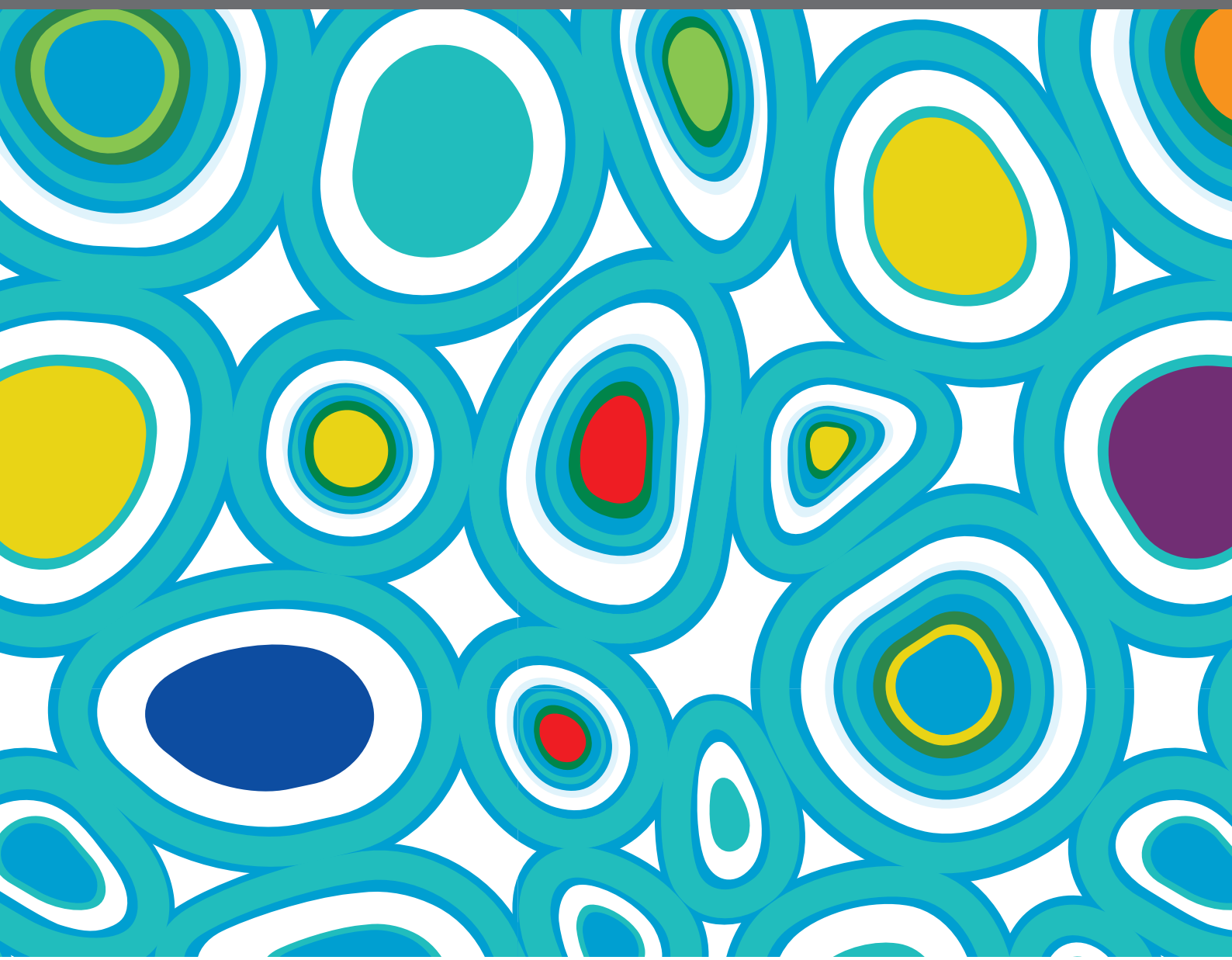
DNA METHYLATION DYNAMICS AND HUMAN DISEASES

EDITED BY: Li Tan, Jiemin Wong, Peter Adams, Chunjie Jiang and Hao Chen

PUBLISHED IN: Frontiers in Cell and Developmental Biology



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-585-0
DOI 10.3389/978-2-88976-585-0

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: frontiersin.org/about/contact

DNA METHYLATION DYNAMICS AND HUMAN DISEASES

Topic Editors:
Li Tan, Fudan University, China
Jiemin Wong, East China Normal University, China
Peter Adams, Sanford Burnham Prebys Medical Discovery Institute, United States
Chunjie Jiang, University of Pennsylvania, United States
Hao Chen, Southern University of Science and Technology, China

Citation: Tan, L., Wong, J., Adams, P., Jiang, C., Chen, H., eds. (2022). DNA Methylation Dynamics and Human Diseases. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88976-585-0

Table of Contents

frontiers | Frontiers in Cell and Developmental Biology

EDITORIAL
published: 23 June 2022
doi: 10.3389/fcell.2022.956286

Editorial: DNA Methylation Dynamics and Human Diseases

Chunjie Jiang[1]* *and Shengli Li*[2]*

[1]Division of Diabetes, Endocrinology, and Metabolism, Department of Medicine, Baylor College of Medicine, Houston, TX, United States, [2]Precision Research Center for Refractory Diseases, Institute for Clinical Research, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Keywords: DNA methylation, next-generation sequencing, human diseases, 5-methylcytosine (5mC), cancer

Editorial on the Research Topic

DNA Methylation Dynamics and Human Diseases

DNA methylation is a biological process that adds methyl groups to DNA molecules. It regulates chromatin architecture and transcription, and plays essential roles in a wide range of biological processes (Li et al., 2013; Li et al., 2018; Roy et al., 2021). Accumulating evidence shows that the dysregulation of DNA methylation is involved in the development of many life-threatening diseases, including cancers, cardiovascular diseases (Pepin et al., 2019; Heery and Schaefer, 2021; Pepin et al., 2021; Dillinger et al., 2022; Rosselló-Tortella et al., 2022). This Research Topic aims to elucidate the correlation between DNA methylation and human diseases, with the hope to deepen our understanding of the underlying molecular mechanisms of DNA methylation in human diseases and pave the way to the development of new strategies or methods for disease prevention, diagnosis, and therapy. In total, there are 9 original research articles that have been published in this Research Topic.

