturation temperature PCR	N	The first few cycles are conventional PCR. Subsequent cycles use a lower denaturation temperature to enrich for DNA molecules that contain mismatches, which occur if there are mutant DNA sequences in the sample.	(Milbury et al., 2011; Castellanos-Rizaldos et al., 2014)
Ice-COLD-PCR	TGT	Preferential denaturation temperature PCR	N	The same as cold-PCR but with the addition of a further blocker oligonucleotide to inhibit amplification of unwanted targets.	(Milbury et al., 2011; Mauger et al., 2018)
High-resolution melt (HRM) curve analysis	TGT	Melt-curve analysis	Y	Following traditional PCR with an intercalating dye (e.g. SYBR green), the PCR product is gradually warmed until the DNA strands denature (melt) apart. DNA melting is detectable by shifts in the level of fluorescent signal over time.	(Wittwer, 2003; Wojdacz and Dobrovic, 2007)
Bis-seq (pyrosequencing)	TGT	Bisulfite conversion specific amplification	Y	"Sequencing by synthesis" method. Sequence readout is obtained by detecting pyrophosphate released during base incorporation during synthesis of the complementary DNA to the target fragment.	(Ronaghi, 1998; Fraga and Esteller, 2002)
EpiTYPER	TGT	Mass-spectrometry	Y	Utilizes mass spectrometry to accurately measure the methylation of PCR-derived amplicons. DNA is converted and amplified by PCR. Incorporation of C/G or T/A bases (methylated or unmethylated) during amplification leads to measurable shifts in molecular weight.	(Ehrich et al., 2005)
Reduced representation bisulfite sequencing (RRBS)	RGW	Enzymatic digest	Y	Genomic DNA digested with methylation-insensitive restriction enzyme (with CpG in the recognition site), followed by size selection prior to bisulfite conversion.	(Meissner et al., 2005)
Combined Bisulfite Restriction Analysis (COBRA)	RGW	Enzymatic digest	Y	Bisulfite-treated genomic DNA is subjected to methylation-insensitive restriction enzyme digest targeting the unconverted amplicon (i.e. originally methylated). Ratio of digested fragments to total fragments correlates with methylation level.	(Xiong and Laird, 1997)
Digital Restriction Enzyme Analysis of Methylation (DREAM)	RGW	Enzymatic digest	N	Methylation specific restriction enzyme (MRSE) variant. DNA digested with two enzymes, one methylation sensitive and one not. Methylation readout based on ratio of cutting.	(Jelinek and Madzo, 2016)
Methylation-sensitive restriction enzyme (MSRE) + qPCR	TGT	Enzymatic digest	N	Unconverted DNA is digested with MRSE. Quantitative PCR is used to establish the efficiency of digestion, which indicates the level of methylation at the target site.	(Hashimoto et al., 2007)
Helper-dependent chain reaction (HDCR)	TGT	Enzymatic digest	N	Genomic DNA digested with methylation dependent restriction enzyme such as GlaI. Gene-specific sequence fragments are tagged with "helper" oligos, while "driver" oligos maintains preferential amplification of tagged fragments.	(Rand et al., 2013)

(*Continued*)

TABLE 3 | Continued

Method name	Class*	Sub-class	Bisulfite-based	Description	Citation
End-specific PCR (ES-PCR)	TGT	Enzymatic digest	N	MRSE variant for detecting unmethylated sequences. DNA is digested with a methylation sensitive enzyme, then specialized oligos are used to add priming sites to the target sequence. Highly useful method for targeting repetitive sequences that are difficult to assay by other methods.	(Rand and Molloy, 2010)
MeDIP	RGW	Affinity capture	N	DNA capture using antibody specific to methylated cytosine. Captured DNA suitable of PCR, array and sequencing based methods.	(Weber et al., 2005)
Various methyl-CpG binding domain (MDB)-based assays	RGW	Affinity capture	N	DNA capture using methylated DNA binding protein (MBD2). Use of salts during elution from MBD can facilitate fractionation on methylation level. Captured DNA suitable of PCR, array and sequencing based methods.	(De Meyer et al., 2013; Aberg et al., 2015)
SuBLiME	RGW	Affinity capture	Y	Biotinylated bases are incorporated into DNA fragments using a PCR-like approach following bisulfite conversion. Biotinylated fragments are captured to enrich for methylated targets. Can be performed in a targeted or genome-wide method.	(Ross et al., 2013)
Bisulfite Specific Padlock Probes (BSPP)	RGW	Molecular inversion probes	Y	BSPP utilizes bisulfite converted DNA and specialized DNA probes. Probes bind two sites in target sequences to form circular DNA structures that can be amplified and sequenced.	(Diep et al., 2012)
Infinium HumanMethylation	RGW	Infinium assay	Y	Bisulfite treated DNA is hybridized to the BeadArray chip, detection using single-base extension and fluorescence ratio between converted and unconverted probes.	(Bibikova et al., 2011)

**GW, genome-wide; RGW, representative genome-wide; TGT, targeted.*

Bisulfite-Treatment

Bisulfite treatment of DNA for diagnostic purposes is not without issues. Foremost, is the significant loss of material due to the harsh chemical and temperature conditions involved (Grunau et al., 2001). This loss reduces sensitivity to detect cancers, especially those releasing low levels of ctDNA. In addition, there is a loss in genome complexity due to the large reduction in the prevalence of cytosine bases in converted DNA, resulting in a largely pseudo-three base genome. Careful PCR primer design is required to specifically amplify rare target molecules in an overwhelming off-target background, such as with a typical methylation-based ctDNA assay.

Bisulfite-treated DNA is typically amplified using conversion-specific PCR (CSP) or methylation-specific PCR (MSP) primers. With CSP, primers are designed to amplify bisulfite-converted DNA regardless of methylation state; while with MSP the primers target unconverted cytosines, such that only methylated DNA is amplified (Herman et al., 1996). To achieve amplification specificity with bisulfite-treated template DNA, nested PCR is sometimes used. However, this is not an optimal fit with an IVD due to exposure of amplified DNA from the first round of PCR into the clinical lab environment. Bisulfite treatment of DNA can be combined with NGS. The entire methylome can be sequenced *via* whole genome bisulfite sequencing (WGBS), or regions targeted by sequencing CSP amplicons. Genome regions can also be targeted using a technique like reduced representation bisulfite sequencing (RRBS), where DNA is digested with a methylation-insensitive restriction enzyme (with CpG in the recognition site), followed by size selection prior to bisulfite conversion (Meissner et al., 2005).

MSP is used to amplify cancer DNA from hypermethylated promoters. With ctDNA assays, the cancer-originating DNA is rare compared to background off-target DNA, so additional measures are often needed such that the PCR assay remains specific even after the large number of amplification cycles needed to observe rare ctDNA. The MethyLight assay is a quantitative MSP with the addition of a TaqMan-based fluorescent probe. It is sensitive for methylation levels as low as 0.01% and has good reproducibility (Eads et al., 2000). The Quantitative Allele-Specific Real-time Target and Signal amplification (QuARTS) method also employs a probe but in addition incorporates a 5′ DNA flap, a flap endonuclease and fluorescence resonance energy transfer (FRET) chemistry for detection of the cleaved products (Zou et al., 2012). The HeavyMethyl method is a quantitative CSP amplification which adds a blocker oligonucleotide that competes for binding across the primer sites to unmethylated DNA, thus preventing efficient amplification of unmethylated DNA (Cottrell et al., 2004).

Other properties of bisulfite-treated DNA can be used to selectively amplify the target molecule, such as preferential amplification using denaturation temperature. This family of methods includes co-amplification at lower denaturation temperature PCR (COLD-PCR) (Milbury et al., 2011; Castellanos-Rizaldos et al., 2014) and bisulfite differential denaturation PCR (Rand et al., 2006), where the basic principle is to select a critical temperature in the PCR to selectively denature unmethylated genomic regions in the presence of an excess of methylated DNA molecules. The methylation-sensitive high-resolution melting (MS-HRM) method uses the difference in melting temperature between methylated versus unmethylated

product after a CSP reaction to quantify methylation (Wojdacz and Dobrovic, 2007). Methylation may also be quantified using a PyroMark pyrosequencer (Qiagen) or the EpiTYPER® mass spectrophotometry instruments (Agena Biosciences). Both approaches can detect small changes in methylation.

Enzyme Cutting

Enzyme-based methods offer an alternative to bisulfite-treatment and are not subject to the same losses of material. The disadvantages are that assayed regions must overlap loci of interest and that incomplete digestion can confound interpretation of the results. Methylation-sensitive restriction enzyme (MSRE) cutting can be coupled with quantitative PCR to estimate DNA methylation, with more product proportional to more methylation at the cut site(s) within the amplicon (Hashimoto et al., 2007). Conversely, a methylation-dependent enzyme such as GlaI can be used to selectively cut only methylated DNA. The selective amplification of DNA with ends cut by GlaI is used in the end-specific PCR (ES-PCR) and helper-dependent chain reaction (HDCR) techniques (Rand and Molloy, 2010; Rand et al., 2013). The Combined Bisulfite Restriction Analysis (COBRA) method is a hybrid which involves cutting DNA that has been first bisulfite-treated and PCR amplified (Xiong and Laird, 1997).

The Digital Restriction Enzyme Analysis of Methylation (DREAM) is a method for mapping DNA methylation levels at a specific set of CpG sites that are contained within the recognition sequence, 5'-CCCGGG-3' for two restriction enzymes, SmaI and XmaI (Jelinek et al., 2012). It relies on the differential sensitivity of the two enzymes to methylation at the central CpG site and their different modes of cutting. Cutting by SmaI is blocked by methylation of the central CpG site, while XmaI cuts whether the CpG site is methylated or not. Thus, methylated sites are scored indirectly as those 5'-CCCGGG sites that are not cut by SmaI.

Affinity Capture

Affinity capture techniques are used to enrich methylated DNA from the overall DNA population. This is usually accomplished by antibody immunoprecipitation methods or with methyl-CpG binding domain (MDB) proteins and there are modifications to the protocol that also enable hydroxymethylation capture (Thomson et al., 2013). Input genomic DNA can be sonicated or enzymatically digested prior to capture and purification, often *via* magnetic beads. Eluted DNA is usually then used as input for the generation of NGS libraries, but also suitable for analysis with microarrays or PCR-based methods. The different variants of this methodological principle result in widely different patterns of the distribution of DNA methylation enrichment (De Meyer et al., 2013; Aberg et al., 2015). An alternative affinity capture technique utilizes the incorporation of biotinylated cytosines during amplification of bisulfite-treated sheared or digested genomic DNA fragments followed by affinity capture using streptavidin-coupled magnetic beads (Ross et al., 2013).

Multiplexed Probe-Based Detection

The Infinium Methylation Assay detects cytosine methylation at CpG dinucleotides using single-base extension of two site-specific probes, one each for the methylated and unmethylated locus in a highly multiplexed reaction on bisulfite-converted genomic DNA. The level of methylation for the interrogated locus can be determined by calculating the ratio of the fluorescent signals from the methylated vs. unmethylated sites. This is by far the most widely used "genome-wide" DNA methylation analysis platform with significant amounts of public data available. The bioinformatics analysis pipelines for this platform are also mature. The currently available third iteration of this platform is the Infinium MethylationEPIC BeadChip, which interrogates 863,904 CpG sites.

Padlock probes are single stranded DNA molecules with two segments complementary to the target DNA connected by a linker sequence, which are hybridized to the DNA target to become circularized (Nilsson et al., 1994). Molecular Inversion Probes (MIP) are derivatives of padlock probes, although they contain a gap in the target sequence, which provides for greater flexibility. These probes can be used for various forms of genomic partitioning, single nucleotide polymorphism (SNP) genotyping, or copy-number variation detection. Bisulfite padlock probes (BSPP) are an adaptation for the analysis of DNA methylation (Ball et al., 2009; Deng et al., 2009; Diep et al., 2012), where padlock probes are hybridized to bisulfite-treated DNA and subsequently interrogated using NGS.

EXISTING REGISTERED ASSAYS

The existing DNA methylation-based ctDNA IVDs with FDA Premarket Approval (PMA) or offered as LDT or European union CE-IVDs are summarized in **Table 1**. A description of these registered tests and upcoming tests on the path to registration follows.

Bladder Cancer

Approximately 70% of bladder cancer cases are non-muscle-invasive (NMIBC). Lifelong post-operative surveillance is essential due to high recurrence rates (50%-70% patients experience recurrence within 5 years), and a moderate chance of disease progression to muscle invasion (10%-15%) (Tilki et al., 2011). The gold standard for diagnosis is cystoscopy and cytology; urinary tests have yet to achieve comparable specificity or sensitivity. However, monitoring for recurrence could be safer and cost effective if the non-invasive test had a high NPV (Witjes et al., 2018).

Bladder EpiCheck® (Nucleix) is a urine assay for NMIBC based on 15 proprietary methylation biomarkers. DNA is extracted from centrifuged cell pellets from 10+ ml of patients' urine, and subjected to methylation-sensitive restriction enzyme digestion before quantitative PCR (qPCR) amplification. The quantitative results are summarized as an EpiScore ranging from 0 to 100 (where scores ≥ 60 are considered positive for recurrence). This test has a reported NPV of 95%-97% and is currently available as a CE-IVD in the EU (Wasserstrom et al., 2016; Witjes et al., 2018; D'Andrea et al., 2019).

Similarly, Bladder CARE™ (Pangea Laboratory) is a urine assay for NMIBC recurrence based on the hypermethylation of a proprietary three-gene panel, likely *SOX1*, *IRAK3*, and methylated LINE1 (Su et al., 2014). According to unpublished

material released by the company, the urine sample (~5 ml) is first mixed in a 1:3 ratio with a stabilization buffer prior to shipment to their clinical lab where DNA is harvested from centrifuged cell pellets, digested with methylation-sensitive restriction enzymes, then amplified with qPCR. Results are summarized as three calls: negative, high-risk, or positive. This LDT currently targets the bladder cancer recurrence market, but promotional materials raise the possibility of early detection due to the high reported PPV and NPV of the test (89% and 92%, respectively).

Hematuria (blood in urine) can be an early sign of bladder cancer, where 3%-28% of patients with hematuria are diagnosed with bladder cancer. AssureMDx™ for Bladder Cancer (MDxHealth) is a urine assay that excludes bladder cancer diagnosis based on a negative result (99% NPV), leading to 77% reduction in diagnostic cystoscopies, resulting in lower diagnostic costs and reduced patient burden (van Kessel et al., 2017). DNA from cells in the urine samples are subjected to a methylation specific PCR targeting three genes (*OTX1*, *ONECUT2*, and *TWIST1*). In addition, the mutation status of three other genes (*FGFR3*, *TERT*, and *HRAS*) provided additional support for the predictive model (Su et al., 2014; van Kessel et al., 2016; van Kessel et al., 2017). This product is currently available as an LDT in the USA.

A promising candidate test for patients presenting with hematuria is UroMark (University College, London), currently in validation studies in the UK. The initial study demonstrated high PPV and NPV (100% and 97%, respectively) (Feber et al., 2017). This test detects the methylation status of 150 loci across the genome, which is obtained from subjecting cell pellets from urine samples to a microdroplet-based PCR amplification of bisulfite-converted DNA.

Breast Cancer

Breast cancer is a highly heterogeneous disease and molecular subtyping has proven effective in reducing mortality. Breast cancer subtypes and treatments are traditionally determined using histopathology for key hormone receptors that are also the targets of most common frontline therapies. Recent IVDs utilizing gene expression and mutational profiles aim to stratify patients into risk/treatment groups (e.g. PAM50/Prosigna, OncotypeDX, and Endopredict). These methods use traditional tissue biopsies and do not make use of DNA methylation. Despite the success of molecular testing in breast tumors, current DNA methylation-based assays and liquid biopsy offerings in breast cancer are sparse and no methylation-based ctDNA assays are available. Current DNA methylation-based offerings are limited to the therascreen® PITX2 RQG test developed by Qiagen/Therawis which is available as a prognostic/predictive CE-IVD in the EU.

Qiagen's therascreen® PITX2 RGQ PCR Kit is a qPCR-based assay that determines the ratio of methylated to unmethylated DNA content in tumor histology sections, where percent methylation ratio (PMR) is indicative of overall survival and patient outcome when anthracyclines are combined with standard therapy (Maier et al., 2007). Anthracyclines carry serious side that may limit treatment (Volkova and Russell, 2012) and patients with less aggressive tumors subtypes or other contraindications may be adequately treated with standard approaches (Turner et al., 2015). By using the therascreen® PITX2 assay, the risk of over-treatment can be minimized without risk to patient outcomes. However, the therascreen® test is limited to estrogen receptor-positive, node-negative tumors only. The more aggressive and/or difficult to treat HER2-positive and triple-negative subtypes or tumors with lymph node involvement do not benefit from this assay.

Cervical Cancer

The screening and detection of cervical cancer has been transformed by the relatively recent discovery of the role of Human Papilloma Virus (HPV) in the initiation and progression of this disease. Traditional cytological screening has now been displaced by modern molecular methods that target HPV. These new approaches are both cheaper and more effective at identifying at risk women, even when screening intervals are increased (Brotherton et al., 2016). The development of the highly successful vaccine against HPV will have a continuing disruptive impact on cervical cancer screening with HPV incidence in young women trending toward zero in nations with effective vaccination programs (Read et al., 2011; Ali et al., 2013; Brotherton et al., 2016).

Unsurprisingly, epigenetic diagnostics available in the market have positioned themselves as triage tests following positive HPV findings. Three competing tests exist in the marketplace, QIAsure (Qiagen), GynTect® (Oncgnostics GmbH), and the CONFIDENCE assay (Neumann Diagnostics). However, the DNA methylation component of the Neumann assay is currently awaiting full certification. All three tests utilize liquid samples from cervical scrapings/smears with minor differences in methodology. GynTect offers a slightly more streamlined protocol when compared to QIAsure, with no dedicated DNA extraction step. QIAsure offers an alternative convenience for patients, in the fact that it offers a process for both physician and self-collected cervical samples without loss of sensitivity (De Strooper et al., 2016) whereas GynTect and CONFIDENCE are limited to physician collected samples only. Target genes are also another source of difference, with QIAsure targeting the promoters of tumor suppressor genes *FAM19A4* and *hsa-mir124-2* and another non-specific positive control. GynTect targets a larger number of genes including *ASTN1*, *DLX1*, *ITGA4*, *RXFP3*, *SOX17*, and *ZNF671* plus two quality control regions. The CONFIDENCE assay targets the fewest sites, measuring methylation at the *POU4F3* gene and one other control region (*COL2A1*) (Kocsis et al., 2017).

The utility of these assays in triaging patients exists in the epigenetic biology of HPV-driven carcinogenesis. HPV detection on its own is not necessarily indicative of the likely presence of cancer, most infections will be benign, and those patients will require no further treatment. Malignant infections will trigger the expression of pro-oncogenic viral genes leading to the formation of the precursor lesion transforming cervical intraepithelial neoplasia (CIN). As CIN progresses from low to high grade (CIN1–3) there is a sequential build-up of DNA methylation aberrations across the genome. By targeting genes

associated with high grade/risk CIN these tests can provide as surrogate for CIN grade that can be used to stratify patients into high/low risk groups. In terms of assay performance, QIAsure's sensitivity of 70.5% for CIN3+ samples exceed GynTect's 61.2%. However, GynTect does have a substantially improved specificity over QIAsure (94.6% and 67.8% for GynTect and QIAsure respectively (De Strooper et al., 2014; Schmitz et al., 2018). The predictive values of both tests are comparable, with NPV for both tests ~90% although QIAsure's reduced specificity does result in better PPV for GynTect. Ultimately, published data shows the two tests to be comparable in performance although ongoing trials may distinguish the two sometime in the future.

Colorectal Cancer

For primary diagnosis of CRC two tests have progressed through to FDA PMA approval, the blood-based Epi proColon® (Epigenomics) and the stool-based Cologuard® (Exact Sciences). Both tests are approved for patients ≥ 50 years of age and require follow colonoscopy for a definitive diagnosis. ColoSure™, a stool-based LDT for primary diagnosis test which detects methylated *VIM* (Ned et al., 2011) has been withdrawn from sale. More recently, COLVERA™, a blood-based test for the detection of CRC recurrence, has been distributed in the USA as an LDT since 2016.

Cologuard® is a stool-based DNA test which consists of a regular FIT together with amplification of methylated *BMP3* and *NDRG4*, β-actin methylation control, and mutant *KRAS*. Cologuard® has been tested in a large asymptomatic screening population consisting of 9,989 patients (Imperiale et al., 2014) and found to have sensitivity for CRC detection similar to that of colonoscopy, and superior sensitivity for advanced precancerous lesions and early stage cancer when compared to FIT (**Table 2**). However, the specificity is lower with Cologuard® in comparison to FIT (**Table 2**). Cologuard® was approved as a screening test for CRC by the FDA in 2014. Cologuard's estimated market share after Q1 2019 is 4.6% and approximately a million tests were ordered in 2018. Exact Sciences is also submitting an application to expand Cologuard's label to include the 45–49 age group in accordance with updated screening guidelines in the USA (American Cancer Society, 2019) to increase the test's market opportunity. Together with researchers at the Mayo Clinic, Exact Sciences is also currently developing an updated version of the test with additional biomarkers.

Epi proColon® has had less market traction than Cologuard®. Epi proColon® detects the presence of methylated *SEPT9* in plasma; it has higher specificity than Cologuard®, but less than FIT and is less sensitive than both. Epi proColon® is not recommended for routine screening of CRC, but is an alternative to patients, 50 years or older, with average risk for CRC, who decline other CRC screening such as FIT or screening colonoscopy.

After surgical resection and subsequent chemotherapy treatment for CRC, there is a 30%-50% chance that the disease will recur within 5 years. This is typically observed as distant metastases of the liver, lung, or locoregional areas (Duffy et al., 2003). Carcinoembryonic antigen (CEA) has historically been the only non-invasive biomarker in routine clinical practice for surveillance of disease recurrence. However, CEA has poor sensitivity (35% with 95% specificity) and blood CEA levels are not elevated in 58% of CRC patients (Goldstein and Mitchell, 2005). Although serial measurements of CEA are widely used in surveillance, there is variable agreement about what constitutes a clinically significant increase. The European Group on Tumour Markers (EGTM) guidelines guardedly define this as at least 30% over the previous value with increase to be followed by a second sample taken within 1 month and a confirmed trend investigated to detect or exclude malignancy (Duffy et al., 2003).

CSIRO co-developed the methylated two-gene (*IKZF1* and *BCAT1*) panel COLVERA™ liquid biopsy test with Clinical Genomics and the Flinders Centre for Innovation in Cancer (Mitchell et al., 2014; Mitchell et al., 2016). COLVERA™ has been available since 2016 in the USA as an LDT to detect residual disease post-surgical resection and for surveillance of recurrent CRC after primary treatment. COLVERA™ is informative with respect to completeness of surgical resection, risk of residual disease, and recurrence-free survival (Murray et al., 2018). It has double the sensitivity of CEA and should allow more judicious use of PET-CT (Young et al., 2016). The *IKZF1*, *BCAT1* marker pair also shows potential for primary diagnosis of CRC and has demonstrated better performance than the Epi proColon® SEPT9 test (**Table 2**).

Glioblastoma

MGMT (O6-methylguanine DNA methyltransferase) promoter methylation is inversely correlated with *MGMT* expression and patients' response to the alkylating agent temozolomide (Esteller et al., 2000) with approximately 50% of grade IV glioma (usually glioblastoma, GBM) exhibiting *MGMT* promoter methylation (Wick et al., 2014). Multiple large-scale clinical studies have identified that patients having hypermethylation of the *MGMT* promoter region experience significant outcome benefit with temozolomide treatment (Hegi et al., 2005; Stupp et al., 2005). As such, testing for hypermethylated *MGMT* has entered standard care and management for patients with glioma and is a key factor for treatment strategy selection for GBM patients (Louis et al., 2016). To date, there is no consensus on the optimal method for detection of *MGMT* promoter methylation. MSP and pyrosequencing of bisulfite-treated DNA are the most common assay methods, with pyrosequencing likely displaying better performance compared to MSP (Havik et al., 2012). Some studies suggest PCR with HRM has better performance than MSP and pyrosequencing with regards to diagnostic accuracy and efficiency but further large-scale trials are needed to be validated (Switzeny et al., 2016).

There are several methylated *MGMT* IVDs on the market. The pyrosequencing-based therascreen® MGMT Pyro® (Qiagen) is a registered CE-IVD and can quantify four CpG sites in the first exon of *MGMT*. The Human MGMT Gene Methylation Detection Kit (Xiamen SpacegenCo) is also a CE-IVD and is based on Xiamen SpacegenCo's proprietary PAP-ARMS® technology which combines the pre-existing Amplification Refractory Mutation System (ARMS) approach with pyrophosphorolysis-activated polymerization (PAP), increasing

specificity by preventing mismatched primer extension. LabCorp also offer PredictMDx™, an MSP-based test for detecting *MGMT* methylation in FFPE biopsies and licensed from MDxHealth.

Researchers in Heidelberg, Germany have developed an innovative methylation profiling tool for classification of central nervous system tumors based on the Illumina Human Methylation BeadChip data of 2,801 reference samples across adult and pediatric tumors (Capper et al., 2018). Classification by the tool resulted in the revision of the initial histopathological diagnosis in 12% of cases. The pathological reinvestigation was ~93% in favor of the machine learning prediction, demonstrating the power of this approach for correct diagnosis. This methylation profiling classification tool, while for research use and not yet clinically validated, is aimed at generating molecular classification results for treating physicians. The authors developed an interactive website (https://www.molecularneuropathology.org/mnp) that allows researchers to upload their own Illumina Human Methylation BeadChip results and have the sample(s) classified against the references and DNA methylation classification, *MGMT* methylation status, and copy number variation (CNV) returned. Since release, this neuropathology classifier web service has already classified more than 16,000 samples (source from website). To register the classifier as an IVD would be arduous, but clearly this approach has clinical utility and is being adopted by the neuro-oncology community.

Liver Cancer

The HCCBloodTest developed by Epigenomics is a diagnostic blood test for the detection of hepatocellular carcinoma in cirrhotic patients. This duplex real-time PCR based CE-IVD qualitatively detects methylated *SEPT9* DNA, where hypermethylation is indicative of liver carcinogenesis. The gene β-actin is measured in parallel and used as an internal control to determine whether there was sufficient DNA input. The sensitivity of this assay to detect hepatocellular carcinoma is 91% with 87% specificity, based on an initial and replication study (Oussalah et al., 2018) which collectively had 289 patients with cirrhosis and 98 of them having HCC. The test now forms the basis of an ongoing clinical trial on an estimated 220 patients with either clinically-diagnosed cirrhosis without HCC (confirmed by medical imaging) or cirrhosis patients with early-stage HCC.

Lung Cancer

Epi proLung®, a CE-IVD DNA methylation test developed by Epigenomics for the detection of lung cancer, has been tested in a validation study of 360 clinical specimens from the US and Europe (Weiss et al., 2017). Of these specimens, 152 patients were diagnosed with lung cancer (pathologically confirmed), while the remainder were not diagnosed with lung cancer either after a CT scan or radiological examination and follow-up of the pulmonary nodule. The Epi proLung® IVD is a triplex PCR assay that detects methylated *PTGER4* and *SHOX2*, while β-actin is measured as an internal control for sufficient DNA input (Weiss et al., 2017). The procedure to use the Epi proLung kit is the same as the HCCBloodTest by Epigenomics (see Liver cancer). To classify the presence of lung cancer requires the calculation of an Epi proLung test score (EPLT-Score) which aggregates real-time PCR cycle threshold (Ct) values for triplicate assays of *SHOX2* and *PTGER4* into a compound formula. Different EPLT score thresholds result in different performance characteristics, where an EPLT score of −0.43 has a sensitivity of 59% and specificity of 95%, while an EPLT score of −1.85 has a sensitivity of 85% and specificity of 50%.

Prostate Cancer

Population screening using blood levels of PSA has long been used for the early detection and treatment prostate cancer. Although originally used as a marker for recurrent prostate cancer, PSA was eventually adopted by the medical community as a standalone screening test. PSA has a reported specificity of 91% and sensitivity of 21% for primary diagnosis of prostate cancer (with cut-off value of 4 ng/ml) (Brawer et al., 1992; Catalona, 1993). While the use of PSA for screening has led to a decrease in mortality rates, this has come at the expense of tremendous over-diagnosis and subsequent over-treatment of the at-risk population. New data shows that prostate cancer treatment may be unnecessary in anywhere from 2% to 67% of cases with PSA detecting a large number of tumors that are unlikely ever to impact the patient. Given the risk of invasive procedures and serious impact on quality of life reported by patients following prostate cancer treatment (radical prostatectomy) there is an urgent need for better biomarkers in the prostate cancer space. Active surveillance of at-risk patients with repeat PSA measures (quarterly), annual examinations by a physician and regular (3 yearly) scans and biopsies has become the method for treating men with an evidently low-grade tumor. Active surveillance minimizes the risk of over-treatment but depends on the rapid detection of changes in tumor grade or growth. At early stages, the probability of a biopsy collecting a tumor sample may be low as any cancer will be just a small percentage of the total prostate mass. This could lead to missed tumor development at biopsy resulting in delayed time to treatment and potentially decreasing rates of survival. As such, the only widely available epigenetic test in prostate cancer has positioned itself to improve cancer detection at biopsy.

The ConfirmMDx test offered as an LDT by MDxHealth targets regions of DNA associated with the genes *GSTP1*, *RASSF1*, and *APC* that exhibit increased methylation in cancer. However, ConfirmMDx does not require a cancer positive biopsy. The target genes all have reported field effects, that is DNA methylation is altered in normal tissue adjacent to the tumor site. By making use of this biology, ConfirmMDx can be used to verify that a tumor negative biopsy is associated with negative risk. MDxHealth report that this results in greater confidence and reduced need for frequent biopsy. ConfirmMDx has an NPV of >90% for high-grade cancers, in Caucasian and African American cohorts (Van Neste et al., 2016; Waterhouse et al., 2019)

Multiple Cancers

IvyGene (Laboratory for Advanced Medicine) is a test that quantifies the presence of four ubiquitous cancers (breast, colon, liver and lung) by assaying the methylation status of cfDNA from

patients' blood samples across a panel of 46 markers (Hao et al., 2017). This test is marketed as an adjunct clinical test, ordered by physicians to bolster patient observations and available in the USA as an LDT.

Cancer of unknown primary (CUP) origin is a highly heterogenous cancer classification and a particularly frustrating diagnosis for oncologists (Fizazi et al., 2015). Historically, CUP has accounted for anywhere from 3%-9% of all cancer diagnoses (Pavlidis and Pentheroudakis, 2012; Varadhachary and Raber, 2014; Fizazi et al., 2015). However, recent years have seen an apparent decrease in CUP rates to <2% (Urban et al., 2013; Rassy et al., 2019). Despite the rapidly decreasing diagnosis, CUP remains difficult to treat with often poor prognostic outcome (Urban et al., 2013; Fizazi et al., 2015). This is largely due to the fact that CUP is diagnosed only after metastasis and without knowledge of the underlying primary tissue biology. The EPICUP™ assay (Moran et al., 2016) as offered by Ferrer has been developed with the intent of offering these patients more specific diagnoses. EPICUP™ received CE marking in 2015 can be performed using either fresh frozen or FFPE biopsy tissue which is assayed using the Illumina HumanMethylation450 BeadChip. The methylome signature from the BeadChip is used to predict the original tissue of origin and other biological features and to facilitate better treatment decisions. In a multicenter retrospective analysis, this tumor type classifier could predict primary cancer of origin in 87% of patients with a CUP diagnosis (Moran et al., 2016). It should be noted that the underlying platform of this assay (HumanMethylation450 BeadChip) has since been superseded by the more comprehensive Infinium MethylationEPIC BeadChip (Pidsley et al., 2016) and that at the time of writing, Ferrer does not seem to offer an updated product. Still, this assay underlines the unique power of methylome analysis to classify tumors.

GRAIL is a company to watch in the multiple cancer DNA methylation-based IVD space. While only formed in January 2016, they are very well resourced, and their research and clinical program is expansive. At the 2018 American Society of Clinical Oncology (ASCO) annual meeting, GRAIL presented data from their Circulating Cell-free Genome Atlas (CCGA) study showing that WGBS outperformed whole-genome sequencing (WGS) in identifying cancer in a large population of 1627 prospectively collected blood cfDNA samples (Klein et al., 2018). The data were from 749 controls and 878 participants with newly diagnosed untreated cancer across 20 tumor types and all stages. For eight tumor types, the reported sensitivity across stage I-III cancers was 66% colorectal (n = 28), 63% esophageal (n = 19), 56% head and neck (n = 5), 80% hepatobiliary (n = 5), 59% lung (n = 73), 77% lymphoma (n = 17), 73% multiple myeloma (n = 11), 90% ovarian (n = 10), and 80% for pancreatic (n = 10) tumors. In each instance, specificity was held at 95%.

TRANSLATION

The path to clinical translation is long and expensive. The steps involved after development typically include initial testing in cohorts, then clinical evaluation in clinical trials, followed by manufacture of the test and development of processes for its use and finally review by regulatory authorities. The proposed IVD must offer a multitude of benefits over current practice to attract the significant investment required to translate. In addition to product-market fit, the strength of intellectual property, the robustness, and quality of the clinical evidence to present to payers and the nature of the regulatory landscape in the proposed marketplace are all crucial factors in attracting investment. Medical professionals must also be willing to adopt the test, so the utility and clinical evidence needs to be published in peer reviewed scientific and medical publications and presented at conferences and seminars. This section discusses the establishment of strong IP, and how to produce high quality clinical evidence for regulators, payers, and medical professionals.

Intellectual Property

Patenting in the epigenetics space has sharply risen since around 2000, driven mostly by the patenting of novel diagnostics and epigenetic techniques (Noonan et al., 2013). Patenting provides 20-year exclusivity for companies to exploit ownership of biomarkers and represents a key strategy for biotech and pharma companies to recover costs associated with developing IVDs for clinical utility, for example through licensing fees that would enable labs to implement their test. While important for commercial translation, patent protection can also discourage innovation as it prevents the clinical research community from improving processes to make testing of a biomarker more efficient, for example, quicker testing times, improvements in sensitivity or test accuracy. In some instances, patent protection can generate a monopoly on testing services encouraging excessively high prices out of reach of the general population, e.g. BRCA testing by Myriad. The social and economic implications of biomarker patenting have long been the subject of philosophical debate (Sawyers, 2008; Hopkins and Hogarth, 2012). The patenting of diagnostics has now become far more difficult in the USA after the Association for Molecular Pathology v. Myriad Genetics, Inc court decision in 2013 and the Mayo Collaborative Services v. Prometheus Labs, Inc court decision in 2012 (Dreyfuss et al., 2018). The ruling finds that diagnostics methods based upon biological correlations are not novel but "laws of nature."

In reaction to the court rulings, the biopharma industry has sought new strategies to describe the uniqueness and inventiveness of their intellectual property. If it is easier to demonstrate novelty and human reasoning in new detection method development, then USA-based diagnostics companies can be expected to formulate strategies coupling promising new biomarkers with a novel (and patentable) detection method. Companies outside of the USA may seek to establish IP in the European and/or Asia-Pacific markets where these court decisions do not apply. For smaller companies seeking capital, uncertainty over the validity of Patent Cooperation Treaty (PCT) claims may make it harder to fundraise in the USA until the key patent claims have been interpreted by the United States Patent and Trademark Office once the PCT reaches national phase. Certainty around the worth of an IP portfolio would come only after commentary is received from the patent examiner and this can be a number of years after initial filing. Some legislators are

now seeking to reduce the influence of the earlier court rulings and favor the patentability of biomarkers with the announcement in May 2019 of a bipartisan, bicameral draft bill intending to reform Section 101 of the Patent Act (US Senate, 2019).

Regulation

While groups such as the International Medical Device Regulators Forum (IMDRF; http://www.imdrf.org) with 10 members jurisdictions are working toward global harmonization of regulation around IVDs, in the foreseeable future large regulatory differences will remain, adding significantly to the complexity of translating a diagnostic assay. The prioritization of geographies is important to consider early in the translation process as the appropriate dossiers of evidence need to be tailored for the regulator. An in-depth analysis of the regulatory systems is beyond the scope of this review, but a brief summary and considerations relating to the USA and European markets will follow. Resources are available elsewhere which consider regulation from an international perspective (Theisz, 2015).

In the United States three different regulatory paths exist for obtaining FDA approval of an IVD. The 510(k) regulatory path is for new tests substantially equivalent to an existing predicate test, while tests with no predicate on the market are subject to *de novo* classification for lower risk tests, or premarket approval (PMA) if they are high risk, such as cancer diagnostics. If the test is done "in-house" in a designated laboratory for patient samples ordered by a physician, then the test can be potentially marketed under "home brew" guidelines, known within the USA as LDTs. Clinical laboratories which run LDTs are regulated by CMS through the Clinical Laboratory Improvement Amendments of 1988 (CLIA) Act. CMS can also approve other methods of certification such as from the state licensing schemes or other organizations such as the College of American Pathologists (CAP). The CLIA regulation concerns the standards of the laboratory and the analytical validity (accuracy and precision) of the test *via* a biennial survey and a laboratory may start distributing test results before evaluation. The CMS' CLIA program does not address the clinical validity of any test; which is the accuracy of the test to identify, measure, or predict the presence or absence of a clinical condition or predisposition in a patient. The FDA has signaled it intends to increasingly regulate LDTs due to their increasing complexity (Food and Drug Administration, 2014). The FDA guidance shows an intention to introduce to LDT regulation earlier, more robust verification of analytical validity and a requirement for clinical validity. The FDA is also introducing the concept of high-, medium-, and low-risk LDTs and does not intend to regulate low-risk LDTs, nor tests for unmet needs or rare diseases. Cancer diagnostic tests will be classified as high-risk LDTs, so diagnostics under development now should prepare to demonstrate both analytical and clinical validity, regardless of the choice of the PMA or LDT pathway.

In Europe, the In Vitro Medical Devices Directive (IVDD) 98/79/EC was established in 1998 to harmonize standards of conformity and assessment procedures and to help create a unified pan-European market for IVDs. CE Marking is required for all IVDs sold in Europe. CE Marking indicates that an IVD device complies with the IVDD. Under this legislation, an IVD manufacturer only has to self-declare that the product complies with the essential requirements of relevant European laws. With continued evolution in the IVD marketplace, the European Commission recognized amendments were necessary. Starting from public consultations from 2008 onwards, the new In Vitro Diagnostic Medical Devices Regulation (IVDR) (EU) 2017/746 emerged and the legislation entered into force on 26 May 2017, with a 5-year transition period to full implementation on 26 May 2022 (European Parliament, 2017). There is no grandfathering on presently regulated IVDs, so all existing regulated IVDs need to be CE Marked again.

The IVDR has more alignment with International Organization for Standardization (ISO) guidelines and introduces a risk-based classification system with increased oversight by Notified Bodies. The classes are based on the Global Harmonization Task Force classification scheme (predecessor to the IMDRF) and identifies four risk classes A-D, with Class D the highest risk. IVDs for screening, diagnostics, and staging of cancer are classified as Class C and require a full quality management system. "In-house" tests made and used within a single health institution do not have to comply with the IVDR but they require laboratory compliance with EN ISO 15189 (Medical laboratories, Requirements for quality and competence) and the health institution must justify the use of such a test by demonstrating that no commercially available alternative exists. The IVDR also requires compliance with General Data Protection Regulation (GDPR) for use of samples for regulatory purposes (European Union, 2018). Compliance with this regulation also needs to be considered early in the planning of clinical trials. Compared to the IVDD, the IVDR also has stronger analytical performance requirements for diagnostic tests, including the requirement for reference materials and methods.

Quality Management Systems

Any IVD seeking registration needs to provision a Quality Management System (QMS) and comply with Good Manufacturing Practice (GMP) requirements. This provides the framework for conformity assessment and ongoing post-market responsibilities such as quality control, external quality assurance, and adverse event reporting. Each jurisdiction has different conformity assessment procedures. With global harmonization in mind, the IMDRF began the Medical Device Single Audit Program (MDSAP) initiative in 2012. Regulatory authorities within the working group have implemented a program where auditing organizations can conduct a single audit of a medical device manufacturer that would be accepted by multiple regulators to address QMS and GMP requirements.

For PMA submissions in the USA, the FDA needs to be satisfied that the appropriate design and manufacturing controls are present and has the power to undertake a pre-approval inspection and will schedule a post-approval inspection with 8–12 months of approval. LDTs do not have to comply with FDA quality system regulation, nor be subject to FDA inspection. LDTs (also known as "in-house" IVDs in other jurisdictions) are regulated around the compliance of the laboratory network.

Many countries have made ISO 15189 part of their mandatory medical laboratory accreditation requirements, however in the USA, accreditation to the ISO 15189 standard does not meet CLIA requirements and cannot replace a CLIA-based accreditation (Schneider et al., 2017). Similar to the regulation pathway, priority of jurisdictions for translation should inform the design of the QMS. The existing standards also change over time and new standards are introduced. Relevant to cancer diagnostics, a new ISO standard on the "Requirements for evaluating the performance of quantification methods for nucleic acid target sequences — qPCR and dPCR" (ISO/FDIS 20395) is now under development (International Organization for Standardization, 2019).

When planning translation of a diagnostic test, it is critical to understand what is the appropriate design and evidence for regulatory bodies that adequately supports analytical and clinical validity. The regulators must also be satisfied that this evidence is gathered from the intended target population of the IVD. For guidance on constructing the appropriate dossier of evidence, the Clinical and Laboratory Standards Institute (CLSI), a non-profit organization, produces a set of guidelines relevant to the diagnostics industry. Their "Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures" guideline document is intended for use by IVD manufacturers, regulators, and clinical laboratories to provide guidance for the evaluation and documentation of the detection limits of clinical laboratory measurement procedures (Clinical and Laboratory Standards Institute, 2012).

FUTURE

Presently, there is a flurry of activity in developing DNA methylation-based IVDs. The attention to this sector will only increase with recent announcements by companies such as GRAIL, who found that methylome sequencing of cfDNA outperformed somatic mutation sequencing for primary diagnosis of cancer. An emerging trend is the incorporation of larger panels of methylated biomarkers for multi-cancer detection and determining the tissue of origin. There are now several studies showing that the methylation state of circulating DNA can be used to predict tissue of origin (Kang et al., 2017; Moss et al., 2018), with spin out companies such as EarlyDiagnostics translating these findings. A large plasma cfDNA panel of 9223 CpG sites designed using The Cancer Genome Atlas (TCGA) data has been shown to detect common advanced cancers and underlying cancer type with high accuracy (Liu et al., 2018). Researchers in partner with AnchorDx Medical (Guangzhou, China) have recently shown that a panel of nine bisulfite sequencing amplicons can detect in plasma early stage lung cancer with high sensitivity (Liang et al., 2019).

Another emerging trend in the cancer IVD sector is the development of multi-analyte tests, such as the CancerSEEK test, which combines somatic mutation detection and immunoassays (Cohen et al., 2018). Recently, Guardant Health acquired Bellwether Bio which will allow them to include nucleosome positioning and fragmentomics information with their NGS ctDNA analysis (Snyder et al., 2016). This study of cfDNA fragment length, an indirect measure of nucleosome positioning, has recently been shown to have good clinical utility (Cristiano et al., 2019).

With the continued reduction in the cost of NGS, the use of whole methylomes for biomarker discovery is becoming more commonplace. With sufficient subjects and sequencing depth, all high utility biomarkers will be identified in a screen. Some tests under development, such as the UroMark 150 biomarker assay for bladder cancer detection, are basing the IVD readout on NGS.

There are many new innovations in determining the methylation state of DNA. Two new enzyme-based DNA conversion methods, Enzymatic Methyl-seq (New England Biolabs, 2019) and TET-assisted pyridine borane sequencing (TAPS) (Liu et al., 2019) make use of enzymes to convert the DNA and with these gentler conditions, may offer more recovery of amplifiable DNA than bisulfite-treatment and the resultant DNA is suitable as input for targeted as well as NGS-based approaches. The continued development of third generation sequencing technology such as that from Pacific Biosciences or Oxford Nanopore Technologies offers new opportunities for direct epigenetic detection. To this end, three groups have trained and tested machine-learning approaches to detect methylated DNA on Oxford Nanopore Technologies MinION devices with reasonable classification success (Rand et al., 2017; Simpson et al., 2017; Ni et al., 2019). However, low-input, short cfDNA fragments are not an optimal fit for these long-read platforms. Methylscape, a new method to directly detect and partition methylated DNA using physicochemical properties is also an exciting innovation and offers the potential for an inexpensive pan-cancer test (Sina et al., 2018).

The continued technological development and increasing commercialization activity in the DNA-methylation IVD sector are leading to a fast-paced, innovative, and competitive environment that will result in significant benefits to patients for the early detection and management of cancer.

AUTHOR CONTRIBUTIONS

JR conceptualized the review and led the writing process. JR, WL, DG, CM, YL, KD, and KF performed the literature search and wrote the manuscript. All the authors have read and approved the manuscript.

FUNDING

This work is completely funded by the Commonwealth Scientific and Industrial Research Organisation (CSIRO), the Australian national science agency. CSIRO Health and Biosecurity receives payment from Clinical Genomics and Rhythm Biosciences for translated IVDs. Clinical Genomics and Rhythm Biosciences had no role in review design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Peter Molloy (CSIRO) for his helpful comments on the manuscript.

REFERENCES

Aberg, K. A., Xie, L., Chan, R. F., Zhao, M., Pandey, A. K., Kumar, G., et al. (2015). Evaluation of methyl-binding domain based enrichment approaches revisited. *PloS One* 10 (7), e0132205. doi: 10.1371/journal.pone.0132205

Adler, A., Geiger, S., Keil, A., Bias, H., Schatz, P., deVos, T., et al. (2014). Improving compliance to colorectal cancer screening using blood and stool based tests in patients refusing screening colonoscopy in Germany. *BMC Gastroenterol.* 14, 183. doi: 10.1186/1471-230X-14-183

Ahlquist, D. A., Taylor, W. R., Mahoney, D. W., Zou, H., Domanico, M., Thibodeau, S. N., et al. (2012). The stool DNA test is more accurate than the plasma septin 9 test in detecting colorectal neoplasia. *Clin. Gastroenterol. Hepatol.* 10 (3), 272–277 e271. doi: 10.1016/j.cgh.2011.10.008

Ali, H., Donovan, B., Wand, H., Read, T. R., Regan, D. G., Grulich, A. E., et al. (2013). Genital warts in young Australians five years into national human papillomavirus vaccination programme: national surveillance data. *BMJ* 346, f2032. doi: 10.1136/bmj.f2032

Alvarez, H., Opalinska, J., Zhou, L., Sohal, D., Fazzari, M. J., Yu, Y., et al. (2011). Widespread hypomethylation occurs early and synergizes with gene amplification during esophageal carcinogenesis. *PloS Genet.* 7 (3), e1001356. doi: 10.1371/journal.pgen.1001356

American Cancer Society (2019). American cancer society guidelines for the early detection of cancer [Online]. Available: https://www.cancer.org/healthy/find-cancer-early/cancer-screening-guidelines/american-cancer-society-guidelines-for-the-early-detection-of-cancer.html [Accessed].

Anderson, G. F., and Frogner, B. K. (2008). Health spending in OECD countries: obtaining value per dollar. *Health Aff. (Millwood)* 27 (6), 1718–1727. doi: 10.1377/hlthaff.27.6.1718

Anderson, J. L., Heidenreich, P. A., Barnett, P. G., Creager, M. A., Fonarow, G. C., Gibbons, R. J., et al. (2014). ACC/AHA statement on cost/value methodology in clinical practice guidelines and performance measures: a report of the American College of Cardiology/American Heart Association Task Force on Performance Measures and Task Force on Practice Guidelines. *J. Am. Coll. Cardiol.* 63 (21), 2304–2322. doi: 10.1016/j.jacc.2014.03.016

Bae, J. M. (2015). Value-based medicine: concepts and application. *Epidemiol. Health* 37, e2015014. doi: 10.4178/epih/e2015014

Ball, M. P., Li, J. B., Gao, Y., Lee, J. H., LeProust, E. M., Park, I. H., et al. (2009). Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.* 27 (4), 361–368. doi: 10.1038/nbt.1533

Belinsky, S. A., Nikula, K. J., Palmisano, W. A., Michels, R., Saccomanno, G., Gabrielson, E., et al. (1998). Aberrant methylation of p16(INK4a) is an early event in lung cancer and a potential biomarker for early diagnosis. *Proc. Natl. Acad. Sci. U. S. A.* 95 (20), 11891–11896. doi: 10.1073/pnas.95.20.11891

Bettegowda, C., Sausen, M., Leary, R. J., Kinde, I., Wang, Y., Agrawal, N., et al. (2014). Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* 6 (224), 224ra224. doi: 10.1126/scitranslmed.3007094

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J. M., et al. (2011). High density DNA methylation array with single CpG site resolution. *Genomics* 98 (4), 288–295. doi: 10.1016/j.ygeno.2011.07.007

Brawer, M. K., Chetner, M. P., Beatie, J., Buchner, D. M., Vessella, R. L., and Lange, P. H. (1992). Screening for prostatic carcinoma with prostate specific antigen. *J. Urol.* 147 (3 Pt 2), 841–845. doi: 10.1016/s0022-5347(17)37401-3

Brotherton, J. M., Gertig, D. M., May, C., Chappell, G., and Saville, M. (2016). HPV vaccine impact in Australian women: ready for an HPV-based screening program. *Med. J. Aust.* 204 (5), 184–184e181. doi: 10.5694/mja15.01038

Bryzgunova, O. E., and Laktionov, P. P. (2015). Extracellular nucleic acids in urine: sources, structure, diagnostic potential. *Acta Naturae* 7 (3), 48–54. doi: 10.32607/20758251-2015-7-3-48-54

Capper, D., Jones, D. T. W., Sill, M., Hovestadt, V., Schrimpf, D., Sturm, D., et al. (2018). DNA methylation-based classification of central nervous system tumours. *Nature* 555 (7697), 469–474. doi: 10.1038/nature26000

Carvalho, A. S., Cuco, C. M., Lavareda, C., Miguel, F., Ventura, M., Almeida, S., et al. (2017). Bronchoalveolar lavage proteomics in patients with suspected lung cancer. *Sci. Rep.* 7, 42190. doi: 10.1038/srep42190

Castellanos-Rizaldos, E., Milbury, C. A., Karatza, E., Chen, C. C., Makrigiorgos, G. M., and Merewood, A. (2014). COLD-PCR amplification of bisulfite-converted DNA allows the enrichment and sequencing of rare un-methylated genomic regions. *PloS One* 9 (4), e94103. doi: 10.1371/journal.pone.0094103

Catalona, W. J. (1993). Detection of organ-confined prostate cancer is increased through prostate-specific antigen—based screening. *JAMA: J. Am. Med. Assoc.* 270 (8), 948–954. doi: 10.1001/jama.1993.03510080052031

Cheng, T. H. T., Jiang, P., Tam, J. C. W., Sun, X., Lee, W. S., Yu, S. C. Y., et al. (2017). Genomewide bisulfite sequencing reveals the origin and time-dependent fragmentation of urinary cfDNA. *Clin. Biochem.* 50 (9), 496–501. doi: 10.1016/j.clinbiochem.2017.02.017

Church, T. R., Wandell, M., Lofton-Day, C., Mongin, S. J., Burger, M., Payne, S. R., et al. (2014). Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. *Gut* 63 (2), 317–325. doi: 10.1136/gutjnl-2012-304149

Clark, S. J., Harrison, J., Paul, C. L., and Frommer, M. (1994). High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* 22 (15), 2990–2997. doi: 10.1093/nar/22.15.2990

Clinical and Laboratory Standards Institute (2012). EP17-A2: Evaluation of detection capability for clinical laboratory measurement procedures, 2nd Edition [Online]. Available: https://clsi.org/standards/products/method-evaluation/documents/ep17/ [Accessed].

Cohen, J. D., Javed, A. A., Thoburn, C., Wong, F., Tie, J., Gibbs, P., et al. (2017). Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers. *Proc. Natl. Acad. Sci. U. S. A.* 114 (38), 10202–10207. doi: 10.1073/pnas.1704961114

Cohen, J. D., Li, L., Wang, Y., Thoburn, C., Afsari, B., Danilova, L., et al. (2018). Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* 359 (6378), 926–930. doi: 10.1126/science.aar3247

Conway, P. H. (2009). Value-driven health care: implications for hospitals and hospitalists. *J. Hosp. Med.* 4 (8), 507–511. doi: 10.1002/jhm.535

Cottrell, S. E., Distler, J., Goodman, N. S., Mooney, S. H., Kluth, A., Olek, A., et al. (2004). A real-time PCR assay for DNA-methylation using methylation-specific blockers. *Nucleic Acids Res.* 32 (1), e10. doi: 10.1093/nar/gnh008

Cristiano, S., Leal, A., Phallen, J., Fiksel, J., Adleff, V., Bruhm, D. C., et al. (2019). Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* 570, 385–389. doi: 10.1038/s41586-019-1272-6

D'Andrea, D., Soria, F., Zehetmayer, S., Gust, K. M., Korn, S., Witjes, J. A., et al. (2019). Diagnostic accuracy, clinical utility and influence on decision-making of a methylation urine biomarker test in the surveillance of non-muscle-invasive bladder cancer. *BJU Int.* 123 (6), 959–967. doi: 10.1111/bju.14673

Dagogo-Jack, I., and Shaw, A. T. (2018). Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.* 15 (2), 81–94. doi: 10.1038/nrclinonc.2017.166

Daskalos, A., Nikolaidis, G., Xinarianos, G., Savvari, P., Cassidy, A., Zakopoulou, R., et al. (2009). Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. *Int. J. Cancer* 124 (1), 81–87. doi: 10.1002/ijc.23849

De Meyer, T., Mampaey, E., Vlemmix, M., Denil, S., Trooskens, G., Renard, J. P., et al. (2013). Quality evaluation of methyl binding domain based kits for enrichment DNA-methylation sequencing. *PloS One* 8 (3), e59068. doi: 10.1371/journal.pone.0059068

De Strooper, L. M., Meijer, C. J., Berkhof, J., Hesselink, A. T., Snijders, P. J., Steenbergen, R. D., et al. (2014). Methylation analysis of the FAM19A4 gene in cervical scrapes is highly efficient in detecting cervical carcinomas and advanced CIN2/3 lesions. *Cancer Prev. Res. (Phila)* 7 (12), 1251–1257. doi: 10.1158/1940-6207.CAPR-14-0237

De Strooper, L. M. A., Verhoef, V. M. J., Berkhof, J., Hesselink, A. T., de Bruin, H. M. E., van Kemenade, F. J., et al. (2016). Validation of the FAM19A4/mir124-2 DNA methylation test for both lavage- and brush-based self-samples to detect cervical (pre)cancer in HPV-positive women. *Gynecol. Oncol.* 141 (2), 341–347. doi: 10.1016/j.ygyno.2016.02.012

Deng, J., Shoemaker, R., Xie, B., Gore, A., LeProust, E. M., Antosiewicz-Bourget, J., et al. (2009). Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat. Biotechnol.* 27 (4), 353–360. doi: 10.1038/nbt.1530

Diefenbach, R. J., Lee, J. H., Kefford, R. F., and Rizos, H. (2018). Evaluation of commercial kits for purification of circulating free DNA. *Cancer Genet.* 228-229, 21–27. doi: 10.1016/j.cancergen.2018.08.005

Diep, D., Plongthongkum, N., Gore, A., Fung, H. L., Shoemaker, R., and Zhang, K. (2012). Library-free methylation sequencing with bisulfite padlock probes. *Nat. Methods* 9 (3), 270–272. doi: 10.1038/nmeth.1871

Dietrich, D., Kneip, C., Raji, O., Liloglou, T., Seegebarth, A., Schlegel, T., et al. (2012). Performance evaluation of the DNA methylation biomarker SHOX2 for the aid in diagnosis of lung cancer based on the analysis of bronchial aspirates. *Int. J. Oncol.* 40 (3), 825–832. doi: 10.3892/ijo.2011.1264

Dreyfuss, R. C., Nielsen, J., and Nicol, D. (2018). Patenting nature-a comparative perspective. *J. Law Biosci.* 5 (3), 550–589. doi: 10.1093/jlb/lsy021

Duffy, M. J., van Dalen, A., Haglund, C., Hansson, L., Klapdor, R., Lamerz, R., et al. (2003). Clinical utility of biochemical markers in colorectal cancer. *Eur. J. Cancer* 39 (6), 718–727. doi: 10.1016/s0959-8049(02)00811-0

Eads, C. A., Danenberg, K. D., Kawakami, K., Saltz, L. B., Blake, C., Shibata, D., et al. (2000). MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res.* 28 (8), E32. doi: 10.1093/nar/28.8.e32

Ehrich, M., Nelson, M. R., Stanssens, P., Zabeau, M., Liloglou, T., Xinarianos, G., et al. (2005). Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* 102 (44), 15785–15790. doi: 10.1073/pnas.0507816102

Elazezy, M., and Joosse, S. A. (2018). Techniques of using circulating tumor DNA as a liquid biopsy component in cancer management. *Comput. Struct. Biotechnol. J.* 16, 370–378. doi: 10.1016/j.csbj.2018.10.002

Esteller, M., Corn, P. G., Urena, J. M., Gabrielson, E., Baylin, S. B., and Herman, J. G. (1998). Inactivation of glutathione S-transferase P1 gene by promoter hypermethylation in human neoplasia. *Cancer Res.* 58 (20), 4515–4518.

Esteller, M., Garcia-Foncillas, J., Andion, E., Goodman, S. N., Hidalgo, O. F., Vanaclocha, V., et al. (2000). Inactivation of the DNA-repair gene MGMT and the clinical response of gliomas to alkylating agents. *N. Engl. J. Med.* 343 (19), 1350–1354. doi: 10.1056/NEJM200011093431901

European Parliament (2017). "Regulation (EU) 2017/746 of the European parliament and of the council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU", (ed.) European Union.).

European Union (2018). 2018 reform of EU data protection rules [Online]. European Commission. Available: https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en [Accessed 03/06/2019].

Feber, A., Dhami, P., Dong, L., de Winter, P., Tan, W. S., Martinez-Fernandez, M., et al. (2017). UroMark-a urinary biomarker assay for the detection of bladder cancer. *Clin. Epigenet.* 9, 8. doi: 10.1186/s13148-016-0303-5

Feinberg, A. P., Ohlsson, R., and Henikoff, S. (2006). The epigenetic progenitor origin of human cancer. *Nat. Rev. Genet.* 7 (1), 21–33. doi: 10.1038/nrg1748

Fizazi, K., Greco, F. A., Pavlidis, N., Daugaard, G., Oien, K., Pentheroudakis, G., et al. (2015). Cancers of unknown primary site: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* 26 Suppl 5, v133–v138. doi: 10.1093/annonc/mdv305

Flekna, G., Schneeweiss, W., Smulders, F. J., Wagner, M., and Hein, I. (2007). Real-time PCR method with statistical analysis to compare the potential of DNA isolation methods to remove PCR inhibitors from samples for diagnostic PCR. *Mol. Cell. Probes* 21 (4), 282–287. doi: 10.1016/j.mcp.2007.02.001

Food and Drug Administration (2014). "Draft guidance for industry, food and drug administration staff, and clinical laboratories: framework for regulatory oversight of laboratory developed tests (LDTs)".).

Fraga, M. F., and Esteller, M. (2002). DNA methylation: a profile of methods and applications. *BioTechniques* 33(3) 632, 634, 636–649. doi: 10.2144/02333rv01

Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., et al. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. U. S. A.* 89 (5), 1827–1831. doi: 10.1073/pnas.89.5.1827

Fung, K. Y., Tabor, B., Buckley, M. J., Priebe, I. K., Purins, L., Pompeia, C., et al. (2015). Blood-based protein biomarker panel for the detection of colorectal cancer. *PloS One* 10 (3), e0120425. doi: 10.1371/journal.pone.0120425

Gai, W., and Sun, K. (2019). Epigenetic biomarkers in cell-free DNA and applications in liquid biopsy. *Genes (Basel)* 10 (1), 32. doi: 10.3390/genes10010032

Gallioli, A., Boissier, R., Territo, A., and Breda, A. (2019). Towards the future of upper tract urothelial carcinoma surveillance: lessons learnt from bladder cancer urinary biomarkers. *World J. Urol.* 37 (9), 1985–1986. doi: 10.1007/s00345-018-2611-1

Gaudet, F., Hodgson, J. G., Eden, A., Jackson-Grusby, L., Dausman, J., Gray, J. W., et al. (2003). Induction of tumors in mice by genomic hypomethylation. *Science* 300 (5618), 489–492. doi: 10.1126/science.1083558

Goldstein, M. J., and Mitchell, E. P. (2005). Carcinoembryonic antigen in the staging and follow-up of patients with colorectal cancer. *Cancer Invest.* 23 (4), 338–351. doi: 10.1081/CNV-200058878

Grunau, C., Clark, S. J., and Rosenthal, A. (2001). Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Res.* 29 (13), E65–E65. doi: 10.1093/nar/29.13.e65

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell* 144 (5), 646–674. doi: 10.1016/j.cell.2011.02.013

Hao, X., Luo, H., Krawczyk, M., Wei, W., Wang, W., Wang, J., et al. (2017). DNA methylation markers for diagnosis and prognosis of common cancers. *Proc. Natl. Acad. Sci. U. S. A.* 114 (28), 7414–7419. doi: 10.1073/pnas.1703577114

Hashimoto, K., Kokubun, S., Itoi, E., and Roach, H. I. (2007). Improved quantification of DNA methylation using methylation-sensitive restriction enzymes and real-time PCR. *Epigenetics* 2 (2), 86–91. doi: 10.4161/epi.2.2.4203

Havik, A. B., Brandal, P., Honne, H., Dahlback, H. S., Scheie, D., Hektoen, M., et al. (2012). MGMT promoter methylation in gliomas-assessment by pyrosequencing and quantitative methylation-specific PCR. *J. Transl. Med.* 10, 36. doi: 10.1186/1479-5876-10-36

Hegi, M. E., Diserens, A. C., Gorlia, T., Hamou, M. F., de Tribolet, N., Weller, M., et al. (2005). MGMT gene silencing and benefit from temozolomide in glioblastoma. *N. Engl. J. Med.* 352 (10), 997–1003. doi: 10.1056/NEJMoa043331

Herman, J. G., Graff, J. R., Myohanen, S., Nelkin, B. D., and Baylin, S. B. (1996). Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc. Natl. Acad. Sci. U. S. A.* 93 (18), 9821–9826. doi: 10.1073/pnas.93.18.9821

Herman, J. G., Umar, A., Polyak, K., Graff, J. R., Ahuja, N., Issa, J. P., et al. (1998). Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* 95 (12), 6870–6875. doi: 10.1073/pnas.95.12.6870

Hoadley, K. A., Yau, C., Hinoue, T., Wolf, D. M., Lazar, A. J., Drill, E., et al. (2018). Cell-of-Origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* 173 (2), 291–304 e296. doi: 10.1016/j.cell.2018.03.022

Hopkins, M. M., and Hogarth, S. (2012). Biomarker patents for diagnostics: problem or solution? *Nat. Biotechnol.* 30 (6), 498–500. doi: 10.1038/nbt.2257

Horvath, I., Barnes, P. J., Loukides, S., Sterk, P. J., Hogman, M., Olin, A. C., et al. (2017). A European Respiratory society technical standard: exhaled biomarkers in lung disease. *Eur. Respir. J.* 49 (4), 1600965. doi: 10.1183/13993003.00965-2016

Huang, Y., Pastor, W. A., Shen, Y., Tahiliani, M., Liu, D. R., and Rao, A. (2010). The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PloS One* 5 (1), e8888. doi: 10.1371/journal.pone.0008888

Hulbert, A., Jusue-Torres, I., Stark, A., Chen, C., Rodgers, K., Lee, B., et al. (2017). Early detection of lung cancer ising DNA promoter hypermethylation in plasma and sputum. *Clin. Cancer. Res.* 23 (8), 1998–2005. doi: 10.1158/1078-0432.CCR-16-1371

Imperiale, T. F., Ransohoff, D. F., Itzkowitz, S. H., Levin, T. R., Lavin, P., Lidgard, G. P., et al. (2014). Multitarget stool DNA testing for colorectal-cancer screening. *New Engl. J. Med.* 370 (14), 1287–1297. doi: 10.1056/NEJMoa1311194

International Organization for Standardization (2019). ISO/FDIS 20395. Biotechnology – requirements for evaluating the performance of quantification methods for nucleic acid target sequences – qPCR and dPCR [Online]. Available: https://www.iso.org/standard/67893.html [Accessed 03/06/2019].

Jacobs, I. J., Menon, U., Ryan, A., Gentry-Maharaj, A., Burnell, M., Kalsi, J. K., et al. (2016). Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. *Lancet* 387 (10022), 945–956. doi: 10.1016/s0140-6736(15)01224-6

Jedi, M., Young, G. P., Pedersen, S. K., and Symonds, E. L. (2018). Methylation and gene expression of BCAT1 and IKZF1 in colorectal cancer tissues. *Clin. Med. Insights Oncol.* 12, 1179554918775064. doi: 10.1177/1179554918775064

Jelinek, J., Liang, S., Lu, Y., He, R., Ramagli, L. S., Shpall, E. J., et al. (2012). Conserved DNA methylation patterns in healthy blood cells and extensive changes in leukemia measured by a new quantitative technique. *Epigenetics* 7 (12), 1368–1378. doi: 10.4161/epi.22552

Jelinek, J., and Madzo, J. (2016). DREAM: a simple method for DNA methylation profiling by high-throughput sequencing. *Methods Mol. Biol.* 1465, 111–127. doi: 10.1007/978-1-4939-4011-0_10

Jeong, I. B., Yoon, Y. S., Park, S. Y., Cha, E. J., Na, M. J., Kwon, S. J., et al. (2018). PCDHGA12 methylation biomarker in bronchial washing specimens as an adjunctive diagnostic tool to bronchoscopy in lung cancer. *Oncol. Lett.* 16 (1), 1039–1045. doi: 10.3892/ol.2018.8699

Kang, S., Li, Q., Chen, Q., Zhou, Y., Park, S., Lee, G., et al. (2017). CancerLocator: non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA. *Genome Biol.* 18 (1), 53. doi: 10.1186/s13059-017-1191-5

Klein, E. A., Hubbell, E., Maddala, T., Aravanis, A., Beausang, J. F., and Filippova, D. (2018). Development of a comprehensive cell-free DNA (cfDNA) assay for early detection of multiple tumor types: The Circulating Cell- free Genome Atlas (CCGA) study. *J. Clin. Oncol.* 36 (15), 12021–12021. doi: 10.1200/JCO.2018.36.15_suppl.12021

Koc, A., Goksel, T., Pelit, L., Korba, K., Dizdas, T. N., Baysal, E., et al. (2019). cfDNA in exhaled breath condensate (EBC) and contamination by ambient air: toward volatile biopsies. *J. Breath Res.* 13 (3), 036006. doi: 10.1088/1752-7163/ab17ff

Kocsis, A., Takacs, T., Jeney, C., Schaff, Z., Koiss, R., Jaray, B., et al. (2017). Performance of a new HPV and biomarker assay in the management of hrHPV positive women: subanalysis of the ongoing multicenter TRACE clinical trial (n > 6,000) to evaluate POU4F3 methylation as a potential biomarker of cervical precancer and cancer. *Int. J. Cancer* 140 (5), 1119–1133. doi: 10.1002/ijc.30534

Kondo, Y. (2009). Epigenetic cross-talk between DNA methylation and histone modifications in human cancers. *Yonsei Med. J.* 50 (4), 455–463. doi: 10.3349/ymj.2009.50.4.455

Konstantinidi, E. M., Lappas, A. S., Tzortzi, A. S., and Behrakis, P. K. (2015). Exhaled breath condensate: technical and diagnostic aspects. *ScientificWorldJournal* 2015, 435160. doi: 10.1155/2015/435160

Kulis, M., and Esteller, M. (2010). DNA methylation and cancer. *Adv. Genet.* 70, 27–56. doi: 10.1016/B978-0-12-380866-0.60002-2

Lawn, S. D., Kerkhoff, A. D., Vogt, M., and Wood, R. (2012). Diagnostic accuracy of a low-cost, urine antigen, point-of-care screening assay for HIV-associated pulmonary tuberculosis before antiretroviral therapy: a descriptive study. *Lancet Infect. Dis.* 12 (3), 201–209. doi: 10.1016/s1473-3099(11)70251-1

Leng, S., Wu, G., Klinge, D. M., Thomas, C. L., Casas, E., Picchi, M. A., et al. (2017). Gene methylation biomarkers in sputum as a classifier for lung cancer risk. *Oncotarget* 8 (38), 63978–63985. doi: 10.18632/oncotarget.19255

Liang, W., Zhao, Y., Huang, W., Gao, Y., Xu, W., Tao, J., et al. (2019). Non-invasive diagnosis of early-stage lung cancer using high-throughput targeted DNA methylation sequencing of circulating tumor DNA (ctDNA). *Theranostics* 9 (7), 2056–2070. doi: 10.7150/thno.28119

Lin, S. Y., Linehan, J. A., Wilson, T. G., and Hoon, D. S. B. (2017). Emerging utility of urinary cell-free nucleic acid biomarkers for prostate, bladder, and renal cancers. *Eur. Urol. Focus* 3 (2-3), 265–272. doi: 10.1016/j.euf.2017.03.009

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462 (7271), 315–322. doi: 10.1038/nature08514

Liu, L., Toung, J. M., Jassowicz, A. F., Vijayaraghavan, R., Kang, H., Zhang, R., et al. (2018). Targeted methylation sequencing of plasma cell-free DNA for cancer detection and classification. *Ann. Oncol.* 29 (6), 1445–1453. doi: 10.1093/annonc/mdy119

Liu, Y., Siejka-Zielinska, P., Velikova, G., Bi, Y., Yuan, F., Tomkova, M., et al. (2019). Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nat. Biotechnol.* 37 (4), 424–429. doi: 10.1038/s41587-019-0041-2

Lotan, Y., Shariat, S. F., Schmitz-Drager, B. J., Sanchez-Carbayo, M., Jankevicius, F., Racioppi, M., et al. (2010). Considerations on implementing diagnostic markers into clinical decision making in bladder cancer. *Urol. Oncol.* 28 (4), 441–448. doi: 10.1016/j.urolonc.2009.11.004

Louis, D. N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W. K., et al. (2016). The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol.* 131 (6), 803–820. doi: 10.1007/s00401-016-1545-1

Maier, S., Nimmrich, I., Koenig, T., Eppenberger-Castori, S., Bohlmann, I., Paradiso, A., et al. (2007). DNA-methylation of the homeodomain transcription factor PITX2 reliably predicts risk of distant disease recurrence in tamoxifen-treated, node-negative breast cancer patients–Technical and clinical validation in a multi-centre setting in collaboration with the European Organisation for Research and Treatment of Cancer (EORTC) PathoBiology group. *Eur. J. Cancer* 43 (11), 1679–1686. doi: 10.1016/j.ejca.2007.04.025

Mauger, F., Kernaleguen, M., Lallemand, C., Kristensen, V. N., Deleuze, J. F., and Tost, J. (2018). Enrichment of methylated molecules using enhanced-ice-co-amplification at lower denaturation temperature-PCR (E-ice-COLD-PCR) for the sensitive detection of disease-related hypermethylation. *Epigenomics* 10 (5), 525–537. doi: 10.2217/epi-2017-0166

Meddeb, R., Pisareva, E., and Thierry, A. R. (2019). Guidelines for the preanalytical conditions for analyzing circulating cell-free DNA. *Clin. Chem.* 65 (5), 623–633. doi: 10.1373/clinchem.2018.298323

Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* 33 (18), 5868–5877. doi: 10.1093/nar/gki901

Milbury, C. A., Li, J., Liu, P., and Makrigiorgos, G. M. (2011). COLD-PCR: improving the sensitivity of molecular diagnostics assays. *Expert Rev. Mol. Diagn.* 11 (2), 159–169. doi: 10.1586/erm.10.115

Mitchell, S. M., Ho, T., Brown, G. S., Baker, R. T., Thomas, M. L., McEvoy, A., et al. (2016). Evaluation of methylation biomarkers for detection of circulating tumor DNA and application to colorectal cancer. *Genes (Basel)* 7 (12), 125. doi: 10.3390/genes7120125

Mitchell, S. M., Ross, J. P., Drew, H. R., Ho, T., Brown, G. S., Saunders, N. F. W., et al. (2014). A panel of genes methylated with high frequency in colorectal cancer. *BMC Cancer* 14, 54. doi: 10.1186/1471-2407-14-54

Moran, S., Martinez-Cardus, A., Sayols, S., Musulen, E., Balana, C., Estival-Gonzalez, A., et al. (2016). Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. *Lancet Oncol.* 17 (10), 1386–1395. doi: 10.1016/S1470-2045(16)30297-2

Moss, J., Magenheim, J., Neiman, D., Zemmour, H., Loyfer, N., Korach, A., et al. (2018). Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat. Commun.* 9 (1), 5068. doi: 10.1038/s41467-018-07466-6

Murray, D. H., Symonds, E. L., Young, G. P., Byrne, S., Rabbitt, P., Roy, A., et al. (2018). Relationship between post-surgery detection of methylated circulating tumor DNA with risk of residual disease and recurrence-free survival. *J. Cancer Res. Clin. Oncol.* 144 (9), 1741–1750. doi: 10.1007/s00432-018-2701-x

Nakagawa, H., and Fujita, M. (2018). Whole genome sequencing analysis for cancer genomics and precision medicine. *Cancer Sci.* 109 (3), 513–522. doi: 10.1111/cas.13505

Nechvatal, J. M., Ram, J. L., Basson, M. D., Namprachan, P., Niec, S. R., Badsha, K. Z., et al. (2008). Fecal collection, ambient preservation, and DNA extraction for PCR amplification of bacterial and human markers from human feces. *J. Microbiol. Methods* 72 (2), 124–132. doi: 10.1016/j.mimet.2007.11.007

Ned, R. M., Melillo, S., and Marrone, M. (2011). Fecal DNA testing for colorectal cancer screening: the ColoSure™ test. *PloS Curr.* 3, RRN1220. doi: 10.1371/currents.RRN1220

New England Biolabs (2019). NEBNext® Enzymatic Methyl-seq (EM-seq™) [Online]. Available: https://www.nebiolabs.com.au/-/media/nebus/files/application-notes/nebnext_em-seq_technicalnote.pdf [Accessed 03/06/2019].

Ni, P., Huang, N., Zhang, Z., Wang, D. P., Liang, F., Miao, Y., et al. (2019). DeepSignal: detecting DNA methylation state from Nanopore sequencing reads using deep-learning. *Bioinformatics*. doi: 10.1093/bioinformatics/btz276

Nilsson, M., Malmgren, H., Samiotaki, M., Kwiatkowski, M., Chowdhary, B., and Landegren, U. (1994). Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science* 265 (5181), 2085–2088. doi: 10.1126/science.7522346

Nino, C. L., Perez, G. F., Isaza, N., Gutierrez, M. J., Gomez, J. L., and Nino, G. (2018). Characterization of sex-based DNA methylation signatures in the airways during early life. *Sci. Rep.* 8 (1), 5526. doi: 10.1038/s41598-018-23063-5

Noonan, W., Dismuke, A. D., and Turker, M. S. (2013). Epigenetic patents: a stressful environment for an emerging science. *Biotechnol. Law Rep.* 32 (5), 302–312. doi: 10.1089/blr.2013.9868

Olson, J., Whitney, D. H., Durkee, K., and Shuber, A. P. (2005). DNA stabilization is critical for maximizing performance of fecal DNA-based colorectal cancer tests. *Diagn. Mol. Pathol.* 14 (3), 183–191. doi: 10.1097/01.pas.0000176768.18423.7e

Osborne, J. M., Wilson, C., Moore, V., Gregory, T., Flight, I., and Young, G. P. (2012). Sample preference for colorectal cancer screening tests: blood or stool? *Open J. Prev. Med.* 02 (03), 326–331. doi: 10.4236/ojpm.2012.23047

Oussalah, A., Rischer, S., Bensenane, M., Conroy, G., Filhine-Tresarrieu, P., Debard, R., et al. (2018). Plasma mSEPT9: a novel circulating cell-free DNA-based epigenetic biomarker to diagnose hepatocellular carcinoma. *EBioMedicine* 30, 138–147. doi: 10.1016/j.ebiom.2018.03.029

Pavlidis, N., and Pentheroudakis, G. (2012). Cancer of unknown primary site. *Lancet* 379 (9824), 1428–1435. doi: 10.1016/S0140-6736(11)61178-1

Payne, S. R. (2010). From discovery to the clinic: the novel DNA methylation biomarker (m)SEPT9 for the detection of colorectal cancer in blood. *Epigenomics* 2 (4), 575–585. doi: 10.2217/epi.10.35

Pearce, C. L., Stram, D. O., Ness, R. B., Stram, D. A., Roman, L. D., Templeman, C., et al. (2015). Population distribution of lifetime risk of ovarian cancer in the United States. *Cancer Epidemiol. Biomarkers Prev.* 24 (4), 671–676. doi: 10.1158/1055-9965.EPI-14-1128

Pedersen, S. K., Baker, R. T., McEvoy, A., Murray, D. H., Thomas, M., Molloy, P. L., et al. (2015a). A two-gene blood test for methylated DNA sensitive for colorectal cancer. *PloS One* 10 (4), e0125041. doi: 10.1371/journal.pone.0125041

Pedersen, S. K., Symonds, E. L., Baker, R. T., Murray, D. H., McEvoy, A., Van Doorn, S. C., et al. (2015b). Evaluation of an assay for methylated BCAT1 and IKZF1 in plasma for detection of colorectal neoplasia. *BMC Cancer* 15, 654. doi: 10.1186/s12885-015-1674-2

Pidsley, R., Zotenko, E., Peters, T. J., Lawrence, M. G., Risbridger, G. P., Molloy, P., et al. (2016). Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol.* 17 (1), 208. doi: 10.1186/s13059-016-1066-1

Potter, N. T., Hurban, P., White, M. N., Whitlock, K. D., Lofton-Day, C. E., Tetzner, R., et al. (2014). Validation of a real-time PCR-based qualitative assay for the detection of methylated SEPT9 DNA in human plasma. *Clin. Chem.* 60 (9), 1183–1191. doi: 10.1373/clinchem.2013.221044

Prada, D., Gonzalez, R., Sanchez, L., Castro, C., Fabian, E., and Herrera, L. A. (2012). Satellite 2 demethylation induced by 5-azacytidine is associated with missegregation of chromosomes 1 and 16 in human somatic cells. *Mutat. Res.* 729 (1-2), 100–105. doi: 10.1016/j.mrfmmm.2011.10.007

Rand, A. C., Jain, M., Eizenga, J. M., Musselman-Brown, A., Olsen, H. E., Akeson, M., et al. (2017). Mapping DNA methylation with high-throughput nanopore sequencing. *Nat. Methods* 14 (4), 411–413. doi: 10.1038/nmeth.4189

Rand, K. N., Mitchell, S. M., Clark, S. J., and Molloy, P. L. (2006). Bisulphite differential denaturation PCR for analysis of DNA methylation. *Epigenetics* 1 (2), 94–100. doi: 10.4161/epi.1.2.2680

Rand, K. N., and Molloy, P. L. (2010). Sensitive measurement of unmethylated repeat DNA sequences by end-specific PCR. *BioTechniques* 49 (4), xiii–xvii. doi: 10.2144/000113494

Rand, K. N., Young, G. P., Ho, T., and Molloy, P. L. (2013). Sensitive and selective amplification of methylated DNA sequences using helper-dependent chain reaction in combination with a methylation-dependent restriction enzymes. *Nucleic Acids Res.* 41 (1), e15. doi: 10.1093/nar/gks831

Rassy, E., Kattan, J., and Pavlidis, N. (2019). Familial cancer of unknown primary. *Int. J. Clin. Oncol.* 24 (10), 1328–1331. doi: 10.1007/s10147-019-01491-1

Read, T. R., Hocking, J. S., Chen, M. Y., Donovan, B., Bradshaw, C. S., and Fairley, C. K. (2011). The near disappearance of genital warts in young women 4 years after commencing a national human papillomavirus (HPV) vaccination programme. *Sex Transm. Infect.* 87 (7), 544–547. doi: 10.1136/sextrans-2011-050234

Rivera, M. P., Mehta, A. C., and Wahidi, M. M. (2013). Establishing the diagnosis of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 143 (5 Suppl), e142S–e165S. doi: 10.1378/chest.12-2353

Ronaghi, M. (1998). DNA SEQUENCING: a sequencing method based on real-time pyrophosphate. *Science* 281 (5375), 363–365. doi: 10.1126/science.281.5375.363

Ross, J. P., Rand, K. N., and Molloy, P. L. (2010). Hypomethylation of repeated DNA sequences in cancer. *Epigenomics* 2 (2), 245–269. doi: 10.2217/epi.10.2

Ross, J. P., Shaw, J. M., and Molloy, P. L. (2013). Identification of differentially methylated regions using streptavidin bisulfite ligand methylation enrichment (SuBLiME), a new method to enrich for methylated DNA prior to deep bisulfite genomic sequencing. *Epigenetics* 8 (1), 113–127. doi: 10.4161/epi.23330

Sawyers, C. L. (2008). The cancer biomarker problem. *Nature* 452 (7187), 548–552. doi: 10.1038/nature06913

Schmitz, M., Eichelkraut, K., Schmidt, D., Zeiser, I., Hilal, Z., Tettenborn, Z., et al. (2018). Performance of a DNA methylation marker panel using liquid-based cervical scrapes to detect cervical cancer and its precancerous stages. *BMC Cancer* 18 (1), 1197. doi: 10.1186/s12885-018-5125-8

Schneider, F., Maurer, C., and Friedberg, R. C. (2017). International Organization for Standardization (ISO) 15189. *Ann. Lab. Med.* 37 (5), 365–370. doi: 10.3343/alm.2017.37.5.365

Schroy, P. C. 3rd, and Heeren, T. C. (2005). Patient perceptions of stool-based DNA testing for colorectal cancer screening. *Am. J. Prev. Med.* 28 (2), 208–214. doi: 10.1016/j.amepre.2004.10.008

Siddiqui, I., Majid, H., and Abid, S. (2017). Update on clinical and research application of fecal biomarkers for gastrointestinal diseases. *World J. Gastrointest. Pharmacol. Ther.* 8 (1), 39–46. doi: 10.4292/wjgpt.v8.i1.39

Simpson, J. T., Workman, R. E., Zuzarte, P. C., David, M., Dursi, L. J., and Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods* 14 (4), 407–410. doi: 10.1038/nmeth.4184

Sina, A. A., Carrascosa, L. G., Liang, Z., Grewal, Y. S., Wardiana, A., Shiddiky, M. J. A., et al. (2018). Epigenetically reprogrammed methylation landscape drives the DNA self-assembly and serves as a universal cancer biomarker. *Nat. Commun.* 9 (1), 4915. doi: 10.1038/s41467-018-07214-w

Snyder, M. W., Kircher, M., Hill, A. J., Daza, R. M., and Shendure, J. (2016). Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. *Cell* 164 (1–2), 57–68. doi: 10.1016/j.cell.2015.11.050

Stone, A., Zotenko, E., Locke, W. J., Korbie, D., Millar, E. K., Pidsley, R., et al. (2015). DNA methylation of oestrogen-regulated enhancers defines endocrine sensitivity in breast cancer. *Nat. Commun.* 6, 7758. doi: 10.1038/ncomms8758

Strelecki ene, G., Reid, H. M., Arnold, N., Bauerschlag, D., and Forster, M. (2018). Quantifying cell free DNA in urine: comparison between commercial kits, impact of gender and inter-individual variation. *BioTechniques* 64 (5), 225–230. doi: 10.2144/btn-2018-0003

Stupp, R., Mason, W. P., van den Bent, M. J., Weller, M., Fisher, B., Taphoorn, M. J., et al. (2005). Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N. Engl. J. Med.* 352 (10), 987–996. doi: 10.1056/NEJMoa043330

Su, S. F., de Castro Abreu, A. L., Chihara, Y., Tsai, Y., Andreu-Vieyra, C., Daneshmand, S., et al. (2014). A panel of three markers hyper- and hypomethylated in urine sediments accurately predicts bladder cancer recurrence. *Clin. Cancer. Res.* 20 (7), 1978–1989. doi: 10.1158/1078-0432.CCR-13-2637

Switzeny, O. J., Christmann, M., Renovanz, M., Giese, A., Sommer, C., and Kaina, B. (2016). MGMT promoter methylation determined by HRM in comparison to MSP and pyrosequencing for predicting high-grade glioma response. *Clin. Epigenet.* 8, 49. doi: 10.1186/s13148-016-0204-7

Symonds, E. L., Pedersen, S. K., Baker, R. T., Murray, D. H., Gaur, S., Cole, S. R., et al. (2016). A Blood Test for Methylated BCAT1 and IKZF1 vs. a Fecal Immunochemical Test for Detection of Colorectal Neoplasia. *Clin. Transl. Gastroenterol.* 7, e137. doi: 10.1038/ctg.2015.67

Symonds, E. L., Pedersen, S. K., Murray, D. H., Jedi, M., Byrne, S. E., Rabbitt, P., et al. (2018). Circulating tumour DNA for monitoring colorectal cancer-a prospective cohort study to assess relationship to tissue methylation, cancer characteristics and surgical resection. *Clin. Epigenet.* 10, 63. doi: 10.1186/s13148-018-0500-5

Tao, Y., Kang, B., Petkovich, D. A., Bhandari, Y. R., In, J., Stein-O'Brien, G., et al. (2019). Aging-like spontaneous epigenetic silencing facilitates Wnt activation, stemness, and Braf(V600E)-Induced tumorigenesis. *Cancer Cell* 35 (2), 315–328 e316. doi: 10.1016/j.ccell.2019.01.005

Theisz, V. (2015). *Medical device regulatory practices: an international perspective*. Singapore: Pan Stanford Publishing.

Thomson, J. P., Hunter, J. M., Nestor, C. E., Dunican, D. S., Terranova, R., Moggs, J. G., et al. (2013). Comparative analysis of affinity-based 5-hydroxymethylation enrichment techniques. *Nucleic Acids Res.* 41 (22), e206. doi: 10.1093/nar/gkt1080

Tilki, D., Burger, M., Dalbagni, G., Grossman, H. B., Hakenberg, O. W., Palou, J., et al. (2011). Urine markers for detection and surveillance of non-muscle-invasive bladder cancer. *Eur. Urol.* 60 (3), 484–492. doi: 10.1016/j.eururo.2011.05.053

Turner, N., Biganzoli, L., and Di Leo, A. (2015). Continued value of adjuvant anthracyclines as treatment for early breast cancer. *Lancet Oncol.* 16 (7), e362–e369. doi: 10.1016/s1470-2045(15)00079-0

Ulrich, B. C., and Paweletz, C. P. (2018). Cell-free DNA in oncology: gearing up for clinic. *Ann. Lab. Med.* 38 (1), 1–8. doi: 10.3343/alm.2018.38.1.1

Ulz, P., Thallinger, G. G., Auer, M., Graf, R., Kashofer, K., Jahn, S. W., et al. (2016). Inferring expressed genes by whole-genome sequencing of plasma DNA. *Nat. Genet.* 48 (10), 1273–1278. doi: 10.1038/ng.3648

Um, S. W., Kim, H. K., Kim, Y., Lee, B. B., Kim, D., Han, J., et al. (2017). Bronchial biopsy specimen as a surrogate for DNA methylation analysis in inoperable lung cancer. *Clin. Epigenet.* 9, 131. doi: 10.1186/s13148-017-0432-5

Urban, D., Rao, A., Bressel, M., Lawrence, Y. R., and Mileshkin, L. (2013). Cancer of unknown primary: a population-based analysis of temporal change and socioeconomic disparities. *Br. J. Cancer* 109 (5), 1318–1324. doi: 10.1038/bjc.2013.386

US Senate (2019). Draft bill text to reform section 101 of the patent act [Online]. Available: https://www.tillis.senate.gov/services/files/E8ED2188-DC15-4876-8F51-A03CF4A63E26 [Accessed 03/06/2019].

van Kessel, K. E., Beukers, W., Lurkin, I., Ziel-van der Made, A., van der Keur, K. A., Boormans, J. L., et al. (2017). Validation of a DNA methylation-mutation urine assay to select patients with hematuria for cystoscopy. *J. Urol.* 197 (3 Pt 1), 590–595. doi: 10.1016/j.juro.2016.09.118

van Kessel, K. E., Van Neste, L., Lurkin, I., Zwarthoff, E. C., and Van Criekinge, W. (2016). Evaluation of an epigenetic profile for the detection of bladder cancer in patients with hematuria. *J. Urol.* 195 (3), 601–607. doi: 10.1016/j.juro.2015.08.085

Van Neste, L., Partin, A. W., Stewart, G. D., Epstein, J. I., Harrison, D. J., and Van Criekinge, W. (2016). Risk score predicts high-grade prostate cancer in DNA-methylation positive, histopathologically negative biopsies. *Prostate* 76 (12), 1078–1087. doi: 10.1002/pros.23191

Varadhachary, G. R., and Raber, M. N. (2014). Cancer of unknown primary site. *N. Engl. J. Med.* 371 (8), 757–765. doi: 10.1056/NEJMra1303917

Volkova, M., and Russell, R. (2012). Anthracycline cardiotoxicity: prevalence, pathogenesis and treatment. *Curr. Cardiol. Rev.* 7 (4), 214–220. doi: 10.2174/157340311799960645

Wang, J., Chang, S., Li, G., and Sun, Y. (2017). Application of liquid biopsy in precision medicine: opportunities and challenges. *Front. Med.* 11 (4), 522–527. doi: 10.1007/s11684-017-0526-7

Wasserstrom, A., Frumkin, D., Dotan, Z., Bukin, E., Gadish, T., Hanuka, S., et al. (2016). Mp13-15 Molecular Urine Cytology – Bladder Epicheck Is a Novel Molecular Diagnostic Tool for Monitoring of Bladder Cancer Patients. *J. Urol.* 195 (4S), e140. doi: 10.1016/j.juro.2016.02.2496

Waterhouse, R. L. Jr., Van Neste, L., Moses, K. A., Barnswell, C., Silberstein, J. L., Jalkut, M., et al. (2019). Evaluation of an epigenetic assay for predicting repeat prostate biopsy outcome in African American men. *Urology* 128, 62–65. doi: 10.1016/j.urology.2018.04.001

Weber, M., Davies, J. J., Wittig, D., Oakeley, E. J., Haase, M., Lam, W. L., et al. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.* 37 (8), 853–862. doi: 10.1038/ng1598

Weiss, G., Schlegel, A., Kottwitz, D., Konig, T., and Tetzner, R. (2017). Validation of the SHOX2/PTGER4 DNA methylation marker panel for plasma-based discrimination between patients with malignant and nonmalignant lung disease. *J. Thorac. Oncol.* 12 (1), 77–84. doi: 10.1016/j.jtho.2016.08.123

Wick, W., Weller, M., van den Bent, M., Sanson, M., Weiler, M., von Deimling, A., et al. (2014). MGMT testing–the challenges for biomarker-based glioma treatment. *Nat. Rev. Neurol.* 10 (7), 372–385. doi: 10.1038/nrneurol.2014.100

Witjes, J. A., Morote, J., Cornel, E. B., Gakis, G., van Valenberg, F. J. P., Lozano, F., et al. (2018). Performance of the bladder EpiCheck™ methylation test for patients under surveillance for non–muscle-invasive bladder cancer: results of a multicenter, prospective, blinded clinical trial. *Eur. Urol. Oncol.* 1 (4), 307–313. doi: 10.1016/j.euo.2018.06.011

Wittwer, C. T. (2003). High-resolution genotyping by amplicon melting analysis using LCGreen. *Clin. Chem.* 49 (6), 853–860. doi: 10.1373/49.6.853

Wojdacz, T. K., and Dobrovic, A. (2007). Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. *Nucleic Acids Res.* 35 (6), e41. doi: 10.1093/nar/gkm013

Xiao, P., Chen, J. R., Zhou, F., Lu, C. X., Yang, Q., Tao, G. H., et al. (2014). Methylation of P16 in exhaled breath condensate for diagnosis of non-small cell lung cancer. *Lung Cancer* 83 (1), 56–60. doi: 10.1016/j.lungcan.2013.09.008

Xiong, Z., and Laird, P. W. (1997). COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res.* 25 (12), 2532–2534. doi: 10.1093/nar/25.12.2532

Yang, I. V., Pedersen, B. S., Liu, A. H., O'Connor, G. T., Pillai, D., Kattan, M., et al. (2017a). The nasal methylome and childhood atopic asthma. *J. Allergy Clin. Immunol.* 139 (5), 1478–1488. doi: 10.1016/j.jaci.2016.07.036

Yang, X., Gao, L., and Zhang, S. (2017b). Comparative pan-cancer DNA methylation analysis reveals cancer common and specific patterns. *Brief. Bioinform.* 18 (5), 761–773. doi: 10.1093/bib/bbw063

Young, G. P., Pedersen, S. K., Mansfield, S., Murray, D. H., Baker, R. T., Rabbitt, P., et al. (2016). A cross-sectional study comparing a blood test for methylated BCAT1 and IKZF1 tumor-derived DNA with CEA for detection of recurrent colorectal cancer. *Cancer Med.* 5 (10), 2763–2772. doi: 10.1002/cam4.868

Zhang, C., Yu, W., Wang, L., Zhao, M., Guo, Q., Lv, S., et al. (2017). DNA methylation analysis of the SHOX2 and RASSF1A panel in bronchoalveolar lavage fluid for lung cancer diagnosis. *J. Cancer* 8 (17), 3585–3591. doi: 10.7150/jca.21368

Zhang, J., and Huang, K. (2017). Pan-cancer analysis of frequent DNA co-methylation patterns reveals consistent epigenetic landscape changes in multiple cancers. *BMC Genomics* 18 (Suppl 1), 1045. doi: 10.1186/s12864-016-3259-0

Zou, H., Allawi, H., Cao, X., Domanico, M., Harrington, J., Taylor, W. R., et al. (2012). Quantification of methylated markers with a multiplex methylation-specific technology. *Clin. Chem.* 58 (2), 375–383. doi: 10.1373/clinchem.2011.171264

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Locke, Guanzon, Ma, Liew, Duesing, Fung and Ross. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

REVIEW
published: 18 November 2019
doi: 10.3389/fgene.2019.01125

Epigenetics of Bladder Cancer: Where Biomarkers and Therapeutic Targets Meet

*Victor G. Martinez[1,2], Ester Munera-Maravilla[1,2,3], Alejandra Bernardini[1,2,3], Carolina Rubio[1,2,3], Cristian Suarez-Cabrera[1,2], Cristina Segovia[1,2], Iris Lodewijk[1,2], Marta Dueñas[1,2,3], Mónica Martínez-Fernández[4] and Jesus Maria Paramio[1,2,3]**

[1] Biomedical Research Institute I+12, University Hospital 12 de Octubre, Madrid, Spain, [2] Molecular Oncology Unit, CIEMAT (Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas), Madrid, Spain, [3] Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain, [4] Genomes & Disease Lab, CiMUS (Center for Research in Molecular Medicine and Chronic Diseases), Universidade de Santiago de Compostela, Santiago de Compostela, Spain

Bladder cancer (BC) is the most common neoplasia of the urothelial tract. Due to its high incidence, prevalence, recurrence and mortality, it remains an unsolved clinical and social problem. The treatment of BC is challenging and, although immunotherapies have revealed potential benefit in a percentage of patients, it remains mostly an incurable disease at its advanced state. Epigenetic alterations, including aberrant DNA methylation, altered chromatin remodeling and deregulated expression of non-coding RNAs are common events in BC and can be driver events in BC pathogenesis. Accordingly, these epigenetic alterations are now being used as potential biomarkers for these disorders and are being envisioned as potential therapeutic targets for the future management of BC. In this review, we summarize the recent findings in these emerging and exciting new aspects paving the way for future clinical treatment of this disease.

Keywords: Epigenetic, chromatin remodelling, bladder cancer, biomarkers, therapeutic target

OPEN ACCESS

Edited by:
Yun Liu,
Fudan University, China

Reviewed by:
Stephanie Michelle Willerth,
University of Victoria, Canada
Daniel B. Lipka,
German Cancer Research Center (DKFZ), Germany
Lucio Lara Santos,
Portuguese Institute of Oncology Francisco Gentil, Portugal

***Correspondence:**
Jesus Maria Paramio
jesusm.paramio@ciemat.es

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *19 June 2019*
Accepted: *17 October 2019*
Published: *18 November 2019*

Citation:
Martinez VG, Munera-Maravilla E, Bernardini A, Rubio C, Suarez-Cabrera C, Segovia C, Lodewijk I, Dueñas M, Martínez-Fernández M and Paramio JM (2019) Epigenetics of Bladder Cancer: Where Biomarkers and Therapeutic Targets Meet. Front. Genet. 10:1125. doi: 10.3389/fgene.2019.01125

INTRODUCTION

BC is a common urogenital cancer which represents a current clinical and social problem. At diagnosis, two thirds of patients present a non-muscle invasive bladder cancer (NMIBC), a relatively limited aggressive disease confined to the bladder and without signs of invasion of the underlying muscle layer. The remaining patients display muscle-invasive bladder cancer (MIBC) (Knowles and Hurst, 2015). This pathological classification also defines clinical management. NMIBC is treated by transurethral resection, which can be followed by intravesical instillation with Bacillus Calmette–Guérin (BCG) or mitomycin (Babjuk et al., 2017). However, a large proportion (60–75%) of NMIBC patients relapse and, in some cases (15–25%), the recurrent tumor shows signs of MIBC indicating tumor progression (van Rhijn et al., 2009). The current therapeutic options for MIBC include radical cystectomy and platin-based chemotherapy in adjuvant or neoadjuvant settings (Stenzl et al., 2011). However, in a high proportion of cases, the disease progresses showing metastatic spread, which is associated with extremely low survival rates (Stenzl et al., 2011; Pal et al., 2013; Witjes et al., 2014a). No major improvement in MIBC management occurred during the last decades, until recent years, in which immunotherapy has been shown to increase survival with responses in 20–30% of the patients presenting advanced and metastatic BC (Powles et al., 2014; Rosenberg et al., 2016; Balar et al., 2017; Bellmunt et al., 2017; Plimack et al., 2017). As in other cancers,

immunotherapy in BC is mainly based on the use of antibodies that prevent PD-1/PD-L1 interaction, the so called immune checkpoint, leading to immune killing of tumor cells (Pardoll, 2012). The limited activity of immune checkpoint inhibitors in the clinics has led to the consideration of possible combinations of different immune and non-immune therapies (Gotwals et al., 2017). Moreover, in the case of BC patients, it is unclear which patients are more likely to benefit from this treatment (Powles et al., 2014; Rosenberg et al., 2016; Balar et al., 2017; Bellmunt et al., 2017; Plimack et al., 2017). Thus, there is a need not only for more effective therapies in patients with advanced BC, but also for new biomarkers that will help to define which patients may benefit from immunotherapy (Havel et al., 2019).

Epigenetics are heritable but reversible modifications that alter gene expression without changing primary DNA sequences. Epigenome functions are fundamental for the normal status of gene expression and their alterations affect basic cellular processes such as proliferation, differentiation and apoptosis, which may lead to important diseases including cancer (Liep et al., 2012; Baylin and Jones, 2016). Therefore, epigenetic-based cancer biomarkers are promising tools for detection, diagnosis, assessment of prognosis, and prediction of response to therapy (Esteller, 2008; Jerónimo and Henrique, 2014). An extraordinary number of alterations in epigenetic machinery have been observed in BC, affecting DNA methylation (Marques-Magalhaes et al., 2018), chromatin organization, histone modifications (Weinstein et al., 2014; Robertson et al., 2018) and non-coding RNAs expression (Pop-Bica et al., 2017; Taheri et al., 2018). This has produced a large body of evidence indicating that epigenetic machinery could represent a putative target for BC management, a source of valuable biomarkers for diagnostic, prognostic and response prediction, and also a novel research field with amazing new insights into the molecular mechanisms of cancer biology governing cell autonomous cancer processes as well as the intricate cross talk between cancer cells and their niche.

CHROMATIN REMODELERS IN BC

The epigenome is defined by changes that do not involve alterations in the DNA nucleotide sequence. These changes are broadly divided into DNA methylation and modifications of the histone tails that allow the opening or closing of the chromatin. The functions of the epigenome are fundamental for normal gene expression, and its alterations affect basic cellular processes (Tsai and Baylin, 2011;Liep et al., 2012). The aberrant epigenetic landscape is a hallmark of human cancer (Han et al., 2012; Mio et al., 2019; Zhao et al., 2019) and, in particular, characterizes BC as an epigenome disease, as studies of complete exome sequencing have shown that it presents frequent alterations in the genes that govern the organization of chromatin and histone modifications, either by mutation or by its expression/altered function (Gui et al., 2011; Weinstein et al., 2014).

Nevertheless, the mechanisms for epigenetic regulation of gene expression are not limited to chromatin modifiers or DNA methylation changes, as non-coding RNAs are also involved (Fabbri and Calin, 2010; Gupta et al., 2010; Kogo et al., 2011) (**Figure 1**).

DNA Methylation in BC

Methylation of DNA is the process by which a methyl group is added by a covalent bound to the 5' position of a cytosine ring of the DNA molecule. The methylation event is a frequent epigenetic episode and usually occurs on a cytosine followed by a guanine (CpG dinucleotide). There are regions of the genome, termed CpG islands, which contain a higher density of the CpG dinucleotide than the rest of the genome (Li et al., 2016a). These CpG islands are located in sites that normally overlap with gene regulatory regions (Baylln et al., 1997). Thereupon, there are CpG islands at promoter/5' regions of 50% of all known genes and they are normally unmethylated (Reinert, 2012) which is associated with (potentially) active transcription (Jones and Liang, 2009). CpG islands are also found in gene bodies and their methylation status positively correlates with gene expression (Yang et al., 2014). DNA methylation is a key process in mammalian development, and its alterations are hallmarks of diseases, including cancer. Changes in normal DNA methylation status exist in approximately 50–90% of BCs, including DNA hypermethylation of promoter sites of *A3BP1*, *NPTX2*, *ZIC4*, *PAX5A*, *MGMT*, *IGSF4*, *GDF15*, *SOX11*, *HOXA9*, *MEIS1*, *VIM*, *STK11*, *MSH6*, *BRCA1*, *TBX2*, *TBX3*, *TERT*, *GATA2*, *DAPK1*, *CDH4*, *CCND2*, *GSTP1*, *CDKN2A*, *CDKN2B*, *WIF1*, *RASSF1A*, among others (Porten, 2018). These genes are mainly tumor suppressors that belong to biological pathways such as DNA repair, cell cycle control, cell invasion and apoptosis (Reinert et al., 2011; Sánchez-Carbayo, 2012). DNA methylation of promoter regions typically negatively affects gene expression, which can promote the development (Costa et al., 2010; Chung et al., 2011) and progression of BC (Yates et al., 2007; Kandimalla et al., 2012; Casadevall et al., 2017), and can predict therapy outcomes (Agundez et al., 2011; Xylinas et al., 2016).

First studies of DNA methylation in BC focused on potential genes which methylated status might correlate with stage, grade and recurrence. More recently, the development of modern whole-genome DNA methylation assays has allowed to analyze in depth the BC methylome. Wolff et al. stablished that most DNA methylation changes happen in early BC and are conserved in carcinoma *in situ*, non-invasive as well as invasive tumors, and are located in CpG islands (Wolff et al., 2010). Furthermore, the degree and extent of hypermethylation correlates with grade and stage since low-grade tumors have less altered methylation loci compared to high-grade and invasive tumors (Catto et al., 2005; Yates et al., 2007; Wolff et al., 2010). DNA methylation also separates mutation status of *FGFR3*. NMIBC *FGFR3* wild-type tumors, which have a poorer prognosis compared to *FGFR3* mutant NMIBC (Van Rhijn et al., 2012), were more methylated than *FGFR3*-mutant tumors (Serizawa et al., 2011; Kandimalla et al., 2012). Besides, in low-grade non-invasive tumors, DNA hypomethylation was more frequent than in invasive tumors (Wolff et al., 2010). Hypermethylation of *ZO2*, *MYOD* and *CDH13* was also detected in normal-appearing urothelium from bladder with cancer compared to urothelium from healthy

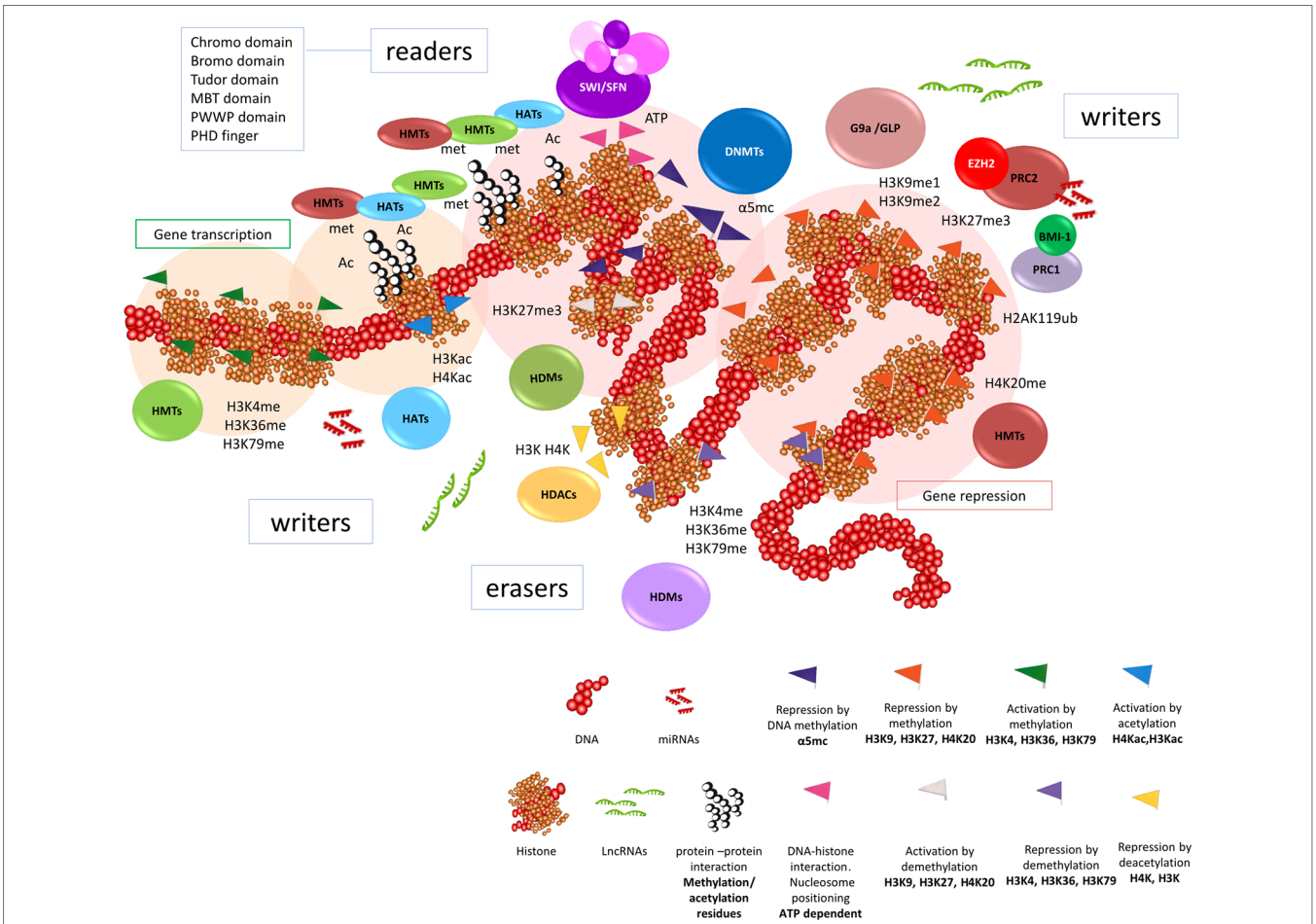


FIGURE 1 | Epigenetic regulation in cancer cells. General scheme of the dynamic interaction of DNA methylation, histone modifications, positioning of nucleosomes, among other factors, that participate in the mechanisms of the epigenome to regulate gene expression. Thus, the tumor cell acquires a particular identity. DNA methylation is present throughout the genome, however we can find aberrant DNA methylations or alterations in the DNMTs enzymes (methyltranferases of DNA nucleotides) in the tumor. The methylation mark H3K27 is the main brand that controls the gene repression in euchromatin. The remodeling enzymes called *writers* (HMT, histone methyltransferase, HAT, histone acetyltransferase), *erasers* (HDM, histone demethylase; HDAC, histone deacetylase) and *readers* (specialized interaction motif containing proteins that recognize post-translational modifications, mostly acetylation and methylation) of the main histone modifications work in a coordinated manner for the regulation of gene transcription. Depending on the genes they regulate, they are recruited to the same place to function together. Therefore, all these molecules are subject of study as possible therapeutic targets.

bladder, indicating an epigenetic 'field defect' and a possible contribution to a loss of epithelial integrity, likely generating a permissive environment for tumor recurrences (Wolff et al., 2010; Majewski et al., 2019).

Since several genes were identified as frequently hypermethylated in primary BC, diagnosis could be performed based on the methylated status of a gene set. For instance, methylation of *IPF1*, *GALR1*, *TAL1*, *PENK* and *TJP2* was found to be higher in MIBC tumors than in NMIBC (Wolff et al., 2010). Sacristan et al. indicated that methylation of *RARB*, *CD44*, *GSTP1*, *IGSF4*, *CHFR*, *PYCARD*, *TP53*, *STK11* and *GATA5* distinguished low-grade versus high-grade tumors, whereas Olkhov-Mitsel et al. stablished that the inclusion of *GP5* and *ZSCAN12* in a methylation panel could feasibly distinguish high-grade and low-grade BC (Olkhov-Mitsel et al., 2017). Unluckily, the overlap between genes found in different studies is limited.

Since 20% of BC patients recur, finding epigenetic markers of progression would be useful to predict recurrence. A wide study reviewed 87 articles reporting the association of epigenetic markers with prognostic outcomes (Casadevall et al., 2017). However, the prognostic influence of epigenetic alterations in BC remains unclear. *CACNA1G* (García-Baquero et al., 2014) and *TBX3* (Kandimalla et al., 2012) were associated with progression and *SFRP5* correlated with recurrence (García-Baquero et al., 2014). *CDNK2A* is methylated in 64% of BCs, however, inconsistent results were found in prognosis (Casadevall et al., 2017). Based on TCGA data, methylation and expression levels of *SOWAHC* were found to be correlated with prognosis (Yang et al., 2019). *HOX* genes appear hypermethylated in almost all aggressive tumors (Reinert et al., 2011; Kandimalla et al., 2012), and *HOXA9* promoter methylation correlated with higher recurrence, progression, and death by cancer in NMIBC and

MIBC (Kitchen et al., 2015) and was associated with cisplatin resistance in BC cell lines (Xylinas et al., 2016). High-risk NMIBC manifest higher rates of progression to invasive tumors than low- and intermediate-risk bladder tumors, which in many cases do not recur or progress. Recently, some investigations proposed multiple CpG sites differentially methylated between high-risk recurrence/progression tumors and less aggressive low-risk no-recurrence tumors (Kitchen et al., 2018; Peng et al., 2018).

A three-gene methylation panel which differentiates between patients with metastatic and free of cancer lymph nodes might also be predictive of metastasis development, and enable the selection of patients that would benefit from lymph node resection and neoadjuvant chemotherapy (Stubendorff et al., 2019). In patients undergoing BCG treatment, methylation status of *MSH6* and *THBS1* may help to distinguish responders to therapy, and methylation of *GATA5* associated with survival (Agundez et al., 2011), allowing the possible identification of patients requiring a more aggressive therapy. After chemotherapeutic treatment, the *MDR1* gene was found to be overexpressed in BC compared with untreated tumors, and in tumors from patients that eventually recurred. This overexpression correlated negatively with methylation of CpG sites in the promoter region (Tada et al., 2000). An interesting study tested gene methylation in second recurrences in bladder of primary upper-tract urothelial carcinomas, and stablished that the methylation rate in certain genes tend to increase with the number of recurrences, which may be a predictive factor for recurrences after surgery (Guan et al., 2018). Nevertheless, the existence of inconsistent results and lack of validation studies hampers at present relevance of these findings (Casadevall et al., 2017; Porten, 2018).

Less data is reported about hypomethylation status in BC. In 1983, a pioneer study reported that hypomethylation could distinguish genes of cancer cells compared with their normal counterparts (Feinberg and Vogelstein, 1983). In normal cells, certain CpG rich satellite repeats are strongly methylated, such as LINE-1 (Schulz, 2006). Interestingly, these regions are strongly hypomethylated in all types of BC (Kreimer et al., 2013) and could translate in genomic instability (Wolff et al., 2010). Besides, as a tissue-fingerprint, the hypomethylation pattern of LINE-1 seems to be specific for each tumor type and tissue (Sharma et al., 2019). Furthermore, a different type of study analyzed global methylation in DNA from blood cells and found that leukocyte DNA hypomethylation is a risk factor for BC (Moore et al., 2008).

DNA methylation is catalyzed by three DNA methyltransferases (DNMT): DNMT1, DNMT3A and DNMT3B. DNMT1 is the keeper of the regular methylation status of the genome after cell replication (Goll and Bestor, 2005), whereas DNMT3a and DNMT3b are *de novo* methyltransferases (Okano et al., 1999). Mutations in chromatin regulatory genes are present in around 76% of BC (Robertson et al., 2018) and are more frequently found in BC than in any other solid tumor (Weinstein et al., 2014). However, the alterations regarding DNMTs in BC are mainly found to be an increase in their expression (Li et al., 2016a). Several genes that are methylated in BC are repressed by polycomb complexes (Wolff et al., 2010; Kandimalla et al., 2012). These complexes composed of EZH2 recruit DNMTs required for DNA methylation (Viré et al., 2006), which suggests an upstream regulation of methylation in BC.

Chromatin Remodeling and Histone Modification in BC

Mutations in chromatin remodeling genes are very frequent in BC (Robertson et al., 2018), affecting 89% of histone remodelers and 64% of nucleosome positioning genes in MIBC (Weinstein et al., 2014; Robertson et al., 2018). The post-translational modifications of histones, such as acetylation, methylation, phosphorylation or ubiquination in specific residues of lysines, arginines and serines (Allis et al., 2007; Rothbart and Strahl, 2014), modulate the dynamic and reversible changes in chromatin structural changes. This "histone code" can be written, erased and read by different molecules modulating transcription (Gillette and Hill, 2015). Therefore, chromatin remodelers can be classified as writers (methyltransferases (HMTs) or acetylases (HATs)), erasers (demethylases (HDMs) and deacetylases (HDACs)) and readers, which are further divided in proteins or effector complexes that interact with specific domains (Gillette and Hill, 2015; Hyun et al., 2017), and nucleosome remodeling multiprotein complexes that are able to alter DNA-histone contacts. The main marks of gene transcription are acetylation of histone 3 and histone 4 (H3Kac, H4Kac) and methylation of histone 3 on lysine 4, 36 and 79 (H3K4me, H3K36me, H3K79me), while methylation of histone 3 on lysine 9 and 27 and histone 4 on lysine 20 (H3K9me, H3K27me, H4K20me) represent important marks for gene repression (Bernstein et al., 2007) (**Figure 1**).

Writers

Histone methyltransferase EZH2 catalyzes H3K27me2 and H3K27me3 marks to regulate the repression of gene expression (Deb et al., 2014), and compacts chromatin with other molecules like BMI-1 (Cao et al., 2005) (**Figure 1**). Its involvement in tumor development and progression is a common characteristic of several human tumors, including BC (Yamaguchi and Hung, 2014). It has been demonstrated that the existence of the oncogenic axis Rb-E2F-EZH2 predicts recurrence and progression in NMIBC (Santos et al., 2014) and promotes global changes in gene expression, including the aberrant expression of lncRNAs such as *HOTAIR* (Martínez-Fernández et al., 2015b), and the silencing of several microRNAs, such as mir-200 family (Martínez-Fernández et al., 2015a). Several studies have shown that EZH2 also interacts with other modifiers such as DNMTs, HDAC or G9a, that could explain some oncological properties of EZH2. The importance of these non-canonical functions of EZH2 in BC is still not well understood, although it could favor intratumoral heterogeneity (Gupta et al., 2011).

Histone methyltransferase G9a (EHMT2) is considered an oncogenic epigenetic factor (Lee et al., 2015), which can be involved in urothelial tumors (Shankar et al., 2013; Cho et al., 2015). This enzyme binds GLP (EHMT1) and catalyzes H3K9me2 leading to gene silencing through physical interaction with cofactors (Bian et al., 2015; Maier et al., 2015; Simon et al., 2015; Hu et al., 2018) and/or non-coding RNAs (Nagano et al., 2008). Additionally, G9a may interact with EZH2 allowing the

silencing of specific loci in a cooperative way (Mozzetta et al., 2014; Mozzetta et al., 2015) becoming a possible target for advanced metastatic BC (Segovia et al., 2019).

Methyltransferase KMT2D (MLL2), which catalyzes H3K4me1 and H3K4me2 (Lee et al., 2013), displays the highest mutation rate among all HMTs in BC (Weinstein et al., 2014) in close association with tumor development, recurrence (Wu et al., 2016a) and resistance to therapy (Lu et al., 2017). *KMT2C* (MLL3) is also commonly mutated in high grade NMIBC (Weinstein et al., 2014; Hurst et al., 2017) and in luminal papillary and basal squamous MIBC subtypes (Robertson et al., 2018), and its silencing affects DNA damage response genes (Rampias et al., 2019). Additionally, somatic mutations of 13 HMT genes, including *NSD1* and *NSD3,* are present in a high proportion of BC tumors (Ding et al., 2019). Moreover, the genes encoding acetyltransferases EP300 and CREBBP are among the genes most frequently inactivated by mutation in human BC (Gui et al., 2011; Duex et al., 2018b).

Erasers

The gene encoding histone demethylase KDM6A (UTX), located on the X chromosome, is one of the genes most frequently mutated in BC (Gui et al., 2011; Nickerson et al., 2014). This demethylase can specifically erase the marks written by EZH2 (Agger et al., 2007; Lee et al., 2007). Mutations in *KDM6A* are more common in NIMBC and in women (Hurst et al., 2017), and tend to be mutually exclusive with *MLL2* alterations (Kim et al., 2015a) suggesting a predominant silenced chromatin during bladder carcinogenesis (Casadevall et al., 2017). In some cases, it has been associated with *RB1* mutation in high grade urothelial tumors (Balbás-Martínez et al., 2013; Ross et al., 2014).

Acetylation of lysine residues in histone tails results in a more open state of the chromatin (Roger et al., 2011) and histone acetylation levels decrease during progression towards MIBC (Ellinger et al., 2016). Furthermore, the deregulated expression of various HDACs, like HDAC1, 2, 3 and 6, has been described in urothelial tumors in close association with malignancy (Chen et al., 2011; Li et al., 2016b; Niegisch et al., 2013; Poyet et al., 2014; Lee and Song, 2017).

Readers

The effects of epigenetic marks are mediated through effector complexes which "read" marks and facilitate the DNA–histone and protein–protein interactions. This provides recruitment platforms for other epigenetic regulators to specific DNA loci (Dawson and Kouzarides, 2012) (**Figure 1**). The methylation and acetylation writers usually have reader domains (predominantly bromodomain (BRD) and plant homeodomain (PHD) finger) that allow recognition of the histone methylation/acetylation status (Dawson and Kouzarides, 2012; Biswas and Rao, 2018).

Methyl CpG sites are recognized by proteins that contain conserved binding domains such as methyl CpG binding domain (MBD), SRA domain and zinc finger (ZnF). These proteins work together with other factors to alter the transcriptional status of DNA (Biswas and Rao, 2018). The histone methylated residues are recognized by conserved binding domains such as PHD finger, Tudor domain, PWWP (Pro-Trp-Trp-Pro) domain, chromodomain, malignant brain tumor domain (MBT), ankyrin repeats (present in G9a and GLP1), ZnFs and WD40 domain, among others. Furthermore, BRDs, double PHD finger and Yeats domains bind specifically to acetylated residues of histones (Dhalluin et al., 1999; Fischle, 2003; Kouzarides, 2007; Taverna et al., 2007; Dawson and Kouzarides, 2012; Biswas and Rao, 2018). BRDs are present in the acetylation writers CBP and p300 along with several protein interaction motifs, both closely related proteins have been deeply investigated since they are able to acetylate the four histones (Dawson and Kouzarides, 2012). Additionally, BRDs of chromatin remodeling enzymes BRM (*SMARCA2*) and BRG1 (*SMARCA4*) recognize multiple acetylation sites at H3 and H4. In BC, the BRD4 histone acetylation reader is overexpressed and can upregulate C-MYC, which controls the expression of cell cycle progression genes, enhancing the recruitment of this factor to the EZH2 promoter and subsequently upregulating EZH2 expression, which has a significant relevance on tumor growth (Wu et al., 2016b). Consequently, EZH2 promotes growth of BC by chromatin modification (Wu et al., 2016b), especially in tumors with loss of *KDM6A* (Ler et al., 2017). Some susceptibilities to EZH2 inhibitors have been found in relation to mutations in components of SWI/SNF complexes such as *ARID1B* (12%), *SMARCA4* (15%) *SMARCA2* (16%) (Helming et al., 2014; Bitler et al., 2015; Kim et al., 2015b). This is relevant in the context of BC, since components of the SWI/SNF complexes are also frequently altered in BC patients (Knowles and Hurst, 2015; Robertson et al., 2018). Other remodelers such as the SWI/SNF nucleosomal complex component, *ARID1A*, often show inactivating mutations or deep eliminations in both MIBC (Weinstein et al., 2014; Robertson et al., 2018) and NIMBC (Hurst et al., 2017).

An additional complexity of chromatin remodeling lies in the fact that many chromatin regulators have more than one type of reader domain, and their binding to chromatin can be further influenced by histone modifications (Ruthenburg et al., 2007). The understanding of the dynamic plasticity of DNA and histone modifications will allow us to open new venues to the management and treatment of BC.

Non-Coding RNAs in BC Etiology and Progression

Non-coding RNAs (ncRNA) represent an important role in the epigenetic changes leading to BC development and progression. Additional to transfer RNA and ribosomal RNA molecules, which represent the most abundant ncRNAs ($3–10 \times 10^7$ and $3–10 \times 10^6$ molecules per cell, respectively), several ncRNA classes can be distinguished, including long non-coding RNA (lncRNA), transcribed ultraconserved region (T-UCR), circular RNA (circRNA), small interfering RNA (siRNA), Y RNA (Y RNA), micro-RNA (miRNA; miR), piwi-interacting RNA (piRNA), small nucleolar RNAs and small nuclear ribonucleic acid (Palazzo and Lee, 2015; Anastasiadou et al., 2017; Gulìa et al., 2017)

NcRNA molecules are specific RNAs which are not translated into proteins, and represent essential regulatory roles in practically every aspect of cellular function. They have been

suggested to exert an essential function in the maintenance of genomic stability, mainly through adjusting DNA expression and complex formation with other ncRNA molecules as well as proteins. Consequently, the description of ncRNA function in isolation is very complicated. Several ncRNAs (like miRNAs) are able to target the messenger RNAs (mRNAs) of multiple other genes, whereas the mRNA of one gene can also be targeted by numerous miRNAs. Furthermore, miRNAs can interact with other ncRNA molecules, like lncRNAs and circRNAs, in order to control their stability, while lncRNAs and circRNAs are able to regulate the abundance of miRNAs. Besides, ncRNAs can interact with individual proteins and protein complexes which might facilitate specific protein targeting or the assembly of protein complexes by providing a scaffold (Anastasiadou et al., 2017; Gulìa et al., 2017).

LncRNAs and miRNAs represent the two main classes of ncRNA involved in BC epigenetic etiology as well as progression, and will be discussed in detail below. Additionally, some other ncRNA molecules associated to this pathology will be briefly described.

Long Non-Coding RNAs

LncRNAs consist of more than 200 nucleotides, and are involved in several essential biochemical processes (Wang and Chang, 2011). Clark et al. examined about 7,200 lncRNA molecules and described a wide variation in stability, ranging from half-lives of less than 30 min for unstable molecules to half-lives of more than 48 h for extremely stable lncRNAs, with a median lncRNA half-life of 3.5 h (Clark et al., 2012). Besides, these lncRNA molecules have been found to be significantly less abundant than, for example, total mRNA ($3–50 \times 10^3$ versus $3–10 \times 10^5$ molecules per cell, respectively) (Palazzo and Lee, 2015). Many lncRNAs were found to be differentially expressed in a wide range of tumor tissues compared to corresponding healthy control tissues, suggesting an important role in carcinogenesis (Martens-Uzunova et al., 2014; Bhan et al., 2017). In BC, deregulation of lncRNAs has been found to contribute to carcinogenesis in several ways including sustained proliferative signaling and induction of invasion as well as metastasis (Bhan et al., 2017).

LncRNA expression in BC has been extensively reviewed (Gulìa et al., 2017; Taheri et al., 2018). Based on their expression patterns and functions in BC tissue compared to healthy control tissue, lncRNA molecules can be classified in two groups, either showing increased (oncogenic lncRNAs) or decreased (tumor suppressor lncRNAs) expression in tumor tissue. For example, oncogenic lncRNA-*UCA1* has been reported to induce epithelial-mesenchymal transition (EMT) and promote BC cell migration and invasion through the miR-145–ZEB1/2–FSCN1 pathway, as well as by targeting miR-582-5p or modulation of the miR-143/HMGBG1 signaling pathway (Xue et al., 2016; Luo et al., 2017; Wu et al., 2019a). Overexpression of *UCA1* has been associated with high risk of poor outcome in BC. Accordingly, the use of *UCA1* as potential biomarker is subject of ongoing research (Wang et al., 2006; Cui et al., 2017). LncRNA-*H19* has been found to be abundantly expressed in BC leading to increased miR-675 expression, thus inhibiting TP53 activation (Ariel et al., 2000; Liu et al., 2016). LncRNA-*H19* has further been described to promote metastasis and EMT through E-cadherin inhibition as well as by targeting miR-29b-3p (Lv et al., 2017; Zhu et al., 2018). Other well-described oncogenic lncRNAs involved in BC include *MALAT1*, *HOTAIR*, *TUG1*, *ANRIL* and *PVT1*, whereas well-known lncRNAs-*MEG3* and *GAS5* represent tumor suppressor lncRNA molecules (Sun et al., 2015; Gulìa et al., 2017; Guo et al., 2018; Liu et al., 2017b; Xie et al., 2017a; Yang et al., 2017; Yu et al., 2019; Jiao et al., 2018; Liu et al., 2018a; Wang et al., 2018b; Huang et al., 2019a; Tian et al., 2019). Even though many other oncogenic and tumor suppressor lncRNAs have recently been identified in BC, they need further investigation to validate their relevance in this disease. Additionally, the use of specific lncRNA as biomarkers or therapeutic targets is subject of ongoing research and will be further discussed below.

Micro-RNAs

As abovementioned, lncRNAs extensively interact with miRNA in the regulation of oncogenic pathways. MiRNAs consist of 21–24 nucleotides and play important roles in the regulation of gene expression (Sohel, 2016). Mature miRNAs have shown high stability reflecting half-lives of approximately 8 hours in the cell, which is reflected in a relatively high abundance of miRNA molecules ($1–3 \times 10^5$ molecules per cell) (Palazzo and Lee, 2015).

Aberrantly expressed miRNAs have been found in BC tissues causing an altered expression of target genes, resulting in BC development and progression (Zhu et al., 2011). As for lncRNAs, miRNA expression in BC has been extensively reviewed (Enokida et al., 2016; Gulìa et al., 2017). The aberrant expression of several miRNAs has been found to alter two main genetic pathways predisposing to BC. Some miRNAs target the FGFR3 pathway (including miR-99a, miR-100, miR-101, and miR-145), while other miRNA molecules modify the TP53 pathway (such as miR-21 and miR-373) (Homami and Ghazi, 2016). Like lncRNAs, miRNA molecules can be divided in oncogenic miRNAs or tumor suppressor miRNAs. For example, the decreased expression of miR-34a in BC has an anti-metastatic function through the CD44/EMT signaling pathway (Yu et al., 2014) and through targeting NOTCH1 and HNF4G also negatively modulates BC cell proliferation and invasion (Zhang et al., 2012; Sun et al., 2015). Accordingly, low expression of miR-34 has been found to be correlated with unfavorable prognosis (Xie et al., 2017b). Besides, downregulation of the tumor suppressor miR-200 family has been proposed to be associated with poor prognosis in BC, and the use of this family as prognostic marker has been indicated (Wiklund et al., 2011; Martínez-Fernández et al., 2015a). The miR-200 family consists of five different members, namely miR-200a, miR-200b, miR-200c, miR-429 and miR-141, and has been suggested to play an essential role in the inhibition of the EMT process by regulation of ZEB1 and ZEB2 transcription factors (Korpal et al., 2008; Park et al., 2008).

Many other tumor suppressor and oncogenic miRNAs have been extensively described or recently discovered as particular players in BC (Enokida et al., 2016; Gulìa et al., 2017). For example, low expression of miR-100, miR-101 and miR-214, as well as high expression of miR-452, miR-21, miR-222, miR-182, miR-133b, miR-155, miR-145, and miR-152 has been correlated with unfavorable prognosis(Xie et al., 2017b).

Given their deregulated expression, the miRNAs have been widely studied as therapeutic target and biomarkers in different pathologies, including several types of cancers (Romero-Cordoba et al., 2014; Shah and Calin, 2014; Chan et al., 2015). Accordingly, the study of miRNAs in liquid biopsy offers great perspective for diagnostic and prognostic purposes. These objectives will be further discussed below.

Other ncRNA

CircRNA molecules represent a type of ncRNA that are covalently closed in a loop at the 3′ and 5′ ends. The lack of free 3′ or 5′ ends provides increased resistance of circRNAs to exoribonuclease-dependent RNA degradation, which results in a prolonged half-life of over 48 h (Jeck and Sharpless, 2014). Even though their cellular functions are still largely unknown, various circRNAs have shown relevance in multiple cancer types (Zhang et al., 2017; Kristensen et al., 2018). Although circRNA research in BC is still scarce, several circRNAs have been shown to be highly expressed in human BC. These endogenous circRNAs competitively target specific miRNAs, thereby suppressing miRNA activity by acting as a miRNA sponge. For example, circTCF25 has been demonstrated to promote cell proliferation and metastasis by acting as a RNA sponge for miR-103a-3p and miR-107, resulting in increased CDK6 levels (Zhong et al., 2016). Besides, circRNA-MYLK and circRNA-CTDP1 competitively bind miR-29a-3p leading to enhanced expression of its target genes *DNMT3B*, *VEGFA*, *HAS3* and *ITGB1*, resulting in angiogenesis, EMT and metastasis (Huang et al., 2016; Zhong et al., 2017). Recently, additional circRNA molecules representing an oncogenic role in BC tumorigenesis and progression have been discovered, including circCEP128, circRNA-VANGL1, circPRMT5 and circRNA-cTFRC (Chen et al., 2018; Wu et al., 2018; Zeng et al., 2019; Su et al., 2019).

Contrarily to the oncogenic role of several circRNAs, some circRNAs act as tumor suppressors and have been shown to be downregulated in human BC. For example, circRNA-ITCH has been shown to suppress the aggressive biological behavior of BC through increased expression of p21 and PTEN by sponging miR-17 and miR-224, whereas circRNA-BCRC-3 has been found to act as a sponge of miR-182-5p resulting in enhanced expression of p27 (Xie et al., 2018; Yang et al., 2018a). Other circRNA molecules which have recently been discovered to mediate anti-oncogenic functions include circRNA-BCRC4, circRNA-Cdr1as and circMTO1 (Li et al., 2017; Li et al., 2018a; Liu et al., 2018b).

Their extensive abundance, stability and tissue-specific expression make circRNAs attractive molecules for clinical research (Barrett and Salzman, 2016). Further research into their regulatory mechanisms on miRNA expression will help us to improve our knowledge regarding their function in carcinogenesis and may provide insights in the use of circRNA molecules as predictive and diagnostic biomarkers as well as novel therapeutic targets (Kulcheski et al., 2016; Han et al., 2017).

Y RNA molecules are small ncRNAs (21–24 nucleotides) necessary for DNA replication through interactions with chromatin and initiation proteins. Four Y RNAs have been identified and found to be highly evolutionary conserved, namely Y1, Y3, Y4 and Y5 (Christov et al., 2006). These ncRNAs are protected from degradation by its interaction with Ro, a ribonucleoprotein particle that provides stability to these molecules, and their abundance has been found to be relatively high (about 1×10^5 molecules per cell) (Christov et al., 2006; Chen et al., 2007). Even though a role for Y RNAs in BC has been indicated by various studies, contradicting observations have been published (Christov et al., 2008; Tolkach et al., 2017). Christov et al. described the significant overexpression of two Y RNAs, Y1 and Y3, whereas Tolkach et al. published the significant downregulation of all four Y RNAs in BC tissue compared to tissue of healthy controls. Accordingly, this emphasizes the need for further studies to clarify the possible role of Y RNA in BC etiology and progression.

PiRNA molecules are short single strands non-coding RNAs (26–31 nucleotides) mediating epigenetic and post-transcriptional gene silencing through interactions with PIWI proteins (Siomi et al., 2011). Their small size suggests particular resistance to degradation, which can result in the presence of relatively high levels of piRNA molecules (Palazzo and Lee, 2015; Pardini and Naccarati, 2017). Deregulated expression of some piRNAs has been found in different cancer types (Chalbatani et al., 2019). In BC, Martinez et al. described the association of high levels of piRNA FR004819 with poorer survival, whereas Taubert et al. defined a significant association between diminished *PIWIL2* expression and poor prognosis (Martinez et al., 2015; Taubert et al., 2015). Additionally, piRABC has been observed to be downregulated in BC tissue and has been identified as an important piRNA in the development and progression of this pathology. Besides, it has been proposed that piRABC may promote cell apoptosis in BC by upregulation of the TNFSF4 protein (Chu et al., 2015; Chalbatani et al., 2019).

EPIGENETIC REGULATION OF THE BC MICROENVIRONMENT

Immune Cell Compartment

Cancer initiation and tumor progression are often associated with the inhibition of anticancer immune response and dysregulation of inflammatory activity (Berraondo et al., 2016; Sukari et al., 2016). Different solid tumors are characterized by the presence of immune cells, such as T and B lymphocytes, natural killer (NK) cells, macrophages, and antigen-presenting cells in the tissue microenvironment (TME). These immune cells exhibit different behaviors and morphologies as a result of aberrant differentiation (Olivieri et al., 2016), sometimes driven by epigenetically regulated lineage-specific changes influencing the expression of genes crucial for the identity of immune cells and promoting cellular responses to stimuli (Herold et al., 2012; Smith and Meissner, 2013; Luperchio et al., 2014) (**Figure 2**).

Recently, some studies have shown that post-translational modification of histones may regulate the behavior of cells involved in the immune response, including tumor associated macrophages (TAMs), regulatory T cells (Tregs), dendritic cells (DCs), NK cells, myeloid-derived suppressor cells (MDSCs), effector T cells (Teffs), and others (Liu et al., 2017a). Based on whole-genome bisulfite sequencing datasets from

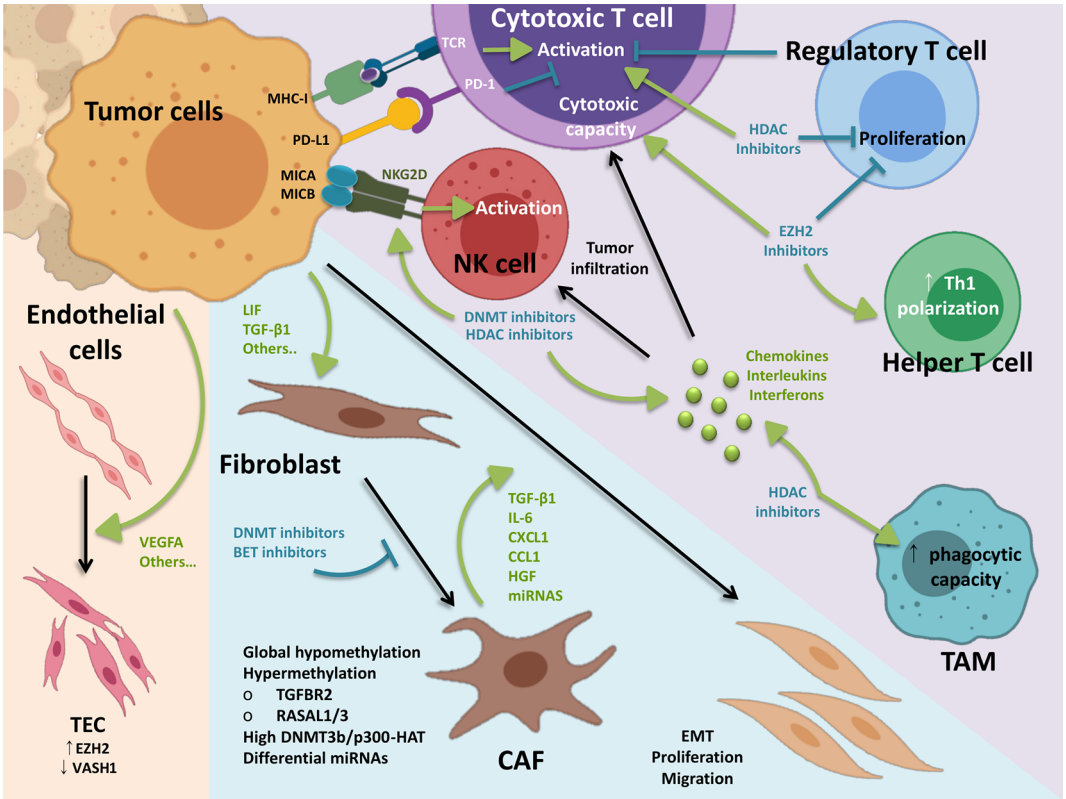


FIGURE 2 | Epigenetic landscape of the tumor microenvironment. Tumor cells can influence the stroma through different factors, being soluble factors the most characterized. Tumor-derived VEGFA induces EZH2 in TEC, which drives hypermethylation of anti-angiogenic Vash1. Also induced by tumor cells, CAF differentiation is associated with several epigenetic features and can be blocked by a number of chromatin remodelers inhibitors. In turn, CAFs promote tumor growth and metastasis *via* secretion of soluble factors and matrix remodeling. On the immune side, cytotoxic T cells and natural killer cells are the main effectors of the anti-cancer immune response. Balance between activating and inhibiting signals coming from tumor targeted cells determines cytotoxic activity of these cells. Other immune cells such as regulatory T cells and macrophages are key in the anti-cancer immune response. Of note, myeloid and lymphoid lineages present inverse methylation patterns in cancer tissues, contributing to aberrant functionality. Inhibition of epigenetic writers can block regulatory T cell differentiation and function, while promoting anti-tumor activity in effector cells. Reverting tumor-driven epigenetic modifications imprinted in the TME may condition the tumor stroma for effective elimination of malignant cells in combination with existing treatments such as immunotherapy. TEC, tumor endothelial cells; CAF, cancer-associated fibroblasts; EMT, epithelial–mesechymal transition; TAM, tumor-associated macrophage.

the BLUEPRINT Epigenome Project (http://www.blueprint-epigenome.eu), Schuyler et al. identified inverse methylation patterns in the myeloid and lymphoid lineages in cancer tissues, where lymphoid-derived neoplasms lose CpG methylation patterns whereas myeloid malignancies significantly increase levels of DNA methylation (Schuyler et al., 2016). These observations have been reproduced by other authors showing that different methylation patterns contribute to the activation of myeloid and lymphoid cancer cells (Bröske et al., 2009; Bock et al., 2012).

The main component of the immune infiltrates present in solid tumors are TAMs, which have been frequently associated with worse prognosis. Compared to the binary M1/M2 classification, TAMs include multiple populations sharing features of both M1 and M2 phenotypes that in many cases do not fit the M1/M2 classification. Nonetheless, it offers a useful working frame for the study of TAMs, in which the overall consensus is that M1 macrophages are anti-tumorigenic, while M2 macrophages can promote tumor growth. M2-macrophage marker genes are epigenetically regulated by reciprocal changes in histone H3 lysine-4 (H3K4) and histone H3 lysine-27 (H3K27) methylation. After IL-4 stimulation, a decrease of H3K27 dimethylation and trimethylation (H3K27me2/3) marks occur as well as the transcriptional activation of specific M2 marker genes. Additionally to methylation, during monocyte to macrophage differentiation, there is a massive reconfiguration of lysine acetylation patterns at gene regulatory elements with a positive correlation between transcriptionally permissive H3 histone acetylation and the activity of regulatory elements (Bistoni et al., 1986).

When analyzing the activation/polarization status of tumor infiltrating lymphocytes (TILs), TAMs and DCs, several studies have shown that the methylation status of immune genes in these cells influences the tumor immune response in the TME, and correlates with the density of TILs and tumor progression. For example, in naïve $CD4^+$ T cells the interferon-γ (*IFN-γ*) gene promoter and upstream enhancer is methylated. However, in Th1 lymphocytes, where the expression of *IFN-γ* is induced,

the *IFN-γ* gene promoter and enhancer are demethylated, suggesting an important role in Th1/Th2 differentiation (Janson et al., 2008). The histone methyl transferase EZH2 has also been shown to play an important role in shaping the function of T cells. Wang et al. demonstrated that accumulation of H3K4me3 in the promoter of *FOXP3* results in the generation of Tregs, and pharmacological or genetic suppression of the activity of EZH2 on tumor-infiltrating Tregs (TI-Tregs) results in the acquisition of pro-inflammatory functions (Wang et al., 2018a) (**Figure 2**). In addition, suppression of EZH2 modulates the TME and enhances the infiltration of $CD8^+$ and $CD4^+$ effector T cells, which can favor tumor eradication (Wang et al., 2018a). Besides H3K27 methylation, G9a-dependent H3K9me2 is an important regulator of inflammatory gene expression and has also been implicated in several aspects of T cell biology. Although genome-wide studies mapping the binding of G9a (or the H3K9me2 mark) in immune cells has not been carried out, a descriptive genome-wide analysis of H3K9me2 marks in resting human lymphocytes using ChIP-on-chip methods demonstrated that this epigenetic mark is enriched on genes that are associated with several specific pathways including T cell receptor signaling, IL-4 signaling, and GATA3 transcription (Zhang et al., 2018a).

In addition to T and B lymphocytes, NK cells are effector lymphocytes of the innate immune system that have been shown to control tumor growth (Vivier et al., 2008). Although studies investigating the role of epigenetic modulation on NK cell activation and cytotoxicity are still scarce, some reports indicate that histone acetylation is involved in the regulation of NK cell activation and effector functions (Schenk et al., 2016; Raulet et al., 2017). Particularly in cancer, HDAC inhibitors have been shown to modulate the expression of NK ligands on the surface of neuroblastoma, melanoma, osteosarcoma, colon and Merkel cell (Zhu et al., 2015; Kiany et al., 2017) (**Figure 2**). Besides, Hicks et al. shows that HDAC inhibitors, in addition to significantly enhancing the expression of multiple NK ligands and death receptors resulting in enhanced NK cell-mediated lysis, also increases tumor cell PD-L1 expression both *in vitro* and in carcinoma xenografts (Hicks et al., 2018). This data offers a rationale for combining HDAC inhibitors with inhibitors of the PD-1/PD-L1 axis, including for patients who are refractory or expected not to respond to these therapies alone due to absent or low PD-L1 tumor expression.

Cancer-Associated Fibroblasts

The tumor stroma is defined as the non-malignant cells and extracellular components that surround tumors, with a fundamental role in growth and progression. Fibroblasts in the tumor microenvironment differentiate into cancer-associated fibroblasts (CAFs), being one of the main components in the tumor stroma (**Figure 2**). CAFs play key roles in all cancerous stages, the vast majority of the studies demonstrating pro-tumoral functions that include extracellular matrix remodeling, angiogenesis, immune suppression and drug resistance (Kalluri, 2016; Tao et al., 2017; Ziani et al., 2018).

The current knowledge on CAF biology in BC is scarce and mostly coming from *in vitro* experiments. Nonetheless, it has been shown that there is a positive correlation between the presence of active CAFs and expression of EMT markers and worse prognosis in BC patients (Schulte et al., 2012; Wu et al., 2017). *In vitro*, BC cells can induce differentiation of healthy fibroblast into CAFs *via* exosomes (Ringuette Goulet et al., 2018; De Palma et al., 2019; Goulet et al., 2019) and other not fully characterized secreted factors (Wang et al., 2007; Grimm et al., 2015; Shi et al., 2015; Yeh et al., 2015). As a result, differentiated CAFs induce motility and migration in cancer cells *via* induction of EMT through secretion of a number of soluble factors which include TGF-β1 (Zhuang et al., 2015; Wu et al., 2017), IL-6 (Yeh et al., 2015; Goulet et al., 2019), and hepatocyte growth factor (HGF) (Wang et al., 2007; Grimm et al., 2015), and/or by direct chemokine attraction through CXCL1 (Shi et al., 2015) and CCL1 (Yeh et al., 2015).

Studies using global methylation analysis have shown that epigenetic modification plays a fundamental role in fibroblast activation and CAF differentiation (Hu et al., 2005; Jiang et al., 2008; Bechtel et al., 2010; Lamprecht et al., 2018). Indeed, an overall hypomethylated status was found in human CAFs (Jiang et al., 2008; Eckert et al., 2019) (**Figure 2**), as well as in functionally related fibrotic fibroblasts (Komatsu et al., 2012). Nevertheless, certain key genes appear hypermethylated in CAFs such as *Tgfbr2* (Banerjee et al., 2014), *RASAL1* and others (Bechtel et al., 2010; Zeisberg and Zeisberg, 2013; Mishra et al., 2018). Seminal work by Cedric Gaggioli's group demonstrated that tumor-derived LIF induces activation of DNMT3b and p300-HAT in CAFs, which sustain JAK1/STAT3 signaling, necessary to maintain a pro-invasive activity (Albrengues et al., 2015). More recently, the nicotinamide N-methyltransferase has been shown as fundamental for CAF′s protumoral behavior *in vitro* and *in vivo*, directly affecting DNA and histone methylation (Eckert et al., 2019). CAF differentiation and activity *in vivo* can be blocked by treating with the DNMT inhibitor 5′-Aza-2′-deoxycytidine (Albrengues et al., 2015; Eckert et al., 2019), acting specifically in pancreatic CAFs compared to normal fibroblasts (Yu et al., 2012). Relevant results when DNMT inhibitors are considered for therapy, which will be further discussed below.

Interestingly, the RasGTP *RASAL3*, negative regulator of the Ras signaling pathway, was also found hypermethylated in prostate cancer (PCa) CAFs (Mishra et al., 2018), increasing Ras signaling in these cells, which drives support of tumor growth and neuroendocrine differentiation. Noteworthy, switch in CAFs towards a Warburg metabolism has been implicated in tumor immune evasion in PCa (Comito et al., 2019), which adds further clinical relevance of epigenetic-mediated changes in CAFs metabolism. Indeed, an *in vitro* 3D-microfluidoc system has shown that CAFs provide metabolic support to proliferation and invasion of BC cells (Shi et al., 2015). The role of epigenetic modifications in this phenomenon and its relevance *in vivo* will require further investigation.

Besides DNA methylation, other epigenetic modifications have been observed in the tumor stroma (Li et al., 2015; Du and Che, 2017; Schoepp et al., 2017; Vafaee et al., 2017; Zhao et al., 2017; Kim et al., 2018). In a proof-of-concept study, Zong et al. showed that overexpression of the non-histone chromosomal high-mobility group protein family member Hmga2 in urogenital sinus mesenchymal cells drives tumorigenesis in a model for

prostatic intraepithelial neoplasia (Zong et al., 2012). In models for pancreatic cancer and *in situ* skin squamous cell carcinomas, an inhibitor of the BRD and extraterminal domain (BET) family proteins decreases tumor growth affecting specifically CAF´s secretome (Yamamoto et al., 2016; Kim et al., 2017). Since targeting histone acetylation has been proposed for combined therapy in BC (Yoon et al., 2011), it would be necessary to characterize the histone acetylation status of stromal cells in BC patients.

Many studies show that miRNAs play fundamental roles in CAF differentiation and function, a subject that has been extensively reviewed (Chou et al., 2013; Kohlhapp et al., 2015; Kuninty et al., 2016; Marks et al., 2016). MiRNAs can be expressed by CAFs or incorporated from other sources, mainly cancer cells *via* exosomes (Pang et al., 2015). The opposite is also possible, when CAFs modulate cancer cell behavior *via* transfer of miRNAs (Josson et al., 2015; Shah et al., 2015). In BC, a study compared miRNA expression between fibroblasts from healthy and tumoral human bladder, finding higher expression of miR-16 and miR-320 (Enkelmann et al., 2011). Which functions are these miRNAs regulating in CAFs and whether they can be used as surrogate markers for stroma abundance would require further investigations.

Tumor Endothelial Cells

In solid cancers, increased *de novo* formation of blood vasculature, known as angiogenesis, is normally observed and provides adequate nourishment for the growing tumor (**Figure 2**). The link between vasculature density and worse prognosis in BC is well documented (Bochner et al., 1995). Indeed, targeting angiogenesis *via* disruption of vascular endothelial growth factor (VEGF) signaling is being considered for treating BC in combination with existing therapies (Petrylak et al., 2016; Sonpavde and Bellmunt, 2016).

Tumor endothelial cells (TECs) display a number of characteristics compared to normal endothelium (Hashizume et al., 2000; Hida et al., 2004). In BC, exacerbated proliferation and sprouting of TECs has been linked to staging and lower survival in patients (Roudnicky et al., 2013; Roudnicky et al., 2017). Invasive BC cell lines show increased adhesion to endothelial cells *via* MUC1 and CD43 binding to ICAM-1, which could be linked to metastatic potential (Laurent et al., 2014; Sundar Rajan et al., 2017). Besides, an *in vitro* study shows that TECs may promote BC cell growth through a paracrine loop involving secretion of epidermal growth factor by TECs in response to tumor-derived VEGFs (Huang et al., 2019b). Finally, TECs have been found in MIBC with aberrant expression of a non-anti-angiogenic thrombospondin-2 variant, also responsible for uncontrolled angiogenesis in these tumors (Roudnicky et al., 2018).

It is well known that epigenetic modifications play a role in endothelial cell (ECs) proliferation, differentiation and pathogenesis (Hulshoff et al., 2018; Nagai et al., 2018; Schlereth et al., 2018; Stone et al., 2018; Nicorescu et al., 2019). In fact, recent work by Wang S. and colleagues shows that response to VEGFA, a master regulator of EC biology, strongly relies on epigenetic mechanisms (Wang et al., 2019). Although less explored, several studies have addressed the role of epigenetic modifications in TECs (Marks et al., 2016). Chromatin remodeling inhibitors reduce tumor growth and angiogenesis by acting on both tumor cells (Kim et al., 2001) and ECs (Deroanne et al., 2002; Hellebrekers et al., 2006). More specifically, high expression of EZH2 in ECs is associated with high-stage and grade, and decreased overall survival in epithelial ovarian cancers (Lu et al., 2010). The authors showed that tumor-derived VEGFs induce expression of EZH2 in ECs, which in turn drives hypermethylation of the anti-angiogenic gene, *Vash1* (Lu et al., 2010) (**Figure 2**). Of note, EZH2 expression in ECs is also under control of the vascular endothelial cadherin, which appears reduced in ovarian TECs (Morini et al., 2018). As new anti-cancer therapies targeting both DNMTs and methylation readers evolve, it is necessary to evaluate their effect in TECs.

Importantly, CAF and TEC biology has a meeting point in what is known as endothelial-to-mesenchymal transition (EndMT) in cancer (Zeisberg et al., 2007). By ChIP-seq, Nagai N. and collaborators found that the transcription factor ERG/FLI1 associates with H3K27ac marks at enhancer/promoter regions of various EC-specific genes, inducing expression of miR-126, which represses EndMT genes. Using available data, the authors also found that lower expression of ERG was significantly related to poor prognosis (Nagai et al., 2018).

NEW THERAPIES IN EPIGENETICS

Epigenetic changes have been suggested as essential for tumor development (Biswas and Rao, 2017). As discussed before, aberrant DNA methylation, histone modifications and chromatin states, as well as aberrant expression of ncRNAs can be used as potential targets by specific drugs and combined with existing therapies. Several molecules targeting epigenetic alterations have been developed and used in different cancers. In the following section, we describe the most recent cancer drugs targeting some epigenetic enzymes. Although in most cases their applications in BC are still in its very early days, we will focus on how they are currently studied in this context.

Drugs Targeting Writers

DNA Methyltransferase Inhibitors

DNMTs inhibitors (DNTMi) are classified in two major subtypes: nucleoside and non-nucleoside inhibitors. Decitabine (5-aza-2'-deoxycytidine) and Azacytidine (5-azacytidine) are cytosine analogues and the best known nucleoside DNMTi. Decitabine and Azacytidine are currently approved by FDA for the treatment of specific forms of myelodysplastic syndromes, chronic myelomonocytic leukemia and acute myeloid leukemia (Lu et al., 2011; Giagounidis et al., 2014). Regarding BC, *in vitro* experiments have demonstrated that Decitabine enhances cisplatin susceptibility, suggesting that combination of both drugs could improve clinical responses (Shang et al., 2008; Wu et al., 2019b). Moreover, Decitabine has completed phase II trials for treatment of BC and phase I trials in combination with tetrahydrouridine (Shang et al., 2008; Bertino and Otterson, 2011).

Second generation nucleoside DNMTi, such as Guadecitabine (SGI-110) or 4'-thio-2'- deoxycytidine, have been developed in order to reduce high toxicity without reducing the therapeutic dose needed. Clinical trials for 4'-thio-2'- deoxycytidine are currently recruiting patients for the treatment of advanced solid tumors (NCT03366116). On the other hand, SGI-110 is in clinical stage for various cancers such as acute myeloid leukemia and myeloidDdysplastic syndrome (NCT03603964), and for different solid tumors like advanced hepatocellular carcinomas (NCT01752933). Also, it has been tested in combination with other therapies such as Ipilimumab in metastatic melanoma (NCT02608437) or with carboplatin in ovarian cancer (NCT01696032), among others. As immune checkpoint inhibitors are currently used in BC, their combination with these second generation DNMTi could represent an attractive scenario to improve the therapeutic response or to expand the number of patients that benfit from immunotherapy. Some others like SGI-1027, a quinoline derivative, and Nanaomycin A, a quinone antibiotic, which are reported to inhibit all three DNMTs or only DNMT3a, respectively, are in preclinical stages for colorectal cancer (Datta et al., 2009; Kuck et al., 2010) (**Figure 3** and **Table 1**).

Apart from these inhibitors, various non-nucleoside DNMTi have been developed and suggested to minimize the direct effect on DNA (Villar-garea et al., 2003). Non-nucleoside analogues, such as Procainamide and MG98, inhibit methylation by binding to the CpG regions of DNA and blocking the activity of DNMTs. MG98, for example, was tested against metastatic renal cell carcinoma but the clinical trial was stopped due to its toxicity (Winquist et al., 2006). However, it has also been evaluated in combination with interferon and results are promising at a specific dose (Amato et al., 2012). Moreover, MG98 was tested in BC patients but the researchers did not find response to the treatment (Plummer et al., 2009).

Histone Lysine Methyltransferase Inhibitors

As it was previously described in this review, HMTs such as G9a and EZH2 are considered oncogenic epigenetic factors in BC (Cho et al., 2015). One of the first histone lysine methyltransferase inhibitors (HKMTi), specific against G9a (EHMT2), was BIX-01294 (Kubicek et al., 2007), which has been shown to inhibit cell proliferation in BC cell lines and induce apoptosis in neuroblastoma cells (Cui et al., 2015). Since then, numerous and improved inhibitors related to G9a blocking have been developed. Various studies have been carried out in molecules like A-366, BRD4770 or UNC0638, in different types of cancer such as neuroblastoma, breast or leukemias (Vedadi et al., 2012; Yuan et al., 2012; Pappano et al., 2015).

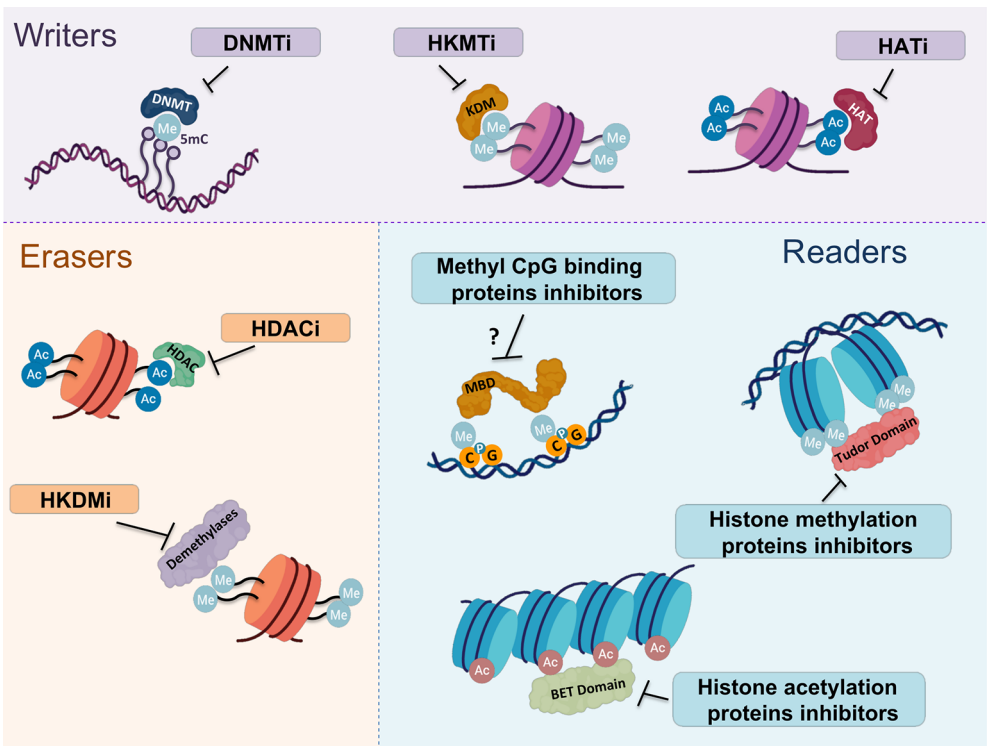


FIGURE 3 | Most representative epigenetic inhibitors targeting writers, readers and erasers. Epigenetic alterations are considered to be reversible and, therefore, all these molecules are subject of study as promising therapeutic targets for cancer treatment. Three main groups of epigenetic drugs can be distinguished according to their targets. The group of compounds targeting epigenetic writers consists mainly of DNMT, HKMT and HAT inhibitors. The second group is directed against epigenetic erasers, which includes HDAC and HKDM inhibitors. Finally, inhibitors of methyl CpG binding proteins, histone methylation and acetylation proteins form the third group targeting epigenetic readers. DNMTi, methyltranferases of DNA inhibitor; HKMTi, histone lysine methyltransferase inhibitor; HATi, histone acetyltransferase inhibitor; HDACi, histone deacetylase inhibitor; HKDMi, histone lysine demethylase inhibitor.

TABLE 1 | A representation of experimental epigenetic drugs targeting writers, readers and erasers.

Drugs Targeting Epigenetic Writers				
Category	**Compound Name**	**Development Stage**	**Cancer Type**	**References**
DNA Methyltransferase inhibitors (DNMTi)				
Nucleoside analogue	Decitabine	Approved	MDS	Giagounidis et al., 2014; Lu et al., 2011
	Azacytidine	Approved	MDS	Giagounidis et al., 2014; Lu et al., 2011
	Guadecitabine	Clinical	MDS, AML	NCT03603964
Non-nucleoside analogue	MG98	Clinical	MRCC	Winquist et al., 2006
	SGI-1027	Preclinical	Colorectal	Datta et al., 2009
	Nanaomycin A	Preclinical	Colorectal	Kuck et al., 2010
Histone Lysine Methyltransferase inhibitors (HKMTi)				
G9a	A-366	Preclinical	Neuroblastoma	Pappano et al., 2015
	BRD4770	Preclinical	Breast	Vedadi et al., 2012
	UNC0638	Preclinical	Leukemia	Yuan et al., 2012
EZH2	UNC1999	Preclinical	Large B-cell lymphoma	Konze et al., 2014
	GSK343	Preclinical	Glioblastoma	Yu et al., 2017
	GSK126	Clinical	Large B-cell lymphomas	NCT02082977
	EPZ6438	Clinical	B-cell Lymphomas	NCT03010982
Histone Acetyltransferase inhibitors (HATi)				
p300	C646	Preclinical	Gastric	Gajer et al., 2015
	PU141	Preclinical	Neuroblastoma	Wang et al., 2017
Drugs Targeting Epigenetic Readers				
Histone Methylation Proteins				
PHD Finger Domain (JARID1A)	Amiodarone	Preclinical	AML	Wagner et al., 2012
MBT Domain	UNC926	Preclinical	Target domain inhibition	Herold et al., 2012
Chromodomain (CBX7)	MS37452	Preclinical	Target domain inhibition	Ren et al., 2015
Histone Acetylation Proteins				
Bet Bromodomain	(+) - JQ1	Preclinical	Colorectal	Zhang et al., 2018b
	OTX015	Clinical	Advanced Solid tumors	NCT02698176
Drugs Targeting Epigenetic Erasers				
Histone Lysine demethylase inhibitors (HKDMi)				
LSD1 Inhibitors	Pargyline	Preclinical	Target domain inhibition	Yang et al., 2018b
	HCI-2509	Preclinical	Neuroblastoma	Gupta et al., 2018
JmjC Domain inhibitors	IOX1	Preclinical	Target domain inhibition	Hopkinson et al., 2013
Histone Deacetylase inhibitors (HDACi)				
Hydroxamic Acid Derivates	Vorinostat	Approved	CTCL	Mann et al., 2007
	Panobinostat	Approved	Blood neoplasias	Eckschlager et al., 2017
	Reminostat	Clinical	Hodgkin's lymphoma	NCT01037478
	Quisinostat	Clinical	Ovarian cancer	NCT02948075

MDS, Myelodysplastic syndromes; MRCC, Metastasic renal cell carcinoma; AML, Acute Myeloid Leukaemia; CTCL, Cutaneous T cell-lymphoma.

Recently, CM272 was described as a novel G9a/DNMT1 dual inhibitor with remarkable antitumor effect in BC *in vitro* and *in vivo* (José-Enériz et al., 2017; Segovia et al., 2019). On the same line, the catalytic subunits of PRC2, EZH1 and EZH2, which catalyze the methylation of H3K27, have been well described in cancer. Some inhibitors of this complex have been studied and they are classified into three groups: (i) pyridone-indazole scaffold like UNC1999 or GSK343 (Konze et al., 2014; Yu et al., 2017) which has been demonstrated to inhibit BC cell lines growth and metastasis (Chen et al., 2019), (ii) pyridone-indole scaffold such as GSK126 (NCT02082977) and (iii) pyridone-phenyl scaffold including EPZ6438 (Brach et al., 2017), known also as Tazemetostat, which has achieved phase I/II trial (NCT03854474) for the treatment of patients with locally advanced or metastatic urothelial carcinoma in combination with pembrolizumab. The potential use of EZH2 in the BC context has been recently reviewed and discussed (Martínez-Fernández et al., 2015c; Segovia and Paramio, 2017).

Histone Acetyltransferase Inhibitors

HATs are typically grouped into three broad families, namely the p300/CBP, the Gcn5 related N-acetyl-transferase and the MYST family. Among them, p300/CBP seems to be frequently mutated in BC (Duex et al., 2018a) and was reported to be associated with doxorubicin resistance (Takeuchi et al., 2012), so it could be a promising molecular therapeutic target for this disease. Accordingly, C646 and PU141 have been demonstrated to be promising in gastric cancer and neuroblastoma, respectively (Gajer et al., 2015; Wang et al., 2017). However, there is very little evidence for useful histone acetyltransferase inhibitors (HATi) being developed and tested (Baell and Miao, 2016), even though the search for new small-molecule HATi has been intense in the last decades (**Figure 3**). Although, to our knowledge no HATi are being tested in BC, it is important to consider that HAT gene deficiencies may confer susceptibilities to other inhibitors, opening new possible therapeutic approaches for various tumors, including BC (Ogiwara et al., 2016).

Drugs Targeting Readers

Methyl CpG Binding Proteins

Sites of DNA methylation recruit two important protein families: MBD and ZnF proteins. The MBD protein family uses its DNA binding domains and other protein-protein domains to alter the transcriptional state of the DNA (Ginder and Williams, 2018). However, the MBD family is not the only protein family that allows the recognition of methylated DNA; for example, the Kaiso protein family (Kaiso/ZBTB33, ZBTB4 and ZBTB38) uses a three-finger zinc motif to bind methylated CGCG (Hendrich and Bird, 2015). Additionally, it has been demonstrated that ZBTB38 promotes cell migration, invasive growth and EMT in BC cell lines (Jing et al., 2018), whereas high MBD2 expression was significantly associated with reduced bladder carcinoma risk (Zhu et al., 2004). Even though different experimental approaches have identified these proteins as good therapeutic targets, inhibitors have not yet been developed to slow down their action (**Figure 3**).

Histone Methylation Proteins

The histone methyl protein family is a large family of proteins that binds differently to methylated lysine and arginine residues and can be divided into several subfamilies: Tudor domain, PHD finger, MBT, chromodomain and BRD. The most studied family among them is the PHD family, which comprises a group of versatile readers of the epigenome that can recognize both methylation and acetylation marks and has been involved in cancer progression (Hayami et al., 2010). Recently, Wagner et al. discovered various compounds that inhibit the PHD of this protein (Wagner et al., 2012). Among them, Amiodarone is able to induce apoptosis in the T24 BC cell line (Bognar et al., 2017). Upregulated UHRF1 (E3 ubiquitin-protein ligase 1), which contains PHDs, has also been shown to promote BC cell invasion *in vitro* and *in vivo* by epigenetic silencing of KiSS1 (Zhang et al., 2014).

Histone Acetylation Proteins

In general, histone acetylation is related to transcriptional activation. Different protein domains that bind specifically to acetylated histones have been identified so far, including the BRD, double PHD finger and Yeats domains. The BRD family identifies acetylated lysine residues, such as those on the *N*-terminal tails of histones, and has been proposed as an attractive therapeutic target due to its involvement in various cancer types. The BET family has been thoroughly investigated (Biswas and Rao, 2018). The first inhibitors of the BET family, I-BET762 (GSK525762) and (+)-JQ1, were reported in 2010 (Filippakopoulos et al., 2010). The inhibitor I-BET762 has recently been studied for dose escalation clinical studies to investigate the safety, pharmacokinetics, pharmacodynamics, and clinical activity in various tumors (NCT01587703), but BC patients were not included in this study. (+)-JQ1 interferes with BRD4 function, blocking the formation of the NUT-BRD4 oncoprotein, and various studies have shown its efficacy in hematological and solid malignancies (Abedin et al., 2016; Ocaña et al., 2017; Gao et al., 2018; Sakaguchi et al., 2018; Tan et al., 2018; Zhang et al., 2018b). Regarding BC, the (+)-JQ1 inhibitor induces autophagy through activation of the LKB1/AMPK pathway, contributing to the inhibition of proliferation of BC cell lines *in vitro* (Li et al., 2019). In combination with Mitomicyn C, (+)-JQ1 enhances cell death, which offers the possibility of a dose reduction of the chemotherapeutic agent (Simm et al., 2018). Hölscher et al. had also shown significant synergistic effects on the induction of apoptosis in urothelial cancer cells by treatment with (+)-JQ1 and Romidepsin, an HDAC inhibitor (HDACi), thus suggesting a promising new combination therapy approach for urothelial cancer (Hölscher et al., 2018).

Even though BRD3 inhibitors have not been studied as much as those of the BRD2/4, it has been observed that I-BET151, a pan-BET inhibitor that targets BRD3 (Picaud et al., 2013), halts the progression of the cell cycle and decreases cell proliferation *in vitro* and *in vivo* by targeting lncRNA *HOTAIR* in glioblastoma (Pastori et al., 2014). Remarkably, *HOTAIR* increased expression is also associated with poor clinical outcome in BC (Martínez-Fernández et al., 2015b), thereby indicating the possible relevance of studying I-BET151 inhibitor in this type of cancer.

Drugs Targeting Erasers

Epigenetic marks can be 'erased', depending on the requirement of the cell, by a group of enzymes that oppose to the writers. Since they also modulate gene expression affecting tumor suppressor genes or oncogenes, they can be considered potential targets.

Histone Lysine Demethylase Inhibitors

Researchers have been exploring inhibitory molecules for the HKDMs KDM1 (LSD1) and KDM2-8 for years (Højfeldt et al., 2013). Early compounds were developed based on the structural characteristics of LSD1 (Yang et al., 2018b). Treatment with LSD1 inhibitor supressed BC cell proliferation and androgen-induced transcription, supporting a novel role for the androgen receptor-KDM (lysine demethylases) complex in BC initiation and progression (Kauffman et al., 2012). Even though numerous LSD1 inhibitors have been reported in the literature, they are in the initial phase of development and there are still many problems that have to be overcome before histone lysine demethylase inhibitors (HKDMi) can reach the clinic (**Figure 3**).

Histone Deacetylase Inhibitors

Various reports have shown that HDACs could be involved in regulating protein function and tumorigenesis. In this line, the use of HDACi has been clinically validated in cancer treatment and, so far, four drugs have been approved by the FDA: Vorinostat, Romidepsin, Panobinostat and Belinostat (**Figure 3**). Vorinostat was the first pan-HDACi approved by the FDA for the treatment of advanced primary cutaneous T-cell lymphoma (Mann et al., 2007). Next, various pharmaceutical companies developed other molecules such as Panobinostat or Belinostat (Eckschlager et al., 2017), all of them intended initially for blood neoplasias.

Moreover, HDACi are being studied for BC therapy (Kaletsch et al., 2018). Romidepsin and Vorinostat have been tested in a phase II trial as monotherapy, and Vorinostat has also completed phase I trials as a combination therapy with docetaxel, but it

was surprisingly toxic and had limited efficacy (Cheung et al., 2008). Additionally, Belinostat has obtained positive responses in BC cells through decreasing cell proliferation *in vitro* and *in vivo* (Buckley et al., 2007) and is being tested in clinical trials against various solid tumors including BC (NCT00413322, NCT00413075).

Apart from the hydroxamic acid derivates, which are approved for the clinic, other molecules are in different phases of study. Some of them are Reminostat (4SC-201) evaluated for Hodgkin's lymphoma (NCT01037478), Quisinostat (JNJ-26481585) for the treatment of ovarian cancer (NCT02948075), or Abexinostat (PCI-24781) which is being evaluated for sarcoma in combination with Doxorubicin (NCT01027910). **Table 1** summarizes HDACi approved and some experimental HDACi in different stages of clinical development.

Epigenetic drugs, as seen previously, have been approved as monotherapy for the treatment of different types of cancer. In addition, the combination of epigenetic drugs with standard chemotherapy or immunotherapy has been explored in recent years with promising results. The basis for this approach comes from results showing that epigenetic drugs reduce the apoptotic threshold, reverse drug resistance and/or induce immune response. Regarding BC, a large proportion of patients are not candidates to chemotherapy due to comorbidities. The use of epigenetic drugs could bring the possibility of a dose reduction, which makes these compounds attractive candidates for combination therapy for these BC patients (Witjes et al., 2014b; Fardi et al., 2018).

Drugs Targeting ncRNAs

LncRNAs

Even though no lncRNA-based targeted BC treatment has been developed so far, modulation of lncRNA expression as a therapy seems promising and has already been described for other cancer types (Bhan et al., 2017). Methods described for the modulation of lncRNA expression include the use of antisense oligonucleotide (ASO) or lncRNA-specific siRNAs for transcript destabilization or degradation, as well as transcript alteration by modulation of lncRNA-encoded promotor activity. Additionally, functional disruption of lncRNAs through aptamers antagonizing the interaction with their binding partners, or the production of synthetic molecules interfering with the association between lncRNAs and regulatory factors, are possible mechanisms to modulate lncRNA expression (Bhan et al., 2017). Finally, these ncRNAs might be valuable in combination therapy and augmentation of therapeutic efficacy since modulation of their expression can enhance the therapeutic sensitivity of tumors (Bhan et al., 2017).

MiRNAs

There are many approaches that have been employed to silence miRNAs in cancer. These include anti-miRNA oligonucleotides (AMOs), miRNA-masking antisense oligonucleotides, peptide nucleic acids and miRNA sponges (Garzon et al., 2010). AMOs mechanism relies on the complementary base pairing of the oligonucleotide sequence to its target miRNA. Therefore, these molecules can repress cellular mRNAs involved in tumor progression and proliferation, and they can also act as competitive inhibitors of miRNAs and impair their interaction with other molecules (Lima et al., 2018b). Joana Filipa and colleagues showed that, using AMOs, they were able to silence the expression of upregulated miR-9 in a cancer cell model of gastric cancer (Lima et al., 2018a).

For BC treatment, there are some indirect therapeutic approaches that affect miRNA expression. For instance, some EZH2 inhibitors act in BC cells modulating the expression of miR-101 (Wang et al., 2014) or miR-143 (Zhang et al., 2015). However, some of these miRNAs are also induced by specific oncogenic insults in BC, indicating the potential problems of considering them as possible targets for treatment (Segovia et al., 2017).

Remarkably, a miRNA-based drug mimicking miR-34a has reached a phase I clinical trial (NCT01829971). MiRNA-34a significance in various human cancers, including BC, is increasingly recognized nowadays (Bader, 2012; Misso et al., 2014), hence the expectation in this new approach.

Other nCRNAs

CircRNAs and piRNAs have been described as a promising therapeutic target in multiple cancer types, including BC (see corresponding section). Potential strategies for the modulation of circRNA expression include the use of ASOs or siRNAs in order to antagonize these ncRNAs, as well as the application of the CRISPR/Cas system to partially or completely remove oncogenic circRNAs (Zhang and Xin, 2018). Regarding the modulation of piRNA expression, possible strategies include the use of synthetic piRNAs at the transcriptional and posttranscriptional level, while antibodies against PIWI proteins might be effective as a posttranscriptional approach (Assumpção et al., 2015). Nonetheless, none of these approaches are being tested in BC therapy so far.

EPIGENETIC ALTERATIONS AS BIOMARKERS IN BC: THE POTENTIAL USE OF LIQUID BIOPSY

Regarding diagnosis and surveillance of BC, a combination of cystoscopy and urine cytology is the most widely used methodology nowadays. Currently, cystoscopy is the gold standard method in clinical practice for detection and follow-up of this disease, with a sensitivity of 85–90% to detect exophytic tumors. However, this technique is highly invasive, showing a big inter-observer and intra-observer variation. On the other hand, BC urinary cytology shows a specificity of approximately 98% but a low sensitivity of 38%. The high rates of recurrence and progression of BC require continuous follow-up of patients by cystoscopy (every 3–6 months during the next 5 years) and urine cytology, making BC one of the most costly malignancies for the National Health systems of developed countries (Lodewijk et al., 2018).

For these reasons, there is a clear need to improve the current systems of diagnosis, prognosis and surveillance of BC patients. Based on the important role of epigenetic modifications in this disease, status evaluation of the involved molecules could

contribute to improve these available systems. In this context, liquid biopsy has emerged as a non-invasive way to determine the genomic landscape of cancer patients, as well as to monitor treatment response, quantify minimal residual disease, and assess therapy resistance (Bardelli and Pantel, 2017; Di Meo et al., 2017; Heitzer et al., 2017; Khetrapal et al., 2018). Liquid biopsy makes reference to the sampling and assessment of biological fluids. In genitourinary cancer, due to the proximity of tumors, urine has been considered a bona fide liquid biopsy sample, being one of the most interesting samples for its easy access and collection. However, in MIBC patients after cystectomy, serum and plasma could be the most appropriate liquid biopsy samples given its invasive and metastatic character (Lodewijk et al., 2018). Currently, there are several systems to detect and follow-up BC using liquid biopsy biomarkers (including sediment cells in urine samples, CTCs in blood samples as well as RNAs and proteins in both cases), which present sensitivity and specificity values within a range of 38–98% and 65–98%, respectively (Lodewijk et al., 2018). The determination of epigenetic alterations in liquid biopsy samples, such as variation in expression levels of ncRNAs or changes in DNA methylation profiles, could improve the predictive values of the current systems of BC diagnosis, prognosis and monitoring. Next, some of the most relevant studies of epigenetic biomarkers in urine and serum/plasma samples are discussed.

Non-Coding RNAs as Epigenetic Biomarkers in Liquid Biopsy of BC Patients

Among the different ncRNAs previously described, miRNAs have been the most widely studied molecules in liquid biopsies so far. MiRNA molecules have several characteristics which make them potential candidates as good biomarkers in liquid biopsy samples: i) they show very homogeneous expression levels among individuals and specific expression profiles in different types of tissue (Liang et al., 2007); ii) they are included in a protein complex and, usually, in exosomes, which confers them high stability, preserving their integrity and preventing their degradation (Weber et al., 2010; Ge et al., 2014; Martínez-Fernández et al., 2016); iii) there are several systems designed to determinate ncRNA expression using RT-qPCR, which allow evaluating a large number of miRNAs from very small amounts of total RNA and at a low cost.

Given the potential of miRNAs, many studies have evaluated their predictive properties, individually or in combination, in the urine of BC patients. In this context, high expression levels of miR-146a-5p and miR-106b have been related with invasion and high grade and stage BC (Zhou et al., 2014; Sasaki et al., 2016). NMIBC patients present high levels of miR-214 in urine samples and, curiously, expression of this miRNA was inversely correlated with risk of recurrence of BC patients (Kim et al., 2013). Besides, some miRNAs such as miR-92a-3p and miR-140-5p have been associated with progression after recurrence (Ingelmo-Torres et al., 2017). Yun and collaborators have demonstrated that urine miR-145 expression levels decrease in BC patients with respect to healthy controls, both in non-invasive and invasive tumors (77.8% and 84.1% sensitivity, respectively, and 61.1% specificity in both cases). They observed an association between downregulation of miR-200a and high risk of recurrence in patients with invasive tumors (Yun et al., 2012). Besides, miR-155 has proved to be a good biomarker in urine samples, distinguishing non-invasive tumors, inflammation and healthy controls with a sensitivity of 80.2% and a specificity of 84.6% (Zhang et al., 2016).

As previously mentioned, detection of miRNA deregulation in serum or plasma may have special relevance in invasive and metastatic tumors. Yang and colleagues observed miR-210 increased expression levels in serum samples of BC patients, being associated with tumor stage, grade, and useful to predict tumor progression (AUC = 0.898) (Yang et al., 2015). Moreover, some studies using plasma have found a positive correlation between upregulation of miR-19a and miR-200b with tumor grade and stage respectively, whereas miR-92 and miR-33 presented inverse association with tumor stage (Adam et al., 2013; Feng et al., 2014).

In recent years, several panels of miRNAs (encompassing profiles from 6 to 25 miRNAs) have been developed in both urine and serum for BC diagnosis, prognosis and monitoring of recurrence. In this context, we have recently gathered some of the main miRNA profiles in BC liquid biopsies which can be consulted in Table 2 at Lodewijk et al. (2018).

Although variation in lncRNA expression levels has not been studied as widely as miRNAs in liquid biopsy samples, altered levels of expression of these molecules have been found in urine and blood samples of BC patients. Increased expression levels of *UCA1* in urine samples has been associated with the presence of high-grade NMIBC, and an integrative meta-analysis including more than 500 BC patients and healthy donors determined that its upregulation may predict BC (81% sensitivity and 86% specificity, AUC = 0.88) (Wang et al., 2006; Cui et al., 2017). In addition, other lncRNAs such as *HOTAIR*, *MALAT1*, *HOX-AS-2*, *OTX2-AS1*, *HYMAI*, *LINC00477* and *LOC100506688*, have shown upregulation in urine exosomes of MIBC patients (Berrondo et al., 2016). In addition, *H19* gene expression is significantly higher in BC patients, and its presence has been detected in the urine of 90.5% of patients *versus* 25.9% of healthy controls (AUC = 0.933) (Gielchinsky et al., 2017).

Additionally, other ncRNAs such as piRNAs and circRNA have been evaluated in biofluids. Both molecules have shown a particular resistance to degradation by exoribonuclease, making them ideal candidates for biomarker development (Pardini and Naccarati, 2017; Vo et al., 2019). Several studies have reported that piRNAs are widely detected in liquid biopsy samples, being especially abundant in urine samples and, therefore, good candidates as new biomarkers for BC. Although no deregulated piRNAs have been found in urine or blood of BC patients so far, expression level alterations of some of these molecules have been shown in liquid biopsy samples of other tumor types (Freedman et al., 2016; Iliev et al., 2016; Yuan et al., 2016; Pardini and Naccarati, 2017). Regarding circRNA, Vo and collaborators have recently developed MiOncoCirc, a technology based on exome capture RNA-seq, which stands as the first cancer-focused circRNA resource to facilitate the study of circRNAs as new biological markers of cancer (Vo et al., 2019). They were able to

identify candidate circRNAs which could serve as biomarkers for prostate cancer, detecting circRNAs in urine (Vo et al., 2019). This technology could open new possibilities to find new biomarkers with predictive values in liquid biopsy samples of BC patients. Nevertheless, even though circRNAs show great potential as valuable biomarker in urine, these RNAs seem to be highly susceptible to circulating RNA endonucleases showing a half-life of only 15 seconds in human serum, which limits their use as a biomarker in this biological fluid (Jeck and Sharpless, 2014).

DNA Methylation Profiles as Epigenetic Biomarkers in Liquid Biopsy of BC Patients.

As mentioned above, changes in methylation are chemically stable and have been broadly reported in BC. Therefore, they are an interesting source of candidate biomarkers to be detected in biofluids including both blood and urine. Currently, there are multiple methods for detecting changes in methylation comprising global genome methylation and specific genes of interest assays. The majority of methods to evaluate specific genes are based on bisulfite conversion followed by PCR and sequencing, pyrosequencing or methylation-specific PCR, among others, which generally show a high sensitivity and specificity and low assay-to-assay variability (Kurdyukov and Bullock, 2016). Already in 2002, Valenzuela and collaborators found that methylation in *p16(lNK4a)* promoter in serum could be useful as diagnostic biomarker with 22% of sensitivity, 95% of specificity and a positive predictive value of 0.98 (Valenzuela et al., 2002). Also in serum, both the methylation in promoters of protocadherin 17 (*PCDH17*) and protocadherin-10 (*PCDH10*) showed an association with BC poor prognosis (Lin et al., 2012; Luo et al., 2014). A slight association between hypermethylation in *p16(lNK4a)* and *DAPK* promoter regions and NMIBC has been also described (Jabłonowski et al., 2011). Finally, the presence of hypermethylated DNA in *APC*, *GSTP1* or *TIG1* in the serum of BC patients was associated with a worse outcome showing 80% sensitivity and 93% specificity for BC detection (Ellinger et al., 2015).

Important for BC, alterations in DNA methylation can be also assessed both in circulating cell-free DNA and in cells shed into urine. In general, it seems that a prevalence of hypermethylated genes is found in urine from BC patients. For instance, the evaluation of methylation in *TWIST1* and *NID2* in urine sediment has shown 90% sensitivity and 93% specificity (Renard et al., 2010; Fantony et al., 2017; van der Heijden et al., 2018). Other studies showed promising results using methylation of *CFTR*, *SALL3* and *TWIST1* genes in urine cell pellets in combination with cytology (van der Heijden et al., 2018). Interestingly, *SOX-1*, *IRAK3*, and *Li-MET* genes methylation status has showed better recurrence predictivity than urine cytology and cystoscopy (80 vs. 35 vs. 15%) (Su et al., 2014). Also in urine sediments, methylation in *p14ARF*, *p16INK4A*, *RASSF1A*, *DAPK*, and *APC* showed a correlation with BC grade and stage (Pietrusiński et al., 2017). Guo et al. used the methylation status for *VIM*, *RASSF1A*, *GDF15*, and *TMEFF2* to identify BC with 82% sensitivity and 53% specificity (Li et al., 2018b). *RBBP8* has been identified as almost exclusively hypermethylated in BC (Mijnes et al., 2018), while Chen et al. showed *CDH13* methylation as a biomarker with prognostic value for BC screening in urine samples (Ren et al., 2016). Using quantitative methylation-specific PCR, a novel two-gene panel with high accuracy in an urine-based test has just been described (Bosschieter et al., 2019). When stratifying in low- or high-risk NMIBC patients, 97.6% sensitivity and 84.8% specificity were obtained using promoter hypermethylation of *HS3ST2*, *SEPTIN9* and *SLIT2* genes in combination with *FGFR3* mutation (Roperch et al., 2016). Interestingly, Patchsung et al. obtained a sensitivity and specificity of 96% for BC screening using a combination of the urinary hypomethylated *LINE-1* loci and the plasma protein carbonyl content (Patchsung et al., 2012). But methylation value has not only been studied in genes and their promoters: for example, last year Shindo et al. reported a study using the methylation of four miRNAs (miR-9-3, miR-124-2, miR-124-3, and miR-137) in voided urine samples, finding an association with recurrence and radical cystectomy (Kitajima et al., 2017).

As a consequence of these new results, there are currently several clinical trials using promising urine-based tests. Among them, Bladder EpiCheck™ (based on the use of methylation-sensitive restriction enzymes followed by RT-PCR) includes a panel of 15 DNA methylation patterns for the identification of recurrent BC from urine samples. First validation results with data from 357 patients showed 88% specificity and a negative predictive value (NPV) of 94.4% for the detection of any cancer, and a NPV of 99.3% for the detection of high-grade cancer (D'Andrea et al., 2019). Another test is AssureMDx, which uses methylation of *OTX1*, *ONECUT2* and *TWIST1* in addition to mutational load of *FGFR3*, *TERT* and *HRAS* in cell pellets from urine samples, showing a sensitivity of 93–97% and a specificity around 81.7–86% (van Kessel et al., 2017). Finally, Uromark was described 2 years ago as a targeted bisulfite next-generation sequencing assay based on 150 CpG loci to diagnose BC from urine with a sensitivity of 98%, specificity of 97% and NPV of 97% for the detection of primary BC (Feber et al., 2017). Following these results, DETECT I and DETECT II are two multi-centre prospective observational studies designed to conduct a robust validation of the UroMark assay. DETECT I will recruit patients having diagnostic investigations for haematuria, while DETECT II will recruit patients with new or recurrent BC to determine respectively the NPV and the sensitivity of UroMark.

As a conclusion, although validation studies are still ongoing, the recent and promising results prompt us to be optimistic and have confidence in a near clinical implementation of a urine methylation test for BC diagnosis and prognosis.

FUTURE PROSPECTS

It is clear that epigenetics has reshaped most of our concepts of biology and, undoubtedly, molecular biology understanding of human pathologies. From the point of view of those researchers interested in BC, or even in cancer in general, it is almost impossible to predict what the future will bring us in this field,

but there are two clear emerging facets at our hands. On the one hand, the use of compounds interfering with many epigenetic processes combined with other therapies currently in the clinics and, on the other hand, the use of these therapies directed not only towards the tumor cells, but also the tumor niche. Obviously, from our current knowledge of immunotherapies, there is a faint border between these two concepts.

Epigenetic drugs, as seen in the previous sections, have been approved as monotherapy for the treatment of different types of cancer. Additionally, they have been shown to synergize with other epigenetic substances or anticancer therapies. The first preclinical investigations focused on the combination of DNMTi and HDACi (Cameron et al., 1999). After a while and due to the development of new epigenetic agents directed to other targets such as HMTs, HDMs or BRDs, new synergistic combinations with DNMTi and/or HDACi are being explored. In addition, due to the importance of immunotherapy in cancer, the combination of epigenetic drugs with standard chemotherapy or immunotherapy has also increased in recent years (Dunn and Rao, 2017). This is based on the theory of using epigenetic drugs to reduce the apoptotic threshold, reverse drug resistance or induce immune responses for further treatment such as chemotherapy or immunotherapy. The concept of partnering epigenetic therapy with reshaping stromal component strategies has generated a wave of translational research that highlights the potential for this approach in many different cancer types. Epigenetic drugs such as DMNTi and HDACi can reverse immune suppression, and modulate stromal cells and extracellular matrix *via* several mechanisms such as enhancing expression of tumor-associated antigens, components of the antigen processing and presenting machinery pathways, immune checkpoint inhibitors, chemokines, and other immune-related genes, as well as changing the CAFs secretomes that will favor or impede the tumor growth. But deep studies of each component interaction are still in their early days. The discoveries in these areas have established a highly promising basis for studies using combined epigenetic and immunotherapeutic agents as anti-cancer therapies with expected long lasting antitumor responses.

Finally, new areas of research such as the use of new gene targeting strategies as therapeutic tools or the potential role of epigenetic mechanisms leading to altered glycosylation, which may clearly impact the liquid biopsy and immunotherapy fields (Dall'Olio and Trinchera, 2017), may represent new horizons in BC management and detection.

AUTHOR CONTRIBUTIONS

All authors contributed equally to review the current literature and write specific sections. The whole work was coordinated by JP. All the authors agreed with the final version.

FUNDING

This study was funded by the following: FEDER cofounded MINECO grant SAF2015-66015-R, grant ISCIII-RETIC RD12/0036/0009, PIE 15/00076 and CB/16/00228 to JP. VM is funded by Consejería de investigación e Innovación, Comunidad de Madrid (ref 2018-T2/BMD-10342).

REFERENCES

Abedin, S. M., Boddy, C. S., and Munshi, H. G. (2016). BET inhibitors in the treatment of hematologic malignancies: current insights and future prospects. *Oncol. Targets Ther.* 9, 5943–5953. doi: 10.2147/OTT.S100515

Adam, L., Wszolek, M. F., Liu, C.-G., Jing, W., Diao, L., Zien, A., et al. (2013). Plasma microRNA profiles for bladder cancer detection. *Urol. Oncol. Semin. Orig. Investig.* 31, 1701–1708. doi: 10.1016/j.urolonc.2012.06.010

Agger, K., Cloos, P. A. C., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., et al. (2007). UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* 449, 731–734. doi: 10.1038/nature06145

Agundez, M., Grau, L., Palou, J., Algaba, F., Villavicencio, H., and Sanchez-Carbayo, M. (2011). Evaluation of the Methylation Status of Tumour Suppressor Genes for Predicting Bacillus Calmette-Guérin Response in Patients With T1G3 High-Risk Bladder Tumours. *Eur. Urol.* 60, 131–140. doi: 10.1016/j.eururo.2011.04.020

Albrengues, J., Bertero, T., Grasset, E., Bonan, S., Maiel, M., Bourget, I., et al. (2015). Epigenetic switch drives the conversion of fibroblasts into proinvasive cancer-associated fibroblasts. *Nat. Commun.* 6, 10204. doi: 10.1038/ncomms10204

Allis, C. D., Berger, S. L., Cote, J., Dent, S., Jenuwien, T., Kouzarides, T., et al. (2007). New Nomenclature for Chromatin-Modifying Enzymes. *Cell* 131, 633–636. doi: 10.1016/j.cell.2007.10.039

Amato, R. J., Stephenson, J., Hotte, S., Nemunaitis, J., Karl, B., Reid, G., et al., (2012). MG98, a Second-Generation DNMT1 Inhibitor, in the Treatment of Advanced Renal Cell Carcinoma. *Cancer Invest.* 2, 415–421. doi: 10.3109/07357907.2012.675381.

Anastasiadou, E., Jacob, L. S., and Slack, F. J. (2017). Non-coding RNA networks in cancer. *Nat. Rev. Cancer* 18, 5–18. doi: 10.1038/nrc.2017.99

Ariel, I., de Groot, N., and Hochberg, A. (2000). Imprinted H19 gene expression in embryogenesis and human cancer: the oncofetal connection. *Am. J. Med. Genet.* 91, 46–50. doi: 10.1002/(sici)1096-8628(20000306)91:1<46::aid-ajmg8>3.0.co;2-i

Assumpção, C. B., Calcagno, D. Q., Araújo, T. M. T., Santos, S. E. B., dos, Santos, Â. K. C. R., dos, Riggins, G. J., et al. (2015). The role of piRNA and its potential clinical implications in cancer. *Epigenomics* 7, 975–984. doi: 10.2217/epi.15.37

Babjuk, M., Böhle, A., Burger, M., Capoun, O., Cohen, D., Compérat, E. M., et al. (2017). EAU Guidelines on Non–Muscle-invasive Urothelial Carcinoma of the Bladder: Update 2016. *Eur. Urol.* 71, 447–461. doi: 10.1016/j.eururo.2016.05.041

Bader, A. G. (2012). MiR-34 - a microRNA replacement therapy is headed to the clinic. *Front. Genet.* 3, 1–9. doi: 10.3389/fgene.2012.00120

Baell, J. B., and Miao, W. (2016). Histone acetyltransferase inhibitors: where art thou? *Futur. Med. Chem.* 8, 1525–1528. doi: 10.4155/fmc-2016-0151

Balar, A. V., Galsky, M. D., Rosenberg, J. E., Powles, T., Petrylak, D. P., Bellmunt, J., et al. (2017). Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: a single-arm, multicentre, phase 2 trial. *Lancet* 389, 67–76. doi: 10.1016/S0140-6736(16)32455-2

Balbás-Martínez, C., Sagrera, A., Carrillo-de-Santa-Pau, E., Earl, J., Márquez, M., Vazquez, M., et al. (2013). Recurrent inactivation of STAG2 in bladder cancer is not associated with aneuploidy. *Nat. Genet.* 45, 1464–1469. doi: 10.1038/ng.2799

Banerjee, J., Mishra, R., Li, X., Jackson, R. S., Sharma, A., and Bhowmick, N. A. (2014). A reciprocal role of prostate cancer on stromal DNA damage. *Oncogene* 33, 4924–4931. doi: 10.1038/onc.2013.431

Bardelli, A., and Pantel, K. (2017). Liquid Biopsies, What We Do Not Know (Yet). *Cancer Cell* 31, 172–179. doi: 10.1016/j.ccell.2017.01.002

Barrett, S. P., and Salzman, J., (2016). Circular RNAs: analysis, expression and potential functions. *Development.* 143, 1838–1847. doi: 10.1242/dev.128074

Baylin, S. B., and Jones, P. A. (2016). Epigenetic Determinants of Cancer. *Cold Spring Harb. Perspect. Biol.* 8, a019505. doi: 10.1101/cshperspect.a019505

Baylln, S. B., Herman, J. G., Graff, J. R., Vertino, P. M., and Issa, J.-P. (1997). Alterations in DNA Methylation: A Fundamental Aspect of Neoplasia. *Adv. Cancer Res.* 72, 141–196. doi: 10.1016/S0065-230X(08)60702-2

Bechtel, W., McGoohan, S., Zeisberg, E. M., Müller, G. A., Kalbacher, H., Salant, D. J., et al. (2010). Methylation determines fibroblast activation and fibrogenesis in the kidney. *Nat. Med.* 16, 544–550. doi: 10.1038/nm.2135

Bellmunt, J., de Wit, R., Vaughn, D. J., Fradet, Y., Lee, J. L., Fong, L., et al. (2017). Pembrolizumab as Second-Line Therapy for Advanced Urothelial Carcinoma. *N. Engl. J. Med.* 376, 1015–1026. doi: 10.1056/NEJMoa1613683

Bernstein, B. E., Meissner, A., and Lander, E. S. (2007). The mammalian epigenome. *Cell* 128, 669–681. doi: 10.1016/j.cell.2007.01.033

Berraondo, P., Minute, L., Ajona, D., Corrales, L., Melero, I., and Pio, R. (2016). Innate immune mediators in cancer: between defense and resistance. *Immunol. Rev.* 274, 290–306. doi: 10.1111/imr.12464

Berrondo, C., Flax, J., Kucherov, V., Siebert, A., Osinski, T., Rosenberg, A., et al. (2016). Expression of the long non-coding RNA HOTAIR correlates with disease progression in Bladder Cancer and Is Contained in Bladder Cancer Patient Urinary Exosomes. *PloS One* 11, e0147236. doi: 10.1371/journal.pone.0147236

Bertino, E. M., and Otterson, G. A. (2011). Romidepsin: a novel histone deacetylase inhibitor for cancer. *Expert Opin. Investig. Drugs* 20, 1151–1158. doi: 10.1517/13543784.2011.594437

Bhan, A., Soleimani, M., and Mandal, S. S. (2017). Long noncoding RNA and cancer: A new paradigm. *Cancer Res.* 77, 3965–3981. doi: 10.1158/0008-5472.CAN-16-2634

Bian, C., Chen, Q., and Yu, X. (2015). The zinc finger proteins ZNF644 and WIZ regulate the G9a/GLP complex for gene repression. *ELife* 4(e05606). doi: 10.7554/elife.05606

Bistoni, F., Vecchiarelli, A., Cenci, E., Puccetti, P., Marconi, P., and Cassone, A. (1986). Evidence for macrophage-mediated protection against lethal Candida albicans infection. *Infect. Immun.* 51, 668–674.

Biswas, S., and Rao, C. M. (2017). Epigenetics in cancer: Fundamentals and Beyond. *Pharmacol. Ther.* 173, 118–134. doi: 10.1016/j.pharmthera.2017.02.011

Biswas, S., and Rao, C. M. (2018). Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. *Eur. J. Pharmacol.* 837, 8–24. doi: 10.1016/j.ejphar.2018.08.021

Bitler, B. G., Aird, K. M., Garipov, A., Li, H., Amatangelo, M., Kossenkov, A. V., et al. (2015). Synthetic lethality by targeting EZH2 methyltransferase activity in ARID1A-mutated cancers. *Nat. Med.* 21, 231–238. doi: 10.1038/nm.3799

Bochner, B. H., Cote, R. J., Weidner, N., Groshen, S., Chen, S.-C., Skinner, D. G., et al. (1995).). Angiogenesis in Bladder Cancer: Relationship Between Microvessel Density and Tumor Prognosis. *JNCI J. Natl. Cancer Inst.* 87, 1603–1612. doi: 10.1093/jnci/87.21.1603

Bock, C., Beerman, I., Lien, W.-H., Smith, Z. D., Gu, H., Boyle, P., et al. (2012). DNA methylation dynamics during *in vivo* differentiation of blood and skin stem cells. *Mol. Cell* 47, 633–647. doi: 10.1016/j.molcel.2012.06.019

Bognar, Z., Fekete, K., Antus, C., Hocsak, E., Bognar, R., Tapodi, A., et al. (2017). Desethylamiodarone—A metabolite of amiodarone—Induces apoptosis on T24 human bladder cancer cells *via* multiple pathways. *PloS One* 12, 1–18. doi: 10.1371/journal.pone.0189470

Bosschieter, J., Nieuwenhuijzen, J. A., Hentschel, A., van Splunter, A. P., Segerink, L. I., Vis, A. N., et al. (2019). A two-gene methylation signature for the diagnosis of bladder cancer in urine. *Epigenomics* 11, 337–347. doi: 10.2217/epi-2018-0094

Brach, D., Johnston-blackwell, D., Drew, A., Lingaraj, T., Motwani, V., Warholic, N. M., et al. (2017). EZH2 Inhibition by Tazemetostat Results in Altered Dependency on B-cell Activation Signaling in DLBCL. *Mol. Cancer Ther*. 16 (11). doi: 10.1158/1535-7163.MCT-16-0840

Bröske, A.-M., Vockentanz, L., Kharazi, S., Huska, M. R., Mancini, E., Scheller, M., et al. (2009). DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. *Nat. Genet.* 41, 1207–1215. doi: 10.1038/ng.463

Buckley, M. T., Yoon, J., Yee, H., Chiriboga, L., Liebes, L., Ara, G., et al. (2007). The histone deacetylase inhibitor belinostat (PXD101) suppresses bladder cancer cell growth *in vitro* and *in vivo*. *J. Transl. Med.* 5, 1–12. doi: 10.1186/1479-5876-5-49

Cameron, E. E., Bachman, K. E., Myöhänen, S., Herman, J. G., and Baylin, S. B. (1999). Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. *Nat. Genet.* 21, 103–107. doi: 10.1038/5047

Cao, R., Tsukada, Y., and Zhang, Y. (2005). Role of Bmi-1 and Ring1A in H2A Ubiquitylation and Hox Gene Silencing. *Mol. Cell* 20, 845–854. doi: 10.1016/J.MOLCEL.2005.12.002

Casadevall, D., Kilian, A. Y., and Bellmunt, J. (2017). The prognostic role of epigenetic dysregulation in bladder cancer: A systematic review. *Cancer Treat. Rev.* 61, 82–93. doi: 10.1016/j.ctrv.2017.10.004

Catto, J. W. F., Azzouzi, A.-R., Rehman, I., Feeley, K. M., Cross, S. S., Amira, N., et al. (2005). Promoter hypermethylation is associated with tumor location, stage, and subsequent progression in transitional cell carcinoma. *J. Clin. Oncol.* 23, 2903–2910. doi: 10.1200/JCO.2005.03.163

Chalbatani, G. M., Dana, H., Memari, F., Gharagozlou, E., Ashjaei, S., Kheirandish, P., et al. (2019). Biological function and molecular mechanism of piRNA in cancer. *Pract. Lab. Med.* 13, 1–6. doi: 10.1016/j.plabm.2018.e00113

Chan, B., Manley, J., Lee, J., and Singh, S. R. (2015). The emerging roles of microRNAs in cancer metabolism. *Cancer Lett.* 356, 301–308. doi: 10.1016/j.canlet.2014.10.011

Chen, W., Wu, W., Li, X., Zhao, L., and Chen, W. (2011). HDAC6 and SIRT2 promote bladder cancer cell migration and invasion by targeting cortactin. *Oncol. Rep.* 27, 819–824. doi: 10.3892/or.2011.1553

Chen, X., Chen, R.-X., Wei, W.-S., Li, Y.-H., Feng, Z.-H., Tan, L., et al. (2018). PRMT5 Circular RNA Promotes Metastasis of Urothelial Carcinoma of the Bladder through Sponging miR-30c to Induce Epithelial–Mesenchymal Transition. *Clin. Cancer Res.* 24, 6319–6330. doi: 10.1158/1078-0432.CCR-18-1270

Chen, X., Wurtmann, E. J., Van Batavia, J., Zybailov, B., Washburn, M. P., and Wolin, S. L. (2007). An ortholog of the Ro autoantigen functions in 23S rRNA maturation in D. radiodurans. *Genes Dev.* 21, 1328–1339. doi: 10.1101/gad.1548207

Chen, Z., Du, Y., Liu, X., Chen, H., Weng, X., Guo, J., et al. (2019). EZH2 inhibition suppresses bladder cancer cell growth and metastasis *via the* JAK2/STAT3 signaling pathway. *Oncol. Lett.* 18, 907–915. doi: 10.3892/ol.2019.10359

Cheung, E. M., Quinn, D. I., Tsao-Wei, D. D., Groshen, S. G., Aparicio, A. M., Twardowski, P., et al. (2008). Phase II study of vorinostat (Suberoylanilide Hydroxamic Acid, SAHA) in patients with advanced transitional cell urothelial cancer (TCC) after platinum-based therapy—California Cancer Consortium/University of Pittsburgh NCI/CTEP-sponsored trial. *J. Clin. Oncol.* 26, 16058–16058. doi: 10.1200/jco.2008.26.15_suppl.16058

Cho, H.-S., Kelly, J. D., Hayami, S., Toyokawa, G., Takawa, M., Yoshimatsu, M., et al. (2015). Enhanced Expression of EHMT2 Is Involved in the Proliferation of Cancer Cells through Negative Regulation of SIAH1. *Neoplasia* 13, 676–IN10. doi: 10.1593/neo.11512

Chou, J., Shahi, P., and Werb, Z. (2013). microRNA-mediated regulation of the tumor microenvironment. *Cell Cycle* 12, 3262–3271. doi: 10.4161/cc.26087

Christov, C. P., Gardiner, T. J., Szuts, D., and Krude, T. (2006). Functional Requirement of Noncoding Y RNAs for Human Chromosomal DNA Replication. *Mol. Cell. Biol.* 26, 6993–7004. doi: 10.1128/mcb.01060-06

Christov, C. P., Trivier, E., and Krude, T. (2008). Noncoding human Y RNAs are overexpressed in tumours and required for cell proliferation. *Br. J. Cancer* 98, 981–988. doi: 10.1038/sj.bjc.6604254

Chu, H., Hui, G., Yuan, L., Shi, D., Wang, Y., Du, M., et al. (2015). Identification of novel piRNAs in bladder cancer. *Cancer Lett.* 356, 561–567. doi: 10.1016/j.canlet.2014.10.004

Chung, W., Bondaruk, J., Jelinek, J., Lotan, Y., Liang, S., Czerniak, B., et al. (2011). Detection of bladder cancer using novel DNA methylation biomarkers in urine sediments. *Cancer Epidemiol. Biomarkers Prev.* 20, 1483–1491. doi: 10.1158/1055-9965.EPI-11-0067

Clark, M. B., Johnston, R. L., Inostroza-Ponta, M., Fox, A. H., Fortini, E., Moscato, P., et al. (2012). Genome-wide analysis of long noncoding RNA stability. *Genome Res.* 22, 885–898. doi: 10.1101/gr.131037.111

Comito, G., Iscaro, A., Bacci, M., Morandi, A., Ippolito, L., Parri, M., et al. (2019). Lactate modulates CD4+ T-cell polarization and induces an immunosuppressive environment, which sustains prostate carcinoma progression *via* TLR8/miR21 axis. *Oncogene* 38, 3681–3695. doi: 10.1038/s41388-019-0688-7

Costa, V. L., Henrique, R., Danielsen, S. A., Duarte-Pereira, S., Eknaes, M., Skotheim, R. I., et al. (2010). Three epigenetic biomarkers, GDF15, TMEFF2, and VIM, accurately predict bladder cancer from DNA-based analyses of urine samples. *Clin. Cancer Res.* 16, 5842–5851. doi: 10.1158/1078-0432.CCR-10-1312

Cui, J., Sun, W., Hao, X., Wei, M., Su, X., Zhang, Y., et al. (2015). EHMT2 inhibitor BIX-01294 induces apoptosis through PMAIP1-USP9X-MCL1 axis in human bladder cancer cells. *Cancer Cell Int.* 15, 1–9. doi: 10.1186/s12935-014-0149-x

Cui, X., Jing, X., Long, C., Yi, Q., Tian, J., and Zhu, J. (2017). Accuracy of the urine UCA1 for diagnosis of bladder cancer: a meta-analysis. *Oncotarget* 8, 35222–35233. doi: 10.18632/oncotarget.16473

D'Andrea, D., Soria, F., Zehetmayer, S., Gust, K. M., Korn, S., Witjes, J. A., et al. (2019). Diagnostic accuracy, clinical utility and influence on decision-making of a methylation urine biomarker test in the surveillance of non-muscle-invasive bladder cancer. *BJU Int.* 123, 959–967. doi: 10.1111/bju.14673

Dall'Olio, F., and Trinchera, M. (2017). Epigenetic Bases of Aberrant Glycosylation in Cancer. *Int. J. Mol. Sci.* 18(5), 998. doi: 10.3390/ijms18050998

Datta, J., Ghoshal, K., Denny, W. A., Gamage, S. A., Darby, G., Phiasivongsa, P., et al. (2009). A New Class of Quinoline-Based DNA Hypomethylating Agents Reactivates Tumor Suppressor Genes by Blocking DNA Methyltransferase 1 Activity and Inducing Its Degradation. *Cancer Res.* 69 (10), 4277–4285. doi: 10.1158/0008-5472.CAN-08-3669

Dawson, M. A., and Kouzarides, T. (2012). Cancer epigenetics: from mechanism to therapy. *Cell* 150, 12–27. doi: 10.1016/j.cell.2012.06.013

De Palma, G., Di Lorenzo, V. F., Krol, S., and Paradiso, A. V. (2019). Urinary exosomal shuttle RNA: Promising cancer diagnosis biomarkers of lower urinary tract. *Int. J. Biol. Markers*, 34 (2), 101–107. doi: 10.1177/1724600819827023

Deb, G., Singh, A. K., and Gupta, S. (2014). EZH2: Not EZHY (Easy) to Deal. *Mol. Cancer Res.* 12, 639–653. doi: 10.1158/1541-7786.MCR-13-0546

Deroanne, C. F., Bonjean, K., Servotte, S., Devy, L., Colige, A., Clausse, N., et al. (2002). Histone deacetylases inhibitors as anti-angiogenic agents altering vascular endothelial growth factor signaling. *Oncogene* 21, 427–436. doi: 10.1038/sj.onc.1205108

Dhalluin, C., Carlson, J. E., Zeng, L., He, C., Aggarwal, A. K., Zhou, M.-M., et al. (1999). Structure and ligand of a histone acetyltransferase bromodomain. *Nature* 399, 491–496. doi: 10.1038/20974

Di Meo, A., Bartlett, J., Cheng, Y., Pasic, M. D., and Yousef, G. M. (2017). Liquid biopsy: A step forward towards precision medicine in urologic malignancies. *Mol. Cancer* 16, 1–14. doi: 10.1186/s12943-017-0644-5

Ding, B., Yan, L., Zhang, Y., Wang, Z., Zhang, Y., Xia, D., et al. (2019). Analysis of the role of mutations in the KMT2D histone lysine methyltransferase in bladder cancer. *FEBS Open Bio* 9, 693–706. doi: 10.1002/2211-5463.12600

Du, H., and Che, G. (2017). Genetic alterations and epigenetic alterations of cancer-associated fibroblasts. *Oncol. Lett.* 13, 3–12. doi: 10.3892/ol.2016.5451

Duex, J. E., Swain, K. E., Dancik, G. M., Paucek, R. D., Owens, C., Churchill, M. E. A., et al. (2018a). Functional Impact of Chromatin Remodeling Gene Mutations and Predictive Signature for Therapeutic Response in Bladder Cancer. *Mol. Cancer Res.* 16, 69–77. doi: 10.1158/1541-7786.MCR-17-0260

Duex, J. E., Swain, K. E., Dancik, G. M., Paucek, R. D., Owens, C., Churchill, M. E. A., et al. (2018b). Functional Impact of Chromatin Remodeling Gene Mutations and Predictive Signature for Therapeutic Response in Bladder Cancer. *Mol. Cancer Res.* 16, 69–77. doi: 10.1158/1541-7786.MCR-17-0260

Dunn, J., and Rao, S. (2017). Epigenetics and immunotherapy: The current state of play. *Mol. Immunol.* 87, 227–239. doi: 10.1016/j.molimm.2017.04.012

Eckert, M. A., Coscia, F., Chryplewicz, A., Chang, J. W., Hernandez, K. M., Pan, S., et al. (2019). Proteomics reveals NNMT as a master metabolic regulator of cancer-associated fibroblasts. *Nature* 569, 723–728. doi: 10.1038/s41586-019-1173-8

Eckschlager, T., Plch, J., Stiborova, M., and Hrabeta, J. (2017). Histone Deacetylase Inhibitors as Anticancer Drugs. *Int. J. Mol. Sci.* 18(7), 1414. doi: 10.3390/ijms18071414

Ellinger, J., Müller, S. C., and Dietrich, D. (2015). Epigenetic biomarkers in the blood of patients with urological malignancies. *Expert Rev. Mol. Diagn.* 15, 505–516. doi: 10.1586/14737159.2015.1019477

Ellinger, J., Schneider, A.-C., Bachmann, A., Kristiansen, G., Müller, S. C., and Rogenhofer, S. (2016). Evaluation of Global Histone Acetylation Levels in Bladder Cancer Patients. *Anticancer Res.* 36, 3961–3964.