In this Research Topic, we received articles in the fields of cancer biology, cardiac biology and covid research. Adding methyl groups to the C-5 position of cytosine, 5-methylcytosine, is the main type of DNA methylation (Choi et al., 2021), whose dysregulation has been reported to play roles in many types of cancer. To understand the 5-methylcytosine in cancer, one of the primary cause of death worldwide (Siegel et al., 2020), Liu et al. developed a 5-methylcytosine score system and evaluated tumor mutation burden, immune check-point inhibitor response, and the clinical prognosis of individual tumors using the score based on the 5-methylcytosine profile of 1,374 lung adenocarcinoma samples. Besides, by analyzing 5-methylcytosine and gene expression profiles from 860 clear cell renal cell carcinoma (ccRCC), Xu et al. found LINC00861 was the potentially intervening target of immunotherapy for prostate cancer patients and was significantly associated with the expression of PD-1 and CTLA4. Li et al. developed predictive diagnostic and prognostic models by using machine-learning and Cox regression approaches based on pancreatic adenocarcinoma (PAAD) datasets. To take the advantage of the third-generation sequencing technique, Zhang et al. collected two pairs of tumor tissues and adjacent normal tissues from hepatocellular carcinoma (HCC) surgical samples, and then conducted Nanopore sequencing. Zhang et al. identified four potential tumor suppressor genes (KCNIP4, CACNA1C, PACRG, and ST6GALNAC3) by the integrative analysis of 5-methycytosine and 6-methyladenine profiling. Their study provided a new solution for epigenetic regulation research and therapy of HCC. LncRNAs have been shown to have high tissue- and disease-specific expression patterns, which endows them the potential in therapy (Wapinski and Chang, 2011; Jiang et al., 2016; Jiang et al., 2019). Zhao et al. focused on lncRNAs differentially expressed in only one and multiple cancer types, and identified 29 lncRNAs as diagnostic biomarkers for ccRCC, the kidney renal papillary cell carcinoma (KIRP), and pan-cancer.

OPEN ACCESS

Edited and reviewed by:
Ramani Ramchandran, Medical College of Wisconsin, United States

***Correspondence:**
Chunjie Jiang
chunjie.jiang917@outlook.com
Shengli Li
shengli.li@shsmu.edu.cn

Specialty section:
This article was submitted to Molecular and Cellular Pathology, a section of the journal Frontiers in Cell and Developmental Biology

Received: *30 May 2022*
Accepted: *07 June 2022*
Published: *23 June 2022*

Citation:
Jiang C and Li S (2022) Editorial: DNA Methylation Dynamics and Human Diseases. Front. Cell Dev. Biol. 10:956286. doi: 10.3389/fcell.2022.956286

Dysregulated regulation of miRNAs has been shown to contribute to the pathogenesis of cardiovascular diseases (Shao et al., 2015; Dong et al., 2020; Qiu et al., 2020; Yang et al., 2020). To investigate the interaction between miR-29b and DNA methylation in cardiovascular diseases, Wu et al. collected heart tissue samples from 17 patients with congenital heart disease (CHD). The authors found an inversely correlation between miR-19a and DNA methyltransferases (DNMT) in the patients. Further exploration in hypomethylated zebrafish showed that miR-29b inhibitor relieved the deformity of hypomethylated zebrafish and restored the DNA methylation patterns in cardiomyocytes, resulting in increased proliferation and renormalization of gene expression, suggesting a mutual regulation between miR-29b and DNMTs in cardiomyocytes and supporting the miRNA-based therapy in cardiomyocytes.

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has become a global public health crisis. To explore the roles of 5-Hydroxymethylcytosine in COVID-19, Chen et al. developed a machine learning model based on genome-wide 5-Hydroxymethylcytosine profiles in plasma cell-free DNA (cfDNA) from 53 healthy volunteers, 66 patients with moderate COVID-19, 99 patients with severe COVID-19, and 38 patients with critical COVID-19. They found 5-Hydroxymethylcytosine detected in cfDNA could be used as early warning markers for the disease progression and myocardial injury of COVID-19.

In summary, by taking the advantage of the state-of-the-art high-throughput sequencing technologies, the authors showed the function role of DNA methylation in various types of cancer, as well as in cardiovascular disease and COVID-19. We would like to thank all authors for their paper published in this Research Topic. These studies made significant contributions in the field, extending our understanding of the roles of DNA methylation in human disease and will facilitate further advancement. Nevertheless, studies investigating the dynamic changes of DNA methylation of human diseases and the potential of DNMT in disease therapeutics are still lacking, which might be the directions of future efforts in this Research Topic. In addition, the rapid development of sequencing technologies, such as nanopore DNA sequencing, will accelerate our research and discovery in this field.

AUTHOR CONTRIBUTIONS

SL and CJ wrote the manuscript. All authors listed approved it for publication.

FUNDING

This work was supported by Shanghai General Hospital Startup Funding (02.06.01.20.06).

ACKNOWLEDGMENTS

We thank L. Tan, J. Wong, P. Adams, and H. Chen for their support and valuable discussions.

REFERENCES

Choi, W. L., Mok, Y. G., and Huh, J. H. (2021). Application of 5-methylcytosine DNA Glycosylase to the Quantitative Analysis of DNA Methylation. *Int. J. Mol. Sci.* 22, 1072. doi:10.3390/ijms22031072

Dillinger, T., Sheibani-Tezerji, R., Pulverer, W., Stelzer, I., Hassler, M. R., Scheibelreiter, J., et al. (2022). Identification of Tumor Tissue-Derived DNA Methylation Biomarkers for the Detection and Therapy Response Evaluation of Metastatic Castration Resistant Prostate Cancer in Liquid Biopsies. *Mol. Cancer* 21, 1–8. doi:10.1186/s12943-021-01445-0

Dong, Y., Xiao, Y., Shi, Q., and Jiang, C. (2020). Dysregulated lncRNA-miRNA-mRNA Network Reveals Patient Survival-Associated Modules and RNA Binding Proteins in Invasive Breast Carcinoma. *Front. Genet.* 10, 1–14. doi:10.3389/fgene.2019.01284

Heery, R., and Schaefer, M. H. (2021). DNA Methylation Variation along the Cancer Epigenome and the Identification of Novel Epigenetic Driver Events. *Nucleic Acids Res.* 49, 12692–12705. doi:10.1093/nar/gkab1167

Jiang, C., Ding, N., Li, J., Jin, X., Li, L., Pan, T., et al. (2019). Landscape of the Long Non-coding RNA Transcriptome in Human Heart. *Briefings Bioinforma.* 20, 1812–1825. doi:10.1093/bib/bby052

Jiang, C., Li, Y., Zhao, Z., Lu, J., Chen, H., Ding, N., et al. (2016). Identifying and Functionally Characterizing Tissue-specific and Ubiquitously Expressed Human lncRNAs. *Oncotarget* 7, 7120–7133. doi:10.18632/oncotarget.6859

Li, S., Zhang, J., Huang, S., and He, X. (2018). Genome-wide Analysis Reveals that Exon Methylation Facilitates its Selective Usage in the Human Transcriptome. *Briefings Bioinforma.* 19, 754–764. doi:10.1093/bib/bbx019

Li, Y., Xu, J., Chen, H., Zhao, Z., Li, S., Bai, J., et al. (2013). Characterizing Genes with Distinct Methylation Patterns in the Context of Protein-Protein Interaction Network: Application to Human Brain Tissues. *PLoS ONE* 8, e65871. doi:10.1371/journal.pone.0065871

Pepin, M. E., Ha, C.-M., Crossman, D. K., Litovsky, S. H., Varambally, S., Barchue, J. P., et al. (2019). Genome-wide DNA Methylation Encodes Cardiac Transcriptional Reprogramming in Human Ischemic Heart Failure. *Lab. Invest.* 99, 371–386. doi:10.1038/s41374-018-0104-x

Pepin, M. E., Ha, C.-M., Potter, L. A., Bakshi, S., Barchue, J. P., Haj Asaad, A., et al. (2021). Racial and Socioeconomic Disparity Associates with Differences in Cardiac DNA Methylation Among Men with End-Stage Heart Failure. *Am. J. Physiology-Heart Circulatory Physiology* 320, H2066–H2079. doi:10.1152/AJPHEART.00036.2021

Qiu, M., Fu, Q., Jiang, C., and Liu, D. (2020). Machine Learning Based Network Analysis Determined Clinically Relevant miRNAs in Breast Cancer. *Front. Genet.* 11, 1–11. doi:10.3389/fgene.2020.615864

Rosselló-Tortella, M., Bueno-Costa, A., Martínez-Verbo, L., Villanueva, L., and Esteller, M. (2022). DNA Methylation-Associated Dysregulation of Transfer RNA Expression in Human Cancer. *Mol. Cancer* 21, 1–8. doi:10.1186/s12943-022-01532-w

Roy, R., Ramamoorthy, S., Shapiro, B. D., Kaileh, M., Hernandez, D., Sarantopoulou, D., et al. (2021). DNA Methylation Signatures Reveal that Distinct Combinations of Transcription Factors Specify Human Immune Cell Epigenetic Identity. *Immunity* 54, 2465–2480. e5. doi:10.1016/j.immuni.2021.10.001

Shao, T., Zhao, Z., Wu, A., Bai, J., Li, Y., Chen, H., et al. (2015). Functional Dissection of Virus-Human Crosstalk Mediated by miRNAs Based on the VmiReg Database. *Mol. Biosyst.* 11, 1319–1328. doi:10.1039/c5mb00095e

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer Statistics, 2020. *CA A Cancer J. Clin.* 70, 7–30. doi:10.3322/caac.21590

Wapinski, O., and Chang, H. Y. (2011). Long Noncoding RNAs and Human Disease. *Trends Cell Biol.* 21, 354–361. doi:10.1016/J.TCB.2011.04.001

Yang, Q., Wu, F., Mi, Y., Wang, F., Cai, K., Yang, X., et al. (2020). Aberrant Expression of miR-29b-3p Influences Heart Development and Cardiomyocyte Proliferation by Targeting NOTCH2. *Cell Prolif.* 53, 1–16. doi:10.1111/cpr.12764

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Cell and Developmental Biology