Enkelmann, A., Heinzelmann, J., von Eggeling, F., Walter, M., Berndt, A., Wunderlich, H., et al. (2011). Specific protein and miRNA patterns characterise tumour-associated fibroblasts in bladder cancer. *J. Cancer Res. Clin. Oncol.* 137, 751–759. doi: 10.1007/s00432-010-0932-6

Enokida, H., Yoshino, H., Matsushita, R., and Nakagawa, M. (2016). The role of microRNAs in bladder cancer. *Investig. Clin. Urol.* 57, S60. doi: 10.4111/icu.2016.57.s1.s60

Esteller, M. (2008). Epigenetics in Cancer. *N. Engl. J. Med.* 358, 1148–1159. doi: 10.1056/NEJMra072067

Fabbri, M., and Calin, G. A. (2010). Epigenetics and miRNAs in Human Cancer. *Adv. Genet.* 70, 87–99. doi: 10.1016/B978-0-12-380866-0.60004-6

Fantony, J. J., Longo, T. A., Gopalakrishna, A., Owusu, R., Lance, R. S., Foo, W.-C., et al. (2017). Urinary NID2 and TWIST1 methylation to augment conventional urine cytology for the detection of bladder cancer. *Cancer Biomark.* 18, 381–387. doi: 10.3233/CBM-160261

Fardi, M., Solali, S., and Farshdousti Hagh, M. (2018). Epigenetic mechanisms as a new approach in cancer treatment: An updated review. *Genes Dis.* 5, 304–311. doi: 10.1016/j.gendis.2018.06.003

Feber, A., Dhami, P., Dong, L., de Winter, P., Tan, W. S., Martínez-Fernández, M., et al. (2017). UroMark—a urinary biomarker assay for the detection of bladder cancer. *Clin. Epigenet.* 9(1), 1–10. doi: 10.1186/s13148-016-0303-5

Feinberg, A. P., and Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 301, 89–92. doi: 10.1038/301089a0

Feng, Y., Liu, J., Kang, Y., He, Y., Liang, B., Yang, P., et al. (2014). miR-19a acts as an oncogenic microRNA and is up-regulated in bladder cancer. *J. Exp. Clin. Cancer Res.* 33, 67. doi: 10.1186/PREACCEPT-9242556491295527

Fischle, W. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev.* 17, 1870–1881. doi: 10.1101/gad.1110503

Freedman, J. E., Gerstein, M., Mick, E., Rozowsky, J., Levy, D., Kitchen, R., et al. (2016). Diverse human extracellular RNAs are widely detected in human plasma. *Nat. Commun.* 7, 11106. doi: 10.1038/ncomms11106

Gajer, J. M., Furdas, S. D., Gründer, A., Gothwal, M., Heinicke, U., Keller, K., et al. (2015). Histone acetyltransferase inhibitors block neuroblastoma cell growth *in vivo*. *Oncogenesis* 4, e137. doi: 10.1038/oncsis.2014.51

Gao, Z., Yuan, T., Zhou, X., Ni, P., Sun, G., Li, P., et al. (2018). Targeting BRD4 proteins suppresses the growth of NSCLC through downregulation of eIF4E expression. *Cancer Biol. Ther.* 19, 407–415. doi: 10.1080/15384047.2018.1423923

García-Baquero, R., Puerta, P., Beltran, M., Alvarez-Mújica, M., Alvarez-Ossorio, J. L., and Sánchez-Carbayo, M. (2014). Methylation of tumor suppressor genes in a novel panel predicts clinical outcome in paraffin-embedded bladder tumors. *Tumor Biol.* 35, 5777–5786. doi: 10.1007/s13277-014-1767-6

Garzon, R., Marcucci, G., and Croce, C. M. (2010). Targeting microRNAs in cancer: rationale, strategies and challenges. *Nat. Rev. Drug Discovery* 9, 775–789. doi: 10.1038/nrd3179

Ge, Q., Zhou, Y., Lu, J., Bai, Y., Xie, X., and Lu, Z. (2014). miRNA in Plasma Exosome is Stable under Different Storage Conditions. *Molecules* 19, 1568–1575. doi: 10.3390/molecules19021568

Giagounidis, A., Schoch, R., Gattermann, N., Sanz, G., List, A., Gore, S. D., et al. (2014). Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study. *Lancet Oncol.* 10, 223–232. doi: 10.1016/S1470-2045(09)70003-8.Efficacy

Gielchinsky, I., Gilon, M., Abu-lail, R., Matouk, I., Hochberg, A., Gofrit, O. N., et al. (2017). *H19* non-coding RNA in urine cells detects urothelial carcinoma: a pilot study. *Biomarkers* 22, 1–6. doi: 10.1080/1354750X.2016.1276625

Gillette, T. G., and Hill, J. A. (2015). Readers, writers, and erasers: chromatin as the whiteboard of heart disease. *Circ. Res.* 116, 1245–1253. doi: 10.1161/CIRCRESAHA.116.303630

Ginder, G. D., and Williams, D. C. (2018). Readers of DNA methylation, the MBD family as potential therapeutic targets. *Pharmacol. Ther.* 184, 98–111. doi: 10.1016/j.pharmthera.2017.11.002

Goll, M. G., and Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* 74, 481–514. doi: 10.1146/annurev.biochem.74.010904.153721

Gotwals, P., Cameron, S., Cipolletta, D., Cremasco, V., Crystal, A., Hewes, B., et al. (2017). Prospects for combining targeted and conventional cancer therapy with immunotherapy. *Nat. Rev. Cancer* 17, 286–301. doi: 10.1038/nrc.2017.17

Goulet, C. R., Champagne, A., Bernard, G., Vandal, D., Chabaud, S., Pouliot, F., et al. (2019). Cancer-associated fibroblasts induce epithelial-mesenchymal transition of bladder cancer cells through paracrine IL-6 signalling. *BMC Cancer* 19, 137. doi: 10.1186/s12885-019-5353-6

Grimm, S., Jennek, S., Singh, R., Enkelmann, A., Junker, K., Rippaus, N., et al. (2015). Malignancy of bladder cancer cells is enhanced by tumor-associated fibroblasts through a multifaceted cytokine-chemokine loop. *Exp. Cell Res.* 335, 1–11. doi: 10.1016/j.yexcr.2015.04.001

Guan, B., Xing, Y., Xiong, G., Cao, Z., Fang, D., Li, Y., et al. (2018). Predictive value of gene methylation for second recurrence following surgical treatment of first bladder recurrence of a primary upper-tract urothelial carcinoma. *Oncol. Lett.* 15, 9397–9405. doi: 10.3892/ol.2018.8498

Gui, Y., Guo, G., Huang, Y., Hu, X., Tang, A., Gao, S., et al. (2011). Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat. Genet.* 43, 875–878. doi: 10.1038/ng.907

Gulia, C., Baldassarra, S., Signore, F., Rigon, G., Pizzuti, V., Gaffi, M., et al. (2017). Role of Non-Coding RNAs in the Etiology of Bladder Cancer. *Genes (Basel)* 8, 339. doi: 10.3390/genes8110339

Guo, P., Zhang, G., Meng, J., He, Q., Li, Z., and Guan, Y. (2018). Upregulation of Long Noncoding RNA TUG1 Promotes Bladder Cancer Cell Proliferation, Migration, and Invasion by Inhibiting miR-29c. *Oncol. Res. Featur. Preclin. Clin. Cancer Ther.* 26, 1083–1091. doi: 10.3727/096504018X15152085755247

Gupta, P. B., Fillmore, C. M., Jiang, G., Shapira, S. D., Tao, K., Kuperwasser, C., et al. (2011). Stochastic State Transitions Give Rise to Phenotypic Equilibrium in Populations of Cancer Cells. *Cell* 146, 633–644. doi: 10.1016/j.cell.2011.07.026

Gupta, S., Doyle, K., Mosbruger, T. L., Butterfield, A., Ast, A., Kaadige, M., et al. (2018). Reversible LSD1 inhibition with HCI-2509 induces the p53 gene expression signature and disrupts the MYCN signature in high- risk neuroblastoma cells. *Oncotarget* 9 (11), 9907–9924.

Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464, 1071–1076. doi: 10.1038/nature08975

Han, C., Seebacher, N. A., Hornicek, F. J., Kan, Q., and Duan, Z. (2017). Regulation of microRNAs function by circular RNAs in human cancer. *Oncotarget* 8, 64622–64637. doi: 10.18632/oncotarget.19930

Han, H., Wolff, E. M., and Liang, G. (2012). Epigenetic Alterations in Bladder Cancer and Their Potential Clinical Implications. *Adv. Urol.* 2012, 1–11. doi: 10.1155/2012/546917

Hashizume, H., Baluk, P., Morikawa, S., McLean, J. W., Thurston, G., Roberge, S., et. al (2000). Openings between Defective Endothelial Cells Explain Tumor Vessel Leakiness. *Am. J. Pathol.* 156, 1363–1380. doi: 10.1016/S0002-9440(10)65006-7

Havel, J. J., Chowell, D., and Chan, T. A. (2019). The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat. Rev. Cancer* 19, 133–150. doi: 10.1038/s41568-019-0116-x

Hayami, S., Yoshimatsu, M., Veerakumarasivam, A., Unoki, M., Iwai, Y., Tsunoda, T., et al. (2010). Overexpression of the JmjC histone demethylase KDM5B in human carcinogenesis: Involvement in the proliferation of cancer cells through the E2F/RB pathway. *Mol. Cancer* 9, 1–14. doi: 10.1186/1476-4598-9-59

Heitzer, E., Perakis, S., Geigl, J. B., and Speicher, M. R. (2017). The potential of liquid biopsies for the early detection of cancer. *NPJ Precis. Oncol.* 1 (36). doi: 10.1038/s41698-017-0039-5

Hellebrekers, D. M. E. I., Jair, K.-W., Viré, E., Eguchi, S., Hoebers, N. T. H., Fraga, M. F., et al. (2006). Angiostatic activity of DNA methyltransferase inhibitors. *Mol. Cancer Ther.* 5, 467–475. doi: 10.1158/1535-7163.MCT-05-0417

Helming, K. C., Wang, X., Wilson, B. G., Vazquez, F., Haswell, J. R., Manchester, H. E., et al. (2014). ARID1B is a specific vulnerability in ARID1A-mutant cancers. *Nat. Med.* 20, 251–254. doi: 10.1038/nm.3480

Hendrich, B., and Bird, A. (2015). Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. *Mol. Cell. Biol.* 18, 6538–6547. doi: 10.1128/mcb.18.11.6538

Herold, J. M., James, L. I., Korboukh, V. K., Gao, C., Coil, K. E., Bua, D. J., et al. (2012). Structure–activity relationships of methyl-lysine reader antagonists. *Med. Chem. Commun.* 3, 45–51. doi: 10.1039/C1MD00195G

Hicks, K. C., Fantini, M., Donahue, R. N., Schwab, A., Knudson, K. M., Tritsch, S. R., et al. (2018). Epigenetic priming of both tumor and NK cells augments antibody-dependent cellular cytotoxicity elicited by the anti-PD-L1 antibody avelumab against multiple carcinoma cell types. *Oncoimmunology* 7, e1466018. doi: 10.1080/2162402X.2018.1466018

Hida, K., Hida, Y., Amin, D. N., Flint, A. F., Panigrahy, D., Morton, C. C., et al. (2004). Tumor-Associated Endothelial Cells with Cytogenetic Abnormalities. *Cancer Res.* 64, 8249–8255. doi: 10.1158/0008-5472.CAN-04-1567

Højfeldt, J. W., Agger, K., and Helin, K. (2013). Histone lysine demethylases as targets. *Nat. Publ. Gr.* 12, 917–930. doi: 10.1038/nrd4154

Hölscher, A. S., Schulz, W. A., Pinkerneil, M., Niegisch, G., and Hoffmann, M. J. (2018). Combined inhibition of BET proteins and class I HDACs synergistically induces apoptosis in urothelial carcinoma cell lines. *Clin. Epigenet.* 10, 1–14. doi: 10.1186/s13148-017-0434-3

Homami, A., and Ghazi, F. (2016). MicroRNAs as biomarkers associated with bladder cancer. *Med. J. Islam. Repub. Iran* 30.

Hopkinson, R.J., Tumber, A., Yapp, C., Chowdhury, R., Aik, W.S., Che, K.H., et al. (2013). 5-Carboxy-8-hydroxyquinoline is a broad spectrum 2-oxoglutarate oxygenase inhibitor which causes iron translocation. *Chem Sci.* 4 (8), 3110–3117. doi: 10.1039/C3SC51122G

Hu, L., Zang, M., Wang, H.-X., Zhang, B.-G., Wang, Z.-Q., Fan, Z.-Y., et al. (2018). G9A promotes gastric cancer metastasis by upregulating ITGB3 in a SET domain-independent manner. *Cell Death Dis.* 9, 278. doi: 10.1038/s41419-018-0322-6

Hu, M., Yao, J., Cai, L., Bachman, K. E., van den Brûle, F., Velculescu, V., et al. (2005). Distinct epigenetic changes in the stromal cells of breast cancers. *Nat. Genet.* 37, 899–905. doi: 10.1038/ng1596

Huang, C., Liao, X., Jin, H., Xie, F., Zheng, F., Li, J., et al. (2019a). MEG3, as a Competing Endogenous RNA, Binds with miR-27a to Promote PHLPP2 Protein Translation and Impairs Bladder Cancer Invasion. *Mol. Ther. Nucleic Acids* 16, 51–62. doi: 10.1016/j.omtn.2019.01.014

Huang, M., Zhong, Z., Lv, M., Shu, J., Tian, Q., and Chen, J. (2016). Comprehensive analysis of differentially expressed profiles of lncRNAs and circRNAs with associated co-expression and ceRNA networks in bladder carcinoma. *Oncotarget* 7, 47186–47200. doi: 10.18632/oncotarget.9706

Huang, Z., Zhang, M., Chen, G., Wang, W., Zhang, P., Yue, Y., et al. (2019b). Bladder cancer cells interact with vascular endothelial cells triggering EGFR signals to promote tumor progression. *Int. J. Oncol.* 54, 1555–1566. doi: 10.3892/ijo.2019.4729

Hulshoff, M. S., Xu, X., Krenning, G., and Zeisberg, E. M. (2018). Epigenetic Regulation of Endothelial-to-Mesenchymal Transition in Chronic Heart Disease. *Arterioscler. Thromb. Vasc. Biol.* 38, 1986–1996. doi: 10.1161/ATVBAHA.118.311276

Hurst, C. D., Alder, O., Platt, F. M., Droop, A., Stead, L. F., Burns, J. E., et al. (2017). Genomic Subtypes of Non-invasive Bladder Cancer with Distinct Metabolic Profile and Female Gender Bias in KDM6A Mutation Frequency. *Cancer Cell* 32, 701–715.e7. doi: 10.1016/j.ccell.2017.08.005

Hyun, K., Jeon, J., Park, K., and Kim, J. (2017). Writing, erasing and reading histone lysine methylations. *Exp. Mol. Med.* 49, e324–e324. doi: 10.1038/emm.2017.11

Iliev, R., Fedorko, M., Machackova, T., Mlcochova, H., Svoboda, M., Pacik, D., et al. (2016). Expression Levels of PIWI-interacting RNA, piR-823, Are Deregulated in Tumor Tissue, Blood Serum and Urine of Patients with Renal Cell Carcinoma. *Anticancer Res.* 36, 6419–6423. doi: 10.21873/anticanres.11239

Ingelmo-Torres, M., Lozano, J. J., Izquierdo, L., Carrion, A., Costa, M., Gomez, L., et al. (2017). Urinary cell microRNA-based prognostic classifier for non-muscle invasive bladder cancer. *Oncotarget* 8, 18238–18247. doi: 10.18632/oncotarget.15315

Jabłonowski, Z., Reszka, E., Gromadzińska, J., Wąsowicz, W., and Sosnowski, M. (2011). Hypermethylation of p16 and DAPK promoter gene regions in patients with non-invasive urinary bladder cancer. *Arch. Med. Sci.* 3, 512–516. doi: 10.5114/aoms.2011.23421

Janson, P. C. J., Marits, P., Thörn, M., Ohlsson, R., Winqvist, O., Gerard, J. R., et al. (2008). CpG methylation of the IFNG gene as a mechanism to induce immunosuppression [correction of immunosupression] in tumor-infiltrating lymphocytes. *J. Immunol.* 181, 2878–2886. doi: 10.4049/jimmunol.181.4.2878

Jeck, W. R., and Sharpless, N. E. (2014). Detecting and characterizing circular RNAs. *Nat. Biotechnol.* 32, 453–461. doi: 10.1038/nbt.2890

Jerónimo, C., and Henrique, R. (2014). Epigenetic biomarkers in urological tumors: A systematic review. *Cancer Lett.* 342, 264–274. doi: 10.1016/j.canlet.2011.12.026

Jiang, L., Gonda, T. A., Gamble, M. V., Salas, M., Seshan, V., Tu, S., et al. (2008). Global Hypomethylation of Genomic DNA in Cancer-Associated

Myofibroblasts. *Cancer Res.* 68, 9900–9908. doi: 10.1158/0008-5472.CAN-08-1319

Jiao, D., Li, Z., Zhu, M., Wang, Y., Wu, G., and Han, X. (2018). LncRNA MALAT1 promotes tumor growth and metastasis by targeting miR-124/foxq1 in bladder transitional cell carcinoma (BTCC). *Am. J. Cancer Res.* 8, 748–760.

Jing, J., Liu, J., Wang, Y., Zhang, M., Yang, L., Shi, F., et al. (2018). The role of ZBTB38 in promoting migration and invasive growth of bladder cancer cells. *Oncol. Rep*. 41 (3), 1980–1990. doi: 10.3892/or.2018.6937

Jones, P. A., and Liang, G. (2009). Rethinking how DNA methylation patterns are maintained. *Nat. Rev. Genet*. 10, 805–811. doi: 10.1038/nrg2651

José-Enériz, E. S., Agirre, X., Rabal, O., Vilas-Zornoza, A., Sanchez-Arias, J. A., Miranda, E., et al. (2017). Discovery of first-in-class reversible dual small molecule inhibitors against G9a and DNMTs in hematological malignancies. *Nat. Commun*. 8, 15424. doi: 10.1038/ncomms15424

Josson, S., Gururajan, M., Sung, S. Y., Hu, P., Shao, C., Zhau, H. E., et al. (2015). Stromal fibroblast-derived miR-409 promotes epithelial-to-mesenchymal transition and prostate tumorigenesis. *Oncogene* 34, 2690–2699. doi: 10.1038/onc.2014.212

Kaletsch, A., Pinkerneil, M., Hoffmann, M. J., Jaguva Vasudevan, A. A., Wang, C., Hansen, F. K., et al. (2018). Effects of novel HDAC inhibitors on urothelial carcinoma cells. *Clin. Epigenet.* 10, 1–18. doi: 10.1186/s13148-018-0531-y

Kalluri, R. (2016). The biology and function of fibroblasts in cancer. *Nat. Rev. Cancer* 16, 582–598. doi: 10.1038/nrc.2016.73

Kandimalla, R., van Tilborg, A. A. G., Kompier, L. C., Stumpel, D. J. P. M., Stam, R. W., Bangma, C. H., et al. (2012). Genome-wide analysis of CpG island methylation in bladder cancer identified TBX2, TBX3, GATA2, and ZIC4 as pTa-specific prognostic markers. *Eur. Urol.* 61, 1245–1256. doi: 10.1016/j.eururo.2012.01.011

Kauffman, E. C., Robinson, B. D., Downes, M., Powell, L. G., Ming, M., Scherr, D. S., et al. (2012). Role of androgen receptor and associated lysine-demethylase coregulators, LSD1 and JMJD2A, in localized and advanced human bladder cancer. *Mol. Carcinog.* 50, 931–944. doi: 10.1002/mc.20758.Role

Khetrapal, P., Lee, M. W. L., Tan, W. S., Dong, L., de Winter, P., Feber, A., et al. (2018). The role of circulating tumour cells and nucleic acids in blood for the detection of bladder cancer: A systematic review. *Cancer Treat. Rev.* 66, 56–63. doi: 10.1016/J.CTRV.2018.03.007

Kiany, S., Huang, G., and Kleinerman, E. S. (2017). Effect of entinostat on NK cell-mediated cytotoxicity against osteosarcoma cells and osteosarcoma lung metastasis. *Oncoimmunology* 6, e1333214. doi: 10.1080/2162402X.2017.1333214

Kim, D. E., Procopio, M.-G., Ghosh, S., Jo, S.-H., Goruppi, S., Magliozzi, F., et al. (2017). Convergent roles of ATF3 and CSL in chromatin control of cancer-associated fibroblast activation. *J. Exp. Med*. 214, 2349–2368. doi: 10.1084/jem.20170724

Kim, D. J., Dunleavey, J. M., Xiao, L., Ollila, D. W., Troester, M. A., Otey, C. A., et. al (2018). Suppression of TGFβ-mediated conversion of endothelial cells and fibroblasts into cancer associated (myo)fibroblasts *via* HDAC inhibition. *Br. J. Cancer* 118, 1359–1368. doi: 10.1038/s41416-018-0072-3

Kim, J., Akbani, R., Creighton, C. J., Lerner, S. P., Weinstein, J. N., Getz, G., et al. (2015a). Invasive Bladder Cancer: Genomic Insights and Therapeutic Promise. *Clin. Cancer Res.* 21, 4514–4524. doi: 10.1158/1078-0432.CCR-14-1215

Kim, K. H., Kim, W., Howard, T. P., Vazquez, F., Tsherniak, A., Wu, J. N., et al. (2015b). SWI/SNF-mutant cancers depend on catalytic and non-catalytic activity of EZH2. *Nat. Med.* 21, 1491–1496. doi: 10.1038/nm.3968

Kim, M. S., Kwon, H. J., Lee, Y. M., Baek, J. H., Jang, J.-E., Lee, S.-W., et al. (2001). Histone deacetylases induce angiogenesis by negative regulation of tumor suppressor genes. *Nat. Med.* 7, 437–443. doi: 10.1038/86507

Kim, S. M., Kang, H. W., Kim, W. T., Kim, Y.-J., Yun, S. J., Lee, S.-C., et al. (2013). Cell-Free microRNA-214 From Urine as a Biomarker for Non-Muscle-Invasive Bladder Cancer. *Korean J. Urol.* 54, 791. doi: 10.4111/kju.2013.54.11.791

Kitajima, H., Itoh, N., Shindo, T., Maruyama, R., Nojima, M., Suzuki, H., et al. (2017). Evaluation of Urinary DNA Methylation as a Marker for Recurrent Bladder Cancer: A 2-Center Prospective Study. *Urology* 113, 71–78. doi: 10.1016/j.urology.2017.11.025

Kitchen, M. O., Bryan, R. T., Emes, R. D., Luscombe, C. J., Cheng, K., Zeegers, M. P., et al. (2018). HumanMethylation450K Array–Identified Biomarkers Predict Tumour Recurrence/Progression at Initial Diagnosis of High-risk Non-muscle Invasive Bladder Cancer. *Biomark. Cancer* 10, 1179299X1775192. doi: 10.1177/1179299x17751920

Kitchen, M. O., Bryan, R. T., Haworth, K. E., Emes, R. D., Luscombe, C., Gommersall, L., et al. (2015). Methylation of HOXA9 and ISL1 Predicts Patient Outcome in High-Grade Non-Invasive Bladder Cancer. *PloS One* 10, e0137003. doi: 10.1371/journal.pone.0137003

Knowles, M. A., and Hurst, C. D. (2015). Molecular biology of bladder cancer: New insights into pathogenesis and clinical diversity. *Nat. Rev. Cancer* 15, 25–41. doi: 10.1038/nrc3817

Kogo, R., Shimamura, T., Mimori, K., Kawahara, K., Imoto, S., Sudo, T., et al. (2011). Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res.* 71, 6320–6326. doi: 10.1158/0008-5472.CAN-11-1021

Kohlhapp, F. J., Mitra, A. K., Lengyel, E., and Peter, M. E. (2015). MicroRNAs as mediators and communicators between cancer cells and the tumor microenvironment. *Oncogene* 34, 5857–5868. doi: 10.1038/onc.2015.89

Komatsu, Y., Waku, T., Iwasaki, N., Ono, W., Yamaguchi, C., and Yanagisawa, J. (2012). Global analysis of DNA methylation in early-stage liver fibrosis. *BMC Med. Genomics* 5, 5. doi: 10.1186/1755-8794-5-5

Konze, K. D., Macnevin, C. J., Liu, F., Gao, C., Huang, X., Kuznetsova, E., et al. (2014). An orally bioavailable chemical probe of the lysine methyltransferase ezh2 and ezh1. *ACS Chemical Biology* 8 (6), 1324–1334. doi: 10.1021/cb400133j

Korpal, M., Lee, E. S., Hu, G., and Kang, Y. (2008). The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2. *J. Biol. Chem.* 283, 14910–14914. doi: 10.1074/jbc.C800074200

Kouzarides, T. (2007). Chromatin Modifications and Their Function. *Cell* 128, 693–705. doi: 10.1016/j.cell.2007.02.005

Kreimer, U., Schulz, W. A., Koch, A., Niegisch, G., and Goering, W. (2013). HERV-K and LINE-1 DNA Methylation and Reexpression in Urothelial Carcinoma. *Front. Oncol.* 3, 255. doi: 10.3389/fonc.2013.00255

Kristensen, L. S., Hansen, T. B., Venø, M. T., and Kjems, J. (2018). Circular RNAs in cancer: Opportunities and challenges in the field. *Oncogene* 37, 555–565. doi: 10.1038/onc.2017.361

Kubicek, S., Sullivan, R. J. O., August, E. M., Hickey, E. R., Zhang, Q., Teodoro, M. L., et al., (2007). Technique Reversal of H3K9me2 by a Small-Molecule Inhibitor for the G9a Histone Methyltransferase. *Mol. Cell* 1, 473–481. doi: 10.1016/j.molcel.2007.01.017

Kuck, D., Caulfield, T., Lyko, F., and Medina-franco, J. L., (2010). Nanaomycin A Selectively Inhibits DNMT3B and Reactivates Silenced Tumor Suppressor Genes in Human Cancer Cells. *Mol. Cancer Ther.* 9 (11), 3015–3024. doi: 10.1158/1535-7163.MCT-10-0609

Kulcheski, F. R., Christoff, A. P., and Margis, R. (2016). Circular RNAs are miRNA sponges and can be used as a new class of biomarker. *J. Biotechnol.* 238, 42–51. doi: 10.1016/j.jbiotec.2016.09.011

Kuninty, P. R., Schnittert, J., Storm, G., and Prakash, J. (2016). MicroRNA Targeting to Modulate Tumor Microenvironment. *Front. Oncol.* 6, 3. doi: 10.3389/fonc.2016.00003

Kurdyukov, S., and Bullock, M. (2016). DNA methylation analysis: Choosing the right method. *Biology (Basel)* 5 (1), 3. doi: 10.3390/biology5010003

Lamprecht, S., Sigal-Batikoff, I., Shany, S., Abu-Freha, N., Ling, E., Delinasios, G. J., et al. (2018). Teaming up for trouble: Cancer cells, transforming growth factor-β1 signaling and the epigenetic corruption of stromal naïve fibroblasts. *Cancers* 10(61). doi: 10.3390/cancers10030061

Laurent, V. M., Duperray, A., Sundar Rajan, V., and Verdier, C. (2014). Atomic Force Microscopy Reveals a Role for Endothelial Cell ICAM-1 Expression in Bladder Cancer Cell Adherence. *PloS One* 9, e98034. doi: 10.1371/journal.pone.0098034

Lee, J.-E., Wang, C., Xu, S., Cho, Y.-W., Wang, L., Feng, X., et al. (2013). H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. *Elife* 2, e01503. doi: 10.7554/eLife.01503

Lee, J.-Y., Lee, S.-H., Heo, S.-H., Kim, K.-S., Kim, C., Kim, D.-K., et al. (2015). Novel Function of Lysine Methyltransferase G9a in the Regulation of Sox2 Protein Stability. *PloS One* 10, e0141118. doi: 10.1371/journal.pone.0141118

Lee, K.-H., and Song, C. G. (2017). Epigenetic regulation in bladder cancer: development of new prognostic targets and therapeutic implications. *Transl. Cancer Res.* 6, S677–S688. doi: 10.21037/14111

Lee, M. G., Villa, R., Trojer, P., Norman, J., Yan, K.-P., Reinberg, D., et al. (2007). Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science* 318, 447–450. doi: 10.1126/science.1149042

Ler, L. D., Ghosh, S., Chai, X., Thike, A. A., Heng, H. L., Siew, E. Y., et al. (2017). Loss of tumor suppressor KDM6A amplifies PRC2-regulated transcriptional repression in bladder cancer and can be targeted through inhibition of EZH2. *Sci. Transl. Med.* 9, eaai8312. doi: 10.1126/scitranslmed.aai8312

Li, B., Xie, F., Zheng, F., Jiang, G., Zeng, F., and Xiao, X. (2017). Overexpression of CircRNA BCRC4 regulates cell apoptosis and MicroRNA-101/EZH2 signaling in bladder cancer. *Curr. Med. Sci.* 37, 886–890. doi: 10.1007/s11596-017-1822-9

Li, F., Yang, C., Zhang, H.-B., Ma, J., Jia, J., Tang, X., et al. (2019). BET inhibitor JQ1 suppresses cell proliferation *via* inducing autophagy and activating LKB1/AMPK in bladder cancer cells. *Cancer Med.* 8 (10) 1–14. doi: 10.1002/cam4.2385

Li, H.-T., Duymich, C. E., Weisenberger, D. J., and Liang, G. (2016a). Genetic and Epigenetic Alterations in Bladder Cancer. *Int. Neurourol. J.* 20, S84–S94. doi: 10.5213/inj.1632752.376

Li, P., Shan, J.-X., Chen, X.-H., Zhang, D., Su, L.-P., Huang, X.-Y., et al. (2015). Epigenetic silencing of microRNA-149 in cancer-associated fibroblasts mediates prostaglandin E2/interleukin-6 signaling in the tumor microenvironment. *Cell Res.* 25, 588–603. doi: 10.1038/cr.2015.51

Li, P., Yang, X., Yuan, W., Yang, C., Zhang, X., Han, J., et al. (2018a). CircRNA-Cdr1as Exerts Anti-Oncogenic Functions in Bladder Cancer by Sponging MicroRNA-135a. *Cell. Physiol. Biochem.* 46, 1606–1616. doi: 10.1159/000489208

Li, Q. Q., Hao, J. J., Zhang, Z., Hsu, I., Liu, Y., Tao, Z., et al. (2016b). Histone deacetylase inhibitor-induced cell death in bladder cancer is associated with chromatin modification and modifying protein expression: A proteomic approach. *Int. J. Oncol.* 48, 2591–2607. doi: 10.3892/ijo.2016.3478

Li, X.-Y., Zhang, K., Li, X.-S., Xiong, G.-Y., He, Q., Yang, K.-W., et al. (2018b). Detection of urothelial carcinoma, upper tract urothelial carcinoma, bladder carcinoma, and urothelial carcinoma with gross hematuria using selected urine-DNA methylation biomarkers: A prospective, single-center study. *Urol. Oncol. Semin. Orig. Investig.* 36, 342.e15–342.e23. doi: 10.1016/j.urolonc.2018.04.001

Liang, Y., Ridzon, D., Wong, L., and Chen, C. (2007). Characterization of microRNA expression profiles in normal human tissues. *BMC Genomics* 8, 166. doi: 10.1186/1471-2164-8-166

Liep, J., Rabien, A., and Jung, K. (2012). Feedback networks between microRNAs and epigenetic modifications in urological tumors. *Epigenetics* 7, 315–325. doi: 10.4161/epi.19464

Lima, J. F., Carvalho, J., Pinto-Ribeiro, I., Almeida, C., Wengel, J., Cerqueira, L., et al. (2018a). Targeting miR-9 in gastric cancer cells using locked nucleic acid oligonucleotides. *BMC Mol. Biol.* 19, 1–13. doi: 10.1186/s12867-018-0107-6

Lima, J. F., Cerqueira, L., Figueiredo, C., Oliveira, C., and Azevedo, N. F. (2018b). Anti-miRNA oligonucleotides: A comprehensive guide for design. *RNA Biol.* 15, 338–352. doi: 10.1080/15476286.2018.1445959

Lin, Y., Li, Z., He, Z., Guan, T., and Ma, J. (2012). Clinical and Prognostic Significance of Protocadherin-10 (PCDH10) Promoter Methylation in Bladder Cancer. *J. Int. Med. Res.* 40, 2117–2123. doi: 10.1177/030006051204000609

Liu, C., Chen, Z., Fang, J., Xu, A., Zhang, W., and Wang, Z. (2016). H19-derived miR-675 contributes to bladder cancer cell proliferation by regulating p53 activation. *Tumour Biol.* 37, 263–270. doi: 10.1007/s13277-015-3779-2

Liu, G., Zhao, X., Zhou, J., Cheng, X., Ye, Z., and Ji, Z. (2018a). Long non-coding RNA MEG3 suppresses the development of bladder urothelial carcinoma by regulating miR-96 and TPM1. *Cancer Biol. Ther.* 19, 1039–1056. doi: 10.1080/15384047.2018.1480279

Liu, H., Bi, J., Dong, W., Yang, M., Shi, J., Jiang, N., et al. (2018b). Invasion-related circular RNA circFNDC3B inhibits bladder cancer progression through the miR-1178-3p/G3BP2/SRC/FAK axis. *Mol. Cancer* 17, 161. doi: 10.1186/s12943-018-0908-8

Liu, M., Zhou, J., Chen, Z., and Cheng, A. S.-L. (2017a). Understanding the epigenetic regulation of tumours and their microenvironments: opportunities and problems for epigenetic therapy. *J. Pathol.* 241, 10–24. doi: 10.1002/path.4832

Liu, Q., Liu, H., Cheng, H., Li, Y., Li, X., and Zhu, C. (2017b). Downregulation of long noncoding RNA TUG1 inhibits proliferation and induces apoptosis through the TUG1/miR-142/ZEB2 axis in bladder cancer cells. *Oncol. Targets Ther.* 10, 2461–2471. doi: 10.2147/OTT.S124595

Lodewijk, I., Dueñas, M., Rubio, C., Munera-Maravilla, E., Segovia, C., Bernardini, A., et al. (2018). Liquid Biopsy Biomarkers in Bladder Cancer: A Current Need for Patient Diagnosis and Monitoring. *Int. J. Mol. Sci.* 19, 2514. doi: 10.3390/ijms19092514

Lu, C., Han, H. D., Mangala, L. S., Ali-Fehmi, R., Newton, C. S., Ozbun, L., et al. (2010). Regulation of Tumor Angiogenesis by EZH2. *Cancer Cell* 18, 185–197. doi: 10.1016/j.ccr.2010.06.016

Lu, C., Paschall, A. V., Shi, H., Savage, N., Waller, J. L., Sabbatini, M. E., et al. (2017). The MLL1-H3K4me3 axis-mediated PD-L1 expression and pancreatic cancer immune evasion. *J. Natl. Cancer Inst.* 109, djw283. doi: 10.1093/jnci/djw283

Lu, M., Selleslag, D., Labar, B., Germing, U., Salih, H. R., Beeldens, F., et al., (2011). Low-Dose Decitabine Versus Best Supportive Care in Elderly Patients With Intermediate- or High-Risk Myelodysplastic Syndrome (MDS) Ineligible for Intensive Chemotherapy : Final Results of the Randomized Phase III Study of the European Organisation for R. *J. Clin. Oncol.* 29, 1987–1996. doi: 10.1200/JCO.2010.30.9245

Luo, J., Chen, J., Li, H., Yang, Y., Yun, H., Yang, S., et al. (2017). LncRNA UCA1 promotes the invasion and EMT of bladder cancer cells by regulating the miR-143/HMGB1 pathway. *Oncol. Lett.* 14, 5556–5562. doi: 10.3892/ol.2017.6886

Luo, Z.-G., Li, Z.-G., Gui, S.-L., Chi, B.-J., and Ma, J.-G. (2014). Protocadherin-17 promoter methylation in serum-derived DNA is associated with poor prognosis of bladder cancer. *J. Int. Med. Res.* 42, 35–41. doi: 10.1177/0300060513504705

Luperchio, T. R., Wong, X., and Reddy, K. L. (2014). Genome regulation at the peripheral zone: lamina associated domains in development and disease. *Curr. Opin. Genet. Dev.* 25, 50–61. doi: 10.1016/J.GDE.2013.11.021

Maier, V. K., Feeney, C. M., Taylor, J. E., Creech, A. L., Qiao, J. W., Szanto, A., et al. (2015). Functional Proteomic Analysis of Repressive Histone Methyltransferase Complexes Reveals ZNF518B as a G9A Regulator. *Mol. Cell. Proteomics* 14, 1435–1446. doi: 10.1074/mcp.M114.044586

Majewski, T., Yao, H., Bondaruk, J., Chung, W., Lee, S., Lee, J. G., et al. (2019). Whole-Organ Genomic Characterization of Mucosal Field Effects Initiating Bladder Carcinogenesis. *Cell Rep.* 26, 2241–2256.e4. doi: 10.1016/j.celrep.2019.01.095

Mann, B. S., Johnson, J. R., Cohen, M. H., Justice, R., and Pazdur, R. (2007). FDA Approval Summary: Vorinostat for Treatment of Advanced Primary Cutaneous T-Cell Lymphoma. *Oncologist* 12 (10), 1247–1252. doi: 10.1634/theoncologist.12-10-1247

Marks, D. L., Olson, R. L., and Fernandez-Zapico, M. E. (2016). Epigenetic control of the tumor microenvironment. *Epigenomics* 8, 1671–1687. doi: 10.2217/epi-2016-0110

Marques-Magalhaes, A., Graca, I., Henrique, R., and Jeronimo, C. (2018). Targeting DNA Methyltranferases in Urological Tumors. *Front. Pharmacol.* 9, 366. doi: 10.3389/fphar.2018.00366

Martens-Uzunova, E. S., Böttcher, R., Croce, C. M., Jenster, G., Visakorpi, T., and Calin, G. A. (2014). Long noncoding RNA in prostate, bladder, and kidney cancer. *Eur. Urol.* 65, 1140–1151. doi: 10.1016/j.eururo.2013.12.003

Martínez-Fernández, M., Dueñas, M., Feber, A., Segovia, C., García-Escudero, R., Rubio, C., et al. (2015a). A Polycomb-mir200 loop regulates clinical outcome in bladder cancer. *Oncotarget* 6, 42258–42275. doi: 10.18632/oncotarget.5546

Martínez-Fernández, M., Feber, A., Dueñas, M., Segovia, C., Rubio, C., Fernandez, M., et al. (2015b). Analysis of the Polycomb-related lncRNAs HOTAIR and ANRIL in bladder cancer. *Clin. Epigenet.* 7, 109. doi: 10.1186/s13148-015-0141-x

Martínez-Fernández, M., Paramio, J. M., and Dueñas, M. (2016). RNA Detection in Urine. *J. Mol. Diagnostics* 18, 15–22. doi: 10.1016/j.jmoldx.2015.07.008

Martínez-Fernández, M., Rubio, C., Segovia, C., López-Calderón, F. F., Dueñas, M., and Paramio, J. M. (2015c). EZH2 in bladder cancer, a promising therapeutic target. *Int. J. Mol. Sci.* 16, 27107–27132. doi: 10.3390/ijms161126000

Martinez, V. D., Vucic, E. A., Thu, K. L., Hubaux, R., Enfield, K. S. S., Pikor, L. A., et al. (2015). Unique somatic and malignant expression patterns implicate PIWI-interacting RNAs in cancer-type specific biology. *Sci. Rep.* 5, 10423. doi: 10.1038/srep10423

Mijnes, J., Veeck, J., Gaisa, N. T., Burghardt, E., de Ruijter, T. C., Gostek, S., et al. (2018). Promoter methylation of DNA damage repair (DDR) genes in human tumor entities: RBBP8/CtIP is almost exclusively methylated in bladder cancer. *Clin. Epigenet.* 10, 1–20. doi: 10.1186/s13148-018-0447-6

Mio, C., Bulotta, S., Russo, D., and Damante, G. (2019). Reading Cancer: Chromatin Readers as Druggable Targets for Cancer Treatment. *Cancers (Basel)* 11, 61. doi: 10.3390/cancers11010061

Mishra, R., Haldar, S., Placencio, V., Madhav, A., Rohena-Rivera, K., Agarwal, P., et al. (2018). Stromal epigenetic alterations drive metabolic and neuroendocrine prostate cancer reprogramming. *J. Clin. Invest.* 128, 4472–4484. doi: 10.1172/JCI99397

Misso, G., Di Martino, M. T., De Rosa, G., Farooqi, A. A., Lombardi, A., Campani, V., et al. (2014). Mir-34: A new weapon against cancer? *Mol. Ther. Nucleic Acids* 3, e195. doi: 10.1038/mtna.2014.47

Moore, L. E., Pfeiffer, R. M., Poscablo, C., Real, F. X., Kogevinas, M., Silverman, D., et al. (2008). Genomic DNA hypomethylation as a biomarker for bladder cancer susceptibility in the Spanish Bladder Cancer Study: a case-control study. *Lancet Oncol.* 9, 359–366. doi: 10.1016/S1470-2045(08)70038-X

Morini, M. F., Giampietro, C., Corada, M., Pisati, F., Lavarone, E., Cunha, S. I., et al (2018). VE-Cadherin–Mediated Epigenetic Regulation of Endothelial Gene Expression. *Circ. Res.* 122, 231–245. doi: 10.1161/CIRCRESAHA.117.312392

Mozzetta, C., Pontis, J., and Ait-Si-Ali, S. (2015). Functional Crosstalk Between Lysine Methyltransferases on Histone Substrates: The Case of G9A/GLP and Polycomb Repressive Complex 2. *Antioxid. Redox Signal.* 22, 1365–1381. doi: 10.1089/ars.2014.6116

Mozzetta, C., Pontis, J., Fritsch, L., Robin, P., Portoso, M., Proux, C., et al. (2014). The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing. *Mol. Cell* 53, 277–289. doi: 10.1016/j.molcel.2013.12.005

Nagai, N., Ohguchi, H., Nakaki, R., Matsumura, Y., Kanki, Y., Sakai, J., et al. (2018). Downregulation of ERG and FLI1 expression in endothelial cells triggers endothelial-to-mesenchymal transition. *PloS Genet.* 14, e1007826. doi: 10.1371/journal.pgen.1007826

Nagano, T., Mitchell, J. A., Sanz, L. A., Pauler, F. M., Ferguson-Smith, A. C., Feil, R., et al. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 322, 1717–1720. doi: 10.1126/science.1163802

Nickerson, M. L., Dancik, G. M., Im, K. M., Edwards, M. G., Turan, S., Brown, J., et al. (2014). Concurrent alterations in TERT, KDM6A, and the BRCA pathway in bladder cancer. *Clin. Cancer Res.* 20, 4935–4948. doi: 10.1158/1078-0432.CCR-14-0330

Nicorescu, I., Dallinga, G. M., de Winther, M. P. J., Stroes, E. S. G., and Bahjat, M. (2019). Potential epigenetic therapeutics for atherosclerosis treatment. *Atherosclerosis* 281, 189–197. doi: 10.1016/j.atherosclerosis.2018.10.006

Niegisch, G., Knievel, J., Koch, A., Hader, C., Fischer, U., Albers, P., et al. (2013). Changes in histone deacetylase (HDAC) expression patterns and activity of HDAC inhibitors in urothelial cancers. *Urol. Oncol. Semin. Orig. Investig.* 31, 1770–1779. doi: 10.1016/j.urolonc.2012.06.015

Ocaña, A., Nieto-Jiménez, C., and Pandiella, A. (2017). BET inhibitors as novel therapeutic agents in breast cancer. *Oncotarget* 8, 71285–71291. doi: 10.18632/oncotarget.19744

Ogiwara, H., Sasaki, M., Mitachi, T., Oike, T., Higuchi, S., Tominaga, Y., et al. (2016). Targeting p300 addiction in cbp-deficient cancers causes synthetic lethality by apoptotic cell death due to abrogation of MYC expression. *Cancer Discovery* 6, 430–445. doi: 10.1158/2159-8290.CD-15-0754

Okano, M., Bell, D. W., Haber, D. A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell* 99, 247–257. doi: 10.1016/S0092-8674(00)81656-6

Olivieri, M., Ferro, M., Terreri, S., Durso, M., Romanelli, A., Avitabile, C., et al. (2016). Long non-coding RNA containing ultraconserved genomic region 8 promotes bladder cancer tumorigenesis. *Oncotarget* 7(15), 20636–20654. doi: 10.18632/oncotarget.7833

Olkhov-Mitsel, E., Savio, A. J., Kron, K. J., Pethe, V. V., Hermanns, T., Fleshner, N. E., et al. (2017). Epigenome-wide DNA methylation profiling identifies differential methylation biomarkers in high-grade bladder cancer. *Transl. Oncol.* 10, 168–177. doi: 10.1016/j.tranon.2017.01.001

Pal, S. K., Milowsky, M. I., and Plimack, E. R. (2013). Optimizing systemic therapy for bladder cancer. *JNCCN J. Natl. Compr. Cancer Netw.* 11, 793–804. doi: 10.6004/jnccn.2013.0101

Palazzo, A. F., and Lee, E. S. (2015). Non-coding RNA: what is functional and what is junk? *Front. Genet.* 6, 2. doi: 10.3389/fgene.2015.00002

Pang, W., Su, J., Wang, Y., Feng, H., Dai, X., Yuan, Y., et al. (2015). Pancreatic cancer-secreted miR-155 implicates in the conversion from normal fibroblasts to cancer-associated fibroblasts. *Cancer Sci.* 106, 1362–1369. doi: 10.1111/cas.12747

Pappano, W. N., Guo, J., He, Y., Ferguson, D., Jagadeeswaran, S., Osterling, D. J., et al. (2015). The histone methyltransferase inhibitor A-366 uncovers a role for G9a/GLP in the epigenetics of leukemia. *PloS One* 10, 1–13. doi: 10.1371/journal.pone.0131716

Pardini, B., and Naccarati, A. (2017). Altered piRNA Profiles in bladder cancer: a new challenge in the next- generation sequencing era? *J. Genet. Genomes* 1 (2), 1000110.

Pardoll, D. M. (2012). The blockade of immune checkpoints in cancer immunotherapy. *Nat. Rev. Cancer* 12, 252–264. doi: 10.1038/nrc3239

Park, S. M., Gaur, A. B., Lengyel, E., and Peter, M. E. (2008). The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes Dev.* 22, 894–907. doi: 10.1101/gad.1640608

Pastori, C., Daniel, M., Penas, C., Volmar, C.-H., Johnstone, A. L., Brothers, S. P., et al. (2014). BET bromodomain proteins are required for glioblastoma cell proliferation. *Epigenetics* 9, 611–620. doi: 10.4161/epi.27906

Patchsung, M., Boonla, C., Amnattrakul, P., Dissayabutra, T., Mutirangura, A., and Tosukhowong, P. (2012). Long interspersed nuclear element-1 hypomethylation and oxidative stress: correlation and bladder cancer diagnostic potential. *PloS One* 7, e37009. doi: 10.1371/journal.pone.0037009

Peng, D., Ge, G., Xu, Z., Ma, Q., Shi, Y., Zhou, Y., et al. (2018). Diagnostic and prognostic biomarkers of common urological cancers based on aberrant DNA methylation. *Epigenomics* 10, 1189–1199. doi: 10.2217/epi-2018-0017

Petrylak, D. P., Tagawa, S. T., Kohli, M., Eisen, A., Canil, C., Sridhar, S. S., et al. (2016). Docetaxel as monotherapy or combined with ramucirumab or icrucumab in second-line treatment for locally advanced or metastatic urothelial carcinoma: an open-label, three-arm, randomized controlled phase II Trial. *J. Clin. Oncol.* 34, 1500–1509. doi: 10.1200/JCO.2015.65.0218

Picaud, S., Wells, C., Felletar, I., Brotherton, D., Martin, S., Savitsky, P., et al. (2013). RVX-208, an inhibitor of BET transcriptional regulators with selectivity for the second bromodomain. *Proc. Natl. Acad. Sci. U. S. A.* 110, 19754–19759. doi: 10.1073/pnas.1310658110

Pietrusiński, M., Kępczyński, Ł, Jędrzejczyk, A., Borkowska, E., Traczyk-Borszyńska, M., Constantinou, M., et al. (2017). Detection of bladder cancer in urine sediments by a hypermethylation panel of selected tumor suppressor genes. *Cancer Biomark.* 18, 47–59. doi: 10.3233/CBM-160673

Plimack, E. R., Bellmunt, J., Gupta, S., Berger, R., Chow, L. Q., Juco, J., et al. (2017). Safety and activity of pembrolizumab in patients with locally advanced or metastatic urothelial cancer (KEYNOTE-012): a non-randomised, open-label, phase 1b study. *Lancet Oncol.* 18, 212–220. doi: 10.1016/S1470-2045(17)30007-4

Plummer, R., Vidal, L., Griffin, M., Lesley, M., De Bono, J., Coulthard, S., et al. (2009). Phase I study of MG98, an oligonucleotide antisense inhibitor of human DNA methyltransferase 1, given as a 7-day infusion in patients with advanced solid tumors. *Clin. Cancer Res.* 15, 3177–3183. doi: 10.1158/1078-0432.CCR-08-2859

Pop-Bica, C., Gulei, D., Cojocneanu-Petric, R., Braicu, C., Petrut, B., and Berindan-Neagoe, I. (2017). Understanding the role of non-coding rnas in bladder cancer: from dark matter to valuable therapeutic targets. *Int. J. Mol. Sci.* 18(7), 1514. doi: 10.3390/ijms18071514

Porten, S. P. (2018). Epigenetic alterations in bladder cancer. *Curr. Urol. Rep.* 19. doi: 10.1007/s11934-018-0861-5

Powles, T., Eder, J. P., Fine, G. D., Braiteh, F. S., Loriot, Y., Cruz, C., et al. (2014). MPDL3280A (anti-PD-L1) treatment leads to clinical activity in metastatic bladder cancer. *Nature* 515, 558–562. doi: 10.1038/nature13904

Poyet, C., Jentsch, B., Hermanns, T., Schweckendiek, D., Seifert, H.-H., Schmidtpeter, M., et al. (2014). Expression of histone deacetylases 1, 2 and 3 in urothelial bladder cancer. *BMC Clin. Pathol.* 14, 10. doi: 10.1186/1472-6890-14-10

Rampias, T., Karagiannis, D., Avgeris, M., Polyzos, A., Kokkalis, A., Kanaki, Z., et al. (2019). The lysine-specific methyltransferase KMT2C/MLL3 regulates DNA repair components in cancer. *EMBO Rep.* 20, e46821. doi: 10.15252/embr.201846821

Raulet, D. H., Marcus, A., and Coscoy, L. (2017). Dysregulated cellular functions and cell stress pathways provide critical cues for activating and targeting natural killer cells to transformed and infected cells. *Immunol. Rev.* 280, 93–101. doi: 10.1111/imr.12600

Reinert, T. (2012). Methylation markers for urine-based detection of bladder cancer: the next generation of urinary markers for diagnosis and surveillance of bladder cancer. *Adv. Urol.* 2012, 503271. doi: 10.1155/2012/503271

Reinert, T., Modin, C., Castano, F. M., Lamy, P., Wojdacz, T. K., Hansen, L. L., et al. (2011). Comprehensive genome methylation analysis in bladder cancer: identification and validation of novel methylated genes and application of these as urinary tumor markers. *Clin. Cancer Res.* 17, 5582–5592. doi: 10.1158/1078-0432.CCR-10-2659

Ren, C., Morohashi, K., Plotnikov, A. N., Jakoncic, J., Smith, S. G., Li, J., et al. (2015). Small-molecule modulators of methyl-lysine binding for the CBX7 chromodomain. *Chem. Biol.* 22 (2), 161–168. doi: 10.1016/j.chembiol.2014.11.021

Ren, Y., Huang, T., Lou, Z., Weng, G., Chen, F., Fan, X., et al. (2016). Clinical significance of CDH13 promoter methylation as a biomarker for bladder cancer: a meta-analysis. *BMC Urol.* 16, 1–9. doi: 10.1186/s12894-016-0171-5

Renard, I., Joniau, S., van Cleynenbreugel, B., Collette, C., Naômé, C., Vlassenbroeck, I., et al. (2010). Identification and validation of the methylated TWIST1 and NID2 genes through real-time methylation-specific polymerase chain reaction assays for the noninvasive detection of primary bladder cancer in urine samples. *Eur. Urol.* 58, 96–104. doi: 10.1016/j.eururo.2009.07.041

Ringuette Goulet, C., Bernard, G., Tremblay, S., Chabaud, S., Bolduc, S., and Pouliot, F. (2018). Exosomes induce fibroblast differentiation into cancer-associated fibroblasts through TGFβ signaling. *Mol. Cancer Res.* 16, 1196–1204. doi: 10.1158/1541-7786.mcr-17-0784

Robertson, A. G., Kim, J., Al-Ahmadie, H., Bellmunt, J., Guo, G., Cherniack, A. D., et al. (2018). Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell* 174, 1033. doi: 10.1016/j.cell.2018.07.036

Roger, T., Lugrin, J., Le Roy, D., Goy, G., Mombelli, M., Koessler, T., et al. (2011). Histone deacetylase inhibitors impair innate immune responses to Toll-like receptor agonists and to infection. *Blood* 117, 1205–1217. doi: 10.1182/blood-2010-05-284711

Romero-Cordoba, S. L., Salido-Guadarrama, I., Rodriguez-Dorantes, M., and Hidalgo-Miranda, A. (2014). miRNA biogenesis: biological impact in the development of cancer. *Cancer Biol. Ther.* 15, 1444–1455. doi: 10.4161/15384047.2014.955442

Roperch, J. P., Grandchamp, B., Desgrandchamps, F., Mongiat-Artus, P., Ravery, V., Ouzaid, I., et al. (2016). Promoter hypermethylation of HS3ST2, SEPTIN9 and SLIT2 combined with FGFR3 mutations as a sensitive/specific urinary assay for diagnosis and surveillance in patients with low or high-risk non-muscle-invasive bladder cancer. *BMC Cancer* 16, 1–9. doi: 10.1186/s12885-016-2748-5

Rosenberg, J. E., Hoffman-Censits, J., Powles, T., van der Heijden, M. S., Balar, A. V., Necchi, A., et al. (2016). Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *Lancet* 387, 1909–1920. doi: 10.1016/S0140-6736(16)00561-4

Ross, J. S., Wang, K., Al-Rohil, R. N., Nazeer, T., Sheehan, C. E., Otto, G. A., et al (2014). Advanced urothelial carcinoma: next-generation sequencing reveals diverse genomic alterations and targets of therapy. *Mod. Pathol.* 27, 271–280. doi: 10.1038/modpathol.2013.135

Rothbart, S. B., and Strahl, B. D. (2014). Interpreting the language of histone and DNA modifications. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1839, 627–643. doi: 10.1016/j.bbagrm.2014.03.001

Roudnicky, F., Dieterich, L. C., Poyet, C., Buser, L., Wild, P., Tang, D., et al. (2017). High expression of insulin receptor on tumour-associated blood vessels in invasive bladder cancer predicts poor overall and progression-free survival. *J. Pathol.* 242, 193–205. doi: 10.1002/path.4892

Roudnicky, F., Poyet, C., Wild, P., Krampitz, S., Negrini, F., Huggenberger, R., et al. (2013). Endocan Is Upregulated on Tumor Vessels in Invasive Bladder Cancer Where It Mediates VEGF-A-Induced Angiogenesis. *Cancer Res.* 73, 1097–1106. doi: 10.1158/0008-5472.CAN-12-1855

Roudnicky, F., Yoon, S. Y., Poghosyan, S., Schwager, S., Poyet, C., Vella, G., et al. (2018). Alternative transcription of a shorter, non-anti-angiogenic thrombospondin-2 variant in cancer-associated blood vessels. *Oncogene* 37, 2573–2585. doi: 10.1038/s41388-018-0129-z

Ruthenburg, A. J., Li, H., Patel, D. J., and David Allis, C. (2007). Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol.* 8, 983–994. doi: 10.1038/nrm2298

Sakaguchi, T., Yoshino, H., Sugita, S., Miyamoto, K., Yonemori, M., Osako, Y., et al. (2018). Bromodomain protein BRD4 inhibitor JQ1 regulates potential prognostic molecules in advanced renal cell carcinoma. *Oncotarget* 9, 23003. doi: 10.18632/ONCOTARGET.25190

Sánchez-Carbayo, M. (2012). Hypermethylation in bladder cancer: biological pathways and translational applications. *Tumor Biol.* 33, 347–361. doi: 10.1007/s13277-011-0310-2

Santos, M., Martínez-Fernández, M., Dueñas, M., García-Escudero, R., Alfaya, B., Villacampa, F., et al. (2014). In vivo disruption of an Rb-E2F-Ezh2 signaling loop causes bladder cancer. *Cancer Res.* 74, 6565–6577. doi: 10.1158/0008-5472.CAN-14-1218

Sasaki, H., Yoshiike, M., Nozawa, S., Usuba, W., Katsuoka, Y., Aida, K., et al. (2016). Expression Level of Urinary MicroRNA-146a-5p Is Increased in Patients With Bladder Cancer and Decreased in Those After Transurethral Resection. *Clin. Genitourin. Cancer* 14, e493–e499. doi: 10.1016/j.clgc.2016.04.002

Schenk, A., Bloch, W., and Zimmer, P. (2016). Natural Killer Cells–An Epigenetic Perspective of Development and Regulation. *Int. J. Mol. Sci.* 17, 326. doi: 10.3390/ijms17030326

Schlereth, K., Weichenhan, D., Bauer, T., Heumann, T., Giannakouri, E., Lipka, D., et al. (2018). The transcriptomic and epigenetic map of vascular quiescence in the continuous lung endothelium. *Elife* 7(e34423). doi: 10.7554/elife.34423

Schoepp, M., Ströse, A. J., and Haier, J. (2017). Dysregulation of miRNA Expression in Cancer Associated Fibroblasts (CAFs) and Its Consequences on the Tumor Microenvironment. *Cancers (Basel)* 9, 54. doi: 10.3390/cancers9060054

Schulte, J., Weidig, M., Balzer, P., Richter, P., Franz, M., Junker, K., et al. (2012). Expression of the E-cadherin repressors Snail, Slug and Zeb1 in urothelial carcinoma of the urinary bladder: relation to stromal fibroblast activation and invasive behaviour of carcinoma cells. *Histochem. Cell Biol.* 138, 847–860. doi: 10.1007/s00418-012-0998-0

Schulz, W. A. (2006). L1 retrotransposons in human cancers. *J. Biomed. Biotechnol.* 2006, 83672. doi: 10.1155/JBB/2006/83672

Schuyler, R. P., Merkel, A., Raineri, E., Altucci, L., Vellenga, E., Martens, J. H. A., et al. (2016). Distinct Trends of DNA Methylation Patterning in the Innate and Adaptive Immune Systems Europe PMC Funders Group. *Cell Rep.* 17, 2101–2111. doi: 10.1016/j.celrep.2016.10.054

Segovia, C., Martínez-Fernández, M., Dueñas, M., Rubio, C., López-Calderón, F. F., Costa, C., et al. (2017). Opposing roles of <i<PIK3CA</i< gene alterations to EZH2 signaling in non-muscle invasive bladder cancer. *Oncotarget* 8, 10531–10542. doi: 10.18632/oncotarget.14453

Segovia, C., and Paramio, J. M. (2017). Ezh2-dependent therapies in bladder cancer: Synthetic lethality. *Ann. Transl. Med.* 5, 3–6. doi: 10.21037/atm.2017.10.08

Segovia, C., San José-Enériz, E., Munera-Maravilla, E., Martínez-Fernández, M., Garate, L., Miranda, E., et al. (2019). Inhibition of a G9a/DNMT network triggers immune-mediated bladder cancer regression. *Nat. Med.* 25, 1073–1081. doi: 10.1038/s41591-019-0499-y

Serizawa, R. R., Ralfkiaer, U., Steven, K., Lam, G. W., Schmiedel, S., Schüz, J., et al. (2011). Integrated genetic and epigenetic analysis of bladder cancer reveals an additive diagnostic value of FGFR3 mutations and hypermethylation events. *Int. J. Cancer* 129, 78–87. doi: 10.1002/ijc.25651

Shah, M. Y., and Calin, G. A. (2014). MicroRNAs as therapeutic targets in human cancers. *Wiley Interdiscip. Rev. RNA* 5, 537–548. doi: 10.1002/wrna.1229

Shah, S. H., Miller, P., Garcia-Contreras, M., Ao, Z., Machlin, L., Issa, E., et al (2015). Hierarchical paracrine interaction of breast cancer associated fibroblasts with cancer cells *via* hMAPK-microRNAs to drive ER-negative breast cancer phenotype. *Cancer Biol. Ther.* 16, 1671–1681. doi: 10.1080/15384047.2015.1071742

Shang, D., Liu, Y., Matsui, Y., Ito, N., Nishiyama, H., Kamoto, T., et al. (2008). Demethylating Agent 5-Aza-2′-Deoxycytidine Enhances Susceptibility of Bladder Transitional Cell Carcinoma to Cisplatin. *Urology* 71, 1220–1225. doi: 10.1016/J.UROLOGY.2007.11.029

Shankar, S. R., Bahirvani, A. G., Rao, V. K., Bharathy, N., Ow, J. R., and Taneja, R. (2013). G9a, a multipotent regulator of gene expression. *Epigenetics* 8, 16–22. doi: 10.4161/epi.23331

Sharma, A., Jamil, M. A., Nuesgen, N., Dauksa, A., Gulbinas, A., Schulz, W. A., et al. (2019). Detailed methylation map of LINE-1 5′-promoter region reveals hypomethylated CpG hotspots associated with tumor tissue specificity. *Mol. Genet. Genomic Med.* 7, e601. doi: 10.1002/mgg3.601

Shi, H., Jiang, H., Wang, L., Cao, Y., Liu, P., Xu, X., et al. (2015). Overexpression of monocarboxylate anion transporter 1 and 4 in T24-induced cancer-associated fibroblasts regulates the progression of bladder cancer cells in a 3D microfluidic device. *Cell Cycle* 14, 3058–3065. doi: 10.1080/15384101.2015.1053666

Simm, C., Caridis, A., Di Maio, A., Knowles, P., Gorman, B., Jones, R., et al. (2018). Targeting of BRD4 with JQ1, combined with mitomycin C as a novel combination therapy for non-muscle invasive bladder cancer. *Eur. Urol. Suppl.* 17, e663. doi: 10.1016/S1569-9056(18)31300-9

Simon, J. M., Parker, J. S., Liu, F., Rothbart, S. B., Ait-Si-ali, S., Strahl, B. D., et al (2015). A role for widely interspaced zinc finger (WIZ) in retention of the G9a methyltransferase on chromatin. *J. Biol. Chem.* 290, 26088–26102. doi: 10.1074/jbc.M115.654459

Siomi, M. C., Sato, K., Pezic, D., and Aravin, A. A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* 12, 246–258. doi: 10.1038/nrm3089

Smith, Z. D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* 14, 204–220. doi: 10.1038/nrg3354

Sohel, M. H. (2016). Extracellular/Circulating MicroRNAs: Release Mechanisms, Functions and Challenges. *Achiev. Life Sci.* 10, 175–186. doi: 10.1016/j.als.2016.11.007

Sonpavde, G., and Bellmunt, J. (2016). Angiogenesis as a therapeutic target in urothelial carcinoma. *Nat. Rev. Urol.* 13, 306–307. doi: 10.1038/nrurol.2016.69

Stenzl, A., Cowan, N. C., De Santis, M., Kuczyk, M. A., Merseburger, A. S., Ribal, M. J., et al. (2011). Treatment of Muscle-invasive and Metastatic Bladder Cancer: Update of the EAU Guidelines. *Eur. Urol.* 59, 1009–1018. doi: 10.1016/j.eururo.2011.03.023

Stone, O. A., El-Brolosy, M., Wilhelm, K., Liu, X., Romão, A. M., Grillo, E., et al (2018). Loss of pyruvate kinase M2 limits growth and triggers innate immune signaling in endothelial cells. *Nat. Commun.* 9, 4077. doi: 10.1038/s41467-018-06406-8

Stubendorff, B., Wilhelm, K., Posselt, K., Catto, J., Hartmann, A., Bertz, S., et al. (2019). A three-gene methylation marker panel for the nodal metastatic risk assessment of muscle-invasive bladder cancer Keywords Muscle-invasive bladder cancer · Methylation · Biomarker · Metastasis. *J. Cancer Res. Clin. Oncol.* 145, 811–820. doi: 10.1007/s00432-018-02829-4

Su, H., Tao, T., Yang, Z., Kang, X., Zhang, X., Kang, D., et al. (2019). Circular RNA cTFRC acts as the sponge of MicroRNA-107 to promote bladder carcinoma progression. *Mol. Cancer* 18, 27. doi: 10.1186/s12943-019-0951-0

Su, S.-F., de Castro Abreu, A. L., Chihara, Y., Tsai, Y., Andreu-Vieyra, C., Daneshmand, S., et al. (2014). A panel of three markers hyper- and hypomethylated in urine sediments accurately predicts bladder cancer recurrence. *Clin. Cancer Res.* 20, 1978–1989. doi: 10.1158/1078-0432.CCR-13-2637

Sukari, A., Nagasaka, M., Al-Hadidi, A., and Lum, L. G. (2016). Cancer immunology and immunotherapy. *Anticancer Res.* 36, 5593–5606. doi: 10.21873/anticanres.11144

Sun, H., Tian, J., Xian, W., Xie, T., and Yang, X. (2015). miR-34a inhibits proliferation and invasion of bladder cancer cells by targeting orphan nuclear receptor HNF4G. *Dis. Markers* 2015, 1–8. doi: 10.1155/2015/879254

Sundar Rajan, V., Laurent, V. M., Verdier, C., and Duperray, A. (2017). Unraveling the receptor-ligand interactions between bladder cancer cells and the endothelium using AFM. *Biophys. J.* 112, 1246–1257. doi: 10.1016/j.bpj.2017.01.033

Tada, Y., Wada, M., Kuroiwa, K., Kinugawa, N., Harada, T., Nagayama, J., et al. (2000). MRD1 gene overexpression and altered degree of methylation at the promoter region in bladder cancer during chemotherapeutic treatment. *Clin. Cancer Res.* 6, 4618–4627.

Taheri, M., Omrani, M. D., and Ghafouri-Fard, S. (2018). Long non-coding RNA expression in bladder cancer. *Biophys. Rev.* 10, 1205–1213. doi: 10.1007/s12551-017-0379-y

Takeuchi, A., Shiota, M., Tatsugami, K., Yokomizo, A., Tanaka, S., Kuroiwa, K., et al. (2012). P300 Mediates cellular resistance to doxorubicin in bladder cancer. *Mol. Med. Rep.* 5, 173–176. doi: 10.3892/mmr.2011.593

Tan, Y., Wang, L., Du, Y., Liu, X., Chen, Z., Weng, X., et al. (2018). Inhibition of BRD4 suppresses tumor growth in prostate cancer *via the* enhancement of FOXO1 expression. *Int. J. Oncol.* 53, 2503–2517. doi: 10.3892/ijo.2018.4577

Tao, L., Huang, G., Song, H., Chen, Y., and Chen, L. (2017). Cancer associated fibroblasts: an essential role in the tumor microenvironment. *Oncol. Lett.* 14, 2611–2620. doi: 10.3892/ol.2017.6497

Taubert, H., Wach, S., Jung, R., Pugia, M., and Keck, B. (2015). Piwil 2 expression is correlated with disease-specific and progression-free survival of chemotherapy-treated bladder cancer patients. *Mol. Med.* 21, 371–380. doi: 10.2119/molmed.2014.00250

Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D., and Patel, D. J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* 14, 1025–1040. doi: 10.1038/nsmb1338

Tian, Z., Cao, S., Li, C., Xu, M., Wei, H., Yang, H., et al. (2019). LncRNA PVT1 regulates growth, migration, and invasion of bladder cancer by miR-31/*CDK1*. *J. Cell. Physiol.* 234, 4799–4811. doi: 10.1002/jcp.27279

Tolkach, Y., Stahl, A. F., Niehoff, E. M., Zhao, C., Kristiansen, G., Müller, S. C., et al. (2017). YRNA expression predicts survival in bladder cancer patients. *BMC Cancer* 17, 1–7. doi: 10.1186/s12885-017-3746-y

Tsai, H.-C., and Baylin, S. B. (2011). Cancer epigenetics: linking basic biology to clinical medicine. *Cell Res.* 21, 502–517. doi: 10.1038/cr.2011.24

Vafaee, F., Colvin, E. K., Mok, S. C., Howell, V. M., and Samimi, G. (2017). Functional prediction of long non-coding RNAs in ovarian cancer-Associated fibroblasts indicate a potential role in metastasis. *Sci. Rep.* 7, 10374. doi: 10.1038/s41598-017-10869-y

Valenzuela, M. T., Galisteo, R., Zuluaga, A., Villalobos, M., Núñez, M. I., Oliver, F. J., et al. (2002). Assessing the use of p16(INK4a) promoter gene methylation in serum for detection of bladder cancer. *Eur. Urol.* 42, 622–628. discussion 628-30.

van der Heijden, A. G., Mengual, L., Ingelmo-Torres, M., Lozano, J. J., van Rijt-van de Westerlo, C. C. M., Baixauli, M., et al. (2018). Urine cell-based DNA methylation classifier for monitoring bladder cancer. *Clin. Epigenet.* 10, 1–10. doi: 10.1186/s13148-018-0496-x

van Kessel, K. E. M., Beukers, W., Lurkin, I., Ziel-van der Made, A., van der Keur, K. A., Boormans, J. L., et al. (2017). Validation of a DNA Methylation-Mutation Urine Assay to Select Patients with Hematuria for Cystoscopy. *J. Urol.* 197, 590–595. doi: 10.1016/j.juro.2016.09.118

van Rhijn, B. W. G., Burger, M., Lotan, Y., Solsona, E., Stief, C. G., Sylvester, R. J., et al. (2009). Recurrence and progression of disease in non-muscle-invasive bladder cancer: from epidemiology to treatment strategy. *Eur. Urol.* 56, 430–442. doi: 10.1016/j.eururo.2009.06.028

Van Rhijn, B. W. G., Van Der Kwast, T. H., Liu, L., Fleshner, N. E., Bostrom, P. J., Vis, A. N., et al. (2012). The FGFR3 mutation is related to favorable pT1 bladder cancer. *J. Urol.* 187, 310–314. doi: 10.1016/j.juro.2011.09.008

Vedadi, M., Barsyte-lovejoy, D., Liu, F., Rival-gervier, S., Allali-hassani, A., Labrie, V., et al. (2012). A chemical probe selectively inhibits G9a and GLP methyltransferase and genanylgeranyltransferase I with antitumor activity against brest cancer *in vivo. Nat. Chem. Biol.* 7, 566–574. doi: 10.1038/nchembio.599.A

Villar-garea, A., Fraga, M. F., Espada, J., and Esteller, M. (2003). Procaine Is a DNA-demethylating Agent with Growth-inhibitory Effects in Human. *Cancer Res.* 63, 4984–4989.

Viré, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., et al. (2006). The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* 439, 871–874. doi: 10.1038/nature04431

Vivier, E., Tomasello, E., Baratin, M., Walzer, T., and Ugolini, S. (2008). Functions of natural killer cells. *Nat. Immunol.* 9, 503–510. doi: 10.1038/ni1582

Vo, J. N., Cieslik, M., Zhang, Y., Shukla, S., Xiao, L., Zhang, Y., et al. (2019). The landscape of circular RNA in cancer. *Cell* 176, 869–881.e13. doi: 10.1016/j.cell.2018.12.021

Wagner, E. K., Nath, N., Flemming, R., Feltenberger, J. B., and Denu, J. M. (2012). Identification and characterization of small molecule inhibitors of a plant homeodomain finger. *Biochemistry* 51, 8293–8306. doi: 10.1021/bi3009278

Wang, D., Quiros, J., Mahuron, K., Pai, C.-C., Ranzani, V., Young, A., et al. (2018a). Targeting EZH2 reprograms intratumoral regulatory T cells to enhance cancer immunity. *Cell Rep.* 23, 3262–3274. doi: 10.1016/j.celrep.2018.05.050

Wang, K. C., and Chang, H. Y. (2011). Molecular mechanisms of long noncoding RNAs. *Mol. Cell* 43, 904–914. doi: 10.1016/j.molcel.2011.08.018

Wang, M., Guo, C., Wang, L., Luo, G., Huang, C., Li, Y., et al. (2018b). Long noncoding RNA GAS5 promotes bladder cancer cells apoptosis through inhibiting EZH2 transcription. *Cell Death Dis.* 9, 238. doi: 10.1038/s41419-018-0264-z

Wang, P., Nishitani, M.-A., Tanimoto, S., Kishimoto, T., Fukumori, T., Takahashi, M., et al. (2007). Bladder cancer cell invasion is enhanced by cross-talk with fibroblasts through hepatocyte growth factor. *Urology* 69, 780–784. doi: 10.1016/j.urology.2007.01.063

Wang, S., Chen, J., Garcia, S. P., Liang, X., Zhang, F., Yan, P., et al. (2019). A dynamic and integrated epigenetic program at distal regions orchestrates transcriptional responses to VEGFA. *Genome Res.* 29, 193–207. doi: 10.1101/gr.239053.118

Wang, X.-S., Zhang, Z., Wang, H.-C., Cai, J.-L., Xu, Q.-W., Li, M.-Q., et al. (2006). Rapid Identification of UCA1 as a very sensitive and specific unique marker for human bladder carcinoma. *Clin. Cancer Res.* 12, 4851–4858. doi: 10.1158/1078-0432.CCR-06-0134

Wang, Y. M., Gu, M. L., Meng, F. S., Jiao, W. R., Zhou, X. X., Yao, H. P., et al. (2017). Histone acetyltransferase p300/CBP inhibitor C646 blocks the survival and invasion pathways of gastric cancer cell lines. *Int. J. Oncol.* 51 (6), 1860–1868. doi: 10.3892/ijo.2017.4176

Wang, Y., Xiang, W., Wang, M., Huang, T., Xiao, X., Wang, L., et al. (2014). Methyl jasmonate sensitizes human bladder cancer cells to gambogic acid-induced apoptosis through down-regulation of EZH2 expression by miR-101. *Br. J. Pharmacol.* 171, 618–635. doi: 10.1111/bph.12501

Weber, J. A., Baxter, D. H., Zhang, S., Huang, D. Y., How Huang, K., Jen Lee, M., et al. (2010). The MicroRNA Spectrum in 12 Body Fluids. *Clin. Chem.* 56, 1733–1741. doi: 10.1373/clinchem.2010.147405

Weinstein, J. N., Akbani, R., Broom, B. M., Wang, W., Verhaak, R. G. W., McConkey, D., et al. (2014). Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* 507, 315–322. doi: 10.1038/nature12965

Wiklund, E. D., Bramsen, J. B., Hulf, T., Dyrskjøt, L., Ramanathan, R., Hansen, T. B., et al. (2011). Coordinated epigenetic repression of the miR-200 family and miR-205 in invasive bladder cancer. *Int. J. Cancer* 128, 1327–1334. doi: 10.1002/ijc.25461

Winquist, E., Knox, J., Wood, L., Wainman, N., Reid, G. K., Pearce, L., et al., (2006). "Phase II trial of DNA methyltransferase 1 inhibition with the antisense oligonucleotide MG98 in patients with metastatic renal carcinoma," in *A National Cancer Institute of Canada Clinical Trials Group investigational new drug study. Invest. New Drugs* 24, 159–160. doi: 10.1007/s10637-006-5938-1

Witjes, J. A., Compérat, E., Cowan, N. C., De Santis, M., Gakis, G., Lebret, T., et al. (2014a). EAU guidelines on muscle-invasive and metastatic bladder cancer: summary of the 2013 guidelines. *Eur. Urol.* 65, 778–792. doi: 10.1016/j.eururo.2013.11.046

Witjes, J. A., Compérat, E., Cowan, N. C., De Santis, M., Gakis, G., Lebret, T., et al. (2014b). EAU guidelines on muscle-invasive and metastatic bladder cancer: summary of the 2013 guidelines. *Eur. Urol.* 65, 778–792. doi: 10.1016/j.eururo.2013.11.046

Wolff, E. M., Chihara, Y., Pan, F., Weisenberger, D. J., Siegmund, K. D., Sugano, K., et al. (2010). Unique DNA methylation patterns distinguish noninvasive and invasive urothelial cancers and establish an epigenetic field defect in premalignant tissue. *Cancer Res.* 70, 8169–8178. doi: 10.1158/0008-5472.CAN-10-1335

Wu, J., Li, W., Ning, J., Yu, W., Rao, T., and Cheng, F. (2019a). Long noncoding RNA UCA1 targets miR-582-5p and contributes to the progression and drug resistance of bladder cancer cells through ATG7-mediated autophagy inhibition. *Oncol. Targets Ther.* 12, 495–508. doi: 10.2147/OTT.S183940

Wu, J., Yu, C., Cai, L., Lu, Y., Jiang, L., Liu, C., et al. (2017). Effects of increased Kindlin-2 expression in bladder cancer stromal fibroblasts. *Oncotarget* 8, 50692–50703. doi: 10.18632/oncotarget.17021

Wu, M., Sheng, L., Cheng, M., Zhang, H., Jiang, Y., Lin, S., et al. (2019b). Low doses of decitabine improve the chemotherapy efficacy against basal-like bladder cancer by targeting cancer stem cells. *Oncogene* 38, 5425–5439. doi: 10.1038/s41388-019-0799-1

Wu, S., Yang, Z., Ye, R., An, D., Li, C., Wang, Y., et al. (2016a). Novel variants in MLL confer to bladder cancer recurrence identified by whole-exome sequencing. *Oncotarget* 7, 2629–2645. doi: 10.18632/oncotarget.6380

Wu, X., Liu, D., Tao, D., Xiang, W., Xiao, X., Wang, M., et al. (2016b). BRD4 Regulates EZH2 transcription through upregulation of C-MYC and represents a novel therapeutic target in bladder cancer. *Mol. Cancer Ther.* 15, 1029–1042. doi: 10.1158/1535-7163.MCT-15-0750

Wu, Z., Huang, W., Wang, X., Wang, T., Chen, Y., Chen, B., et al. (2018). Circular RNA CEP128 acts as a sponge of miR-145-5p in promoting the bladder cancer progression *via* regulating SOX11. *Mol. Med.* 24, 40. doi: 10.1186/s10020-018-0039-0

Xie, F., Li, Y., Wang, M., Huang, C., Tao, D., Zheng, F., et al. (2018). Circular RNA BCRC-3 suppresses bladder cancer proliferation through miR-182-5p/p27 axis. *Mol. Cancer* 17, 144. doi: 10.1186/s12943-018-0892-z

Xie, H., Liao, X., Chen, Z., Fang, Y., He, A., Zhong, Y., et al. (2017a). LncRNA MALAT1 inhibits apoptosis and promotes invasion by antagonizing miR-125b in bladder cancer cells. *J. Cancer* 8, 3803–3811. doi: 10.7150/jca.21228

Xie, Y., Ma, X., Chen, L., Li, H., Gu, L., Gao, Y., et al. (2017b). MicroRNAs with prognostic significance in bladder cancer: A systematic review and meta-analysis. *Sci. Rep.* 7, 5619. doi: 10.1038/s41598-017-05801-3

Xue, M., Pang, H., Li, X., Li, H., Pan, J., and Chen, W. (2016). Long non-coding RNA urothelial cancer-associated 1 promotes bladder cancer cell migration and invasion by way of the hsa-miR-145-ZEB1/2-FSCN1 pathway. *Cancer Sci.* 107, 18–27. doi: 10.1111/cas.12844

Xylinas, E., Hassler, M. R., Zhuang, D., Krzywinski, M., Erdem, Z., Robinson, B. D., et al. (2016). An epigenomic approach to improving response to neoadjuvant cisplatin chemotherapy in bladder cancer. *Biomolecules* 6, 37. doi: 10.3390/biom6030037

Yamaguchi, H., and Hung, M.-C. (2014). Regulation and Role of EZH2 in Cancer. *Cancer Res. Treat.* 46, 209–222. doi: 10.4143/crt.2014.46.3.209

Yamamoto, K., Tateishi, K., Kudo, Y., Hoshikawa, M., Tanaka, M., Nakatsuka, T., et al. (2016). Stromal remodeling by the BET bromodomain inhibitor JQ1 suppresses the progression of human pancreatic cancer. *Oncotarget* 7, 61469–61484. doi: 10.18632/oncotarget.11129

Yang, C., Yuan, W., Yang, X., Li, P., Wang, J., Han, J., et al. (2018a). Circular RNA circ-ITCH inhibits bladder cancer progression by sponging miR-17/miR-224 and regulating p21, PTEN expression. *Mol. Cancer* 17, 19. doi: 10.1186/s12943-018-0771-7

Yang, G.-J., Lei, P.-M., Wong, S., Ma, D.-L., and Leung, C.-H. (2018b). Pharmacological inhibition of LSD1 for cancer treatment. *Molecules* 23 (12), 3194. doi: 10.3390/molecules23123194

Yang, X., Han, H., De Carvalho, D. D., Lay, F. D., Jones, P. A., and Liang, G. (2014). Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* 26, 577–590. doi: 10.1016/j.ccr.2014.07.028

Yang, X., Yuan, W., Tao, J., Li, P., Yang, C., Deng, X., et al. (2017). Identification of circular RNA signature in bladder cancer. *J. Cancer* 8, 3456–3463. doi: 10.7150/jca.19345

Yang, Y., Qu, A., Liu, J., Wang, R., Liu, Y., Li, G., et al. (2015). Serum miR-210 contributes to tumor detection, stage prediction and dynamic surveillance in patients with bladder cancer. *PloS One* 10, e0135168. doi: 10.1371/journal.pone.0135168

Yang, Z., Liu, A., Xiong, Q., Xue, Y., Liu, F., Zeng, S., et al. (2019). Prognostic value of differentially methylated gene profiles in bladder cancer. *J. Cell. Physiol.*, 28515. doi: 10.1002/jcp.28515

Yates, D. R., Rehman, I., Abbod, M. F., Meuth, M., Cross, S. S., Linkens, D. A., et al. (2007). Promoter hypermethylation identifies progression risk in bladder cancer. *Clin. Cancer Res.* 13, 2046–2053. doi: 10.1158/1078-0432.CCR-06-2476

Yeh, C. R., Hsu, I., Song, W., Chang, H., Miyamoto, H., Xiao, G. Q., et al. (2015). Fibroblast ERalpha promotes bladder cancer invasion *via* increasing the CCL1 and IL-6 signals in the tumor microenvironment. *Am. J. Cancer Res.* 5, 1146–1157.

Yoon, C. Y., Park, M. J., Lee, J. S., Lee, S. C., Oh, J. J., Park, H., et al. (2011). The histone deacetylase inhibitor trichostatin a synergistically resensitizes a cisplatin resistant human bladder cancer cell line. *J. Urol.* 185, 1102–1111. doi: 10.1016/j.juro.2010.10.034

Yu, C., Longfei, L., Long, W., Feng, Z., Chen, J., Chao, L., et al. (2019). LncRNA PVT1 regulates VEGFC through inhibiting miR-128 in bladder cancer cells. *J. Cell. Physiol.* 234, 1346–1353. doi: 10.1002/jcp.26929

Yu, G., Yao, W., Xiao, W., Li, H., Xu, H., and Lang, B. (2014). MicroRNA-34a functions as an anti-metastatic microRNA and suppresses angiogenesis in bladder cancer by directly targeting CD44. *J. Exp. Clin. Cancer Res.* 33, 779. doi: 10.1186/s13046-014-0115-4

Yu, J., Walter, K., Omura, N., Hong, S.-M., Young, A., Li, A., et al. (2012). Unlike pancreatic cancer cells pancreatic cancer associated fibroblasts display minimal gene induction after 5-Aza-2′-Deoxycytidine. *PloS One* 7, e43456. doi: 10.1371/journal.pone.0043456

Yu, T., Wang, Y., Hu, Q., Wu, W., Wu, Y., and Wei, W., (2017). The EZH2 inhibitor GSK343 suppresses cancer stem-like phenotypes and reverses mesenchymal transition in glioma cells. *Oncotarget* 8 (58), 98348–98359. doi: 10.18632/oncotarget.21311

Yuan, T., Huang, X., Woodcock, M., Du, M., Dittmar, R., Wang, Y., et al. (2016). Plasma extracellular RNA profiles in healthy and cancer patients. *Sci. Rep.* 6, 19413. doi: 10.1038/srep19413

Yuan, Y., Wang, Q., Paulk, J., Kubicek, S., Kemp, M. M., Adams, D. J., et al. (2012). A small-molecule probe of the histone methyltransferase G9a induces cellular senescence in pancreatic adenocarcinoma. *ACS Chem. Biol.* 7, 1152–1157. doi: 10.1021/cb300139y

Yun, S. J., Jeong, P., Kim, W.-T., Kim, T. H., Lee, Y.-S., Song, P. H., et al. (2012). Cell-free microRNAs in urine as diagnostic and prognostic biomarkers of bladder cancer. *Int. J. Oncol.* 41, 1871–1878. doi: 10.3892/ijo.2012.1622

Zeisberg, E. M., Potenta, S., Xie, L., Zeisberg, M., and Kalluri, R. (2007). discovery of endothelial to mesenchymal transition as a source for carcinoma-associated fibroblasts. *Cancer Res.* 67, 10123–10128. doi: 10.1158/0008-5472.CAN-07-3127

Zeisberg, E. M., and Zeisberg, M. (2013). The role of promoter hypermethylation in fibroblast activation and fibrogenesis. *J. Pathol.* 229, 264–273. doi: 10.1002/path.4120

Zeng, Z., Zhou, W., Duan, L., Zhang, J., Lu, X., Jin, L., et al. (2019). Circular RNA circ-VANGL1 as a competing endogenous RNA contributes to bladder cancer progression by regulating miR-605-3p/VANGL1 pathway. *J. Cell. Physiol.* 234, 3887–3896. doi: 10.1002/jcp.27162

Zhang, C., Yao, Z., Zhu, M., Ma, X., Shi, T., Li, H., et al. (2012). Inhibitory effects of MicroRNA-34a on cell migration and invasion of invasive urothelial bladder carcinoma by targeting Notch1. *J. Huazhong Univ. Sci. Technol. Med. Sci.* 32, 375–382. doi: 10.1007/s11596-012-0065-z

Zhang, M., and Xin, Y. (2018). Circular RNAs: a new frontier for cancer diagnosis and therapy. *J. Hematol. Oncol.* 11, 21. doi: 10.1186/s13045-018-0569-5

Zhang, Q., Miao, S., Han, X., Li, C., Zhang, M., Cui, K., et al. (2018a). MicroRNA-3619-5p suppresses bladder carcinoma progression by directly targeting β-catenin and CDK2 and activating p21. *Cell Death Dis.* 9, 960. doi: 10.1038/s41419-018-0986-y

Zhang, Q., Zhao, W., Ye, C., Zhuang, J., Chang, C., Li, Y., et al. (2015). Honokiol inhibits bladder tumor growth by suppressing EZH2/miR-143 axis. *Oncotarget* 6, 37335–37348. doi: 10.18632/oncotarget.6135

Zhang, X., Zhang, Y., Liu, X., Fang, A., Wang, J., Yang, Y., et al. (2016). Direct quantitative detection for cell-free miR-155 in urine: a potential role in diagnosis and prognosis for non-muscle invasive bladder cancer. *Oncotarget* 7, 3255–3266. doi: 10.18632/oncotarget.6487

Zhang, Y., Huang, Z., Zhu, Z., Zheng, X., Liu, J., Han, Z., et al. (2014). Upregulated UHRF1 promotes bladder cancer cell invasion by epigenetic silencing of KiSS1. *PloS One* 9(10), e104252. doi: 10.1371/journal.pone.0104252

Zhang, Y., Liang, W., Zhang, P., Chen, J., Qian, H., Zhang, X., et al. (2017). Circular RNAs: emerging cancer biomarkers and targets. *J. Exp. Clin. Cancer Res.* 36, 152. doi: 10.1186/s13046-017-0624-z

Zhang, Y., Tian, S., Xiong, J., Zhou, Y., Song, H., and Liu, C. (2018b). JQ-1 Inhibits colon cancer proliferation *via* suppressing Wnt/β-catenin signaling and miR-21. *Chem. Res. Toxicol.* 31, 302–307. doi: 10.1021/acs.chemrestox.7b00346

Zhao, L., Duan, Y.-T., Lu, P., Zhang, Z.-J., Zheng, X.-K., Wang, J.-L., et al. (2019). Epigenetic targets and their inhibitors in cancer therapy. *Curr. Top. Med. Chem.* 18, 2395–2419. doi: 10.2174/1568026619666181224095449

Zhao, L., Ji, G., Le, X., Wang, C., Xu, L., Feng, M., et al. (2017). Long noncoding rna linc00092 acts in cancer-associated fibroblasts to drive glycolysis and progression of ovarian cancer. *Cancer Res.* 77, 1369–1382. doi: 10.1158/0008-5472.CAN-16-1615

Zhong, Z., Huang, M., Lv, M., He, Y., Duan, C., Zhang, L., et al. (2017). Circular RNA MYLK as a competing endogenous RNA promotes bladder cancer progression through modulating VEGFA/VEGFR2 signaling pathway. *Cancer Lett.* 403, 305–317. doi: 10.1016/j.canlet.2017.06.027

Zhong, Z., Lv, M., and Chen, J. (2016). Screening differential circular RNA expression profiles reveals the regulatory role of circTCF25-miR-103a-3p/miR-107-CDK6 pathway in bladder carcinoma. *Sci. Rep.* 6, 30919. doi: 10.1038/srep30919

Zhou, X., Zhang, X., Yang, Y., Li, Z., Du, L., Dong, Z., et al. (2014). Urinary cell-free microRNA-106b as a novel biomarker for detection of bladder cancer. *Med. Oncol.* 31, 1–7. doi: 10.1007/s12032-014-0197-z

Zhu, J., Jiang, Z., Gao, F., Hu, X., Zhou, L., Chen, J., et al. (2011). A systematic analysis on DNA methylation and the expression of both mRNA and microRNA in bladder cancer. *PloS One* 6, e28223. doi: 10.1371/journal.pone.0028223

Zhu, S., Denman, C. J., Cobanoglu, Z. S., Kiany, S., Lau, C. C., Gottschalk, S. M., et al. (2015). The narrow-spectrum HDAC inhibitor entinostat enhances NKG2D expression without NK cell toxicity, leading to enhanced recognition of cancer cells. *Pharm. Res.* 32, 779–792. doi: 10.1007/s11095-013-1231-0

Zhu, Y., Spitz, M. R., Zhang, H., Grossman, H. B., Frazier, M. L., and Wu, X. (2004). Methyl-CpG-binding domain 2: a protective role in Bladder Carcinoma. *Cancer* 100, 1853–1858. doi: 10.1002/cncr.20199

Zhuang, J., Lu, Q., Shen, B., Huang, X., Shen, L., Zheng, X., et al. (2015). TGFβ1 secreted by cancer-associated fibroblasts induces epithelial-mesenchymal transition of bladder cancer cells through lncRNA-ZEB2NAT. *Sci. Rep.* 5, 11924. doi: 10.1038/srep11924

Ziani, L., Chouaib, S., and Thiery, J. (2018). alteration of the antitumor immune response by cancer-associated fibroblasts. *Front. Immunol.* 9, 414. doi: 10.3389/FIMMU.2018.00414

Zong, Y., Huang, J., Sankarasharma, D., Morikawa, T., Fukayama, M., Epstein, J. I., et al. (2012). Stromal epigenetic dysregulation is sufficient to initiate mouse prostate cancer *via* paracrine Wnt signaling. *Proc. Natl. Acad. Sci.* 109, E3395–E3404. doi: 10.1073/pnas.1217982109

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Martinez, Munera-Maravilla, Bernardini, Rubio, Suarez-Cabrera, Segovia, Lodewijk, Dueñas, Martínez-Fernández and Paramio. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 20 November 2019
doi: 10.3389/fgene.2019.00803

Comprehensive Analysis of Therapy-Related Messenger RNAs and Long Noncoding RNAs as Novel Biomarkers for Advanced Colorectal Cancer

Jibin Li[1], Siping Ma[1], Tao Lin[1], Yanxi Li[1], Shihua Yang[2], Wanchuan Zhang[2], Rui Zhang[1] and Yongpeng Wang[1]**

[1] Department of Colorectal Surgery, Liaoning Cancer Hospital, Cancer Hospital of China Medical University, Shenyang, China, [2] China Medical University, Shenyang, China

OPEN ACCESS

Edited by:
Nejat Dalay,
Istanbul University, Turkey

Reviewed by:
Kenneth K.W. To,
The Chinese University of HongKong, Hong Kong
Alice Hudder,
Lake Erie College of Osteopathic Medicine, United States

***Correspondence:**
Yongpeng Wang
puhao9502147@163.com
Rui Zhang
tanghuangsouo@163.com

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: 07 May 2019
Accepted: 31 July 2019
Published: 20 November 2019

Citation:
Li J, Ma S, Lin T, Li Y, Yang S, Zhang W, Zhang R and Wang Y (2019) Comprehensive Analysis of Therapy-Related Messenger RNAs and Long Noncoding RNAs as Novel Biomarkers for Advanced Colorectal Cancer. Front. Genet. 10:803. doi: 10.3389/fgene.2019.00803

Colorectal cancer (CRC) is one of the most common types of human cancers. However, the mechanisms underlying CRC progression remained elusive. This study identified differently expressed messenger RNAs (mRNAs), long noncoding RNAs (lncRNAs), and small nucleolar RNAs (snoRNAs) between pre-therapeutic biopsies and post-therapeutic resections of locally advanced CRC by analyzing a public dataset, GSE94104. We identified 427 dysregulated mRNAs, 4 dysregulated lncRNAs, and 19 dysregulated snoRNAs between pre- and post-therapeutic locally advanced CRC samples. By constructing a protein–protein interaction network and co-expressing networks, we identified 10 key mRNAs, 4 key lncRNAs, and 7 key snoRNAs. Bioinformatics analysis showed therapy-related mRNAs were associated with nucleosome assembly, chromatin silencing at recombinant DNA, negative regulation of gene expression, and DNA replication. Therapy-related lncRNAs were associated with cell adhesion, extracellular matrix organization, angiogenesis, and sister chromatid cohesion. In addition, therapy-related snoRNAs were associated with DNA replication, nucleosome assembly, and telomere organization. We thought this study provided useful information for identifying novel biomarkers for CRC.

Keywords: long noncoding RNA, snoRNAs, prognostic markers, expression profiling, protein–protein interaction analysis, co-expression analysis, colorectal cancer

INTRODUCTION

Colorectal cancer (CRC) is one of the most common types of human cancers (Ma et al., 2014). The morbidity and mortality of CRC have increased rapidly in recent years (Budai et al., 2004). In 2016, a total of 134,490 new cases of CRC and 49,190 deaths caused by CRC were reported worldwide. In the past decades, the diagnostic technologies and therapeutic strategies of CRC have made significant progress (Ress et al., 2015). However, the prognosis of CRC remained poor with 5-year survival rates being only 10–15%, and the recurrent disease rates of CRC remained high. Therefore, there was still an urgent need to understand the mechanisms underlying CRC progression and identify novel potential biomarkers for the prognosis of CRC.

Emerging studies had demonstrated that noncoding RNAs played crucial roles in the progression of CRC (Rezanejad Bardaji et al., 2018), including microRNAs, long noncoding RNAs (lncRNAs), and small nucleolar RNA (snoRNAs). The important roles of microRNAs in CRC had been studied clearly (Zhang et al., 2012a). lncRNAs are a large class of transcripts longer than 200 bases, with no protein-coding potential. Previous studies had showed that lncRNAs were associated with CRC progression and prognosis. For example, overexpression of lncRNA TUSC7 reduces cell migration and invasion in CRC (Xu J. et al., 2017). lncRNA KCNQ1OT1 enhanced the methotrexate resistance of CRC cells by regulating miR-760/PPP1R1B (Sunamura et al., 2011). LINC01354 interacting with hnRNP-D contributes to the proliferation and metastasis in CRC through activating Wnt/β-catenin signalling (Zhang et al., 2016). Recent studies have also indicated that snoRNAs were also associated with the progression of CRC, for example, Yoshida et al. (2017).

In the present study, we re-annotated a Gene Expression Omnibus (GEO) dataset GSE94104 to identify CRC-related mRNAs and lncRNAs. Bioinformatics analysis was also performed to understand the potential roles of these lncRNAs in CRC. This study could provide novel clues to prove that CRC-related lncRNAs could serve as biomarkers for CRC.

MATERIALS AND METHODS

lncRNA Classification Pipeline

We used a pipeline described by Zhang et al. to re-annotate microarray data using the following criteria (Zhang et al., 2012b). Briefly, first, GPL570 platform of Affymetrix Human Genome U133 Plus 2.0 Array (Affymetrix Inc., Santa Clara, California, USA) probe set ID was mapped to the NetAffx Annotation Files (HG-U133 Plus 2.0 Annotations, CSV format, release 31, 08/23/10). The annotations included the probe set ID, gene symbol, and Refseq transcript ID. Second, the probe sets that were assigned with a Refseq transcript ID in the NetAffx annotations were extracted. In this study, we only retained those labeled as "NR_" (NR indicates noncoding RNA in the Refseq database). Finally, 2,448 annotated lncRNA transcripts with corresponding Affymetrix probe IDs were generated.