ORIGINAL RESEARCH
published: 11 November 2021
doi: 10.3389/fcell.2021.779367

Molecular Characterization of the Clinical and Tumor Immune Microenvironment Signature of 5-methylcytosine-Related Regulators in non-small Cell Lung Cancer

Taisheng Liu[1†], Liyi Guo[2†], Guihong Liu[3†], Xiaoshan Hu[4†], Xiaoning Li[1], Jinye Zhang[1], Zili Dai[5], Peng Yu[5], Ming Jiang[6], Jian Wang[1] and Jian Zhang[5]**

[1]Department of Thoracic Surgery, Affiliated Cancer Hospital and Institute of Guangzhou Medical University, Guangzhou, China, [2]The Sixth People's Hospital of Huizhou City, Huiyang Hospital Affiliated to Southern Medical University, Huizhou, China, [3]Department of Radiation Oncology, DongGuan Tungwah Hospital, Dongguan, China, [4]Department of Internal Medicine of Oncology, Affiliated Cancer Hospital and Institute of Guangzhou Medical University, Guangzhou, China, [5]Department of Radiation Oncology, Affiliated Cancer Hospital and Institute of Guangzhou Medical University, State Key Laboratory of Respiratory Diseases, Guangzhou Institute of Respiratory Disease, Guangzhou, China, [6]Department of Breast Surgery, Affiliated Cancer Hospital and Institute of Guangzhou Medical University, Guangzhou, China

Background: DNA methylation is an important epigenetic modification, among which 5-methylcytosine methylation (5mC) is generally associated with tumorigenesis. Nonetheless, the potential roles of 5mC regulators in the tumor microenvironment (TME) remain unclear.

Methods: The 5mC modification patterns of 1,374 lung adenocarcinoma samples were analyzed systematically. The correlation between the 5mC modification and tumor microenvironment cell infiltration was further assessed. The 5mCscore was developed to evaluate tumor mutation burden, immune check-point inhibitor response, and the clinical prognosis of individual tumors.

Results: Three 5mC modification patterns were established based on the clinical characteristics of 21 5mC regulators. According to the differential expression of 5mC regulators, three distinct 5mC gene cluster were also identified, which showed distinct TME immune cell infiltration patterns and clinical prognoses. The 5mCscore was constructed to evaluate the tumor mutation burden, immune check-point inhibitor response, and prognosis characteristics. We found that patients with a low 5mCscore had significant immune cell infiltration and increased clinical benefit.

OPEN ACCESS

Edited by:
Chunjie Jiang, University of Pennsylvania, United States

Reviewed by:
Weiru Liu, University of Pennsylvania, United States
Yang Xie, Brigham and Women's Hospital, United States
Chong Jin, University of Pennsylvania, United States

***Correspondence:**
Jian Zhang
zhangjian@gzhmu.edu.cn
Jian Wang
wangjian469@163.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular and Cellular Pathology, a section of the journal Frontiers in Cell and Developmental Biology

Received: *18 September 2021*
Accepted: *19 October 2021*
Published: *11 November 2021*

Citation:
Liu T, Guo L, Liu G, Hu X, Li X, Zhang J, Dai Z, Yu P, Jiang M, Wang J and Zhang J (2021) Molecular Characterization of the Clinical and Tumor Immune Microenvironment Signature of 5-methylcytosine-Related Regulators in non-small Cell Lung Cancer. Front. Cell Dev. Biol. 9:779367. doi: 10.3389/fcell.2021.779367

Abbreviations: 5mC: 5-methylcytosine methylation; TME: tumor microenvironment; NSCLC: (Siegel et al., 2020); LUAD: lung adenocarcinoma; LUSC: lung squamous carcinoma; ICI: The immune checkpoint inhibitors; PD-L1: programmed death-ligand 1; PD-1: programmed cell death 1; CTLA-4: cytotoxic T-lymphocyte antigen-4; GEO: Gene Expression Omnibus; SNV: single nucleotide variant; CNV: copy number variation; GSVA: gene set variation analysis; ssGSEA: single-sample gene-set enrichment analysis; DEGs: differentially expressed genes; ANOVA: one-way analysis of variance; KM: Kaplan–Meier; TMB: tumor mutation burden; GO: Gene ontology; BP: biological process; CC: cellular component; MF: molecular function; MDSCs: myeloid-derived suppressor cells.

Conclusion: This study indicated that the 5mC modification is involved in regulating TME infiltration remodeling. Targeting 5mC modification regulators might be a novel strategy to treat lung cancer.

Keywords: lung adenocarcinoma, 5mC, tumour microenvironment, immunotherapy, mutation burden

INTRODUCTION

Lung cancer is the primary cause of cancer-related deaths worldwide (Siegel et al., 2020) (NSCLC), accounting approximately for 85% of newly diagnosed lung cancer cases, is classified into lung adenocarcinoma (LUAD) and lung squamous carcinoma (LUSC) (Curran et al., 2011). For unresectable advanced NSCLC, a combination of radiotherapy and chemotherapy has been the most common first-line treatment (Yoda et al., 2019), and impressive clinical success has been observed using targeted therapies (Treat, 2005; Yuan et al., 2019; Alexander et al., 2020). Unfortunately, most NSCLC patients will suffer the relapse within 1 year (Fountzilas et al., 2021). Thus, understanding the mechanism and identifying novel targets to treat NSCLC remain an urgent clinical need.