Microarray Data and Data Preprocessing

By screening colon cancer-related public datasets in GEO database, we selected GSE94104 dataset for further study, which contained the largest number of therapy-related colon cancer samples. In the present study, we downloaded GSE94104 datasets (Tsukamoto et al., 2011) from GEO database to identify differently expressed mRNAs and lncRNAs. A total of 40 matched formalin-fixed paraffin-embedded pre-therapeutic locally advanced rectal cancer biopsy and post-therapeutic locally advanced rectal cancer biopsy samples were included in this study. All samples were provided by the Northern Ireland Biobank and arrayed using the Illumina HumanHT-12 WG-DASL V4 expression beadchip. The raw data were normalized using robust multi-array average method under R 3.4.2 statistical software with affy package from BioConductor. Normalization was separately performed for LCM dataset and homogenized tissue dataset. The normalized gene expression levels were presented as log2-transformed values by robust multi-array average. lncRNAs with fold changes ≥2 and P values <0.05 were considered as differentially expressed lncRNAs.

Co-Expression Network Construction and Analysis

In this study, the Pearson correlation coefficient of different expressed gene–lncRNA pairs was calculated according to the expression value of them. The co-expressed differentially expressed gene–lncRNA pairs with the absolute value of Pearson correlation coefficient ≥0.6 were selected, and the co-expression network was established by using cytoscape software.

Functional Group Analysis

The DAVID system (http://david.ncifcrf.gov/) was used to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses. GO analyses included biological process, cellular component, and molecular function. GO terms and KEGG pathways with a P value of <0.05 were considered as significantly enriched function annotations.

Protein–Protein Interaction Network and Module Analysis

STRING online software was used to construct a protein–protein interaction (PPI) network (Liu et al., 2009) (https://string-db.org/cgi/input.pl?sessionId = AUH42ZEZwajP&input_page_show_search = on). PPI with the combined score >0.4 was considered as significant. Cytoscape software was used to visualize the PPI network.

RESULTS

Transcriptional Analysis of Therapy-Related Messenger RNAs in Pre-Therapeutic Biopsies and Post-Therapeutic Resections of Locally Advanced Colorectal Cancer

The present study aimed to identify therapy-related mRNAs in advanced CRC using a public dataset, GSE94104. A total of 40 pre-therapeutic advanced CRC samples and 40 post-therapeutic advanced CRC samples were included in this dataset. We identified 427 dysregulated mRNAs between pre- and post-therapeutic locally advanced CRC (LACC) samples, including 235 upregulated mRNAs and 192 downregulated mRNAs after therapy in LACC. Hierarchical clustering was used to show differentially expressed mRNAs in post-therapeutic LACC (**Figure 1A**).

Transcriptional Analysis of Therapy-Related Noncoding RNAs in Pre-Therapeutic Biopsies and Post-Therapeutic Resections of Locally Advanced Colorectal Cancer

Next, we focused on identifying noncoding RNAs between pre- and post-therapeutic LACC samples. A total of 19 snoRNAs and 4 lncRNAs were found to be differently expressed (**Figure 1B**).

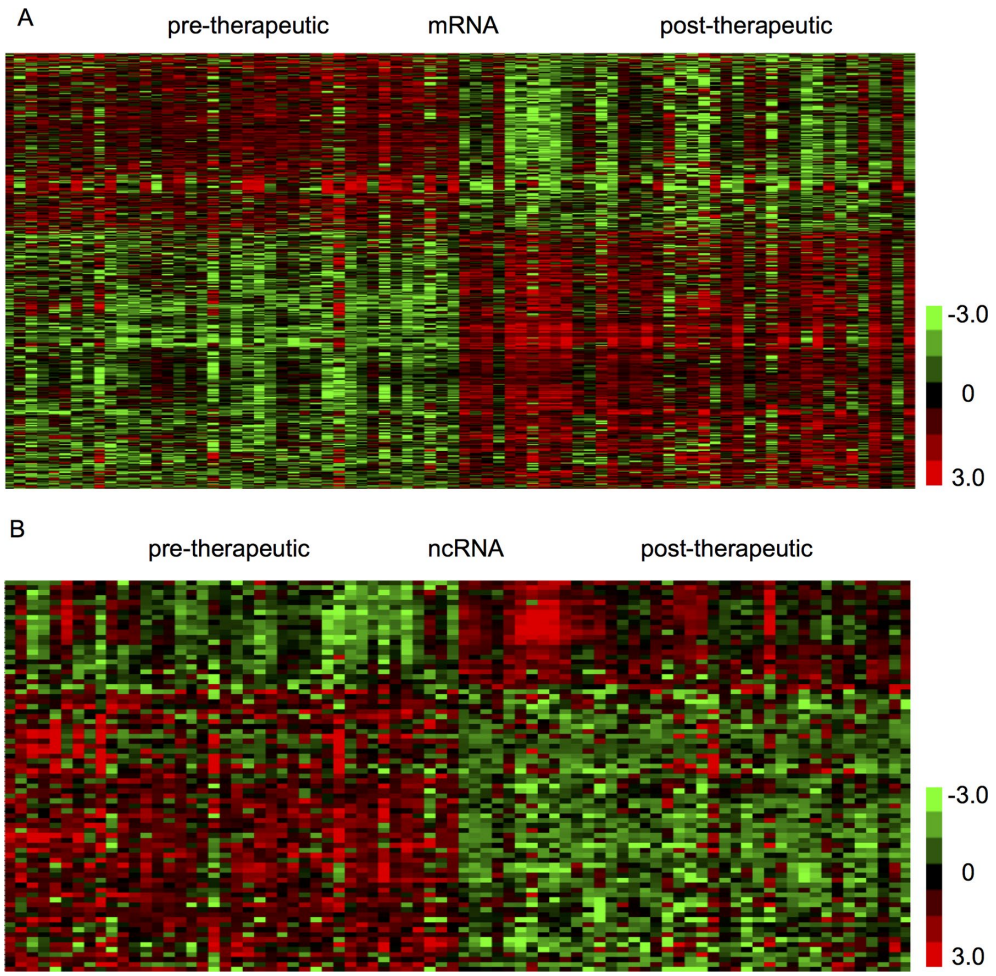


FIGURE 1 | Identification of therapy-related mRNAs and ncRNAs in CRC. **(A)** Hierarchical clustering analysis showed differential mRNAs expression in the CRC by using GSE94104. **(B)** Hierarchical clustering analysis showed differential ncRNAs expression in the CRC by using GSE94104.

Among these lncRNAs, we found SERTAD4-AS1 and MIR100HG were upregulated, whereas PCAT18 and KRTAP5-AS1 were downregulated in post-therapeutic LACC samples compared with those in pre-therapeutic LACC samples. Interestingly, we found most of these affected snoRNAs (18/19) were upregulated in post-therapeutic LACC samples compared with those in pre-therapeutic LACC samples, including SNORD116-4, SNORD116-2, SNORD 107, SNORD61, SNORD112, SNORD109A, SNORD113-5, SNOR D113-8, SNORD113-7, SNORD114-1, SNORD114-11, SNORD 113-6, SNORD114-17, SNORD113-9, SNORD113-3, SNORD114-3, SNORD113-2, and SNORD114-13.

Protein–Protein Interaction Network Analysis of Therapy-Related Messenger RNAs in Locally Advanced Colorectal Cancer

In order to reveal the relationships among therapy-related mRNAs in LACC, we constructed PPI networks using STRING database. The combined score >0.4 was used as the cutoff criterion. As shown in **Figure 2**, a total of 348 nodes and 1,047 edges were included in this PPI network. The nodes that had higher degrees were identified as hub genes, including FN1, CDC20, SPP1, HIST1H3B, ZWINT, CENPF, HIST1H3C, CXCR4, HIST1H3G, and RFC3.

Construction of Therapy-Related Long Noncoding RNAs and Small Nucleolar RNAs Regulating Co-Expression Network in Locally Advanced Colorectal Cancer

In order to reveal the potential functions of therapy-related lncRNAs and snoRNAs in LACC, we first performed Pearson correlation calculation between lncRNAs or snoRNAs and mRNAs in LACC. Based on the correlation analysis results, we constructed mRNA–lncRNA/snoRNAs co-expression networks (p-value < 0.05 and absolute value of correlation coefficient >0.7).

As shown in **Figure 3**, the mRNAs–lncRNAs co-expression network included 4 lncRNAs (MIR100HG, SERTAD4-AS1, KRTAP5-AS1, and PCAT18) and 226 mRNAs. MIR100HG was

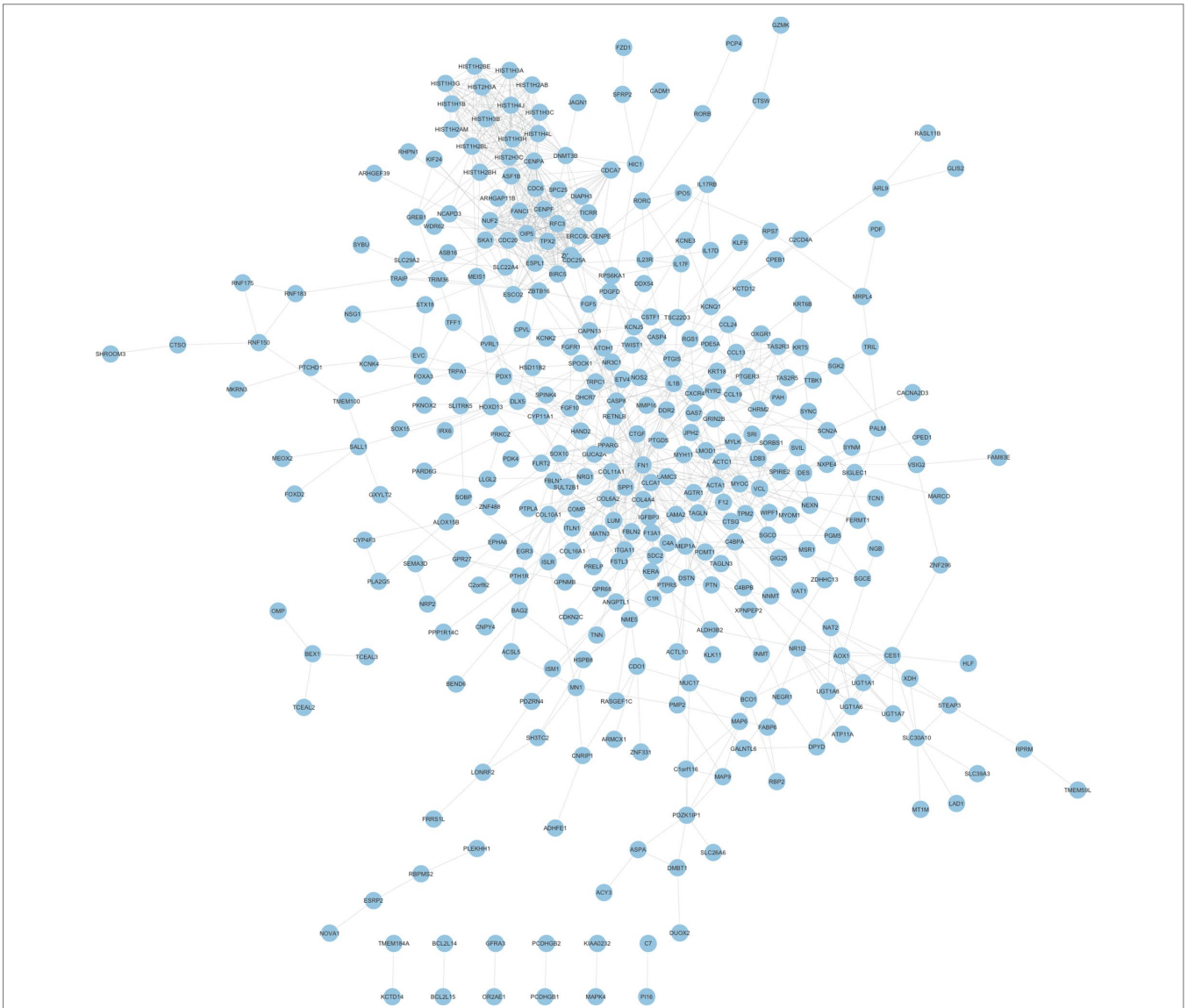


FIGURE 2 | Construction of therapy-related PPI networks in CRC. We constructed therapy-related PPI networks in CRC, including a total of 348 nodes and 1,047 edges.

the key lncRNA in this network by co-expressing with more than 200 mRNAs, including FZD1, FGFR1, FN1, and KLF9. The top 10 most co-expressing genes of SERTAD4-AS1 included COL16A1, ISLR, ZNF626, COL6A2, SOX15, FRMD6, PCDHGA9, CDC6, TPM2, and C1R. The top 10 most co-expressing genes of KRTAP5-AS1 included DPYD, GAL3ST2, CSTL1, HSD11B2, LRCH2, DIAPH3, FERMT1, MRPL4, NEGR1, and LAMA2. The top 10 most co-expressing genes of PCAT18 included ZNF626, OR2AE1, WIPF1, CTGF, IL17F, L3HYPDH, COL16A1, KCNIP3, PCDHGA9, and COL6A2.

As shown in **Figure 4**, the mRNAs–snoRNAs co-expression network included 19 snoRNAs and 360 mRNAs. Several snoRNAs were identified as key regulators by co-expressing with more than 150 mRNAs, including SNORD114-3, SNORD114-1, SNORD113-5, SNORD88B, SNORD113-8, SNORD114-11, and SNORD113-2.

Bioinformatics Analysis of Therapy-Related Messenger RNAs in Locally Advanced Colorectal Cancer

Furthermore, we performed GO and KEGG analysis for therapy-related mRNAs in LACC (**Figures 5 A**, **B**). Bioinformatics analysis showed that the therapy-related mRNAs were mainly involved in regulating nucleosome assembly, chromatin silencing at recombinant DNA (rDNA), negative regulation of gene expression, DNA replication-dependent nucleosome assembly, extracellular matrix organization, cellular protein metabolic process, telomere

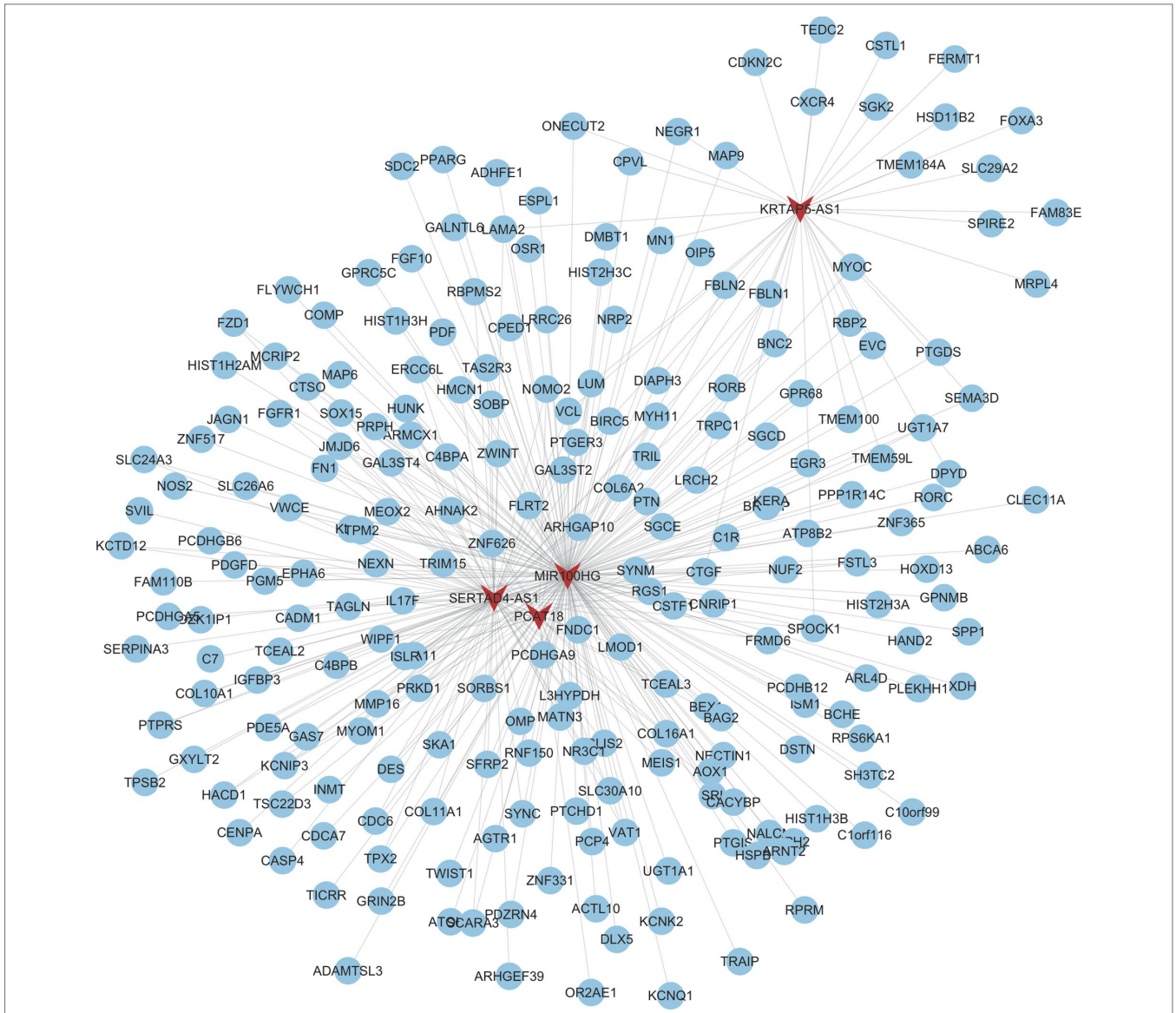


FIGURE 3 | Construction of therapy-related lncRNA regulating co-expression networks in CRC. We constructed therapy-related lncRNA regulating co-expression networks in CRC, including a total of 4 lncRNAs and 226 mRNAs. Red node, lncRNAs; blue nodes, mRNAs.

organization, regulation of gene silencing, positive regulation of gene expression, and muscle organ development. KEGG pathway analysis revealed that therapy-related mRNAs were mainly involved in regulating systemic lupus erythematosus, drug metabolism—other enzymes, alcoholism, transcriptional misregulation in cancer, and extracellular matrix (ECM)–receptor interaction.

Bioinformatics Analysis for Related Long Noncoding RNAs and Small Nucleolar RNAs in Locally Advanced Colorectal Cancer

Then, bioinformatics analysis for related lncRNAs and snoRNAs in LACC was performed using their regulating targets in LACC (**Figures 5 C–F**). GO analysis showed that differentially expressed lncRNAs were associated with cell adhesion, extracellular matrix organization, angiogenesis, sister chromatid cohesion, positive regulation of transcription, apoptotic process, chromatin silencing at rDNA, epithelial cell differentiation, cell division, and cellular protein metabolic process. KEGG pathway analysis indicated therapy-related lncRNAs were associated with ECM–receptor interaction, transcriptional misregulation in cancer, focal adhesion, pathways in cancer, and PI3K-Akt signaling pathway.

GO analysis showed that differentially expressed snoRNAs were associated with chromatin silencing at rDNA, DNA replication, nucleosome assembly, telomere organization, regulation of gene silencing, muscle organ development, cellular

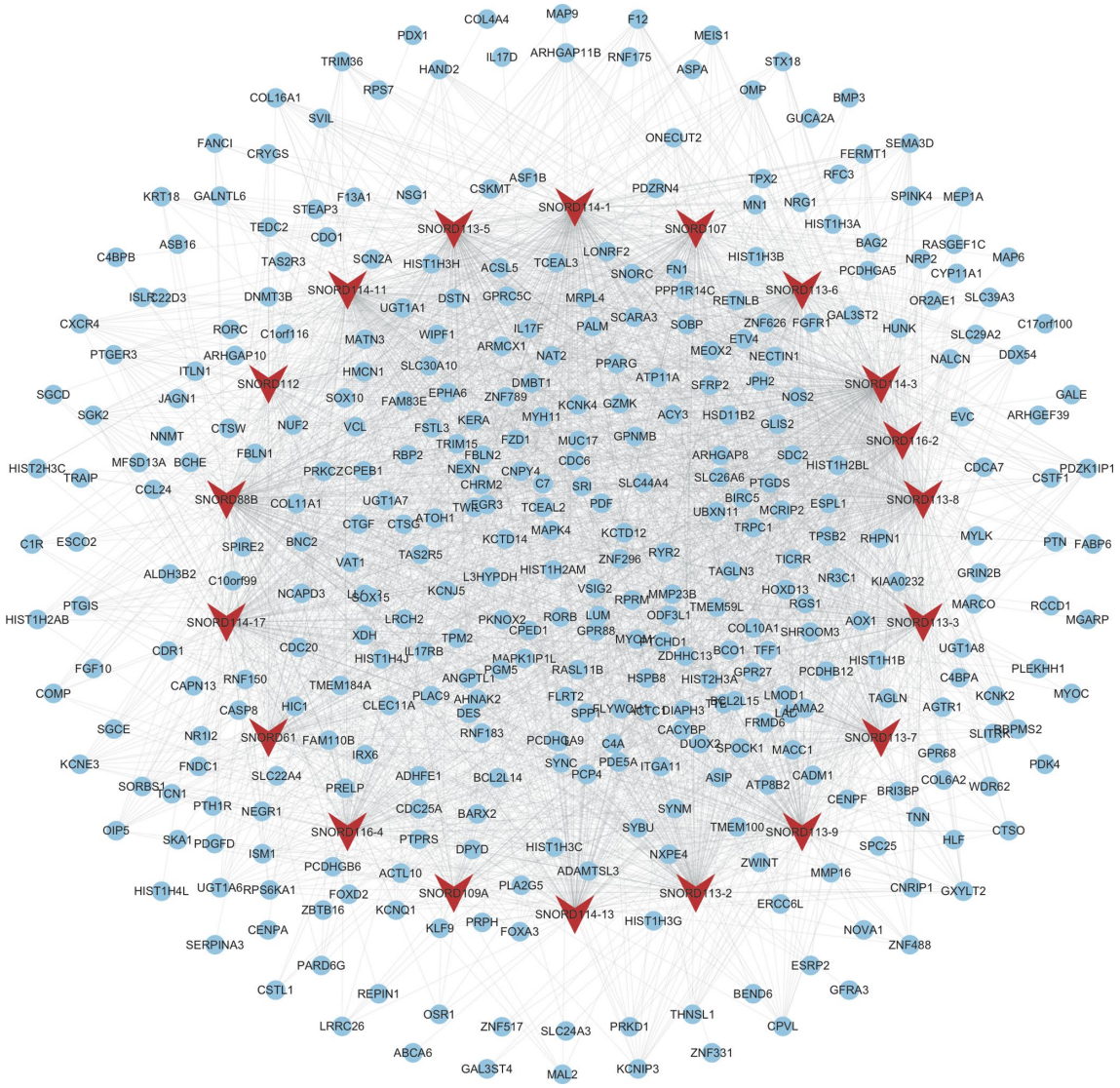


FIGURE 4 | Construction of therapy-related snoRNAs regulating co-expression networks in CRC. We constructed therapy-related snoRNAs regulating co-expression networks in CRC, including a total of 19 snoRNAs and 360 mRNAs. Red node, snoRNAs; blue nodes, mRNAs.

protein metabolic process, protein heterotetramerization, extracellular matrix organization, and cell division. KEGG pathway analysis indicated that therapy-related snoRNAs were associated with systemic lupus erythematosus, alcoholism, drug metabolism, ECM–receptor interaction, and transcriptional misregulation in cancer.

Expression of Key lncRNAs Were Dysregulated in Colorectal Cancer Samples

In order to investigate the prognostic value of key lncRNAs in CRC, we analyzed an independent public dataset, the Gene Expression Profiling Interactive Analysis (GEPIA) database. By analyzing the GEPIA database, we found that the expression levels of MIR100HG, SERTAD4-AS1, and PCAT18 were significantly downregulated; however, KRTAP5-AS1 was upregulated in both colon adenocarcinoma (COAD) and rectum adenocarcinoma (READ) samples compared with that in normal tissues (**Figure 6**).

DISCUSSION

CRC is one of the most common types of human cancer, which is caused by multiple genetic and epigenetic aberrations. However, the mechanisms underlying CRC remained largely unclear. This study identified differently expressed mRNAs, lncRNAs, and snoRNAs between pre-therapeutic biopsies and post-therapeutic resections of locally advanced CRC by analyzing a public dataset, GSE94104. Then, we constructed a PPI network to identify

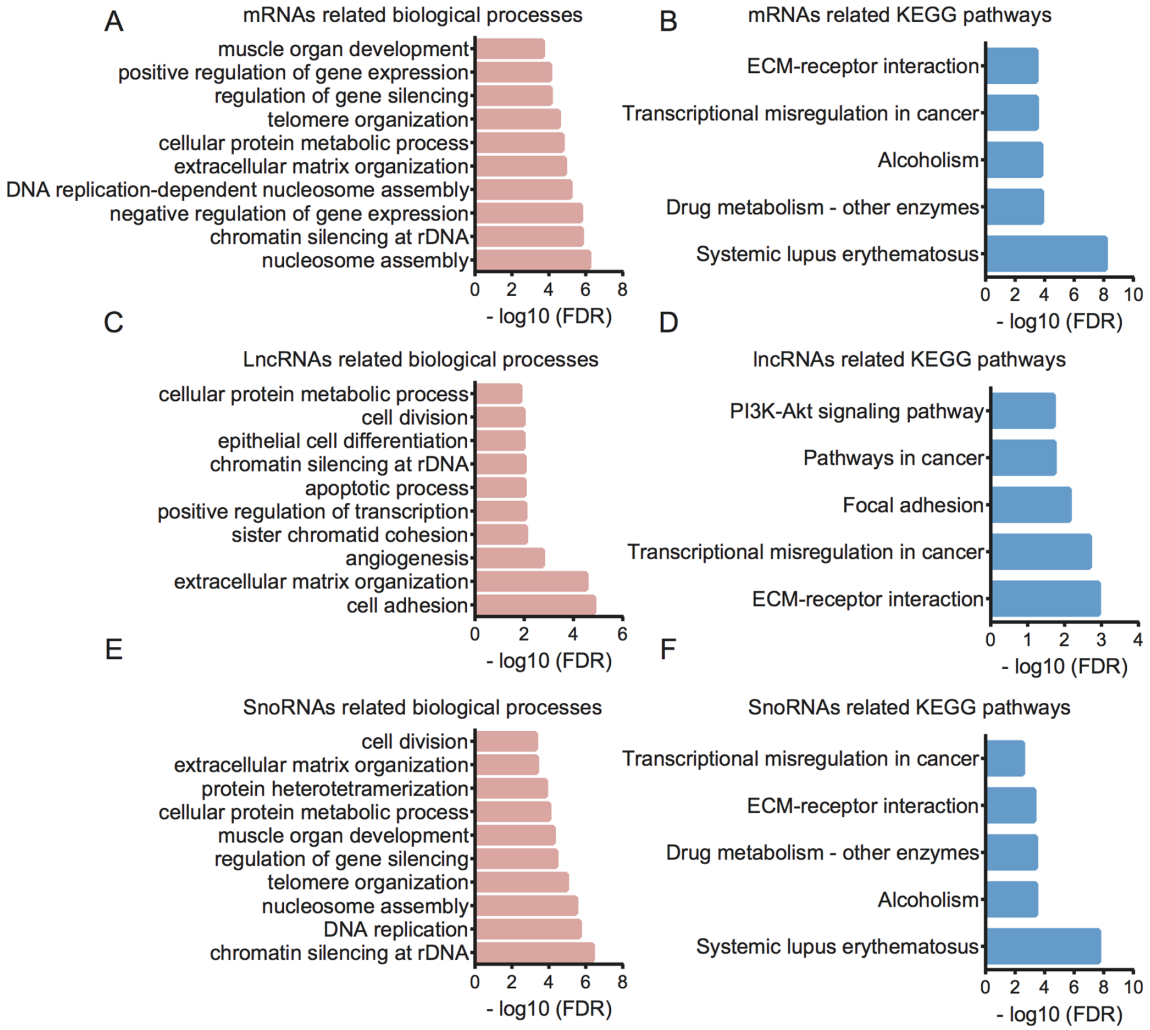


FIGURE 5 | Bioinformatics analysis for therapy-related mRNAs, lncRNAs, and snoRNAs in CRC. **(A)** GO analysis showed therapy-related mRNA-associated biological processes. **(B)** KEGG pathway analysis showed therapy-related mRNA-associated pathways. **(C)** GO analysis showed therapy-related lncRNA-associated biological processes. **(D)** KEGG pathway analysis showed therapy-related lncRNA-associated pathways. **(E)** GO analysis showed therapy-related snoRNA-associated biological processes. **(F)** KEGG pathway analysis showed therapy-related snoRNA-associated pathways.

key therapy-related proteins in LACC. Next, we constructed snoRNAs and lncRNAs regulating co-expression networks to identify key therapy-related snoRNAs and lncRNAs in LACC. Finally, GO and KEGG pathway analyses were conducted to predict their potential functions in LACC.

The present study identified a total of 235 upregulated mRNAs and 192 downregulated mRNAs after therapy in LACC. Bioinformatics analysis showed that these mRNAs were associated with nucleosome assembly, chromatin silencing at rDNA, negative regulation of gene expression, and DNA replication. Furthermore, a PPI network including 348 proteins and 1,037 edges were constructed to reveal the relationship among therapy-related proteins. Ten proteins were identified as key regulators in this network, including FN1, CDC20, SPP1, HIST1H3B, ZWINT, CENPF, HIST1H3C, CXCR4, HIST1H3G, and RFC3. FN1 is a novel protein involved in regulating cancer progression (Ifon et al., 2005). FN1 was found to be dysregulated in multiple human cancers, including colon cancer (Cai et al., 2018). In CRC, a single nucleotide polymorphism in FN1 was found to be associated with tumor shape. FN1 was transcriptionally activated by HMGA2, and the suppression of FN1 inhibited CRC growth and metastasis. CDC20 is a key E3 ligase that binds to APC and recognizes D-box or KEN box substrates to promote proteasomal degradation (Paul et al., 2017). CDC20 was frequently overexpressed in malignant tumors, such as prostate cancer, hepatocellular carcinoma, and ovarian cancer. SPP1 was reported to be overexpressed in numerous tumors, such as lung cancer, colon cancer, breast cancer, and prostate cancer (Xu C. et al., 2017). SPP1 was associated with tumor metastasis in gastric cancer and esophageal adenocarcinoma. Zwint is an important regulatory protein for chromosome movement and mitotic checkpoints

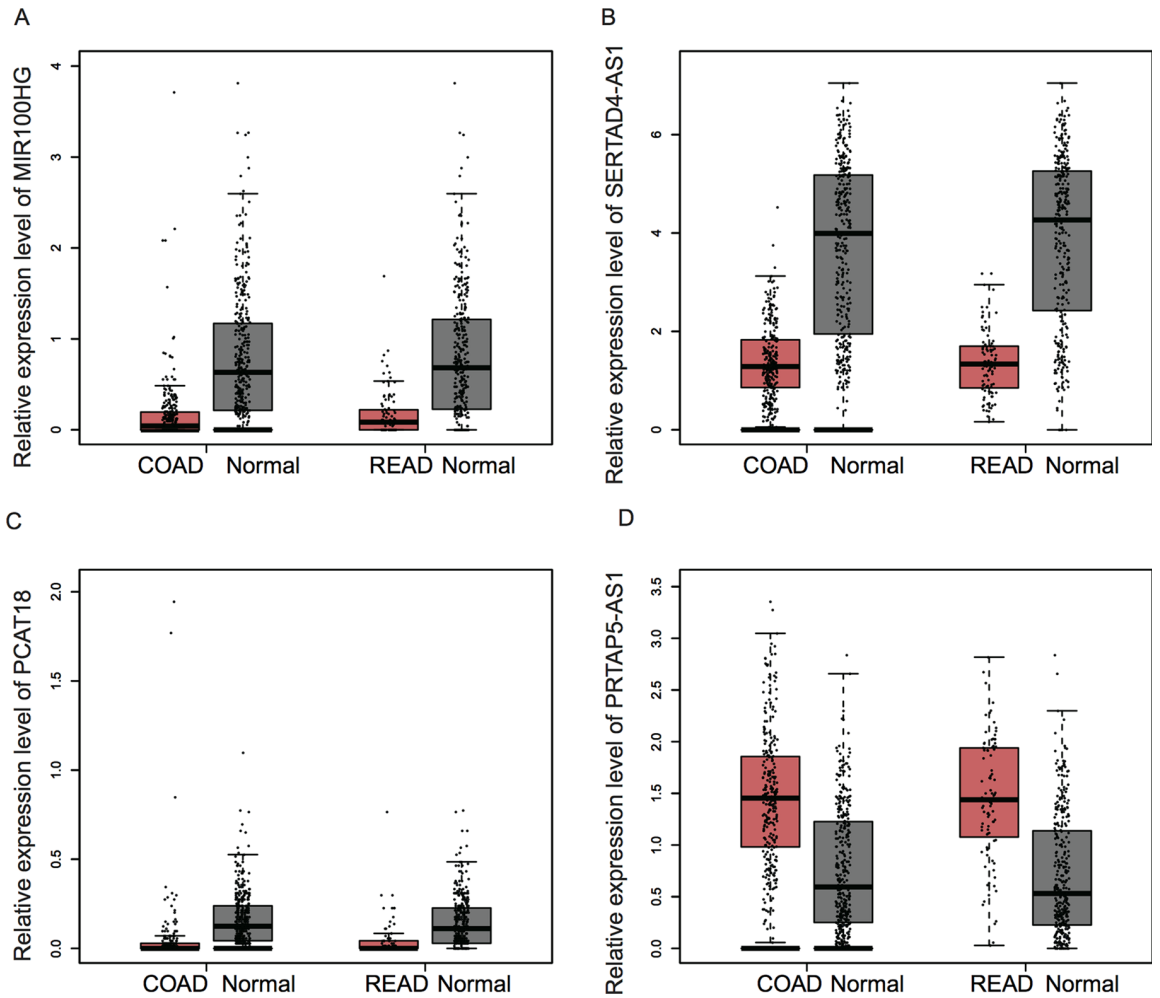


FIGURE 6 | The expression of key lncRNAs were dysregulated in CRC samples. **(A–D)** The expression levels of MIR100HG **(A)**, SERTAD4-AS1 **(B)**, and PCAT18 **(C)** were significantly downregulated; however, KRTAP5-AS1 **(D)** was upregulated in both COAD and READ samples compared with that in normal tissues.

(Kasuboski et al., 2011). Previous studies have identified Zwint overexpression in breast and ovarian cancers. CENPF is a part of the centromere–kinetochore complex and is a component of the nuclear matrix during G2 of interphase (Sugimoto et al., 1999). Recent studies showed that CENPF played crucial roles in the progression of human cancers. For example, the altered phosphorylation of CENPF affected glutamine uptake in colon cancers (Michalak et al., 2019). CXCR4 is a transmembrane G-protein-couple receptor and played a central role in the neurotropism of cells (Xu et al., 2015). RFC3 was a member of RFC family, which played a key role in DNA replication, DNA damage repair, and checkpoint control. Multiple studies indicated RFC3 was overexpressed and correlated to the progression of human cancers (Shen et al., 2014). These reports together with our findings suggested that these key regulators may play key roles in regulating the therapy-related biological processes in LACC.

Recent studies showed ncRNAs were involved in regulating multiple cancer-related biological processes, such as cell proliferation, apoptosis, and invasion. For example, HAND2-AS1 was observed to suppress CRC proliferation though sponging miR-1275 (Zhou et al., 2018). SNORA21 played as an oncogenic snoRNA in CRC with a prognostic biomarker potential. However, the ncRNAs involved in CRC therapy remained largely unclear. The present study identified 4 lncRNAs and 19 snoRNAs as therapy-related ncRNAs in LACC. Next, lncRNA–mRNA and snoRNAs–mRNA co-expression networks were constructed. Four lncRNAs, including MIR100HG, SERTAD4-AS1, KRTAP5-AS1, and PCAT18, were found to play crucial roles in this progression. KRTAP5-AS1 was reported as a potential biomarker for papillary thyroid carcinoma. *PCAT18* was found to be associated with the progression of gastric cancer (Foroughi et al., 2018) and prostate cancer. For example, PCAT18 silencing inhibited prostate cancer proliferation, migration, and invasion (Zhan et al., 2018). MIR100HG was identified as a key regulator in LACC (Li et al., 2019). A recent study showed that MIR100HG regulates cell cycle by modulating

the interaction between HuR and its target mRNAs (Sun et al., 2018). The present study showed that MIR100HG regulated more than 200 mRNAs, including FZD1, FGFR1, FN1, and KLF9. These genes had been demonstrated to be related to CRC progression. For example, KLF9 prevents CRC through inhibition of interferon-related signaling. Downregulation of FN1 suppressed CRC proliferation, migration, and invasion. FZD1 was a key regulator of wnt signaling and involved in regulating CRC metastasis. SERTAD4-AS1 was involved in regulating COL16A1, ISLR, ZNF626, COL6A2, SOX15, FRMD6, PCDHGA9, CDC6, TPM2, and C1R. Among these genes, TPM2 knockdown had been reported to promote CRC progression upon RhoA activation. We also found KRTAP5-AS1 might regulate DPYD, GAL3ST2, CSTL1, HSD11B2, LRCH2, DIAPH3, FERMT1, MRPL4, NEGR1, and LAMA2 in CRC. Among these mRNAs, DPYD variants was reported to be a predictor of 5-fluorouracil toxicity in adjuvant colon cancer treatment. FERMT1 promoted colon cancer metastasis and epithelial–mesenchymal transition progression *via* modulation of β-catenin transcriptional activity. By analyzing the GEPIA database, we found that the expression levels of MIR100HG, SERTAD4-AS1, and PCAT18 were significantly downregulated; however, KRTAP5-AS1 was upregulated in both COAD and READ samples compared with that in normal tissues. Furthermore, we conducted bioinformatics analysis for these therapy-related lncRNAs and snoRNAs. Our results showed therapy-related lncRNAs were associated with cell adhesion, extracellular matrix organization, angiogenesis, and sister chromatid cohesion. In addition, therapy-related snoRNAs were associated with DNA replication, nucleosome assembly, and telomere organization.

Of note, several limitations should be noted in this study. First, the number of samples used in present study were limited. In the further study, more samples should be included to identify therapy-related lncRNAs, snoRNAs, and mRNAs. Second, the detail molecular functions and mechanisms of these key lncRNAs and snoRNAs were unclear. The further validation of these genes should be further investigated. Finally, with the development of next-generation sequence methods, RNA-seq would be a more powerful method to identify novel therapy-related lncRNAs, snoRNAs, and mRNAs in LACC.

In conclusion, we identified 427 dysregulated mRNAs, 4 dysregulated lncRNAs, and 19 dysregulated snoRNAs between pre- and post-therapeutic LACC samples. By constructing a PPI network and co-expressing networks, we identified 10 key mRNAs, 4 key lncRNAs, and 7 key snoRNAs. Bioinformatics analysis showed that therapy-related mRNAs were associated with nucleosome assembly, chromatin silencing at rDNA, negative regulation of gene expression, and DNA replication. Therapy-related lncRNAs were associated with cell adhesion, extracellular matrix organization, angiogenesis, and sister chromatid cohesion. Furthermore, therapy-related snoRNAs were associated with DNA replication, nucleosome assembly, and telomere organization. We think this study provided useful information for identifying novel biomarkers for CRC.

DATA AVAILABILITY

The datasets generated for this study can be found in the GSE94104.

AUTHOR CONTRIBUTIONS

(I) Conception and design: YW, RZ; (II) Administrative support: YW, RZ; (III) Provision of study materials or patients: JL, SM, TL; (IV) Collection and assembly of data: JL, YL, SY; (V) Data analysis and interpretation: JL, WZ; (VI) Manuscript writing: All authors: (VII) Final approval of manuscript: All authors.

REFERENCES

Budai, B., Hitre, E., Adleff, V., Czegledi, F., Gyergyay, F., Lang, I., (2004). The clinical importance of methylenetetrahydrofolate reductase (MTHFR) C677T polymorphism in the 5-fluoropyrimidine-based therapy of metastatic colorectal tumours. *Magy Onkola* 48, 253–257. doi: HUON.2004.48.3.0253

Cai, X., Liu, C., Zhang, T. N., Zhu, Y. W., Dong, X., and Xue, P. (2018). Down-regulation of FN1 inhibits colorectal carcinogenesis by suppressing proliferation, migration, and invasion. *J. Cell. Biochem.* 119 (6), 4717–4728. doi: 10.1002/jcb.26651

Foroughi, K., Amini, M., Atashi, A., Mahmoodzadeh, H., Hamann, U., and Manoochehri, M. (2018). Tissue-specific down-regulation of the long non-coding RNAs PCAT18 and LINC01133 in gastric cancer development. *Int. J. Mol. Sci.* 19 (12), 3881. doi: 10.3390/ijms19123881

Ifon, E. T., Pang, A. L., Johnson, W., Cashman, K., Zimmerman, S., Muralidhar, S., et al.(2005). U94 alters FN1 and ANGPTL4 gene expression and inhibits tumorigenesis of prostate cancer cell line PC3. *Cancer Cell Int.* 5 (1), 19. doi: 10.1186/1475-2867-5-19

Kasuboski, J. M., Bader, J. R., Vaughan, P. S., Tauhata, S. B., Winding, M., Morrissey, M. A., et al. (2011). Zwint-1 is a novel Aurora B substrate required for the assembly of a dynein-binding platform on kinetochores. *Mol. Biol. Cell* 22 (18), 3318–3330. doi: 10.1091/mbc.e11-03-0213

Li, J., Xu, Q., Wang, W., and Sun, S. (2019). MIR100HG: a credible prognostic biomarker and an oncogenic lncRNA in gastric cancer. *Biosci. Rep.* 39 (4). doi: 10.1042/BSR20190171

Liu, G., Wong, L., and Chua, H. N. (2009). Complex discovery from weighted PPI networks. *Bioinformatics* 25, 1891–1897. doi: 10.1093/bioinformatics/btp311

Ma, Y., Yan, F., Li, L., Liu, L., and Sun. J., et al. (2014). Deletion and down-regulation of SMAD4 gene in colorectal cancers in a Chinese population. *Chin. J. Cancer Res.* 26, 525–531. doi: 10.3978/j.issn.1000-9604.2014.09.02

Michalak, M., Warnken, U., Schnölzer, M., Gabius, H. J., and Kopitz, J. (2019). Detection of malignancy-associated phosphoproteome changes in human colorectal cancer induced by cell surface binding of growth-inhibitory galectin-4. *IUBMB Life* 71 (3), 364–375. doi: 10.1002/iub.1987

Paul, D., Ghorai, S., Dinesh, U. S., Shetty, P., Chattopadhyay, S., and Santra, M. K. (2017). Cdc20 directs proteasome-mediated degradation of the tumor suppressor SMAR1 in higher grades of cancer through the anaphase promoting complex. *Cell Death Dis.* 8 (6), e2882. doi: 10.1038/cddis.2017.270

Ress, A. L., Perakis, S. and Pichler, M., microRNAs and Colorectal Cancer, *Oxygen Transport to Tissue XXXIII*, 889 (2015), 89-103. doi: 10.1007/978-3-319-23730-5_6

Rezanejad Bardaji, H., Asadi, M. H., and Yaghoobi, M. M. (2018). Long noncoding RNA VIM-AS1 promotes colorectal cancer progression and metastasis by inducing EMT. *Eur. J. Cell Biol.* 97, 279–288. doi: 10.1016/j.ejcb.2018.04.004

Shen, H., Cai, M., Zhao, S., Wang, H., Li, M., Yao, S., et al. (2014). Overexpression of RFC3 is correlated with ovarian tumor development and poor prognosis. *Tumour Biol.* 35 (10), 10259–10266. doi: 10.1007/s13277-014-2216-2

Sugimoto, K., Tsutsui, M., AuCoin, D., and Vig, B. K. (1999). Visualization of prekinetochore locus on the centromeric region of highly extended chromatin fibers: does kinetochore autoantigen CENP-C constitute a kinetochore organizing center? *Chromosome Res.* 7 (1), 9–19. doi: 10.1023/A:1009267010071

Sun, Q., Tripathi, V., Yoon, J. H., Singh, D. K., Hao, Q., Min, K. W., et al. (2018). MIR100 host gene-encoded lncRNAs regulate cell cycle by modulating the interaction between HuR and its target mRNAs. *Nucleic Acids Res.* 46 (19), 10405–10416. doi: 10.1093/nar/gky696

Sunamura, N., Ohira, T., Kataoka, M., Inaoka, D., Tanabe, H., Nakayama, Y., et al. (2011). Regulation of functional KCNQ1OT1 lncRNA by beta-catenin. *Sci. Rep.* 6, 20690. doi: 10.1038/srep20690

Tsukamoto, S., Ishikawa, T., Iida, S., Ishiguro, M., Mogushi, K., Mizushima, H., et al. (2011). Clinical significance of osteoprotegerin expression in human colorectal cancer. *Clin. Cancer Res.* 17 (8), 2444–2450. doi: 10.1158/1078-0432.CCR-10-2884

Xu, Q., Wang, Z., Chen, X., et al. (2015). Stromal-derived factor-1alpha/CXCL12-CXCR4 chemotactic pathway promotes perineural invasion in pancreatic cancer. *Oncotarget 6,* 4717–4732. doi: 10.18632/oncotarget.3069

Xu, C., Sun, L., Jiang, C., Zhou, H., Gu, L., Liu, Y., et al. (2017a). SPP1, analyzed by bioinformatics methods, promotes the metastasis in colorectal cancer by activating EMT pathway. *Biomed. Pharmacother.* 91, 1167–1177. doi: 10.1016/j.biopha.2017.05.056

Xu, J., Zhang, R., and Zhao, J. (2017b). The novel long noncoding RNA TUSC7 inhibits proliferation by sponging MiR-211 in colorectal cancer. *Cell. Physiol. Biochem.* 41 (2), 635–644. doi: 10.1159/000457938

Yoshida, K., Toden, S., Weng, W., Shigeyasu, K., Miyoshi, J., Turner, J., et al. (2017). SNORA21 - An oncogenic small nucleolar RNA, with a prognostic biomarker potential in human colorectal cancer. *EBioMedicine* 22, 68–77. doi: 10.1016/j.ebiom.2017.07.009

Zhan, F., Shen, J., Wang, R., Wang, L., Dai, Y., Zhang, Y., et al. (2018). Role of exosomal small RNA in prostate cancer metastasis. *Cancer Manag. Res.* 10, 4029. doi: 10.2147/CMAR.S170610

Zhang, P., Ma, Y., Wang, F., Yang, J., Liu, Z., Peng, J., et al. (2012a). Comprehensive gene and microRNA expression profiling reveals the crucial role of hsa-let-7i and its target genes in colorectal cancer metastasis. *Mol. Biol. Rep.* 39 (2), 1471–1478. doi: 10.1007/s11033-011-0884-1

Zhang, X., Sun, S., Pu, J. K., Tsang, A. C., Lee, D., Man, V. O., et al. (2012b). Long non-coding RNA expression profiles predict clinical phenotypes in glioma. *Neurobiol. Dis.* 48 (1), 1–8. doi: 10.1016/j.nbd.2012.06.004

Zhang, Z., Zhou, C., Chang, Y., Zhang, Z., Hu, Y., Zhang, F., et al. (2016). Long non-coding RNA CASC11 interacts with hnRNP-K and activates the WNT/beta-catenin pathway to promote growth and metastasis in colorectal cancer. *Cancer Lett.* 376 (1), 62–73. doi: 10.1016/j.canlet.2016.03.022

Zhou, J., Lin, J., Zhang, H., Zhu, F., and Xie, R. (2018). LncRNA HAND2-AS1 sponging miR-1275 suppresses colorectal cancer progression by upregulating KLF14. Biochem. *Biophys. Res. Commun.* 503 (3), 1848–1853. doi: 10.1016/j.bbrc.2018.07.125

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Li, Ma, Lin, Li, Yang, Zhang, Zhang and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

METHODS
published: 22 November 2019
doi: 10.3389/fgene.2019.01195

Testing Mediation Effects in High-Dimensional Epigenetic Studies

*Yuzhao Gao[1], Haitao Yang[2], Ruiling Fang[1], Yanbo Zhang[1], Ellen L. Goode[3] and Yuehua Cui[4]**

[1] Division of Health Statistics, School of Public Health, Shanxi Medical University, Taiyuan, China, [2] Division of Health Statistics, School of Public Health, Hebei Medical University, Shijiazhuang, China, [3] Department of Health Sciences Research, College of Medicine, Mayo Clinic, Rochester, MN, United States, [4] Department of Statistics and Probability, Michigan State University, East Lansing, MI, United States

Mediation analysis has been a powerful tool to identify factors mediating the association between exposure variables and outcomes. It has been applied to various genomic applications with the hope to gain novel insights into the underlying mechanism of various diseases. Given the high-dimensional nature of epigenetic data, recent effort on epigenetic mediation analysis is to first reduce the data dimension by applying high-dimensional variable selection techniques, then conducting testing in a low dimensional setup. In this paper, we propose to assess the mediation effect by adopting a high-dimensional testing procedure which can produce unbiased estimates of the regression coefficients and can properly handle correlations between variables. When the data dimension is ultra-high, we first reduce the data dimension from ultra-high to high by adopting a sure independence screening (SIS) method. We apply the method to two high-dimensional epigenetic studies: one is to assess how DNA methylations mediate the association between alcohol consumption and epithelial ovarian cancer (EOC) status; the other one is to assess how methylation signatures mediate the association between childhood maltreatment and post-traumatic stress disorder (PTSD) in adulthood. We compare the performance of the method with its counterpart *via* simulation studies. Our method can be applied to other high-dimensional mediation studies where high-dimensional mediation variables are collected.

Keywords: de-sparsify, DNA methylation, high-dimensional testing, high-dimensional mediation, mediation analysis

OPEN ACCESS

Edited by:
Xiangqin Cui, Emory University, United States

Reviewed by:
Jingying Zhou, The Chinese University of Hong Kong, China
Hao Wu, Emory University, United States

****Correspondence:***
Yuehua Cui
cuiy@msu.edu

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *10 April 2019*
Accepted: *29 October 2019*
Published: *22 November 2019*

Citation:
Gao Y, Yang H, Fang R, Zhang Y, Goode EL and Cui Y (2019) Testing Mediation Effects in High-Dimensional Epigenetic Studies. Front. Genet. 10:1195. doi: 10.3389/fgene.2019.01195

INTRODUCTION

Introduced by Baron and Kenny in 1986 (Baron and Kenny, 1986), mediation analysis has been broadly applied in many scientific disciplines, such as sociology, psychology, behavioral science, economics, epidemiology, public health science, and genetics (e.g., E.Shrout and Bolger, 2002; Preacher and Hayes, 2008; Hafeman and Schwartz, 2009; Pfeffer and Devoe, 2009; Imai et al., 2010; Rocca et al., 2010; Pearl, 2012; Pierce et al., 2014). Through solving a chain of relations between an exposure variable and an outcome, it helps to understand how the effect of one variable is transmitted to another variable. Thus, mediation analysis offers researchers a unique statistical tool to reveal the underlying mechanism or process of various scientific questions, especially when designing an intervention strategy. It has been further extended and developed *via* taking nonlinearity, interactions, various types of mediating and outcome variables, as well as missing data into account

in recent developments (e.g., Imai et al., 2010; Vanderweele and Vansteelandt, 2010; Pearl, 2012; Zhang and Wang, 2013).

Recently, mediation analysis has been applied to genetic association studies in which one can evaluate how genetic variants (e.g., single nucleotide polymorphisms (SNPs)) pass effects to mediators such as gene expression or DNA methylation (DNAm) to affect a disease risk (e.g., Liu et al., 2013; Huang et al., 2014; Huang et al., 2015). The genome-wide mediation analysis provides additional insight into the causal mechanisms of complex diseases. DNAm is an epigenetic phenomenon. Its status change reflects environmental exposures on the genome. DNAm can regulate gene expressions and can be potential biomarkers for the early prevention of stress-related disorders (Klengel et al., 2014). Properly maintained DNAms are necessary for regulating chromosomal stability and gene expressions. However, they can change the DNA activity when things go wrong, and lead to unexpected consequences. A growing body of literature shows that different environmental factors can alter the level of DNAm among individuals (e.g., Guida et al., 2015; Dongen et al., 2016). Abdolmaleky et al. (2004) showed that DNAm may modulate gene-environment interactions on psychiatry disorder. Li et al. (2003) reported that exposure to xenobiotics in early life can persistently change the pattern of DNAm, resulting in potentially adverse biological effects which may explain the increased risk in adulthood of some chronic diseases. All evidences demonstrate the important role of DNAm in mediating the effect of environmental exposures on disease outcomes. Successful identification of causal DNAm as potential biomarkers can offer novel insights into the early prevention of some diseases such as stress-related disorders.

In a typical DNAm study, the number of DNAm can be much larger than the number of sample size. Mediation analysis focusing on one mediator at a time is not efficient enough to handle thousands of mediators (e.g., CpG sites). Methods for multiple mediators have been proposed assuming different data distributions with different methods. Focusing on continuous mediators, Huang and Pan (2016) developed a testing procedure using Monte-Carlo resampling method to evaluate the statistical significance. However, it is time consuming when the computing resource is limited.

Let X be an exposure variable; M_j, j=1,…,k be the jth mediator; and Y be an outcome variable. **Figure 1** illustrates the mediation model with a single mediator (a) and multiple mediators (b). In an epigenetic study, multiple mediators could be potentially correlated. For example, methylation signals in a given gene or region are typically correlated. Such correlation, if not properly handled, can lead to potential false positives or false negatives in traditional mediation analysis.

The high-dimensional and correlation nature of DNAm signatures (**Figure 1B**) motivates us to consider a high-dimensional mediation model, which is not a trivial extension of a low dimensional multiple mediator model studied in literature. Methodology development for mediation analysis with high-dimensional mediators is still in its infancy. Zhang et al. (2016) proposed a high-dimensional mediation analysis method. They first applied a sure independence screening (SIS) method to reduce the data dimension from ultra-high to high, then adopted a penalized regression to shrinkage coefficients of irrelevant variables to zero. After the shrinkage, those mediators with non-zero coefficients were refit in a low-dimensional regression model for further hypothesis testing. Such penalized regression methods typically produce biased estimators, especially when correlations between predictors exist. This method thus could face potential issues with either false positives or false negatives. Huang and Pan (2016) proposed to transform the correlated mediators into independent ones, then performed the mediation analysis on the transformed variables. Such a method solves the correlation issue but faces the difficulty of interpretation, since the transformed variable is a linear combination of the original mediators and does not have a direct interpretation.

High-dimensional data analysis is typically formulated with high-dimensional penalized regression models, with the purpose to select important features that can minimize the prediction error. Popular methods include LASSO (Tibshiranit, 1996), adaptive LASSO (Zou, 2006), and elastic net (Zou and Hastie, 2005). Although these methods can do variable estimation and selection simultaneously, they cannot quantify the estimation uncertainty. There has been a flourish of recent literature on testing low-dimensional coefficients in high-dimensional sparse regression models (e.g., Zhang and Zhang, 2014; Dezeure et al., 2015; Zhang and Cheng, 2017; Wang and Samworth, 2018). These methods essentially implement a debias technique, then perform hypothesis testing using the debiased estimators (Zhang and Zhang, 2014). Following the asymptotic normality, one can obtain a p-value or construct a confidence interval for each coefficient (Van de Geer et al., 2014). Taking the high dimensionality and correlation issue into account, in this article, we adopt a high-dimensional testing framework and conduct simultaneous inference under a high-dimensional sparse mediation model based on the recent de-sparsifying LASSO estimators (Zhang

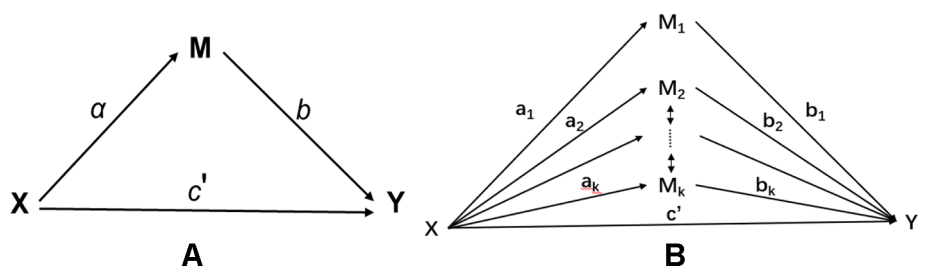


FIGURE 1 | Mediation model: **(A)** single mediator model; **(B)** multiple mediator model with correlated mediators.

and Zhang, 2014). High-dimensional testing is embedded in the mediation model to handle the high dimensionality and correlation issues between mediators. We conduct extensive simulations to evaluate the performance of the methods and compare it with its counterpart. Application to two real data sets is given. Our method can be extended to other mediation analysis where high-dimensional mediators are observed.

STATISTICAL METHOD

Figure 1A demonstrates a single mediation model. There are two types of effect from X to Y: (1) the direct effect from X to Y, denoted as c'; and (2) the indirect effect from X to Y *via* the intermediate mediation variable M. The indirect effect measures the amount of mediation which comes from two sources: i) the effect from X to M, denoted as a; and ii) the effect from M to Y, denoted as b. The product of a and b defines the indirect effect. The total effect c from X to Y contains two parts, i.e., $c = c' + ab$. By fitting three different regression models, one can use the Sobel's method (Sobel, 1982) to estimate the standard error of $\hat{a}\hat{b}$ from which the significance of mediation effect can be assessed.

The single mediator model shown in **Figure 1A** can be extended to a multiple mediator model by fitting a multiple regression model involving both the exposure and the mediator variables. The multiple mediator model is given as follows,

$$\begin{aligned} Y &= \theta_1 + cX + e_1 \\ M_j &= \theta'_j + a_j X + \varepsilon_j, j = 1,...,k, \\ Y &= \theta_2 + c'X + \sum_{j=1}^{k} b_j M_j + e_2, \end{aligned} \quad (1)$$

where M_j, j=1,..,k is the jth mediator variable; c represents the total effect from the independent variable X to the dependent variable Y; c' represents the direct effect from X to Y adjusting for the effects of multiple mediators; the indirect effect from X to Y mediated by M_j is denoted by $a_j b_j$. The total mediation effect can be obtained as $c - c'$ or $\sum_{j=1}^{k} a_j b_j$. When the response variable Y is a categorical variable, method to estimate the total mediation effect based on the product measure, $a_j b_j$, is less susceptible to the scaling problem since only the b_j coefficient is from a categorical regression analysis (MacKinnon, 2008). Model (1) is for continuous Y variable. For a categorical response, Model (1) becomes,

$$\begin{aligned} E(Y) &= \theta_1 + cX, \\ M_j &= \theta'_j + a_j X + \varepsilon_j, j = 1,...,k, \\ E(Y) &= \theta_2 + c'X + \sum_{j=1}^{k} b_j M_j. \end{aligned} \quad (2)$$

As we mentioned in the *Introduction* section, a genomic mediation study often involves high-dimensional mediators. In many cases, the number of mediators is far beyond the sample size (k>>n). For example, the number of DNAm loci can be nearly half million, far more than the sample size. Another phenomenon for genomic mediators is that they are often correlated. Both the curse of dimensionality and correlation between mediators cause estimation problems in Model (1) and (2). Classical regression analysis cannot be directly adopted to deal with the estimation and testing problem appeared in the third equation in Model (1) and (2). To solve both the high dimensionality and correlation problem, we propose to adopt a high-dimensional testing framework which is focused on de-sparsified LASSO estimators (Zhang and Zhang, 2014). The detailed estimation and testing procedure for the proposed high-dimensional mediation testing framework is given as follows:

Step 1: First apply an SIS procedure to reduce the methylation dimension from ultra-high to high dimension (Fan and Lv, 2008). According to the SIS algorithm, the top d=n/log(n) methylation variables with the largest effects were remained in the model when the response Y is a continuous variable. For a binary response, the top d=n/log(n) variables can be kept in the model. SIS theoretically guarantees that no true signals are removed from the model. The SIS step can be based on the third or the second regression equation in Model (2). For a binary response Y, Zhang et al. (2016) suggested that SIS can be done based on the second equation in Model (2). For a continuous response variable, the SIS step can be done based on the third regression equation in (2). After SIS, the number of methylation loci is reduced from k to d. We then focused our analysis to these d methylation variables to test mediation effects. Denote the remaining methylation loci after the SIS step as M_j,j=1,…,d.

Step 2: In the second step, we fit the following model,

$$E(Y) = \theta_2 + c'X + \sum_{j=1}^{d} b_j M_j \quad (3)$$

Other covariates can also be fitted to this model. Since the dimension d can still be relatively large after the SIS step, regular least squares estimation will not work well. For high-dimensional data, penalized regressions are commonly applied for simultaneous variable selection and estimation. However, penalized estimators are biased and cannot be directly used for testing or confidence interval construction. Zhang and Zhang (2014) first time proposed a de-biased estimator for high-dimensional data. Let $\hat{b}_{lasso}$ be the LASSO estimators. For a continuous response variable Y, A de-biased estimator, also called a de-sparsified estimator, is a bias-corrected estimator which can be given as,

$$\hat{b}_j = \frac{Z_j^T Y}{Z_j^T M_j} - \sum j \neq l \frac{Z_j^T M_l}{Z_j^T M_j} \hat{b}_{lasso,l} \quad (4)$$

where $\hat{b}_j$ is the bias-corrected coefficient of the jth methylation M_j; $\hat{b}_{lasso,l}$ is the coefficient of the lth M_l estimated by fitting a LASSO regression; Z_j is the regularized residuals obtained by $Z_j = M_j - M_{-j}\hat{\gamma}_{lasso}$, where $\hat{\gamma}_{lasso}$ is the regression coefficients obtained based on a LASSO regression by regressing M_j on all other M except the jth M_j denoted as M_{-j}. Van de Geer et al.

(2014) proved the asymptotic normality of the de-sparsified estimate, i.e.,

$$u_j = \frac{\sqrt{n}\left(\hat{b}_j - \gamma_j^0\right)}{\sigma_\epsilon \sqrt{\Omega_{jj}}} \xrightarrow{d} N(0,1) \quad as \quad p \geq n \to \infty$$

where γ_j^0 represents the true regression coefficient; σ_ϵ can be calculated by using the scaled LASSO algorithm (Sun and Zhang, 2012), and Ω_{jj} can be calculated by,

$$\Omega_{jj} = \frac{nZ_j^T Z_j}{\left[Z_j^T Z_j\right]\left[Z_j^T Z_j\right]}$$

Under the null that $H_0 : \gamma_j^0 = 0$, we can get *p*-values for all the *d* methylation loci based on the asymptotic normality (Van De Geer et al., 2014).

For a binary response, Van de Geer et al. (2014) also proved the asymptotic normality for the de-sparsified estimates. Let $W = \left(X, M\right)^T$, $\boldsymbol{\beta} = \left(c', \boldsymbol{b}\right)^T$, and $L_{\boldsymbol{\beta}}\left(y, W\right) = L\left(y, W\boldsymbol{\beta}\right)$ be a loss function, and define $\dot{L}_\beta = \frac{\partial}{\partial \beta} L_\beta$ and $\ddot{L}_\beta = \frac{\partial^2}{\partial \beta \partial \beta^T} L_\beta$, and further define $\varphi_L := \sum_{i=1}^{n} L\left(y_i, w_i^T \boldsymbol{\beta}\right) / n :=$ The LASSO estimator for the mediation coefficients β is given as $\hat{\boldsymbol{\beta}} = \arg\min\left(\varphi_L + \lambda \forall \| \boldsymbol{\beta} \|_1\right)$, where λ is a tuning parameter. Define $\hat{\Sigma} := \ddot{\varphi}_{L_{\hat{\beta}}}$ and construct $\hat{\Theta} = \hat{\Theta}_{LASSO}$ by doing a nodewise LASSO with $\hat{\Sigma}$ as input. Then the de-sparsified LASSO estimator is given as $\tilde{\boldsymbol{\beta}} := \widehat{\boldsymbol{\beta}} - \hat{\Theta} \dot{\varphi}_{L_{\hat{\beta}}}$. van de Geer et al. (2014) provided a detailed algorithm for computing the de-sparsified LASSO estimators in a generalized linear model framework. They also proved the asymptotic normality of the de-sparsified estimate, i.e.,

$$u_j = \frac{\sqrt{n}\left(\tilde{\beta}_j - \gamma_j^0\right)}{\hat{\sigma}_j} \xrightarrow{d} N(0,1) \quad as \quad p \geq n \to \infty$$

where $\hat{\sigma}_j^2 = \left(\hat{\Theta} P_{\ddot{L}_{\hat{\beta}} i \dot{L}_{\hat{\beta}}^T} \hat{\Theta}^T\right)_{j,j}$. Similarly, we can get a p-value for each mediator based on the asymptotic normality property.

Let the p-values for all the *d* methylation loci denoted as $P_b = (P_{1,b}, P_{2,b}, \ldots, P_{d,b})$ where $P_{j,b}$ can be calculated as $P_{j,b} = 2\left\{1 - \Phi\left(\frac{\sqrt{n}\hat{b}_j}{\sigma_\epsilon \sqrt{\Omega_{jj}}}\right)\right\}$ for a continuous *Y* or as $P_{j,b} = 2\left\{1 - \Phi\left(\frac{\sqrt{n}\hat{b}_j}{\hat{\sigma}_j}\right)\right\}$ for a discrete *Y*.

Step 3: Let $S = \{t : P_{t,b} < 0.05\}$, which is based on the high-dimensional inference in the second step. For testing $H_0 : a_t = 0$, we denote the testing p-value as $P_{t,a}$

$$P_{t,a} = 2\left\{1 - \Phi\left(\frac{|\hat{a}_t|}{\hat{\sigma}_t}\right)\right\},$$

where $t \in S$, $\hat{a}_t$ is the ordinary least squares estimator for a_t and $\hat{\sigma}_t$ is the corresponding estimated standard error, by fitting the 2nd regression equation in Model (2).

Step 4: We reject the null hypothesis of no mediation effect for M_t only if both a_t and b_t are significant. The p-value for the joint significance test is defined as,

$$P_t^* = \max\left(P_{t,a}, P_{t,b}\right)$$

A methylation locus has a significant mediation effect if $P_t^* < 0.05$. This is also a so called intersection-union test (Berger and Hsu, 1996).

Remark 1: To make the paper self-contained, here we briefly introduce the High-dimensional mediation analysis (HIMA) method proposed by Zhang et al. (2016). The HIMA method involves three major steps:

Step 1: (Screening) Use the SIS (Fan and Lv, 2008) to identify a subset of top mediators.

Step 2. (MCP-penalized estimate). Apply the MCP-based penalized regression to do simultaneous variable selection and estimation based on the variables from step 1.

Step 3. (Joint significance test). For those mediators with non-zero coefficients from step 2, fit a regression model again and get a p-value for testing each coefficient, then, taking the maximum of this p-value and the p-value for testing the α effect as the final p-value to assess the significance of the mediation effect.

Remark 2: Our method has two advantages: 1) It fits multiple mediators in one regression model and do the testing, rather than fitting and testing mediation effect one at a time. Statistically speaking, this yields more robust and efficient estimation and testing results; and 2) Different from Zhang et al. (2016), our method is a simultaneous inference in a high-dimensional sparse regression model implemented with a de-biasing technique. The de-sparsifying strategy can well handle correlations between methylation loci, as demonstrated in the simulation study.

SIMULATION STUDIES

We conduct extensive simulations to evaluate the performance of the proposed method and compare it with the HIMA method proposed by Zhang et al. (2016). In the follows, we denote our method as HDMA (high dimensional mediation analysis) and the method by Zhang et al. (2016) as HIMA. Data are generated following Model (2), where the exposure variable *X* is generated from a binomial distribution, i.e., *B*(*n*,0.74) in which the probability 0.74 is determined based on the proportion of drinking in the first real set (see the real data analysis section for details). To have a fair comparison, we follow the simulation setup for the

regression coefficients as given in Zhang et al. (2016). The first 8 elements of $b(b_j, j = 1,\ldots,8)$ are given as $(0.8,0.7,0.6,0.5,0,0,0.5,0.5)^T$, and the first 8 elements of $a(a_j, j = 1,\ldots,8)$are given as $(0.35,0.25, 0.35,0.55,0.55,0.55,0,0)^T$. The rest of *as* and $b's$ are all set to zero. Under this setting, the first four methylation loci have significant mediation effects while the rest have no effect.

For the intercept terms, we set $\theta_2 = -4.5$ and $\theta_j^{'} = 1$. We also consider different correlations among the mediators, i.e., $\rho = 0$, and 0.8. When the direct effect $c' = 0$, the model is a complete mediation model in which exposures affect outcome only through mediators. In this case, the total effect $c = c' + \sum_{j=1}^{k} a_j b_j = 0.94$. When the direct effect $c' > 0$, the model is a partial mediation model. For the partial mediation model, we set $c' = 0.5$ and the total effect $c = c' + \sum_{j=1}^{k} a_j b_j = 1.44$.

We simulate k methylation loci which follow a multivariate normal distribution, i.e., $M_i \sim MVN\left(1 + a_i X_i, \Sigma\right)$, where $a_i = \left(0.35,\ 0.25,\ 0.35,\ 0.55,\ 0.55,\ 0.55, \underbrace{0,\ldots,0}_{k-6}\right)^T$ and $\Sigma_{st} = \rho^{|s-t|}$. Then we sample the response $Y_i \sim Ber(1, p_i)$, where $p_i = \exp(\eta_i)/\left(1 + \exp\left(\eta_i\right)\right)$ and $\eta_i = -4.5 + c' X_i + \sum_{j=1}^{k} b_j M_{ij}$.

We evaluate the performance of our method (HDMA) in terms of false positive rate and power and compare with HIMA. We report the power ($M_1 \sim M_4$) and the type I error ($M_5 \sim M_8$) for each locus. For the rest of the k-8 loci, we report the averaged type I error rate. All simulations are based on 1000 replications under different sample sizes, i.e., $n = 300$ and 600 and different correlations, i.e., $\rho = 0$ and 0.8.

Table 1 lists the results for binary responses assuming a complete mediation effect, i.e., $c' = 0$. There are several observations: (i) HIMA and HDMA have very similar power and size when there are no correlations between M ($\rho = 0$) under different scenarios. However, HDMA has substantially higher power than HIMA does when $\rho = 0.8$; (ii) The testing power decreases as the data dimension increases for both methods. For example, the power of testing M_1 is 0.754 for HDMA with $k = 100$, but decreases to 0.721 with $k = 5000$, when fixing $n = 300$ and $\rho = 0$; (iii) The power increases as the sample size increases. For example, when fixing $\rho = 0.8$ and $k = 1000$, the power increases from 0.598 to 0.951 for testing M_1 when the sample size increases from 300 to 600, a 59% increase; and (iv) HDMA is not sensitive to the correlation structures while HIMA suffers significantly from power loss when there are high correlations between the M variables. The difference is even more striking when the sample size increases from 300 to 600. For example, the power difference for testing M_1 is 0.014 for HDMA compared to 0.238 for HIMA when ρ is increased from 0 to 0.8, when fixing $n = 600$ and $k = 1000$. Similar patterns were observed for the other three M variables.

Figure 2 summarizes the results with partial mediation, i.e., $c' = 0.5$. We consider $N = 300$ and 600, $p = 100$, 1000 and 5000, and $\rho = 0$ and 0.8. Corresponding to each mediator, there are four power bars. The left two correspond to the case with correlation $\rho = 0$, while the right two correspond to the case with $\rho = 0.8$. For a fixed sample size, the power typically decreases as the data dimension (p) increases. This is because of the increase of the noise features. When $\rho = 0$ (the independent case), HIMA and HDMA perform very similarly under different scenarios. However, when the correlation increases to $\rho = 0.8$, we observe a power gain by HDMA compared

TABLE 1 | List of the power and type I error rate under different sample sizes and correlations with data analyzed with HDMA and HIMA.

n	*k*		Method	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_{other}
300	100	0	HIMA	0.754	0.467	0.723	0.849	0.025	0.022	0.034	0.047	0.001
			HDMA	0.754	0.460	0.713	0.825	0.021	0.017	0.034	0.046	0.001
		0.8	HIMA	0.502	0.241	0.362	0.377	0.075	0.070	0.028	0.019	0.001
			HDMA	0.649	0.348	0.445	0.422	0.062	0.062	0.023	0.012	0.000
	1000	0	HIMA	0.763	0.478	0.653	0.702	0.008	0.008	0.049	0.029	0.001
			HDMA	0.763	0.476	0.660	0.697	0.008	0.006	0.044	0.032	0.000
		0.8	HIMA	0.513	0.194	0.370	0.386	0.078	0.072	0.013	0.023	0.000
			HDMA	0.598	0.297	0.399	0.417	0.060	0.055	0.012	0.019	0.000
	5000	0	HIMA	0.714	0.437	0.590	0.528	0.003	0.002	0.024	0.029	0.000
			HDMA	0.721	0.440	0.589	0.549	0.002	0.002	0.027	0.027	0.000
		0.8	HIMA	0.545	0.182	0.374	0.386	0.081	0.076	0.024	0.017	0.000
			HDMA	0.577	0.267	0.413	0.388	0.047	0.045	0.017	0.013	0.000
600	100	0	HIMA	0.957	0.769	0.969	0.990	0.008	0.010	0.046	0.051	0.001
			HDMA	0.957	0.769	0.969	0.996	0.019	0.015	0.046	0.052	0.001
		0.8	HIMA	0.776	0.352	0.505	0.476	0.044	0.047	0.027	0.018	0.001
			HDMA	0.950	0.686	0.781	0.602	0.069	0.059	0.022	0.021	0.001
	1000	0	HIMA	0.965	0.770	0.967	0.979	0.004	0.004	0.039	0.043	0.000
			HDMA	0.965	0.770	0.966	0.977	0.013	0.008	0.040	0.043	0.001
		0.8	HIMA	0.727	0.366	0.494	0.443	0.052	0.046	0.037	0.015	0.000
			HDMA	0.951	0.685	0.790	0.632	0.060	0.071	0.026	0.021	0.000
	5000	0	HIMA	0.962	0.760	0.959	0.945	0.005	0.007	0.057	0.054	0.000
			HDMA	0.963	0.761	0.960	0.941	0.005	0.007	0.054	0.054	0.000
		0.8	HIMA	0.733	0.391	0.503	0.472	0.068	0.058	0.041	0.025	0.000
			HDMA	0.924	0.666	0.759	0.604	0.070	0.067	0.039	0.024	0.000

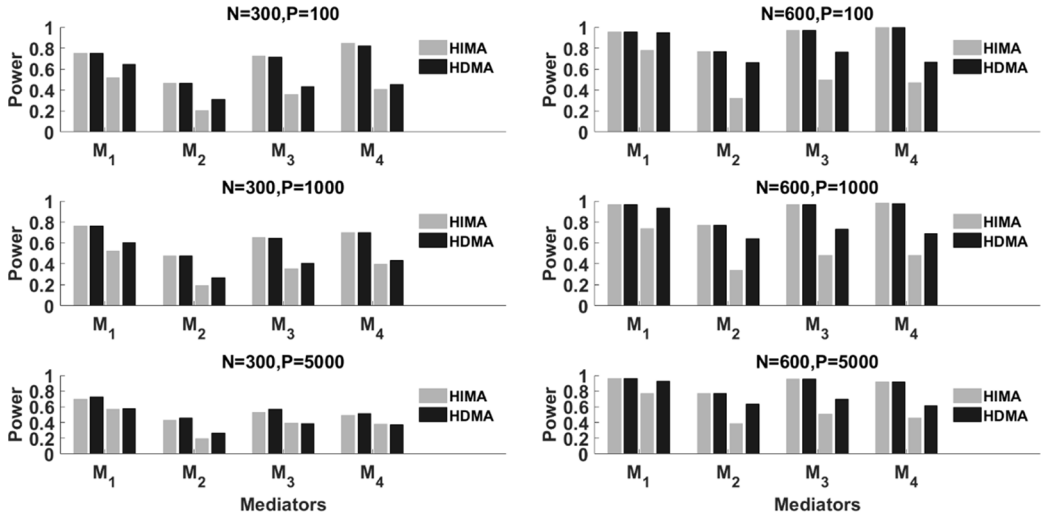


FIGURE 2 | The power of HIMA (light gray) and HDMA (black) under different sample sizes, data dimensions, and correlations. M_1~M_4 refer to the first four significant mediators. There are four power bars corresponding to each mediator. The left and right two bars correspond to the case with correlation ρ=0 and 0.8 respectively.

to HIMA under a sample size of 300. As the sample size increases from 300 to 600, we observe substantial power gain for HDMA. This shows the advantage of HDMA which can take care of the high correlation structure among the mediators.

Figure 3 displays the type I error rate of the two methods. M_{other} represents all p-8 zero effect mediators. The type I error for M_{other} is calculated as the average type I error of the p-8 mediators. Again, each mediator has four bars. The left two correspond to ρ=0 while the right two correspond to ρ=0.8. Overall, the type I errors for the two methods are reasonably controlled, especially under a large sample size (N = 600). When the correlation is high, i.e., ρ=0.8, for some mediators such as M_5 and M_6, HIMA has a higher false positive rate than HDMA does. This indicates the advantage of HDMA in false positive control when there are high correlations among mediators.

In summary, HDMA shows relative advantages over HIMA under different scenarios, especially when there are high correlations among mediators. As correlations are highly expected in real methylation data, HDMA can be an alternative strategy to HIMA and is generally safe to apply.

REAL DATA ANALYSIS

We apply the HDMA method to two real data sets with methylation loci as the mediators. DNAms play key roles in

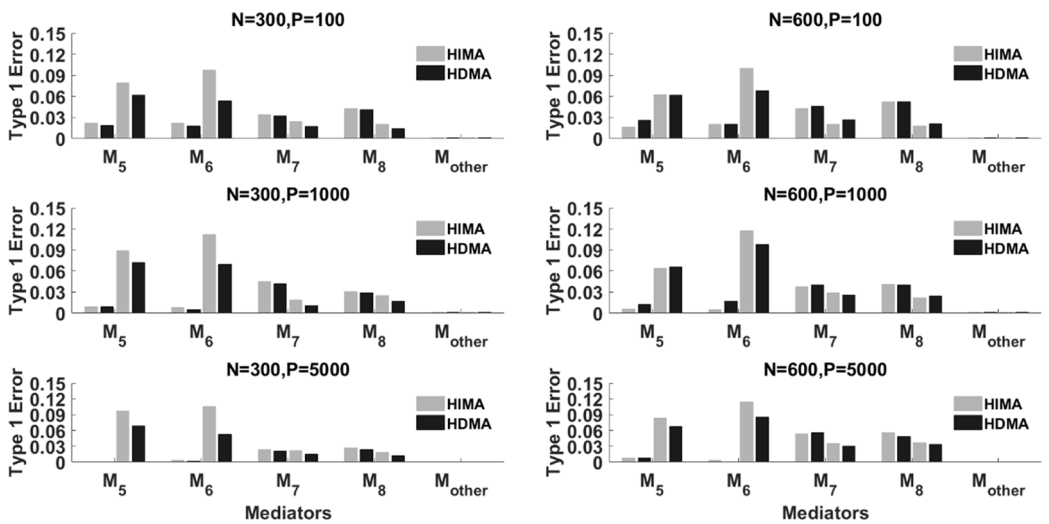


FIGURE 3 | The type I error of HIMA (light gray) and HDMA (black) under different sample sizes, data dimensions, and correlations. M_1 and M_4 refer to the first four significant mediators. There are four power bars corresponding to each mediator. The left and right two bars correspond to the case with correlation ρ=0 and 0.8 respectively.

regulating many cellular processes and are associated with human diseases (Robertson 2005). The first data set involves DNAm mediating the effect of alcohol consumption on epithelial ovarian cancer (EOC) status. Alcohol may induce DNAm alterations, which could trigger alcohol-induced carcinogenesis (Varela-Rey et al., 2013). In the second data set, we evaluate the effect of childhood maltreatment on post-traumatic stress disorder (PTSD) in adulthood, mediated by DNAms. It is hypothesized that childhood maltreatment affects biological processes *via* DNAm, which can have negative consequences late in life (e.g., Mehta et al., 2013; Klengel et al., 2016).

Case Study 1: Mediation Analysis of Alcohol Consumption, DNam, and EOC Status

The participants with age ranging from 27 to 91 were recruited between the year 1999 and year 2007 in the Mayo Clinic Ovarian Cancer. They were women of European ancestry who were invasive EOC cases and controls one-to-one matched on the basis of age (within 1-year). After eliminating missing values and other quality control, 196 cases and 202 controls were retained for further analysis. The exposure variable is alcohol consumption. Information on alcohol use was obtained *via* a written questionnaire asking "Do you currently drink alcoholic beverages?". DNAms are the mediators and EOC status is the outcome. We would like to identify the mediators and further quantify the mediation effect. Readers are referred to Koestler et al. (2014) and Wu et al. (2018) for more details about the data.

Table 2 summarizes the lifestyle and demographic characteristics of the study population. The Student *t*-test or Chi-square test is used for comparisons between groups for continuous or categorical variables, respectively. As can be seen in the table, alcohol consumption is significantly lower in cases compared to controls. Enrollment year shows a significant difference in proportions between cases and controls. Thus, we include the enrollment year as a covariate in further mediation analysis.

Leukocyte-derived DNA was assayed with the Illumina Infinium HumanMethylation27 Beadchip platform and underwent quality control procedures at the Mayo Clinic Molecular Genome Facility (Koestler et al., 2014). The methylation beta values (β) of each CpG locus was logit-transformed ($\log(\beta/(1-\beta))$) to get the M-value for further analysis. A total of 25,926 CpG sites were remained for analysis after normalization and adjusting for any batch or plate effects. Study shows that heterogeneity in white blood cells has the potential to confound DNAm measurements and statistical treatment is needed to correct for this confounding effect (Adalsteinsson et al., 2012). Similarly, variation in cell-type proportions across samples has the potential to confound the mediation effect of DNAm on the association of alcohol consumption and EOC status (Titus et al., 2017). We thus include the predicted proportions of the leukocyte sub-types for each of the study samples as covariates in the analysis, following a mixture deconvolution method by Houseman et al. (2012).

Since the response is a binary variable, we apply a logistic regression for the first and third regression equation in Model (2), while including enrollment year as a covariate. Note that the cell type data should be included whenever methylation signals are included in the model. Including the enrollment year (Enroll) and the proportion of cell type (CellType), Model (2) becomes,

$$\begin{aligned}
&\text{logit}(P)=\theta_1+c_{Alcohol}\text{Alcohol}+\lambda_1^T\text{Enroll}\\
&\text{CpG}_j=\theta_j'+a_j\text{Alcohol}+\lambda_2^T\text{Enroll}+\delta_1^T\text{CellType}+\varepsilon j.j=1,\ldots,k,\\
&\text{logit}(P)=\theta_2+c'_{Alcohol}\text{Alcohol}+\sum_{j=1}^{k}b_jCpG_j+\lambda_3^T\text{Enroll}+\delta_2^T\text{CellType}
\end{aligned}$$

The coefficient estimates for the total effect is given as $\hat{c}_{Alcohol}$=-1.310 (p-value < 0.001), indicating a significant protective effect of alcohol consumption on EOC status.

We apply the SIS algorithm to reduce the methylation dimension to 34 ($n/2\log(n)$), then apply the HDMA and HIMA methods for further inference. **Table 3** lists the findings by the two methods. Our method identified four CpGs with important

TABLE 2 | Partial list of covariates and their association with case/control status.

	Case (N = 196)	Control (N = 202)	Total (N = 398)	p value
Age at diagnosis/interview				
Mean(SD)	62.31 (12.36)	62.37 (12.69)	62.34 (12.51)	0.965
Enrollment year				
1999–2002 year	76 (38.78%)	91 (45.05%)	167 (41.96%)	<0.001
2003 year	17 (8.67%)	27 (13.37%)	44 (11.06%)	
2004 year	25 (12.76%)	42 (20.79%)	67 (16.83%)	
2005 year	30 (15.31%)	17 (8.42%)	47 (11.81%)	
2006–2007 year	48 (24.49%)	25 (12.38%)	73 (18.34%)	
Alcohol use at study enrollment				
Yes	123 (62.76%)	172 (85.15%)	295 (74.1%)	<0.001
No	73 (37.24%)	30 (14.85%)	103 (25.9%)	
Minnesota (MN) state				
Other	93 (47.45%)	82 (40.59%)	175 (43.97%)	0.202
MN	103 (52.55%)	120 (59.41%)	223 (56.03%)	
Smoking at study enrollment				
No	178 (90.82%)	192 (95.05%)	370 (92.96%)	0.145
Yes	18 (9.18%)	10 (4.95%)	28 (7.04%)	

mediation effects while HIMA identified two CpGs. Two CpGs, namely cg12278770 and cg03012280, overlap in two methods. A heatmap in **Figure 4** shows that there are moderate correlations among the 34 CpG sites. Thus, it is not surprising to see that HDMA identifies more CpG mediators than HIMA does.

CpG site cg18394848 resides in gene *K-RAS*. Nakayama et al. (2008) examined the *K-RAS* mutations in relation to extracellular signal-regulated protein kinase (*ERK*) activation in 58 ovarian carcinomas. Auner et al. (2009) drew a conclusion that *K-RAS* mutation is a common event in ovarian cancer primarily in carcinomas of lower grade, lower FIGO stage, and mucinous histotype. KEGG pathway shows that this gene is involved in the pathogenesis of ovarian cancer (**Figure 5**). This evidence indicates that cg18394848 could be an important epigenetic marker which mediates the effect of alcohol consumption on EOC pathogenesis.

Elgaaen et al. (2010) found that gene *KSP37* correlates strongly with histology, stage, and outcome in ovarian carcinomas. Thus, cg08132711 (in gene *KSP37*) can also be a potential epigenetic marker associated with the EOC status. Although we do not find direct literature support about the two genes *FAM167B* and *ZFYVE19* where cg12278770 and cg03012280 are respectively located in, a two samples t-test results show that there are significant differences on methylation signals of cg12278770 and cg03012280 between cases and controls. The *t*-test statistics (p-value) are $t_{cg12278770}$=4.881(P<0.001) and $t_{cg0301220}$=5.415(P<0.001). It suggests that these two CpG sites may act as important players to mediate the effect of alcohol intake on EOC status (**Figure 6**).

Case Study 2: Mediation Analysis of Childhood Maltreatment, Dnam, and PTSD

The data came from the Grady Trauma Project study recruiting Afro-American participants from Atlanta inner-city residents, approved by the Institutional Review Board of Emory University School of Medicine and Grady Memorial Hospital (Wingo et al., 2018). A growing body of literature indicates that DNAm plays pivotal roles in the disease process of PTSD and in vulnerability and resilience to PTSD (Uddin et al., 2011; Lutz and Turecki, 2014). Studies also show that childhood maltreatment is associated with DNAm changes of multiple loci in adulthood (Mehta et al., 2013). We apply the proposed method to establish the link between childhood maltreatment and PTSD and further evaluate the mediating role of DNAm. The data set contains baseline information, cell composition, and DNAm. We adopt the modified PTSD Symptom Scale (PSS) and the Beck Depression Inventory (BDI) to classify cases and controls. Cases with current symptoms of comorbid PTSD and depression are

TABLE 3 | List of significant CpGs identified by HDMA and HIMA.

Method	CpG	Chr	Gene name	$\hat{a}$	$\hat{b}$	$\hat{a}\hat{b}$	% of total effect	*p*-value
HDMA	cg18394848	12	*K-RAS*	−0.076	1.772	−0.136	10.343	0.008
	cg08132711	4	*KSP37*	−0.080	1.207	−0.096	7.313	0.033
	cg12278770	1	*FAM167B*	−0.071	−2.045	0.144	11.012	0.005
	cg03012280	15	*ZFYVE19*	−0.175	−0.878	0.153	11.709	0.002
HIMA	cg12278770	1	*FAM167B*	−0.071	−0.828	0.058	4.461	0.002
	cg03012280	15	*ZFYVE19*	−0.175	−0.525	0.092	7.010	0.004

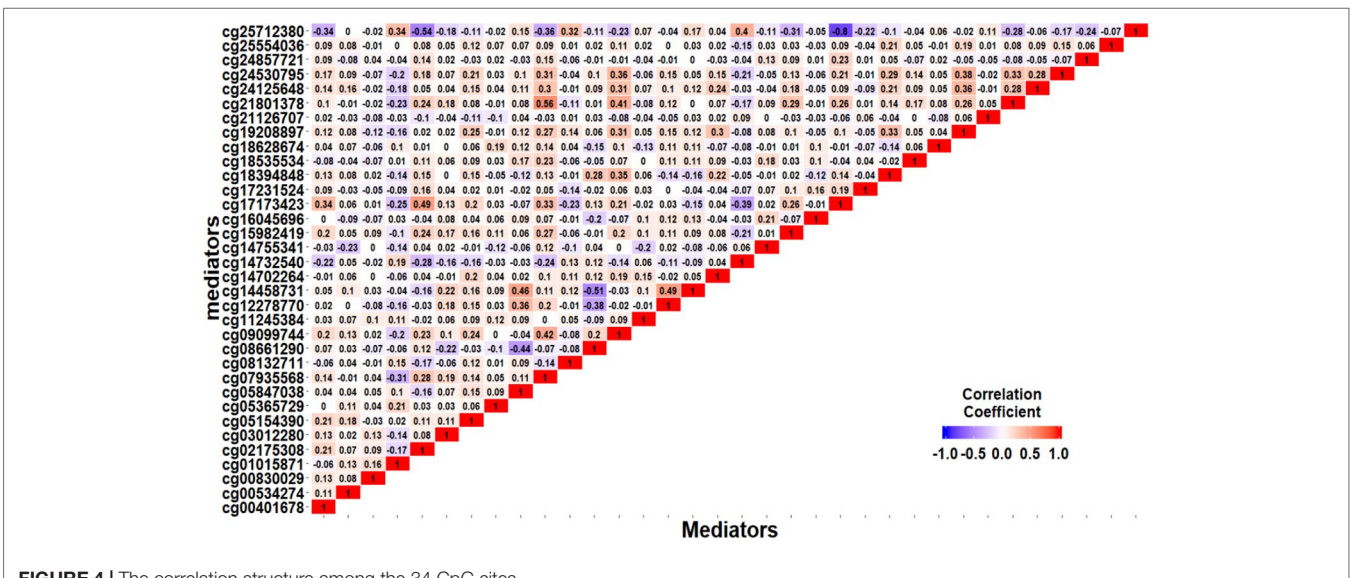


FIGURE 4 | The correlation structure among the 34 CpG sites.

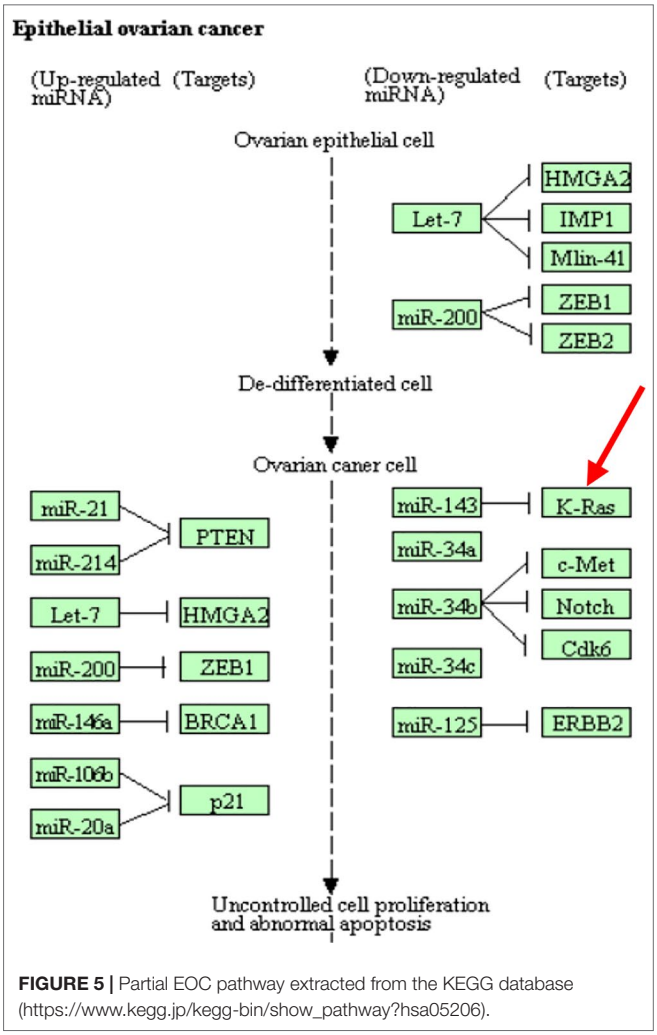


FIGURE 5 | Partial EOC pathway extracted from the KEGG database (https://www.kegg.jp/kegg-bin/show_pathway?hsa05206).

defined as having a PSS score ≥14 and a BDI score ≥14. Controls are defined as having neither PTSD nor depressive symptoms, as mirrored by a PSS score ≤7 and BDI score ≤7, despite being exposed to trauma (Beck et al., 1961; Foa et al., 2000; Wingo et al., 2018). We eliminate observations with missing values and exclude those with PTSD treated since the treatment might affect DNAm changes which can complicate the mediation effect. Finally, 54 controls and 74 cases are retained for further analysis.

Table 4 summarizes the demographic characteristics of the study population. Ranges of age in case and control are (27.97, 57.97) and (30.69, 56.79), respectively. There is no statistical significance among the selected variables such as age, sex, and body mass index (BMI), but childhood sexual/physical abuse moderate to extreme is significantly higher for cases compared to controls. The same analysis plan as detailed in Case Study 1 is applied here. Since no clinical factors show statistical significance, we do not include any covariates in our mediation model. Next, we apply HDMA and HIMA to test which DNAm plays a mediating role between childhood maltreatment and PTSD.

The raw methylation beta values from the HumanMethylation 450k BeadChip (Illumina) are obtained *via* the Illumina Beadstudio program. Samples with probe detection call rates <90% and those with an average intensity value of either <50% of the experiment-wide sample mean or <2,000 arbitrary units (AU) are excluded from further analysis. The beta values are further converted to M-values and a total of 335,669 CpG sites are used for subsequent analysis. For the details of the data, readers are referred to the website http://gradytraumaproject.com/. The data set can be downloaded at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE72680.

Lutz and Turecki (2014) reviewed human studies indicating that early-life experiences (e.g., childhood maltreatment) regulate life-long stress activities (e.g. psychopathological disorders) through epigenetic regulations (e.g., DNAms). Klengel et al. (2014) found that exposure to stress can induce long-lasting changes in DNAs, which may relate to the pathophysiology of depression and PTSD. This evidence suggests that a mediation model can help to understand how childhood maltreatment can alter long lasting DNAm changes which further affect phycological disorders such as PTSD. We fit the following mediation model while adjusting for the cell type effect whenever CpG sites are involved, i.e.,

$$\text{logit}(P)=\theta_1+c_{Maktreatment}\text{Maltreatment},$$
$$\text{CpG}_j=\theta'_j+a_j\text{Maltreatment}+\lambda_2^T+\delta_1^T\text{CellType}+\varepsilon j.j=1,\ldots,k,$$
$$\text{logit}(P)=\theta_2+c'_{Maltreatment}\text{Maltreatment}+\sum_{j=1}^{k}b_jCpG_j+\lambda_2^T\text{CellType}.$$

Based on the first regression model, we identify an existing relationship between childhood maltreatment and PTSD with $\hat{c}_{Maltreament}=1.866$ (95% CI: [1.091, 2.698]) by fitting a logistic regression model. When doing the SIS step to screen CpG sites,

TABLE 4 | Partial list of covariates and their association with PTSD case/control status.

Variables	Case (N = 74)	Control (N = 54)	Total (N = 128)	p-value
Age				
Mean (SD)	40.97 (13.00)	43.74 (13.05)	42.141 (13.04)	0.238
Sex				
Male	52 (70.27%)	36 (66.67%)	88 (68.75)	0.771
Female	22 (29.73%)	18 (33.33%)	40 (31.25)	
BMI				
Mean (SD)	31.433 (7.82)	31.614 (8.10)	31.510 (7.91)	0.899
Childhood sexual/physical abuse moderate to extreme				
No	26 (35.14%)	42 (77.78)	68 (53.13%)	<0.001
Yes	48 (64.87%)	12 (22.22)	60 (46.88%)	

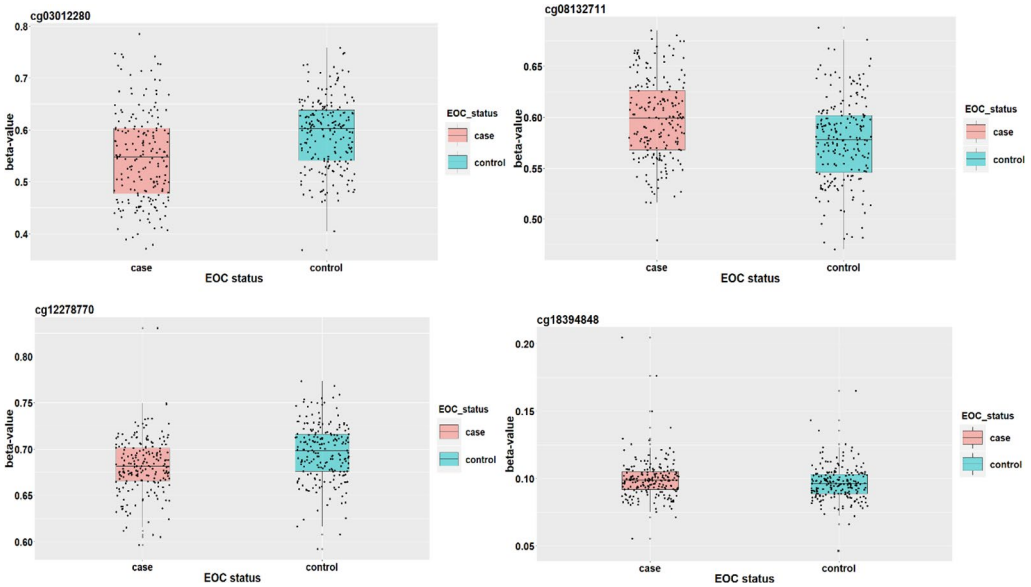


FIGURE 6 | The DNAm status (β value) of cg18394848, cg08132711, cg12278770, and cg03012280 between EOC cases and controls.

we keep $n/\log(n)$ mediators rather than $n/2\log(n)$ to avoid missing important loci, due to the small sample size. After the SIS step, 27 DNAm sites are left in the model for further analysis. **Table 5** summarizes the results. HDMA identifies two significant CpG sites (cg06998765 and cg16928335) which reside in gene *RPS6KL1* on chromosome 12 and gene *SH2D1A* on chromosome X, respectively. The two CpG sites, cg06998765 and cg16928335, respectively explain 22.73% and 19.95% of the total mediation effect. HIMA identifies one CpG site which is a subset of what HDMA detected. A heatmap of the 27 methylation signals after SIS is shown in **Figure 7**. It is clear that there are strong correlations between some CpG sites and it is not surprising that HDMA identified one more CpG site since it can handle correlation well. We further test the methylation signal difference between cases and controls for the two CpG sites and the results show significant differences for cg06998765 ($t = 4.109$, $P<0.001$) and cg16928335 ($t = 2.242$, $P = 0.027$).

Figure 8 plots the methylation signals between cases and controls for the two CpG sites. Ward et al. (2017) applied a genome-wide analysis method to analyze UK Biobank data and identified four loci associated with mood instability. Gene *RPS6KL1* is located nearby one of these regions, suggesting a potential role of this DNAm on PTSD. Although we cannot find evidence to support the association between PTSD and gene *SH2D1A* where cg06998765 is located, a two samples t-test shows that there is a significant difference on methylation signal of cg06998765 between cases and controls. The upshot suggests that this CpG site may have an important role to mediate the effect of childhood maltreatment on PTSD (**Figure 8**).

DISCUSSION

A large body of literature has suggested that environmental exposures can leave epigenetic tags such as DNAm changes which further affect disease risks. Such a causal relationship can be better understood with a causal mediation model, with the hope to identify important epigenetic players (e.g., DNAm) that mediate the relationship between an exposure and a disease outcome. As biotechnology getting cheaper and cheaper, the pace of generating epigenetic data becomes faster and faster. In many applications, the number of epigenetic features can be much larger than the sample size, resulting in the so-called (ultra-) high dimensional data. These high-dimensional data provide unprecedented opportunity to reveal the molecular mechanism of many diseases. In the meantime, they also challenge the traditional mediation analysis methods which are developed for low-dimensional data.

In this work, we propose a high-dimensional mediation model to tackle issues due to high dimensionality and high correlation. Different from the HIMA approach developed by Zhang et al.

TABLE 5 | List of significant CpGs identified by HDMA and HIMA.

Method	CpG	Chr	Gene name	$\hat{a}$	$\hat{b}$	$\hat{a}\hat{b}$	% of total effect	*p*-value
HDMA	cg06998765	12	RPS6KL1	0.266	1.594	0.424	22.748	0.020
	cg16928335	X	SH2D1A	−0.222	−1.674	0.372	19.933	0.046
HIMA	cg06998765	12	RPS6KL1	0.266	0.535	0.142	7.635	0.030

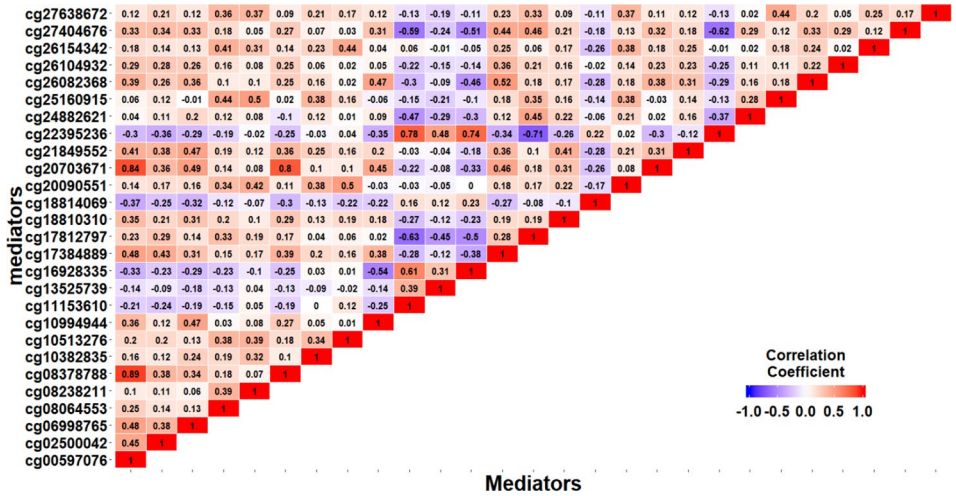


FIGURE 7 | Heatmap of 27 methylation signals after screening with the SIS procedure.

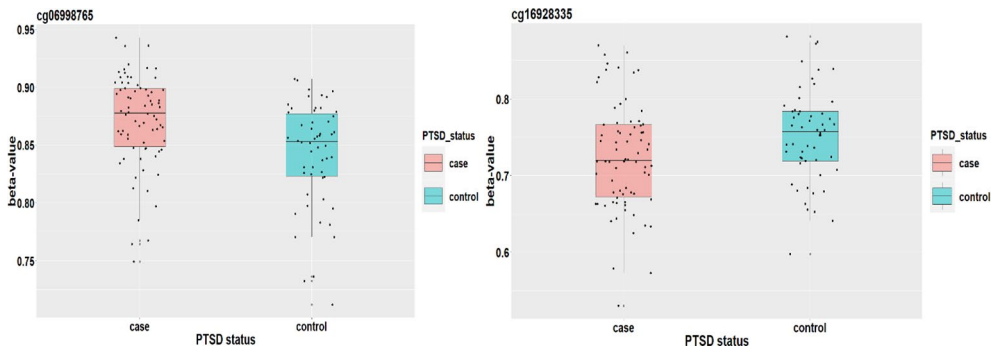


FIGURE 8 | The DNAm status (β value) of cg06998765 and cg16928335 between PTSD cases and controls.

(2016), our method is built under a high-dimensional inference framework where we can simultaneously estimate and test the effect of regression coefficients in a regression model. The high-dimensional testing method implements a debias approach and the de-sparsified estimates can well take care of correlations between mediators (Zhang and Zhang, 2014). Such correlations are naturally arising due to the nature of the epigenetic data. We illustrate the performance of the proposed method *via* simulations and case studies and compare with the HIMA method (Zhang et al, 2016). The simulation studies show that our method (HDMA) outperforms the HIMA method when there are high correlations between mediators. Thus, HDMA can be safely used in a high-dimensional mediation analysis from population studies.

In the first real data analysis, four CpG sites are identified to mediate the effects between alcohol consumption and EOC status. HDMA identifies two more CpG sites than HIMA does. In the second real data analysis, of the two CpG sites identified by HDMA, one overlaps with HIMA. These CpG sites may mediate the effect of childhood maltreatment to PTSD risk in adulthood. In both real data analysis, HDMA identifies more CpG sites than HIMA does, demonstrating the superior power of HDMA over HIMA. However, further biological verification is needed to validate the results, since statistical significance does not guarantee a biological significance.

Philibert et al. (2012) found that alcohol intake is linked to widespread changes in DNAm in women. Cvetkovic (2003) showed that DNAm alterations are an early step in carcinogenesis and could represent a mechanism of disease. Many such pieces of evidence point to the proper linkage of DNAm mediating the relationship between alcohol consumption and EOC status. Similar evidence also supports the linkage between childhood maltreatment and PTSD mediated by DNAm. Mehta et al. (2013) provided epigenetic support that childhood maltreatment is likely to carve long-lasting epigenetic marks, leading to adverse health outcomes such as PTSD in adulthood. Childhood abuse can increase the risk of neuropsychiatric and cardiometabolic disease via changes in epigenetic marks

(Szyf, 2012; Yang et al., 2013). These studies support the mediation role of DNAm between childhood maltreatment and the risk of developing PTSD in adulthood.

The mediation effect in this study is based on a linear effect assumption, while effects such as interactions including magnitude epistasis and sign epistasis are not considered. Such kinds of complex interactive mechanisms can complicate the model, especially under a high-dimensional setup. For example, if there are antagonistic epistatic interactions among mediators, the mediation effects between exposure and the outcome can be weakened, leading to the failure to detect the mediation effects. If there are synergistic epistatic interactions among mediators, the existence of mediators can produce a synergistic effect to enhance their mediation effect. In the event of multiple exposures, models can be even more complicated. Under these situations, it is not clear on how to model and assess the mediation effect in a high-dimensional setup. These issues imply the simplicity of the current method and also raise modeling challenges for further methodological development. We will take these into consideration in our future studies. The R code that implements the method can be found in github with weblink: https://github.com/YuzhaoGao/High-dimensional-mediation-analysis-R/blob/master/HDMA.R.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, to any qualified researcher. Requests to access the datasets should be directed to Yuehua Cui cuiy@msu.edu.

AUTHOR CONTRIBUTIONS

YG implemented the method and drafted the manuscript. HY and RF were involved in the data analysis. YZ and EG participated in the study. YC conceived the idea, designed the study, and drafted the manuscript. All authors read and approved the final manuscript.

REFERENCES

Abdolmaleky, H. M., Smith, C. L., Faraone, S. V, Shafa, R., Stone, W., Glatt, S. J., et al. (2004). Methylomics in psychiatry: modulation of gene-environment interactions may be through DNA methylation. *Am. J. Med. Genet. Part B (Neuropsychiatric Genetics)* 127B, 51–59. doi: 10.1002/ajmg.b.20142

Adalsteinsson, B. T., Gudnason, H., Aspelund, T., Harris, T. B., Launer, L. J., Eiriksdottir, G. et al. (2012). Heterogeneity in white blood cells has potential to confound DNA methylation measurements. *PLoS One*, 7 (10), 1–9. doi: 10.1371/journal.pone.0046705

Adalsteinsson, B. T., Gudnason, H., and Aspelund, T. et al. (2012). Heterogeneity in white blood cells has potential to confound DNA methylation measurements. *PLoS One* 7 (10), 1–9. doi: 10.1371/journal.pone.0046705

Auner, V., Kriegshäuser, G., Tong, D., Horvat, R., Reinthaller, A., Mustea, A. et al. (2009). KRAS mutation analysis in ovarian samples using a high sensitivity biochip assay. *BMC Cancer* 9, 1–8. doi: 10.1186/1471-2407-9-111

Baron, R. M., and Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical Considerations. *J. Pers. Soc. Psychol* 51 (6), 1173–1182

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., and Erbaugh, J. (1961). Inventory for measuring depression. *Arch. Gen. Psychiatry* 4 (6), 561–571. doi: 10.1001/archpsyc.1961.01710120031004

Berger, R. L., and Hsu, J. C. (1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Stat. Sci.* 11 (4), 283–319. doi: 10.1214/ss/1032280304

Cvetkovic, D. (2003). Early events in ovarian oncogenesis. *Reprod. Biol. Endocrin.* 1, 68. doi: 10.1186/1477-7827-1-68

Dezeure, R., Bühlmann, P., Meier, L., and Meinshausen, N. (2015). High-dimensional inference: confidence intervals, p-values and R-software hdi. *Stat. Sci.* 30 (4), 533–558. doi: 10.1214/15-STS527

Dongen, J. Van, Nivard, M. G., Willemsen, G., Hottenga, J., Helmer, Q., Dolan, C. V. et al. (2016). Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat. Commun.* 7, 1–13. doi: 10.1038/ncomms11115

E. Shrout, P., and Bolger, N. (2002). Mediation in experimental and nonexperimental studies: new procedures and recommendations. *Psychol. Meth.* 7 (4), 422–445. doi: 10.1037//1082-989X.7.4.422

Elgaaen, B. V., Haug, K. B. F., Wang, J., Olstad, O. K., Fortunati, D., Onsrud, M. et al. (2010). POLD2 and KSP37 (FGFBP2) correlate strongly with histology, stage and outcome in ovarian carcinomas. *PLoS One*, 5 (11), e13837. doi: 10.1371/journal.pone.0013837

Fan, J., and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B: Stat. Methodol.* 70 (5), 849–911. doi: 10.1111/j.1467-9868.2008.00674.x

Foa, E. B., and Tolin, D. F. (2000). Comparison of the PTSD symptom scale-interview version and the clinician-administered PTSD scale. *J. Trauma. Stress* 13 (2), 181–191.

Guida, F., Sandanger, T. M., Castagné, R., Campanella, G., Polidoro, S., Palli, D., et al. (2015). Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. *Hum. Mol. Genet.* 24 (8), 2349–2359. doi: 10.1093/hmg/ddu751

Hafeman, D. M., and Schwartz, S. (2009). Opening the black box : a motivation for the assessment of mediation. *Int. J. Epidemiol.* (38), 838–845. doi: 10.1093/ije/dyn372

Huang, Y.-T., J.VanderWeele, T., and Lin, X. (2014). Joint analysis of SNP and gene expression data in genetic association studies of comples disease. *Ann. Appl. Stat.* 8 (1), 352–376. doi: 10.1214/13-AOAS690

Huang, Y., Liang, L., Moffatt, M. F., Cookson, W. O. C., and Lin, X. (2015). iGWAS:integrative genome-wide association studies of genetic and genomic data for disease susceptibility using mediation analysis. *Genet. Epidemiol.* 39 (5), 347–356. doi: 10.1002/gepi.21905

Huang, Y., and Pan, W. (2016). Hypothesis test of mediation effect in causal mediation model with high-dimensional continuous mediators. *Biometrics*, (72), 402–413. doi: 10.1111/biom.12421

Houseman, E. A., Accomando, W. P., Koestler, D. C., Christensen, B. C., Marsit, C. J., Nelson, H. H. et al. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 13, 86. doi: 10.1186/1471-2105-13-86

Imai, K., Keele, L., and Tingley, D. (2010). A general approach to causal mediation analysis. *Psychol. Meth.* 15 (4), 309–334. doi: 10.1037/a0020761

Klengel, T., Pape, J., Binder, E. B., and Mehta, D. (2014). The role of DNA methylation in stress-related psychiatric disorders. *Neuropharmacology* 80, 115–132. doi: 10.1016/j.neuropharm.2014.01.013

Klengel, T., Mehta, D., Anacker, C., Rex-haffner, M., Pruessner, J. C., Pariante, C. M. et al. (2016). Allele-specific FKBP5 DNA demethylation mediates gene-childhood trauma interactions. *Cornell Law Rev.* 101 (6), 1533–1595. doi: 10.1038/nn.3275

Koestler, D. C., Chalise, P., Cicek, M. S., Cunningham, J. M., Armasu, S., Larson, M. C. et al. (2014). Integrative genomic analysis identifies epigenetic marks that mediate genetic risk for epithelial ovarian cancer. *BMC Medical Genomics* 7 (8), 1–14. doi: 10.1186/1755-8794-7-8

Li, S., D.Hursting, S., J.Davis, B., McLachlan, J. A., and Barrett, J. car. (2003). Environmental exposure, DNA methylation, and gene regulation: lessons from diethylstilbesterol-induced cancers. *N. Y. Acad. Sci.* 983 (1), 161–169.

Liu, Y., Aryee, M. J., Padyukov, L., Fallin, M. D., Hesselberg, E., Runarsson, A. et al. (2013). Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat. Biotechnol.* 31 (2), 142–147. doi: 10.1038/nbt.2487

Lutz, P. E., and Turecki, G. (2014). DNA methylation and childhood maltreatment: From animal models to human studies. *Neuroscience* 264, 142–156. doi: 10.1016/j.neuroscience.2013.07.069

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, London:Taylor & Francis Group.

Mehta, D., Klengel, T., Conneely, K. N., Smith, A. K., Altmann, A., Pace, T. W. et al. (2013). Childhood maltreatment is associated with distinct genomic and epigenetic profiles in posttraumatic stress disorder. *Proc. Nat. Acad. Sci.* 110 (20), 8302–8307. doi: 10.1073/pnas.1217750110

Nakayama, N., Nakayama, K., Yeasmin, S., Ishibashi, M., Katagiri, A., Iida, K. et al. (2008). KRAS or BRAF mutation status is a useful predictor of sensitivity to MEK inhibition in ovarian cancer. *Br. J. Cancer* 99 (12), 2020–2028. doi: 10.1038/sj.bjc.6604783

Pearl, J. (2012). The causal mediation formula – a guide to the assessment of pathways and mechanisms. *Soc. Prev. Res,* (13), 426–436. doi: 10.1007/s11121-011-0270-1

Pfeffer, J., and Devoe, S. E. (2009). Economic evaluation: the effect of money and economics on attitudes about volunteering. *J. Econ. Psychol.* (30), 500–508. doi: 10.1016/j.joep.2008.08.006

Philibert, R. A., Plume, J. M., Gibbons, F. X., Brody, G. H., and Beach, S. R. H. (2012). The impact of recent alcohol use on genome wide DNA methylation signatures. *Front. Genet.* 3 (APR), 1–8. doi: 10.3389/fgene.2012.00054

Pierce, B. L., Tong, L., Chen, L. S., Rahaman, R., Argos, M., Jasmine, F. et al. (2014). Mediation analysis demonstrates that trans-eQTLs are often explained by cis-mediation: a genome-wide analysis among 1800 South Asians. *PLoS Genetics* 10 (12), e1004818. doi: 10.1371/journal.pgen.1004818

Preacher, K., and Hayes, A. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav. Res, Meth.* 40 (3), 879–891. doi: 10.3758/BRM.40.3.879

Robertson, K. D. (2005). DNA methylation and human disease. *Nat. Rev. Genet.* 6 (8), 597–610. doi: 10.1038/nrg1655

Rocca, C. H., Doherty, I., Padian, N. S., Hubbard, A. E., and Minnis, A. M. (2010). Pregnancy intentions and teenage pregnancy among latinas : A Mediation Analysis. *Perspect. Sex. Reprod. Health* 42 (3), 186–196. doi: 10.1363/4218610

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociol. Methodol.* 13, 290–312. doi: 10.2307/270723

Sun, T., and Zhang, C. (2012). Scaled sparse linear regression. *Biometrika* 99 (4), 879–898. doi: 10.1093/biomet/ass043

Szyf, M. (2012). The early-life social environment and DNA methylation. *Clin. Genet.* (81), 341–349. doi: 10.1111/j.1399-0004.2012.01843.x

Tibshiranit, R. (1996). Regression shrinkage and selection *via* the lasso. *J. R. Stat. Soc. Ser. B: Stat. Methodol.* (1), 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x

Titus, A. J., Gallimore, R. M., Salas, L. A., and Christensen, B. C. (2017). Cell-type deconvolution from DNA methylation: a review of recent applications. *Hum. Mol. Genet.* 26 (R2), R216-R224. doi: 10.1093/hmg/ddx275

Uddin, M., Galea, S., Chang, S. C., Aiello, A. E., Wildman, D. E., De Los Santos, R. et al. (2011). Gene expression and methylation signatures of MAN2C1 are associated with PTSD. *Dis. Markers* 30 (2–3), 111–121. doi: 10.3233/DMA-2011-0750

Van De Geer, S., Bühlmann, P., Ritov, Y., and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Stat.* 42 (3), 1166–1202. doi: 10.1214/14-AOS1221

Vanderweele, T. J., and Vansteelandt, S. (2010). Odds ratios for mediation analysis for a dichotomous outcome. *Am. J. Epidemiol.* 172 (12), 1339–1348. doi: 10.1093/aje/kwq332

Varela-Rey, M., Woodhoo, A., Martinez-Chantar, M.-L., Mato, J., and Lu, S. C. (2013). Alcohol, DNA methylation, and cancer. *Alcohol Research* 35, 25–35. doi: 10.4067/S0370-41062008000700008

Wang, T., and Samworth, R. J. (2018). High dimensional change point estimation via sparse projection. *J. R. Stat. Soc. Ser. B: Stat. Methodol.* 80 (1), 57–83. doi: 10.1111/rssb.12243

Ward, J., Strawbridge, R. J., Bailey, M. E. S., Graham, N., Ferguson, A., Lyall, D. M. et al. (2017). Genome-wide analysis in UK Biobank identifies four loci associated with mood instability and genetic correlation with major depressive disorder, anxiety disorder and schizophrenia. *Trans. Psychiatry* 7, 1264. doi: 10.1038/s41398-017-0012-7

Wingo, A. P., Velasco, E. R., Florido, A., Lori, A., Choi, D. C., Jovanovic, T. et al. (2018). Expression of the PPM1F gene is regulated by stress and associated with anxiety and depression. *Biol. Psychiatry* 83 (3), 284–295. doi: 10.1016/j.biopsych.2017.08.013

Wu, D., Yang, H., Winham, S.J., Natanzon, Y., Koestler, D.C., Luo, T. et al. (2018). Mediation analysis of alcohol consumption, DNA methylation, and epithelial ovarian cancer. *J. Hum. Genet*. 63, 339–348. doi:10.1038/s10038-017-0385-8

Yang, B., Zhang, H., Ge, W., Weder, N., Douglas-palumberi, H., Perepletchikova, F. et al. (2013). Child abuse and epigenetic mechanisms of disease risk. *Am. J. Prev. Med.* 44 (2), 101–107. doi: 10.1016/j.amepre.2012.10.012

Zhang, X., and Cheng, G. (2017). Simultaneous Inference for High-Dimensional Linear Models. *J. Am. Stat. Assoc.* 112 (518), 757–768. doi: 10.1080/01621459.2016.1166114

Zhang, Z., and Wang, L. (2013). Methods for mediation analysis with missing data. *Psychom. Soc.* 78 (1), 154–184. doi: 10.1007/s11336-012-9301-5

Zhang, C., and Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B: Stat. Methodol.* 76 (1), 217–242. doi: 10.1111/rssb.12026

Zhang, H., Zheng, Y., Zhang, Z., Gao, T., Joyce, B., Yoon, G., et al. (2016). Estimating and testing high-dimensional mediation effects in epigenetic studies. *Bioinformatics* 32 (20), 3150–3154. doi: 10.1093/bioinformatics/btw351

Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B: Stat. Methodol.* 67 (2), 301–320. doi: 10.1111/j.1467-9868.2005.00503.x

Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.* 101 (476), 1418–1429. doi: 10.1198/016214506000000735

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Gao, Yang, Fang, Zhang, Goode and Cui. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 22 November 2019
doi: 10.3389/fgene.2019.01042

Circulating Serum MicroRNAs as Potential Diagnostic Biomarkers of Posttraumatic Stress Disorder: A Pilot Study

Clara Snijders[1], Julian Krauskopf[2], Ehsan Pishva[1,3], Lars Eijssen[1,4], Barbie Machiels[1], Jos Kleinjans[2], Gunter Kenis[1], Daniel van den Hove[1], Myeong Ok Kim[5], Marco P. M. Boks[6], Christiaan H. Vinkers[7,8], Eric Vermetten[9,10,11,12], Elbert Geuze[6,11], Bart P. F. Rutten[1†] and Laurence de Nijs[1†]*

[1] Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastricht University, Maastricht, Netherlands, [2] Department of Toxicogenomics, Maastricht University, Maastricht, Netherlands, [3] College of Medicine and Health, University of Exeter Medical School, Exeter, United Kingdom, [4] Department of Bioinformatics (BiGCaT), NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University, Maastricht, Netherlands, [5] Division of Applied Life Science (BK 21), College of Natural Sciences, Gyeongsang National University, Jinju, South Korea, [6] UMC Utrecht Brain Center, Department of Psychiatry, Utrecht, Netherlands, [7] Amsterdam UMC (location VUmc), Department of Anatomy and Neurosciences, Amsterdam, Netherlands, [8] Amsterdam UMC (location VUmc), Department of Psychiatry, Amsterdam, Netherlands, [9] Arq, Psychotrauma Research Expert Group, Diemen, Netherlands, [10] Department of Psychiatry, Leiden University Medical Center, Leiden, Netherlands, [11] Military Mental Healthcare, Netherlands Ministry of Defense, Utrecht, Netherlands, [12] Department of Psychiatry, New York University School of Medicine, New York, United States

OPEN ACCESS

Edited by:
Yun Liu,
Fudan University, China

Reviewed by:
Alice Hudder,
Lake Erie College of Osteopathic Medicine, United States
Fouad Janat,
Independent researcher, Waverly, RI, United States

***Correspondence:**
Laurence de Nijs
Laurence.denijs@maastrichtuniversity.nl

†*These authors share senior authorship*

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *29 July 2019*
Accepted: *30 September 2019*
Published: *22 November 2019*

Citation:
Snijders C, Krauskopf J, Pishva E, Eijssen L, Machiels B, Kleinjans J, Kenis G, van den Hove D, Kim MO, Boks MPM, Vinkers CH, Vermetten E, Geuze E, Rutten BPF and de Nijs L (2019) Circulating Serum MicroRNAs as Potential Diagnostic Biomarkers of Posttraumatic Stress Disorder: A Pilot Study. Front. Genet. 10:1042. doi: 10.3389/fgene.2019.01042

Posttraumatic stress disorder (PTSD) is a psychiatric disorder that can develop upon exposure to a traumatic event. While most people are able to recover promptly, others are at increased risk of developing PTSD. However, the exact underlying biological mechanisms of differential susceptibility are unknown. Identifying biomarkers of PTSD could assist in its diagnosis and facilitate treatment planning. Here, we identified serum microRNAs (miRNAs) of subjects that underwent a traumatic event and aimed to assess their potential to serve as diagnostic biomarkers of PTSD. Next-generation sequencing was performed to examine circulating miRNA profiles of 24 members belonging to the Dutch military cohort Prospective Research in Stress-Related Military Operations (PRISMO). Three groups were selected: "susceptible" subjects who developed PTSD after combat exposure, "resilient" subjects without PTSD, and nonexposed control subjects ($N = 8$ per group). Differential expression analysis revealed 22 differentially expressed miRNAs in PTSD subjects compared to controls and 1 in PTSD subjects compared to resilient individuals (after multiple testing correction and a log2 fold-change cutoff of $\geq|1|$). Weighted Gene Coexpression Network Analysis (WGCNA) identified a module of coexpressed miRNAs which could distinguish between the three groups. In addition, receiver operating characteristic curve analyses suggest that the miRNAs with the highest module memberships could have a strong diagnostic accuracy as reflected by high areas under the curves. Overall, the results of our pilot study suggest that serum miRNAs could potentially serve as diagnostic biomarkers of PTSD, both individually or grouped within a cluster of coexpressed miRNAs. Larger studies are now needed to validate and build upon these preliminary findings.

Keywords: posttraumatic stress disorder, circulating miRNAs, diagnostic biomarker, trauma, susceptibility

BACKGROUND

Posttraumatic stress disorder (PTSD) is a psychiatric disorder that can develop upon exposure to a life-threatening traumatic event, i.e., an event capable of producing intense feelings of fear, helplessness, and horror (Association, 2013). Symptoms associated with PTSD include re-experiencing of the traumatic event, avoidance behavior, overall negative mood, and hyperarousal (Association, 2013). The economic burden associated with PTSD is substantial, and patients with PTSD are at increased risk of committing suicide and having familial issues such as marital problems (Fontana and Rosenheck, 1994; Tarrier and Gregg, 2004; Nock and Kessler, 2006; Ferry et al., 2015). Although ~60% of individuals within Western Europe will one day be exposed to a traumatic event, only ~6% of these individuals develop PTSD while others show a positive psychological adaptation process denoted as resilience (Kalisch et al., 2017; Koenen et al., 2017). However, some populations such as military soldiers are at elevated risk for trauma exposure, making PTSD a relatively common chronic disorder in the combat Veteran population (Thomas et al., 2017). Currently, a variety of treatment options exist for PTSD, without one being clearly superior to another (Yehuda et al., 2014). Moreover, pharmacological treatment options for PTSD are at best moderately effective and only work for a subset of patients (Richter-Levin et al., 2018). Therefore, increasing efforts are being made to unravel the biological underpinnings of PTSD in order to develop more efficient therapeutic strategies. It is now becoming clear that epigenetic mechanisms are involved in the lasting behavioral and molecular effects of trauma exposure (Schmidt et al., 2011; Snijders et al., 2018a).

Epigenetics refers to a variety of processes that are triggered by environmental factors and cause lasting but reversible alterations in gene expression (Goldberg et al., 2007). Among epigenetic mechanisms, noncoding RNA molecules such as microRNAs (miRNAs) are involved in the posttranscriptional regulation of gene expression by binding to specific messenger RNAs (Peschansky and Wahlestedt, 2014). Several miRNAs have been found implicated in PTSD, shedding much needed light on the underlying pathophysiological underpinnings of this disorder (Wingo et al., 2015; Bam et al., 2016a, Bam et al., 2016b; Martin et al., 2017). Such findings emphasize the notion that expression profiles of miRNAs could one day serve as relatively easily accessible biomarkers or be embedded within a network of several relevant biological processes that together could more accurately reflect the complexity of PTSD. For those individuals who have difficulties recognizing or properly describing their symptoms, identifying such markers could be of use in clinical contexts in order to objectively confirm the presence of the disorder and establish appropriate treatment plans when needed (Lehrner and Yehuda, 2014). Using these markers could be equally relevant during postdeployment medical screenings since military service members may have secondary reasons to not fully disclose their symptoms (Yehuda et al., 2013).

Here, we aimed to identify serum miRNAs that could one day serve as diagnostic biomarkers of PTSD. We further aimed to gain insights in the coexpression patterns of these miRNAs, their predicted gene targets and underlying biological pathways, along with their diagnostic accuracy. We hypothesized that specific miRNAs are differentially expressed between subjects with PTSD, trauma-exposed healthy individuals (referred to as "resilient" subjects in this paper), and nonexposed healthy controls. For this, we performed next-generation sequencing (NGS) on serum samples of 24 military members belonging to a Dutch military cohort, and we compared miRNA profiles between the three groups. Our findings suggest that miRNAs could potentially serve as biomarkers of PTSD, both individually or grouped within a cluster of coexpressed miRNAs. Larger studies are now needed in order to further validate and build upon these preliminary findings.

MATERIALS AND METHODS

Participants

A subset of military personnel (24 males) was selected from the larger Prospective Research in Stress-Related Military Operations (PRISMO) study, a prospective cohort of Dutch military members deployed to Afghanistan for 4 months (Reijnen et al., 2015; Eekhout et al., 2016). Based on the level of combat exposure during deployment and the severity of postdeployment PTSD symptoms, three subgroups were identified: 1) susceptible individuals, i.e., trauma-exposed subjects with deployment-related PTSD symptoms at 6 months follow-up; 2) resilient individuals, i.e., trauma-exposed soldiers with no PTSD diagnosis at follow-up; and 3) controls, i.e., deployed, but nonexposed and mentally healthy military members. Blood samples were collected at the Utrecht University Medical Center at 6 months postdeployment. Trauma exposure was assessed using a 19-item deployment experiences checklist (van Zuiden et al., 2011). The severity of PTSD symptoms was established using the 22-item Self-Rating Inventory for PTSD (SRIP) (Hovens et al., 2002). Information on smoking and alcohol was collected using self-report measures. This study was approved by the ethical committee of University Medical Center Utrecht (01-333/0) and conducted in accordance with the Declaration of Helsinki. All participants gave written informed consent.

RNA Isolation

Total RNA was isolated from 300µl human serum using the *mir*Vana PARIS kit (Ambion) according to the manufacturer's instructions. Briefly, the samples were incubated with an equal volume of denaturing solution, acid-phenol/chloroform was added, and the samples spun for 5min at 10,000×*g*. The aqueous phase was recovered and passed through a filter which was washed three times with the provided wash solutions. Final RNA was eluted in 100µl nuclease-free water. The concentrations and quality of the recovered RNA were measured using the Agilent Bioanalyzer 2100 (Agilent Technologies, Inc., CA, USA). All eluates were stored at −80°C until further use.

Small RNA Library Preparation and Next-Generation Sequencing

Barcoded libraries (N = 24, 8 per group) were prepared with an input of 25ng total RNA using the Illumina Small RNA TruSeq kit (Illumina, CA, USA). Briefly, 3′ and 5′ RNA adapters were added, the samples were reverse transcribed and amplified using 11 PCR cycles. All samples were processed in parallel and received a unique barcode. The complementary DNA constructs were gel purified and concentrated by ethanol precipitation. The quality control was performed using Agilent's 2100 Bioanalyzer with a High-Sensitivity DNA Chip. The 24 samples were pooled (N = 8 per group) and sequenced in duplicate using the Illumina HiSeq 2000 DNA sequence platform according to the manufacturer's protocol (GEO accession: GSE137624).

Small RNA Sequencing Data Analysis

Quality control of the raw sequences was done using FastQC (v. 0.11.3), and reads were preprocessed and mapped to the latest release of miRBase (v. 21) (Baras et al., 2015) utilizing miRge with default settings (Kozomara and Griffiths-Jones, 2014). In order to compensate for bias introduced by very low abundant sequences, only those miRNAs with an average of 50 counts (or more) across samples were considered for further analyses.

Differential Expression Analysis

Data normalization and differential expression analysis was conducted with the DESeq2 package in R (v. 3.5.2) (Love et al., 2014) thereby correcting for age, alcohol use, and smoking status. Resulting p-values were controlled by the false discovery rate (FDR) at 5% (Benjamini and Hochberg, 1995).

Weighted Gene Coexpression Network Construction and Module Detection

The identified miRNAs were used to construct coexpression networks using the Weighted Gene Coexpression Network Analysis (WGCNA) R package (Langfelder and Horvath, 2008). Normalized miRNA data was used as input. An adjacency matrix was generated by calculating Pearson's correlations between all miRNAs. Next, topological overlap between miRNAs was calculated using a power of 9. We performed 200 rounds of bootstrapping in order to construct a network that is robust to outliers. The cutreeDynamic function in the dynamicTreeCut R package was then used to identify coexpression modules of positively correlated miRNAs with high topological overlap. Modules with at least 30 miRNAs were assigned a color. Modules with highly correlated eigengenes were merged using the mergeCloseModules function in R. Pearson correlations between module eigengenes, age, smoking status, and alcohol were calculated. Welch's t-tests were performed in order to detect differences between module eigengenes of the control subjects and the trauma-exposed individuals. One-way ANOVAs were performed to detect differences between the three groups. When significant, the post-hoc Tukey HSD test was used to detect pairwise group differences.

Target Gene Pathway and Enrichment Analyses

The experimentally validated miRNA–target interactions database miRTarBase 6.0 (Chou et al., 2018) was used to identify gene targets of miRNAs. In order to narrow down the amount of target genes for further analyses, one-sided Fisher tests (with FDR multiple correction) were performed to evaluate whether the amount of miRNAs targeting a specific gene was significantly higher than expected by chance. Those genes were then analyzed for enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology terms (GO terms) using the online Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (Huang da et al., 2009; Huang da et al., 2009).

Statistical Analyses

To detect differences in age, number of previous deployments, cigarette smoking, alcohol use, trauma exposure scores, and SRIP scores between the groups, the Welch ANOVA with Games–Howell post-hoc test was applied. Since data on alcohol use at the 6 months follow-up time point was not available for all subjects, predeployment values were used instead. For each individual, smoking status was estimated based on their unique methylation patterns in 183CpGs, as previously described (Zeilinger et al., 2013). Finally, the classification accuracy of specific miRNAs was determined by calculating the area under the receiver operating characteristic (ROC) curve (AUC) in R.

RESULTS

Demographic Characteristics

A total of 24 subjects were included in the present study, of which 8 developed PTSD following deployment, 8 were resilient, and 8 were nonexposed controls (**Supplementary Table 1**). Based on the sequencing results, four subjects were excluded due to having a distinctively lower amount of reads causing great variation in expression data between samples. The three groups did not differ in terms of age, number of previous deployments, smoking status, and alcohol use (**Table 1**). On average, subjects with PTSD and resilient individuals were exposed to a similar amount of traumatic events, which was significantly more than the nonexposed controls [$F(2, 8.8) = 54.67$, $p < 0.001$. Games–Howell post-hoc showed $p < 0.001$ for PTSD versus control, and resilient versus control]. Finally, resilient and control subjects had similar postdeployment PTSD scores as measured by the SRIP, which were significantly lower than the average score of the PTSD group [$F(2, 11.15) = 25.23$, $p < 0.001$. Games–Howell post-hoc showed $p < 0.001$ for PTSD versus resilient, and PTSD versus control].

miRNAs Sequencing and Differential Expression Analysis

Small RNA sequencing yielded an average of 9.5 million unfiltered sequencing reads across all samples. After adaptor trimming and

TABLE 1 | Demographic characteristics of the 20 subjects remaining after outlier exclusion.

	Susceptible (*N* = 8)	Resilient (*N* = 6)	Control (*N* = 6)	*p*-value
Age when deployed	22.13 (0.61)	34.17 (4.88)	27.50 (3.62)	0.083
Number of previous deployments	0.29 (0.18)	0.83 (0.48)	0.17 (0.17)	0.465
Cigarette smoking	2.79 (1.65)	–1.43 (1.30)	–0.15 (1.89)	0.194
Alcohol use	2.86 (0.67)	1.17 (0.40)	1.83 (0.60)	0.155
Trauma exposure score	7.75 (0.98)	7.17 (0.75)	0.5 (0.22)	<0.001
SRIP PTSD score	55.25 (4.01)	25.50 (1.63)	24.50 (1.46)	<0.001

Data are presented as mean (SE). SRIP, self-rating inventory for posttraumatic stress disorder.

size selection, 1.9 million high-quality reads remained, which were aligned to miRNA sequences from miRBase (release 21). As mentioned earlier, principal component analysis revealed the presence of four outliers, which were excluded from further analysis. The count data were then filtered for miRNAs that showed an average of 50 reads or more across all samples. This resulted in the identification of 306 different miRNAs. Differential expression analysis in DESeq2 revealed that a total of 123 miRNAs showed differential expression between PTSD cases and nonexposed controls, while 4 were downregulated in PTSD cases compared to resilient individuals (**Supplementary Table 2**). Selecting those miRNAs with a log2 fold-change (FC) value≥|1.0| and FDR adjusted $p < 0.05$ revealed that one miRNA, miR-1246, was downregulated in PTSD subjects compared to resilient subjects and 22 were differentially expressed between PTSD subjects and nonexposed controls (**Table 2**, **Figure 1**). Of these, 4 were downregulated and 18 were upregulated. We used the Venn tool to identify those differentially expressed miRNAs that are specific for PTSD only (**Figure 2**). Two miRNAs were identified at the intersection of the blue and yellow circles, i.e., miR-4454 and miR-210-3p. Both miRNAs were significantly downregulated in PTSD subjects compared to resilient subjects and controls and not differentially expressed between resilient subjects and controls, suggesting that these could be more specific to PTSD (**Supplementary Table 2**). However, both miRNAs had log2 FC values of −0.61 and −0.54, which does not pass our threshold of ≥|1.0|.

Weighted Gene Coexpression Network Analysis

WGCNA was applied using the 306 identified miRNAs in order to detect clusters of coexpressed miRNAs. Based on the sample dendrogram, one outlier was removed from further analyses (**Supplementary Figure 1**). We identified three miRNA modules (**Figure 3**). The turquoise, blue, and brown modules each had

TABLE 2 | Differentially expressed microRNAs (miRNAs) between posttraumatic stress disorder (PTSD) cases versus controls and PTSD cases versus resilient individuals with a log2 fold-change value≥|1.0| and FDR adjusted $p < 0.05$.

Nr	PTSD vs. control				PTSD vs. resilient			
	miRNA	*Log2 FC*	p-*value*	*FDR adj* p-*value*	*miRNA*	*Log2 FC*	p-*value*	*FDR adj* p-*value*
1	hsa-miR-218-2-3p	3.22	1.65E–02	4.20E–02	hsa-miR-1246	–1.06	3.54E–04	3.05E–02
2	hsa-miR-3609	3.05	8.22E–06	6.62E–05				
3	hsa-miR-432-5p	2.37	8.44E–04	3.31E–03				
4	hsa-miR-138-5p	2.29	2.18E–16	9.54E–15				
5	hsa-miR-221-5p	2.06	6.11E–13	1.34E–11				
6	hsa-miR-4485-3p	1.98	3.45E–15	9.59E–14				
7	hsa-miR-31-5p	1.92	2.16E–15	7.35E–14				
8	hsa-miR-146b-5p	1.67	1.86E–25	2.85E–23				
9	hsa-miR-5096	1.62	1.81E–06	1.84E–05				
10	hsa-miR-222-3p	1.56	2.81E–15	8.61E–14				
11	hsa-miR-1273g-3p	1.55	7.23E–05	3.75E–04				
12	hsa-miR-302a-5p	1.49	6.61E–09	1.06E–07				
13	hsa-miR-221-3p	1.45	1.03E–14	2.62E–13				
14	hsa-miR-619-5p	1.40	1.74E–04	8.31E–04				
15	hsa-miR-335-5p	1.28	7.18E–28	2.20E–25				
16	hsa-miR-146b-3p	1.25	6.30E–11	1.13E–09				
17	hsa-miR-3175	1.17	1.95E–04	9.20E–04				
18	hsa-miR-3656	1.00	1.07E–02	2.86E–02				
19	hsa-miR-184	–1.20	1.58E–05	1.01E–04				
20	hsa-let-7d-5p	–1.27	1.22E–18	9.33E–17				
21	hsa-miR-98-5p	–1.33	1.57E–22	1.60E–20				
22	hsa-miR-146a-5p	–2.04	1.52E–08	2.22E–07				

The table is organized based on decreasing log2 fold-change values. miRNA, microRNA; log2 FC, log2 fold-change.

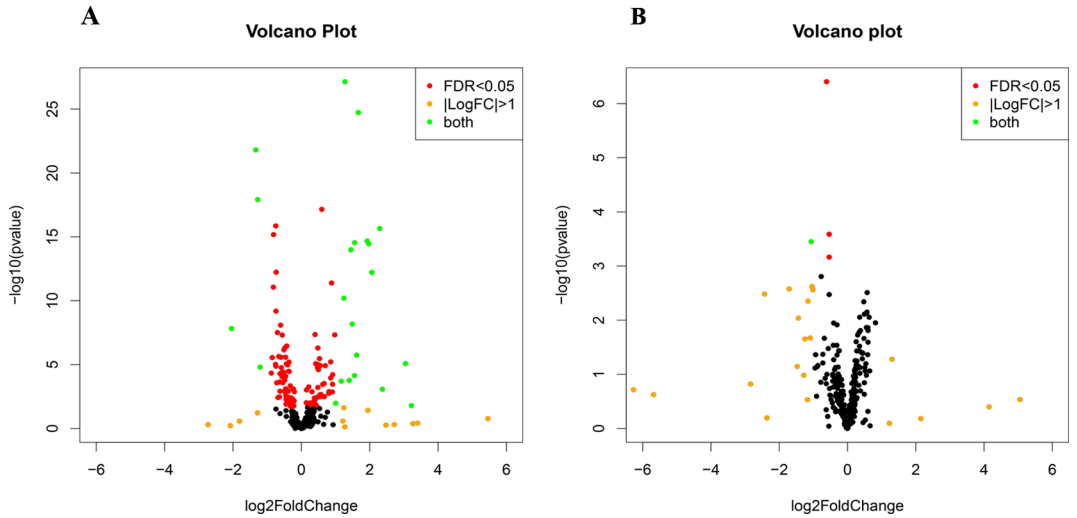


FIGURE 1 | Volcano plots of differentially expressed microRNAs (miRNAs) between posttraumatic stress disorder (PTSD) cases and controls **(A)** and PTSD cases and resilient subjects **(B)**. Black dots represent nonsignificantly differentially expressed miRNAs, red dots represent significant miRNAs with a log2 FC<|1|, orange dots represent nonsignificant miRNAs with a log2 FC≥|1|, and green dots represent significantly differentially expressed miRNAs with a log2 FC≥|1|. Significance is declared when adjusted p<0.05.

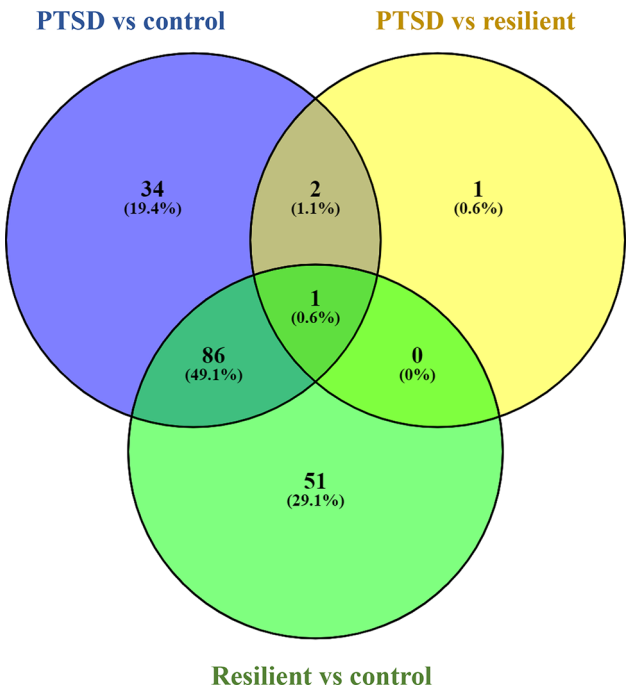


FIGURE 2 | Venn diagram showing overlapping microRNAs (miRNAs). Of particular interest here are the two miRNAs at the intersection of the blue and yellow circle (but not the green), i.e., miR-4454 and miR-210-3p. Obtained using https://bioinfogp.cnb.csic.es/tools/venny/index.html

84, 79, and 65 miRNAs, respectively. None of the modules were associated with the potential covariates age, smoking status, or alcohol use (**Supplementary Figure 2**). Within each module, the module eigengenes were significantly different between trauma-exposed individuals and nonexposed controls for the turquoise and blue modules (p = 2.67×10^{-04}, p = 2.51×10^{-06}, respectively) but not for the brown module (p = 0.196; **Figure 3A**). When stratifying the trauma-exposed individuals into PTSD subjects and resilient subjects, the individual eigengenes of the blue module were significantly different between PTSD subjects and resilient individuals (p = 1.46×10^{-03}; **Figure 3B**), which was not the case for the other modules. We therefore focused on the blue module for further analyses.

Out of the 79 miRNAs belonging to this module, 67 were differentially expressed between PTSD subjects and controls (**Table 1**), including miR-138-5p, the hub miRNA (**Table 3**). In order to evaluate the diagnostic accuracy of some of these miRNAs, we performed ROC analysis for those miRNAs with the highest absolute module memberships. The five most contributing miRNAs, i.e., miR-221-3p, miR-335-5p, miR-138-5p, miR-222-3p, and miR-146-5p (**Table 3**), could perfectly distinguish PTSD subjects and controls (AUC of 1 for all miRNAs; **Supplementary Figure 3** A.1 and A.2 for miR-221-3p). These miRNAs could equally well differentiate PTSD subjects from resilient subjects, except for miR-221-3p and miR-222-3p (AUC of 0.95 and 0.98, respectively). When obtaining ROC curves using miRNA expression levels adjusted for confounders (i.e., age, smoking, and alcohol use), all miRNAs could still distinguish PTSD subjects from controls (**Supplementary Figure 3** B.1 and B.2 for miR-221-3p). However, differentiating PTSD from resilience was less accurate as reflected by AUCs of 0.625, 0.775, 0.725, 0.675, and 0.775 for miR-221-3p, miR-335-5p, miR-138-5p, miR-222-3p, and miR-146-5p, respectively (**Supplementary Figure 3** B.1 and B.2 for miR-221-3p).

Target Gene Pathway and GO Enrichment Analyses

Validated gene targets of the 79 miRNAs comprised within the blue module were obtained from the online database miRTarBase

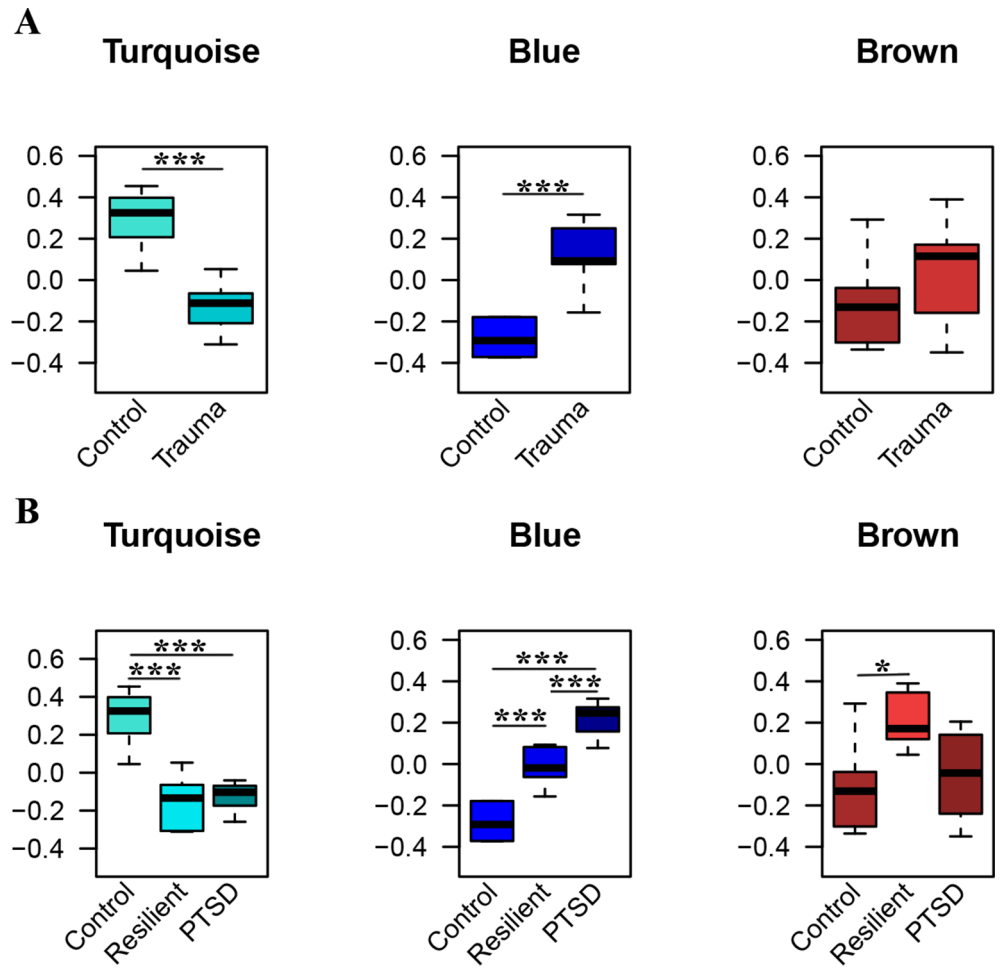


FIGURE 3 | Significant modules of coexpressed miRNAs. The y-axis displays the module eigengene values. Groups are stratified by trauma exposure **(A)**, and trauma exposure and PTSD status **(B)**. Significances were detected using Welch tests **(A)** or one-way ANOVAs **(B)** $*p < 0.05$, $***p < 0.001$.

(Chou et al., 2018). In order to narrow down this extensive set of target genes ($N = 9270$), Fisher tests were performed to select only those genes that were targeted by significantly more miRNAs than expected by chance. This revealed a set of 146 genes, which were considered for pathway and enrichment analyses (**Supplementary Table 3**). After FDR adjustment, 15 significantly enriched KEGG pathways were identified of which most were cancer-related (**Table 4**). GO enrichment analyses of these target genes further identified eight significant biological processes, five molecular functions, and six cellular components (**Table 5**). The most enriched GO terms were related to apoptotic processes, protein binding, and intracellular compartments, respectively (**Table 5**).

DISCUSSION

In this study, we aimed to identify the diagnostic biomarker potential of circulating miRNAs for PTSD using serum samples from Dutch military subjects. We further aimed to gain insights in the coexpression patterns of these miRNAs, their predicted gene targets, and underlying biological pathways. Our preliminary findings suggest that 1) certain miRNAs could potentially serve as individual biomarkers of susceptibility, and 2) the coexpression of a specific set of miRNAs could accurately distinguish between subjects with PTSD, resilient individuals, and nonexposed controls. Such markers could be useful in clinical settings for accurate diagnosis and treatment planning, which is especially relevant for individuals who have that have difficulties associating their symptoms to a traumatic event, are unable to describe their symptoms, or are unwilling to fully disclose them (Yehuda et al., 2013).

Differential expression analysis identified 1 differentially expressed miRNA between subjects with PTSD and resilient individuals and 22 between subjects with PTSD and nonexposed controls (after multiple testing correction and a log2 FC cutoff of $\geq|1|$). Of these, miR-138-5p was significantly overexpressed in subjects with PTSD as compared to controls, and WGCNA revealed that this was the hub miRNA of the blue module. Serum levels of this miRNA were previously found altered in a rat model

TABLE 3 | MicroRNAs (miRNAs) belonging to the Weighted Gene Coexpression Network Analysis (WGCNA) blue module.

Nr	miRNA	Nr	miRNA
1	hsa-miR-221-3p	41	hsa-miR-641
2	hsa-miR-335-5p	42	hsa-miR-208a-3p
3	hsa-miR-138-5p	43	hsa-miR-18a-3p
4	hsa-miR-222-3p	44	hsa-miR-193a-5p
5	hsa-miR-146b-5p	45	hsa-miR-411-5p*
6	hsa-miR-31-5p	46	hsa-miR-148a-5p
7	hsa-miR-340-5p	47	hsa-miR-505-3p
8	hsa-miR-210-3p	48	hsa-miR-214-3p
9	hsa-miR-208b-3p	49	hsa-miR-335-3p
10	hsa-miR-302a-5p	50	hsa-miR-4485-3p
11	hsa-let-7i-5p	51	hsa-miR-10a-5p
12	hsa-miR-4454	52	hsa-miR-212-3p
13	hsa-miR-146b-3p	53	hsa-miR-331-3p
14	hsa-miR-99b-3p	54	hsa-miR-490-3p*
15	hsa-let-7a-3p	55	hsa-miR-20a-5p
16	hsa-miR-221-5p	56	hsa-miR-455-5p
17	hsa-miR-27a-3p/27b-3p	57	hsa-miR-874-5p
18	hsa-miR-127-3p	58	hsa-miR-675-5p
19	hsa-miR-3200-3p	59	hsa-miR-504-5p
20	hsa-miR-128-3p	60	hsa-miR-654-3p*
21	hsa-miR-20b-5p	61	hsa-miR-30a-3p*
22	hsa-miR-199a-3p	62	hsa-miR-425-3p
23	hsa-miR-181c-5p	63	hsa-miR-183-5p
24	hsa-miR-652-3p	64	hsa-miR-532-3p*
25	hsa-miR-4662a-5p	65	hsa-miR-193b-3p*
26	hsa-miR-17-5p/106a-5p	66	hsa-miR-130a-3p
27	hsa-miR-146a-5p	67	hsa-miR-7151-5p*
28	hsa-miR-148a-3p	68	hsa-miR-23a-3p
29	hsa-miR-708-5p	69	hsa-miR-877-3p*
30	hsa-miR-34a-5p	70	hsa-miR-424-3p
31	hsa-miR-574-3p	71	hsa-miR-151a-3p
32	hsa-miR-145-3p	72	hsa-miR-3175
33	hsa-miR-490-5p	73	hsa-miR-361-3p
34	hsa-miR-148b-5p	74	hsa-miR-92b-5p
35	hsa-miR-143-3p	75	hsa-miR-181a-3p
36	hsa-miR-184	76	hsa-miR-328-3p*
37	hsa-miR-199b-3p	77	hsa-miR-423-3p*
38	hsa-miR-628-5p	78	hsa-miR-139-5p*
39	hsa-miR-132-3p	79	hsa-miR-339-3p*
40	hsa-miR-181c-3p		

miRNAs with a star were not differentially expressed in PTSD cases versus controls or versus resilient subjects. The miRNAs are ranked based on their absolute module membership (highest to lowest).

TABLE 4 | Significant Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.

Nr	Term	*p* value	FDR adj *p* value
1	Pathways in cancer	2.94E–10	3.66E–07
2	Pancreatic cancer	6.27E–09	7.79E–06
3	HTLV-I infection	1.26E–08	1.56E–05
4	Small cell lung cancer	8.93E–08	1.11E–04
5	Melanoma	2.14E–07	2.65E–04
6	Hepatitis B	2.23E–07	2.77E–04
7	FoxO signaling pathway	8.06E–07	1.00E–03
8	Colorectal cancer	9.16E–07	1.14E–03
9	Prostate cancer	1.38E–06	1.71E–03
10	MicroRNAs in cancer	2.34E–06	2.91E–03
11	Chronic myeloid leukemia	2.93E–06	3.63E–03
12	Apoptosis	1.14E–05	1.41E–02
13	Central carbon metabolism in cancer	1.41E–05	1.75E–02
14	MAPK signaling pathway	1.49E–05	1.85E–02
15	Glioma	1.56E–05	1.94E–02

These pathways were identified using the online Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8.

of restraint stress (Balakathiresan et al., 2014), while hippocampal miR-138-5p levels were associated with the formation of fear memories in mice (Li et al., 2018). Another miRNA, miR-1246, was the only significant miRNA that was downregulated in PTSD cases compared to resilient subjects and had a log2 FC>|1|. This miRNA was previously found downregulated in peripheral blood mononuclear cells of war veterans suffering from PTSD as compared to healthy nontrauma-exposed controls (Bam et al., 2016b). Such findings suggest that these miRNAs could be implicated in PTSD and potentially aid in diagnosing this disorder.

The three modules of coexpressed miRNAs identified by WGCNA revealed that most of the detected miRNAs could be clustered based on similarities in their expression patterns. The blue module contained 79 miRNAs which could significantly differentiate between trauma-exposed individuals and nonexposed controls. Interestingly, within the trauma-exposed individuals, the expression profiles of these miRNAs were significantly different between individuals with and without PTSD. This highlights the importance of including and studying not only non-trauma exposed controls but also trauma-exposed healthy individuals in order to disentangle PTSD effects from trauma-related effects. Moreover, 67 of the miRNAs of the blue module, including its hub miRNA, were significantly differentially expressed between PTSD cases and controls, which enhances the notion that these miRNAs could be relevant for PTSD.

Of the five miRNAs with the highest module membership, we calculated the AUCs to assess their diagnostic accuracy (Grund and Sabin, 2010). In order to determine the biomarker potential of these miRNAs, i.e., their potential to reflect PTSD regardless of any other confounding condition, we used uncorrected miRNA expression values. Interestingly, the results suggest that these miRNAs could almost perfectly distinguish PTSD subjects from resilient individuals and controls. However, these results were not reflected by the DESeq2 analyses in which the expression levels of these miRNAs were not different between PTSD cases and resilient individuals. Part of this discrepancy can most likely be attributed to confounding effects, as DESeq2 analyses were corrected for age, alcohol, and smoking status. When obtaining the ROC curves using confounder-adjusted miRNA expression values, the AUCs more accurately corresponded to the DESeq2 results. Although these results suggest that the expression of our selected miRNAs fluctuates with confounders, they mostly strengthen the need of replication in larger cohorts. This will further be valuable in determining whether these miRNAs could be specific for PTSD only as opposed to trauma more broadly.

Enrichment of GO terms indicated that target genes of the coexpressed miRNAs in the blue module are enriched in several KEGG pathways of which most were cancer-related. This suggests that these miRNAs could be implicated in cancer pathways that are also involved in signaling cascades possibly related to PTSD.

TABLE 5 | Significant gene ontology (GO) terms enriched for a subset of target genes (*N* = 218) of the coexpressed microRNAs (miRNAs) from the blue module (*N* = 70).

Nr	GO ID	GO term	*p*-value	FDR adj *p*-value
	Biological process			
1	GO:0008630	Intrinsic apoptotic signaling pathway in response to DNA damage	5.11E–09	8.60E–06
2	GO:0071456	Cellular response to hypoxia	1.19E–07	2.00E–04
3	GO:0071260	Cellular response to mechanical stimulus	1.46E–07	2.45E–04
4	GO:0006919	Activation of cysteine-type endopeptidase activity involved in apoptotic process	4.96E–07	8.35E–04
5	GO:0043066	Negative regulation of apoptotic process	1.54E–06	2.59E–03
6	GO:0045944	Positive regulation of transcription from RNA polymerase II promoter	2.35E–06	3.95E–03
7	GO:0030308	Negative regulation of cell growth	8.64E–06	1.45E–02
8	GO:0097192	Extrinsic apoptotic signaling pathway in absence of ligand	9.20E–06	1.55E–02
	Molecular function			
1	GO:0005515	Protein binding	8.15E–13	1.13E–09
2	GO:0044822	Poly(A) RNA binding	3.81E–07	5.29E–04
3	GO:0008134	Transcription factor binding	3.77E–06	5.23E–03
4	GO:0042802	Identical protein binding	1.11E–05	1.54E–02
5	GO:0031625	Ubiquitin protein ligase binding	2.41E–05	3.34E–02
	Cellular component			
1	GO:0005829	Cytosol	6.27E–12	8.32E–09
2	GO:0005654	Nucleoplasm	8.76E–08	1.16E–04
3	GO:0005741	Mitochondrial outer membrane	2.53E–06	3.36E–03
4	GO:0005634	Nucleus	1.73E–05	2.29E–02
5	GO:0016020	Membrane	2.49E–05	3.31E–02
6	GO:0005739	Mitochondrion	3.34E–05	4.43E–02

Identified using the online Database for Annotation, Visualization, and Integrated Discovery (DAVID) v6.8.

The target genes were also involved in several biological processes of which most were involved in apoptotic processes. Previous studies found reduced level of apoptotic markers in the serum of subjects with PTSD (Mkrtchian et al., 2013) and abnormal apoptosis in specific brain regions of animals undergoing single prolonged stress as a model for PTSD (Han et al., 2013; Li et al., 2013; Jia et al., 2018). These findings indicate a potential apoptosis dysfunction that could contribute to the inflammation pattern frequently observed within PTSD (Mkrtchian et al., 2013). Furthermore, the involvement of the identified genes in cellular responses after mechanical stimuli could indicate the need to correct for traumatic brain injuries, which are not uncommon among military members. Unfortunately, this information was not available for the present study. Finally, enriched molecular function GO terms suggest their involvement in the binding of proteins and RNA, while the significant cellular component GO terms show involvement of intracellular compartments such as the cytosol and the nucleoplasm.

Of note, the present paper refers to trauma-exposed healthy individuals as being "resilient" in order to create a clear differentiation between trauma-exposed healthy subjects and nonexposed control subjects. However, we do acknowledge and emphasize that resilience is more than just the reverse side of PTSD or the absence of symptomatology (Kalisch et al., 2017; Snijders et al., 2018b). Instead, resilience is an active and dynamic process that needs to remain separated from the multifaceted and complex nature of PTSD. This complexity further suggests that identifying one true and valid biomarker of susceptibility is likely not realistic. We therefore urge future studies to combine findings such as the ones presented in this paper with several other biological networks and phenotypic profiles in order to develop a cross-dimensional, global understanding of PTSD.

The main strength of this study lies in the inclusion of three different groups, i.e., PTSD subjects, resilient subjects, and nonexposed healthy controls, which allows us to disentangle PTSD- from trauma-related effects. However, the study is mainly limited by its relatively small sample size consisting of male subjects only. Given the existence of female- and male-biased miRNAs as recently reported by Cui, Yang et al. (2018) (Cui et al., 2018), these findings may not be applicable to the female population. This study population may also differ from other cohorts such as civilians in terms of demographics, psychological characteristics, and type of experienced trauma, which limits the extrapolation potential. Next, one could question the validity of self-report PTSD measures and whether the observed markers are specific to PTSD since certain comorbidities such as (history of) traumatic brain injuries were not available and thus not accounted for.

In conclusion, this paper presents preliminary evidence for using specific miRNAs as diagnostic biomarkers of PTSD, either individually or grouped within coexpressed clusters. Identifying reliable biomarkers of PTSD is essential for accurate diagnosis and treatment planning. We therefore encourage future studies to build upon these findings by aiming to replicate these in larger cohorts and thus pave the way for functional studies to gain insights into the precise roles of these miRNAs in stress susceptibility.

DATA AVAILABILITY STATEMENT

The Data is available at GEO, accession: GSE137624.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The ethical committee of University Medical Center Utrecht (01-333/0). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LN, BR, GK, DH, JKl, and MK participated in the conception and the design of the study. MB, CV, EV, and EG recruited the participants and collected the blood samples. CS and BM performed the experiments. CS, JKr, EP, and LE performed the bioinformatics analysis. CS wrote the manuscript. LN, BR, and MK provided the necessary fundings. All authors critically read, commented, provided scientific content and approved the final manuscript.

FUNDING

This work has been funded by the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement N0. 707362 (LN) and by a VIDI award number 91718336 from the Netherlands Scientific Organization (BR).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01042/full#supplementary-material

SUPPLEMENTARY FIGURE S1 | Sample dendrogram to detect outliers. Clustering was based on miRNA expression data. Sample names were relabeled as 1-20.

SUPPLEMENTARY FIGURE S2 | Correlations of modules detected by WGCNA and the following potential covariates: age, smoking status and alcohol use. P-values are presented between brackets. ME: module eigengene.

SUPPLEMENTARY FIGURE S3 | Receiver operating characteristic (ROC) curve for the miRNA with the highest module membership in the blue module, i.e. miR-221-3p. The graphs represent PTSD vs control without confounders (A.1) or with confounders (B.1), and PTSD vs resilient without confounders (A.2) or with confounders (B.2).

REFERENCES

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Arlington, VA: American Psychiatric Association.

Balakathiresan, N. S., Chandran, R., Bhomia M., Jia M., Li H., Maheshwari, R. K., et al. (2014). Serum and amygdala microRNA signatures of posttraumatic stress: fear correlation and biomarker potential. *J. Psychiatric Res.* 57, 65–73. doi: 10.1016/j.jpsychires.2014.05.020

Bam, M., Yang. X., Zumbrun, EE., Zhong, Y., Zhou, J., Ginsberg, J. P., et al. (2016).Dysregulated immune system networks in war veterans with PTSD is an outcome of altered miRNA expression and DNA methylation. *Sci. Rep.* 6, 31209. doi: 10.1038/srep31209

Bam, M., Yang, X., Zhou, J., Ginsberg, J. P., Leyden, Q., Nagarkatti, P. S., et al. (2016). Evidence for epigenetic regulation of pro-inflammatory cytokines, interleukin-12 and interferon gamma, in peripheral blood mononuclear cells from PTSD patients. *J. Neuroimmune Pharmacol.* 11 (1), 168–181. doi: 10.1007/s11481-015-9643-8

Baras, A. S., Mitchell, C. J., Myers, J. R., Gupta, S., Weng, L. C., Ashton, J. M., et al. (2015). miRge - A Multiplexed Method of Processing Small RNA-Seq Data to Determine MicroRNA Entropy. *PLoS One* 10p (11), e0143066. doi: 10.1371/journal.pone.0143066

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Chou, C. H., Shrestha, S., Yang, C. D., Chang, N. W., Lin, Y. L., Liao, K. W., et al. (2018). miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* 46, D296–D302. doi: 10.1093/nar/gkx1067

Cui, C., Yang, W., Shi, J., Zhou, Y., Yang, J., Cui, Q., et al. (2018). Identification and analysis of human sex-biased MicroRNAs. *Genomics Proteomics Bioinf.* 16 (3), 200–211. doi: 10.1016/j.gpb.2018.03.004

Eekhout, I., Reijnen, A., Vermetten, E., and Geuze, E., (2016). Post-traumatic stress symptoms 5 years after military deployment to Afghanistan: an observational cohort study. *Lancet Psychiatry* 3 (1), 58–64. doi: 10.1016/S2215-0366(15)00368-5

Ferry, F. R., Brady, S. E., Bunting, B. P., Murphy, S. D., Bolton, D., O'Neill, S. M., et al. (2015). The economic burden of PTSD in Northern Ireland. *J Trauma Stress* 28 (3), 191–197. doi: 10.1002/jts.22008

Fontana, A., and Rosenheck, R. (1994). Traumatic war stressors and psychiatric symptoms among World War II, Korean, and Vietnam War veterans. *Psychology Aging* 9 (1), 27–33. doi: 10.1037/0882-7974.9.1.27

Goldberg, A. D., Allis, C. D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. *Cell* 128 (4), 635–638. doi: 10.1016/j.cell.2007.02.006

Grund, B., and Sabin, C. (2010). Analysis of biomarker data: logs, odds ratios, and receiver operating characteristic curves. *Curr. Opin. HIV AIDS* 5 (6), 473–479. doi: 10.1097/COH.0b013e32833ed742

Han, F., Yan, S., and Shi, Y. (2013). Single-prolonged stress induces endoplasmic reticulum-dependent apoptosis in the hippocampus in a rat model of post-traumatic stress disorder. *PLoS One* 8 (7), e69340. doi: 10.1371/journal.pone.0069340

Hovens, J. E., Bramsen, I., and van der Ploeg, H. M. (2002). Self-rating inventory for posttraumatic stress disorder: review of the psychometric properties of a new brief Dutch screening instrument. *Percept. Motor Skills* 94 (3), 996–1008. doi: 10.2466/pms.2002.94.3.996

Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4 (1), 44–57. doi: 10.1038/nprot.2008.211

Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37 (1), 1–13. doi: 10.1093/nar/gkn923

Jia, Y., Han, Y., Wang, X., and Han, F. (2018). Role of apoptosis in the post-traumatic stress disorder model-single prolonged stressed rats. *Psychoneuroendocrinology* 95, 97–105. doi: 10.1016/j.psyneuen.2018.05.015

Kalisch, R., Baker, D. G., Basten, U., Boks, M. P., Bonanno, G. A., Brummelman, E., et al. (2017). The resilience framework as a strategy to combat stress-related disorders. *Nat. Hum. Behav.* 1 (11), 784–790. doi: 10.1038/s41562-017-0200-8

Koenen, K. C., Ratanatharathorn, A., Ng, L., McLaughlin, K. A., Bromet, E. J., Stein, D. J., et al. (2017). Posttraumatic stress disorder in the world mental health surveys. *Psychol. Med.* 47 (13), 2260–2274. doi: 10.1017/S0033291717000708

Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, D68–D73. doi: 10.1093/nar/gkt1181

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559. doi: 10.1186/1471-2105-9-559

Lehrner, A., and Yehuda, R. (2014). Biomarkers of PTSD: military applications and considerations. *Eur. J. Psychotraumatol.* 5, 10.3402/ejpt.v5.23797. doi: 10.3402/ejpt.v5.23797

Li, Y., Han, F., and Shi, Y. (2013). Increased neuronal apoptosis in medial prefrontal cortex is accompanied with changes of Bcl-2 and bax in a rat model of post-traumatic stress disorder. *J. Mol. Neurosci.* 51 (1), 127–137. doi: 10.1007/s12031-013-9965-z

Li, D. W., Liu, J. Z., Li, S. C., Yang, J. B., Sun, H. H., Wang, A. H., et al. (2018). Fear conditioning downregulates miR-138 expression in the hippocampus to facilitate the formation of fear memory. *Neuroreport* 29 (16), 1418–1424. doi: 10.1097/WNR.0000000000001129

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi: 10.1186/s13059-014-0550-8

Martin, C. G., Kim, H., Yun, S., Livingston, W., Fetta, J., Mysliwiec, V., et al. (2017). Circulating miRNA associated with posttraumatic stress disorder in a cohort of military combat veterans. *Psychiatry Res.* 251, 261–265. doi: 10.1016/j.psychres.2017.01.081

Mkrtchyan, G. M., Boyadzhyan, A. S., Avetyan, D. G., and Sukiasyan, S. G., (2014). The involvement of abnormal apoptosis in the disturbance of synaptic plasticity in posttraumatic stress disorder. *Neurosci. Behav. Physiol.* 44 (4), 26–29.

Nock, M. K., and Kessler, R. C. (2006). Prevalence of and risk factors for suicide attempts versus suicide gestures: analysis of the national comorbidity survey. *J. Abnorm. Psychol.* 115 (3), 616–623. doi: 10.1037/0021-843X.115.3.616

Peschansky, V. J., and Wahlestedt, C. (2014). Non-coding RNAs as direct and indirect modulators of epigenetic regulation. *Epigenetics* 9 (1), 3–12. doi: 10.4161/epi.27473

Reijnen, A., Rademaker, A. R., Vermetten, E., and Geuze, E. (2015). Prevalence of mental health symptoms in Dutch military personnel returning from deployment to Afghanistan: a 2-year longitudinal analysis. *Eur. Psychiatry* 30 (2), 341–346. doi: 10.1016/j.eurpsy.2014.05.003

Richter-Levin, G., Stork, O., and Schmidt, M. V. (2018). Animal models of PTSD: a challenge to be met. *Mol. Psychiatry* 24, 1135–1156. doi: 10.1038/s41380-018-0272-5

Schmidt, U., Holsboer, F., and Rein, T. (2011). Epigenetic aspects of posttraumatic stress disorder. *Dis. Markers* 30 (2-3), 77–87. doi: 10.1155/2011/343616

Snijders, C., de Nijs, L., Baker, D. G., Hauger, R. L., van den Hove, D., Kenis, G., et al. (2018a). MicroRNAs in post-traumatic stress disorder. *Curr. Top Behav. Neurosci.* 38, 23–46. doi: 10.1007/7854_2017_32

Snijders, C., Pries, L. K., Sgammeglia, N., Al Jowf, G., Youssef, N. A., de Nijs, L., et al. (2018b). Resilience against traumatic stress: current developments and future directions. *Front. Psychiatry* 9, 676. doi: 10.3389/fpsyt.2018.00676

Tarrier, N., and Gregg, L. (2004). Suicide risk in civilian PTSD patients–predictors of suicidal ideation, planning and attempts. *Soc. Psychiatry Psychiatr. Epidemiol.* 39 (8), 655–661. doi: 10.1007/s00127-004-0799-4

Thomas, M. M., Harpaz-Rotem, I., Tsai, J., Southwick, S. M., and Pietrzak, R.H. (2017). Mental and physical health conditions in us combat veterans: results from the national health and resilience in veterans study. *Prim. Care Companion CNS Disord.* 19 (3). doi: 10.4088/PCC.17m02118

van Zuiden, M., Geuze, E., Willemen, H. L., Vermetten, E., Maas, M., Heijnen, C. J., et al. (2011). Pre-existing high glucocorticoid receptor number predicting development of posttraumatic stress symptoms after military deployment. *Am. J. Psychiatry* 168 (1), 89–96. doi: 10.1176/appi.ajp.2010.10050706

Wingo, A. P., Almli, L. M., Stevens, J. S., Klengel. T., Uddin, M., Li, Y., et al. (2015). DICER1 and microRNA regulation in post-traumatic stress disorder with comorbid depression. *Nat. Commun.* 6, 10106. doi: 10.1038/ncomms10106

Yehuda, R., Neylan, T. C., Flory, J. D., and McFarlane, A. C. (2013). The use of biomarkers in the military: from theory to practice. *Psychoneuroendocrinology* 38 (9), 1912–1922. doi: 10.1016/j.psyneuen.2013.06.009

Yehuda, R., Vermetten, E., McFarlane, A. C., and Lehrner, A. (2014). PTSD in the military: special considerations for understanding prevalence, pathophysiology and treatment following deployment. *Eur. J. Psychotraumatol.* 5. doi: 10.3402/ejpt.v5.25322

Zeilinger, S., Kühnel, B., Klopp, N., Baurecht, H., Kleinschmidt, A., Gieger, C., et al. (2013). Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One* 8 (5), e63812. doi: 10.1371/journal.pone.0063812

Zhou, J., Nagarkatti, P., Zhong, Y., Ginsberg, J. P., Singh, N. P., Zhang, J., et al. (2014). Dysregulation in microRNA expression is associated with alterations in immune functions in combat veterans with post-traumatic stress disorder. *PLoS One* 9 (4), e94075. doi: 10.1371/journal.pone.0094075

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Snijders, Krauskopf, Pishva, Eijssen, Machiels, Kleinjans, Kenis, van den Hove, Kim, Boks, Vinkers, Vermetten, Geuze, Rutten and de Nijs. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 29 November 2019
doi: 10.3389/fgene.2019.01242

E-Cadherin Downregulation is Mediated by Promoter Methylation in Canine Prostate Cancer

Carlos Eduardo Fonseca-Alves[1,2], Priscila Emiko Kobayashi[3], Antonio Fernando Leis-Filho[3], Patricia de Faria Lainetti[2], Valeria Grieco[4], Hellen Kuasne[5], Silvia Regina Rogatto[6†] and Renee Laufer-Amorim[2†]*

[1] Institute of Health Sciences, Paulista University—UNIP, Bauru, Brazil, [2] Department of Veterinary Surgery and Anesthesiology, School of Veterinary Medicine and Animal Science, Sao Paulo State University—UNESP, Botucatu, Brazil, [3] Department of Veterinary Clinic, School of Veterinary Medicine and Animal Science, Sao Paulo State University—UNESP, Botucatu, Brazil, [4] Department of Veterinary Medicine, Università degli studi di Milano, Milan, Italy, [5] International Center for Research (CIPE), AC Camargo Cancer Center, Sao Paulo, Brazil, [6] Department of Clinical Genetics, University Hospital of Southern Denmark, Institute of Regional Health Research, University of Southern Denmark, Vejle, Denmark

OPEN ACCESS

Edited by:
Mojgan Rastegar, University of Manitoba, Canada

Reviewed by:
Sanjay Gupta, Case Western Reserve University, United States
Mónica Hebe Vazquez-Levin, National Council for Scientific and Technical Research (CONICET), Argentina

***Correspondence:**
Carlos Eduardo Fonseca-Alves
carlos.e.alves@unsp.br.

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *10 May 2019*
Accepted: *11 November 2019*
Published: *29 November 2019*

Citation:
Fonseca-Alves CE, Kobayashi PE, Leis-Filho AF, Lainetti PdF, Grieco V, Kuasne H, Rogatto SR and Laufer-Amorim R (2019) E-Cadherin Downregulation is Mediated by Promoter Methylation in Canine Prostate Cancer. Front. Genet. 10:1242. doi: 10.3389/fgene.2019.01242

E-cadherin is a transmembrane glycoprotein responsible for cell-to-cell adhesion, and its loss has been associated with metastasis development. Although E-cadherin downregulation was previously reported in canine prostate cancer (PC), the mechanism involved in this process is unclear. It is well established that dogs, besides humans, spontaneously develop PC with high frequency; therefore, canine PC is an interesting model to study human PC. In human PC, *CDH1* methylation has been associated with E-cadherin downregulation. However, no previous studies have described the methylation pattern of *CDH1* promoter in canine PC. Herein, we evaluated the E-cadherin protein and gene expression in canine PC compared to normal tissues. DNA methylation pattern was investigated as a regulatory mechanism of *CDH1* silencing. Our cohort is composed of 20 normal prostates, 20 proliferative inflammatory atrophy (PIA) lesions, 20 PC, and 11 metastases from 60 dogs. The E-cadherin protein expression was assessed by immunohistochemistry and western blotting and gene expression by qPCR. Bisulfite-pyrosequencing assay was performed to investigate the *CDH1* promoter methylation pattern. Membranous E-cadherin expression was observed in all prostatic tissues. A higher number of E-cadherin negative cells was detected more frequently in PC compared to normal and PIA samples. High-grade PC showed a diffuse membranous positive immunostaining. Furthermore, PC patients with a higher number of E-cadherin negative cells presented shorter survival time and higher Gleason scores. Western blotting and qPCR assays confirmed the immunohistochemical results, showing lower E-cadherin protein and gene expression levels in PC compared to normal samples. We identified *CDH1* promoter hypermethylation in PIA and PC samples. An in vitro assay with two canine prostate cancer cells (PC1 and PC2 cell lines) was performed to confirm the methylation as a regulatory mechanism of E-cadherin expression. PC1 cell line presented *CDH1* hypermethylation and after 5-Aza-dC treatment, a decreased *CDH1* methylation and increased gene expression levels were observed. Positive E-cadherin cells were massively found in metastases (mean of 90.6%). In conclusion, low levels of E-cadherin

protein, gene downregulation and *CDH1* hypermethylation was detected in canine PC. However, in metastatic foci occur E-cadherin re-expression confirming its relevance in these processes.

Keywords: dog, *CDH1*, prostate, hypermethylation, surface protein

INTRODUCTION

Human prostate cancer (PC), the second cause of male cancer-related death in North America, has a variable behavior (Siegel et al., 2019). The mortality rate is associated with metastasis (Huynh et al., 2016), which more commonly affects bone, lymph node, and lung (Siegel et al., 2019). Canine PC is a very aggressive disease associated with high metastatic rate at the diagnosis (more than 85%) being bones, lungs, and iliac lymph nodes, the most common metastatic sites disease-associated (Cornell et al., 2000; Fonseca-Alves et al., 2015a).

Dogs have been reported as a model for human PC and the knowledge regarding molecular aspects of canine PC has increased in recent years (Fonseca-Alves et al., 2018a; Costa et al., 2019; Laufer-Amorim et al., 2019; Rivera-Calderón et al., 2019). These recent studies bring new evidence that canine PC can represent a model to human castration-resistant prostate cancer (CRPC) (Laufer-Amorim et al., 2019). Usually, canine PC lacks NKX3.1, PTEN (Fonseca-Alves et al., 2013; Fonseca-Alves et al., 2018a; Fonseca-Alves et al., 2018b), and androgen receptor expression (Laufer-Amorim et al., 2019) resembling human CRPC. Besides that, canine PC shows alterations in TP53, C-MYC, and MDM2 protein expression (Fonseca-Alves et al., 2013; Fonseca-Alves et al., 2018b). These findings pointed out that the clinical behavior and molecular alterations are similar in both species, making dogs an exciting model in comparative initiatives.

The carcinogenic process, from normal to pre-neoplastic and invasive carcinoma, involves the ability of epithelial cells to detach one another, survive and invade the surrounding tissues (Friedl and Wolf, 2003). Metastasis of PC is a complex process associated with loss of epithelial markers, acquirement of a mesenchymal phenotype, and ability of cells to spread through the lymphatic system or bloodstream (Staník et al., 2014). E-cadherin is a transmembrane protein that has a crucial role in cell adhesion and migration (Debelec-Butuner et al., 2014). E-cadherin also is involved in the β catenin/APC pathway, which is related to cell proliferation and epithelial-mesenchymal transition (EMT) (Tsui et al., 2016). Loss of E-cadherin is associated with poor prognosis in patients with high-grade prostate tumors in both humans (Umbas et al., 1992; Umbas et al., 1994; Abdelrahman et al., 2017; Dhar et al., 2017; Wang et al., 2017; Li et al., 2019) and canine (Fonseca-Alves et al., 2013; Fonseca-Alves et al., 2015a; Kobayashi et al., 2018).

Different mechanisms have been implicated with E-cadherin downregulation in human medicine, including copy number loss (Saramaki and Visakorpi, 2007), somatic mutations (Busch et al., 2017), methylation (Graff et al., 1995; Yoshiura et al., 1995; Li et al., 2001; Mostafavi-Pour et al., 2015), and suppression mediated by ZEB1 and SRC family kinases (Mostafavi-Pour et al., 2015). *CDH1* gene repression promoted by its promoter hypermethylation, plays a crucial role in tumor invasion and spread (Graff et al., 1995; Yoshiura et al., 1995; Li et al., 2001; Mostafavi-Pour et al., 2015). *CDH1* hypermethylation and E-cadherin downregulation have been reported in more than 75% of patients with metastatic PC (Maruyama et al., 2002; Singal et al., 2004; Hoque et al., 2005). Also, *CDH1* promoter methylation is widely studied as a cause of E-cadherin down-regulation in human PC (Graff et al., 1995; Yoshiura et al., 1995; Li et al., 2001; Mostafavi-Pour et al., 2015). However, conflicting results have been reported due to the difficulties in studying methylation (Zhang et al., 2016b). Disparities among methodologies, sample quality, regions of prostatic biopsy, and promoter region evaluated make difficult comparisons among the published studies (Zhang et al., 2016b). Besides that, neoplastic cells can induce hypomethylation and re-express the transcript and its respective protein (Chao et al., 2010), which is compatible with the reversibility phenomenon described in the methylation process.

Transcriptional E-cadherin downregulation mediated by its promoter methylation is widely investigated in human PC (Graff et al., 1995; Yoshiura et al., 1995; Li et al., 2001; Mostafavi-Pour et al., 2015), and E-cadherin plasticity has been proposed during the metastatic progression in human PC (Bae et al., 2011). In high-grade human PC, E-cadherin loss leads to the invasion of metastatic cells to lymph nodes and bones (Putzke et al., 2011). Interestingly, bone metastasis seems to express more E-cadherin than soft tissue metastasis (Putzke et al., 2011). However, few studies evaluating the molecular mechanisms related to *CDH1* silencing have been reported in dogs. Loss of E-cadherin during the lymphatic invasion by neoplastic epithelial cells and E-cadherin re-expression in metastatic foci were previously reported in canine PC (Fonseca-Alves et al., 2015a).

Herein, we investigated E-cadherin gene and protein expression in canine proliferative inflammatory atrophy (PIA), PC and its metastasis as well the methylation status of *CDH1* as a silencing mechanism responsible for the dynamic E-cadherin expression.

MATERIALS AND METHODS

Tissue Selection and Histopathological Evaluation

This cohort is composed of 60 dogs of different breeds, varying from 8 to 14 years old. We selected 20 normal canine prostates, 20 PIA lesions, and 9 PC formalin-fixed embedded-paraffin (FFPE) from the archives from the Department Veterinary Pathology, Sao Paulo State University- UNESP, Brazil. In addition, 11FFPE

prostate cancer matched with 11 metastases from the same subjects were selected. All metastases were morphologically analyzed and presented PSA protein expression, as previously described (Fonseca-Alves et al., 2018b). The correspondent fresh frozen tissues from 20 normal canine prostates, 20 PIA lesions, 20 PC samples were used for pyrosequencing and Western blot. All FFPE samples were evaluated by protein and gene expression using immunohistochemistry and qPCR, respectively.

PC samples were collected during surgical or biopsy procedures from animals showing clinical signs. The metastases were identified by imaging tests (X-ray or computed tomography) followed by a biopsy. Normal and PIA samples were collected during necropsies from animals without clinical signs of prostatic disease, with an interval between death and necropsy less than 6 h. All prostate samples were from intact dogs.

The histopathological classification was performed according to the human WHO classification of Tumors of the Urinary System and Male Genital Organs (Humphrey et al., 2016). The Gleason-like system was applied according to Palmieri and Grieco (Palmieri and Grieco, 2015). Briefly, the architectural patterns are evaluated, and the sum of the primary and secondary grades is determined to result in a final Gleason score.

The study was approved by the Animal Ethics Committee according to the national and international guidelines for using animals in research. All animal owners gave written informed consent for the dog's material, clinical information and examination results to be used for research and academic matters under protocol #107/2015.

E-cadherin Expression Analysis by Immunohistochemistry

Five-micron thick sections were obtained from FFPE blocks, dewaxed in xylol and rehydrated in graded ethanol. For antigen retrieval, the slides containing the samples were incubated with citrate buffer (pH 6.0) in a pressure cooker (Pascal®; Dako, Carpinteria, CA, USA). The samples were then treated with freshly prepared 3% hydrogen peroxide in methanol for 20 min and further washed in Tris-buffered saline. The slides were incubated overnight at 4°C with 0.01µg/µL monoclonal mouse Anti-Human E-cadherin antibody (catalog number GA059, Dako, Carpinteria, CA, USA). A polymer system (catalog number K406511-2, Envision, Dako, Carpinteria, CA, USA) was applied as a secondary antibody conjugated to peroxidase. DAB (3′-diaminobenzidine tetrahydrochloride, Dako, Carpinteria, CA, USA) was used as the chromogen, for 5 min, followed by Harris hematoxylin counterstain. Negative control using mouse universal negative control (Dako, Carpinteria, CA, USA) was included according to the manufacturer's recommendation. Positive E-cadherin cells in adjacent epithelial cells were considered positive internal controls.

E-cadherin immunoexpression was evaluated according to the number of negative cells. Slides were analyzed under a light microscope (Leica Microsystems, Germany) and 10 images were taken for each slide (Leica QWin V3 software; Leica Microsystems, Germany) at high-power (40X objective) field. Representative areas were qualitatively selected for immunostaining analysis. We choose areas with minimal inflammatory cells, necrosis or connective tissue and with lower E-cadherin staining. Samples were scored based on an assessment of the number of negative cells per the total of cells in 10 high power fields (HPF), according to Hong et al. (2011). These results were expressed in a percentage of negative cells.

E-cadherin/Ki67 Double Immunostaining

E-cadherin and Ki67 double immunoexpression were performed to exclude cell proliferation as a mechanism associated with E-cadherin focal loss. The procedures were performed as previously reported (Fonseca-Alves et al., 2015b). Briefly, the paraffin sections were deparaffinated in xylol for 15 min and antigen retravel was performed using citrate buffer pH 6.0 solution in a pressure cooker (Pascal, Dako, Carpinteria, CA, USA). Then, endogenous peroxidase was blocked using 8% of hydrogen peroxidase (Dinamica, São Paulo, SP, Brazil), diluted in methanol (Dinamica, São Paulo, SP, Brazil). We used 0.02µg/µL of mouse monoclonal anti-Ki67 antibody (catalog number GA62661-2, Dako, Carpinteria, CA, USA) overnight at 4°C. The polymer system was applied as a secondary antibody for 1 h (catalog number K406511-2, Envision, Dako, Carpinteria, CA, USA) and 3′-diaminobenzidine tetrahydrochloride (DAB, Dako, Carpinteria, CA, USA) was used as the chromogen, for 5 min. The tissue sections were washed with immunohistochemistry buffer (Dako, Carpinteria, CA, USA) and 0.01µg/µL of mouse monoclonal anti-E-cadherin antibody (catalog number GA059, Dako, Carpinteria, CA, USA) was applied overnight at 4°C. After, the HRP magenta chromogen (catalog number GV925, Dako, Carpinteria, CA, USA) was used for 5 min and counterstained with Harris hematoxylin. The positive and negative controls were performed, as described above.

Immunoblotting

Western blotting was performed to quantify E-cadherin protein expression in seven normal prostates, seven PIA lesions, and seven PC. The frozen prostate samples were sectioned in a cryostat and re-analyzed to confirm the previous diagnosis. The samples were mechanically homogenized, prepared and transferred to nitrocellulose membranes, as previously described (Rivera-Calderón et al., 2016). The blots were blocked with 6% skimmed milk in TBS-T (BioRad, Hercules, CA, USA) for 2 h. Next, the Mouse monoclonal anti-human E-cadherin (0.002µg/µL; catalog number GA059, Dako, Carpinteria, CA, USA) antibody was applied and the slides were incubated at 4°C for 18 h. Goat polyclonal anti-β-actin antibody (0.001µg/µL, catalog number sc-1615, Santa Cruz Biotechnology, Santa Cruz, CA, USA) was used as a loading control. After incubation with the corresponding horseradish peroxidase-conjugated sheep anti-mouse (catalog number NA931, GE Healthcare, Chicago, IL, USA) and donkey anti-goat (catalog number NA9340, GE Healthcare, Chicago, IL, USA) secondary antibodies (0.001µg/µL), the blots were detected by means of chemiluminescence (Amersham ECL Select Western Blotting Detection Reagent, GE Healthcare). Protein bands were quantified by densitometry analysis (Imagequant LAS 500,

GE Healthcare, Chicago, IL, USA) and expressed as integrated optical density (IOD). E-cadherin protein expression was normalized using the β-actin values. Normalized data were expressed in means and standard deviations (SD).

Tumor-Derived Cell Cultures

Two cell lines (PC1 and PC2) were established in our previous study (Zhang et al., 2016). The PC1 cell line was from a 10-years-old, intact, mixed breed dog with non-metastatic PC (cribriform pattern and Gleason score 10). PC2 cell line was from an 11-year-old, intact, poodle dog with metastatic PC (tumor showed cribriform pattern and Gleason score 10). Both cell lines were cultured (the passage 30) in DMEM medium (Lonza, Basel, Switzerland) containing 10% fetal bovine serum (FBS) (LGC Bio, Cotia, SP, Brazil), 1% of penicillin-streptomycin (Thermo Fischer Scientific, Waltham, MA, USA) and amphotericin B (Thermo Fischer Scientific, Waltham, MA, USA) at 37°C in a humidified atmosphere containing 5% CO2. After reaching a minimum of 80% of confluence, both cell lines were processed to obtain DNA. DNA extraction were also performed in their respective primary tumors (fresh frozen samples) followed by pyrosequencing to evaluate the *CDH1* methylation status.

Methyl Thiazolyl Tetrazolium (MTT) Assay

The 5-Aza 2′deoxycytidine (5-Aza-dC) toxicity was investigated in canine prostatic cells based on the MTT assay. The IC_{50} values were calculated from the dose-response curves to establish the in vitro dosage that will induce demethylation instead of cell death. We used 96-well plates to grow the cancer cells at a density of 2,500 cells per well. The medium was changed every 48 h, and 5-Aza-dC (Sigma-Aldrich, Saint Louis, MO, USA) was added every 24 h. MTT analysis was performed on day 7. The medium was removed, the cells were washed with 3X PBS, and fresh medium was added in each well followed by incubation at 37°C for 4 h. The medium was removed and 200μL of dimethyl sulfoxide (DMSO) (Sigma-Aldrich, Saint Louis, MO, USA) was added in each well and formazan (Sigma-Aldrich, Saint Louis, MO, USA) was solubilized. The optical density (OD) level was measured at 570 wavelengths. Each treatment was performed in triplicate and the experiment in duplicate. Cell viability was calculated into a percentage.

CDH1 Gene Expression

Gene expression analysis was performed in our set of samples and both cell lines prior and after 5-Aza-dC treatment. Macrodissection was performed in normal, PIA, PC, and metastatic samples (FFEP) using 16-gauge needles, as previously described (Hoque et al., 2005). mRNA was extracted using RecoverAll™ Total Nucleic Acid Kit (Ambion, Life Technologies, MA, USA) according to the manufacturer's instructions. cDNA synthesis was performed using total RNA (Applied Biosystems, Foster City, CA, USA), according to the manufacturer's recommendations. The primers set for *CDH1* (Gene ID: 442858) (Forward: 5′-CAGCATGGACTCAGAAGACAGAAG-3′ and Reverse: 5′-TTCCGGGCAGCTGATAGG-3′) and *ACTB* (Gene ID: 403580) used as endogenous (ACTB, Forward: 5′-GGCATCCTGACCCTCAAGTA-3′ and Reverse: 5′-CTTCT CCATGTCGTCCCAGT-3′) genes were used for RT-qPCR assays. The reaction was conducted in a total volume of 10 μL containing Power SYBR Green PCR Master Mix (Applied Biosystems; Foster City, CA, USA), 1 μL of cDNA (1:10) and 0.3 μM of each primer pair in triplicate using QuantStudio 12K Flex Thermal Cycler equipment (Applied Biosystems; Foster City, CA, USA). A dissociation curve was included in all experiments to determine the PCR product specificity. Relative gene expression was quantified using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001).

5-Aza-2′-Deoxycytidine Treatment

To investigate if hypermethylation is associated with *CDH1* silencing, we treated the PC cell lines with 5-Aza-dC and compared with untreated cells. As previously established by MTT assay, we added 1μg of 5-Aza-dC to the culture medium every 24 h (due to 5-Aza-dC stability) and for seven days. Treated cells were washed with PBS three times. All procedures were performed in duplicate, according to da Costa Prando et al. 2011). Subsequently, mRNA and DNA were extracted to perform RT-qPCR and pyrosequencing analysis, respectively.

Quantitative Bisulfite Pyrosequencing

The pyrosequencing analysis was performed to evaluate the frequency of *CDH1* gene promoter methylation in all frozen tissue samples (20 normal prostates, 20 PIA samples, and 20 PC) and cell lines (prior and after 5-Aza-dC treatment). Prostate samples were sectioned in a cryostat to confirm the diagnosis. The bisulfite conversion of the genomic DNA was performed using EZ DNA Methylation-Gold Kit (Zymo Research Corporation, Irvine, CA, USA). The forward (5′ TTTGGGAAGAGGAGGGGG 3′) and reverse primer (5′ CCCTTCCCCTCTCTCTCTC - BIOTIN 3′) of *CDH1* CpG island (Gene ID: 442858) were amplified by PCR (HotStarTaq Master Mix kit - Qiagen). The pyrosequencing was performed using a sequencing primer (5′ TTTGGGAAGAGGAGGGGG 3′) following the manufacturer's instructions (PyroMark ID Q96, Qiagen and Biotage, Uppsala, Sweden).

Statistical Analysis

Statistical analysis was performed using GraphPad Prism v.8.1.0 (GraphPad Software Inc., La Jolla, CA, USA). The column test was performed to evaluate data normality. For statistical purposes, the mean of E-cadherin negative cells was used as a threshold to compare the overall survival between patients with over and lower protein expression. Variance analysis (ANOVA) was applied to compare CDH1 transcript levels among normal, PIA and PC samples. Mann-Whitney test was used to evaluate the association of E-cadherin protein and gene expression between two categorical variables. Correlation among the IHC score and clinical parameters, protein expression and transcript levels were also investigated. Mann-Whitney test was applied to evaluate the differences in the methylation levels among the groups. The

samples were grouped according to the Gleason score in "low Gleason score" (Gleason score 6 and 8) and "high Gleason score" (Gleason score 10).

RESULTS

Clinical Features

The clinical features of the 20 PC-affected dogs are described in **Table 1**. Survival information was not available in two of 20 PC patients. The 20 canine PC preseted Gleason scores 6 (30% of cases), 8 (15%) and 10 (55%). Eleven of 20 dogs with PC had metastasis (55%); eight of them (8/11) presented bone and lung metastasis while pelvic bones, intestine and liver were observed in one patient each. From the patients with multiple metastatic sites (bone and lung), only the bone biopsy was evaluated. Seventy-three percent (8/11) of PC patients showing Gleason score 10 had metastasis. Dogs with PC Gleason 8 had no metastasis (n = 3), while 50% (3/6) of cases with Gleason 6 showed metastasis at diagnosis. Patients with lower Gleason score (6 and 8) experienced a higher survival time (P = 0.003) than those with Gleason score 10 (**Figure 1A**).

E-cadherin Immunoexpression

We found positive epithelial cells with membranous staining in normal, PIA, PC (**Figure 2**), and metastasis. Cases with less than 10% of negative cells showed a higher survival time (P = 0.004) (**Figure 1B**). A higher number of negative cells was observed in PC (**Figure 1C**) compared to normal and PIA samples. Normal samples showed 100% E-cadherin positive cells; while a mean of 2.1% and 10.5% of negative cells was detected in PIA and PC samples, respectively. Metastases had a mean of 9.5% of negative cells. Tumors showing Gleason score 10 had a higher percentage of negative E-cadherin neoplastic cells compared to PC Gleason scores 6 and 8 and normal samples (P = 0.0003). Metastases had a higher number of negative cells in comparison with normal samples (P = 0.0003) and no statistical difference was observed between all PC samples and metastases (P > 0.05). E-cadherin pattern in each histological subtype is detailed in **Table 1**. The

TABLE 1 | Clinical information of 20 canine prostate cancer-affected patients evaluated in this study.

Case	Breed	Age (years)	Histological Pattern	Gleason-like score*	Treatment	Metastasis**	E-cadherin Negative Cells (%)	E-cadherin Positive Cells (%)	Methylation (%)	Follow-up (days)
1	Boxer	14	Cribriform	10	Piroxicam	Lung, Bone and Liver	30	70	94	90
2	Boxer	12	Cribriform	10	LDMT	Bone, Lung	15	85	94	278
3	German Shepherd	8	Small acinar	6	RP	No	2	98	95	453
4	American Cocker Spaniel	9	Small acinar	6	LDMT	No	1	99	95	523
5	Poodle	13	Small acinar	6	N/T	Bone, Lung	10	90	95	321
6	Poodle	14	Small acinar	6	LDMT	No	20	80	94	674
7	Boxer	11	Small acinar	6	Piroxicam	Bone	1	99	92	52
8	MBD	14	Small acinar	6	LDMT	Lung, Intestine	0	100	93	132
9	Poodle	13	Papillary	8	Piroxicam	No	10	90	95	463
10	MBD	12	Papillary	8	Carboplatin + Piroxicam	No	0	100	92	567
11	MBD	11	Papillary	8	LDMT	No	8	92	93	368
12	American Cocker Spaniel	10	Small acinar	10	RP	No	2	98	98	32
13	Poodle	10	Solid	10	RP	No	17	83	100	213
14	MBD	10	Solid	10	Piroxicam	Lung, Liver	0	100	95	55
15	MBD	15	Cribriform with comedonecrosis	10	Doxorrubicin + Piroxicam	Bone, Lung	10	90	94	75
16	American Cocker Spaniel	10	Solid	10	Doxorrubicin	Bone, Lung	22	78	100	78
17	MBD	9	Cribriform	10	LDMT	Bone, Lung	15	85	98	375
18	MBD	7	Cribriform with comedonecrosis	10	N/A	Bone, Lung	25	75	96	N/A
19	MBD	13	Cribriform with comedonecrosis	10	RP	Bone, Lung	5	95	97	45
20	Teckel	11	Cribriform	10	N/A	No	9	91	95	N/A

*PC, prostate cancer; MBD, Mixed Breed dog; N/A, Not Available; N/T, No Treatment; RP, Radical Prostatectomy; LDMT, Low-dose metronomic therapy. * Gleason like score was evaluated according to Palmieri and Grieco (2015). ** Metastasis identified at the diagnosis or during the follow-up.*

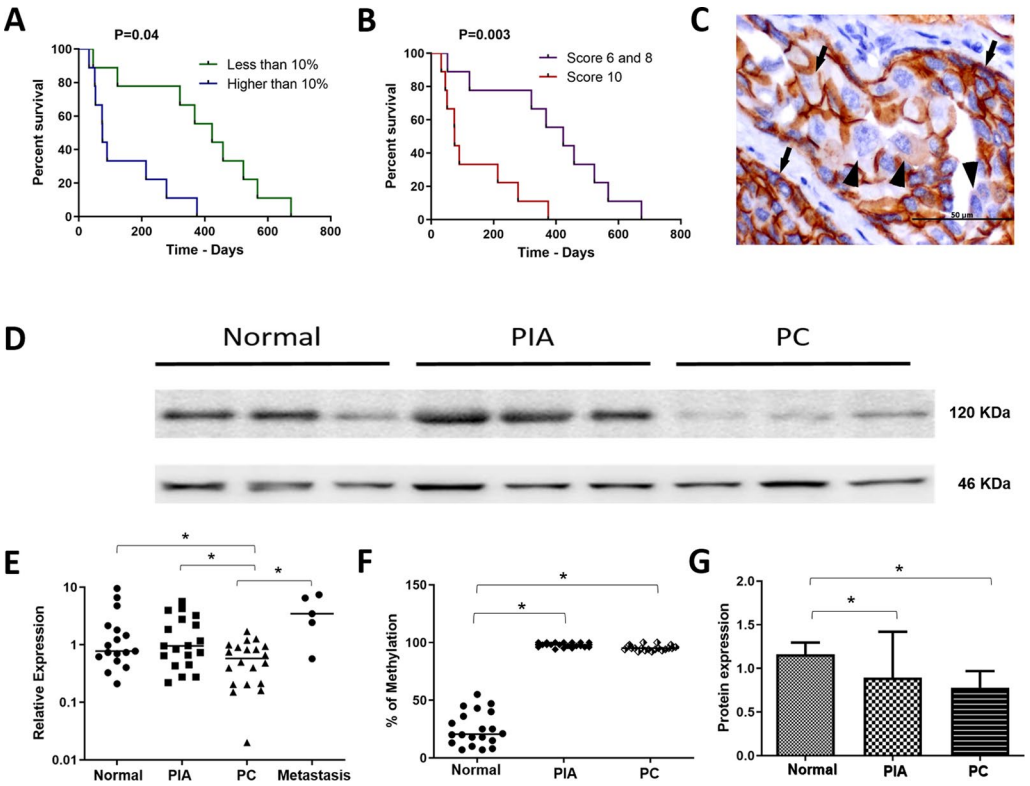


FIGURE 1 | (A) survival analysis according to the percentage of E-cadherin negative cells. Patients with over than 10% o E-cadherin negative cells experienced a shorter survival time. **(B)** survival analysis of the canine prostate cancer affected patients according to the Gleason score. Patients with Gleason score 10 experienced a shorter survival time. **(C)** E-cadherin immunohistochemistry showing positive membranous staining (arrows) in neoplastic epithelial cells. Cells were considered E-cadherin negative when partial or total (arrowhead) lack of expression. **(D)** Western blotting showing E-cadherin expression in normal, proliferative inflammatory atrophy and prostate cancer (PC) samples. It is possible to observe E-cadherin down expression in PC samples. **(E)** ANOVA analysis of *CDH1* transcripts in the different canine samples. The prostate cancer (PC) samples showed a lower *CDH1* transcript levels among normal, proliferative inflammatory atrophy (PIA) and metastasis. **(F)** Graphic representation of the percentage of methylation in normal, PIA and PC samples. PIA and PC samples were hypermethylated compared to normal samples. **(G)** graphic representation of E-cadherin protein expression by Western blotting after normalization with β-actin. It is possible to observe lack in both PIA and PC compared to normal samples. *Statistical difference between two variable comparisons.

comparison between E-cadherin expression clinical-pathological data is summarized in **Table 2**.

Comparing the E-cadherin immunoexpression between the primary tumors and its paired metastasis, no statistical difference was found (P > 0.05). The mean of E-cadherin negative cells was similar in primary PC and its paired metastasis (15 ± 7.09 and 16.8 ± 5.25, respectively). No correlation was found between the number of E-cadherin negative cells in the primary PC samples (N = 11) and its respective metastasis (N = 11) (r = 0.076, P = 0.8223). In addition, this comparison was no significant by regression analysis [F (1, 9) = 0.01838, P = 0.08951, R^2 = 0.7071]. Although in a limited number of cases, a significant difference was observed comparing E-cadherin negative cells in bone metastasis (N = 9; 14.77 ± 4.02) with those in soft tissues (N = 2; 26 ± 5.0) A positive correlation between E-cadherin negative cells (r = 0.8565, P = 0.0052) was found comparing only primary tumors with the respective paired bone metastasis. We also found a significant regression equation (F (1, 7) = 25.08, P = 0.0016, R^2 = 0.7818), comparing primary tumors with their respective metastasis. We observed a positive correlation between the Gleason score and the number of negative E-cadherin neoplastic cells (R = 0.8505 and P < 0.0001) and a significant regression equation [F (1, 18) = 36.18, P < 0.0001), R^2 = 0.6678]. Overall, prostate cancer with a high Gleason score showed a higher number of negative E-cadherin cells in comparison with those with lower Gleason scores. The linear regression graphics are shown in **Supplementary Figure 1**.

We also investigated the proliferative index in E-cadherin negative areas using E-cadherin/Ki67 double immunoexpression. All normal samples (N = 20) showed only membranous E-cadherin with no nuclear Ki67 expression. On the other hand, it was identified a higher number of double-stained epithelial cells in PIA samples (N = 20). In PC samples, areas with E-cadherin downregulation showed only scattered Ki67 expression, indicating a low proliferative index (**Supplementary Figure 2**).

Western Blotting

A strong 120 KDa band was identified in normal prostate tissues (**Figures 1D**, **G**). No statistical difference was observed comparing the E-cadherin expression in normal prostates with PIA samples.