Immunotherapies represent a promising advance in cancer treatment (Lussier et al., 2021). The immune checkpoint inhibitors (ICI), including programmed death-ligand 1 (PD-L1), programmed cell death 1 (PD-1), and cytotoxic T-lymphocyte antigen-4 (CTLA-4), combined with chemoradiotherapy, have been approved or are being widely evaluated in clinical trials (Grant et al., 2021). However, targeting PD-1 or PD-L1 has demonstrated durable efficacy only in a subset of patients with NSCLC (Jazieh et al., 2021). Thus, it is important to determine the underlying mechanisms with the aim of improving the curative effect.

DNA methylation is an epigenetic modification that is associated with regulating cell differentiation and tissue development (Smith and Meissner, 2013; Slieker et al., 2015). Dysregulation of DNA methylation patterns are important characteristics of several diseases, including cancers (Li et al., 2013; Božić et al., 2021; Cristall et al., 2021; Miyakuni et al., 2021). 5-Methylcytosine (5mC), a type of DNA methylation, was firstly reported by Wyatt, (1951). DNA 5mC methylation is the classic epigenetic process, which is controlled by "writers" (DNA methyltransferases), "erasers" (DNA methyltransferases), and "readers" (Ito et al., 2011; Du et al., 2015; Lio et al., 2020). With the discovery of 5mC regulators, recent studies suggested that DNA cytosine modifications may act as epigenetic markers in tumorigenesis (Wu and Zhang, 2010; Cavalcante et al., 2020; Jiang, 2020; Mo et al., 2020) and can regulate tumor microenvironment (TME) infiltrating cells (Chen et al., 2020; Zhao et al., 2021; Onodera et al., 2021). However, the comprehensive roles of TME cell infiltration directed by 5mC regulators in NSCLC remain unclear.

In this study, we evaluated 5mC methylation patterns comprehensively by analyzing genomic information of 1374 LUAD samples, and correlated the 5mC methylation pattern with the characteristics of TME cell infiltration. We identified three 5mC methylation patterns, and revealed that 5mC methylation mediation of TME cell infiltration characteristics was closely associated with the immune response phenotype, indicating the 5mC methylation played an important role in modifying TME characteristics. Furthermore, the 5mCscore could be applied as a promising biomarker to predict immune response and clinical outcome in NSCLC.

MATERIALS AND METHODS

Dataset Acquisition and Processing

Supplementary Figure S1 shows the workflow of the this study. mRNA expression with clinical and survival information were downloaded from Gene Expression Omnibus (GEO) and GDC data portal. Patients without clinical survival information were excluded. Five eligible lung adenocarcinoma cohorts (GSE19188, GSE31210, GSE37745, GSE50081, and TCGA-LUAD [lung adenocarcinoma data from The Cancer genome Atlas (TGCA)]) were included for further analysis (**Supplementary Table S1**). For background correction and normalization, the Robust Multichip Average algorithm was used to uniformly process the raw. CEL files of the four GEO datasets (Gautier et al., 2004). Next, a GEO meta-cohort were created by merging the GEO datasets using the R sva package (Leek et al., 2012).

Twenty-one 5mc regulators, including three writers (DNA methyltransferase 1 (DNMT1), DNA methyltransferase 3 Alpha (DNMT3A), DNA methyltransferase 3 beta (DNMT3B)), three erasers (tet methylcytosine dioxygenase 1 (TET1), tet methylcytosine dioxygenase 2 (TET2), tet methylcytosine dioxygenase 3 (TET3)), and 15 readers (methyl-cpg binding domain protein 1 (MBD1), methyl-cpg binding domain protein 2 (MBD2), methyl-cpg binding domain protein 3 (MBD3), methyl-cpg binding domain protein 4 (MBD4), methyl-cpg binding protein 2 (MECP2), nei like dna glycosylase 1 (NEIL1), *n*th like dna glycosylase 1 (NTHL1), single-strand-selective monofunctional uracil-dna glycosylase 1 (SMUG1), thymine dna glycosylase (TDG), ubiquitin like with phd and ring finger domains 1 (UHRF1), ubiquitin like with phd and ring finger domains 2 (UHRF2), uracil dna glycosylase (UNG), zinc finger and btb domain containing 33 (ZBTB33), zinc finger and btb domain containing 34 (ZBTB34), zinc finger and btb domain containing 4 (ZBTB4)) (Chen et al., 2020), and 23 tumor immune related cells from published studies (Zhang et al., 2020a; Zhao et al., 2021), were included for analysis. The transcriptomics data, single nucleotide variant (SNV), copy number variation (CNV), and 5mC phenotypic data were collected using the UCSC Xena database (https://xenabrowser.net) and the GDC data portal.

Unsupervised Clustering of 21 5mC Regulators

To identify 5mC regulator-mediated modification sub-clusters, unsupervised consensus clustering was used to cluster tumor samples into sub-clusters based on the expression levels of the 21 5mC regulators. To ensure the stability of the clusters, the parameters of clustering were as follows: number of repetitions = 1,000 bootstraps, clustering algorithm = k-means method, pFeature = 1.0, pItem = 0.8. The cluster with the most significant survival difference was included for further analysis.

Gene Set Variation Analysis and Functional Annotation

To explore the biological behavior among the different 5mC modification patterns, their pathway scores were evaluated using gene set variation analysis (GSVA) using the R GSVA package (Hänzelmann et al., 2013), with the "c2. cp.kegg.v7.4. symbols" gene set as the background. Differential pathways were further screened using $p < 0.05$ in the R package limma.