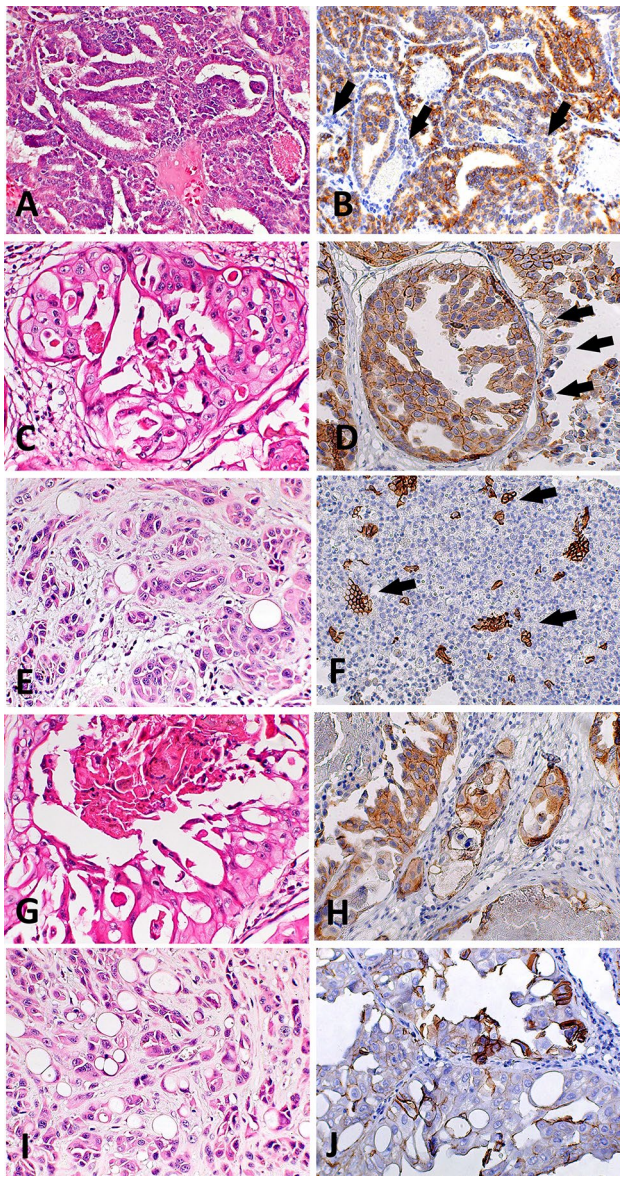


FIGURE 2 | Histological and immunohistochemical E-cadherin evaluation in canine prostate cancer (PC). **(A)** canine PC presenting a papillary pattern. It is possible to observe multifocal areas of E-cadherin loss **(B)** (arrows) in this pattern. **(C)** Canine PC with cribriform patter. Note E-cadherin membranous diffuse expression **(D)** in neoplastic cells and areas of E-cadherin loss (arrows). **(E)** Canine PC with solid pattern. **(F)** area of E-cadherin loss in canine PC with solid pattern. There are only few remaining positive cells (arrows). **(G)** Canine PC showing cribriform with central comedonecrosis pattern. **(H)** is possible to observe membranous E-cadherin expression in neoplastic cells with only few cells showing no E-cadherin expression. **(I)** Canine PC with signet ring pattern. **(J)** It is possible to observe multifocal areas with E-cadherin loss.

However, a lower E-cadherin expression was detected in PC compared to normal prostate (P = 0.0003) and PIA samples (P = 0.0001). **Supplementary Figure 3** is representative of the Western blotting assays performed in normal prostate, PIA, and PC samples.

CDH1 Gene Expression

PC samples showed lower *CDH1* transcript levels in comparison with PIA (P = 0.0038) and normal samples (P = 0.0427) (**Figure 1E**). No statistical difference was observed between the transcript levels in PIA and normal samples. Unfortunately, only five metastatic samples (5/11) were evaluated by RT-qPCR, mainly due to poor mRNA quality. The median of *CDH1* relative quantification (RQ) was 0.7 (0.2–9.5), 0.9 (0.2–5.6), 0.5 (0.02–1.7), and 3.45 (0.6–2.4) in normal, PIA, PC and metastases samples, respectively. In prostate cancer, a strong positive correlation was observed between high levels of E-cadherin protein expression and *CDH1* transcript levels (Spearman R = 0.9429; P = 0.0167) (Significant regression equation: F (1, 4) = 9.654, P= 0.036, R^2 = 0.7071). *CDH1* gene expression between the primary tumors (N = 5) and its paired metastasis (N = 5) showed no correlation (r = 0.2000, P = 0.7833) and no significant regression equation [F (1, 2) = 0.06048, P = 0.8216, R^2 = 0.01976]. A higher methylation pattern was detected in samples with lower levels of *CDH1* transcripts and a higher number of E-cadherin negative cells, which revealed a direct association of the methylation pattern with gene and protein down expression.

Quantitative Bisulfite Pyrosequencing

CDH1 promoter hypermethylation was identified in PIA and PC compared to normal samples (P < 0.0001). The median of methylation was 20.5% (7–55%), 98% (94–100%) and 95% (94–100%) in normal, PIA and PC samples, respectively (**Figure 1F**) (**Supplementary Figure 4**).

InVitro Assays

CDH1 was hypermethylated and presented lower transcript levels (0.86±0.04) in the PC1 cell line. After the 5-Aza-dC treatment, this cell line presented an inverted methylation pattern and increased gene expression level (1.7 ±0.2).

DISCUSSION

In this study, E-cadherin gene and protein expression findings were associated with *CDH1* methylation in canine PC, which gives evidence of the regulatory mechanism of *CDH1* in canine

TABLE 2 | Mean percentage of E-cadherin negative and positive cells according to the diagnosis and Gleason score.

IHC results	Normal	PIA	PC	Metastasis	Gleason 6	Gleason 8	Gleason 10
Positive cells (%)	100±0	97.9	89.5±4.7	90.5±4.2	98.6±7.3	91%±6.9	82.6±7.1
Negative cells (%)	0±0	2.1	10.5±4.6	9.5±4.2	1.4±7.2	9±6.5	17.4±7.4

IHC' protein expression by immunohistochemistry; PIA, Proliferative inflammatory atrophy; PC, Prostate cancer.

PC. E-cadherin is a cell-to-cell adhesion molecule and its loss correlates with epithelial-mesenchymal transition, metastasis and poor prognosis (Putzke et al., 2011; Fonseca-Alves et al., 2015a). Considering the high variation among the different semi-quantitative scores for immunohistochemical evaluation, we counted the number of negative cells and provided a score. We found a higher number of negative E-cadherin cells in PC compared to PIA and normal prostate. Also, a lower number of positive cells was correlated with survival.

Similarly to our findings, Fonseca et al. (2013) and Tsui et al. (2016) reported that PIA presented a lack of E-cadherin expression compared with normal samples. Using Western blot, we confirmed these previous data. Moreover, no statistical difference was observed in PIA compared to PC in cases with a higher E-cadherin expression, which could be explained by the lack of metastatic potential and malignancy of these preneoplastic lesions. Furthermore, during cell proliferation, it is expected the presence of E-cadherin loss by epithelia cells related to cell division instead of a migration (Tsui et al., 2016). For this reason, we performed E-cadherin/Ki67 double staining and confirmed that tumor areas presented E-cadherin losses with no proliferative activity. This result strongly suggests that E-cadherin downregulation is more related to cell migration instead of proliferation. In human PC, *CDH1* hypermethylation and E-cadherin loss is more frequent in metastatic tumors with higher Gleason score (Maruyama et al., 2002). Similar results were observed in our canine PC samples. Although the Gleason score is relatively new in veterinary practice, our study is the first to associate Gleason score with overall survival and E-cadherin downregulation.

In human PC, E-cadherin downregulation is frequent in later stages of the disease and poorly differentiated tumors (Ipekci et al., 2015; Zhang et al., 2016a). Considering the dynamic process of E -cadherin expression, a group of cases with negative cells could also be associated with worse prognosis in canine PC. We showed an association between a higher number of E-cadherin negative cells with shorter survival time, suggesting that the number of E-cadherin negative cells could be used as a prognostic factor. To our knowledge, no previous studies presented the percentage of E-cadherin negative cells and their association with the prognosis in human PC (Graff et al., 1995; Yoshiura et al., 1995; Li et al., 2001; Mostafavi-Pour et al., 2015). On the other hand, in human pancreatic adenocarcinomas, Hong et al. (Hong et al., 2011) described the lowest survival time in patients with total a loss of E-cadherin compared with those with partial loss of the protein expression. The authors suggested that partial and total loss of E-cadherin are an independent negative prognostic factor. In human breast cancer, different authors associated E-cadherin decreased expression with worse prognosis, such as lower overall survival, disease-free interval, positive lymph node (Tang et al., 2012; Ricciardi et al., 2015; Wang et al., 2018), and higher proliferative rate evaluated by Ki-67 (Kashiwagi et al., 2011). In 103 prostate carcinomas, Ipekci et al. (Ipekci et al., 2015) showed E-cadherin decreased expression, but no correlation was found with disease-free survival. The authors suggested that epithelial-mesenchymal transition evaluated by E-cadherin, β-catenin, vimentin and Wnt is a late event in tumor progression. These proteins could not be detected in the primary tumor and, therefore, would not be good predictors of metastasis (Ipekci et al., 2015).

We found a strong positive correlation (r = 0.9424) between E-cadherin protein and gene expression in PC samples. Interestingly, we also found an association between the *CHD1* hypermethylation pattern with gene downregulation. The PC1 cell line was densely hypermethylated and associated with low transcript levels. After 5-Aza-dC treatment, *CDH1* hypomethylation and restoration of gene expression were detected. These results indicated an epigenetic regulation of *CDH1* in canine PC. Similar results were previously described in two prostatic cell lines, DuPro and TSUPr1 (Graff et al., 1995). Considering that DNA methylation is a reversible process, the 5-Aza-dC treatment was efficient in inducing gene demethylation, which suggested that hypermethylated tumors could be sensitive to epigenetic drugs. The hypomethylating agents have been used to treat acute myeloid leukemia (AML) with promising results (Cruijsen et al., 2014). Although our findings are preliminary, dogs could be a preclinical model in precision medicine for testing epigenetic agents in PC patients.

Although cells lacking E-cadherin expression acquire motility and show an invasive and migratory phenotype, only a few cells with no E-cadherin expression are required to develop micrometastasis (Umbas et al., 1994; Canel et al., 2013). Thus, the evaluation of this cell group is relevant for a better understanding of the metastatic process. E-cadherin downregulation occurs in most cases by posttranscriptional mechanisms (Canel et al., 2013). *CDH1* promoter hypermethylation is widely studied in many human cancers, including prostate cancer (Graff et al., 1995; Yoshiura et al., 1995; Li et al., 2001; Mostafavi-Pour et al., 2015). Interestingly, a mean of 90.5% of E-cadherin positive cells was detected in the metastasis. Our data reinforce that the modulation of the metastatic foci and adhesion molecules re-expression are pivotal for the metastasis development (Welch, 2007). A higher number of metastatic cases was observed (N = 3) in patients showing Gleason 10 (N = 8). These samples presented a mean of 17.4% of negative cells. Overall, these results suggest that a group of cells showing lack of E-cadherin expression in primary tumors would have the potential to invade and re-express E-cadherin in metastatic foci.

There is limited information regarding E-cadherin expression in human PC and its paired metastasis (Bae et al., 2011). During the invasion of an artificial basal cell membrane, prostatic cells presented loss of E-cadherin expression and re-expressed after overtaking the membrane (Bae et al., 2011). In dogs, the lack of E-cadherin expression was previously demonstrated in PCs and a complete E-cadherin loss was observed in the neoplastic emboli (Fonseca-Alves et al., 2015a). Interestingly, the paired metastasis showed E-cadherin re-expression. Thus, a dynamic E-cadherin expression occurs during the tumor progression to metastasis. Further studies to evaluate the *CDH1* methylation analysis in circulating prostate cancer cells and its prognostic value could be relevant for clinical purposes.

CONCLUSION

Our results suggested an epigenetic regulation of the E-cadherin promoter leading to E-cadherin downregulation in canine PC. The number of negative E-cadherin cells investigated by immunohistochemistry demonstrated the importance of these cells to PC prognosis. Overall, our results indicate that dogs could be a preclinical model for testing hypomethylating agents in precision medicine.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/ **Supplementary Material**.

ETHICS STATEMENT

This study was approved by the Animal Ethics Committee of the University of Sao Paulo State, UNESP, Botucatu, Brazil (#107/2015).

AUTHOR CONTRIBUTIONS

CF-A wrote the first manuscript draft. CF-A performed the immunohistochemistry and qPCR experiments. CF-A, AL-F, PL, and PK performed the cell culture experiments. CF-A and RL-A conceived the project and grant funding. CF-A, SR, and HK conceived and performed the pyrosequencing experiments. VG contributed constructive comments. RL-A and SR supervised the project and revised the manuscript. All authors read and approved the final version of the manuscript.

FUNDING

This research was funded by the Sao Paulo Research Foundation (FAPESP) grant (#2012/18426-1 and 2019/24649-2). National Council for Scientific and Technological Development (CNPq) (#422139/2018-1). We also would like to thank the research grant from the National Council for Scientific and Technological Development (CNPq) (#422139/2018-1).

ACKNOWLEDGMENTS

We would like to thank Dr. Marcio Carvalho for his technical support. We also thank to the A.C. Camargo Cancer Center, SP, Brazil for allowing the access to its institutional infrastructure.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01242/full#supplementary-material

REFERENCES

Abdelrahman, A. E., Arafa, S. A., and Ahmed, R. A. (2017). Prognostic value of twist-1, E-cadherin and EZH2 in prostate cancer: an immunohistochemical study. *Turk Patoloji Derg.* 4, 198–210. doi: 10.5146/tjpath.2016.01392

Bae, K. M., Parker, N. N., Dai, Y., Vieweg, J., and Siemann, D. W. (2011). E-cadherin plasticity in prostate cancer stem cell invasion. *Am. J. Cancer Res.* 1 (1), 71–84.

Busch, E. L., Hornick, J. L., Umeton, R., Albayrak, A., Lindeman, N. I., MacConaill, L. E., et al. (2017). Somatic mutations in CDH1 and CTNNB1 in primary carcinomas at 13 anatomic sites. *Oncotarget* 8 (49), 85680–85691. doi: 10.18632/oncotarget.21115

Canel, M., Serrels, A., Frame, M. C., and Brunton, V. G. (2013). E-cadherin-integrin crosstalk incancer invasion and metastasis. *J. Cell Sci.* 126 (Pt 2), 393–401. doi: 10.1242/jcs.100115

Chao, Y. L., Shepard, C. R., and Wells, A. (2010). Breast carcinoma cells re-express E-cadherin during mesenchymal to epithelial reverting transition. *Mol. Cancer* 9, 179. doi: 10.1186/1476-4598-9-179

Cornell, K. K., Bostwick, D. G., Cooley, D. M., Hall, G., Harvey, H. J., Hendrick, M. J., et al. (2000). Clinical and pathologic aspects of spontaneous canine prostate carcinoma: a retrospective analysis of 76 cases. *Prostate* 45 (2), 173–183. doi: 10.1002/1097-0045(20001001)45:2<173::AID-PROS12>3.0.CO;2-R

Costa, C. D., Justo, A. A., Kobayashi, P. E., Story, M. M., Palmieri, C., Laufer-Amorim, R., et al. (2019). Characterization of OCT3/4, Nestin, NANOG, CD44 and CD24 as stem cell markers in canine prostate cancer. *Int. J. Biochem. Cell Biol.* 108, 21–28. doi: 10.1016/j.biocel.2019.01.002

Cruijsen, M., Lübbert, M., Wijermans, P., and Huls, G. (2014). Clinical results of hypomethylating agents in AML treatment. *J. Clin. Med.* 4 (1), 1–17. doi: 10.3390/jcm4010001

da Costa-Prando, E., Cavalli, L. R., and Rainho, C. A. (2011). Evidence of epigenetic regulation of the tumor suppressor gene cluster flanking RASSF1 in breast cancer cell lines. *Epigenetics* 6 (12), 1413–1424. doi: 10.4161/epi.6.12.18271

Debelec-Butuner, B., Alapinar, C., Ertunc, N., Gonen-Korkmaz, C., Yörükoğlu, K., and Korkmaz, K. S. (2014). TNFα-mediated loss of β-catenin/E-cadherin association and subsequent increase in cell migration is partially restored by NKX3.1 expression in prostate cells. *PloS One* 9 (10), 1–12. doi: 10.1371/journal.pone.0109868

Dhar, S., Kumar, A., Gomez, C. R., Akhtar, I., Hancock, J. C., Lage, J. M., et al. (2017). MTA1-activated Epi-microRNA-22 regulates E-cadherin and prostate cancer invasiveness. *FEBS Lett.* 591 (6), 924–933. doi: 10.1002/1873-3468.12603

Fonseca-Alves, C. E., Rodrigues, M. M., de Moura, V. M., Rogatto, S. R., and Laufer-Amorim, R. (2013). Alterations of C-MYC, NKX3.1, and E-cadherin expression in canine prostate carcinogenesis. *Microsc. Res. Tech.* 76 (12), 1250–1256. doi: 10.1002/jemt.22292

Fonseca-Alves, C. E., Kobayashi, P. E., Rivera-Calderón, L. G., and Laufer-Amorim, R. (2015a). Evidence of epithelial-mesenchymal transition in canine prostate cancer metastasis. *Res. Vet. Sci.* 100, 176–181. doi: 10.1016/j.rvsc.2015.03.001

Fonseca-Alves, C. E., Bento, D. D., Torres-Neto, R., Werner, J., Kitchell, B., and Laufer-Amorim, R. (2015b). Ki67/KIT double immunohistochemical staining in cutaneous mast cell tumors from Boxer dogs. *Res. Vet. Sci.* 102, 122–126. doi: 10.1016/j.rvsc.2015.08.007

Fonseca-Alves, C. E., Kobayashi, P. E., and Laufer-Amorim, R. (2018a). Evaluation of NKX3.1 and C-MYC expression in canine prostatic cancer. *Res. Vet. Sci.* 118, 365–370. doi: 10.1016/j.rvsc.2018.04.001

Fonseca-Alves, C. E., Kobayashi, P. E., Rivera-Calderón, L. G., Felisbino, S. L., Rinaldi, J. C., Drigo, S. A., et al. (2018b). Immunohistochemical panel to characterize canine prostate carcinomas according to aberrant p63 expression. *PloS One* 13 (6), 1–16. doi: 10.1371/journal.pone.0199173

Friedl, P., and Wolf, K. (2003). Tumour-cell invasion and migration: diversity and escapemechanisms. *Nat. Rev. Cancer.* 3 (5), 362–374. doi: 10.1038/nrc1075

Graff, J. R., Herman, J. G., Lapidus, R. G., Chopra, H., Xu, R., Jarrard, D. F., et al. (1995). E-cadherin expression is silenced by DNA hypermethylation in human breast and prostate carcinomas. *Cancer Res.* 55 (22), 5195–5199.

Hong, S. M., Li, A., Olino, K., Wolfgang, C. L., Herman, J. M., Schulick, R. D., et al. (2011). Loss of E-cadherin expression and outcome among patientswith resectable pancreatic adenocarcinomas. *Mod. Pathol.* 24 (9), 1237–1247. doi: 10.1038/modpathol.2011.74

Hoque, M. O., Topaloglu, O., Begum, S., Henrique, R., Rosenbaum, E., Van Criekinge, W., et al. (2005). Quantitative methylation-specific polymerase chain reaction gene patterns in urine sediment distinguish prostate cancer patients from control subjects. *J. Clin. Oncol.* 23 (27), 6569–6575. doi: 10.1200/JCO.2005.07.009

Humphrey, P. A., Moch, H., Cubilla, A. L., Ulbright, T. M., and Reuter, V. E. (2016). The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part B: Prostate and Bladder Tumours. *Eur. Urol.* 17, 106–119. doi: 10.1016/j.eururo.2016.02.028

Huynh, M. A., Chen, M. H., Wu, J., Braccioforte, M. H., Moran, B. J., and D'Amico, A. V. (2016). Influence of comorbidity on the risk of mortality in men with unfavorable-risk prostate cancer undergoing high-dose radiation therapy alone. *Int. J. Radiat. Oncol. Biol. Phys.* 11, 1158–1167. doi: 10.1016/j.ijrobp.2016.03.004

Ipekci, T., Ozden, F., Unal, B., Saygin, C., Uzunaslan, D., and Ates, E. (2015). Epithelial-Mesenchymal transition markers β-catenin, snail, and E-Cadherin do not predict disease free survival in prostate adenocarcinoma: a prospective study. *Pathol. Oncol. Res.* 21 (4), 1209–1216. doi: 10.1007/s12253-015-9958-z

Kashiwagi, S., Yashiro, M., Takashima, T., Aomatsu, N., Ikeda, K., Ogawa, Y., et al. (2011). Advantages of adjuvant chemotherapy for patients with triple-negativebreast cancer at Stage II: usefulness of prognostic markers E-cadherin and Ki67. *Breast Cancer Res.* 13 (6), R122. doi: 10.1186/bcr3068

Kobayashi, P. E., Fonseca-Alves, C. E., Rivera-Calderón, L. G., Carvalho, M., Kuasne, H., Rogatto, S. R., et al. (2018). Deregulation of E-cadherin, β-catenin, APC and Caveolin-1 expression occurs in canine prostate cancer and metastatic processes. *Res. Vet. Sci.* 118, 254–261. doi: 10.1016/j.rvsc.2018.03.004

Laufer-Amorim, R., Fonseca-Alves, C. E., Villacis, R. A. R., Linde, S. A. D., Carvalho, M., Larsen, S. J., et al. (2019). Comprehensive genomic profiling of androgen-receptor-negative canine prostate cancer. *Int. J. Mol. Sci.* 28, 20(7. doi: 10.3390/ijms20071555

Li, L. C., Zhao, H., Nakajima, K., Oh, B. R., Ribeiro-Filho, L. A., Carroll, P., et al. (2001). Methylation of the E-cadherin gene promoter correlates with progression of prostate cancer. *J. Urol.* 166 (2), 705–709. doi: 10.1016/S0022-5347(05)66047-8

Li, F., Pascal, L. E., Stolz, D. B., Wang, K., Zhou, Y., Chen, W., et al. (2019). E-cadherin is downregulated in benign prostatic hyperplasia and required for tight junction formation and permeability barrier in the prostatic epithelial cell monolayer. *Prostate* 79 (11), 1226–1237. doi: 10.1002/pros.23806

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Maruyama, R., Toyooka, S., Toyooka, K. O., Virmani, A. K., Zöchbauer-Müller, S., Farinas, A. J., et al. (2002). Aberrant promoter methylation profile of prostate cancers and its relationship to clinicopathological features. *Clin. Cancer Res.* 8 (2), 514–519.

Mostafavi-Pour, Z., Kianpour, S., Dehghani, M., Mokarram, P., Torabinejad, S., and Monabati, A. (2015). Methylation of Integrin α4 and E-Cadherin genes in human prostate cancer. *Pathol. Oncol. Res.* 21 (4), 921–927. doi: 10.1007/s12253-015-9917-8

Palmieri, C., and Grieco, V. (2015). Proposal of Gleason-like grading system of canine prostate carcinoma in veterinary pathology practice. *Res. Vet. Sci.* 103, 11–15. doi: 10.1016/j.rvsc.2015.09.004

Putzke, A. P., Ventura, A. P., Bailey, A. M., Akture, C., Opoku-Ansah, J., Celiktaş, M., et al. (2011). Metastatic progression of prostate cancer and e-cadherin regulation by zeb1 and SRC family kinases. *Am. J. Pathol.* 179 (1), 400–410. doi: 10.1016/j.ajpath.2011.03.028

Ricciardi, G. R., Adamo, B., Ieni, A., Licata, L., Cardia, R., Ferraro, G., et al. (2015). Androgen Receptor (AR), E-Cadherin, and Ki-67 as emergingtargets and novel prognostic markers in triple-negative breast cancer (TNBC) patients. *PloS One* 10 (6), 1–11. doi: 10.1371/journal.pone.0128368

Rivera-Calderón, L. G., Fonseca-Alves, C. E., Kobayashi, P. E., Carvalho, M., Drigo, S. A., de Oliveira-Vasconcelos, R., et al. (2016). Alterations in PTEN, MDM2, TP53 and AR protein and gene expression are associated with canine prostate carcinogenesis. *Res. Vet. Sci.* 106, 56–61. doi: 10.1016/j.rvsc.2016.03.008

Rivera-Calderón, L. G., Fonseca-Alves, C. E., Kobayashi, P. E., Carvalho, M., Vasconcelos, R. O., and Laufer-Amorim, R. (2019). p-mTOR, p-4EBP-1 and eIF4E expression in canine prostatic carcinoma. *Res. Vet. Sci.* 122, 86–92. doi: 10.1016/j.rvsc.2018.11.006

Saramaki, O., and Visakorpi, T. (2007). Chromosomal aberrations in prostate cancer. *Front. Biosci.* 12, 3287–3301. doi: 10.2741/2312

Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer statistics, 2019. *CA Cancer J. Clin.* 69 (1), 7–34. doi: 10.3322/caac.21551

Singal, R., Ferdinand, L., Reis, I. M., and Schlesselman, J. J. (2004). Methylation of multiple genes in prostate cancer and the relationship with clinicopathological features of disease. *Oncol. Rep.* 12, 631–637. doi: 10.3892/or.12.3.631

Staník, M., Čapák, I., Macík, D., Vašina, J., Lžičařová, E., Jarkovský, J., et al. (2014). Sentinel lymph node dissection combined with meticuloushistology increases the detection rate of nodal metastases in prostate cancer. *Int. Urol. Nephrol.* 46 (8), 1543–1549. doi: 10.1007/s11255-014-0704-3

Tang, D., Xu, S., Zhang, Q., and Zhao, W. (2012). The expression and clinical significance of the androgen receptor and E-cadherin in triple-negative breast cancer. *Med. Oncol.* 29 (2), 526–533. doi: 10.1007/s12032-011-9948-2

Tsui, K. H., Lin, Y. H., Chung, L. C., Chuang, S. T., Feng, T. H., Chiang, K. C., et al. (2016). Prostate-derived ets factor represses tumorigenesis and modulates epithelial-to-mesenchymal transition in bladder carcinoma cells. *Cancer Lett.* 375 (1), 142–151. doi: 10.1016/j.canlet.2016.02.056

Umbas, R., Schalken, J. A., Aalders, T. W., Carter, B. S., Karthaus, H. F., Schaafsma, H. E., et al. (1992). Expression of the cellular adhesion molecule E-cadherinis reduced or absent in high-grade prostate cancer. *Cancer Res.* 52 (18), 5104–5109.

Umbas, R., Isaacs, W. B., Bringuier, P. P., Schaafsma, H. E., Karthaus, H. F., Oosterhof, G. O., et al. (1994). Decreased E-cadherin expression is associated with poor prognosis in patients with prostate cancer. *Cancer Res.* 54 (14), 3929–3933.

Wang, W., Wang, L., Mizokami, A., Shi, J., Zou, C., Dai, J., et al. (2017). Down-regulation of E-cadherin enhances prostate cancer chemoresistance via Notch signaling. *Chin. J. Cancer* 36 (1), 35. doi: 10.1186/s40880-017-0203-x

Wang, X., Dong, B., Zhang, K., Ji, Z., Cheng, C., Zhao, H., et al. (2018). E-cadherin bridges cell polarity and spindle orientation to ensure prostate epithelial integrity and prevent carcinogenesis in vivo. *PloS Genet.* 14 (8), 1–23. doi: 10.1371/journal.pgen.1007609

Welch, D. R. (2007). Do we need to redefine a cancer metastasis and staging definitions? *Breast Dis.* 26, 3–12. doi: 10.3233/BD-2007-26102

Yoshiura, K., Kanai, Y., Ochiai, A., Shimoyama, Y., Sugimura, T., and Hirohashi, S. (1995). Silencing of the E-cadherin invasion-suppressor gene by CpG methylation in human carcinomas. *Proc. Natl. Acad. Sci. U. S. A.* 92 (16), 7416–7419. doi: 10.1073/pnas.92.16.7416

Zhang, S. Q., Zhang, G. Q., and Zhang, L. (2016a). Correlation between methylation of the E-Cadherin gene and malignancy of prostate cancer. *Genet. Mol. Res.* 15 (2), 1–9. doi: 10.4238/gmr.15028046

Zhang, J., Chen, X. Y., Huang, K. J., Wu, W. D., Jiang, T., Cao, J., et al. (2016b). Expression of FoxM1 and the EMT-associated protein E-cadherin in gastric cancer and its clinical significance. *Oncol. Lett.* 12 (4), 2445–2450. doi: 10.3892/ol.2016.4917

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Fonseca-Alves, Kobayashi, Leis-Filho, Lainetti, Grieco, Kuasne, Rogatto and Laufer-Amorim. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

REVIEW
published: 10 January 2020
doi: 10.3389/fgene.2019.01247

The Current State of MicroRNAs as Restenosis Biomarkers

Nelson Varela[1], *Fernando Lanas*[2,3], *Luis A. Salazar*[3] *and Tomás Zambrano*[4]*

[1] Laboratory of Chemical Carcinogenesis and Pharmacogenetics, Department of Basic-Clinical Oncology, Faculty of Medicine, Universidad de Chile, Santiago, Chile, [2] Department of Internal Medicine, Faculty of Medicine, Universidad de La Frontera, Temuco, Chile, [3] Center of Molecular Biology and Pharmacogenetics, Scientific and Technological Bioresource Nucleus, Universidad de La Frontera, Temuco, Chile, [4] Department of Medical Technology, Faculty of Medicine, Universidad de Chile, Santiago, Chile

In-stent restenosis corresponds to the diameter reduction of coronary vessels following percutaneous coronary intervention (PCI), an invasive procedure in which a stent is deployed into the coronary arteries, producing profuse neointimal hyperplasia. The reasons for this process to occur still lack a clear answer, which is partly why it remains as a clinically significant problem. As a consequence, there is a vigorous need to identify useful non-invasive biomarkers to differentiate and follow-up subjects at risk of developing restenosis, and due to their extraordinary stability in several bodily fluids, microRNA research has received extensive attention to accomplish this task. This review depicts the current understanding, diagnostic potential and clinical challenges of microRNA molecules as possible blood-based restenosis biomarkers.

Keywords: epigenetics, microRNAs, in-stent restenosis, biomarkers, personalized and precision medicine

OPEN ACCESS

Edited by:
Rui Henrique,
Portuguese Oncology Institute, Portugal

Reviewed by:
Fouad Janat,
Independent researcher, Waverly, Rhode Island, United States
Girdhari Lal,
National Centre for Cell Science, India

***Correspondence:**
Tomás Zambrano
tomas.zambrano@uchile.cl

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *14 June 2019*
Accepted: *13 November 2019*
Published: *10 January 2020*

Citation:
Varela N, Lanas F, Salazar LA and Zambrano T (2020) The Current State of MicroRNAs as Restenosis Biomarkers. Front. Genet. 10:1247. doi: 10.3389/fgene.2019.01247

INTRODUCTION

Cardiovascular disease (CVD) refers to a group of pathologies initiated by an underlying process known as atherosclerosis and ultimately affecting the heart and blood vessels. Atherosclerosis plaques build-up inside the coronary arteries, consequently limiting the blood flow and resulting in coronary artery disease (CAD). Atherosclerosis is an inflammatory disease (Ross, 1999) able to produce two morphologically opposite lesions within the coronary arteries, stenotic and non-stenotic. The last may be asymptomatic for years and clinical management is generally supported on lifestyle modifications and, eventually, pharmacological interventions in high-risk individuals. In contrast, stenotic lesions have clinical manifestations like angina pectoris, and common medical management includes revascularization procedures such as coronary artery bypass grafting (CABG) and percutaneous transluminal coronary angioplasty (PTCA), a widely performed techniques since the late 1970s to correct serious coronary atherosclerotic lesions (Gruntzig et al., 1979), restoring myocardial blood flow and reducing angina symptoms. Despite its massive use, however, elevated restenosis rates affected almost half of the patients treated (Fischman et al., 1994) and established one of the main problems of current cardiology. Restenosis is arbitrarily defined as a narrowing of vessel diameter greater than 50% to that of the reference vessel (Marx et al., 2011), and results from excessive proliferation and migration of vascular smooth muscle cells (VSMC) to the intima, eventually leading to re-narrowing of the arterial lumen (Chaabane et al., 2013). Since this problem was identified, interventional cardiology has moved from PTCA to percutaneous coronary intervention (PCI), a technique involving the placement of a stent. This procedure is the most widely performed treatment for

symptomatic coronary disease patients (Serruys et al., 1994). The use of bare metal stents (BMS) has made it possible to eliminate factors that favor restenosis, such as elastic recoil and negative remodeling (Lowe et al., 2002), reducing the prevalence of restenosis from 50% to 20-30% (Kastrati et al., 1997). As the main cause of restenosis was attributed to the excessive proliferation of VSMC, the development of new technologies determined the arrival of drug-eluting stents (DES), reducing the restenosis rate below 10% (Morice et al., 2002). Despite the implementation of new stenting technologies, along with novel pharmacological or mechanical approaches to reduce restenosis incidence, this problem is still considered an important drawback, especially in high-risk patients, limiting the overall success of DES.

PATHOPHYSIOLOGY OF RESTENOSIS IN STENTED ARTERIES

Stent placement produces a mechanical vascular lesion that can be briefly divided into 3 phases:

a) Early phase: the stent produces an injury to the endothelium, damaging or totally destroying the endothelial cells (EC) that line the intimal arterial tunic, resulting in consecutive endothelial stripping, re-endothelization and subsequent generation of neo-endothelium (Grewe et al., 2000). The above is followed by an inflammatory response, including platelet activation and recruitment of circulating leukocytes, releasing cytokines and growth factors (Mitra and Agrawal, 2006).

b) Intermediate phase: characterized by the migration and proliferation of VSMC.

c) Late phase or tissue remodeling: VSMCs change from a contractile and quiescent non-proliferative G0 phase phenotype towards a highly active synthetic phenotype, with extracellular matrix (ECM) deposition in the arterial intima. Various growth factors, such as fibroblast growth factor (FGF-2), epidermal growth factor (EGF), platelet-derived growth factor (PDGF), and insulin-like growth factor (IGF) initiate VSMCs proliferation through the tyrosine kinase receptor, activating the Mitogen-activated protein kinases (MAPK) pathway. While the ECM allows the inflammatory infiltrate to adhere, the VSMC secrete hyaluronic acid and proteoglycans that interact and stabilize the fibrin-enriched ECM (Grewe et al., 2000; Mitra and Agrawal, 2006). These vascular responses, characterized by neointima proliferation and vascular remodeling, are responsible for the elevated frequency of post-PTCA restenosis. Anatomopathological studies in post-PCI restenosis demonstrated the same proliferative response of the neointima (Farb et al., 2002; Farb et al., 2004). In addition to VSMC proliferation and ECM synthesis, there is also neointima colonization by extravascular cells, e.g., endothelial progenitors or dendritic cells, together with compensatory mechanisms of apoptosis (Tuleta et al., 2008; Tuleta et al., 2010).

TYPES OF STENT: BARE METAL STENT AND DRUG-ELUTING STENT

Since the introduction of BMS in 1987 (Sigwart et al., 1987), important PTCA limitations such as restenosis and sudden narrowing of diseased arteries after angioplasty were reduced. Serruys et al. demonstrated that stent implantation reduces the need for a second coronary angioplasty compared with standard balloon angioplasty (RR 0.58, CI 0.40–0.85) mainly due to a low restenosis rate, going from 32% to 22% (Serruys et al., 1994). However, the benefit was accompanied by an increased risk of cardiovascular complications and longer hospitalization time. Stent-induced injury causes greater damage than damaged produced by standard balloon angioplasty, delineating processes of thrombosis, inflammation and proliferation (Edelman and Rogers, 1998) followed by the deposition of platelet-rich thrombi, which occur from the first days (Farb et al., 1999) until 1 month post-PCI (Komatsu et al., 1998), with additional accumulation of acute inflammatory cells such as neutrophils during the first 30 days, together with chronic inflammatory cells e.g., lymphocytes and macrophages (Farb et al., 1999). There is a correlation between the type of inflammatory reaction and the degree of injury (Rogers and Edelman, 1995), indicating that the surface of the material together with the geometric configuration of the stent contributes to neointimal hyperplasia and thrombosis. Other factors that favor restenosis development are tunica media damage, and the penetration of the stent edges into the lipid core of the atherosclerotic plaque. Both factors increase the inflammatory process within the artery and, therefore, increase intima proliferation (Farb et al., 2002). Since the introduction of combination therapy with P2Y platelet receptor antagonists (ticlopidine, clopidogrel) and acetylsalicylic acid, the incidence of post-stent thrombosis has been significantly reduced (Bertrand et al., 2000), and the majority of thrombotic events occurred within the first 10 days post-PCI. Additionally, post-stent thrombosis with BMS after the first month is considered rare (Farb et al., 2003). DES development is based on a coating containing an antiproliferative drug, so both post-PCI proliferation of the tunica intima and subsequent restenosis can be reduced (Garg and Serruys, 2010). In this way, first-generation DES devices were developed, which are

Abbreviations: AUC, area under the curve; AGO2, argonaute protein family 2; BMS, bare metal stent; CABG, coronary artery bypass grafting; CHD, coronary heart disease; CVD, cardiovascular disease; DES, drug-eluting stent; DGCR8, DiGeorge syndrome critical region 8; EC, endothelial cells; ECM, extracellular matrix; EDTA, ethylene diamine tetra-acetic acid; EGF, epidermal growth factor; FGF-2, fibroblast growth factor 2; IGF, insulin-like growth factor; ISR, in-stent restenosis; LEAOD, lower extremity arterial occlusive disease; LDL-C, low-density lipoprotein cholesterol; miRNA, micro ribonucleic acid; MAPK, mitogen-activated protein kinase; MIQE, minimum information for publication of quantitative real-time PCR experiments; mRNA, messenger ribonucleic acid; ncRNAs, non-coding RNAs; NGS, next-generation sequencing; PAD, peripheral artery disease; PCI, percutaneous coronary intervention; pre-miRNA, precursor micro ribonucleic acid; pri-miRNA, primary micro ribonucleic acid; PTCA, percutaneous transluminal coronary angioplasty; qPCR, quantitative polymerase chain reaction; RASP, rapid angiographic stenotic progression; RISC, RNA-induced silencing complex; ROC, receiver operating characteristic; Rnases, ribonucleases; snRNA, small nuclear RNAs; snoRNAs, small nucleolar RNAs; TLR, target lesion restenosis; TVR, target vessel revascularization; VSMC, vascular smooth muscle cells; XPO5, exportin 5; 3′-UTR, three prime untranslated region; 5′-UTR, five prime untranslated region.

coated with a drug-containing polymer designed to interrupt cell replication and reduce neointimal hyperplasia, markedly decreasing the occurrence of post-PCI restenosis (Herdeg et al., 2000; Curfman, 2002). However, after DES implantation, intervention centers have identified an increase in thrombosis associated with the stent placement process for up to 3 years after stent implantation, an additional complication rarely caused by the use of BMS (Luscher et al., 2007). Several reports show the occurrence of acute (<24 hours), sub-acute (<30 days), late (>30 days), and very late (>12 months) thrombosis after DES placement (McFadden et al., 2004; Pfisterer et al., 2006; Brodie et al., 2012). A large observational study revealed that from a total of 2229 consecutive patients receiving a total of 4495 DES, 29 had stent- associated thrombosis, occurring more than 30 days after stent placement, also revealing a 45% mortality rate (Iakovou et al., 2005).

MICRORNAS

In 1993, a key report involving the study of the *Caenorhabditis elegans* roundworm and showing downregulation of the LIN-14 protein by a small transcript namely lin-4 through antisense interaction between RNAs due to sequence complementarity between lin-4 and the 3'-untranslated region (3'-UTR) of the lin-14 mRNA (Lee et al., 1993) suggested a novel gene silencing mechanism affecting protein levels. Afterward, a second 21 nucleotides (nt) small RNA identified as let-7 was also implicated in the regulation of heterochronic genes related to *C. elegans* development (Reinhart et al., 2000). Moreover, the small let-7 RNA was shown to be highly conserved, indicating that its sequence is critical for functional purposes (Pasquinelli et al., 2000). These small RNAs were the 2 first of a family currently known as microRNAs (miRNAs) and further characterized as endogenous non-coding RNAs (ncRNAs) evolutionarily conserved between species with a size comprised between 20 and 23 nt, existing in both plants and animals. Their main function is to control gene expression by cleaving messenger RNA (mRNA) or through translational repression, preventing mRNA translation to its corresponding protein (Bartel, 2004). This control mechanism fine-tunes gene expression through the complementary matching of a segment comprised between nucleotides 2 to 7 of the miRNA i.e., seed region, with both the 3′- and 5′-UTR regions of target mRNAs (Lytle et al., 2007). It is estimated that miRNAs control more than 30% of the human genome (Lewis et al., 2005), through an interaction that can be reversible (Wu and Belasco, 2008). The canonical pathway of miRNA biogenesis begins with transcription from miRNA genes by RNA polymerase II, producing primary miRNAs (pri-miRNA) that undergo subsequent processing by the Drosha-DGCR8 (DiGeorge syndrome critical region 8) microprocessor complex, producing a miRNA precursor (pre-miRNA) of approximately 70 nt transported to the cytoplasm *via* exportin 5 (XPO5). Once in the cytoplasm, pre-miRNAs are further processed by RNase III (Dicer) into a double-stranded 21-23 nt miRNA. One strand of the miRNA is charged into the RNA-induced silencing complex (RISC) in conjunction with members of the Argonaute protein family (AGO2), a nuclear protein essential for miRNA maturation and functionality (Bushati and Cohen, 2007; Winter et al., 2009).

Different studies show that miRNAs orchestrate a wide network of cellular activities and are deeply involved in almost every biological pathway, regulating processes such as cell division and apoptosis (Ng et al., 2012), metabolism (Wilfred et al., 2007), intracellular signaling (Zhang et al., 2012), immune response (Taganov et al., 2006) and cell movement (Png et al., 2011). Similarly, miRNAs have been associated with restenosis-related processes, such as VSMC proliferation, migration and neointima formation (Chen et al., 2012; Yamakuchi, 2012; Gareri et al., 2016), revealing the great potential for diagnostic, prognostic, therapeutics or additional clinical manipulation. In fact, by examining the hypothesis that miRNAs produced by the placenta can be released into circulation, a set of placental miRNAs was successfully identified in maternal plasma (Chim et al., 2008), shedding light into another possible role as blood-based biomarkers, a crucial finding confirmed during the same year by a meticulous characterization of a large number of exceptionally stable miRNAs in both serum and plasma (Chen et al., 2008b). Since then, numerous reports have shown that miRNAs can be detectable in multiple fluids including urine, saliva, and cerebrospinal fluid, and even though the extracellular environment is rich in ribonucleases (RNases), miRNAs can be especially stable in serum and plasma as well, representing an enormous potential as non-invasive biomarkers for several pathologies (Gilad et al., 2008; Mitchell et al., 2008; Gupta et al., 2010). The mechanisms by which miRNAs remain particularly unaffected in circulation are due to their association with different carrier particles that confer protection against the potent blood RNases (**Figure 1**). It was first proposed that miRNAs circulate in the bloodstream by a cellular discharge mechanism through membrane-bound vesicles such as exosomes (Valadi et al., 2007; Kosaka et al., 2010), which are 50 to 100 nm vesicles released by exocytosis (Fevrier and Raposo, 2004). However, reports indicated that the abundant majority of miRNAs are exosome free and associated with Ago2 (Arroyo et al., 2011; Turchinovich et al., 2011). Importantly, as the RISC constitutes the effector component of the gene-silencing mechanism portrayed by miRNAs, it has been suggested that the miRNA-Ago2 complex is functional in circulation. Moreover, in 2011, Vickers et al. showed that miRNAs are associated with HDL in plasma not only for transport but these complexes maintain also the functional gene repression role of miRNAs directed to their cell target through delivery by a scavenger receptor BI (SR-BI)-dependent mechanism (Vickers et al., 2011). Considering that miRNAs are associated with dissimilar transport molecules, further classification of extracellular miRNAs according to their transporting molecules has been provided elsewhere (Russo et al., 2012).

BIOMARKER DISCOVERY

In general, biomarkers are classified as: (1) Diagnostic biomarkers for a specific pathology, disease or syndrome; (2) predictive biomarkers for the response to a given medication or treatment; (3) biomarkers of predictions about the probable

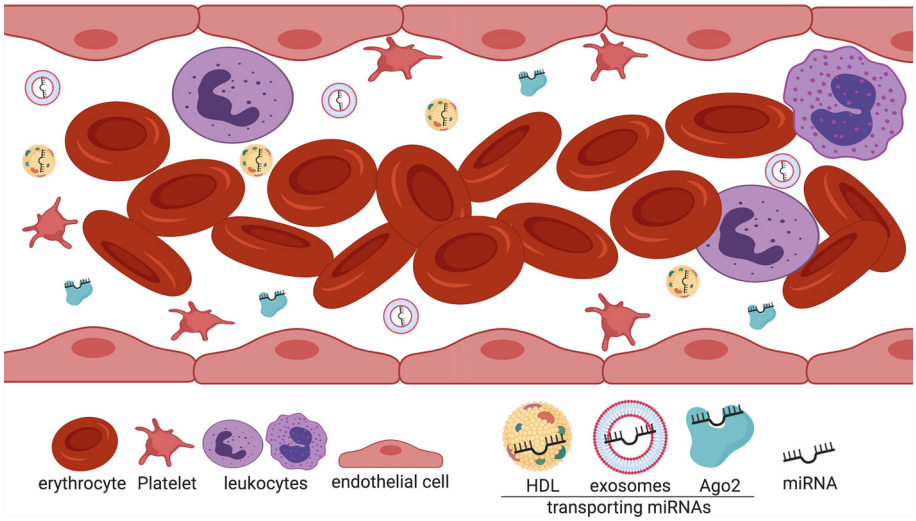


FIGURE 1 | Extracellular miRNAs are associated with different transport molecules in the bloodstream. Cell-free miRNAs are transported in the blood mainly associated with Ago2 (green) and HDL particles (yellow). To a lesser extent, miRNAs are transported *via* exosomes (light blue). Created with BioRender.

course of a disease; and (4) biomarkers of predisposition or susceptibility to a disease (Simon, 2011). According to the World Health Organization, a biomarker is "any substance, structure, process or products that can be measured in the body and influence or predict the outcome or incidence of the disease". Different study models have used cell lines, animals, patient cohorts, biopsies, biobank samples, or prospective studies as the starting point for biomarker development (Vargas and Harris, 2016). On the other hand, "omics" technologies are a particularly suitable tool for biomarkers discovery as they take advantage of the potential of the transcriptome, proteome and metabolome readings, facilitating detailed molecular characterization in a particular biological sample. Examples of these technologies are microarrays and next-generation sequencing (NGS) used for genomic and transcriptomic studies. In this sense, the usual strategy has been to describe large amounts of data from a specific molecule (e.g., miRNAs), in samples such as cell lines, animals and most importantly patients with a specific condition (McShane and Polley, 2013; Ghai and Wang, 2016) in order to generate hypotheses based on the large data available following bioinformatic analysis (Simon, 2010). Thus, new proposed biomarkers are capable, for example, of facilitating diagnosis of a certain disease or predicting the response of therapeutic interventions, such as post-stenting restenosis. Currently, a number of reasons have proposed circulating miRNAs as one of the most attractive candidates molecules to be explored as diagnosis, prognosis, and treatment biomarkers for various pathologies, mainly their extraordinary stability in blood circulation, the relative ease of extraction from the most common non-invasive matrices, and their susceptibility to sensitive detection through quantitative polymerase chain reaction (qPCR) (Gilad et al., 2008; Moldovan et al., 2014; Ghai and Wang, 2016) and rapid multiplexing platforms (Jiang et al., 2014).

CIRCULATING MIRNAS AS RESTENOSIS BIOMARKERS

Few investigations have examined the utility of cell-free miRNAs as potential in-stent restenosis (ISR) biomarkers. One of the pioneer reports was a case-control study revealing a series of 4 miRNAs -miRNA-21, miRNA-100, miRNA-143 and miRNA-145 - as candidate ISR markers, with the two latter showing the highest sensitivity and specificity according to receiver operating characteristic (ROC) curves (He et al., 2014) (**Table 1**). Consistent with their newfound role, these 4 miRNAs have been previously related to the pathogenesis of vascular diseases such as neointimal lesion formation (Ji et al., 2007), and VSMC proliferation, migration, and differentiation (Davis et al., 2008; Cordes et al., 2009; Grundmann et al., 2011; O'Sullivan et al., 2011). Interestingly, miRNA-21, miRNA-100, miRNA-143 and miRNA-145 were also able to significantly discriminate between diffuse vs. focal ISR, yet, this last finding should be interpreted with attention as it originated from additional analyses performed on a fraction of the total sample, probably introducing bias such as loss of randomization i.e. cases and controls are no longer balanced groups or lesser power related to the smaller sample. Another recent report showed that miRNA-93-5p was differentially expressed between ISR and non-ISR patients, proposing miRNA-93-5p as a robust independent ISR predictor (O'Sullivan et al., 2019). Additionally, they found

TABLE 1 | Studies reporting extracellular miRNAs as restenosis biomarkers.

miRNA	Performance	95% CI	P	Sens. (%)	Spec. (%)	Expression in restenosis patients	Groups	Grouping	Stent type	Follow-up (years)	ISR	Sample	Ethnicity	Ref
miRNA-21	AUC: 0.568	0.372-0.757	<0.05	50.1	68.6	UR	51 ISR; 130 NISR; 52 Controls	Randomization	DES	0.5 - 1	Coronary	Plasma	Chinese	He et al., 2014
miRNA-100	AUC: 0.608	0.372-0.757	<0.05	60.2	68.9	DR								
miRNA-143	AUC: 0.818	0.755-0.963	<0.001	82.1	80.1	DR								
miRNA-145	AUC: 0.880	0.791-0.987	<0.001	88.7	83.1	DR								
miRNA-93-5p	C: 0.734	not shown	0.0001	not shown	not shown	UR	39 ISR; 39 NISR	Consecutive, matched for age and diabetes	DES/ BMS	4.9	Coronary	Plasma	Caucasian	O'Sullivan et al., 2019
miRNA-93 + FHSRF + SL and SD	C: 0.769	0.00001												
miRNA-19a, miRNA-126, miRNA-210 miRNA-378	AUC: 0.776	0.722-0.831	not shown	not shown	not shown	DR	222 NISR; 64 ISR	Consecutive	DES	1	Coronary	Plasma	Chinese	Dai et al., 2019
miRNA-146a	AUC: 0674	0.567-0.781	not shown	not shown	not shown	UR	232 NISR; 23 ISR	Consecutive	DES	1	Coronary	Plasma	Chinese	Zhang et al., 2019
miRNA-146b	AUC: 0.801	0.729-0.875												
miRNA-92a (TLR)	HR: 0.55	0.34-0.88	0.013	not shown	not shown	DR	26 primary endpoint; 21 secondary endpoint; 36 controls	Consecutive	not shown	1 and 2	PAD	Serum	Caucasian	Stojkovic et al., 2018
miRNA-195 (TLR)	HR: 0.40	0.23-0.68	0.001											
miRNA-195 (TVR)	HR: 0.40	0.22-0.75	0.005											
miRNA-92a + CF (TLR)	C: 0.70	0.60-0.80	0.130											
miRNA-195 + CF (TLR)	C: 0.75	0.66-0.85	0.030											
miRNA-143	AUC: 0.866	not shown	not shown	83.7	82.6	DR	74 ISR; 91 NISR	Consecutive	BMS???	not shown	LEAOD	Serum? Plasma?	Chinese	Yu et al., 2017
miRNA-21	AUC: 0.938	0.898-0.977	not shown	83.5	98.2	DR	79 ISR; 327 NISR	Consecutive, randomly chosen	BMS	0.5	LEAOD	Plasma	Chinese	Zhang et al., 2017
miRNA-320a	AUC: 0.766	not shown	not shown	82.1	63.8	UR	78 ISR; 68 NISR; 62 controls	Consecutive	Not shown	not shown	LEAOD	Plasma	Chinese	Yuan et al., 2019
miRNA-572	AUC: 0.690	69.2	68.9											

AUC, area under the curve; BMS, bare metal stent; C, C-statistic (a comparable measure to AUC); CF, clinical factors; DR, down-regulated; DES, drug eluting stent; FHSRF, Framingham heart study risk factors; HR, hazard ratio; ISR, in-stent restenosis; LEAOD, lower extremity arterial occlusive disease; NISR, non in-stent restenosis; PAD, peripheral artery disease; Sens., sensibility; Spec., specificity; SL, stent length; SD, stent diameter; TLR, target lesion restenosis; TVR, target vessel revascularization; UR, up-regulated.

that the predictive performance of a model including main risk factors for ISR e.g., diabetes, stent length and diameter, together with common risk factors for CAD development such as age, sex, active smoking, diabetes, hypertension, and hyperlipidemia was further improved by adding miRNA-93-5p levels. Even though the results shown are encouraging, an important weakness of the study lies in the lack of additional validation in an independent cohort, restricting the extent of the results. Very recently, Dai et al. selected 14 angiogenesis-related candidate miRNAs (Dai et al., 2019) and reported 4 as independently associated with decreased restenosis risk (miRNA-19a, miRNA-126, miRNA-210, and miRNA-378). ROC curves showed that this subgroup of miRNAs had better predictive values for restenosis occurrence in Chinese population than each on its own (AUC: 0.776; 95% CI: 0.722-0.831). Moreover, they found that 2 additional miRNAs (let-7f and miR-296) correlated with a lower risk of rapid angiographic stenotic progression (RASP), and together with the previous miRNAs, the model exhibited greater performance for RASP prediction (AUC: 0.879; 95% CI: 0.841-0.917). Another similar study also performed in Chinese population recently reported that miRNA-146a and miRNA-146b were overexpressed in restenosis vs non-restenosis patients (P = 0.006), both holding prognostic value for restenosis risk in subjects with coronary heart disease (CHD) (Zhang et al., 2019). Analogously to the previous work, Zhang and colleagues also found that these miRNAs were up-regulated in RASP patients, and were both individually able to predict RASP occurrence in CHD subjects.

In the case of peripheral artery disease (PAD) ISR, the role of 11 restenosis-related circulating miRNAs (miRNA-17, miRNA-21, miRNA-92a, miRNA-126, miRNA-143, miRNA-145, miRNA-195, miRNA-221, miRNA-222, miRNA-223, and miRNA-424) was examined in a primary endpoint constituted by target lesion restenosis (TLR) and atherothrombotic events, and a secondary endpoint represented by target vessel revascularization (TVR) (Stojkovic et al., 2018). Findings showed that miRNA-92a and miRNA-195 were independent predictors of the primary endpoint, but only miRNA-195 was able to independently predict TVR. Interestingly, miRNA-143 and miRNA-145 were detected at very low expression levels and were excluded from additional analyses even though they were previously suggested as ISR markers (He et al., 2014). Nonetheless, and similarly to the report from O'Sullivan and colleagues, adding miRNA-195 to clinical factors not only improved the ability to distinguish TLR from non-TLR subjects against a model considering miRNA-92a (P = 0.012), but also proved superior to a model integrating clinical risk factors plus both miRNA-92a and miRNA-195 (Stojkovic et al., 2018) (**Table 1**).

A series of studies exploring the utility of predictive miRNAs for lower extremity arterial occlusive disease (LEAOD) restenosis have been performed. One of them identified low levels of circulating miRNA-143 in restenosis vs. non-restenosis patients, correlating this measure with smoking status, history of diabetes, glucose, and low-density lipoprotein cholesterol (LDL-C) (Yu et al., 2017). Even though a low expression of the restenosis-related miRNA-143 is consistent with the findings from He and colleagues, it is unknown if the expression pattern remains the same for the rest of the previously reported restenosis-associated miRNAs, as Yu et al. evaluated miRNA-143 only. Additionally, it is also uncertain if the analysis was performed either in serum or plasma due to authors referred to both biological fluids as interchangeable concepts, an unfortunate but substantial ambiguity hampering a clear interpretation of the results and that will be later described. Another study showed overexpression of the coronary ISR-associated miRNA-21 in LEAOD restenosis patients (Zhang et al., 2017), constituting an excellent predictor of vascular restenosis according to ROC curve analysis, with an AUC of 0.938. Moreover, miRNA-21 was correlated with age, diabetes, and hypertension, and together with diabetes, miRNA-21 represented the main risk factors for LEAOD restenosis occurrence. Lastly, a very recent report found that circulating levels of miRNA-320a and miRNA-572 were significantly overexpressed in restenosis-developing LEAOD patients (Yuan et al., 2019). ROC curves also showed that these miRNAs were capable of discerning between patients developing ISR versus patients that not, with AUC values of 0.766 and 0.690, respectively. Although the results provided are auspicious, the sample size enrolled was relatively small. Additionally, the study fails to report minimal but very relevant clinical data like the stent types used, follow-up time, and important statistical estimates (**Table 1**), however, one of the strengths lies in the inclusion of a second control group made up by healthy volunteers besides the classical non-ISR group, a similar methodological approach than the study of He et al., allowing to better discriminate miRNA behavior between these 2 conditions. In this sense, it is noteworthy that the relative expression among miRNAs evaluated by Yuan et al. was very similar between the non-ISR group and healthy volunteers. Furthermore, 2 miRNAs showed comparable levels between ISR, non-ISR and healthy volunteers, which is also consistent with previous studies, and represents an interesting outcome if we consider that current extracellular miRNA normalization is commonly based on the addition of exogenous spike-in miRNAs, a technical issue that could be more suitably replaced by analyzing endogenous miRNAs stable enough for the discovery of restenosis biomarkers. Still, additional experimentation is needed to clarify this observation.

Technical Challenges

Besides requiring a feasible and reliable analyte associated with a particular condition, miRNA routine sample analysis needs to cautiously overcome a significant amount of potentially detrimental obstacles that, if not properly managed, will not only affect miRNA analyses but most importantly, can compromise the patient's diagnosis and clinical management by inaccurate lab determinations. Consequently, before implementing miRNA measurement into day-to-day laboratory testing, a large number of technical issues must be correctly addressed. One of the very first concerns related to miRNA analysis comes from the collecting tubes employed for blood withdrawal. For instance, EDTA-collecting tubes can alter circulating miRNA detection, especially

if samples are not immediately processed. Moreover, the longer it takes for sample processing, the stronger the effect on extracellular miRNA patterns (Leidinger et al., 2015). Studies show that proper attention must be paid when selecting the type of biological matrix for further miRNA analysis. For instance, serum samples were reported to contain a higher number of miRNAs than their corresponding plasma counterparts, even if analyzing the same individual, an outcome highly dependent on the measurement platforms used (Wang et al., 2012). However, various reports argue in favor of the opposing scenario, where not only plasma was reported to contain higher miRNA concentrations (McDonald et al., 2011), but miRNAs diversity was far more restricted in serum samples (Foye et al., 2017). Findings have also shown that serum- and plasma-abundant miRNAs such as miRNA-451a, miRNA-16-5p, miRNA-223-3p, and miRNA-25-3p are differentially expressed between these 2 biological fluids (Foye et al., 2017), reinforcing the idea that both fluids cannot be assumed to be interchangeable concepts regarding miRNA concentrations. Also, hemolysis affects directly the concentration of a small number of miRNAs (Kirschner et al., 2011; McDonald et al., 2011) which is consistent with reports showing particular and specific erythrocytes-derived miRNAs (Chen et al., 2008a; Kannan and Atreya, 2010). Deepening in this area, hemolysis was demonstrated to affect a far greater number of miRNAs than previously reported, compromising miRNAs previously recognized as important biomarkers for various diseases (Kirschner et al., 2013). Interestingly, the use of a ratio between the hemolysis-dependent and -independent miRNA-451 and miRNA-23a, respectively, has been proposed to better assess the degree of erythrocyte lysis and therefore, diminish the effect of hemolysis on blood-based miRNA determinations (Blondal et al., 2013).

On the other hand, important variations for biomarker discovery can be introduced at the analytical stage, which is highly dependent on the measurement platform selected. In the case of miRNAs, the most common and widely used technique corresponds to qPCR largely due to its robustness, relative ease, elevated specificity and sensitivity, broad dynamic range and high resolution, among others. However, each step required for qPCR assays can introduce a different cause of variation that can mask the biological differences we are looking to determine, and to date, several unsolved questions can affect circulating miRNA analysis when using this system. For example, to date, there is no predetermined or consensus set of extracellular miRNAs that can be used for normalization (Roberts et al., 2014), which would be the ideal scenario to allow proper comparisons between a target miRNA against a normalizer miRNA to obtain reliable miRNA expression levels. In contrast, nowadays normalization is frequently achieved by using synthetic alternatives such as exogenous miRNAs that are spiked in during RNA isolation in an attempt to avoid technical differences regarding the extraction procedure. Another different normalization strategy is the use of ncRNAs such as small nuclear RNAs (snRNA) or small nucleolar RNAs (snoRNAs), however, to select the proper normalizer, a set of these ncRNAs must be previously analyzed in each lab for validation purposes to obtain accurate results. Importantly, RNA quality is one the most fundamental determinants of reproducibility for qPCR results, and improper sample handling regarding the collection, transport or storage can affect RNA integrity and unambiguously lead to irreproducible experiments. Therefore, every RNA preparation must be meticulously assessed to ensure that nucleic acids present have not been degraded. In general, a standardized qPCR protocol should be closely followed to ensure consistency between diverse laboratories, as suggested in the MIQE guidelines (minimum information for publication of quantitative real-time PCR experiments) (Bustin et al., 2009).

Future Perspectives

miRNA research as ISR biomarkers is still at an early stage. The scarce findings reported so far include conventional flaws in study designs such as small samples, lack of proper control groups or validation cohorts, and the absence of clinical, technical and statistical data that may be crucial for a correct interpretation and reproducibility. The tolerant operative consensus at the time of reporting putative restenosis biomarkers leads to inconsistent or unreliable candidate miRNAs, and to advance the field, investigations must meet minimal and uniform conditions. Clinical outcomes should be very well defined to turn them into quantifiable events, or at least easily measurable. In this sense, large randomized, multicenter, prospective trials capable of establishing whether miRNAs can effectively predict clinical features are greatly needed.

A persistent but reasonable shortcoming regarding miRNA research as biomarkers for restenosis is represented by the candidate approach, i.e., handpicking specific RNA molecules exclusively based on previous reports. Although most of these investigations have solid grounds since they are based on the choice of miRNAs previously associated with restenosis-related mechanisms, the success rates contrast with what one might expect, since they are not even close to 100%, as the case of different studies mentioned above (Stojkovic et al., 2018; Dai et al., 2019). The candidate approach is predominantly used because is significantly less expensive than other wide-ranging strategies, such as microarray or NGS, but the loss of information can be excessive. On the contrary, using, for example, NGS allows having the clearest depiction of the total amount of miRNAs that may be relevant or even participate in the endpoint and that we could be missing when executing the candidate methodology, which ultimately points to the cost-benefit relation.

Even though important technical difficulties can further delay the arrival of miRNAs into the clinic, ongoing research in the matter has allowed the most common problems to be properly identified and therefore, prone to correction with an adequate and strictly controlled standardization of laboratory practices, including pre-analytical, analytical and post-analytical procedures, eliminating as many possible variables affecting routine miRNA determinations. But even if we carefully consider the aforementioned arguments, the complete potential of miRNAs as

clinically applicable biomarkers is rather far from becoming an imminent reality, as some additional and significant questions remain poorly explored, for example, the existence of circadian oscillations of human miRNAs. A recent and very stimulating line of research has demonstrated important diurnal variations in miRNA levels (Rekker et al., 2015; Heegaard et al., 2016; Hicks et al., 2018), with important and detrimental implications for an exceedingly trivial preanalytical issue such as establishing the proper moment to collect blood samples.

AUTHOR CONTRIBUTIONS

NV wrote sections of the manuscript. TZ contributed conception, design and wrote sections of the manuscript. FL and LS contributed critical analysis and wrote sections of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

This work was supported by FONDECYT-Chile (grant number 3170785).

ACKNOWLEDGMENTS

We thank Isabel Castro Massó for her collaboration in obtaining the figure.

REFERENCES

Arroyo, J. D., Chevillet, J. R., Kroh, E. M., Ruf, I. K., Pritchard, C. C., Gibson, D. F., et al. (2011). Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc. Natl. Acad. Sci. U. S. A.* 108 (12), 5003–5008. doi: 10.1073/pnas.1019055108

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116 (2), 281–297. doi: 10.1016/s0092-8674(04)00045-5

Bertrand, M. E., Rupprecht, H. J., Urban, P., Gershlick, A. H., and Investigators, C. (2000). Double-blind study of the safety of clopidogrel with and without a loading dose in combination with aspirin compared with ticlopidine in combination with aspirin after coronary stenting: the clopidogrel aspirin stent international cooperative study (CLASSICS). *Circulation* 102 (6), 624–629. doi: 10.1161/01.cir.102.6.624

Blondal, T., Jensby Nielsen, S., Baker, A., Andreasen, D., Mouritzen, P., Wrang Teilum, M., et al. (2013). Assessing sample and miRNA profile quality in serum and plasma or other biofluids. *Methods* 59 (1), S1–S6. doi: 10.1016/j.ymeth.2012.09.015

Brodie, B., Pokharel, Y., Garg, A., Kissling, G., Hansen, C., Milks, S., et al. (2012). Predictors of early, late, and very late stent thrombosis after primary percutaneous coronary intervention with bare-metal and drug-eluting stents for ST-segment elevation myocardial infarction. *JACC Cardiovasc. Interv.* 5 (10), 1043–1051. doi: 10.1016/j.jcin.2012.06.013

Bushati, N., and Cohen, S. M. (2007). microRNA functions. *Annu. Rev. Cell Dev. Biol.* 23, 175–205. doi: 10.1146/annurev.cellbio.23.090506.123406

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., et al. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55 (4), 611–622. doi: 10.1373/clinchem.2008.112797

Chaabane, C., Otsuka, F., Virmani, R., and Bochaton-Piallat, M. L. (2013). Biological responses in stented arteries. *Cardiovasc. Res.* 99 (2), 353–363. doi: 10.1093/cvr/cvt115

Chen, S. Y., Wang, Y., Telen, M. J., and Chi, J. T. (2008a). The genomic analysis of erythrocyte microRNA expression in sickle cell diseases. *PloS One* 3 (6), e2360. doi: 10.1371/journal.pone.0002360

Chen, X., Ba, Y., Ma, L., Cai, X., Yin, Y., Wang, K., et al. (2008b). Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res.* 18 (10), 997–1006. doi: 10.1038/cr.2008.282

Chen, L. J., Lim, S. H., Yeh, Y. T., Lien, S. C., and Chiu, J. J. (2012). Roles of microRNAs in atherosclerosis and restenosis. *J. BioMed. Sci.* 19 (1), 79. doi: 10.1186/1423-0127-19-79

Chim, S. S., Shing, T. K., Hung, E. C., Leung, T. Y., Lau, T. K., Chiu, R. W., et al. (2008). Detection and characterization of placental microRNAs in maternal plasma. *Clin. Chem.* 54 (3), 482–490. doi: 10.1373/clinchem.2007.097972

Cordes, K. R., Sheehy, N. T., White, M. P., Berry, E. C., Morton, S. U., Muth, A. N., et al. (2009). miR-145 and miR-143 regulate smooth muscle cell fate and plasticity. *Nature* 460 (7256), 705–710. doi: 10.1038/nature08195

Curfman, G. D. (2002). Sirolimus-eluting coronary stents. *N. Engl. J. Med.* 346 (23), 1770–1771. doi: 10.1056/NEJM200206063462302

Dai, R., Liu, Y., Zhou, Y., Xiong, X., Zhou, W., Li, W., et al. (2019). Potential of circulating pro-angiogenic microRNA expressions as biomarkers for rapid angiographic stenotic progression and restenosis risks in coronary artery disease patients underwent percutaneous coronary intervention. *J. Clin. Lab. Anal.* e23013. doi: 10.1002/jcla.23013

Davis, B. N., Hilyard, A. C., Lagna, G., and Hata, A. (2008). SMAD proteins control DROSHA-mediated microRNA maturation. *Nature* 454 (7200), 56–61. doi: 10.1038/nature07086

Edelman, E. R., and Rogers, C. (1998). Pathobiologic responses to stenting. *Am. J. Cardiol.* 81 (7A), 4E–6E. doi: 10.1016/s0002-9149(98)00189-1

Farb, A., Sangiorgi, G., Carter, A. J., Walley, V. M., Edwards, W. D., Schwartz, R. S., et al. (1999). Pathology of acute and chronic coronary stenting in humans. *Circulation* 99 (1), 44–52. doi: 10.1161/01.cir.99.1.44

Farb, A., Weber, D. K., Kolodgie, F. D., Burke, A. P., and Virmani, R. (2002). Morphological predictors of restenosis after coronary stenting in humans. *Circulation* 105 (25), 2974–2980. doi: 10.1161/01.cir.0000019071.72887.bd

Farb, A., Burke, A. P., Kolodgie, F. D., and Virmani, R. (2003). Pathological mechanisms of fatal late coronary stent thrombosis in humans. *Circulation* 108 (14), 1701–1706. doi: 10.1161/01.CIR.0000091115.05480.B0

Farb, A., Kolodgie, F. D., Hwang, J. Y., Burke, A. P., Tefera, K., Weber, D. K., et al. (2004). Extracellular matrix changes in stented human coronary arteries. *Circulation* 110 (8), 940–947. doi: 10.1161/01.CIR.0000139337.56084.30

Fevrier, B., and Raposo, G. (2004). Exosomes: endosomal-derived vesicles shipping extracellular messages. *Curr. Opin. Cell Biol.* 16 (4), 415–421. doi: 10.1016/j.ceb.2004.06.003

Fischman, D. L., Leon, M. B., Baim, D. S., Schatz, R. A., Savage, M. P., Penn, I., et al. (1994). A randomized comparison of coronary-stent placement and balloon angioplasty in the treatment of coronary artery disease. Stent Restenosis Study Investigators. *N. Engl. J. Med.* 331 (8), 496–501. doi: 10.1056/NEJM199408253310802

Foye, C., Yan, I. K., David, W., Shukla, N., Habboush, Y., Chase, L., et al. (2017). Comparison of miRNA quantitation by Nanostring in serum and plasma samples. *PloS One* 12 (12), e0189165. doi: 10.1371/journal.pone.0189165

Gareri, C., De Rosa, S., and Indolfi, C. (2016). MicroRNAs for Restenosis and Thrombosis After Vascular Injury. *Circ. Res.* 118 (7), 1170–1184. doi: 10.1161/CIRCRESAHA.115.308237

Garg, S., and Serruys, P. W. (2010). Coronary stents: current status. *J. Am. Coll. Cardiol.* 56 (10 Suppl), S1–42. doi: 10.1016/j.jacc.2010.06.007

Ghai, V., and Wang, K. (2016). Recent progress toward the use of circulating microRNAs as clinical biomarkers. *Arch. Toxicol.* 90 (12), 2959–2978. doi: 10.1007/s00204-016-1828-2

Gilad, S., Meiri, E., Yogev, Y., Benjamin, S., Lebanony, D., Yerushalmi, N., et al. (2008). Serum microRNAs are promising novel biomarkers. *PloS One* 3 (9), e3148. doi: 10.1371/journal.pone.0003148

Grewe, P. H., Deneke, T., Machraoui, A., Barmeyer, J., and Muller, K. M. (2000). Acute and chronic tissue response to coronary stent implantation: pathologic findings in human specimen. *J. Am. Coll. Cardiol.* 35 (1), 157–163. doi: 10.1016/s0735-1097(99)00486-6

Grundmann, S., Hans, F. P., Kinniry, S., Heinke, J., Helbing, T., Bluhm, F., et al. (2011). MicroRNA-100 regulates neovascularization by suppression of mammalian target of rapamycin in endothelial and vascular smooth muscle cells. *Circulation* 123 (9), 999–1009. doi: 10.1161/CIRCULATIONAHA.110.000323

Gruntzig, A. R., Senning, A., and Siegenthaler, W. E. (1979). Nonoperative dilatation of coronary-artery stenosis: percutaneous transluminal coronary angioplasty. *N. Engl. J. Med.* 301 (2), 61–68. doi: 10.1056/NEJM197907123010201

Gupta, S. K., Bang, C., and Thum, T. (2010). Circulating microRNAs as biomarkers and potential paracrine mediators of cardiovascular disease. *Circ. Cardiovasc. Genet.* 3 (5), 484–488. doi: 10.1161/CIRCGENETICS.110.958363

He, M., Gong, Y., Shi, J., Pan, Z., Zou, H., Sun, D., et al. (2014). Plasma microRNAs as potential noninvasive biomarkers for in-stent restenosis. *PloS One* 9 (11), e112043. doi: 10.1371/journal.pone.0112043

Heegaard, N. H., Carlsen, A. L., Lilje, B., Ng, K. L., Ronne, M. E., Jorgensen, H. L., et al. (2016). Diurnal Variations of Human Circulating Cell-Free Micro-RNA. *PloS One* 11 (8), e0160577. doi: 10.1371/journal.pone.0160577

Herdeg, C., Oberhoff, M., Baumbach, A., Blattner, A., Axel, D. I., Schroder, S., et al. (2000). Local paclitaxel delivery for the prevention of restenosis: biological effects and efficacy *in vivo*. *J. Am. Coll. Cardiol.* 35 (7), 1969–1976. doi: 10.1016/s0735-1097(00)00614-8

Hicks, S. D., Khurana, N., Williams, J., Dowd Greene, C., Uhlig, R., and Middleton, F. A. (2018). Diurnal oscillations in human salivary microRNA and microbial transcription: Implications for human health and disease. *PloS One* 13 (7), e0198288. doi: 10.1371/journal.pone.0198288

Iakovou, I., Schmidt, T., Bonizzoni, E., Ge, L., Sangiorgi, G. M., Stankovic, G., et al. (2005). Incidence, predictors, and outcome of thrombosis after successful implantation of drug-eluting stents. *JAMA* 293 (17), 2126–2130. doi: 10.1001/jama.293.17.2126

Ji, R., Cheng, Y., Yue, J., Yang, J., Liu, X., Chen, H., et al. (2007). MicroRNA expression signature and antisense-mediated depletion reveal an essential role of MicroRNA in vascular neointimal lesion formation. *Circ. Res.* 100 (11), 1579–1588. doi: 10.1161/CIRCRESAHA.106.141986

Jiang, L., Shen, Y., Zheng, K., and Li, J. (2014). Rapid and multiplex microRNA detection on graphically encoded silica suspension array. *Biosens. Bioelectron* 61, 222–226. doi: 10.1016/j.bios.2014.05.020

Kannan, M., and Atreya, C. (2010). Differential profiling of human red blood cells during storage for 52 selected microRNAs. *Transfusion* 50 (7), 1581–1588. doi: 10.1111/j.1537-2995.2010.02585.x

Kastrati, A., Schomig, A., Elezi, S., Schuhlen, H., Dirschinger, J., Hadamitzky, M., et al. (1997). Predictive factors of restenosis after coronary stent placement. *J. Am. Coll. Cardiol.* 30 (6), 1428–1436. doi: 10.1016/s0735-1097(97)00334-3

Kirschner, M. B., Kao, S. C., Edelman, J. J., Armstrong, N. J., Vallely, M. P., van Zandwijk, N., et al. (2011). Haemolysis during sample preparation alters microRNA content of plasma. *PloS One* 6 (9), e24145. doi: 10.1371/journal.pone.0024145

Kirschner, M. B., Edelman, J. J., Kao, S. C., Vallely, M. P., van Zandwijk, N., and Reid, G. (2013). The Impact of Hemolysis on Cell-Free microRNA Biomarkers. *Front. Genet.* 4, 94. doi: 10.3389/fgene.2013.00094

Komatsu, R., Ueda, M., Naruko, T., Kojima, A., and Becker, A. E. (1998). Neointimal tissue response at sites of coronary stenting in humans: macroscopic, histological, and immunohistochemical analyses. *Circulation* 98 (3), 224–233. doi: 10.1161/01.cir.98.3.224

Kosaka, N., Iguchi, H., Yoshioka, Y., Takeshita, F., Matsuki, Y., and Ochiya, T. (2010). Secretory mechanisms and intercellular transfer of microRNAs in living cells. *J. Biol. Chem.* 285 (23), 17442–17452. doi: 10.1074/jbc.M110.107821

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75 (5), 843–854. doi: 10.1016/0092-8674(93)90529-y

Leidinger, P., Backes, C., Rheinheimer, S., Keller, A., and Meese, E. (2015). Towards Clinical Applications of Blood-Borne miRNA Signatures: The Influence of the Anticoagulant EDTA on miRNA Abundance. *PloS One* 10 (11), e0143321. doi: 10.1371/journal.pone.0143321

Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120 (1), 15–20. doi: 10.1016/j.cell.2004.12.035

Lowe, H. C., Oesterle, S. N., and Khachigian, L. M. (2002). Coronary in-stent restenosis: current status and future strategies. *J. Am. Coll. Cardiol.* 39 (2), 183–193. doi: 10.1016/s0735-1097(01)01742-9

Luscher, T. F., Steffel, J., Eberli, F. R., Joner, M., Nakazawa, G., Tanner, F. C., et al. (2007). Drug-eluting stent and coronary thrombosis: biological mechanisms and clinical implications. *Circulation* 115 (8), 1051–1058. doi: 10.1161/CIRCULATIONAHA.106.675934

Lytle, J. R., Yario, T. A., and Steitz, J. A. (2007). Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proc. Natl. Acad. Sci. U. S. A.* 104 (23), 9667–9672. doi: 10.1073/pnas.0703820104

Marx, S. O., Totary-Jain, H., and Marks, A. R. (2011). Vascular smooth muscle cell proliferation in restenosis. *Circ. Cardiovasc. Interv.* 4 (1), 104–111. doi: 10.1161/CIRCINTERVENTIONS.110.957332

McDonald, J. S., Milosevic, D., Reddi, H. V., Grebe, S. K., and Algeciras-Schimnich, A. (2011). Analysis of circulating microRNA: preanalytical and analytical challenges. *Clin. Chem.* 57 (6), 833–840. doi: 10.1373/clinchem.2010.157198

McFadden, E. P., Stabile, E., Regar, E., Cheneau, E., Ong, A. T., Kinnaird, T., et al. (2004). Late thrombosis in drug-eluting coronary stents after discontinuation of antiplatelet therapy. *Lancet* 364 (9444), 1519–1521. doi: 10.1016/S0140-6736(04)17275-9

McShane, L. M., and Polley, M. Y. (2013). Development of omics-based clinical tests for prognosis and therapy selection: the challenge of achieving statistical robustness and clinical utility. *Clin. Trials* 10 (5), 653–665. doi: 10.1177/1740774513499458

Mitchell, P. S., Parkin, R. K., Kroh, E. M., Fritz, B. R., Wyman, S. K., Pogosova-Agadjanyan, E. L., et al. (2008). Circulating microRNAs as stable blood-based markers for cancer detection. *Proc. Natl. Acad. Sci. U. S. A.* 105 (30), 10513–10518. doi: 10.1073/pnas.0804549105

Mitra, A. K., and Agrawal, D. K. (2006). In stent restenosis: bane of the stent era. *J. Clin. Pathol.* 59 (3), 232–239. doi: 10.1136/jcp.2005.025742

Moldovan, L., Batte, K. E., Trgovcich, J., Wisler, J., Marsh, C. B., and Piper, M. (2014). Methodological challenges in utilizing miRNAs as circulating biomarkers. *J. Cell Mol. Med.* 18 (3), 371–390. doi: 10.1111/jcmm.12236

Morice, M. C., Serruys, P. W., Sousa, J. E., Fajadet, J., Ban Hayashi, E., Perin, M., et al. (2002). A randomized comparison of a sirolimus-eluting stent with a standard stent for coronary revascularization. *N. Engl. J. Med.* 346 (23), 1773–1780. doi: 10.1056/NEJMoa012843

Ng, R., Song, G., Roll, G. R., Frandsen, N. M., and Willenbring, H. (2012). A microRNA-21 surge facilitates rapid cyclin D1 translation and cell cycle progression in mouse liver regeneration. *J. Clin. Invest.* 122 (3), 1097–1108. doi: 10.1172/JCI46039

O'Sullivan, J. F., Martin, K., and Caplice, N. M. (2011). Microribonucleic acids for prevention of plaque rupture and in-stent restenosis: "a finger in the dam". *J. Am. Coll. Cardiol.* 57 (4), 383–389. doi: 10.1016/j.jacc.2010.09.029

O'Sullivan, J. F., Neylon, A., Fahy, E. F., Yang, P., McGorrian, C., and Blake, G. J. (2019). MiR-93-5p is a novel predictor of coronary in-stent restenosis. *Heart Asia* 11 (1), e011134. doi: 10.1136/heartasia-2018-011134

Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., et al. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* 408 (6808), 86–89. doi: 10.1038/35040556

Pfisterer, M., Brunner-La Rocca, H. P., Buser, P. T., Rickenbacher, P., Hunziker, P., Mueller, C., et al. (2006). Late clinical events after clopidogrel discontinuation may limit the benefit of drug-eluting stents: an observational study of drug-eluting versus bare-metal stents. *J. Am. Coll. Cardiol.* 48 (12), 2584–2591. doi: 10.1016/j.jacc.2006.10.026

Png, K. J., Halberg, N., Yoshida, M., and Tavazoie, S. F. (2011). A microRNA regulon that mediates endothelial recruitment and metastasis by cancer cells. *Nature* 481 (7380), 190–194. doi: 10.1038/nature10661

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., et al. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* 403 (6772), 901–906. doi: 10.1038/35002607

Rekker, K., Saare, M., Roost, A. M., Kaart, T., Soritsa, D., Karro, H., et al. (2015). Circulating miR-200-family micro-RNAs have altered plasma levels in patients with endometriosis and vary with blood collection time. *Fertil. Steril.* 104938-946 (4), e932. doi: 10.1016/j.fertnstert.2015.06.029

Roberts, T. C., Coenen-Stass, A. M., and Wood, M. J. (2014). Assessment of RT-qPCR normalization strategies for accurate quantification of extracellular microRNAs in murine serum. *PloS One* 9 (2), e89237. doi: 10.1371/journal.pone.0089237

Rogers, C., and Edelman, E. R. (1995). Endovascular stent design dictates experimental restenosis and thrombosis. *Circulation* 91 (12), 2995–3001. doi: 10.1161/01.cir.91.12.2995

Ross, R. (1999). Atherosclerosis–an inflammatory disease. *N. Engl. J. Med.* 340 (2), 115–126. doi: 10.1056/NEJM199901143400207

Russo, F., Di Bella, S., Nigita, G., Macca, V., Lagana, A., Giugno, R., et al. (2012). miRandola: extracellular circulating microRNAs database. *PloS One* 7 (10), e47786. doi: 10.1371/journal.pone.0047786

Serruys, P. W., de Jaegere, P., Kiemeneij, F., Macaya, C., Rutsch, W., Heyndrickx, G., et al. (1994). A comparison of balloon-expandable-stent implantation with balloon angioplasty in patients with coronary artery disease. Benestent Study Group. *N. Engl. J. Med.* 331 (8), 489–495. doi: 10.1056/NEJM199408253310801

Sigwart, U., Puel, J., Mirkovitch, V., Joffre, F., and Kappenberger, L. (1987). Intravascular stents to prevent occlusion and restenosis after transluminal angioplasty. *N. Engl. J. Med.* 316 (12), 701–706. doi: 10.1056/NEJM198703193161201

Simon, R. (2010). Clinical trials for predictive medicine: new challenges and paradigms. *Clin. Trials* 7 (5), 516–524. doi: 10.1177/1740774510366454

Simon, R. (2011). Genomic biomarkers in predictive medicine: an interim analysis. *EMBO Mol. Med.* 3 (8), 429–435. doi: 10.1002/emmm.201100153

Stojkovic, S., Jurisic, M., Kopp, C. W., Koppensteiner, R., Huber, K., Wojta, J., et al. (2018). Circulating microRNAs identify patients at increased risk of in-stent restenosis after peripheral angioplasty with stent implantation. *Atherosclerosis* 269, 197–203. doi: 10.1016/j.atherosclerosis.2018.01.020

Taganov, K. D., Boldin, M. P., Chang, K. J., and Baltimore, D. (2006). NF-kappaB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses. *Proc. Natl. Acad. Sci. U.S.A.* 103 (33), 12481–12486. doi: 10.1073/pnas.0605298103

Tuleta, I., Skowasch, D., Peuster, M., Nickenig, G., and Bauriedel, G. (2008). Cells of primarily extravascular origin in neointima formation following stent implantation: coordinated expression of endothelial progenitor, dendritic and neural crest-derived cells. *Cardiology* 110 (3), 199–205. doi: 10.1159/000111930

Tuleta, I., Bauriedel, G., Steinmetz, M., Pabst, S., Peuster, M., Welsch, U., et al. (2010). Apoptosis-regulated survival of primarily extravascular cells in proliferative active poststent neointima. *Cardiovasc. Pathol.* 19 (6), 353–360. doi: 10.1016/j.carpath.2009.07.006

Turchinovich, A., Weiz, L., Langheinz, A., and Burwinkel, B. (2011). Characterization of extracellular circulating microRNA. *Nucleic Acids Res.* 39 (16), 7223–7233. doi: 10.1093/nar/gkr254

Valadi, H., Ekstrom, K., Bossios, A., Sjostrand, M., Lee, J. J., and Lotvall, J. O. (2007). Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat. Cell Biol.* 9 (6), 654–659. doi: 10.1038/ncb1596

Vargas, A. J., and Harris, C. C. (2016). Biomarker development in the precision medicine era: lung cancer as a case study. *Nat. Rev. Cancer* 16 (8), 525–537. doi: 10.1038/nrc.2016.56

Vickers, K. C., Palmisano, B. T., Shoucri, B. M., Shamburek, R. D., and Remaley, A. T. (2011). MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. *Nat. Cell Biol.* 13 (4), 423–433. doi: 10.1038/ncb2210

Wang, K., Yuan, Y., Cho, J. H., McClarty, S., Baxter, D., and Galas, D. J. (2012). Comparing the MicroRNA spectrum between serum and plasma. *PloS One* 7 (7), e41561. doi: 10.1371/journal.pone.0041561

Wilfred, B. R., Wang, W. X., and Nelson, P. T. (2007). Energizing miRNA research: a review of the role of miRNAs in lipid metabolism, with a prediction that miR-103/107 regulates human metabolic pathways. *Mol. Genet. Metab.* 91 (3), 209–217. doi: 10.1016/j.ymgme.2007.03.011

Winter, J., Jung, S., Keller, S., Gregory, R. I., and Diederichs, S. (2009). Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat. Cell Biol.* 11 (3), 228–234. doi: 10.1038/ncb0309-228

Wu, L., and Belasco, J. G. (2008). Let me count the ways: mechanisms of gene regulation by miRNAs and siRNAs. *Mol. Cell* 29 (1), 1–7. doi: 10.1016/j.molcel.2007.12.010

Yamakuchi, M. (2012). MicroRNAs in Vascular Biology. *Int. J. Vasc. Med.* 2012, 13. doi: 10.1155/2012/794898

Yu, Z. H., Wang, H. T., and Tu, C. (2017). Diagnostic value of microRNA-143 in predicting in-stent restenosis for patients with lower extremity arterial occlusive disease. *Eur. J. Med. Res.* 22 (1), 2. doi: 10.1186/s40001-016-0240-y

Yuan, L., Dong, J., Zhu, G., Bao, J., Lu, Q., Zhou, J., et al. (2019). Diagnostic Value of Circulating microRNAs for In-Stent Restenosis in Patients with Lower Extremity Arterial Occlusive Disease. *Sci. Rep.* 9 (1), 1402. doi: 10.1038/s41598-018-36295-2

Zhang, P., Bill, K., Liu, J., Young, E., Peng, T., Bolshakov, S., et al. (2012). MiR-155 is a liposarcoma oncogene that targets casein kinase-1alpha and enhances beta-catenin signaling. *Cancer Res.* 72 (7), 1751–1762. doi: 10.1158/0008-5472.CAN-11-3027

Zhang, B., Yao, Y., Sun, Q. F., Liu, S. Q., Jing, B., Yuan, C., et al. (2017). Circulating mircoRNA-21 as a predictor for vascular restenosis after interventional therapy in patients with lower extremity arterial occlusive disease. *Biosci. Rep.* 37 (2), 1–8. doi: 10.1042/BSR20160502

Zhang, H., Zhang, Q., Liu, Y., and Xue, T. (2019). miR-146a and miR-146b predict increased restenosis and rapid angiographic stenotic progression risk in coronary heart disease patients who underwent percutaneous coronary intervention. *Ir. J. Med. Sci.* 1–8. doi: 10.1007/s11845-019-02101-9

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Varela, Lanas, Salazar and Zambrano. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

REVIEW
published: 10 March 2020
doi: 10.3389/fgene.2020.00171

DNA Methylation Biomarkers in Aging and Age-Related Diseases

*Yasmeen Salameh†, Yosra Bejaoui† and Nady El Hajj**

College of Health and Life Sciences, Hamad Bin Khalifa University, Doha, Qatar

Recent research efforts provided compelling evidence of genome-wide DNA methylation alterations in aging and age-related disease. It is currently well established that DNA methylation biomarkers can determine biological age of any tissue across the entire human lifespan, even during development. There is growing evidence suggesting epigenetic age acceleration to be strongly linked to common diseases or occurring in response to various environmental factors. DNA methylation based clocks are proposed as biomarkers of early disease risk as well as predictors of life expectancy and mortality. In this review, we will summarize key advances in epigenetic clocks and their potential application in precision health. We will also provide an overview of progresses in epigenetic biomarker discovery in Alzheimer's, type 2 diabetes, and cardiovascular disease. Furthermore, we will highlight the importance of prospective study designs to identify and confirm epigenetic biomarkers of disease.

Keywords: aging, DNA methylation, epigenetic clocks, biomarkers, Alzheimer's disease, diabetes, cardiovascular diseases

OPEN ACCESS

Edited by:
Trygve Tollefsbol, University of Alabama at Birmingham, United States

Reviewed by:
Francesco Marabita, Karolinska Institutet (KI), Sweden
Simonetta Friso, University of Verona, Italy

***Correspondence:**
Nady El Hajj
nelhajj@hbku.edu.qa

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *25 June 2019*
Accepted: *13 February 2020*
Published: *10 March 2020*

Citation:
Salameh Y, Bejaoui Y and El Hajj N (2020) DNA Methylation Biomarkers in Aging and Age-Related Diseases. Front. Genet. 11:171. doi: 10.3389/fgene.2020.00171

INTRODUCTION

Aging is a complex and time-dependent deterioration of physiological process occurring in the majority of living organisms (Galloway, 1993). In humans, life expectancy has increased rapidly in the last few centuries due to a significant improvement in medical care and public health awareness (Crimmins, 2015). Consequently, increased life expectancy caused higher morbidity rates since advanced age is a predominant risk factor for several diseases including cancer, dementia, diabetes, and cardiovascular disease (CVD) (Jaul and Barron, 2017; Franceschi et al., 2018). Currently, there is an urgent need to improve health and longevity to increase not just the life span but also the health span of the elderly population. In recent years, several molecular and cellular processes have been reported to be linked to aging and contribute to its phenotype. Scientists proposed nine hallmarks of aging that can be classified into three categories: primary, antagonistic, or integrative (López-Otín et al., 2013). The primary hallmarks are defined as key factors causing cellular damage including genomic instability, telomere attrition, loss of proteostasis, and epigenetic alterations (López-Otín et al., 2013). During aging, there is a continuous accumulation of epigenetic changes, which might give rise to multiple age-related pathologies. A number of epidemiological studies revealed that monozygotic twins exhibit an increased rate of phenotypic discordance particularly for age-related diseases among older siblings (Frederiksen et al., 2002; Reynolds et al., 2005; Zwijnenburg et al., 2010; Greenwood et al., 2011; Castillo-Fernandez et al., 2014). This may be due to a gradual decrease in methylation conservation rates with successive cell divisions, a phenomenon referred to as "Epigenetic Drift" (Poulsen et al., 2007; Issa, 2014). This notion proposes an increased rate of stochastic methylation errors across the entire genome during aging. Indeed, several reports provided compelling evidence that older monozygotic twins exhibit global differences in DNA methylation (DNAm) patterns when compared to their younger counterparts

(Fraga et al., 2005; Lévesque et al., 2014; Tan et al., 2016; Wang et al., 2018). Similarly, a centenarian's methylome displays reduced DNA methylation levels as well as a decreased pair-wise correlation in the methylation status of neighboring CpG sites relative to the methylome of a newborn (Heyn et al., 2012).

In 1973, Vanyushin et al. (1973) were the first to describe global 5-methylcytosine (5mC) variations during aging in rats. Now, vast literature have revealed genome-wide DNA methylation changes that occur in response to aging across multiple species. These age-related epigenetic alterations either arise systemically or are restricted to a specific tissue/cell type. Age-related DNA methylation changes also take place in germ cells and might be possibly transmitted to the offspring (Atsem et al., 2016; Potabattula et al., 2018). Since the sequencing of the human genome the scientific community has been trying to elucidate how the genetic code controls the spatial and temporal expression of genes. The essence of DNA lies within the dynamic interaction between the genetic sequence (i.e. genome) and the epigenome. In many ways, environmental influences alter gene expression through various mechanisms such as DNA methylation, hydroxymethylation, histone modifications, alternative splicing, etc. (Edwards and Myers, 2007). Recent advances in "omics" technologies availed new avenues toward implementing precision medicine based on the genetic, environmental, and lifestyle factors of each individual. Similarly, treatments of complex diseases is demanding better diagnostic and screening tools for early detection particularly in the initial phase of the disease. DNA methylation (5-methylcytosine) is a covalent epigenetic modification to the DNA by addition of a methyl group to the C-5 position of the cytosine ring by DNA methyltransferases (Dnmts). Whereas, DNA hydroxymethylation (5-hydroxymethylcytosine) is a more recently discovered modification involving the addition of a hydroxymethyl group to the 5′ position of cytosine.DNA hydroxymethylation has been reported to be enriched in the brain especially in the proximity of synaptic genes (Kriaucionis and Heintz, 2009; Khare et al., 2012). The role 5-hmC plays in various biological processes remains elusive, nevertheless scientists are starting to appreciate its importance in gene expression regulation. Methylation and demethylation processes are not only important for transcription regulation but also play a crucial role during development and cell differentiation (Moore et al., 2013). Recently, DNA methylation measurements were shown to be valuable age prediction tools, even surpassing in accuracy the age prediction models based on telomere length (Horvath et al., 2016a). DNA methylation-based age prediction models are not only accurate in predicting chronological age but can also estimate biological aging rates (Chen et al., 2016; Christiansen et al., 2016).

EPIGENETIC-BASED AGING CLOCKS

It is only 6 years since Steve Horvath inaugurated a new era in epigenetics and aging research. In a landmark study, he developed a multivariate age predictor based on DNA methylation values of 353 individual CpG sites (Horvath, 2013). One of the main advantages of the Horvath clock is its ability to predict age systemically in all human cell types and tissues, excluding sperm. This is in contrast to other clocks that can be only applied to a single tissue (Hannum et al., 2013; **Figure 1**). Interestingly, the clock starts ticking early during development where fetal tissues as well as embryonic and induced pluripotent stem cells reveal a DNA methylation age (DNAm age) between −1 and 0 years (Horvath, 2013; Spiers et al., 2015). Till now, the biological mechanisms underlying changes measured by the epigenetic age clock have not been clearly identified. Therefore, recognizing genes that influence the rate of epigenetic aging might help determine such biological processes. Recent genome-wide association studies revealed tissue-specific association of variants in metabolism, immune system, aging, and autophagy -related genes with epigenetic age acceleration (Kananen et al., 2016; Lu et al., 2016, 2017, 2018). Epigenetic clocks have been also proposed to measure molecular processes involved in development and tissue homeostasis particularly those affecting stem cell differentiation as well as replenishment of committed cells (Horvath and Raj, 2018).

By regressing DNAm age on chronological age, epigenetic clocks can determine whether biological age acceleration occurs in certain diseases or in response to environmental factors (Horvath and Raj, 2018). Using this approach, age acceleration measurements in blood were associated with body mass index (BMI), obesity, physical fitness, Huntington's disease, Parkinson's disease, sleep, and smoking (Horvath et al., 2014; Horvath and Ritz, 2015; Horvath et al., 2016b; Carroll et al., 2017; Quach et al., 2017; Levine et al., 2018). Epigenetic clocks are highly valuable age prediction tools nevertheless their true value as diagnostic biomarkers requires further confirmation (**Figure 2**). Such biomarkers are epigenetic modifications/marks used as a risk assessment and diagnostic tool to uncover sequence of events preceding the manifestation of disease. Biomarkers can be measured within tissue or body fluid, in the context of disease vs health state, for the purpose of disease detection, disease prognosis, response to therapy, and therapy monitoring (García-Giménez et al., 2016).

Evidently, epigenetic clocks were employed to study epigenetic age acceleration in age-related disorders. For e.g. several reports showed DNAm age acceleration associated with incidence, future onset, and mortality across several types of cancer (Levine et al., 2015a; Zheng et al., 2016; Ambatipudi et al., 2017). Similarly, DNAm age was reported to be a useful biomarker for predicting physical and mental fitness in elderly individuals (Marioni et al., 2015) and was shown to be associated with cholesterol (High Density Lipoprotein: HDL), insulin, glucose, and triglycerides levels (Quach et al., 2017; Levine et al., 2018). The adult progeroid disease, Werner syndrome, which mimics aging at a faster rate, also revealed DNAm age acceleration of >6 years (Maierhofer et al., 2017). Recently, the Horvath lab developed the DNAm PhenoAge clock by training their predictor on phenotypic age rather than chronological age (Levine et al., 2018). The DNAm PhenoAge is a powerful biomarker for measuring health- and life- span that relies on measurements from 513 CpG sites (Levine et al., 2018). This clock could conclusively predict CVD incidence using whole

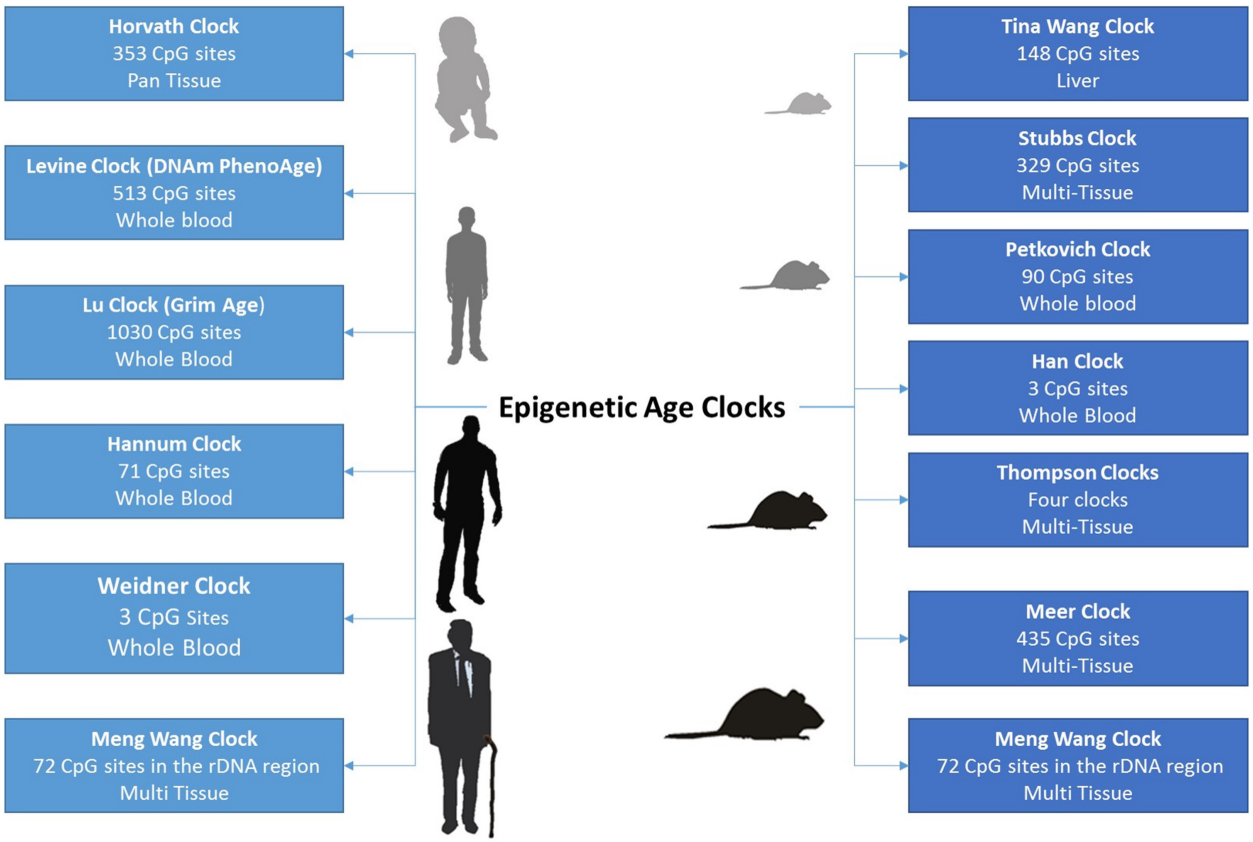


FIGURE 1 | The growing number of epigenetic age clocks developed for both humans and mice, including the number of CpG sites comprising the age-prediction model, as well as the tissues in which age can be estimated.

blood DNA methylation values. In 2019, the DNAm GrimAge clock was released where it was reported to predict mortality, cancer, and coronary heart disease (CHD) to a high level of accuracy (Lu et al., 2019). Epigenetic clocks that can estimate gestational age of neonates are also available (Knight et al., 2016). Using these clocks, we have demonstrated that DNAm age of children born via intracytoplasmic sperm injection (ICSI) lags half a week behind their naturally conceived counterparts (El Hajj et al., 2017).

In mice, epigenetic aging clocks were recently developed by relying on reduced representation bisulfite sequencing (RRBS) or whole genome bisulfite sequencing (WGBS) data (BI Ageing Clock Team et al., 2017; Petkovich et al., 2017; Wang et al., 2017; Meer et al., 2018; Thompson et al., 2018). These clocks provide useful biomarkers for measuring whether experimental interventions are able to slow the aging process in mice. Current research is focused on identifying evolutionary conserved pan-mammalian clocks that can calculate age across multiple species with varying lifespans. In addition, efforts are being invested in identifying clocks based on a handful of CpG sites since methylation arrays, RRBS, or WGBS remain relatively expensive compared to bisulfite pyrosequencing. In this aspect, Wolfgang Wagner's group has shown that measurements from just three CpG sites can accurately readout lifespan in both humans and mice (Weidner et al., 2014; Han et al., 2018). More recently, an epigenetic clock based on ribosomal DNA methylation was reported to be evolutionary conserved across several species (Wang and Lemos, 2019).