Estimation of the Tumor Microenvironment

To identify the TME cell infiltration in LUAD samples, the relative abundances of immune cells were quantified using the single-sample gene-set enrichment analysis (ssGSEA) algorithm. According to the method revealed by Charoentong et al. (2017b), various kinds of immune cells, including regulatory T cells, activated $CD8^+$ T cells, dendritic cells, and B cells, were evaluated. The relative abundance of TME infiltrating cells in clinical samples was represented by the enrichment scores.

Differentially Expressed Genes

To identify 5mC-related differentially expressed genes (DEGs), based on the expression levels of 21 5mC regulators, three distinct 5mC modification patterns were identified in the patients with LUAD. The empirical Bayesian approach of R package limma package was used for the difference analysis (Ritchie et al., 2015), which screened out 324 DEGs, 246 DEGs and 144 DEGs according to $p < 0.001$, $p < 0.0005$ and $p < 0.0001$. $p < 0.0005$ was most suitable for subsequent analysis.

Construction of 5mC Gene Signatures

Considering the heterogeneity and complexity of tumors, and according to the method used by Zhang J. et al. (2020), the 5mCscore was developed to quantify the modification pattern of individual patients with LUAD based on the identified DEGs. A univariate Cox regression model was used for the prognostic analysis of each gene in the 5mC signatures. We obtained 103 genes related to prognosis from among the 246 DEGs, and then principal component analysis (PCA) was performed, scored as PCi_1 and PCi2. This approach had advantage of focusing the score on the set with the largest block of well correlated (or anticorrelated) genes in the set, while down-weighting contributions from genes that do not track with other set members. The 5mC score of each patient was calculated as follows:

$$5mCscore = PCi_1 + PCi_2$$

Evaluation of Immune-Checkpoint Inhibitor Genomic and Clinical Information

To explore the application of the 5mC score to predict immune-checkpoint inhibitor (ICI) efficacy, the expression data and clinical annotations of the immunotherapeutic cohort of atezolizumab (IMvigor210 cohort) were downloaded from the website based on the Creative Commons 3.0 License (http://research-pub.Gene.com/imvigor210corebiologies) (Mariathasan et al., 2018).

Statistical Analysis

Correlation coefficients between the expression of 5mC regulators and the TME immune infiltration cells was conducted using the Spearman method and distance correlation analysis. The Wilcoxon test was used to analyze the difference between two groups. The Kruskal–Wallis test and one-way analysis of variance (ANOVA) were used to analyze difference among three or more groups. The log-rank test and the Kaplan–Meier (KM) method were applied to evaluate the survival time. A statistical two-sided p value < 0.05 was considered as having significance. All data processing in this study was done using R 3.6.1 software.

RESULTS

Genetic Variation and Expression Analysis of 5mC Methylation Regulators

According to the map described in **Figure 1A**, in this study, 21 5 mC methylation regulators (writers: DNMT1, DNMT3A and DNMT3B; erasers: TET1, TET2, TET3; readers: MBD1, MBD2, MBD3, MBD4, MECP2, NEIL1, NTHL1, SMUG1, TDG, UHRF1, UHRF2, UNG, ZBTB33, ZBTB38, and ZBTB4) were identified (**Supplementary Table S2**). To determine genetic alternations, we firstly evaluated the SNV variation frequency of the genes encoding the 21 5mC methylation regulators. As shown in **Figure 1B**, Among 561 LUAD samples, 21.39% of 5mC regulators had mutations. The main mutation type was missense_mutation. However, the mutation frequency of individual regulators only ranged from 0 to 4%. The CNV frequency of the 5 mC regulators showed that *MECP2*, *SMUG1*, *DNMT3B*, *ZBTB33*, and *NTHL1* had distinct CNV amplification, with frequencies of 11.71, 6.13, 5.58, 4.68, and 4.68%, respectively. *MBD3*, *UHRF1*, *MBD1*, *UHRF2*, and *TDG* had a CNV deletion, with frequencies of 6.30, 5.40, 5.58, 5.23, and 4.86%, respectively (**Figure 1C** and **Supplementary Table S3**). The distribution analysis of CNV alterations on 23 chromosomes showed that their distribution among the 21 5mC regulators was scattered and unorganized (**Figure 1D**). Survival analysis indicated that high expression of *DNMT3B*, *MDB2*, *MDB3*, *SMUG1*, *TDG*, *HURF1*, *UNG*, and *ZBTB38* were associated with poor survival of LUAD ($p < 0.05$); while, high expression of *MDB4*, *MECP2*, *NEIL1*, *TET2*, *UHRF2*, and *ZBTB4* were

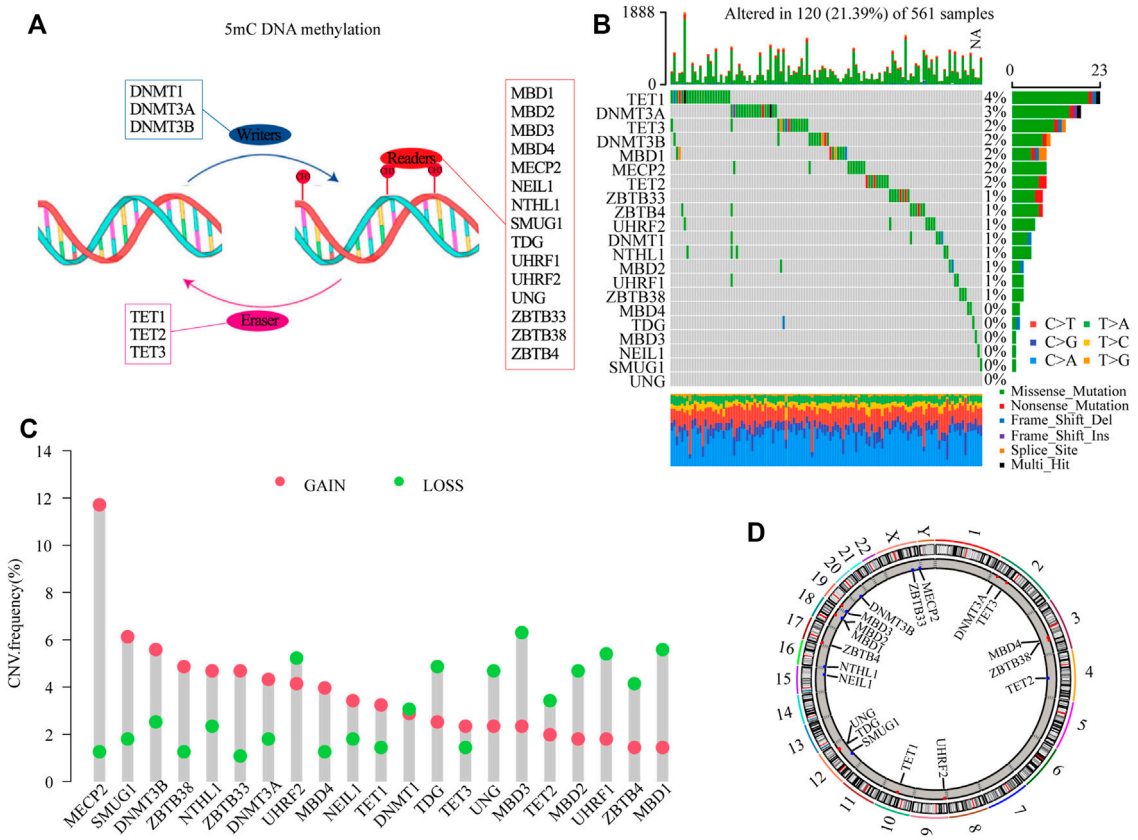


FIGURE 1 | Genetic landscape and expression analysis of 5mC regulators in LUAD. **(A)** Schematic diagram of 5mC DNA methylation mediated by 21 5mC regulators. **(B)** The mutation frequency of 21 5mC regulators in the TCGA-LUAD cohort. The column indicates individual patients. The upper barplot shows the TMB, The number on the right indicates the mutation frequency. The right barplot shows the percentage of mutation type in each regulator. The stacked barplot shows the fraction of conversions. **(C)** The CNV variation frequency of 5mC regulators in the TCGA-LUAD cohort. The height of the column indicates the alteration frequency of the regulators. The green dot is the deletion frequency; The red dot is the amplification frequency. **(D)** The locations of CNV alterations of 5mC regulators in the TCGA-LUAD cohort. CNV, copy number variation; 5mC, 5-methylcytosine; LUAD, lung adenocarcinoma; TCGA, The Cancer Genome Atlas; TMB, tumor mutation burden.

associated with better survival of LUAD ($p < 0.05$, **Supplementary Figure S2** and **Supplementary Table S4**).

Identification of 5mC Methylation-Related Phenotypes

To determine the roles of interaction among 5 mC methylation regulators in LUAD, correlation analysis among the 21 5 mC regulators was performed, which showed that there was a strong positive correlation among most of the regulators (**Supplementary Figure S3A** and **Supplementary Table S5**). The prognostic values of the 21 5 mC regulators in LUAD were evaluated using a univariate Cox regression model (**Supplementary Figure S3B**). As shown in **Figure 2A**, *MDB4*, *MECP2*, *NEIL1*, *TET2*, *ZBTB4*, and *ZBTB33* were favorable factors for overall survival (OS), while *DNMT1*, *DNMT3A*, *DNMT3B*, *TET1*, *TET3*, *SMUG1*, *TDG*, *UHRF1*, *UHRF2*, *UNG*, and *ZBTB38* were risk factors for OS. Significant negative correlations were obtained for UHRF1 and DNMT1, TDG and DNMT3A, TDG and UNG, MECP2 and ZBTB33, MECP2 and TET2, and TET2 and UHRF2 ($p < 0.001$). On the other hand, several erasers and readers also showed significant negative correlations: NTHL1 and TET2, NTHL1 and TET3, and MBD3 and TET2 ($p < 0.001$) (**Supplementary Tables S6–7**). Using unsupervised clustering analysis, three distinct 5mC modification patterns were identified based on the expression of 21 5mC regulators (**Supplementary Figure S4**). Prognostic analysis of the three 5mC modification clusters revealed a particularly prominent survival advantage for the 5mC cluster-B modification pattern (**Figure 2B** and **Supplementary Table S8**; $p = 0.001$). The results showed that cross-talk among the 5mC modification regulators might be involved in the formation of the 5mC modification and in the characteristics of TME cell infiltration.