EPIGENETIC DYSREGULATION IN TYPE 2 DIABETES, ALZHEIMER'S DISEASE, AND CARDIOVASCULAR DISEASE

The dynamic change between methylation and demethylation states introduces flexibility to the rigidly stable DNA code, allowing controlled changes in gene expression in response to external and internal environmental cues. These moldable, yet generally stable processes are becoming valuable tools for distinguishing healthy versus diseased states. In cancer, despite the genome-wide hypomethylation, CpG islands are hypermethylated and can serve as a biomarker for early cancer detection (Anglim et al., 2008). Recent studies have shown that changes in global content of 5mC and 5hmC are not only useful as early detection tools but also a valuable source for understanding the underlying mechanisms of cancer development and patient

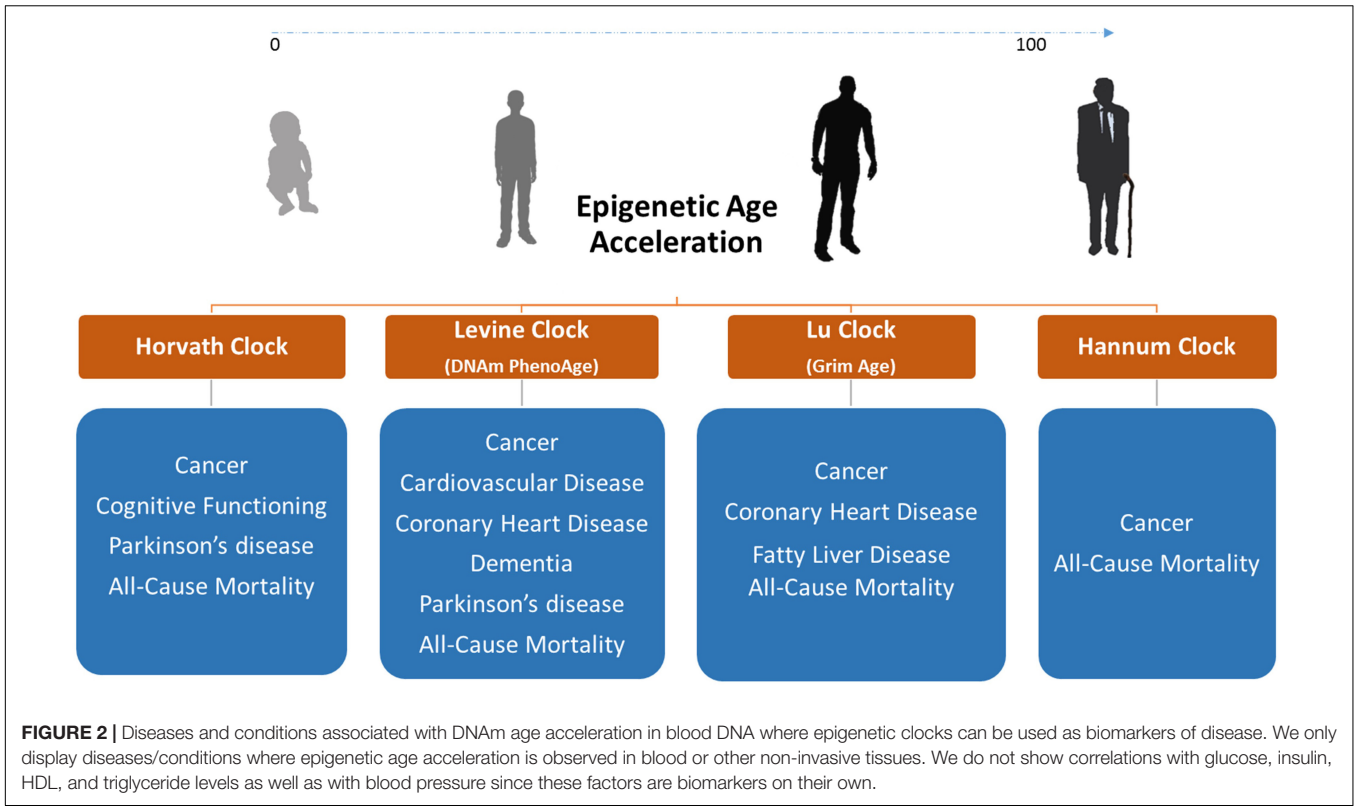


FIGURE 2 | Diseases and conditions associated with DNAm age acceleration in blood DNA where epigenetic clocks can be used as biomarkers of disease. We only display diseases/conditions where epigenetic age acceleration is observed in blood or other non-invasive tissues. We do not show correlations with glucose, insulin, HDL, and triglyceride levels as well as with blood pressure since these factors are biomarkers on their own.

prognosis (Liu et al., 2019). There are several published reviews discussing epigenetic biomarkers in cancer, however, this review's main focus will be on DNA methylation biomarkers in type 2 diabetes (T2D), Alzheimer's disease (AD), and cardiovascular disease (CVD) (**Supplementary Table S1**). Here, it is important to mention that these biomarkers are independent of the epigenetic clock described in the previous section.

Type 2 Diabetes

According to the World Health Organization (WHO), >420 million adults suffer from diabetes where 1.6 million deaths per year are directly attributed to the disease (Chan, 2014). The increased lifespan in humans is one of the main contributors to the rising prevalence of diabetes in the older population. Currently, more than third the United States population above the age of 65 are diabetics with numbers projected to increase in the next decade. Type 2 Diabetes (T2D) is a metabolic disorder characterized by abnormally elevated blood glucose levels due to β-cells dysfunction and insulin resistance (Chatterjee et al., 2017). T2D is a complex multifactorial disease where a variety of genetic, epigenetic, and environmental factors contribute to its etiology (McCarthy, 2010). Common complications of diabetes include cardiovascular problems, neuropathy, nephropathy, and retinopathy due to high blood glucose levels (Jacobs et al., 2017). Therefore, prevention or early treatment are very important to prevent damage to several of the body's systems. Despite the availability of well-established measures for diagnosing diabetes such as hemoglobin A1c (HbA1c) and fasting glucose, additional DNA-methylation based biomarkers can help complement current tests for screening and diagnosis. Identifying an individual during the pre-diabetic stage is very important for the management of the disease since ∼70% of persons with intermediate hyperglycaemia tend to develop T2D later in life.

Recently, efforts have focused on defining epigenetic risk factors associated with T2D as well as its major risk factors. Published reports have identified DNAm alterations in various tissues of T2D patients including blood, liver, pancreas, skeletal muscle, and adipose tissue (Ling and Rönn, 2019). These studies employed different approaches to quantify methylation changes including candidate gene analysis, global 5mC measurements, DNA methylation arrays, as well as WGBS (Volkov et al., 2017; Ling and Rönn, 2019). Evidently, the first reports describing epigenetic dyrsegulation in skeletal muscle and pancreatic islets of T2D patients applied a candidate gene approach. These studies identified increased DNA methylation and reduced gene expression in T2D-related genes such as *INS*, *PDX1*, *PPARGC1A*, and *GLP1R* (Ling et al., 2008; Barrès et al., 2009; Yang et al., 2012; Hall et al., 2013). Similarly, bisulfite pyrosequencing and methylation-specific PCR were employed to study methylation of key T2D genes in blood DNA. Investigated genes included *KCNJ11, PPARgamma, PDK4, KCNQ1, PDX1, FTO, PEG3, TCF7L2, GCK, PRKCZ, BCL11A,GIPR, SLC30A8, IGFBP-7, PTPPN1, CAMK1D, CRY2, CALM2, TLR2, TLR4,* and *FFAR3* [reviewed in Willmer et al. (2018)]. Most of those studies suffered from low sample size apart of a report by Seman et al. (2015), which quantified methylation in the solute carrier family 30 member 8 *(SLC30A8)*. Here, the authors detected

hypermethylation at several CpG sites in *SLC30A8* in 516 T2D subjects vs 476 individuals with normal glucose tolerance (Seman et al., 2015). Global changes in DNAm levels were also investigated using bisulfite pyrosequencing of *ALU* and *LINE-1* elements, liquid chromatography mass spectrometry, Imprint Methylated DNA Quantification kit (Sigma-Aldrich), and High Performance Liquid Chromatography (HPLC). Conflicting results were reported which might be inherently related to low sample size and lack of replication in independent cohorts [reviewed in Willmer et al. (2018)].

The development of Infinium Methylation arrays and NGS-based methylation sequencing allowed simultaneous quantification of methylation at thousands of CpG sites. Several case-control array studies compared DNA methylation abnormalities in pancreatic islets, liver, and subcutaneous adipose tissue of T2D patients. The focus of this review is on methylation-based biomarkers therefore we will mainly describe changes reported in blood or other accessible tissues. One impressive example of such alterations is the occurrence of dynamic DNA methylation changes in Peripheral Blood Mononuclear Cells (PBMCs) ∼80–90 days prior to elevated glucose levels. This was observed by Chen et al. (2018) after longitudinally following a healthy individual over the course of 3 years while measuring DNA methylation levels using WGBS at 28 selected time-points. Another study by Toperoff et al. (2012) used a pooling-based methylation screen followed by individual-level replication in a prospective cohort to identify CpGs that can predict future T2D risk. The authors reported a single CpG site in the first intron of the fat mass and obesity-associated (FTO) gene to be hypomethylated prior to the appearance of T2D (Toperoff et al., 2012). DNA methylation alterations were also measured in concordant and discordant monozygotic twins for T2D using genome-wide methylated DNA immunoprecipitation sequencing (MeDIP-seq). This elegantly designed study uncovered differentially methylated regions (DMRs) located in the promoters of *MALT1* and *GPR61* (Yuan et al., 2014).

In addition to age, BMI is a major risk factor contributing to T2D and has been the focus of multiple epigenome-wide association studies (EWAS) studies. A large study on >10,000 samples identified DNA methylation changes across 187 loci correlating with high BMI levels. Out of the 187 "sentinel obesity biomarkers," 62 loci were associated with T2D incidence including a probe in *ABCG1* with the strongest significance. A methylation risk score based on the sum of these markers exhibited a higher predictive power of future T2D onset when compared to traditional risk factors such as obesity, fasting glucose, and hyperinsulinemia (Wahl et al., 2017). Similarly, a longitudinal follow-up study on Indian Asians and Europeans discovered five T2D methylation markers in whole blood DNA collected at baseline prior to diabetes onset. These markers located in *ABCG1*, *PHOSPHO1*, *SOCS3*, *SREBF1*, and *TXNIP* were associated with metabolic measures of insulin resistance including glucose concentration, BMI, waist-to-hip ratio, and homeostatic model assessment for insulin resistance (HOMA-IR) (Chambers et al., 2015). A conceptually related study tried to replicate the association between T2D and the five previously mentioned genes in subjects from the Botnia prospective cohort. Nonetheless, they could only confirm *ABCG1* and *PHOSPHO1* methylation as predictors of future T2D risk (Dayeh et al., 2016). This association was also observed in healthy individuals where *ABCG1* methylation was reported to correlate with fasting insulin and HOMA-IR (Hidalgo et al., 2014).

Further EWAS studies could confirm methylation aberrations in some of the previously mentioned genes. A large EWAS analysis in Mexican-American individuals unraveled five CpG sites linked to T2D-related traits out of which 3 were located in *TXNIP (cg19693031)*, *ABCG1*, and *SAMD12* (Kulkarni et al., 2015). Two separate studies from Spain and Germany confirmed the association between decreasing methylation levels at *TXNIP (cg19693031)* and T2D, as well as with fasting glucose and HbA1c concentrations (Florath et al., 2016; Soriano-Tárraga et al., 2016). To end with EWAS, it is important to mention a meta-analysis by Walaszczyk et al. (2018) that took the initiative to confirm potential glycemic trait and T2D biomarkers. In this replication analyses, the authors concluded that a significant association between T2D and methylation sites in *ABCG1*, *TXNIP*, and *SREBF1* exists, which makes them promising biomarkers for early T2D detection. As a final point, we have to emphasize the significance of non-genetic elements including blood sugar levels, patient age, BMI, and gender in predicting future diabetes risk. Thus, such factors should be integrated into a T2D predictive model that includes genetic and epigenetic biomarkers to improve early T2D detection and allow better disease prognosis.

Alzheimer's Disease

Accumulation of errors in the epigenetic machinery during aging progression increases the risk for onset of age-related pathologies, such of those involving brain deterioration and neurodegeneration. The most common brain disorders affecting elderly individuals are those causing dementia through loss of synaptic plasticity, leading to memory impairment and defective learning capabilities. Alzheimer's disease (AD) affects 45–60% of the population with dementia and its burden is expected to double by the year 2060 (Finder, 2010; Duong et al., 2017). AD is a polygenic, complex and age-related neurodegenerative disease clinically characterized by progressive memory loss and cognitive impairment. Its pathological features include accumulation of β-amyloid (Aβ) in senile plaques, the formation of neurofibrillary tangles (NFTs) composed of hyperphosphorylated protein tau, and massive neuronal loss mainly in the hippocampus as well as associated regions of the neocortex (Hardy, 2006). Several clinical and epidemiological aspects of AD indicate a role for epigenetic factors in its etiology. This is evident in monozygotic twins discordant for the disease where prognosis and age-of onset could vary by >10 years. Indeed, a broad spectrum of epigenetic pathways such as DNA methylation, histone modification, and non-coding RNAs (ncRNAs) appear to be aberrant. For e.g. Wang et al. (2008) reported that Alzheimer's susceptibility loci have an age-specific epigenetic drift in brain and blood of individuals with late-onset AD. Several studies were conducted to identify epigenetic aberrations, as well as to differentiate specific methylation changes occurring in AD vs non-AD dementias [reviewed in: Lardenoije et al. (2015)]. Using southern blot

analysis, West et al. (1995) first showed loss of methylation at a single site in the amyloid precursor protein (APP) gene in postmortem human brain of a single individual with AD. This was confirmed by Tohgi et al. (1999) who reported that hypomethylation of cytosine residues within the APP promoter with age results in Aβ deposition in the cerebral cortex of human autopsy brain samples. Nevertheless, new studies using bisulfite sequencing failed to replicate these findings (Brohede et al., 2010). Recently, neuronal fractions from postmortem brains of Alzheimer's patients were reported to display significantly up-regulated expression of *BRCA1*, consistent with hypomethylation of a CpG island (CGI) in its promoter region. BRCA1 protein levels were also increased in response to Aβ deposition and became mislocalized to the cytoplasm, in both *in vitro* cellular and *in vivo* mouse models (Mano et al., 2017).

After the introduction of methylation arrays, a large study on >700 autopsied brain samples revealed methylation and expression changes in *ANK1*, *CDH23*, *DIP2A*, *RHBDF2*, *RPL13*, *SERPINF1*, and *SERPINF2* (De Jager et al., 2014). Similarly, Lunnon and collaborators performed a large EWAS analysis on four brain regions where they reported a significant hypermethylation of *ANK1* in the entorhinal cortex, superior temporal gyrus, and prefrontal cortex of AD individuals. The authors went on to measure methylation in pre-mortem blood DNA where they identified distinct differentially methylated probes (DMPs) to those in AD brains (Lunnon et al., 2014). The top ranked AD-associated blood DMPs were located in *DAPK1*, *GAS1*, and *NDUFS5*. Furthermore, epigenetic age acceleration was shown to be associated with AD neuropathological markers such as neuritic plaques, diffuse plaques, and amyloid load in the dorsolateral prefrontal cortex (Levine et al., 2015b). Down's syndrome patients, predisposed to early onset AD, also display DNAm age acceleration in blood and brain tissue starting early during *in utero* development (Horvath et al., 2015; El Hajj et al., 2016) in addition to epigenetic dysregulation at the clustered protocadherin locus (Almenar-Queralt et al., 2019).

Presently, a definitive AD diagnosis is only possible through neuropathological examination of brain tissue after death. Therefore, it is important to identify clinical biomarkers that can help in early disease detection. In addition, the effectiveness of available FDA-approved treatments for AD increases when administered during early stages of the disease. Currently, ongoing research efforts are mainly focused on delineating AD-related epigenetic changes that occur in various brain regions. So far, only a limited number of studies have assessed DNA methylation changes in blood cells. These articles will be the subject of the next section, where we will first summarize findings observed using a candidate gene approach. In one of these studies, blood DNA methylation of the Brain-derived neurotrophic factor gene (*BDNF*) promoter and a tag SNP (rs6265) were shown to have a significant role in the progression of the amnestic mild cognitive impairment (aMCI) to AD. Here, the interaction between DNA methylation of CpG5 and AA genotype of rs6265 had a role in the progression of aMCI to AD (p = 0.003, OR = 1.399, 95% CI: 1.198–1.477) (Xie et al., 2017a). A 5-year longitudinal study also revealed *BDNF* promoter methylation as a significant independent predictor of aMCI to AD transformation (Xie et al., 2017b). Similarly, Nagata et al. (2015) reported higher DNA methylation affecting a single CpG site in the *BDNF* promoter of patients with AD. Nevertheless, it is important to note that Carboni et al. (2015) could not confirm methylation alterations in the *BDNF* promoter in peripheral blood of Alzheimer's disease patients. Therefore, doubts remain as to whether *BDNF* promoter methylation changes occur in AD patients. Besides, DNA methylation levels were demonstrated to be significantly elevated in Coenzyme A Synthase (*COASY*) and Serine Peptidase Inhibitor (*SPINT1*) gene promoter regions in AD and aMCI (Kobayashi et al., 2016). DNA methylation at the *NCAPH2/LMF2* promoter region was also found to be a useful biomarker for the diagnosis of AD and aMCI where it was shown to be associated with hippocampal atrophy through apoptosis (Shinagawa et al., 2016). Furthermore, Ozaki et al. (2017) could show that a decline in DNA methylation in intron 1 of Triggering receptor expressed on myeloid cells 2 gene (*TREM2*) causes higher mRNA expression in the leukocytes of AD subjects versus controls. Phosphatidylinositol Binding Clathrin Assembly Protein (*PICALM*) was another candidate gene whose methylation associated with cognitive decline in blood cells of AD patients (Mercorio et al., 2018). Higher global DNA methylation levels were also observed in the peripheral blood mononuclear cells of late onset Alzheimer disease (LOAD) patients. This hypermethylation was associated with APOEε4 allele (p = 0.0043) and APOEε3 carriers (p = 0.05) (Di Francesco et al., 2015). In the same way, Bollati et al. (2011) observed a hypermethylation of *LINE-1* elements in AD patients after measuring DNA methylation at *ALU*, *LINE-1*, and alpha satellite repetitive elements.

In AD, epigenome-wide association studies (EWAS) on prospective cohorts are still lacking. To address this limitation, the German Study on Aging, Cognition and Dementia in Primary Care Patients (AgeCoDe) recruited >3300 healthy individuals at baseline to investigate markers for early detection of dementia and cognitive impairment. From this cohort, Lardenoije et al. (2019) identified 55 converters healthy at baseline that developed AD dementia at follow-up. Using DNA methylation arrays, several differentially methylated regions were spotted in blood of AD converters at baseline. By focusing on those regions, we could discern epigenetic dysregulation at six DMPs in blood DNA of Down's syndrome patients who are at high risk of developing early onset AD. One of the DMPs mapped to *ADAM10*, a major alpha-secretase, responsible for APP cleavage in neurons (Haertle et al., 2019). It is still challenging to find a non-invasive biomarker that reflects AD pathogenesis in the brain. Nonetheless, the previously described epigenetic alterations might be considered potential biomarkers that require further research to assess their efficacy.

Cardiovascular Disease

Cardiovascular disease (CVD) is an umbrella term for a range of conditions that affect the heart or blood vessels. The main determinants of a person's cardiovascular health is age, as well as several risk factors including diabetes, smoking, obesity, and high blood pressure. Epigenetic aging biomarkers based on "The Horvath Clock," "DNAm PhenoAge," and "DNAm

GrimAge" were recently reported to be associated with CVD risk (Levine et al., 2018; Lind et al., 2018; Lu et al., 2019). Even though not much research is published on the epigenetics of CVD, however, the impact of epigenetics has been extensively studied in the aforementioned risk factors. The complex interplay of genetics, epigenetics, and environment have an important role in the pathogenesis and progress of these conditions. For e.g. a *trans*-ancestry genome wide association study (GWAS) identified 12 genetic variants associated with methylation levels, which influences susceptibility for hypertension (Kato et al., 2015). Similarly, elevated global DNA methylation levels were reported to be positively associated with CVD and its predisposing risk factors (Sharma et al., 2008; Kim et al., 2010). Another example by Infante et al. (2019) investigated DNA methylation and expression changes in coronary heart disease patients undergoing Cardiac Computed Tomography (CCT). They could show that genes involved in cholesterol bioactivity such as *LDLR* promoter have higher methylation in PBMNCs of CHD patients compared to healthy controls. *LDLR* promoter methylation was also associated with calcified plaque volume and total plague burden measured via CCT. A case-control study using Human CpG 12K Array (HCGI12K) revealed 72 DMRs hyper-methylated in patients with coronary artery disease (CAD) (Sharma et al., 2014). More recently, an EWAS analyses using the HumanMethylation450 BeadChips reported 211 CpG sites located on 196 genes to be differentially methylated in patients with a history of myocardial infarction (MI) (Rask-Andersen et al., 2016). A similar EWAS study on acute coronary syndrome revealed associations with blood methylation levels of 47 CpG sites located in genes involved in atherogenic signaling and immune response (Li et al., 2017). Nakatochi et al. (2017) also performed an EWAS analyses on blood DNA of patients suffering from MI which revealed three differentially methylated CpG sites in *SGK1, SMARC4,* and *ZFHX3*. A large EWAS study on the Women's Health Initiative (discovery set) and Framingham Heart Study (FHS) – (replication set) identified three DMRs in *SLC9A1*, *SLC1A5*, and *TNRC6C* linked to CVD incidence (Westerman et al., 2018). The authors also performed a module based epigenetic analysis, which revealed three modules associated with CVD and its risk factors out of which two had strong concordance in both cohorts (Westerman et al., 2018).

A growing number of studies reported a possible role for DNA methylation in atherosclerosis pathogenesis (Newman, 1999; Napoli et al., 2012; Aavik et al., 2015; Liu et al., 2018). Atherosclerotic lesions are known to harbor differentially methylated CpGs in genes involved in endothelial and smooth muscle functions (Zaina et al., 2014). Circulating concentrations of tumor necrosis factor α, a pro-inflammatory cytokine linked to atherosclerosis, were recently shown to be associated with methylation changes in the immune response-related genes *DTX3L-PARP9* and *NLRC5*. DNA methylation levels of those genes were also shown to negatively correlate with CHD incidence (Aslibekyan et al., 2018). Similarly, a large EWAS meta-analysis on serum C-reactive protein (CRP), an inflammation biomarker predicting heart failure, identified 58 CpG sites related to CRP levels. Several of those CpGs (51 sites) were associated with cardio-metabolic traits including CHD prevalence and incidence (Ligthart et al., 2016). More recently, focus shifted toward understanding the role of 5-Hydroxymethylcytosine in CVD, where reports have shown that global DNA hydroxymethylation levels could be better predictors of MI and CHD when compared to 5-mC. In elderly individuals, the incidence and degree of coronary atherosclerosis (CA) were linked to increased DNA hydroxymethylation levels in PBMCs (Jiang et al., 2019a). This lead the authors to propose a novel CA biomarker based on integrating carotid plaques scores, as well as DNA methylation and hydroxymethylation data (Jiang et al., 2019b).

From a precision health perspective, a machine learning based framework focused on the FHS cohort could detect CHD presence and foresee its incidence by implementing genetic, epigenetic and phenotypic data (Dogan et al., 2018a,b). Similarly, DNA methylation levels in the *TRAF3* gene were reported to predict recurrence of ischemic events in patients treated with Clopidogrel (Gallego-Fabrega et al., 2016b). A conceptually related study from the same group identified *PPM1A* methylation to be associated with vascular recurrence after stroke in aspirin treated patients (Gallego-Fabrega et al., 2016a). Nonetheless, there must be a more concerted effort to establish whether the reported epigenetic alterations can be reliable CVD biomarkers.

CONCLUSION AND FUTURE PERSPECTIVES

Despite the extensive plethora of epigenetic modifications, measuring DNA methylation of specific CpG sites remains the most promising epigenetic biomarker. DNA methylation modifications are highly stable compared to RNA- or protein-based biomarkers, relatively easy to measure using non-invasive biospecimen, and are quantifiable marks on the DNA that can track the influences of various environmental and lifestyle factors (Berdasco and Esteller, 2019). Nevertheless, epigenetic biomarkers are still in the nascent stage and more research is warranted to move toward applications in healthcare. Still, efforts invested in developing biomarkers based on the epigenetic clocks has accelerated discoveries in the field. Furthermore, GRAIL a multi-billion dollar investment has chosen DNA methylation as its preferred approach for a non-invasive test for early cancer detection.

A key factor in the development of epigenetic clocks was the advent of Infinium Methylation arrays that enabled simultaneous quantification of DNA methylation starting from ∼27,000 individual CpG sites (Infinium HumanMethylation27 BeadChip) up to 850,000 sites via EPIC arrays. These methylation arrays provide a cost-effective approach for large-scale epigenetic epidemiology studies. Nevertheless, the human genome is comprised of 28 million CpG sites out of which 3% are measured using Epic Arrays. Even though, a few reports have mentioned that whole genome bisulfite sequencing (WGBS) is potentially inefficient due to non-dynamic methylation across a large fraction of CpG cites as well as the majority of WGBS reads being non-informative (Ziller et al., 2013). Nevertheless, sequencing costs are

decreasing dramatically and more comprehensive DNA methylation datasets would become publicly available once whole-genome bisulfite and oxidative bisulfite sequencing becomes mainstream. Development of more accurate epigenetic biomarkers by relying on whole genome sequencing data will be a hot topic in the next years. Future work based on these data should be even more exciting and would have important implications for human health.

AUTHOR CONTRIBUTIONS

All authors were involved in literature review, writing the manuscript, and figure preparation.

ACKNOWLEDGMENTS

The authors would like to thank Aya Abdelaal for helpful remarks to improve the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00171/full#supplementary-material

TABLE S1 | Summary of the potential blood-based epigenetic biomarkers for Alzheimer's disease, cardiovascular disease, and Type 2 diabetes.

REFERENCES

Aavik, E., Lumivuori, H., Leppanen, O., Wirth, T., Hakkinen, S.-K., Brasen, J.-H., et al. (2015). Global DNA methylation analysis of human atherosclerotic plaques reveals extensive genomic hypomethylation and reactivation at imprinted locus 14q32 involving induction of a miRNA cluster. *Eur. Heart J.* 36, 993–1000. doi: 10.1093/eurheartj/ehu437

Almenar-Queralt, A., Merkurjev, D., Kim, H. S., Navarro, M., Ma, Q., Chaves, R. S., et al. (2019). Chromatin establishes an immature version of neuronal protocadherin selection during the naive-to-primed conversion of pluripotent stem cells. *Nat. Genet.* 51, 1691–1701. doi: 10.1038/s41588-019-0526-4

Ambatipudi, S., Horvath, S., Perrier, F., Cuenin, C., Hernandez-Vargas, H., Le Calvez-Kelm, F., et al. (2017). DNA methylome analysis identifies accelerated epigenetic ageing associated with postmenopausal breast cancer susceptibility. *Eur. J. Cancer* 75, 299–307. doi: 10.1016/j.ejca.2017.01.014

Anglim, P. P., Alonzo, T. A., and Laird-Offringa, I. A. (2008). DNA methylation-based biomarkers for early detection of non-small cell lung cancer: an update. *Mol. Cancer* 7:81. doi: 10.1186/1476-4598-7-81

Aslibekyan, S., Agha, G., Colicino, E., Do, A. N., Lahti, J., Ligthart, S., et al. (2018). Association of methylation signals with incident coronary heart disease in an epigenome-wide assessment of circulating tumor necrosis factor α. *JAMA Cardiol.* 3, 463–472. doi: 10.1001/jamacardio.2018.0510

Atsem, S., Reichenbach, J., Potabattula, R., Dittrich, M., Nava, C., Depienne, C., et al. (2016). Paternal age effects on sperm FOXK1 and KCNA7 methylation and transmission into the next generation. *Hum. Mol. Genet.* 25, 4996–5005. doi: 10.1093/hmg/ddw328

Barrès, R., Osler, M. E., Yan, J., Rune, A., Fritz, T., Caidahl, K., et al. (2009). Non-CpG methylation of the PGC-1α promoter through DNMT3B controls mitochondrial density. *Cell Metab.* 10, 189–198. doi: 10.1016/j.cmet.2009.07.011

Berdasco, M., and Esteller, M. (2019). Clinical epigenetics: seizing opportunities for translation. *Nat. Rev. Genet.* 20, 109–127. doi: 10.1038/s41576-018-0074-2

BI Ageing Clock Team, Stubbs, T. M., Bonder, M. J., Stark, A.-K., Krueger, F., von Meyenn, F., et al. (2017). Multi-tissue DNA methylation age predictor in mouse. *Genome Biol.* 18:68. doi: 10.1186/s13059-017-1203-5

Bollati, V., Galimberti, D., Pergoli, L., Dalla Valle, E., Barretta, F., Cortini, F., et al. (2011). DNA methylation in repetitive elements and Alzheimer disease. *Brain Behav. Immun.* 25, 1078–1083. doi: 10.1016/j.bbi.2011.01.017

Brohede, J., Rinde, M., Winblad, B., and Graff, C. (2010). A DNA methylation study of the amyloid precursor protein gene in several brain regions from patients with familial Alzheimer disease. *J. Neurogenet.* 24, 179–181. doi: 10.3109/01677063.2010.503978

Carboni, L., Lattanzio, F., Candeletti, S., Porcellini, E., Raschi, E., Licastro, F., et al. (2015). Peripheral leukocyte expression of the potential biomarker proteins Bdnf, Sirt1, and Psen1 is not regulated by promoter methylation in Alzheimer's disease patients. *Neurosci. Lett.* 605, 44–48. doi: 10.1016/j.neulet.2015.08.012

Carroll, J. E., Irwin, M. R., Levine, M., Seeman, T. E., Absher, D., Assimes, T., et al. (2017). Epigenetic aging and immune senescence in women with insomnia symptoms: findings from the women's health initiative study. *Biol. Psychiatry* 81, 136–144. doi: 10.1016/j.biopsych.2016.07.008

Castillo-Fernandez, J. E., Spector, T. D., and Bell, J. T. (2014). Epigenetics of discordant monozygotic twins: implications for disease. *Genome Med.* 6:60. doi: 10.1186/s13073-014-0060-z

Chambers, J. C., Loh, M., Lehne, B., Drong, A., Kriebel, J., Motta, V., et al. (2015). Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. *Lancet Diabetes Endocrinol.* 3, 526–534. doi: 10.1016/S2213-8587(15)00127-8

Chan, M. (2014). Global report on diabetes. *World Health Org.* 58, 1–88. doi: 10.1128/AAC.03728-14

Chatterjee, S., Khunti, K., and Davies, M. J. (2017). Type 2 diabetes. *Lancet* 389, 2239–2251. doi: 10.1016/S0140-6736(17)30058-2

Chen, B. H., Marioni, R. E., Colicino, E., Peters, M. J., Ward-Caviness, C. K., Tsai, P.-C., et al. (2016). DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging (Albany NY)* 8, 1844–1865. doi: 10.18632/aging.101020

Chen, R., Xia, L., Tu, K., Duan, M., Kukurba, K., Li-Pook-Than, J., et al. (2018). Longitudinal personal DNA methylome dynamics in a human with a chronic condition. *Nat. Med.* 24, 1930–1939. doi: 10.1038/s41591-018-0237-x

Christiansen, L., Lenart, A., Tan, Q., Vaupel, J. W., Aviv, A., McGue, M., et al. (2016). DNA methylation age is associated with mortality in a longitudinal Danish twin study. *Aging Cell* 15, 149–154. doi: 10.1111/acel.12421

Crimmins, E. M. (2015). Lifespan and healthspan: past, present, and promise. *Gerontologist* 55, 901–911. doi: 10.1093/geront/gnv130

Dayeh, T., Tuomi, T., Almgren, P., Perfilyev, A., Jansson, P.-A., de Mello, V. D., et al. (2016). DNA methylation of loci within *ABCG1* and *PHOSPHO1* in blood DNA is associated with future type 2 diabetes risk. *Epigenetics* 11, 482–488. doi: 10.1080/15592294.2016.1178418

De Jager, P. L., Srivastava, G., Lunnon, K., Burgess, J., Schalkwyk, L. C., Yu, L., et al. (2014). Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat. Neurosci.* 17, 1156–1163. doi: 10.1038/nn.3786

Di Francesco, A., Arosio, B., Falconi, A., Micioni Di Bonaventura, M. V., Karimi, M., Mari, D., et al. (2015). Global changes in DNA methylation in Alzheimer's disease peripheral blood mononuclear cells. *Brain Behav. Immun.* 45, 139–144. doi: 10.1016/j.bbi.2014.11.002

Dogan, M. V., Beach, S., Simons, R., Lendasse, A., Penaluna, B., and Philibert, R. (2018a). Blood-based biomarkers for predicting the risk for five-year incident coronary heart disease in the Framingham Heart Study via machine learning. *Genes* 9:641. doi: 10.3390/genes9120641

Dogan, M. V., Grumbach, I. M., Michaelson, J. J., and Philibert, R. A. (2018b). Integrated genetic and epigenetic prediction of coronary heart disease in the Framingham Heart Study. *PLoS One* 13:e0190549. doi: 10.1371/journal.pone.0190549

Duong, S., Patel, T., and Chang, F. (2017). Dementia: what pharmacists need to know. *Can. Pharm. J. (Ott)* 150, 118–129. doi: 10.1177/1715163517690745

Edwards, T. M., and Myers, J. P. (2007). Environmental exposures and gene regulation in disease etiology. *Environ. Health Perspect* 115, 1264–1270. doi: 10.1289/ehp.9951

El Hajj, N., Dittrich, M., Böck, J., Kraus, T. F. J., Nanda, I., Müller, T., et al. (2016). Epigenetic dysregulation in the developing Down syndrome cortex. *Epigenetics* 11, 563–578. doi: 10.1080/15592294.2016.1192736

El Hajj, N., Haertle, L., Dittrich, M., Denk, S., Lehnen, H., Hahn, T., et al. (2017). DNA methylation signatures in cord blood of ICSI children. *Hum. Reprod.* 32, 1761–1769. doi: 10.1093/humrep/dex209

Finder, V. H. (2010). Alzheimer's disease: a general introduction and pathomechanism. *J. Alzheimers Dis.* 22, S5–S19. doi: 10.3233/JAD-2010-100975

Florath, I., Butterbach, K., Heiss, J., Bewerunge-Hudler, M., Zhang, Y., Schöttker, B., et al. (2016). Type 2 diabetes and leucocyte DNA methylation: an epigenome-wide association study in over 1,500 older adults. *Diabetologia* 59, 130–138. doi: 10.1007/s00125-015-3773-7

Fraga, M. F., Ballestar, E., Paz, M. F., Ropero, S., Setien, F., Ballestar, M. L., et al. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl. Acad. Sci. U.S.A.* 102, 10604–10609. doi: 10.1073/pnas.0500398102

Franceschi, C., Garagnani, P., Morsiani, C., Conte, M., Santoro, A., Grignolio, A., et al. (2018). The continuum of aging and age-related diseases: common mechanisms but different rates. *Front. Med. (Lausanne)* 5:61. doi: 10.3389/fmed.2018.00061

Frederiksen, H., Gaist, D., Christian Petersen, H., Hjelmborg, J., McGue, M., Vaupel, J. W., et al. (2002). Hand grip strength: a phenotype suitable for identifying genetic variants affecting mid- and late-life physical functioning. *Genet. Epidemiol.* 23, 110–122. doi: 10.1002/gepi.1127

Gallego-Fabrega, C., Carrera, C., Reny, J.-L., Fontana, P., Slowik, A., Pera, J., et al. (2016a). *PPM1A* methylation is associated with vascular recurrence in aspirin-treated patients. *Stroke* 47, 1926–1929. doi: 10.1161/STROKEAHA.116.013340

Gallego-Fabrega, C., Carrera, C., Reny, J.-L., Fontana, P., Slowik, A., Pera, J., et al. (2016b). *TRAF3* epigenetic regulation is associated with vascular recurrence in patients with ischemic stroke. *Stroke* 47, 1180–1186. doi: 10.1161/STROKEAHA.115.012237

Galloway, A. (1993). The evolutionary biology of aging. By Michael R. Rose, New York: Oxford University Press. 1991. ix + 221 pp. ISBN 0-19-506133-0. $35.00 (cloth). *Am. J. Phys. Anthropol.* 91, 260–262. doi: 10.1002/ajpa.1330910217

García-Giménez, J. L., Ushijima, T., and Tollefsbol, T. O. (2016). "Epigenetic biomarkers: new findings, perspectives, and future directions in diagnostics," in *Epigenetic Biomarkers and Diagnostics*, ed. J. L. García-Giménez (Amsterdam: Elsevier), 1–18. doi: 10.1016/B978-0-12-801899-6.00001-2

Greenwood, T. A., Beeri, M. S., Schmeidler, J., Valerio, D., Raventós, H., Mora-Villalobos, L., et al. (2011). Heritability of cognitive functions in families of successful cognitive aging probands from the Central Valley of Costa Rica. *J. Alzheimers Dis.* 27, 897–907. doi: 10.3233/JAD-2011-110782

Haertle, L., Müller, T., Lardenoije, R., Maierhofer, A., Dittrich, M., Riemens, R. J. M., et al. (2019). Methylomic profiling in trisomy 21 identifies cognition- and Alzheimer's disease-related dysregulation. *Clin. Epigenet.* 11:195. doi: 10.1186/s13148-019-0787-x

Hall, E., Dayeh, T., Kirkpatrick, C. L., Wollheim, C. B., Dekker Nitert, M., and Ling, C. (2013). DNA methylation of the glucagon-like peptide 1 receptor (GLP1R) in human pancreatic islets. *BMC Med. Genet.* 14:76. doi: 10.1186/1471-2350-14-76

Han, Y., Eipel, M., Franzen, J., Sakk, V., Dethmers-Ausema, B., Yndriago, L., et al. (2018). Epigenetic age-predictor for mice based on three CpG sites. *eLife* 7:e37462. doi: 10.7554/eLife.37462

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., et al. (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* 49, 359–367. doi: 10.1016/j.molcel.2012.10.016

Hardy, J. (2006). A hundred years of Alzheimer's disease research. *Neuron* 52, 3–13. doi: 10.1016/j.neuron.2006.09.016

Heyn, H., Li, N., Ferreira, H. J., Moran, S., Pisano, D. G., Gomez, A., et al. (2012). Distinct DNA methylomes of newborns and centenarians. *Proc. Natl. Acad. Sci. U.S.A.* 109, 10522–10527. doi: 10.1073/pnas.1120658109

Hidalgo, B., Irvin, M. R., Sha, J., Zhi, D., Aslibekyan, S., Absher, D., et al. (2014). Epigenome-wide association study of fasting measures of glucose, insulin, and HOMA-IR in the genetics of lipid lowering drugs and diet network study. *Diabetes* 63, 801–807. doi: 10.2337/db13-1100

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biol.* 14:R115. doi: 10.1186/gb-2013-14-10-r115

Horvath, S., Erhart, W., Brosch, M., Ammerpohl, O., von Schonfels, W., Ahrens, M., et al. (2014). Obesity accelerates epigenetic aging of human liver. *Proc. Natl. Acad. Sci. U.S.A.* 111, 15538–15543. doi: 10.1073/pnas.1412759111

Horvath, S., Garagnani, P., Bacalini, M. G., Pirazzini, C., Salvioli, S., Gentilini, D., et al. (2015). Accelerated epigenetic aging in Down syndrome. *Aging Cell* 14, 491–495. doi: 10.1111/acel.12325

Horvath, S., Gurven, M., Levine, M. E., Trumble, B. C., Kaplan, H., Allayee, H., et al. (2016a). An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biol.* 17:171. doi: 10.1186/s13059-016-1030-0

Horvath, S., Langfelder, P., Kwak, S., Aaronson, J., Rosinski, J., Vogt, T. F., et al. (2016b). Huntington's disease accelerates epigenetic aging of human brain and disrupts DNA methylation levels. *Aging* 8, 1485–1512. doi: 10.18632/aging.101005

Horvath, S., and Raj, K. (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* 19, 371–384. doi: 10.1038/s41576-018-0004-3

Horvath, S., and Ritz, B. R. (2015). Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. *Aging* 7, 1130–1142. doi: 10.18632/aging.100859

Infante, T., Forte, E., Schiano, C., Punzo, B., Cademartiri, F., Cavaliere, C., et al. (2019). Evidence of association of circulating epigenetic-sensitive biomarkers with suspected coronary heart disease evaluated by Cardiac Computed Tomography. *PLoS One* 14:e0210909. doi: 10.1371/journal.pone.0210909

Issa, J.-P. (2014). Aging and epigenetic drift: a vicious cycle. *J. Clin. Invest.* 124, 24–29. doi: 10.1172/JCI69735

Jacobs, E., Hoyer, A., Brinks, R., Kuss, O., and Rathmann, W. (2017). Burden of mortality attributable to diagnosed diabetes: a nationwide analysis based on claims data from 65 million people in Germany. *Diabetes Care* 40, 1703–1709. doi: 10.2337/dc17-0954

Jaul, E., and Barron, J. (2017). Age-related diseases and clinical and public health implications for the 85 years old and over population. *Front. Public Health* 5:335. doi: 10.3389/fpubh.2017.00335

Jiang, D., Sun, M., You, L., Lu, K., Gao, L., Hu, C., et al. (2019a). DNA methylation and hydroxymethylation are associated with the degree of coronary atherosclerosis in elderly patients with coronary heart disease. *Life Sci.* 224, 241–248. doi: 10.1016/j.lfs.2019.03.021

Jiang, D., Wang, Y., Chang, G., Duan, Q., You, L., Sun, M., et al. (2019b). DNA hydroxymethylation combined with carotid plaques as a novel biomarker for coronary atherosclerosis. *Aging* 11, 3170–3181. doi: 10.18632/aging.101972

Kananen, L., Marttila, S., Nevalainen, T., Kummola, L., Junttila, I., Mononen, N., et al. (2016). The trajectory of the blood DNA methylome ageing rate is largely set before adulthood: evidence from two longitudinal studies. *Age* 38:65. doi: 10.1007/s11357-016-9927-9

Kato, N., Loh, M., Takeuchi, F., Verweij, N., Wang, X., Zhang, W., et al. (2015). Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. *Nat. Genet.* 47, 1282–1293. doi: 10.1038/ng.3405

Khare, T., Pai, S., Koncevicius, K., Pal, M., Kriukiene, E., Liutkeviciute, Z., et al. (2012). 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nat. Struct. Mol. Biol.* 19, 1037–1043. doi: 10.1038/nsmb.2372

Kim, M., Long, T. I., Arakawa, K., Wang, R., Yu, M. C., and Laird, P. W. (2010). DNA methylation as a biomarker for cardiovascular disease risk. *PLoS One* 5:e9692. doi: 10.1371/journal.pone.0009692

Knight, A. K., Craig, J. M., Theda, C., Bækvad-Hansen, M., Bybjerg-Grauholm, J., Hansen, C. S., et al. (2016). An epigenetic clock for gestational age at birth based on blood methylation data. *Genome Biol.* 17:206. doi: 10.1186/s13059-016-1068-z

Kobayashi, N., Shinagawa, S., Nagata, T., Shimada, K., Shibata, N., Ohnuma, T., et al. (2016). Usefulness of DNA methylation levels in COASY and SPINT1 gene promoter regions as biomarkers in diagnosis of Alzheimer's disease and amnestic mild cognitive impairment. *PLoS One* 11:e0168816. doi: 10.1371/journal.pone.0168816

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA Base 5-hydroxymethylcytosine is present in purkinje neurons and the brain. *Science* 324, 929–930. doi: 10.1126/science.1169786

Kulkarni, H., Kos, M. Z., Neary, J., Dyer, T. D., Kent, J. W., Göring, H. H. H., et al. (2015). Novel epigenetic determinants of type 2 diabetes in Mexican-American families. *Hum. Mol. Genet.* 24, 5330–5344. doi: 10.1093/hmg/ddv232

Lardenoije, R., Iatrou, A., Kenis, G., Kompotis, K., Steinbusch, H. W. M., Mastroeni, D., et al. (2015). The epigenetics of aging and neurodegeneration. *Prog. Neurobiol.* 131, 21–64. doi: 10.1016/j.pneurobio.2015.05.002

Lardenoije, R., Roubroeks, J. A. Y., Pishva, E., Leber, M., Wagner, H., Iatrou, A., et al. (2019). Alzheimer's disease-associated (hydroxy)methylomic changes in the brain and blood. *Clin. Epigenet.* 11:164. doi: 10.1186/s13148-019-0755-5

Lévesque, M. L., Casey, K. F., Szyf, M., Ismaylova, E., Ly, V., Verner, M.-P., et al. (2014). Genome-wide DNA methylation variability in adolescent monozygotic twins followed since birth. *Epigenetics* 9, 1410–1421. doi: 10.4161/15592294.2014.970060

Levine, M. E., Hosgood, H. D., Chen, B., Absher, D., Assimes, T., and Horvath, S. (2015a). DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. *Aging* 7, 690–700. doi: 10.18632/aging.100809

Levine, M. E., Lu, A. T., Bennett, D. A., and Horvath, S. (2015b). Epigenetic age of the pre-frontal cortex is associated with neuritic plaques, amyloid load, and Alzheimer's disease related cognitive functioning. *Aging* 7, 1198–1211. doi: 10.18632/aging.100864

Levine, M. E., Lu, A. T., Quach, A., Chen, B. H., Assimes, T. L., Bandinelli, S., et al. (2018). An epigenetic biomarker of aging for lifespan and healthspan. *Aging* 10, 573–591. doi: 10.18632/aging.101414

Li, J., Zhu, X., Yu, K., Jiang, H., Zhang, Y., Deng, S., et al. (2017). Genome-wide analysis of DNA methylation and acute coronary syndrome. *Cir. Res.* 120, 1754–1767. doi: 10.1161/CIRCRESAHA.116.310324

Ligthart, S., Marzi, C., Aslibekyan, S., Mendelson, M. M., Conneely, K. N., Tanaka, T., et al. (2016). DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. *Genome Biol.* 17:255. doi: 10.1186/s13059-016-1119-5

Lind, L., Ingelsson, E., Sundström, J., Siegbahn, A., and Lampa, E. (2018). Methylation-based estimated biological age and cardiovascular disease. *Eur. J. Clin. Invest.* 48:e12872. doi: 10.1111/eci.12872

Ling, C., Del Guerra, S., Lupi, R., Rönn, T., Granhall, C., Luthman, H., et al. (2008). Epigenetic regulation of PPARGC1A in human type 2 diabetic islets and effect on insulin secretion. *Diabetologia* 51, 615–622. doi: 10.1007/s00125-007-0916-5

Ling, C., and Rönn, T. (2019). Epigenetics in human obesity and type 2 diabetes. *Cell Metab.* 29, 1028–1044. doi: 10.1016/j.cmet.2019.03.009

Liu, J., Jiang, J., Mo, J., Liu, D., Cao, D., Wang, H., et al. (2019). Global DNA 5-hydroxymethylcytosine and 5-formylcytosine contents are decreased in the early stage of hepatocellular carcinoma. *Hepatology* 69, 196–208. doi: 10.1002/hep.30146

Liu, Y., Peng, W., Qu, K., Lin, X., Zeng, Z., Chen, J., et al. (2018). *TET2*: a novel epigenetic regulator and potential intervention target for atherosclerosis. *DNA Cell Biol.* 37, 517–523. doi: 10.1089/dna.2017.4118

López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. *Cell* 153, 1194–1217. doi: 10.1016/j.cell.2013.05.039

Lu, A. T., Hannon, E., Levine, M. E., Crimmins, E. M., Lunnon, K., Mill, J., et al. (2017). Genetic architecture of epigenetic and neuronal ageing rates in human brain regions. *Nat. Commun.* 8:15353. doi: 10.1038/ncomms15353

Lu, A. T., Hannon, E., Levine, M. E., Hao, K., Crimmins, E. M., Lunnon, K., et al. (2016). Genetic variants near MLST8 and DHX57 affect the epigenetic age of the cerebellum. *Nat. Commun.* 7:10561. doi: 10.1038/ncomms10561

Lu, A. T., Quach, A., Wilson, J. G., Reiner, A. P., Aviv, A., Raj, K., et al. (2019). DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging* 11, 303–327. doi: 10.18632/aging.101684

Lu, A. T., Xue, L., Salfati, E. L., Chen, B. H., Ferrucci, L., Levy, D., et al. (2018). GWAS of epigenetic aging rates in blood reveals a critical role for TERT. *Nat. Commun.* 9:387. doi: 10.1038/s41467-017-02697-5

Lunnon, K., Smith, R., Hannon, E., De Jager, P. L., Srivastava, G., Volta, M., et al. (2014). Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat. Neurosci.* 17, 1164–1170. doi: 10.1038/nn.3782

Maierhofer, A., Flunkert, J., Oshima, J., Martin, G. M., Haaf, T., and Horvath, S. (2017). Accelerated epigenetic aging in Werner syndrome. *Aging* 9, 1143–1152. doi: 10.18632/aging.101217

Mano, T., Nagata, K., Nonaka, T., Tarutani, A., Imamura, T., Hashimoto, T., et al. (2017). Neuron-specific methylome analysis reveals epigenetic regulation and tau-related dysfunction of BRCA1 in Alzheimer's disease. *Proc. Natl. Acad. Sci.* 114, E9645–E9654. doi: 10.1073/pnas.1707151114

Marioni, R. E., Shah, S., McRae, A. F., Chen, B. H., Colicino, E., Harris, S. E., et al. (2015). DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol.* 16:25. doi: 10.1186/s13059-015-0584-6

McCarthy, M. I. (2010). Genomics, type 2 diabetes, and obesity. *N. Engl. J. Med.* 363, 2339–2350. doi: 10.1056/NEJMra0906948

Meer, M. V., Podolskiy, D. I, Tyshkovskiy, A., and Gladyshev, V. N. (2018). A whole lifespan mouse multi-tissue DNA methylation clock. *eLife* 7:e40675. doi: 10.7554/eLife.40675

Mercorio, R., Pergoli, L., Galimberti, D., Favero, C., Carugno, M., Dalla Valle, E., et al. (2018). PICALM gene methylation in blood of Alzheimer's disease patients is associated with cognitive decline. *J. Alzheimers Dis.* 65, 283–292. doi: 10.3233/JAD-180242

Moore, L. D., Le, T., and Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology* 38, 23–38. doi: 10.1038/npp.2012.112

Nagata, T., Kobayashi, N., Ishii, J., Shinagawa, S., Nakayama, R., Shibata, N., et al. (2015). Association between DNA methylation of the BDNF promoter region and clinical presentation in Alzheimer's disease. *Dement. Geriatr. Cogn. Dis. Extra* 5, 64–73. doi: 10.1159/000375367

Nakatochi, M., Ichihara, S., Yamamoto, K., Naruse, K., Yokota, S., Asano, H., et al. (2017). Epigenome-wide association of myocardial infarction with DNA methylation sites at loci related to cardiovascular disease. *Clin. Epigenetics* 9:54. doi: 10.1186/s13148-017-0353-3

Napoli, C., Crudele, V., Soricelli, A., Al-Omran, M., Vitale, N., Infante, T., et al. (2012). Primary prevention of atherosclerosis: a clinical challenge for the reversal of epigenetic mechanisms? *Circulation* 125, 2363–2373. doi: 10.1161/CIRCULATIONAHA.111.085787

Newman, P. E. (1999). Can reduced folic acid and vitamin B12 levels cause deficient DNA methylation producing mutations which initiate atherosclerosis? *Med. Hypotheses* 53, 421–424. doi: 10.1054/mehy.1998.0794

Ozaki, Y., Yoshino, Y., Yamazaki, K., Sao, T., Mori, Y., Ochi, S., et al. (2017). DNA methylation changes at TREM2 intron 1 and TREM2 mRNA expression in patients with Alzheimer's disease. *J. Psychiatr. Res.* 92, 74–80. doi: 10.1016/j.jpsychires.2017.04.003

Petkovich, D. A., Podolskiy, D. I, Lobanov, A. V., Lee, S.-G., Miller, R. A., and Gladyshev, V. N. (2017). Using DNA methylation profiling to evaluate biological age and longevity interventions. *Cell Metab.* 25, 954–960.e6. doi: 10.1016/j.cmet.2017.03.016

Potabattula, R., Dittrich, M., Böck, J., Haertle, L., Müller, T., Hahn, T., et al. (2018). Allele-specific methylation of imprinted genes in fetal cord blood is influenced by cis-acting genetic variants and parental factors. *Epigenomics* 10, 1315–1326. doi: 10.2217/epi-2018-0059

Poulsen, P., Esteller, M., Vaag, A., and Fraga, M. F. (2007). The epigenetic basis of twin discordance in age-related diseases. *Pediatr. Res.* 61, 38R–42R. doi: 10.1203/pdr.0b013e31803c7b98

Quach, A., Levine, M. E., Tanaka, T., Lu, A. T., Chen, B. H., Ferrucci, L., et al. (2017). Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. *Aging* 9, 419–446. doi: 10.18632/aging.101168

Rask-Andersen, M., Martinsson, D., Ahsan, M., Enroth, S., Ek, W. E., Gyllensten, U., et al. (2016). Epigenome-wide association study reveals differential DNA methylation in individuals with a history of myocardial infarction. *Hum. Mol. Genet.* 25, 4739–4748. doi: 10.1093/hmg/ddw302

Reynolds, C. A., Finkel, D., McArdle, J. J., Gatz, M., Berg, S., and Pedersen, N. L. (2005). Quantitative genetic analysis of latent growth curve models of cognitive abilities in adulthood. *Dev. Psychol.* 41, 3–16. doi: 10.1037/0012-1649.41.1.3

Seman, N., Mohamud, W., Östenson, C.-G., Brismar, K., and Gu, H. F. (2015). Increased DNA methylation of the SLC30A8 gene promoter is associated with type 2 diabetes in a Malay population. *Clin. Epigenetics* 7:30. doi: 10.1186/s13148-015-0049-5

Sharma, P., Garg, G., Kumar, A., Mohammad, F., Kumar, S. R., Tanwar, V. S., et al. (2014). Genome wide DNA methylation profiling for epigenetic alteration in coronary artery disease patients. *Gene* 541, 31–40. doi: 10.1016/j.gene.2014.02.034

Sharma, P., Kumar, J., Garg, G., Kumar, A., Patowary, A., Karthikeyan, G., et al. (2008). Detection of altered global dna methylation in coronary artery disease patients. *DNA Cell Biol.* 27, 357–365. doi: 10.1089/dna.2007.0694

Shinagawa, S., Kobayashi, N., Nagata, T., Kusaka, A., Yamada, H., Kondo, K., et al. (2016). DNA methylation in the NCAPH2/LMF2 promoter region is associated with hippocampal atrophy in Alzheimer's disease and amnesic mild cognitive impairment patients. *Neurosci. Lett.* 629, 33–37. doi: 10.1016/j.neulet.2016.06.055

Soriano-Tárraga, C., Jiménez-Conde, J., Giralt-Steinhauer, E., Mola-Caminal, M., Vivanco-Hidalgo, R. M., Ois, A., et al. (2016). Epigenome-wide association study identifies *TXNIP* gene associated with type 2 diabetes mellitus and sustained hyperglycemia. *Hum. Mol. Genet.* 25, 609–619. doi: 10.1093/hmg/ddv493

Spiers, H., Hannon, E., Schalkwyk, L. C., Smith, R., Wong, C. C. Y., O'Donovan, M. C., et al. (2015). Methylomic trajectories across human fetal brain development. *Genome Res.* 25, 338–352. doi: 10.1101/gr.180273.114

Tan, Q., Heijmans, B. T., Hjelmborg, J. V. B., Soerensen, M., Christensen, K., and Christiansen, L. (2016). Epigenetic drift in the aging genome: a ten-year follow-up in an elderly twin cohort. *Int. J. Epidemiol.* 45, 1146–1158. doi: 10.1093/ije/dyw132

Thompson, M. J., Chwiałkowska, K., Rubbi, L., Lusis, A. J., Davis, R. C., Srivastava, A., et al. (2018). A multi-tissue full lifespan epigenetic clock for mice. *Aging* 10, 2832–2854. doi: 10.18632/aging.101590

Tohgi, H., Utsugisawa, K., Nagane, Y., Yoshimura, M., Genda, Y., and Ukitsu, M. (1999). Reduction with age in methylcytosine in the promoter region -224 approximately -101 of the amyloid precursor protein gene in autopsy human cortex. *Brain Res. Mol. Brain Res.* 70, 288–292. doi: 10.1016/s0169-328x(99)00163-1

Toperoff, G., Aran, D., Kark, J. D., Rosenberg, M., Dubnikov, T., Nissan, B., et al. (2012). Genome-wide survey reveals predisposing diabetes type 2-related DNA methylation variations in human peripheral blood. *Hum. Mol. Genet.* 21, 371–383. doi: 10.1093/hmg/ddr472

Vanyushin, B. F., Nemirovsky, L. E., Klimenko, V. V., Vasiliev, V. K., and Belozersky, A. N. (1973). The 5-methylcytosine in DNA of rats. Tissue and age specificity and the changes induced by hydrocortisone and other agents. *Gerontologia* 19, 138–152.

Volkov, P., Bacos, K., Ofori, J. K., Esguerra, J. L. S., Eliasson, L., Rönn, T., et al. (2017). Whole-genome bisulfite sequencing of human pancreatic islets reveals novel differentially methylated regions in type 2 diabetes pathogenesis. *Diabetes* 66, 1074–1085. doi: 10.2337/db16-0996

Wahl, S., Drong, A., Lehne, B., Loh, M., Scott, W. R., Kunze, S., et al. (2017). Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* 541, 81–86. doi: 10.1038/nature20784

Walaszczyk, E., Luijten, M., Spijkerman, A. M. W., Bonder, M. J., Lutgers, H. L., Snieder, H., et al. (2018). DNA methylation markers associated with type 2 diabetes, fasting glucose and HbA1c levels: a systematic review and replication in a case–control sample of the Lifelines study. *Diabetologia* 61, 354–368. doi: 10.1007/s00125-017-4497-7

Wang, M., and Lemos, B. (2019). Ribosomal DNA harbors an evolutionarily conserved clock of biological aging. *Genome Res.* 29, 325–333. doi: 10.1101/gr.241745.118

Wang, S.-C., Oelze, B., and Schumacher, A. (2008). Age-specific epigenetic drift in late-onset Alzheimer's disease. *PLoS One* 3:e2698. doi: 10.1371/journal.pone.0002698

Wang, T., Tsui, B., Kreisberg, J. F., Robertson, N. A., Gross, A. M., Yu, M. K., et al. (2017). Epigenetic aging signatures in mice livers are slowed by dwarfism, calorie restriction and rapamycin treatment. *Genome Biol.* 18:57. doi: 10.1186/s13059-017-1186-2

Wang, Y., Karlsson, R., Lampa, E., Zhang, Q., Hedman, ÅK., Almgren, M., et al. (2018). Epigenetic influences on aging: a longitudinal genome-wide methylation study in old Swedish twins. *Epigenetics* 13, 975–987. doi: 10.1080/15592294.2018.1526028

Weidner, C., Lin, Q., Koch, C., Eisele, L., Beier, F., Ziegler, P., et al. (2014). Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol.* 15:R24. doi: 10.1186/gb-2014-15-2-r24

West, R. L., Lee, J. M., and Maroun, L. E. (1995). Hypomethylation of the amyloid precursor protein gene in the brain of an Alzheimer's disease patient. *J. Mol. Neurosci.* 6, 141–146. doi: 10.1007/BF02736773

Westerman, K., Sebastiani, P., Jacques, P., Liu, S., DeMeo, D., and Ordovás, J. M. (2018). DNA methylation modules associate with incident cardiovascular disease and cumulative risk factor exposure. *bioRxiv* [Preprint]. doi: 10.1101/471722

Willmer, T., Johnson, R., Louw, J., and Pheiffer, C. (2018). Blood-based DNA methylation biomarkers for type 2 diabetes: potential for clinical applications. *Front. Endocrinol.* 9:744. doi: 10.3389/fendo.2018.00744

Xie, B., Liu, Z., Liu, W., Jiang, L., Zhang, R., Cui, D., et al. (2017a). DNA methylation and tag SNPs of the BDNF gene in conversion of amnestic mild cognitive impairment into Alzheimer's disease: a cross-sectional cohort study. *J. Alzheimers Dis.* 58, 263–274. doi: 10.3233/JAD-170007

Xie, B., Xu, Y., Liu, Z., Liu, W., Jiang, L., Zhang, R., et al. (2017b). Elevation of peripheral BDNF promoter methylation predicts conversion from amnestic mild cognitive impairment to Alzheimer's disease: a 5-year longitudinal study. *J. Alzheimers Dis.* 56, 391–401. doi: 10.3233/JAD-160954

Yang, B. T., Dayeh, T. A., Volkov, P. A., Kirkpatrick, C. L., Malmgren, S., Jing, X., et al. (2012). Increased DNA methylation and decreased expression of PDX-1 in pancreatic islets from patients with type 2 diabetes. *Mol. Endocrinol.* 26, 1203–1212. doi: 10.1210/me.2012-1004

Yuan, W., Xia, Y., Bell, C. G., Yet, I., Ferreira, T., Ward, K. J., et al. (2014). An integrated epigenomic analysis for type 2 diabetes susceptibility loci in monozygotic twins. *Nat. Commun.* 5:5719. doi: 10.1038/ncomms6719

Zaina, S., Heyn, H., Carmona, F. J., Varol, N., Sayols, S., Condom, E., et al. (2014). DNA methylation map of human atherosclerosis. *Circ. Cardiovasc. Genet.* 7, 692–700. doi: 10.1161/CIRCGENETICS.113.000441

Zheng, Y., Joyce, B. T., Colicino, E., Liu, L., Zhang, W., Dai, Q., et al. (2016). Blood epigenetic age may predict cancer incidence and mortality. *EBioMedicine* 5, 68–73. doi: 10.1016/j.ebiom.2016.02.008

Ziller, M. J., Gu, H., Müller, F., Donaghey, J., Tsai, L. T.-Y., Kohlbacher, O., et al. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500, 477–481. doi: 10.1038/nature12433

Zwijnenburg, P. J. G., Meijers-Heijboer, H., and Boomsma, D. I. (2010). Identical but not the same: the value of discordant monozygotic twins in genetic research. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 153B, 1134–1149. doi: 10.1002/ajmg.b.31091

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Salameh, Bejaoui and El Hajj. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 27 March 2020
doi: 10.3389/fgene.2020.00238

Detection and Comparative Analysis of Methylomic Biomarkers of Rheumatoid Arthritis

Xin Feng[1,2,3], Xubing Hao[4], Ruoyao Shi[5], Zhiqiang Xia[3], Lan Huang[6], Qiong Yu[1] and Fengfeng Zhou[3]**

[1] Department of Epidemiology and Biostatistics, School of Public Health, Jilin University, Changchun, China, [2] Jilin Institute of Chemical Technology, Jilin, China, [3] BioKnow Health Informatics Lab, College of Computer Science and Technology, and Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, China, [4] BioKnow Health Informatics Lab, College of Software, and Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, China, [5] BioKnow Health Informatics Lab, College of Life Sciences, Jilin University, Changchun, China, [6] College of Computer Science and Technology, and Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, China

Rheumatoid arthritis (RA) is a common autoimmune disorder influenced by both genetic and environmental factors. To investigate possible contributions of DNA methylation to the etiology of RA with minimum confounding genetic heterogeneity, we investigated genome-wide DNA methylation in disease-discordant monozygotic twin pairs. This study hypothesized that methylomic biomarkers might facilitate accurate RA detection. A comprehensive series of biomarker detection algorithms were utilized to find the best methylomic biomarkers for detecting RA patients using the methylomic data of the peripheral blood samples. The best model achieved 100.00% in accuracy (Acc) with 81 methylomic biomarkers and a 10-fold cross-validation (10FCV) strategy. Some of the methylomic biomarkers were experimentally confirmed to be associated with the onset or development of RA. It is also interesting to observe that many of the detected biomarkers were from chromosome Y, supporting the knowledge that RA has a significant gender discrepancy.

Keywords: feature selection, rheumatoid arthritis, methylation biomarker, methylome, chromosome Y

OPEN ACCESS

Edited by:
Yun Liu,
Fudan University, China

Reviewed by:
Nan Lin,
Regeneron Genetic Center, United States
Ling-Qing Yuan,
Central South University, China

****Correspondence:***
Qiong Yu
yuqiong@jlu.edu.cn
Fengfeng Zhou
FengfengZhou@gmail.com;
ffzhou@jlu.edu.cn

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *11 June 2019*
Accepted: *28 February 2020*
Published: *27 March 2020*

Citation:
Feng X, Hao X, Shi R, Xia Z, Huang L, Yu Q and Zhou F (2020) Detection and Comparative Analysis of Methylomic Biomarkers of Rheumatoid Arthritis. Front. Genet. 11:238. doi: 10.3389/fgene.2020.00238

INTRODUCTION

The chronic autoimmune disease rheumatoid arthritis (RA) demonstrates significant changes to joints, with major symptoms like joint pain and swollenness (Triantafyllias et al., 2016). RA is strongly associated with the inflammation around major organs like lungs (Chatzidionisyou and Catrina, 2016; Farquhar et al., 2019) and heart (Crowson et al., 2013; Lazzerini et al., 2017). RA may be developed in about 1% of the population in the developed countries (Smolen et al., 2016). Moreover, females have a 2.5 times high risk than males to develop RA (Alam et al., 2011).

The cause of RA remained unclear and was hypothesized to be under the orchestrated regulation of both genetic and epigenetic factors (Villanueva-Romero et al., 2018; Khan et al., 2019). Various genetic biomarkers were detected through genome-wide association studies (Massey et al., 2018; Shadrina et al., 2018; Lopez-Mejias et al., 2019). Multiple genetic mutations were detected to be statistically associated with the susceptibility for RA, including the SNPs in the genes interferon

regulatory factor 4 (IRF-4) (Lopez-Isac et al., 2016) and Solute Carrier family 8 (SLC8A3) (Julia et al., 2016). Genetic factors were also observed to be associated with the treatment responses of the tumor necrosis factor alpha inhibitors (TNFi) (Massey et al., 2018) and the methotrexate (MTX) monotherapy (Taylor et al., 2018).

Recent studies also demonstrated that the differential status of the epigenomic loci was also statistically significantly associated with RA even in a small population (Julia et al., 2017; Carnero-Montoro and Alarcon-Riquelme, 2018). The RA pathogenesis was observed to be actively regulated by the epigenetic modifications of the immune machineries in the joint tissues (Ibanez-Cabellos et al., 2019). Various environmental factors like cigarette smoking and certain oral pathogens may induce RA through epigenetic modifications (Brandt et al., 2019). Novel treatment plans were proposed to use epigenetic modulators to reverse the differentially methylated regions (Petralia et al., 2019). So the detection of RA methylation biomarkers may both facilitate the understanding of RA pathogenesis and propose more epigenetic drug targets.

There were two main types of computer algorithms to detect biomarkers, i.e., filters and wrappers (Xie et al., 2013; Singh et al., 2018; Verde and De Pietro, 2019). A filter tries to rank the features by each feature's statistical association significance with the phenotype, assuming the features are independent of each other (Lyu et al., 2017). The filter algorithm has a linear time complexity and runs fast enough for many large datasets (Xu et al., 2018). A wrapper utilizes a few heuristic rules to generate a feature subset with a performance evaluation iteratively, and the final feature subset is output if the stop criterion is met (Tekin Erguzel et al., 2015). The strategies of both filters and wrappers may be integrated to generate a hybrid feature selection algorithm (Kumar and Nirmalkumar, 2019; Wu et al., 2019).

This study hypothesized that methylomic features might reflect both the genetic and epigenetic status of RA. So a comprehensive biomarker detection procedure was carried out to find a biomarker set with the satisfying RA prediction accuracy (Acc). The best RA prediction model was also compared with the two sets of methylomic biomarkers from the previous studies. Our model demonstrated a better RA prediction Acc and interesting biological observations.

MATERIALS AND METHODS

Summary of the Dataset

This study screened 485,577 methylomic features detected from 79 RA children and their 79 healthy monozygotic twin siblings (Webster et al., 2018). The twin pairs were identified from the TwinsUK register (Moayyeri et al., 2013) and the RA status was detected in a questionnaire between 1997 and 2002. The twin volunteers were recruited after an advertisement in the National RA Society newsletter in 2013. The RA status was clinical confirmed after these twins were recruited, and only those twins with one healthy and the other RA status were kept for this study. The blood samples were stored at −80°C for DNA extraction.

The methylome was generated by the Illumina HumanMethylation450 BeadChip 15017482 v1.1. The raw data were available at the ArrayExpress database (Athar et al., 2019) with the accession number E-MTAB-6988. This methylomic dataset was formulated as a binary classification problem between the pediatric RA patients and the controls.

The data were provided in the raw format of IDAT, and the methylation level was calculated using the function getBeta() of the R package minfi version 1.28.3 (Aryee et al., 2014).

Pre-screening the Methylomic Features

Many feature selection algorithms run slow on a large dataset, and each methylome has almost half a million features. The downstream feature selection algorithms may crash if they were used directly on the methylomic datasets. So we carried out a pre-screening step to reduce the number of features to be within the capacity of the feature selection algorithms. So the classifier LinearSVC was used to select features for further feature screening. The Python package sklearn has a module SelectFromModel() for this purpose. The model can select features based on the indicators given by the LinearSVC trained on the dataset and the user may determine the number of features screened for further analysis.

Filter Algorithms

Four widely used filter algorithms were used to rank the features, assuming the features were independent of each other. *T*-test (Ttest) assumed that the data followed a normal distribution and were widely used in bioOMIC data. Ttest evaluated the statistical significance of a feature's differential values between two groups of samples (Kim, 2015; Gharbali et al., 2018; Jankowski et al., 2018). This study focused on the differential methylated residues between the RA patients and the siblings and assumed the independences between the two groups of samples (Lotsch et al., 2013; Kahl et al., 2018).

Chi-squared test (Chi2) can be used to select features with the highest values of the chi-squared statistics from a vector × relative to the classes. The chi-square test measures dependence between stochastic variables. It also checked whether a feature was statistically significantly associated with the class label under the assumption of a chi-squared distribution (Bangdiwala, 2016; Fernandez Rojas et al., 2019).

Mutual information (MI) measured the mutual dependency between a feature and the class label (Wei and Stocker, 2016; Meng et al., 2019). MI is equal to zero if and only if two random variables are independent, and a higher value means a higher dependency between the two random variables. The function relies on non-parametric methods based on entropy estimation from k-nearest-neighbor (KNN) distances.

Pearson correlation coefficient (PCC) evaluated the linear correlation between a feature and the class label with the assumption of sample independence (Liu et al., 2017). The PCC measures the linear relationship between two variables. PCC assumed that each variable be normally distributed, and do not necessarily have a zero-mean. Like the other correlation coefficients, PCC varies between −1 and +1 with 0 implying no correlation between the two variables. Correlations of −1

or +1 imply an exact negative or positive linear relationship. Positive correlations imply that as x increases, so does y. Negative correlations imply that as x increases, y decreases. The p-value roughly indicates the probability of an uncorrelated system producing variables that have a Pearson correlation at least as extreme as the one computed from these variables.

Recursive Feature Elimination Strategy

Recursive feature elimination (RFE) was a strategy to iteratively remove a feature with the least weight from the training of a classification model. The following four classification models were used to build the RFE feature selection procedure. Logistic regression (LR) (rfeLR) was a popular binary classifier and may be embedded in the RFE strategy (Pandey et al., 2018). LR is also known in the literature as logit regression, maximum-entropy classification (MaxEnt), or the log-linear classifier. In this model, the probabilities describing the possible outcomes of a single trial are modeled using a logistic function.

Lasso was a regression model and may be used to assign weights to features after a model training (rfeLasso) (Wang et al., 2019). The Lasso is a linear model that estimates sparse coefficients. It is useful in some contexts due to its tendency to prefer solutions with fewer non-zero coefficients, so Lasso can effectively reduce the number of features upon which the given solution is dependent. For this reason, Lasso and its variants are fundamental to the field of compressed sensing (Angelosante et al., 2009). Mathematically, it consists of a linear model with an added regularization term. The objective function to minimize is:

$$\min_{w} \frac{1}{2n_{\text{samples}}} ||X_{w-y}||_2^2 + \alpha||w||_1.$$

The lasso estimate thus solves the minimization of the least-squares penalty with αw_1 added, where α is a constant and w_1 is the l1-norm of the coefficient vector.

The Naïve Bayes method calculated the association probability of each feature with the class label under the assumption of inter-feature independence (rfeNBayes) (Youn and Jeong, 2009). Naive Bayes methods are a set of supervised learning algorithms based on applying Bayes' theorem with the "naive" assumption of conditional independence between every pair of features given the value of the class variable. Naive Bayes learners and classifiers can be extremely fast compared to more sophisticated methods. The decoupling of the class conditional feature distributions means that each distribution can be independently estimated as a one-dimensional distribution. This in turn helps to alleviate problems stemming from the curse of dimensionality.

The ridge regressor (rfeRidge) tried to assign minimized weights to non-associated features to a model (Barker and Brown, 2001; Rottmann and Berbeco, 2014). Ridge regression addresses some of the problems of ordinary least squares by imposing a penalty on the size of the coefficients. The ridge coefficients minimize a penalized residual sum of squares:

$$\min_{w} ||X_{w-y}||_2^2 + \alpha||w||_2^2.$$

The complexity parameter $\alpha \geq 0$ controls the amount of shrinkage: the larger the value of α, the greater the amount of shrinkage and thus the coefficients become more robust to collinearity.

Heuristic Feature Selection Strategies

Three heuristic feature selection strategies were used to generate a feature subset. The ascending feature screening (AFS) strategy started with an empty feature subset and selected the next feature with the best rank or largest weight after a model training. Then this chosen feature was removed from the remaining feature list. While the descending feature screening (DFS) strategy started with all the features and removed the next feature with the lowest rank or the least weight after a model training. Cawley and Talbot (2010) suggested that a classification model may be over-fitted if the number of training samples was smaller than that of features. We proposed a feature removal procedure BackFS to carry out an iterative removal of a feature that contributed the least prediction performance improvement. The feature subset with the best prediction performance was kept for further analysis.

All the computational experiments were conducted in the Python programming language version 3.6.5. Chi2 and MI were provided in the python sklearn version 0.19.1. PCC and Ttest were provided in the python scipy version 1.1.0. The four RFE procedures were programmed using the python sklearn version 0.19.1.

Classification Algorithms

Five widely used classifiers were utilized to measure the prediction performance of a feature subset. The discriminative power of a feature subset may be evaluated by a multivariate LR (Inzaule et al., 2018). The support vector machine (SVM) with the linear kernel function was another binary classifier that had been widely used for biomedical datasets (Citak-Er et al., 2018). SVMs are a set of supervised learning methods used for classification, regression, and outlier detection which can analyze data in classification and regression analysis. Given a set of training instances, each training instance is marked as belonging to one of the two categories, and the SVM training algorithm creates a model that assigns new instances to one of the two categories, making it a non-probability two Meta linear classifier. The SVM model represents instances as points in space, so that the mapping allows the instances of the individual categories to be separated by as wide an apparent interval as possible. Then, map new instances to the same space and predict which category they belong to based on which side of the interval they fall on. SVM may also be used to select biomarkers. After an SVM model was trained on a dataset, each input feature was assigned with a weight and the features with the default weight threshold $1e{-}5$ may be chosen for further analysis.

The simple classifier KNN had demonstrated very good prediction accuracies in some cases (Nejadgholi and Bolic, 2015; Yang et al., 2017). Neighbors-based classification is a type of instance-based learning or non-generalizing learning. It does not attempt to construct a general internal model, but simply stores instances of the training data. Classification is computed from a simple majority vote of the nearest neighbors of each point: a query point is assigned the data class which has the most representatives within the nearest neighbors of the point.

The ensembled classifier Random Forest (RF) integrated the final decision based on the prediction results of multiple random trees (Lu et al., 2017; Olsen et al., 2018; Rahman et al., 2018). The RandomForest algorithm is perturb-and-combine techniques specifically designed for trees. This means a diverse set of classifiers is created by introducing randomness in the classifier construction. The prediction of the ensemble is given as the averaged prediction of the individual classifiers. In RFs, each tree in the ensemble is built from a sample drawn with replacement (i.e., a bootstrap sample) from the training set. The Gaussian naïve Bayes classifier was used in this study as an evaluator of a feature subset (Cao et al., 2017). GaussianNB implements the Gaussian Naive Bayes algorithm for classification. The likelihood of the features is assumed to be Gaussian:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\,\sigma_y^2}}\exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right).$$

The parameters σ_y and μ_y are estimated using maximum likelihood.

The python sklearn version 0.19.1 provided the code of these five classifiers.

Performance Measurements

Three classification performance measurements, i.e., accuracy (Acc), sensitivity (Sn), and specificity (Sp), were used to evaluate how well a feature subset performed (Ye et al., 2017; Xu et al., 2018; Yokoi et al., 2018; Zhao et al., 2018). The RA children were regarded as the positive samples (P) while the matched controls were the negative samples (N). P and N were also denoted as the numbers of positive and negative samples. Sensitivity (Sn) was defined as the correctly predicted ratio of positive samples, i.e., Sn = TP/(TP + FN) = TP/P, where TP and FN were the numbers of correctly and incorrectly predicted positive samples, respectively. Specificity (Sp) was the correct prediction ratio of negative samples, i.e., Sp = TN/(TN + FP) = TN/N, where TN and FP were the numbers of negative samples with correct and incorrect predictions, respectively. The overall prediction Acc was defined as Acc = (TP + TN)/(P + N).

These measurements were used in various prediction models like the DNA and RNA functional elements (He et al., 2018; Feng et al., 2019). And they were calculated using the 10-fold cross-validation (10FCV) strategy as similar in Ye et al. (2017) and Zhao et al. (2018).

Experimental Design

The experiments were carried out in three major steps, as illustrated in **Figure 1**. The first step was to find 20,000 features with the largest variations. A methylation residue with a large variation was easier to be detected while a residue with a stable methylation level required a high-resolution technology to measure. And the downstream feature selection algorithms may crash on a dataset with a large number of features. So we have to reduce the feature dimensions to be within the capacity of the eight feature selection algorithms. So LinearSVC was used to select 147 features for further feature screening.

Then the two steps of feature selection and classification were carried out iteratively to find the best classification model using the selected features, as shown in **Figure 1**.

RESULTS AND DISCUSSION

Data Preprocessing

The raw data of this methylomic dataset was provided in the format IDAT, and was processed using the function getBeta() of the R package minfi version 1.28.3 (Aryee et al., 2014). There were 485,577 methylation features for each sample, among which 65 probes designed to interrogate SNPs within the samples and was ignored in the R package minfi. Some methylation residues had many missing values, e.g., the feature cg01550828 has no values in all the 158 samples. The feature cg01550828 was a cysteine in the N termini of the gene Ring Finger Protein 168 (RNF168), which encoded an E3 ubiquitin ligase protein. After the preprocessing, 485,511 methylomic features were detected for the following analysis.

We hypothesized that methylated residues with larger beta-value fluctuations may be easier to detect in the clinical practice. Therefore, we calculated the standard deviation of the beta-values of each methylated residue, and sorted the features in the descendental order. The top-ranked 20,000 features of the 158 samples were kept for further analysis.

Limitations the Variation Threshold 20,000

We performed the 10FCV of the classifier LinearSVC on the features with different variation thresholds, as shown in **Figure 2**. Due to that the number of features were much larger than the number of samples, only the features with the LinearSVC model weight larger than the default weight threshold $1e{-}5$ were kept for model performance evaluation. **Figure 2** demonstrated the running time and 10FCV classification Acc of different numbers of features, i.e., 1000, 2000, 3000, . . ., 22,000. As shown in the figure, the variance threshold 20,000 achieved 0.9873 in Acc while costed a very relatively small running time 17.6620 s. But the procedure of feature selection and classification was not optimized for the final classification Acc. So the other choice of variance threshold may achieve a better final classification Acc.

The evaluation procedure was carried out in a computer with the Windows 7 operating system and Python 3.7 programming language. The computer had a 3.30GHz CPU, 32 Gb memory, and 1Tb hard disk.

Optimizing LinearSVC to Select Features

Firstly, the feature selection procedure SelectFromModel() was used to find the initial feature subset with a reasonable prediction accuracy, as shown in **Figure 3**. The screening procedure was provided by the Python package scikit-learn version 0.21.2 and Python version 3.6. The penalization was carried out by the L1 penalty. In the Python package sklearn.svm.LinearSVC, the parameter *C* was a float with default = 1.0. It was a regularization parameter. The strength of the regularization was inversely

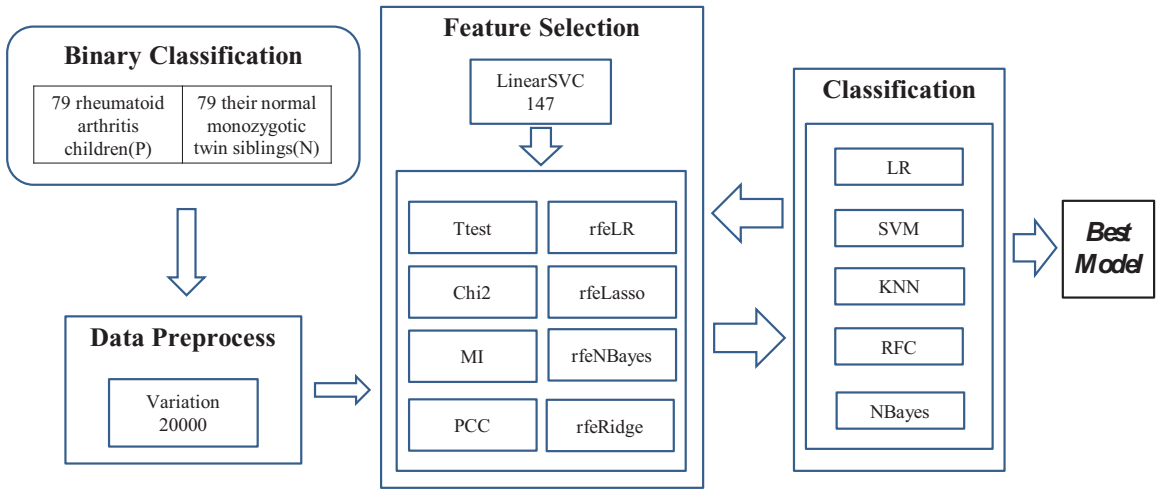


FIGURE 1 | Experiment flowchart of this study. Three major steps were carried out to find the best classification model. The first step was to find the 20,000 features with the largest variation. Then a subset of 147 features was detected using LinearSVC, and 10 feature selection algorithms were utilized to find a better feature subset. The prediction performance was evaluated using five popular binary classifiers.

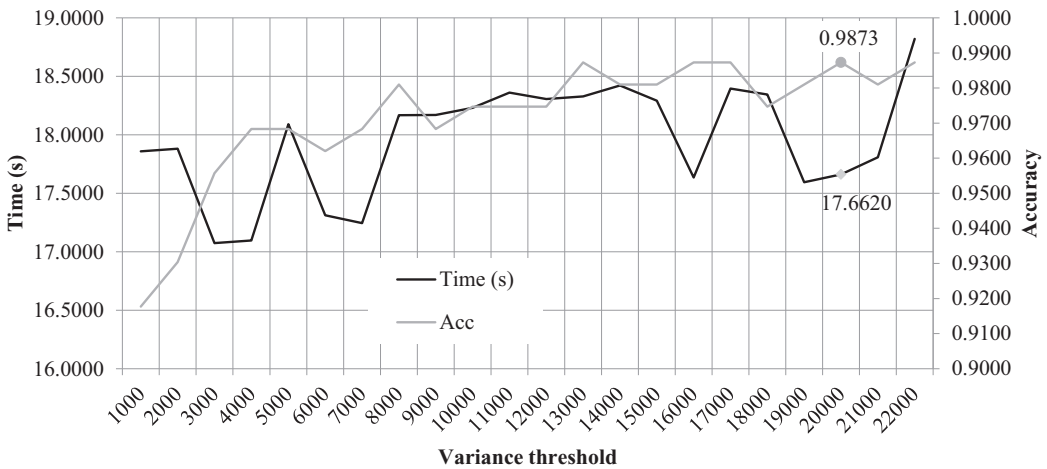


FIGURE 2 | Classification accuracy and time cost of the classifier LinearSVC with different variance thresholds. The horizontal axis was the "variance threshold." The left and right vertical axises were the computational time cost (seconds) and the classification accuracy, respectively.

proportional to C and this parameter must be strictly positive. The parameter C was screened by the values between [0.10, 5.00] with the step size 0.10, as shown in **Figure 3**.

Figure 3 demonstrated that after C reached the value 1.8, the prediction accuracy remained stable. The classifier LinearSVC achieved Acc = 0.9873 with C = 1.8 and 140 features. The best prediction accuracy 0.9937 was achieved by C = 2.4, 3.2, 3.4, 3.5, 4.3, 4.4, 4.6, and 4.7. The data demonstrated that the best Acc = 0.9937 was achieved by many choices of the parameter C, but no better performance was achieved. A smaller number of features suggested a simpler model. So C = 2.4 may be the best choice based on **Figure 3**. Its also interesting to observe that at least 155 features were chosen when C = 3.2, 3.4 and 3.5. So the following sections tried to find a smaller feature subset from this list of 147 features, which were listed in the **Supplementary Table S1**.

Selecting Features by Filters

A filter algorithm assumed the inter-feature independence and evaluated each feature separately for its association with the phenotype. So the AFS strategy selected the k-feature subset as the top-ranked k features. While the DFS strategy removed the least-ranked feature from a $(k + 1)$-feature subset based on the filter-calculated single-feature association with the class label. That is to say, the k-feature subset generated by the DFS strategy was also the top-ranked k features. The ascending and DFS strategies of a filter algorithm selected the same features for a given number of features. So this section only investigated the AFS() strategy of the four filter algorithms. The details of the

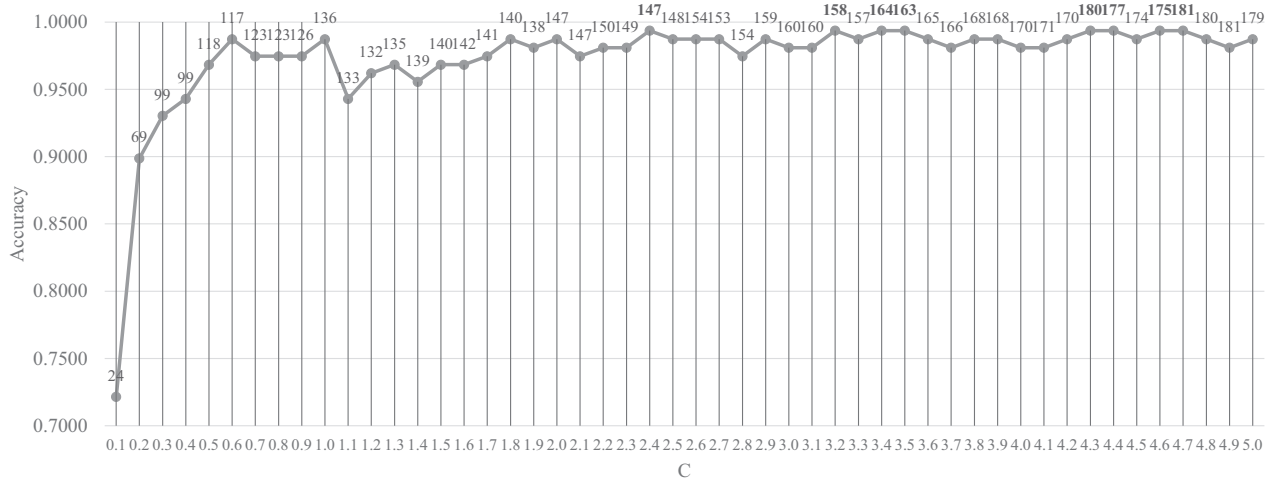


FIGURE 3 | Classification performance of the classifier LinearSVC with different values of the penalty parameter *C*. The parameter *C* was screened by the values between [0.10, 5.00] with the step size 0.10. The vertical axis was for the classification accuracy and the horizontal axis was for the values of the parameter *C*. The number of features selected was given on top of each point in the curve.

AFS strategy were described in the section "Heuristic Feature Selection Strategies."

Our data suggested that all the five classifiers performed similarly well on a feature subset with a size <50, as shown in **Figure 4**. However, the two classifiers LR and SVM kept improving the classification accuracies by adding more features. And SVM achieved the best classification accuracies on features selected by all the four filter algorithms. The best model with Acc = 1.0000 was achieved by the classifier SVM with 144 Chi2-selected methylomic features. The other three classifiers (KNN, RFC and NBayes) reached the plateau of about 0.7000 in Acc after the number of features reached 50.

Selecting Features by the RFE Strategies

We firstly evaluated the two feature selection procedures AFS(rfeLR) and DFS(rfeLR), as shown in **Supplementary Figure S1**. Filter algorithms had the assumption of the inter-feature independence. Although filters usually ran faster than the other algorithms like wrappers and RFE strategies, filters usually selected more features to achieve similar classification accuracies as the other feature selection algorithms (Srivastava et al., 2014; Suto et al., 2016).

When almost all the 147 features were kept, AFS(rfeLR) and DFS(rfeLR) performed similarly well for each of the five classifiers. The same pattern as in the previous section was observed that the two classifiers LR and SVM outperformed the other three with significantly improved accuracies, and the classifier SVM performed the best. **Supplementary Figure S1** illustrated a novel pattern that the descendent feature removal strategy (DFS) performed much better than the ascendant feature addition strategy (AFS). AFS(rfeLR) required at least 116 features to achieve Acc > 0.9000. While DFS(rfeLR) only needed 41 features to achieve Acc = 0.9114.

DFS(rfeRidge) performed even better than AFS(rfeRidge), as shown in **Figure 5** and **Supplementary Figure S4**. AFS(rfeRidge) selected 97 features to train an SVM model with Acc = 0.9051. But only 37 methylomic features were selected by DFS(rfeRidge) to train an SVM model with Acc = 0.9114. And the SVM model performed very stably with more features selected by DFS(rfeRidge), as shown in **Figure 5**. The strategy BackFS required many more features to achieve a similar prediction accuracy, as in **Figure 5C**. The classifier NBayes assumed the inter-feature independence, which may not be the case in the dataset used in this study. This might be the reason that the classifier NBayes didn't perform very well in this study, as shown in **Figure 5**.

Also, DFS(rfeLasso) performed better than AFS(rfeLasso), as shown in **Supplementary Figure S2**. AFS(rfeLasso) selected 144 features to train an SVM model with Acc = 0.9684. But 144 methylomic features were selected by DFS(rfeLasso) to train an SVM model with Acc = 0.9810. And the SVM model performed very stably with more features selected by DFS(rfeLasso).

DFS(rfeNBayes) performed similarly well for each of the five classifiers as AFS(rfeNBayes), as shown in **Supplementary Figure S3**. Both AFS(rfeNBayes) and DFS(rfeNBayes) achieved Acc = 0.9177 when selecting 101 features to train an SVM model. And the SVM model performed very stably with more features selected.

Overall, the best model achieved in this study was the SVM model (Acc = 1.0000) using the 81 features selected by the strategy DFS(rfeRidge), as shown in **Figure 5**.

Another evaluation procedure was carried out for the above-selected features. The stratified splitting strategy was used to split the samples into one-third training, one-third validation, and one-third test datasets. The SVM parameter C was evaluated for its different values from 0.1 to 3.0 with the step size 0.1, as shown in **Figure 6**. After the 81 methylomic features were selected by the strategy DFS(rfeRidge), the binary classification SVM models with different *C* values were trained on the training dataset and evaluated for the classification accuracies on the validation

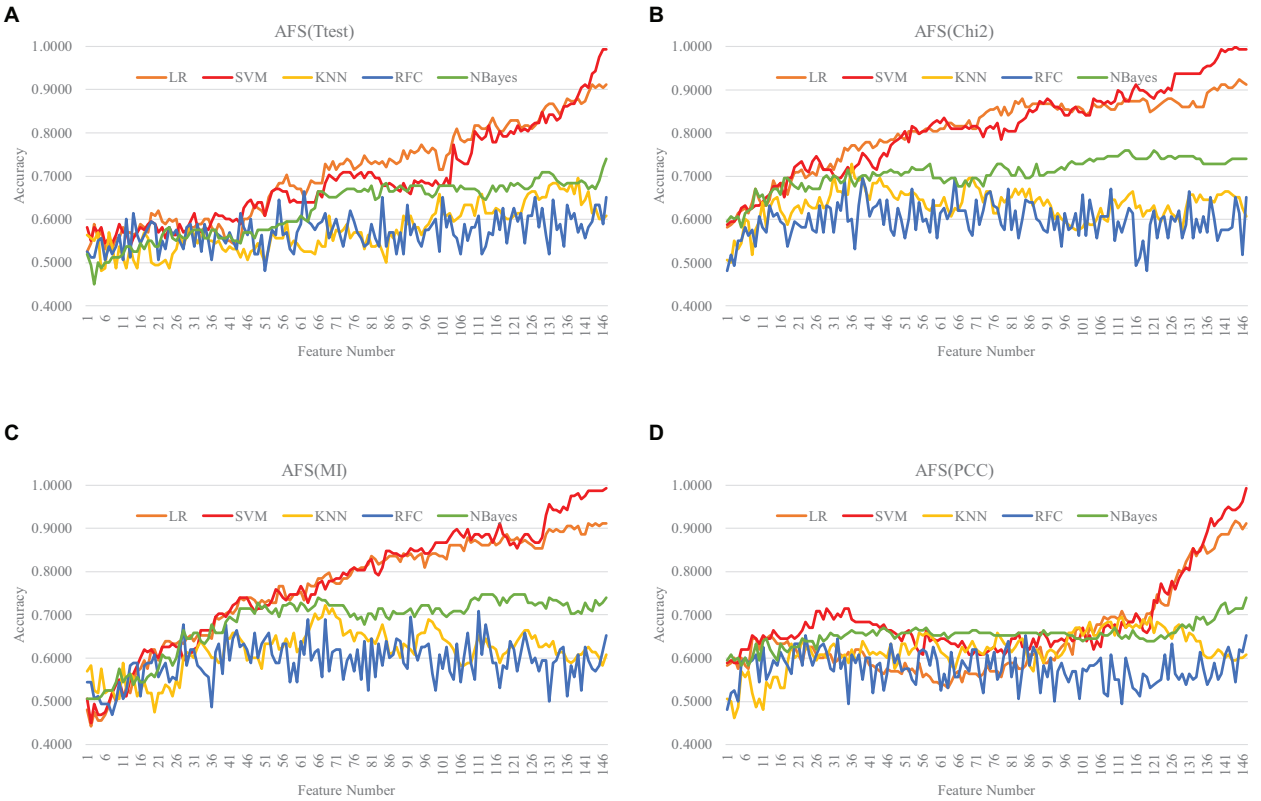


FIGURE 4 | Ascending feature screening (AFS) of four filter algorithms. The classification performance of each filter algorithm was evaluated by five classifiers. The four filter algorithms were **(A)** Ttest, **(B)** Chi2, **(C)** MI, and **(D)** PCC, and the five classifiers were LR, SVM, KNN, RFC, and NBayes.

dataset, as shown in **Figure 6**. When the parameter was 0.5, the validation accuracy reached the best value 0.8868. A similar classification accuracy 0.8679 was achieved on the test dataset. This suggested the model stability for the classification algorithm.

Refining the 147 Features With Two Other Regression Algorithms

This study evaluated how the regression-based feature selection algorithms might be improved by two other regression algorithms, i.e., sliced inverse regression (SIR) (Cook and Weisberg, 1991; Li, 1991) and group lasso (GroupLasso) (Yuan and Lin, 2006; Yuan et al., 2011). **Figure 1** demonstrated that the LinearSVC model selected 147 features and then the filters and regression-based RFE algorithms were applied. So SIR and GroupLasso were utilized to further refine the subset of 147 features.

Sliced inverse regression doesn't need to optimize the parametric or non-parametric model training process and demonstrates a significant capability to reduce the feature dimensions (Cook and Weisberg, 1991; Li, 1991). This study utilized the SIR in the Python package sliced version 0.1 (Li, 1991). Its interesting to observe that the classifier SVM from the best model achieved again Acc = 1.0000 using only the first feature engineered by SIR. Our experimental data demonstrated that SIR and the proposed feature selection procedure achieved the same classification performances on the investigated problem in this study. But the best model used only 81 original methylated residues while SIR used the one feature engineered from the 147 features.

GroupLasso is another widely used feature selection algorithm that assigns non-zero weights to groups of features instead of the individual ones like the regular lasso (Yuan and Lin, 2006; Yuan et al., 2011). This study utilized GroupLasso in the Python package group-lasso version 1.1.1 (Yuan and Lin, 2006; Yuan et al., 2011). Unfortunately no features were selected by GroupLasso.

Refining Differentially Methylated and Variable Biomarkers

Twenty differentially methylated residues were detected in the previous study, but all of them were not statistically significantly associated with RA by the adjusted *p*-values (Webster et al., 2018). This study further refined this subset of 20 methylation residues with the classification accuracy as the optimization goal.

The AFS strategy of the four filter algorithms was applied to the 20 differentially methylated residues, as shown in **Supplementary Figure S5**. The classifier NBayes achieved the best Acc = 0.7532 on the original subset of 20 features. This model may be further improved to Acc = 0.7658 using only 10 features, which was selected by the algorithm AFS(MI).

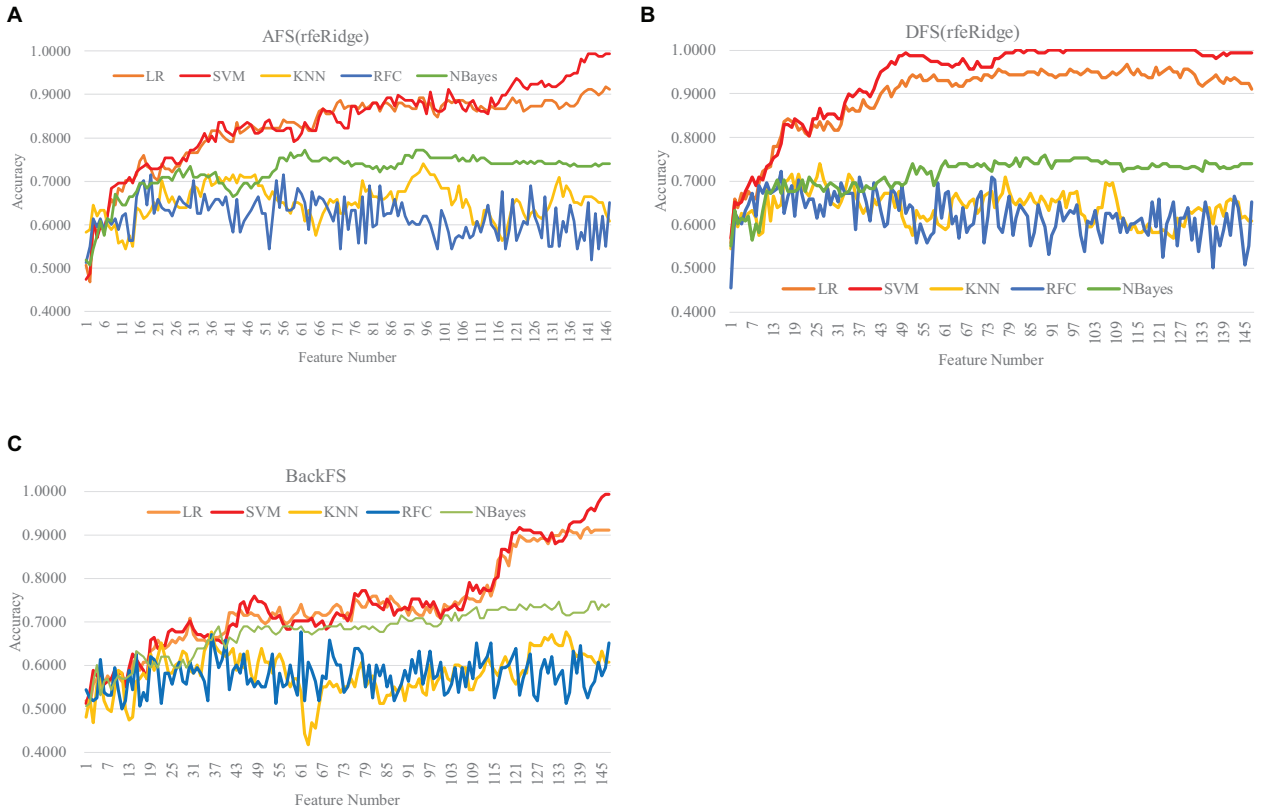


FIGURE 5 | Comparison of the three feature screening strategies of the feature selection algorithm rfeRidge. The classification performance was evaluated by five classifiers. The five classifiers were LR, SVM, KNN, RFC, and NBayes. And the three feature screening strategies were **(A)** AFS(rfeRidge), **(B)** DFS(rfeRidge), and **(C)** BackFS.

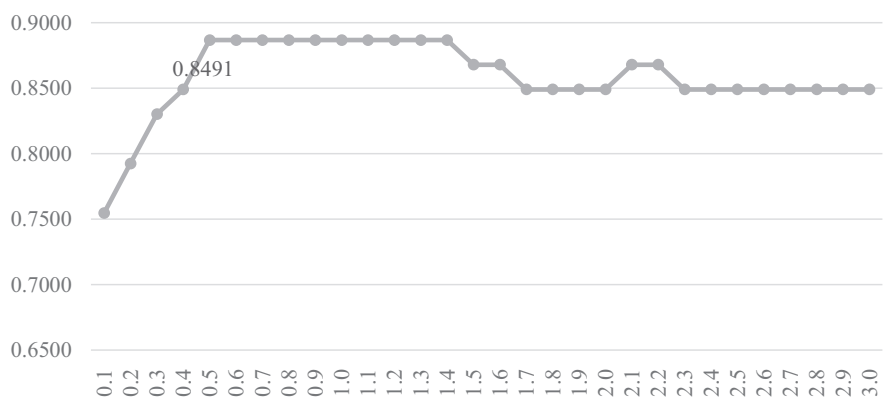


FIGURE 6 | Parameter tuning for the SVM parameter *C*. The horizontal axis was the value of the parameter *C*, and the vertical axis was the classification accuracy. Each model was trained on the training dataset and evaluated on the validation dataset.

Another algorithm AFS(Ttest) achieved the same prediction Acc = 0.7532 using only 4 and 10 features for the classifiers KNN and NBayes, respectively.

An even better improvement may be achieved by both AFS(rfeLasso) and DFS(rfeLasso), as shown in **Supplementary Figure S6**. Firstly, the original list of 20 differentially methylated residues may be reduced to 11 features to achieve Acc = 0.7658. Secondly, the best model achieved Acc = 0.8038 using only 18 features.

Webster et al. (2018) also evaluated a list of two differentially variable residues, which were refined in the same way in this study, as shown in **Supplementary Figures S7, S8**. The similar patterns were observed, and the best improved SVM model achieved Acc = 0.7722 with 12 features selected by AFS(Chi2).