Tumor Microenvironment Cell Infiltration Characteristics in the 5mC Methylation Clusters

To identify the potential function of the differentially expressed 5mC regulators, cluster analysis was first performed. As shown in **Figure 2C**, the 21 5mC regulators had a distinct distribution

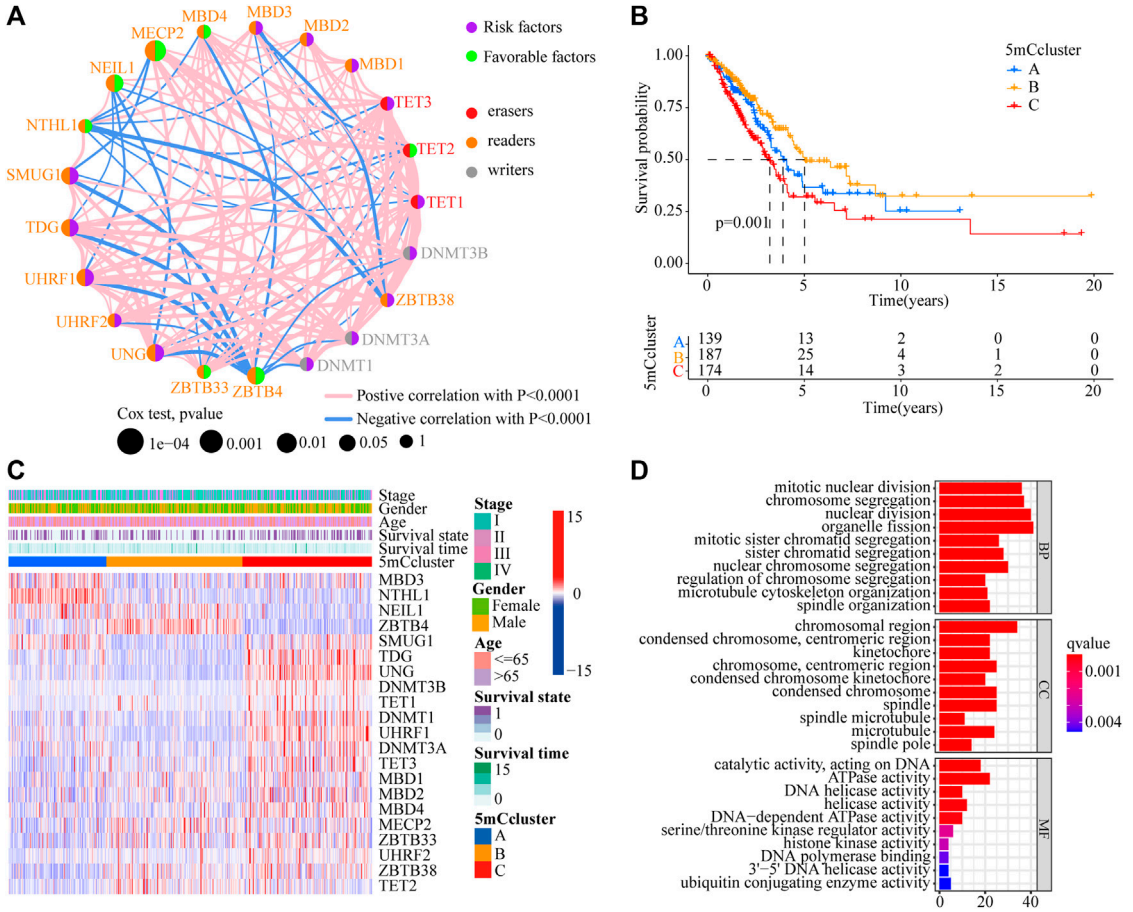


FIGURE 2 | Prognostic and biological characteristics of 5mC modification patterns in LUAD. **(A)** The interaction among 5mC regulators in the TCGA-LUAD cohort. The circle size indicates the effect of each regulator on prognosis. Green dots indicate favorable factors for prognosis; Purple dots indicate risk factor for prognosis. The lines linking regulators indicate their interactions, and thickness show the correlation strength between the regulators. Negative correlations are marked in blue and positive correlation in red. **(B)** Survival analyses of 5mC modification patterns in the TCGA-LUAD cohort, including 500 cases in 5mC cluster A (n = 139), 5mC cluster B (n = 187), and 5mC cluster C (n = 174) (p < 0.0001, Log-rank test). **(C)** Cluster analysis of 21 5mC regulators among the three 5mC modification patterns. **(D)** Gene ontology (GO) analysis of 21 5mC regulators among the three 5mC modification patterns.

among the three 5mC clusters. Gene ontology (GO) analysis was performed to identify the biological process (BP), cellular component (CC), and molecular function (MF) of the regulators. The aberrantly expressed 5mC regulators were mainly enriched for GO terms related to regulation of mitotic nuclear division, chromosome segregation, and nuclear division (BP); chromosomal region, condensed chromosome/centromeric region, and kinetochore (CC); and ATPase activity, DNA helicase activity, and helicase activity (MF) (**Figure 2D**). To further identify the potential behaviors, GSVA enrichment analysis was performed, as shown in **Supplementary Table S9**. To further exlpore unsupervised consensus clustering of all tumor samples for the molecular classification of LUAD. The optimal number of clusters was determined by the K value. After assessing relative changes in the area under the cumulative distribution function curve and consensus matrix heatmap, we selected a three-cluster solution (K = 3), which showed no appreciable increase in the area under the cumulative distribution function curve (**Supplementary Figure S4**). 5mC cluster A was markedly enriched in damage repair-related pathways, such as base excision repair, DNA replication, spliceosome, and RNA polymerase. 5mC cluster B was prominently related to immune activation-related pathways, such as the JAK-STAT signaling pathway, the T cell receptor signaling pathway, and the calcium signaling pathway. 5mC cluster C was mostly associated with carcinogenic activation and damage repair pathways, such as, the p53 signaling pathway, basal transcription factors, spliceosome, RNA degradation, DNA replication, base excision repair, homologous recombination, DNA replication, and mismatch repair (**Figures 3A–C**). Based on the expression levels of these 21 5mC regulators, the three 5mC modification patterns could be partially differentiated using PCA (**Figure 3D**). TME cell infiltration analysis showed 5mC cluster B was associated with activated B cells, activated dendritic cells, mast cells, natural killer T cells, and