Refining the Previous Biomarkers by BackFS

The two lists of RA biomarkers were further refined by a simple iterative feature elimination procedure BackFS, as shown in **Figure 7**. BackFS exhaustively removed the redundant features, so only the subset of features achieving the best prediction accuracy was kept for further analysis. The original list of 20 differentially methylated features may be further selected to achieve a better prediction Acc = 0.7658 using only 18 features for the classifier NBayes, as shown in **Figure 7A**. While the list of 20 differentially variable features may be reduced to 15 with a better prediction Acc = 0.7595 for the same classifier NBayes, as shown in **Figure 7B**.

Independent Effectiveness Evaluation of the Proposed Biomarker Detection Procedure

We further evaluated the effectiveness of the proposed biomarker detection procedure on an independent dataset. There is no simulation tool for the array-based methylomes. So another independent dataset TCGA-BRCA (Berger et al., 2018) was chosen to evaluate our biomarker detection procedure, as shown in **Figure 8**. There were 982 samples and each sample had 485,577 methylated residues. Multiple samples were extracted from some patients and only sample was randomly chosen to represent this patient. 763 samples were collected to have the clinical annotation "tumor_stage" (I/II/III/IV). The binary classification problem was formulated between the class Positive (555 samples from the stages I and II) and Negative (208 samples from the stages III and IV).

The same biomarker detection procedure was carried out on the methylomic dataset TCGA-BRCA, as shown in **Figure 6**. The initial 20,000 top-ranked features with the largest standard-deviations were screened to find the best value of the parameter *C*, as shown in **Figure 6**. The binary classification problem for the dataset TCGA-BRCA seemed to reach the classification accuracy 1.0000 with the parameter $C = 0.3$. There were 499 features selected in this step. Then the four filter algorithms were evaluated using the AFS strategy and the four RFE algorithms were evaluated by both AFS and DFS strategies, in the same procedure as the above. The features screened by DFS(rfeLR) achieved the best classification accuracy 1.0000 using only 240 features. Among the five classifiers, SVM achieved the best performance, as the same in the RA biomarker detection problem. The best feature selection algorithm DFS(rfeRidge) for the RA biomarker detection problem achieved a similar classification accuracy (0.9882) for the dataset TCGA-BRCA.

So overall the biomarker detection procedure in this study effectively detected methylated residues for the methylome-based classification problems.

Biological Observations of Methylomic Biomarkers

This study selected 81 methylated residues as biomarkers to separate the RA patients from their controls, as shown in **Supplementary Table S1**. Its interesting to observe that 38 of these 81 methylated residues were from the chromosome Y and many of them were within the transcriptional start sites (TSS) of non-coding RNA gene family Testis-Specific Transcript, Y-Linked (TTTY). This supported the observations in the literature about the gender discrepancy on autoimmune diseases like RA (Jansson and Holmdahl, 1994). Many of these methylated residues were in the TSS regions of these non-coding RNAs, suggesting that methylation may have played a regulatory role in the onset and development of RA (Relle et al., 2015; Houtman et al., 2018). Such reversible epigenetic modifications may serve as therapeutic candidates (Cribbs et al., 2015; Doody et al., 2017).

Another RA-associated gene HLA-DRB1 (Major Histocompatibility Complex, Class II, DR Beta 1) was also a methylation biomarker (cg27107292) detected in this study (Conigliaro et al., 2019; Okada et al., 2019). HLA-DRB1 was one of the first few RA biomarkers discovered four decades ago and harbored more than 100 RA-associated loci (Okada et al., 2019). Recently, HLA-DRB1 was also observed to be differentially methylated in RA (Liu et al., 2013) and had significant associations with the mortality and prognosis of RA (Ruyssen-Witrand et al., 2012; Viatte et al., 2015) and other autoimmune diseases (Bettencourt et al., 2012; Okayama et al., 2018). Furthermore, the pathway analysis through the KEGG Database (Kanehisa et al., 2017) demonstrated that various immune pathways were associated with HLA-DRB1 such as hsa04612 (Antigen processing and presentation pathway), hsa04659 (Th17 cell differentiation pathway), and hsa05323 (RA pathway). This suggested that the detected biomarker HLA-DRB1 was strongly connected to the autoimmune disease RA.

Furthermore, C5orf30 (a methylation biomarker cg17605604) was reported as a damaging regulator of tissue in RA, which is highly expressed in RA synovial fibroblast (RASF) involving joint destruction (Muthana et al., 2015). The clinical data analysis also demonstrated that the variant rs26232 in C5orf30 locus was testified to be associated with RA susceptibility and radiologic damage severity. These observations from the literature supported that C5orf30 may play a significant role in the progression of arthrosis damage (Teare et al., 2013).

Two gender-specific methylation biomarker genes DDX3Y and UTY which have been reported as sex-affected differentially expressed genes for inflammatory arthritis through the Wnt signaling (Kudryavtseva et al., 2012). This situation exactly matched to the gender-biased disease condition for RA. Besides DDX3Y was suggested to be differentially expressed in cartilage tissues of RA patients versus control groups with potential association with miRNA (Toraih et al., 2016). Many other genes like RPS4Y2, KDM5D, EIF1AY, and CYorf15A have also been shown as important biomarker genes in RA via the Monte Carlo cross-validation (Song et al., 2017).

Supplementary Table S1 also illustrated that the methylated biomarkers were from various genic sites, i.e., TSS, 5′-untranslated region (UTR), 3′-UTR, first exon, and genic body. This suggested that these RA methylation biomarkers contributed their regulatory roles through different biological mechanisms. Those frequently appeared genes, and non-coding RNA genes may need further wet-lab investigations of their potential biological mechanisms.

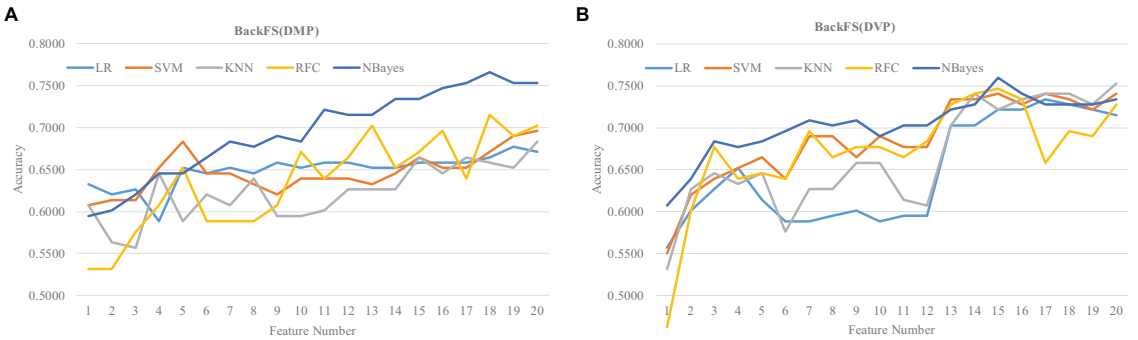


FIGURE 7 | Refine the two lists of previous methylation biomarkers of RA. The classification performance was evaluated by five classifiers. The five classifiers were LR, SVM, KNN, RFC, and NBayes. Refining procedures of **(A)** the 20 differentially methylated positions (DMP) and **(B)** the 20 differentially variable positions (DVP).

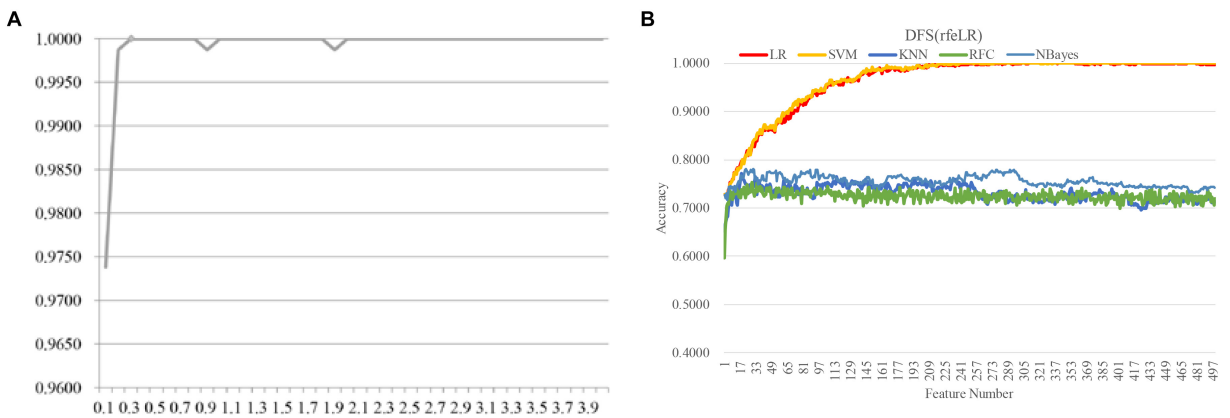


FIGURE 8 | Evaluating the proposed biomarker detection procedure with a new dataset TCGA-BRCA. The classification performance was evaluated by five classifiers. The five classifiers were LR, SVM, KNN, RFC, and NBayes. **(A)** The classification accuracies (vertical axis) of the classifier LinearSVC with different values of the parameter *C* between [0.10, 4.00] with the step size 0.10. **(B)** The classification accuracy plots of the feature selection strategy DFS(rfeLR) using the five classifiers, i.e., LR, SVM, KNN, RFC, and NBayes.

CONCLUSION

This study comprehensively utilized the widely used modeling algorithms to find the set of methylomic features with the best RA prediction accuracy. The best model used the features selected by the DFS(rfeRidge) strategy and the classifier SVM. The best accuracy 100.00% was achieved with the 81 detected methylomic biomarkers using the 10FCV strategy. The 81 methylomic biomarkers may accurately separate the RA patients from their matched controls. These biomarkers also demonstrated that chromosome Y contributed 38 methylated residues to the final model, supporting the literature about the gender-specific discrepancy. These 81 methylated biomarkers came from both regulatory regions and the gene body. So the biological mechanisms of how these 81 methylated residues were involved in RA's onset and development may vary from the transcriptional regulation to the epigenetic modifications.

The number of biomarker features was still too large for the clinical practice. Clinical data other than the methylomic features may be integrated to improve the proposed RA detection model. A weakened model may also be considered using fewer features. For example, if only 37 methylomic features selected by DFS(rfeRidge) were used to train the SVM model, the detection accuracy reached Acc = 0.9114, an acceptable accuracy in some cases. RA was a complex human disease and the subtypes may be described by fewer biomarkers. So the detection models for the RA subtypes may also use fewer biomarkers to achieve satisfying accuracies.

The samples were 70 pairs of monozygotic twins. Each twin shared the same genetic background that might reduce the noise information induced by the methylation status of genetic variations. This sample setting suggested that the detected methylomic biomarkers mainly reflected the epigenetic status of RA. Independent validation datasets might also further improve our models.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: E-MTAB-6988 at the ArrayExpress database.

AUTHOR CONTRIBUTIONS

FZ and XF conceived the project and designed the experiments. XF, XH, RS, ZX, LH, and QY wrote the codes and conducted the experiments. XF, XH, RS, and ZX generated the experimental results and drafted the discussions. FZ and XF discussed the experimental design and polished the manuscript. FZ and XF drafted and polished the manuscript. FZ, QY, and XF designed and carried out the additional experiments according to the reviewers' comments. FZ, QY, and XF also revised and polished the revised version of the manuscript.

FUNDING

This work was supported by the Jilin Provincial Key Laboratory of Big Data Intelligent Computing (20180622002JC), Jilin Science and Technology Bureau (20190104130), the Education Department of Jilin Province (JJKH20180145KJ), and the startup grant of the Jilin University. This work was also partially supported by the Bioknow MedAI Institute (BMCPP-2018-001), the High Performance Computing Center of Jilin University, and by the Fundamental Research Funds for the Central Universities, JLU.

ACKNOWLEDGMENTS

Constructive comments from the two reviewers were much appreciated.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00238/full#supplementary-material

REFERENCES

Alam, S. M., Kidwai, A. A., Jafri, S. R., Qureshi, B. M., Sami, A., Qureshi, H. H., et al. (2011). Epidemiology of rheumatoid arthritis in a tertiary care unit, Karachi, Pakistan. *J. Pak. Med. Assoc.* 61, 123–126.

Angelosante, D., Giannakis, G. B., and Grossi, E. (2009). "Compressed sensing of time-varying signals," in *Proceedings of the 2009 16th International Conference on Digital Signal Processing* (Santorini-Hellas: IEEE), 1–8.

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., et al. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 30, 1363–1369. doi: 10.1093/bioinformatics/btu049

Athar, A., Fullgrabe, A., George, N., Iqbal, H., Huerta, L., Ali, A., et al. (2019). ArrayExpress update – from bulk to single-cell expression data. *Nucleic Acids Res.* 47, D711–D715. doi: 10.1093/nar/gky964

Bangdiwala, S. I. (2016). Chi-squared statistics of association and homogeneity. *Int. J. Inj. Contr. Saf. Promot.* 23, 444–446. doi: 10.1080/17457300.2016.1228144

Barker, L., and Brown, C. (2001). Logistic regression when binary predictor variables are highly correlated. *Stat. Med.* 20, 1431–1442. doi: 10.1002/sim.680

Berger, A. C., Korkut, A., Kanchi, R. S., Hegde, A. M., Lenoir, W., Liu, W., et al. (2018). A comprehensive pan-cancer molecular study of gynecologic and breast cancers. *Cancer Cell* 33, 690–705e699. doi: 10.1016/j.ccell.2018.03.014

Bettencourt, A., da Silva, A., Pinho, E. C. P., and Martins Silva, B. (2012). Molecular genetic studies of multiple sclerosis in the portuguese population. *Acta Med. Port.* 25, 224–230.

Brandt, B., Rashidiani, S., Ban, A., and Rauch, T. A. (2019). DNA methylation-governed gene expression in autoimmune arthritis. *Int. J. Mol. Sci.* 20:5646. doi: 10.3390/ijms20225646

Cao, J., Wu, Z., Ye, W., and Wang, H. (2017). "Learning functional embedding of genes governed by pair-wised labels," in *Proceedings of the 2017 2nd IEEE International Conference on Computational Intelligence and Applications* (Beijing: IEEE), 397–401.

Carnero-Montoro, E., and Alarcon-Riquelme, M. E. (2018). Epigenome-wide association studies for systemic autoimmune diseases: the road behind and the road ahead. *Clin. Immunol.* 196, 21–33. doi: 10.1016/j.clim.2018.03.014

Cawley, G. C., and Talbot, N. L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.* 11, 2079–2107.

Chatzidionisyou, A., and Catrina, A. I. (2016). The lung in rheumatoid arthritis, cause or consequence? *Curr. Opin. Rheumatol.* 28, 76–82. doi: 10.1097/BOR.0000000000000238

Citak-Er, F., Firat, Z., Kovanlikaya, I., Ture, U., and Ozturk-Isik, E. (2018). Machine-learning in grading of gliomas based on multi-parametric magnetic resonance imaging at 3T. *Comput. Biol. Med.* 99, 154–160. doi: 10.1016/j.compbiomed.2018.06.009

Conigliaro, P., Triggianese, P., De Martino, E., Fonti, G. L., Chimenti, M. S., Sunzini, F., et al. (2019). Challenges in the treatment of rheumatoid arthritis. *Autoimmun. Rev.* 18, 706–713. doi: 10.1016/j.autrev.2019.05.007

Cook, R. D., and Weisberg, S. (1991). Sliced inverse regression for dimension reduction: comment. *J. Am. Stat. Assoc.* 86, 328–332.

Cribbs, A., Feldmann, M., and Oppermann, U. (2015). Towards an understanding of the role of DNA methylation in rheumatoid arthritis: therapeutic and diagnostic implications. *Ther. Adv. Musculoskelet. Dis.* 7, 206–219. doi: 10.1177/1759720X15598307

Crowson, C. S., Liao, K. P., Davis, J. M. III, Solomon, D. H., Matteson, E. L., Knutson, K. L., et al. (2013). Rheumatoid arthritis and cardiovascular disease. *Am. Heart J.* 166, 622.e1–628.e1. doi: 10.1016/j.ahj.2013.07.010

Doody, K. M., Bottini, N., and Firestein, G. S. (2017). Epigenetic alterations in rheumatoid arthritis fibroblast-like synoviocytes. *Epigenomics* 9, 479–492. doi: 10.2217/epi-2016-0151

Farquhar, H., Vassallo, R., Edwards, A. L., and Matteson, E. L. (2019). Pulmonary complications of rheumatoid arthritis. *Semin. Respir. Crit. Care Med.* 40, 194–207. doi: 10.1055/s-0039-1683995

Feng, X., Zhang, R., Liu, M., Liu, Q., Li, F., Yan, Z., et al. (2019). An accurate regression of developmental stages for breast cancer based on transcriptomic biomarkers. *Biomark. Med.* 13, 5–15. doi: 10.2217/bmm-2018-0305

Fernandez Rojas, R., Huang, X., and Ou, K. L. (2019). A machine learning approach for the identification of a biomarker of human pain using fNIRS. *Sci. Rep.* 9:5645. doi: 10.1038/s41598-019-42098-w

Gharbali, A. A., Najdi, S., and Fonseca, J. M. (2018). Investigating the contribution of distance-based features to automatic sleep stage classification. *Comput. Biol. Med.* 96, 8–23. doi: 10.1016/j.compbiomed.2018.03.001

He, J., Fang, T., Zhang, Z., Huang, B., Zhu, X., and Xiong, Y. (2018). PseUI: pseudouridine sites identification based on RNA sequence information. *BMC Bioinformatics* 19:306. doi: 10.1186/s12859-018-2321-0

Houtman, M., Shchetynsky, K., Chemin, K., Hensvold, A. H., Ramskold, D., Tandre, K., et al. (2018). T cells are influenced by a long non-coding RNA in the autoimmune associated PTPN2 locus. *J. Autoimmun.* 90, 28–38. doi: 10.1016/j.jaut.2018.01.003

Ibanez-Cabellos, J. S., Seco-Cervera, M., Osca-Verdegal, R., Pallardo, F. V., and Garcia-Gimenez, J. L. (2019). Epigenetic regulation in the pathogenesis of sjogren syndrome and rheumatoid arthritis. *Front. Genet.* 10:1104. doi: 10.3389/fgene.2019.01104

Inzaule, S. C., Kityo, C. M., Siwale, M., Akanmu, A. S., Wellington, M., de Jager, M., et al. (2018). Previous antiretroviral drug use compromises standard first-line HIV therapy and is mediated through drug-resistance. *Sci. Rep.* 8:15751. doi: 10.1038/s41598-018-33538-0

Jankowski, K. R. B., Flannelly, K. J., and Flannelly, L. T. (2018). The t-test: an influential inferential tool in chaplaincy and other healthcare research. *J. Health Care Chaplain.* 24, 30–39. doi: 10.1080/08854726.2017.1335050

Jansson, L., and Holmdahl, R. (1994). The Y chromosome-linked "autoimmune accelerating" yaa gene suppresses collagen-induced arthritis. *Eur. J. Immunol.* 24, 1213–1217. doi: 10.1002/eji.1830240531

Julia, A., Absher, D., Lopez-Lasanta, M., Palau, N., Pluma, A., Waite Jones, L., et al. (2017). Epigenome-wide association study of rheumatoid arthritis identifies differentially methylated loci in B cells. *Hum. Mol. Genet.* 26, 2803–2811. doi: 10.1093/hmg/ddx177

Julia, A., Gonzalez, I., Fernandez-Nebro, A., Blanco, F., Rodriguez, L., Gonzalez, A., et al. (2016). A genome-wide association study identifies SLC8A3 as a susceptibility locus for ACPA-positive rheumatoid arthritis. *Rheumatology (Oxford)* 55, 1106–1111. doi: 10.1093/rheumatology/kew035

Kahl, V. F. S., Dhillon, V. S., Simon, D., da Silva, F. R., Salvador, M., Branco, C. D. S., et al. (2018). Chronic occupational exposure endured by tobacco farmers from Brazil and association with DNA damage. *Mutagenesis* 33, 119–128. doi: 10.1093/mutage/gex045

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361. doi: 10.1093/nar/gkw1092

Khan, H., Sureda, A., Belwal, T., Cetinkaya, S., Suntar, I., Tejada, S., et al. (2019). Polyphenols in the treatment of autoimmune diseases. *Autoimmun. Rev.* 18, 647–657. doi: 10.1016/j.autrev.2019.05.001

Kim, T. K. (2015). T test as a parametric statistic. *Korean J. Anesthesiol.* 68, 540–546. doi: 10.4097/kjae.2015.68.6.540

Kudryavtseva, E., Forde, T. S., Pucker, A. D., and Adarichev, V. A. (2012). Wnt signaling genes of murine chromosome 15 are involved in sex-affected pathways of inflammatory arthritis. *Arthritis Rheum.* 64, 1057–1068. doi: 10.1002/art.33414

Kumar, N. S., and Nirmalkumar, P. (2019). A robust decision support system for wireless healthcare based on hybrid prediction algorithm. *J. Med. Syst.* 43:170. doi: 10.1007/s10916-019-1304-7

Lazzerini, P. E., Capecchi, P. L., and Laghi-Pasini, F. (2017). Systemic inflammation and arrhythmic risk: lessons from rheumatoid arthritis. *Eur. Heart J.* 38, 1717–1727. doi: 10.1093/eurheartj/ehw208

Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *J. Am. Stat. Assoc.* 86, 316–327.

Liu, A. N., Wang, L. L., Li, H. P., Gong, J., and Liu, X. H. (2017). Correlation between posttraumatic growth and posttraumatic stress disorder symptoms based on pearson correlation coefficient: a meta-analysis. *J. Nerv. Ment. Dis.* 205, 380–389. doi: 10.1097/NMD.0000000000000605

Liu, Y., Aryee, M. J., Padyukov, L., Fallin, M. D., Hesselberg, E., Runarsson, A., et al. (2013). Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat. Biotechnol.* 31, 142–147. doi: 10.1038/nbt.2487

Lopez-Isac, E., Martin, J. E., Assassi, S., Simeon, C. P., Carreira, P., Ortego-Centeno, N., et al. (2016). Brief report: IRF4 newly identified as a common susceptibility locus for systemic sclerosis and rheumatoid arthritis in a cross-disease meta-analysis of genome-wide association studies. *Arthritis Rheumatol.* 68, 2338–2344. doi: 10.1002/art.39730

Lopez-Mejias, R., Carmona, F. D., Genre, F., Remuzgo-Martinez, S., Gonzalez-Juanatey, C., Corrales, A., et al. (2019). Identification of a 3'-untranslated genetic variant of RARB associated with carotid intima-media thickness in rheumatoid arthritis: a genome-wide association study. *Arthritis Rheumatol.* 71, 351–360. doi: 10.1002/art.40734

Lotsch, D., Ghanim, B., Laaber, M., Wurm, G., Weis, S., Lenz, S., et al. (2013). Prognostic significance of telomerase-associated parameters in glioblastoma: effect of patient age. *Neuro Oncol.* 15, 423–432. doi: 10.1093/neuonc/nos329

Lu, S., Xia, Y., Cai, W., Fulham, M., Feng, D. D., and Neuroimaging Initiative, (2017). Early identification of mild cognitive impairment using incomplete random forest-robust support vector machine and FDG-PET imaging. *Comput. Med. Imaging Graph* 60, 35–41. doi: 10.1016/j.compmedimag.2017.01.001

Lyu, H., Wan, M., Han, J., Liu, R., and Wang, C. (2017). A filter feature selection method based on the maximal information coefficient and gram-schmidt orthogonalization for biomedical data mining. *Comput. Biol. Med.* 89, 264–274. doi: 10.1016/j.compbiomed.2017.08.021

Massey, J., Plant, D., Hyrich, K., Morgan, A. W., Wilson, A. G., Spiliopoulou, A., et al. (2018). Genome-wide association study of response to tumour necrosis factor inhibitor therapy in rheumatoid arthritis. *Pharmacogenomics J.* 18, 657–664. doi: 10.1038/s41397-018-0040-6

Meng, C., Wang, Q., Guan, S., Sun, K., and Liu, B. J. I. A. (2019). 2D-3D registration with weighted local mutual information in vascular interventions. *IEEE Access* 7, 162629–162638. doi: 10.1109/access.2019.2905345

Moayyeri, A., Hammond, C. J., Valdes, A. M., and Spector, T. D. (2013). Cohort profile: twinsuk and healthy ageing twin study. *Int. J. Epidemiol.* 42, 76–85. doi: 10.1093/ije/dyr207

Muthana, M., Hawtree, S., Wilshaw, A., Linehan, E., Roberts, H., Khetan, S., et al. (2015). C5orf30 is a negative regulator of tissue damage in rheumatoid arthritis. *Proc. Natl. Acad. Sci. U.S.A.* 112, 11618–11623. doi: 10.1073/pnas.1501947112

Nejadgholi, I., and Bolic, M. (2015). A comparative study of PCA, SIMCA and Cole model for classification of bioimpedance spectroscopy measurements. *Comput. Biol. Med.* 63, 42–51. doi: 10.1016/j.compbiomed.2015.05.004

Okada, Y., Eyre, S., Suzuki, A., Kochi, Y., and Yamamoto, K. (2019). Genetics of rheumatoid arthritis: 2018 status. *Ann. Rheum. Dis.* 78, 446–453. doi: 10.1136/annrheumdis-2018-213678

Okayama, T., Hashiguchi, Y., Kikuyama, H., Yoneda, H., and Kanazawa, T. (2018). Next-generation sequencing analysis of multiplex families with atypical psychosis. *Transl. Psychiatry* 8:221. doi: 10.1038/s41398-018-0272-x

Olsen, R. M., Aasvang, E. K., Meyhoff, C. S., and Dissing Sorensen, H. B. (2018). Towards an automated multimodal clinical decision support system at the post anesthesia care unit. *Comput. Biol. Med.* 101, 15–21. doi: 10.1016/j.compbiomed.2018.07.018

Pandey, G., Pandey, O. P., Rogers, A. J., Ahsen, M. E., Hoffman, G. E., Raby, B. A., et al. (2018). A nasal brush-based classifier of asthma identified by machine learning analysis of nasal RNA sequence data. *Sci. Rep.* 8:8826.

Petralia, M. C., Mazzon, E., Basile, M. S., Cutuli, M., Di Marco, R., Scandurra, F., et al. (2019). Effects of treatment with the hypomethylating agent 5-aza-2'-deoxycytidine in murine type II collagen-induced arthritis. *Pharmaceuticals (Basel)* 12:174. doi: 10.3390/ph12040174

Rahman, M. M., Bhuiyan, M. I. H., and Hassan, A. R. (2018). Sleep stage classification using single-channel EOG. *Comput. Biol. Med.* 102, 211–220. doi: 10.1016/j.compbiomed.2018.08.022

Relle, M., Foehr, B., and Schwarting, A. (2015). Epigenetic aspects of systemic lupus erythematosus. *Rheumatol. Ther.* 2, 33–46. doi: 10.1007/s40744-015-0014-y

Rottmann, J., and Berbeco, R. (2014). Using an external surrogate for predictor model training in real-time motion management of lung tumors. *Med. Phys.* 41:121706. doi: 10.1118/1.4901252

Ruyssen-Witrand, A., Constantin, A., Cambon-Thomsen, A., and Thomsen, M. (2012). New insights into the genetics of immune responses in rheumatoid arthritis. *Tissue Antigens* 80, 105–118. doi: 10.1111/j.1399-0039.2012.01939.x

Shadrina, A., Tsepilov, Y., Sokolova, E., Smetanina, M., Voronina, E., Pakhomov, E., et al. (2018). Genome-wide association study in ethnic Russians suggests an association of the MHC class III genomic region with the risk of primary varicose veins. *Gene* 659, 93–99. doi: 10.1016/j.gene.2018.03.039

Singh, N. P., Bapi, R. S., and Vinod, P. K. (2018). Machine learning models to predict the progression from early to late stages of papillary renal cell carcinoma. *Comput. Biol. Med.* 100, 92–99. doi: 10.1016/j.compbiomed.2018.06.030

Smolen, J. S., Aletaha, D., and McInnes, I. B. (2016). Rheumatoid arthritis. *Lancet* 388, 2023–2038. doi: 10.1016/S0140-6736(16)30173-8

Song, W., Zhang, Y. M., Ma, T., Wang, J., and Wang, K. Z. (2017). Identification of significant pathway cross-talk in rheumatoid arthritis by the Monte Carlo cross-validation method. *Genet. Mol. Res.* 16:gmr16029142. doi: 10.4238/gmr16029142

Srivastava, B., Srivastava, R., and Jangid, M. (2014). "Filter vs. wrapper approach for optimum gene selection of high dimensional gene expression dataset: an analysis with cancer datasets," in *Proceedings of the 2014 International Conference on High Performance Computing and Applications* (Bhubaneswar: IEEE), 1–6.

Suto, J., Oniga, S., and Sitar, P. P. (2016). "Comparison of wrapper and filter feature selection algorithms on human activity recognition," in *Proceedings of the 2016 6th International Conference on Computers Communications and Control* (Oradea: IEEE), 124–129.

Taylor, J. C., Bongartz, T., Massey, J., Mifsud, B., Spiliopoulou, A., Scott, I. C., et al. (2018). Genome-wide association study of response to methotrexate in early rheumatoid arthritis patients. *Pharmacogenomics J.* 18, 528–538. doi: 10.1038/s41397-018-0025-5

Teare, M. D., Knevel, R., Morgan, M. D., Kleszcz, A., Emery, P., Moore, D. J., et al. (2013). Allele-dose association of the C5orf30 rs26232 variant with joint damage in rheumatoid arthritis. *Arthritis Rheum.* 65, 2555–2561. doi: 10.1002/art.38064

Tekin Erguzel, T., Tas, C., and Cebi, M. (2015). A wrapper-based approach for feature selection and classification of major depressive disorder-bipolar disorders. *Comput. Biol. Med.* 64, 127–137. doi: 10.1016/j.compbiomed.2015.06.021

Toraih, E. A., Ismail, N. M., Toraih, A. A., Hussein, M. H., and Fawzy, M. S. (2016). Precursor miR-499a variant but not miR-196a2 is associated with rheumatoid arthritis susceptibility in an Egyptian population. *Mol. Diagn. Ther.* 20, 279–295. doi: 10.1007/s40291-016-0194-3

Triantafyllias, K., De Blasi, M., Hoffmann, I., Thomaidis, T., Drees, P., and Schwarting, A. (2016). The count of tender rather than swollen joints correlates with aortic stiffness in patients with rheumatoid arthritis. *Springerplus* 5:428. doi: 10.1186/s40064-016-2066-z

Verde, L., and De Pietro, G. (2019). A neural network approach to classify carotid disorders from heart rate variability analysis. *Comput. Biol. Med.* 109, 226–234. doi: 10.1016/j.compbiomed.2019.04.036

Viatte, S., Plant, D., Han, B., Fu, B., Yarwood, A., Thomson, W., et al. (2015). Association of HLA-DRB1 haplotypes with rheumatoid arthritis severity, mortality, and treatment response. *JAMA* 313, 1645–1656. doi: 10.1001/jama.2015.3435

Villanueva-Romero, R., Gutierrez-Canas, I., Carrion, M., Perez-Garcia, S., Seoane, I. V., Martinez, C., et al. (2018). The anti-inflammatory mediator, vasoactive intestinal peptide, modulates the differentiation and function of Th subsets in rheumatoid arthritis. *J. Immunol. Res.* 2018:6043710. doi: 10.1155/2018/6043710

Wang, Y., Deng, H., Xin, S., Zhang, K., Shi, R., and Bao, X. (2019). Prognostic and predictive value of three DNA methylation signatures in lung adenocarcinoma. *Front. Genet.* 10:349. doi: 10.3389/fgene.2019.00349

Webster, A. P., Plant, D., Ecker, S., Zufferey, F., Bell, J. T., Feber, A., et al. (2018). Increased DNA methylation variability in rheumatoid arthritis-discordant monozygotic twins. *Genome Med.* 10:64. doi: 10.1186/s13073-018-0575-9

Wei, X. X., and Stocker, A. A. (2016). Mutual information, fisher information, and efficient coding. *Neural Comput.* 28, 305–326. doi: 10.1162/NECO_a_00804

Wu, C., Chen, J., Liu, Y., and Hu, X. (2019). Improved prediction of regulatory element using hybrid abelian complexity features with DNA sequences. *Int. J. Mol. Sci.* 20:1704. doi: 10.3390/ijms20071704

Xie, J., Lei, J., Xie, W., Shi, Y., and Liu, X. (2013). Two-stage hybrid feature selection algorithms for diagnosing erythemato-squamous diseases. *Health Inf. Sci. Syst.* 1:10. doi: 10.1186/2047-2501-1-10

Xu, C., Liu, J., Yang, W., Shu, Y., Wei, Z., Zheng, W., et al. (2018). An OMIC biomarker detection algorithm TriVote and its application in methylomic biomarker detection. *Epigenomics* 10, 335–347. doi: 10.2217/epi-2017-0097

Yang, C. H., Weng, Z. J., Chuang, L. Y., and Yang, C. S. (2017). Identification of SNP-SNP interaction for chronic dialysis patients. *Comput. Biol. Med.* 83, 94–101. doi: 10.1016/j.compbiomed.2017.02.004

Ye, Y., Zhang, R., Zheng, W., Liu, S., and Zhou, F. (2017). RIFS: a randomly restarted incremental feature selection algorithm. *Sci. Rep.* 7:13013. doi: 10.1038/s41598-017-13259-6

Yokoi, A., Matsuzaki, J., Yamamoto, Y., Yoneoka, Y., Takahashi, K., Shimizu, H., et al. (2018). Integrated extracellular microRNA profiling for ovarian cancer screening. *Nat. Commun.* 9:4319. doi: 10.1038/s41467-018-06434-4

Youn, E., and Jeong, M. K. (2009). Class dependent feature scaling method using naive Bayes classifier for text datamining. *Pattern Recognit. Lett.* 30, 477–485. doi: 10.1016/j.patrec.2008.11.013

Yuan, L., Liu, J., and Ye, J. (2011). Efficient methods for overlapping group lasso. *Adv. Neural Inf. Process. Syst.* 35, 352–360.

Yuan, M., and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Series B Stat. Methodol.* 68, 49–67. doi: 10.1111/j.1467-9868.2005.00532.x

Zhao, R., Zhang, R., Tang, T., Feng, X., Li, J., Liu, Y., et al. (2018). TriZ-a rotation-tolerant image feature and its application in endoscope-based disease diagnosis. *Comput. Biol. Med.* 99, 182–190. doi: 10.1016/j.compbiomed.2018.06.006

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Feng, Hao, Shi, Xia, Huang, Yu and Zhou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 18 June 2020
doi: 10.3389/fgene.2020.00648

New Analysis Framework Incorporating Mixed Mutual Information and Scalable Bayesian Networks for Multimodal High Dimensional Genomic and Epigenomic Cancer Data

*Xichun Wang, Sergio Branciamore, Grigoriy Gogoshin, Shuyu Ding and Andrei S. Rodin**

Department of Computational and Quantitative Medicine, Beckman Research Institute and Diabetes and Metabolism Research Institute of the City of Hope, Duarte, CA, United States

OPEN ACCESS

Edited by:
Yun Liu,
Fudan University, China

Reviewed by:
Asaf Salamov,
Lawrence Berkeley National Laboratory, United States
Weihao Gao,
University of Illinois at Urbana-Champaign, United States

***Correspondence:**
Andrei S. Rodin
arodin@coh.org

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *27 February 2019*
Accepted: *28 May 2020*
Published: *18 June 2020*

Citation:
Wang X, Branciamore S, Gogoshin G, Ding S and Rodin AS (2020) New Analysis Framework Incorporating Mixed Mutual Information and Scalable Bayesian Networks for Multimodal High Dimensional Genomic and Epigenomic Cancer Data. Front. Genet. 11:648. doi: 10.3389/fgene.2020.00648

We propose a novel two-stage analysis strategy to discover candidate genes associated with the particular cancer outcomes in large multimodal genomic cancers databases, such as The Cancer Genome Atlas (TCGA). During the first stage, we use mixed mutual information to perform variable selection; during the second stage, we use scalable Bayesian network (BN) modeling to identify candidate genes and their interactions. Two crucial features of the proposed approach are (i) the ability to handle mixed data types (continuous and discrete, genomic, epigenomic, etc.) and (ii) a flexible boundary between the variable selection and network modeling stages — the boundary that can be adjusted in accordance with the investigators' BN software scalability and hardware implementation. These two aspects result in high generalizability of the proposed analytical framework. We apply the above strategy to three different TCGA datasets (LGG, Brain Lower Grade Glioma; HNSC, Head and Neck Squamous Cell Carcinoma; STES, Stomach and Esophageal Carcinoma), linking multimodal molecular information (SNPs, mRNA expression, DNA methylation) to two clinical outcome variables (tumor status and patient survival). We identify 11 candidate genes, of which 6 have already been directly implicated in the cancer literature. One novel LGG prognostic factor suggested by our analysis, methylation of TMPRSS11F type II transmembrane serine protease, presents intriguing direction for the follow-up studies.

Keywords: The Cancer Genome Atlas, Bayesian networks, multimodal big data, variable selection, mixed mutual information, methylation, genomic and epigenomic molecular data

INTRODUCTION

The Cancer Genome Atlas (TCGA) resource contains genomic data compiled for more than 30 different types/subtypes of cancer (Tomczak et al., 2015). For each type, clinical outcome/progression data (e.g., tumor status and patient survival) for a considerable number of patients is matched to the large-scale molecular data. The latter is multimodal, ranging from genetic (e.g., somatic mutations) to expression (e.g., RNA-seq gene expression) to epigenetic

(e.g., promoter methylation) data. Not surprisingly, there is substantial enthusiasm for causally linking the latter to the former using various modeling and secondary data analysis techniques (Jeong et al., 2015; Phan et al., 2016; Hou et al., 2018; Tian et al., 2018; Xu et al., 2018). The ultimate goals of these analyses are (i) to gain better mechanistic understanding of the underlying molecular biology of cancer, primarily by identifying important genes and their interactions; (ii) to construct compact and efficient clinical predictors (e.g., prognostic scores, indices and signatures); (iii) to associate the latter with the particular patient groups and subgroups, in the context of personalized/precision medicine. One of the more attractive and popular methods for such multivariate analysis is Bayesian networks (BNs) (Heckerman, 1995), a well-established fixture in computational systems biology (Friedman et al., 2000). Among the BN advantages are their probabilistic nature, model flexibility, ability to handle non-additive, higher-order, interactions, and ease of the result interpretation. However, applications of BNs to the TCGA (and TCGA-like) data (Gevaert et al., 2006; Xu et al., 2012, 2014; Wang et al., 2013; Huang et al., 2015; Zhu et al., 2015; Kaiser et al., 2016; Wu et al., 2017) face two principal difficulties: combining mixed data types in a single analysis framework, *and* achieving sufficient (for genomic data) scalability, *simultaneously*. (These, of course, are the two fundamental, and interconnected, BN modeling challenges in general, not just in the TCGA application). The latest developments in addressing these two challenges encompass more efficient computational approaches (Gogoshin et al., 2017; Ramsey et al., 2017), and mathematically rigorous and robust methods for handling mixed data, such as mixed local probability models and/or adaptive discretization (Gogoshin et al., 2017; Andrews et al., 2018; Sedgewick et al., 2018). Nevertheless, resolving both difficulties simultaneously in a generalizable toolkit (seamlessly applicable, for example, across the individual TCGA datasets) remains elusive. A promising approach to devising such a toolkit would be to precede the comparatively exhaustive NP-hard BN modeling with a variable selection procedure [for example (Zhang et al., 2014)], where the full dataset is pared down to a subset of variables most relevant to a particular clinical outcome or phenotype. While alleviating the scalability issue, this, however, could potentially "throw away the wheat with the chaff," especially if the variable selection process (Blum and Langley, 1997; Guyon and Elisseeff, 2003) is of a simplistic and overly too restrictive kind (e.g., a statistically conservative univariate filter). There are three possible ways to address this, namely: (i) increase the scalability of the BN modeling to genomic data levels (possible, but impractical for frequent/serial analyses), (ii) incorporate higher-order interactions into the variable selection step (thus "upgrading" it from the simple filter to the wrapper [Kohavi and John, 1997; Guyon and Elisseeff, 2003; Leng et al., 2010) — this is the solution implemented in Zhang et al. (2014)], or (iii) adjust the transition boundary between the variable selection step and the BN modeling step, depending on the investigators' computational resources and the nature (dimensionality, sparseness, heterogeneity) of the actual data. It is the third analytical strategy that we propose in this study, with the goal to achieve the optimal compromise between the computational practicality and modeling exhaustiveness.

In our analysis pipeline, we start with the variable selection procedure based on the mixed-type Mixed Mutual Information (MMI) forward selection filter. We compute the MMI values for all available gene-outcome (specifically, tumor status and patient survival) pairs, and use the MMI frequency distribution to select top variables/genes (or, alternatively, to remove bottom variables/genes) before moving on to the BN modeling. This mixed-type measure-based approach to gene selection is the principal innovation of this paper. We then use the maximum entropy (ME) – based discretization to construct the mixed-type BNs using our previously reported scalable BN modeling algorithm and software (Gogoshin et al., 2017). Subsequently, we concentrate on the sub-networks centered around the clinical outcome variables of interest, and identify the molecular gene components belonging to these sub-networks.

The proposed analysis strategy has been applied by us to 12 different TCGA cancer datasets. This allowed us to check for robustness, scalability and generalizability. Here, we present the results for the Brain Lower Grade Glioma (LGG), Head and Neck Squamous Cell Carcinoma (HNSC) and Stomach and Esophageal Carcinoma (STES) datasets (all three datasets being reasonably well-populated and proportionally balanced across the different outcomes and molecular data types). For the purposes of this particular analysis, we decided to concentrate on three types of molecular data, one discrete (somatic mutations) and two – continuous (RNA-seq gene expression, and promoter methylation). This selection is reflective of the recent trends in multimodal cancer data analyses (Zhang et al., 2014; Yoo et al., 2017), makes sense in the broad cancer genetics context (Phipps et al., 2016; Fang et al., 2017; Liang et al., 2017; Rajesh et al., 2017; Zhang C. et al., 2017; Koch et al., 2018), and underscores the comparative importance of the methylation molecular data (Koch et al., 2018). While focusing solely on the gene-centric modalities is inherently limiting (many disease-linked SNPs are localized in the non-coding regions), one of the primary purposes of this study was to showcase the MMI approach (enjoining three different modalities in a single measure/score), which necessitated the gene-centric analysis. In future, we plan to generalize our analytical framework to other, non-gene-centric, data.

We conclude by identifying a compact list of genes potentially associated with cancer-related clinical phenotypes (tumor status and patient survival), scrutinizing these genes in light of the current literature, and discussing the generalizability of our approach to the different datasets, diseases and molecular data types.

MATERIALS AND METHODS

Data Preprocessing

The Cancer Genome Atlas, LGG, HNSC, and STES datasets were downloaded for the clinical data ["Clinical_Pick_Tier1

(MD5)"], SNP data ["Mutation_Packager_Calls (MD5)"], expression data ["mRNAseq_Preprocess (MD5)"] and promoter-centric methylation data ["Methylation_Preprocess (MD5)"]. Patients were further subdivided into (i) two disease progression categories (according to the "tumor status" variable), and (ii) two patient survival categories (high death risk, with survival less than 2 years, and low death risk, with survival more than 2 years, which is a common cutoff point in recent cancer literature). We further excluded patients with ambiguous or missing outcome variable values (e.g., no survival status, survival status as "living" with survival time less than 2 years, tumor status neither "tumor-free" nor "with tumor," etc.). These clinical variables ("tumor status" and "2-year survival") were subsequently used for the variable selection purposes, and, eventually, to extract "tumor status" and "survival" – centered sub-networks from the full BNs. Expression data and methylation data (designated by "E" and "M" below, for brevity) were not discretized at this stage, as both variable selection and BN construction tools in our computational pipeline can, by design, accept mixed (continuous and discreet) variable types. SNP (somatic mutation) data (designated by "S" below) were compressed into a binary variable (presence or absence of at least one non-synonymous mutation in at least one sample of the particular gene).

After filtering out patient records with incomplete, partially missing, or ambiguously labeled data, the final datasets consisted of 4782 genes (LGG), 12516 genes (HNSC) and 16164 genes (STES). 273 patient records were available for LGG/tumor status analysis (140 patients with tumor, 133 without); 213 patients – for LGG/survival (120 patients with survival less than 2 years, 93 with long-term survival). Similarly, 260 patient records were available for HNSC/tumor status analysis (94 patients with tumor, 166 without); 139 patients – for HNSC/survival (40 patients with survival less than 2 years, 99 with long-term survival). Finally, 403 patient records were available for STES/tumor status analysis (147 patients with tumor, 256 without); 258 patients – for STES/survival (191 patients with survival less than 2 years, 67 with long-term survival).

Here we would like to re-emphasize that it is possible to include other different molecular data types and outcome variables, both continuous and discrete, into the proposed framework without substantial alterations to the analysis pipeline, except for some rudimentary data preprocessing.

Variable Selection

There are very few BN algorithms/software solutions that scale up to (epi)genomic levels (tens to hundreds of thousands of variables) (Gogoshin et al., 2017; Ramsey et al., 2017). Even with these, exhaustive analyses require dedicated hardware and weeks of processing time. This might be acceptable for a one-off, "final" analysis, but is clearly impractical for the exploratory research. This is why it is a common practice to carry out variable selection (or feature selection, or feature set reduction) in order to generate a comparatively compact subset of variables to be subsequently fed into the network modeling algorithm/software (Guyon and Elisseeff, 2003). Variable selection approaches range from the very simple (univariate filters) to increasingly more sophisticated; at some point, the latter become essentially indistinguishable from the multivariate modeling methods *per se*. Depending on the dataset to be analyzed, different "couplings" of variable selection and multivariate modeling methods might prove to be more or less effective, and it is difficult to devise *a priori* the objectively optimal combination for each new dataset. For a principally network-centric data analysis approach (innate to the systems biology), it would make sense to feed as many variables into the network-building module as possible, thus "delegating" the resolution of the higher-order / non-additive interactions and conditional independence relationships to the BN algorithm itself. Therefore, for the exploratory research, we suggest that the investigators first define the upper BN scalability limit that they are comfortable with (given the available software/hardware), and then adjust the variable selection cutoff point accordingly. For more "finalized" analysis, that limit should be raised higher (and the variable selection process, consequently, be made less restrictive).

In TCGA dataset (and other similar (epi)genomic resources), there are tens of thousands of potentially predictive/relevant variables (roughly proportional to the number of genes in the human genome). The "hand off" point between the variable selection and BN analysis steps should therefore vary between 100s of variables (for the exploratory and preliminary analyses) and 1,000s of variables (for the final analyses). The actual number might also depend on the shape of the variable selection curve, or on the statistical significance criteria–we stop adding increasingly less significant variables during the forward variable selection process (or stop removing increasingly more significant variables during the backward variable elimination process) when a certain statistical significance cutoff point is reached (Rodin et al., 2009). The above considerations were taken into account in the course of this study, as detailed in the section "Results" below.

It is difficult to integrate the multimodal, mixed-type, data into the variable selection process (filter or wrapper) as, until recently, there has been a paucity of the usable mixed-type metrics. In this study, a recently developed measure, Mixed Mutual Information (MMI) (Gao et al., 2018), was used to link the gene information (a mixed-type vector consisting of the S, E, and M molecular data components for each gene) to the clinical variable (tumor status or 2-year survival) in a "forward-selection-filter" variable selection procedure. MMI is a non-parametric and distribution-free measure [which makes it more attractive than the alternatives, such as linear correlation – especially in the biological networks context (Margolin et al., 2006; Asur et al., 2007)] that is based on the entropy estimates from k-nearest neighbor (k-NN) distances (Kraskov et al., 2004). It is, therefore, sensitive to the choice of the k parameter. Lower values of k (1–4) tend to lead to higher dispersion, while much higher values (>20) are associated with unnecessarily increased computational complexity and possible overfitting [, personal communication from Gao et al. (2018)]. We have evaluated different values of k on the actual TCGA datasets by measuring the Jaccard index for the pairs of consecutive (in k) post-selection variable sets as a function of k. The index appeared to stabilize in the 8–20 range in 12 different TCGA datasets

analyzed (see section "Results" below); therefore, k was set at 15 throughout this study.

Bayesian Networks Modeling

Bayesian networks modeling, in its basic form, reconstructs a sparse graphical representation of a joint multivariate probability distribution of random variables from a "flat" dataset. Nodes in the network represent random variables, edges – dependencies. Absence of an edge between the two nodes indicates conditional independence between them. Recent work in BN methodology refinement led to significant progress in scalability – our latest BN modeling software implementation (Gogoshin et al., 2017) easily processes datasets up to $\sim$ 1 mln variables $\times$ 1 mln datapoints. Handling mixed variable types (both continuous and discrete, in a typical application) is still not entirely seamless; it was recently suggested (Gogoshin et al., 2017; Andrews et al., 2018) that adaptive discretization (of continuous variables) might be preferable to forcing mixed local probability models. Consequently, we were using maximum entropy – based three-bin discretization throughout this study – expression data ("E" molecular data component) and methylation data ("M" molecular data component) were discretized into three bins – which has attractive mathematical properties, and has been shown by us earlier to maintain near-optimal over/under-fitting balance (Gogoshin et al., 2017).

Detailed description of the BN methodology in general and of our implementation (including applications to other types of high-dimensional biological data) in particular can be found in Gogoshin et al. (2017), Zhang X. et al. (2017); here we will only note that (1) our BN implementation uses a hybrid "sparse candidates" + "search-and-score" graduate descent algorithm coupled with various model scoring metrics and maximum entropy-based adaptive discretization; (2) in the resulting BN visualizations, numbers next to the edges and edge "thickness" indicate relative edge strengths (the numbers are the model scores' ratios for the models with/without corresponding edges, which are proportional to the marginal likelihood ratios); (3) directionality in the network (arrow points attached to the edges, when present) does not necessarily imply the causality flow, and is used predominantly for the mathematical convenience (to avoid cyclic dependencies); (4) when deciphering conditional dependence and independence patterns, it is useful to concentrate on the immediate Markov neighborhood (MN) of a particular variable of interest (such as a clinical outcome). This neighborhood can be roughly defined as all the nodes that are in immediate contact with ("one degree of separation" from) the node representing the aforementioned variable of interest. Under certain conditions, given its MN, the variable of interest is conditionally independent of the remaining variables (rest of the network). Therefore, deriving a MN for a variable of interest is analogous to the variable selection activity, specifically of the embedded variety (Guyon and Elisseeff, 2003). The central step in our computational analysis pipeline is using full BN reconstruction to generate the MN for the clinical outcome variable, and then ascertaining the interplay of the (small number of) gene-related variables (S, E and M molecular data components) within that MN. (It should be noted that MN is a simplification of the more rigorous concept of Markov Blanket – meaning, for our purposes, that sometimes "two degrees of separation" are needed for encapsulating a variable/node of interest).

RESULTS

Figure 1 depicts the variable selection process for six possible combinations of two clinical variables ("tumor status" and "survival") and three TCGA cancer datasets (LGG, HNSC, and STES). MMI (mixed mutual information) between (S, E, M) and tumor status/survival was computed for 4782 genes (LGG), 12516 genes (HNSC), and 16164 genes (STES). (All six gene lists, with corresponding MMI values, are available in **Supplementary Tables S1–S6**). The histogram representation of the MMI distribution, as shown in **Figure 1**, is convenient, as it allows to evaluate (both visually and quantitatively) the relative predictive values of the top-ranking genes with respect to the outcome variable classification. For the purposes of this study, and to make the resulting full BNs "observable," we have chosen the "top genes" cutoff value of 99.5% MMI CDF (cumulative distribution function), which leads to the selection of 24 genes (72 future BN nodes/variables in total, comprising 24 S, 24 E, and 24 M components) out of 4728 for two LGG networks, 63 genes (189 nodes/variables) out of 12516 for two HNSC networks, and 81 genes (243 nodes/variables) out of 16164 for two STES networks. Note that the S, E, and M components of each gene vector were considered as the separate nodes/variables in the subsequent BN construction, as at this time we do not have a BN scoring function that can incorporate mixed multivariate distance measures. It should also be noted that although MMI, intuitively, should not be negative, due to the way it is computed it can get into the negative range when (i) continuous variables are involved, and (ii) the number of dimensions is more than two (four, in our case). This said, all the negative MMI values in **Figure 1** reside well within the allowed algorithmic negative deviation range, and should not influence the variable rankings [personal communication from Gao et al. (2018)].

Interestingly, every histogram in **Figure 1** has a heavy right tail, which sometimes appears to follow a clear "knee point" – for example, at MMI $\sim$ = 0.08 in **Figures 1A–C**. This suggests that MMI $>$0.08 could also be used as a "natural" cutoff value, at least in these three datasets.

The variable selection distributions shown in **Figure 1** were derived with the MMI parameter k set at 15. **Figure 2** illustrates the motivation behind that choice, using the LGG/survival dataset example. Shown is the plot of the Jaccard index (JI, a.k.a. set "Intersection over Union," which is a common measure of sample set similarity) comparing the gene/variable sets resulting from the above variable selection procedure, with cutoff set at 99.5% MMI CDF, where JI(k) compares the sets obtained with k and $k+1$. It is clear that as k reaches $\sim$15, the set composition somewhat stabilizes; further increase in k does not seem to offer any advantages. (JI plots for the other datasets exhibit a similar pattern).

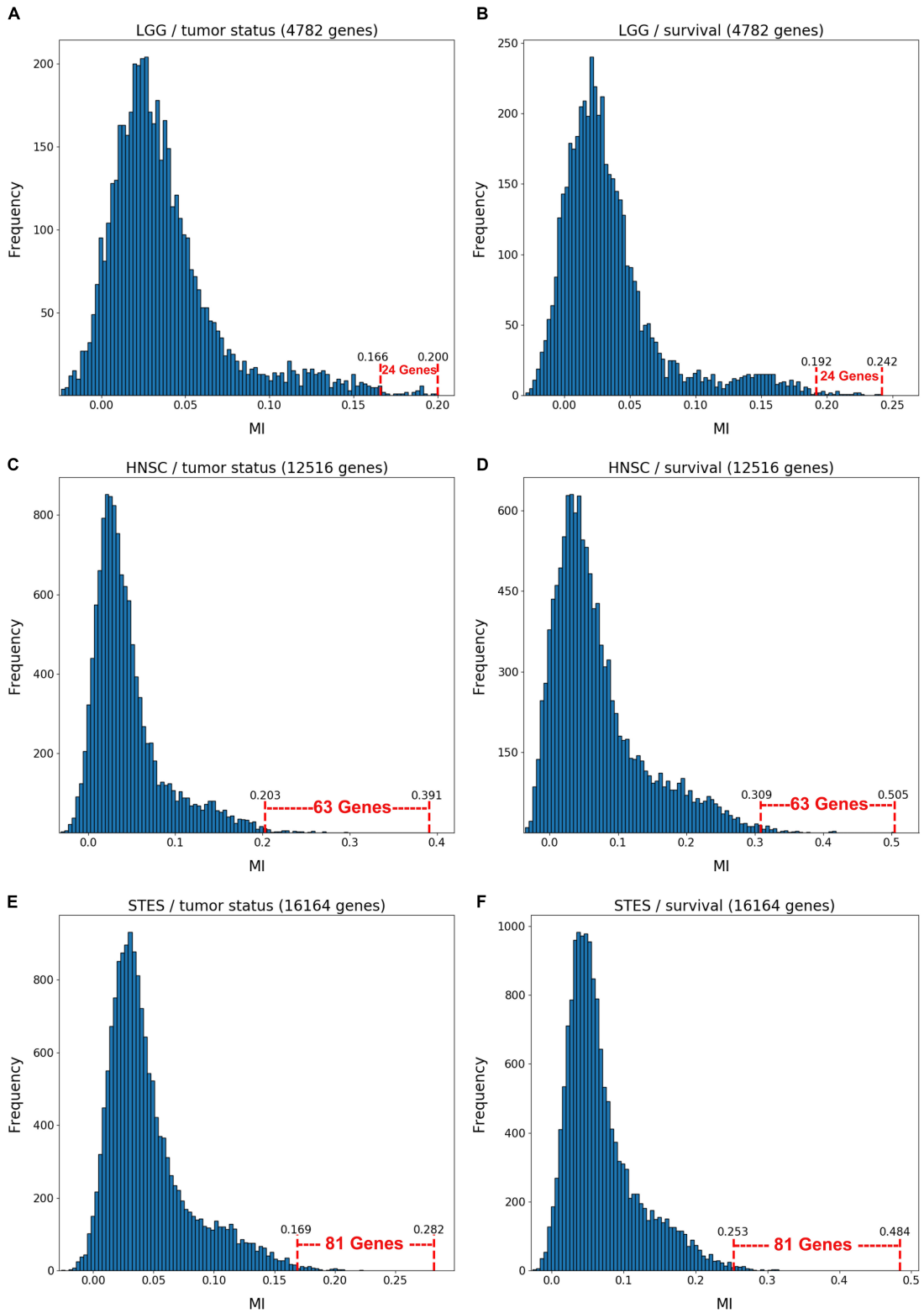


FIGURE 1 | Variable selection process for six combinations of two clinical outcome variables ("tumor status" and "survival") and three TCGA cancer datasets (LGG, HNSC, STES). MMI between the (S, E, M) molecular data vector and tumor status/survival was computed for 4782 genes (LGG), 12516 genes (HNSC) and 16164 genes (STES). The histogram representation of the MMI distribution is shown with the selection of "top" (i.e., with the MMI CDF >99.5%) genes superimposed on the right tail of the MMI frequency distribution. **(A)** LGG/tumor status; **(B)** LGG/survival; **(C)** HNSC/tumor status; **(D)** HNSC/survival; **(E)** STES/tumor status; **(F)** STES/survival.

Figures 3, **4** depict the full BNs obtained from the LGG/tumor status and LGG/survival datasets. **Supplementary Data Sheets S1–S6** depict, in PDF format, the full BNs obtained from the LGG/tumor status, LGG/survival, HNSC/tumor status, HNSC/survival, STES/tumor status, and STES/survival datasets, respectively. Six corresponding DOT (standard network / causal graphical models format) files can be found in the **Supplementary Tables S7–S12**.

While the resulting full BNs, in PDF format, are zoomable and searchable, and the DOT files can be exported into the specialized network-oriented software, the full BNs tend to be visually overwhelming for the number of variables/nodes >100. Consequently, **Figures 5–10** depict the immediate MNs of the clinical variables/nodes in the corresponding six BNs: LGG/tumor status (**Figure 5**), LGG/survival (**Figure 6**), HNSC/tumor status (**Figure 7**), HNSC/survival (**Figure 8**), STES/tumor status (**Figure 9**) and STES/survival (**Figure 10**).

It is noticeable in **Figures 5–10** that all three molecular data components (S, E, and M) are represented in the MNs. This testifies to the efficacy and proportionality of both the MMI measure (during the variable selection stage) and the maximum entropy - based discretization (during the BN construction stage). Also of note, for some genes, more than one component is present (HTR4 E and S for STES/tumor status, CHIA E and S for LGG/tumor status, AFP E and S for LGG/tumor status). Conversely, some genes are associated with both tumor status and survival (MUC4 for HNSC, TMPRSS11F, SLC6A18, and DEFB119 for LGG).

The performance of our BN reconstruction algorithm or software is discussed in general terms in Gogoshin et al. (2017); here, we will evaluate the statistical significance of the resulting MNs. While the edge strength estimates in **Figures 5–10** are useful in the relative sense, they do not immediately translate into the statistical significance measurements (such as *p*-values). Therefore, we have augmented the edge strengths with the *p*-values obtained *via* two-sample Kolmogorov–Smirnov (KS) probability distribution equality test (for continuous E and M molecular component variables) and two-sided Fisher's exact test (for discrete S molecular component variable). To illustrate the KS test application, **Figure 11** shows CDFs, separately for two "tumor status" groups, for seven continuous variables present in the MN depicted in **Figure 5** (LGG/tumor status), in order of decreasing edge strength (**Figure 11A**, MMP1_M; **Figure 11B**, DDX4_E; **Figure 11C**, AFP_E; **Figure 11D**, CHIA_E; **Figure 11E**, TMPRSS11F_M; **Figure 11F**, KERA_E; **Figure 11G**, MUC16_E). Only MMP1_M and DDX4_E appear to be statistically highly significant, with TMPRSS11F_M being arguably a borderline case.

Table 1 lists the *p*-values for all 55 potentially predictive molecular gene components present in six MNs depicted in **Figures 5–10**, in order of decreasing edge strength for each network / MN. 12 gene components were found to be statistically significant (marked with an asterisk in **Table 1**), however, we decide to exclude LCT_S (marked with ** in **Table 1**) from further scrutiny because of the very low mutation counts in both survival groups.

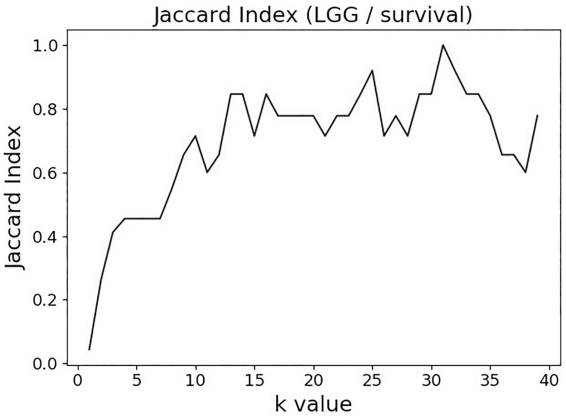


FIGURE 2 | Jaccard index (JI) ("Intersection over Union") comparing the gene/variable sets resulting from the LGG/survival dataset variable selection with the cutoff set at 99.5% MMI CDF. JI(*k*) compares the sets obtained with *k* and *k+1*.

Subsequently, we performed manual literature / database search to ascertain if any of the remaining 11 genes were previously reported in the cancer context. The following resources were used: GeneCards (Stelzer et al., 2016) and DisGeNET (Pinero et al., 2017) databases, PubMed, and Google Scholar. Six genes were found to be implicated in cancer etiology / progression / clinical outcomes with high degree of certainty: MMP1, DDX4, TRPM3, DPP6, KCNA1, and MUC17 (Senapati et al., 2010; Saied et al., 2012; Lallet-Daher et al., 2013; Kawal et al., 2016; Park et al., 2016; Schudrowitz et al., 2017). Four genes (SLC7A14, LRRIQ, SLCO1B3, and SLC9A4) were supported by weaker, circumstantial evidence (Chan-On et al., 2013; Matullo et al., 2013; Fridley et al., 2016; Tanaka et al., 2017). One gene, TMPRSS11F, has not been discussed in the cancer context before, to the best of our knowledge [see also (Kataoka et al., 2018)]. However, increased expression levels of a similar type II transmembrane serine protease, TMPRSS11D, were found to be a significant non-small cell lung cancer survival predictor (Cao et al., 2017). Therefore, we suggest that TMPRSS11F should be further investigated as a strong predictive factor playing a role in LGG patients' clinical characteristics – survival, especially. Lower TMPRSS11F methylation values correspond to a poorer long-term (2-year) survival. One possible mechanism is *via* the proteolysis of extracellular matrix which, in turn, is linked to the metastatic processes (Cao et al., 2017).

In summary, our analysis framework confirmed six well-known cancer-related genes, supplied additional evidence to support four other suspected cancer-related genes, and identified one novel potentially strongly predictive factor, methylation of TMPRSS11F.

DISCUSSION

Systems biology approach to the complex genetic and epigenetic cancer data analysis is arguably superior to the simpler single-gene (or even single-data type) alternatives. However, it is

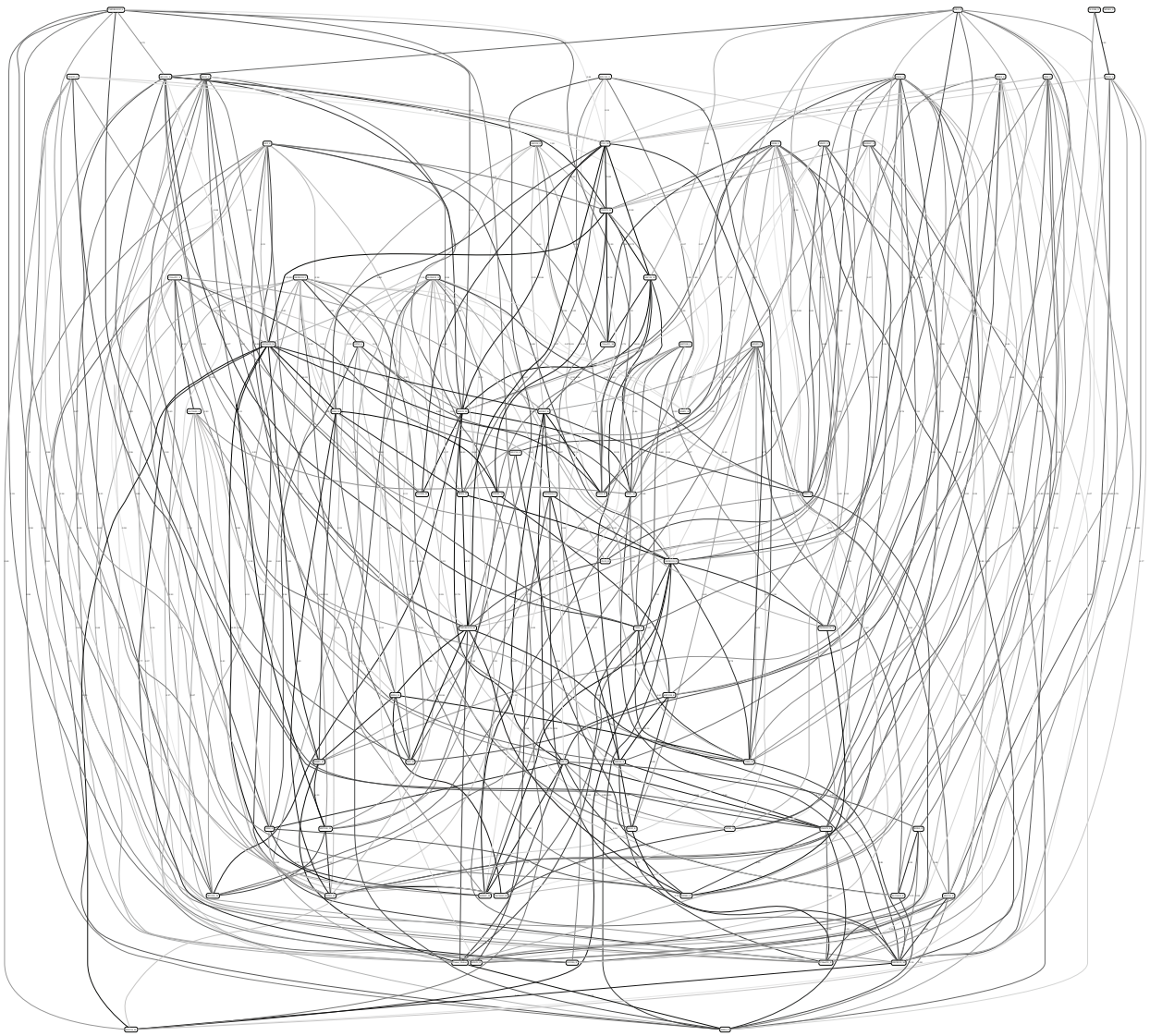


FIGURE 3 | Full BN derived from the LGG/tumor status data. "Tumor_Status" node in the BN is self-explanatory. Other nodes in the networks correspond to the genes/molecular components (gene name_S/E/M). Edges in the network correspond to the dependencies between the nodes. Directionality of the edge (arrow) is for mathematical convenience only and does not imply causation. "Boldness" of the edge is proportional to the dependency strength, also indicated by the number shown next to the edge. Note that the BN pdf image files are searchable (using gene manes), and that all the BN pdf and source files (in DOT format) are included as part of the **Supplementary Material**.

intrinsically linked to the fundamental, interrelated, challenges – scalability, "curse of dimensionality," accounting for non-additive, higher-order interactions, and visualization of the results (i.e., translation of the massive network graphs into concrete biomedical insights). In this study we propose a flexible and generalizable approach to the BN-based systems biology analysis of the multi-modal cancer data, using the TCGA database as an example. It consists of the variable selection step (which is not computationally demanding) and the BN reconstruction step (which is substantially computationally demanding). Ideally, the investigators would simply feed the complete dataset (all variables) into the BN software, obtain the full graphical model (no matter how large and complex), and then "zoom in" on the MN of the variable(s) of interest, such as a clinical outcome or a cancer phenotype. However, this is impractical for most real datasets and available hardware configurations.

Consequently, we propose starting with the variable selection step to select a (relatively) small subset of genes that are associated with the variable(s) of interest (tumor status and 2-year survival in the present study). The principal novelty of our approach lies in using the MMI measure for the variable/gene selection, in which all possible types of molecular information (discrete and continuous, genetic and epigenetic) are considered

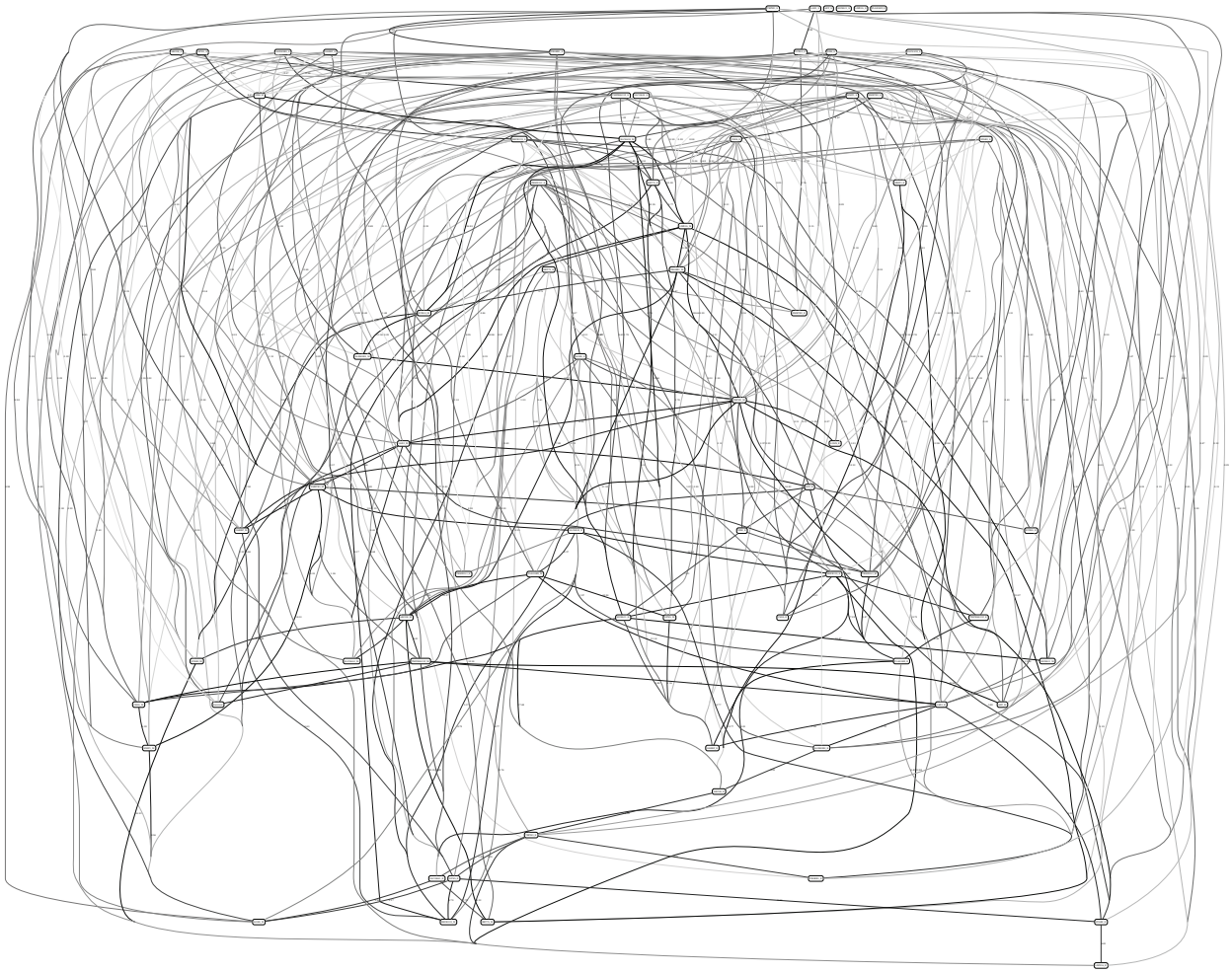


FIGURE 4 | Full BN derived from the LGG/survival data. "Survival" node in the BN is self-explanatory. Other designations are as in **Figure 3**.

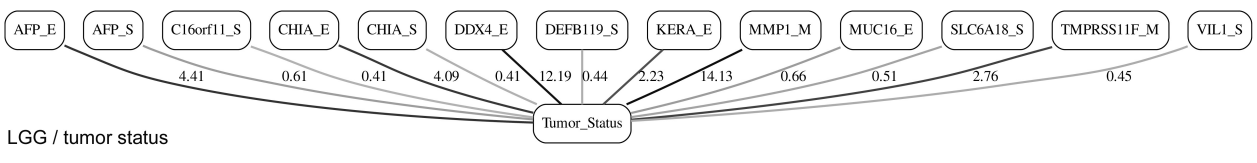


FIGURE 5 | MN of the "Tumor_Status" node in the LGG/tumor status BN.

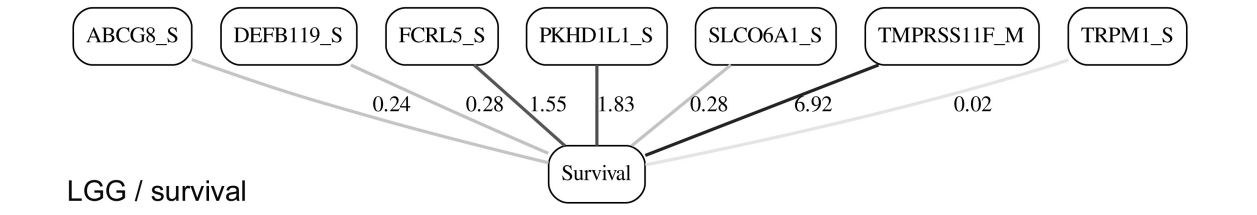


FIGURE 6 | MN of the "Survival" node in the LGG/survival BN.

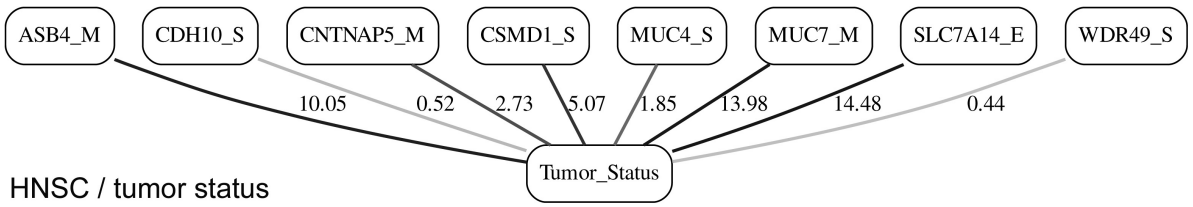


FIGURE 7 | MN of the "Tumor_Status" node in the HNSC/tumor status BN.

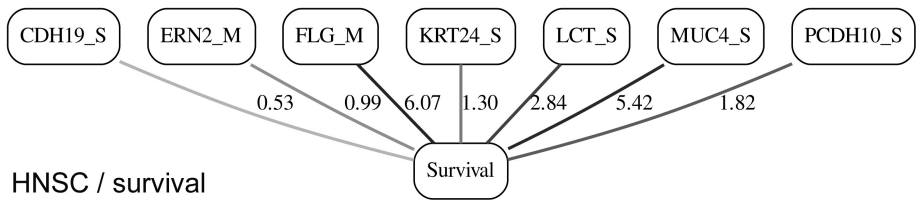


FIGURE 8 | MN of the "Survival" node in the HNSC/survival BN.

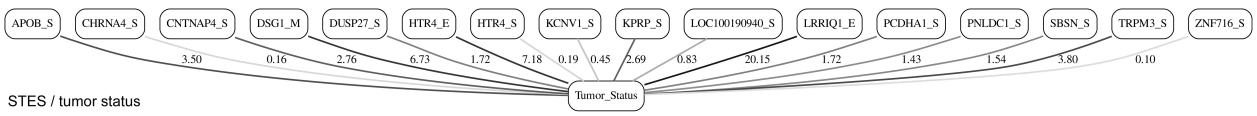


FIGURE 9 | MN of the "Tumor_Status" node in the STES/tumor status BN.

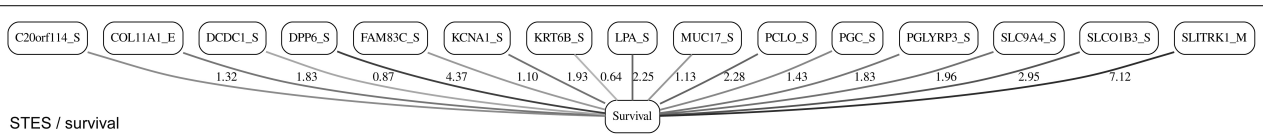


FIGURE 10 | MN of the "Survival" node in the STES/survival BN.

simultaneously. The other innovative aspect of our approach lies in the adjustability of the "hand-off" point between the variable selection and BN modeling steps. This hand-off point can depend on the investigators' computational resources, the shape of the variable selection curves, or the predefined statistical cutoff points. For example, ∼20 K genes can be reduced to 100–200 genes for the subsequent BNs construction, in which case the complete analysis takes less than an hour on a mid-level PC. When feeding the complete datasets (10,000–15,000 genes, in case of TCGA and similar genomic resources) into our BN software (Gogoshin et al., 2017), without the preliminary variable selection step, it takes about 3 days to build a full BN on a dedicated multi-core workstation. Therefore, the investigators can choose the appropriate balance depending on whether they are interested in a quick, exploratory analysis or a finalized, exhaustive one.

In our analyses, the final predictive gene sets (such as shown in **Table 1**) were different from the sets (of comparable sizes) of "top" genes obtained in the variable selection step alone (otherwise there would be no need to invoke the computationally expensive BN modeling step). This was to be expected, because BN modeling is a multivariate modeling tool (which aims to reconstruct the most fitting pattern of conditional independencies in the MN of a clinical variable), while MMI ranking is a univariate variable selection "filter" that does not account for the dependencies between the (top) genes. Another reason that the two corresponding gene sets tend to be different has to do with the fact that the first analysis stage is gene-centric, whereas the second analysis stage separates the three molecular modalities. Limiting our analysis pipeline to just the first stage (MMI filter/ranking) would therefore miss the strong one-modality (but week multiple-modalities) predictors. In future, we plan to study the extent of intersection of such two sets as a function of the "hand-off" point (between MMI pre-ranking and full BN analysis) parameter.

Our computational pipeline is inherently generalizable, as it can be directly applied to any large multimodal genetic/epigenetic dataset with minimal preprocessing. The only two changeable parameters are the aforementioned variable selection / BN modeling hand-off point, and the BN discretization mechanism. The latter is currently set as the 3-bin maximum entropy-based discretization coupled with the multinomial local probability

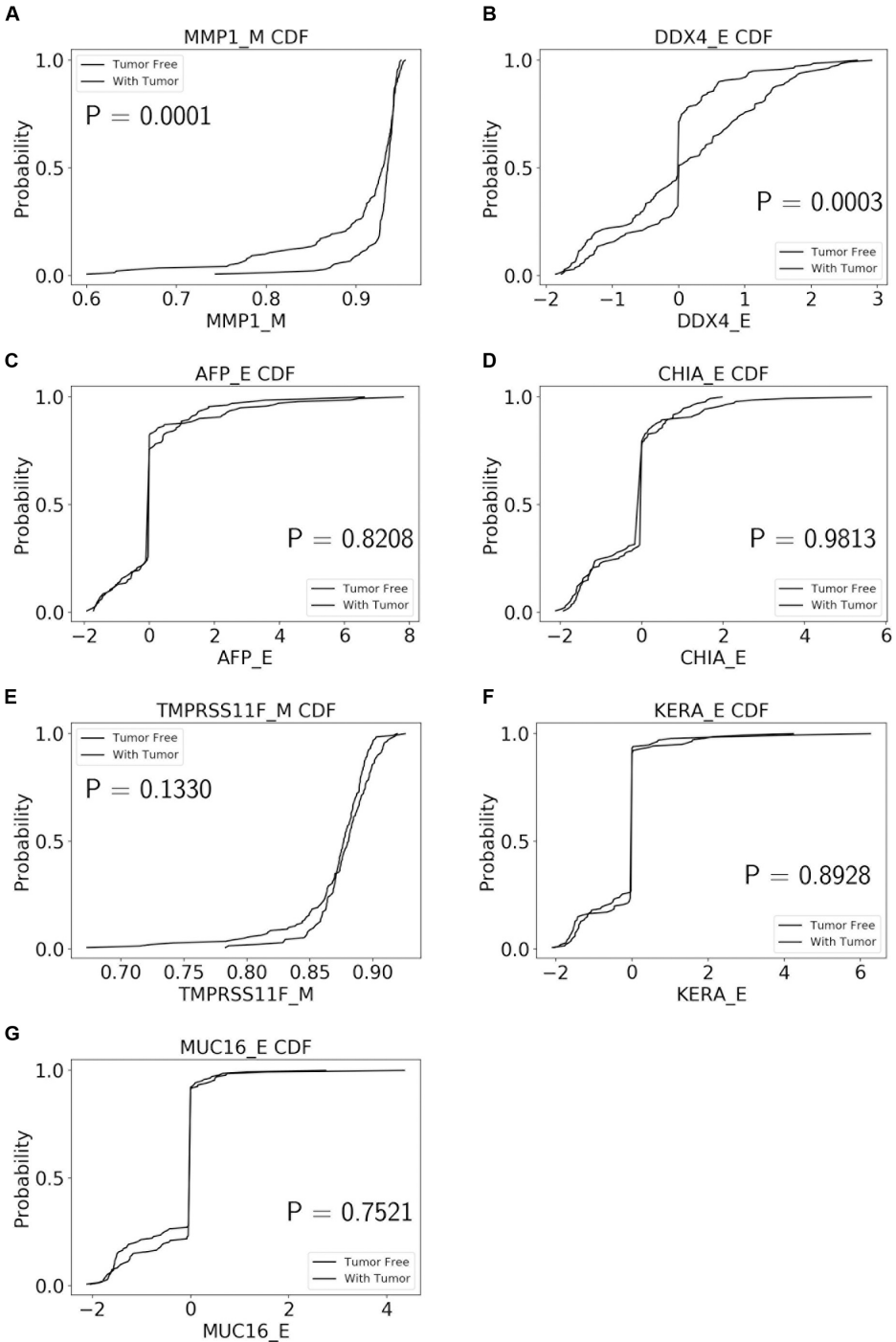


FIGURE 11 | CDFs, shown separately for two "tumor status" groups, for seven continuous variables present in the MN depicted in **Figure 5** (LGG/tumor status), in order of decreasing edge strength. **(A)** MMP1_M; **(B)** DDX4_E; **(C)** AFP_E; **(D)** CHIA_E; **(E)** TMPRSS11F_M; **(F)** KERA_E; **(G)** MUC16_E. *P*-values for the two-sample Kolmogorov-Smirnov test are shown in each chart.

model (Gogoshin et al., 2017). This is not the most elegant, or universally applicable, solution. In future, we plan to develop a novel BN model scoring function derived from a mixed distance measure (such as the MMI), or a similar metric that expresses divergence between the current network model and the data *via* mixed-type distances. The resulting two-stage analytical strategy will thus fully automatically deal with the mixed variables, in both of its stages. This has not been done before, so we plan to implement and test the MMI-based BN algorithm alongside the more established mixed-type BN solutions (hybrid local

TABLE 1 | *P*-values for 55 potentially predictive molecular gene components present in six MNs depicted in **Figures 5–10**, subdivided by six datasets, in order of decreasing edge strength for each dataset/MN.

(A)		
Gene_component	**BN edge strength**	***p*-value**
LGG/tumor status		
MMP1_M	14.13	0.0001*
DDX4_E	12.19	0.0003*
AFP_E	4.41	0.8208
CHIA_E	4.09	0.9813
TMPRSS11F_M	2.76	0.1330
KERA_E	2.23	0.8928
MUC16_E	0.66	0.7521
AFP_S	0.61	0.4985
SLC6A18_S	0.51	∼1.0
VIL1_S	0.45	0.2364
DEFB119_S	0.44	0.4872
C16orf11_S	0.41	∼1.0
CHIA_S	0.41	∼1.0
(B)		
LGG/survival		
TMPRSS11F_M	6.92	0.0003*
PKHD1L1_S	1.83	0.5818
FCRL5_S	1.55	0.1895
DEFB119_S	0.28	∼1.0
SLCO6A1_S	0.28	∼1.0
ABCG8_S	0.24	0.4366
TRPM1_S	0.02	0.1895
(C)		
HNSC/tumor status		
SLC7A14_E	14.48	2.1230e-05*
MUC7_M	13.98	0.3432
ASB4_M	10.05	0.2832
CSMD1_S	5.07	∼1.0
CNTNAP5_M	2.73	0.5261
MUC4_S	1.85	0.1496
CDH10_S	0.52	∼1.0
WDR49_S	0.44	∼1.0
(D)		
HNSC/survival		
FLG_M	6.07	0.2427
MUC4_S	5.42	0.4103
LCT_S	2.84	0.0226**
PCDH10_S	1.82	0.3544
KRT24_S	1.3	0.4942
ERN2_M	0.99	0.4832
CDH19_S	0.53	0.5787
(E)		
STES/tumor status		
LRRIQ_E	20.15	0.0004*
HTR4_E	7.18	0.2960
DSG1_M	6.73	0.1204
TRPM3_S	3.80	0.0245*
APOB_S	3.50	0.2723
CNTNAP4_S	2.76	0.7784
KPRP_S	2.69	∼1.0
DUSP27_S	1.72	0.5480
PCDHA1_S	1.72	0.8074
SBSN_S	1.54	0.6248
PNLDC1_S	1.43	∼1.0
LOC100190940_S	0.83	∼1.0
KCNV1_S	0.45	∼1.0
HTR4_S	0.19	0.3594
CHRNA4_S	0.16	0.3968
ZNF716_S	0.10	∼1.0
(F)		
STES/survival		
SLITRK1_M	7.12	0.1082
DPP6_S	4.37	0.0280*
SLCO1B3_S	2.95	0.0059*
PCLO_S	2.28	0.7238
LPA_S	2.25	0.0549
SLC9A4_S	1.96	0.0059*
KCNA1_S	1.93	0.0108*
COL11A1_E	1.83	0.2016
PGLYRP3_S	1.83	0.1119
PGC_S	1.43	0.4527
C20orf114_S	1.32	∼1.0
MUC17_S	1.13	0.0008*
FAM83C_S	1.10	0.6515
DCDC1_S	0.87	0.2051
KRT6B_S	0.64	∼1.0

*Twelve gene components were found to be statistically significant (marked with *); LCT_S (marked with **) was excluded from further analysis because of the very low mutation counts (zero mutations in >2-year survival group, three mutations in <2-year survival group).*

probability models, adaptive discretization), and use both real and simulated data to investigate which method is preferable.

Another limitation of the present study has to do with its primary focus on the clinical outcomes / phenotypes; at this time, we decided to largely concentrate on the MNs of the clinical variables/nodes. In future, we intend to analyze the resulting full BNs more "holistically," paying attention to the general network topological properties, gene clusters, hub and bottleneck genes, etc. Consequently, one useful extension of our analytical framework would be to incorporate multiple clinical outcomes / phenotypes into the network analyses, to see if the inter-outcome dependencies are reflected in the resulting networks, and if they are mediated by other nodes/variables.

Application of our pipeline to TCGA data resulted in the identification of a number of candidate genes for the different clinical cancer characteristics, *via* varied molecular components. It is well known that epigenetic processes / DNA methylation play

an important role in many cancers' diagnosis, progression, and outcome; our results support that notion, as many of the most statistically significant predictors generated in the present study were in fact the methylation molecular components (**Table 1**). Notably, the one novel candidate gene pinpointed in this study, TMPRSS11F, likely would not have been identified *via* any other (non-epigenetic) modality. Our results, therefore, underscore the essentiality of the simultaneous analysis of different molecular modalities, including the epigenetic ones, for the precision or personalized medicine to be effective in cancer treatment.

DATA AVAILABILITY STATEMENT

All the intermediate datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

XW, SB, and AR conceptualized the study, carried out the analyses, and wrote the manuscript. GG and SD contributed to carrying out the analyses. All authors contributed to the article and approved the submitted version.

FUNDING

AR, SB, and GG are supported by NIH NCI U01 CA232216. AR is supported by the Susumu Ohno Chair in Theoretical and Computational Biology. GG is supported by Susumu Ohno Distinguished Investigator fellowship. This work was also partially supported by City of Hope funds (XW, SB, GG, and AR). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

The authors are grateful to Arthur D. Riggs, Russell Rockne, Dustin Schones, Wendong Huang, and Weihao Gao for many stimulating discussions and useful suggestions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00648/full#supplementary-material

TABLES S1–S6 | Data files, in Excel format, listing the genes analyzed in this study, together with the corresponding MMI values, in order of decreasing MMI values. **(1)** LGG/tumor status, **(2)** LGG/survival, **(3)** HNSC/tumor status, **(4)** HNSC/survival, **(5)** STES/tumor status, **(6)** STES/survival. Genes with the MMI values >99.5% MMI CDF (i.e., genes selected for further BN analyses) are shown in bold.

TABLES S7–S12 | Full Bayesian networks, in Word / DOT format, for the six datasets in this study. **(1)** LGG/tumor status, **(2)** LGG/survival, **(3)** HNSC/tumor status, **(4)** HNSC/survival, **(5)** STES/tumor status, **(6)** STES/survival.

DATA SHEETS S1–S6 | Full Bayesian networks, in PDF format, for the six datasets in this study. **(1)** LGG/tumor status, **(2)** LGG/survival, **(3)** HNSC/tumor status, **(4)** HNSC/survival, **(5)** STES/tumor status, **(6)** STES/survival. Designations are as in main text **Figure 3**.

REFERENCES

Andrews, B., Ramsey, J., and Cooper, G. F. (2018). Scoring Bayesian Networks of Mixed Variables. *Int. J. Data Sci. Anal.* 6, 3–18. doi: 10.1007/s41060-017-0085-7

Asur, S., Ucar, D., and Parthasarathy, S. (2007). An ensemble framework for clustering protein-protein interaction networks. *Bioinformatics* 23, i29–i40. doi: 10.1093/bioinformatics/btm212

Blum, A. L., and Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artif. Intell.* 97, 245–271. doi: 10.1016/s0004-3702(97)00063-5

Cao, X., Tang, Z., Huang, F., Jin, Q., Zhou, X., Shi, J., et al. (2017). High TMPRSS11D protein expression predicts poor overall survival in non-small cell lung cancer. *Oncotarget* 8, 12812–12819. doi: 10.18632/oncotarget.14559

Chan-On, W., Nairismagi, M. L., Ong, C. K., Lim, W. K., Dima, S., Pairojkul, C., et al. (2013). Exome sequencing identifies distinct mutational patterns in liver fluke-related and non-infection-related bile duct cancers. *Nat. Genet.* 45, 1474–1478. doi: 10.1038/ng.2806

Fang, X., Yin, Z., Li, X., Xia, L., Quan, X., Zhao, Y., et al. (2017). Multiple functional SNPs in differentially expressed genes modify risk and survival of non-small cell lung cancer in chinese female non-smokers. *Oncotarget* 8, 18924–18934. doi: 10.18632/oncotarget.14836

Fridley, B. L., Ghosh, T. M., Wang, A., Raghavan, R., Dai, J., Goode, E. L., et al. (2016). Genome-wide study of response to platinum. taxane, and combination therapy in ovarian cancer: in vitro phenotypes, inherited variation, and disease recurrence. *Front Genet* 7:37. doi: 10.3389/fgene.2016.00037

Friedman, N., Linial, M., Nachman, I., and Pe'er, D. (2000). Using Bayesian networks to analyze expression data. *J. Comput. Biol.* 7, 601–620. doi: 10.1089/106652700750050961

Gao, W., Kannan, S., Oh, S., and Viswanath, P. (2018). Estimating mutual information for discrete-continuous mixtures. *arXiv [Preprint]*. Available online at: https://arxiv.org/abs/1709.06212

Gevaert, O., De Smet, F., Timmerman, D., Moreau, Y., and De Moor, B. (2006). Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. *Bioinformatics* 22, e184–e190. doi: 10.1093/bioinformatics/btl230

Gogoshin, G., Boerwinkle, E., and Rodin, A. S. (2017). New Algorithm and Software (BNOmics) for Inferring and Visualizing Bayesian Networks from Heterogeneous Big Biological and Genetic Data. *J. Comput. Biol.* 24, 340–356. doi: 10.1089/cmb.2016.0100

Guyon, I., and Elisseeff, A. (2003). An introduction to variable and feature selection. *J. Mach. Learn. Res.* 3, 1157–1182.

Heckerman, D. (1995). *Tutorial on Learning with Bayesian Networks*. Technical Report MSR-TR-95–96. Redmond, WA: Microsoft Research.

Hou, X., He, X., Wang, K., Hou, N., Fu, J., et al. (2018). Genome-wide network-based analysis of colorectal cancer identifies novel prognostic factors and an integrative prognostic index. *Cell Physiol. Biochem.* 49, 1703–1716. doi: 10.1159/000493614

Huang, T., Yang, J., and Cai, Y. D. (2015). Novel candidate key drivers in the integrative network of genes, microRNAs, methylations, and copy number variations in squamous cell lung carcinoma. *Biomed. Res. Int.* 2015, 358125.

Jeong, H., Leem, S., Wee, K., and Sohn, K. A. (2015). Integrative network analysis for survival-associated gene-gene interactions across multiple genomic profiles in ovarian cancer. *J. Ovarian. Res.* 8:42.

Kaiser, J. L., Bland, C. L., and Klinke, D. J. II (2016). Identifying causal networks linking cancer processes and anti-tumor immunity using Bayesian network inference and metagene constructs. *Biotechnol. Prog.* 32, 470–479. doi: 10.1002/btpr.2230

Kataoka, H., Kawaguchi, M., Fukushima, T., and Shimomura, T. (2018). Hepatocyte growth factor activator inhibitors (HAI-1 and HAI-2): emerging key players in epithelial integrity and cancer. *Pathol. Int.* 68, 145–158. doi: 10.1111/pin.12647

Kawal, P., Chandra, A., Rajkumar, Dhole, T. N., and Ojha, B. (2016). Correlations of polymorphisms in matrix metalloproteinase-1, -2, and -7 promoters to susceptibility to malignant gliomas. *Asian J. Neurosurg.* 11, 160–166.

Koch, A., Joosten, S. C., Feng, Z., de Ruijter, T. C., Draht, M. X., Melotte, V., et al. (2018). Analysis of DNA methylation in cancer: location revisited. *Nat. Rev. Clin. Oncol.* 15, 459–466. doi: 10.1038/s41571-018-0004-4

Kohavi, R., and John, G. H. (1997). Wrappers for feature subset selection. *Artif. Intell.* 97, 273–324. doi: 10.1016/s0004-3702(97)00043-x

Kraskov, A., Stögbauer, H., and Grassberger, P. (2004). Estimating mutual information. *Phys. Rev. E* 69, 066138.

Lallet-Daher, H., Wiel, C., Gitenay, D., Navaratnam, N., Augert, A., Calvé, B. L., et al. (2013). Potassium channel KCNA1 modulates oncogene-induced senescence and transformation. *Cancer Res.* 73, 5253–5265. doi: 10.1158/0008-5472.can-12-3690

Leng, J., Valli, C., and Armstrong, L. (2010). "A wrapper-based feature selection for analysis of large data sets," in *Proceedings of the Conference: 3rd International Conference on Computer and Electrical Engineering (ICCEE 2010)*, Chengdu.

Liang, A., Zhou, B., and Sun, W. (2017). Integrated genomic characterization of cancer genes in glioma. *Cancer Cell Int.* 17:90.

Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Favera, R. D., et al. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7(Suppl. 1):S7. doi: 10.1186/1471-2105-7-S1-S7

Matullo, G., Guarrera, S., Betti, M., Fiorito, G., Ferrante, D., Voglino, F., et al. (2013). Genetic variants associated with increased risk of malignant pleural mesothelioma: a genome-wide association study. *PLoS One* 8:e61253. doi: 10.1371/journal.pone.0061253

Park, Y. R., Chun, J. N., So, I., Kim, H. J., Baek, S., Jeon, J. H, et al. (2016). Data-driven Analysis of TRP channels in cancer: linking variation in gene expression to clinical significance. *Cancer Genom. Proteom.* 13, 83–90.

Phan, J. H., Hoffman, R., Kothari, S., Wu, P. Y., and Wang, M. D. (2016). Integration of multi-modal biomedical data to predict cancer grade and patient survival. *IEEE EMBS Int. Conf. Biomed. Health Inform.* 2016, 577–580.

Phipps, A. I., Passarelli, M. N., Chan, A. T., Harrison, T. A., Jeon, J., Hutter, C. M., et al. (2016). Common genetic variation and survival after colorectal cancer diagnosis: a genome-wide analysis. *Carcinogenesis* 37, 87–95.

Pinero, J., Bravo, A., Queralt-Rosinach, N., Gutierrez-Sacristan, A., Deu-Pons, J., Centeno, E., et al. (2017). DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* 45, D833–D839.

Rajesh, Y., Pal, I., Banik, P., Chakraborty, S., Borkar, S. A., Dey, G., et al. (2017). Insights into molecular therapy of glioma: current challenges and next generation blueprint. *Acta Pharmacol. Sin.* 38, 591–613. doi: 10.1038/aps.2016.167

Ramsey, J., Glymour, M., Sanchez-Romero, R., and Glymour, C. (2017). A million variables and more: the Fast Greedy Equivalence Search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images. *Int. J. Data Sci. Anal.* 3, 121–129. doi: 10.1007/s41060-016-0032-z

Rodin, A. S., Litvinenko, A., Klos, K., Morrison, A. C., Woodage, T., Coresh, J., et al. (2009). Use of wrapper algorithms coupled with a random forests classifier for variable selection in large-scale genomic association studies. *J. Comput. Biol.* 16, 1705–1718. doi: 10.1089/cmb.2008.0037

Saied, M. H., Marzec, J., Khalid, S., Smith, P., Down, T. A., Rakyan, V. K., et al. (2012). Genome wide analysis of acute myeloid leukemia reveal leukemia specific methylome and subtype specific hypomethylation of repeats. *PLoS One* 7:e33213. doi: 10.1371/journal.pone.0033213

Schudrowitz, N., Takagi, S., Wessel, G. M., and Yajima, M. (2017). Germline factor DDX4 functions in blood-derived cancer cell phenotypes. *Cancer Sci.* 108, 1612–1619. doi: 10.1111/cas.13299

Sedgewick, A. J., Buschur, K., Shi, I., Ramsey, J. D., Raghu, V. K., Manatakis, D. V., et al. (2018). Mixed graphical models for integrative causal analysis with application to chronic lung disease diagnosis and prognosis. *Bioinformatics* 35, 1204–1212. doi: 10.1093/bioinformatics/bty769

Senapati, S., Ho, S. B., Sharma, P., Das, S., Chakraborty, S., Kaur, S., et al. (2010). Expression of intestinal MUC17 membrane-bound mucin in inflammatory and neoplastic diseases of the colon. *J. Clin. Pathol.* 63, 702–707. doi: 10.1136/jcp.2010.078717

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., et al. (2016). The genecards suite: from gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinformatics* 54:1.30.1-1.30.33.

Tanaka, H., Kanda, M., Shimizu, D., Tanaka, C., Kobayashi, D., Hayashi, M., et al. (2017). FAM46C serves as a predictor of hepatic recurrence in patients with resectable gastric cancer. *Ann. Surg. Oncol.* 24, 3438–3445. doi: 10.1245/s10434-016-5636-y

Tian, W., Li, Y., Zhang, J., Li, J., and Gao, J. (2018). Combined analysis of DNA methylation and gene expression profiles of osteosarcoma identified several prognosis signatures. *Gene* 650, 7–14. doi: 10.1016/j.gene.2018.01.093

Tomczak, K., Czerwinska, P., and Wiznerowicz, M. (2015). The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp. Oncol.* 19, A68–A77.

Wang, W., Baladandayuthapani, V., Holmes, C. C., and Do, K. A. (2013). Integrative network-based Bayesian analysis of diverse genomics data. *BMC Bioinformatics* 14(Suppl. 13):S8. doi: 10.1186/1471-2105-14-S13-S8

Wu, S., Law, A., and Whipple, M. E. (2017). A bayesian network model of head and neck squamous cell carcinoma incorporating gene expression profiles. *Stud. Health Technol. Inform.* 245, 634–638.

Xu, P., Yang, J., Liu, J., Yang, X., Liao, J., Yuan, F., et al. (2018). Identification of glioblastoma gene prognosis modules based on weighted gene co-expression network analysis. *BMC Med. Genomics* 11:96. doi: 10.1186/s12920-018-0407-1

Xu, Y., Zhang, J., Yuan, Y., Mitra, R., Muller, P., Ji, Y., et al. (2012). A bayesian graphical model for integrative analysis of TCGA Data. *IEEE Int. Workshop Genomic Signal Process. Stat.* 2012, 135–138.

Xu, Y., Zhu, Y., Muller, P., Mitra, R., and Ji, Y. (2014). Characterizing Cancer-Specific Networks by Integrating TCGA Data. *Cancer Inform.* 13, 125–131.

Yoo, B. C., Kim, K. H., Woo, S. M., and Myung, J. K. (2017). Clinical multi-omics strategies for the effective cancer management. *J. Proteomics* 188, 97–106. doi: 10.1016/j.jprot.2017.08.010

Zhang, Q., Burdette, J. E., and Wang, J. P. (2014). Integrative network analysis of TCGA data for ovarian cancer. *BMC Syst Biol* 8:1338. doi: 10.1186/s12918-014-0136-9

Zhang, X., Branciamore, S., Gogoshin, G., Rodin, A. S., and Riggs, A. D. (2017). Analysis of high-resolution 3D intrachromosomal interactions aided by Bayesian network modeling. *Proc. Natl. Acad. Sci. U.S.A.* 114, E10359–E10368.

Zhang, C., Cheng, W., Ren, X., Wang, Z., Liu, X., Li, G., et al. (2017). Tumor purity as an underlying key factor in Glioma. *Clin. Cancer Res.* 23, 6279–6291. doi: 10.1158/1078-0432.ccr-16-2598

Zhu, Y., Xu, Y., Helseth, D. L. Jr., Gulukota, K., Yang, S., Pesce, L. L., et al. (2015). Zodiac: a comprehensive depiction of genetic interactions in cancer by integrating TCGA data. *J. Natl. Cancer Inst.* 107:djv129.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Wang, Branciamore, Gogoshin, Ding and Rodin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership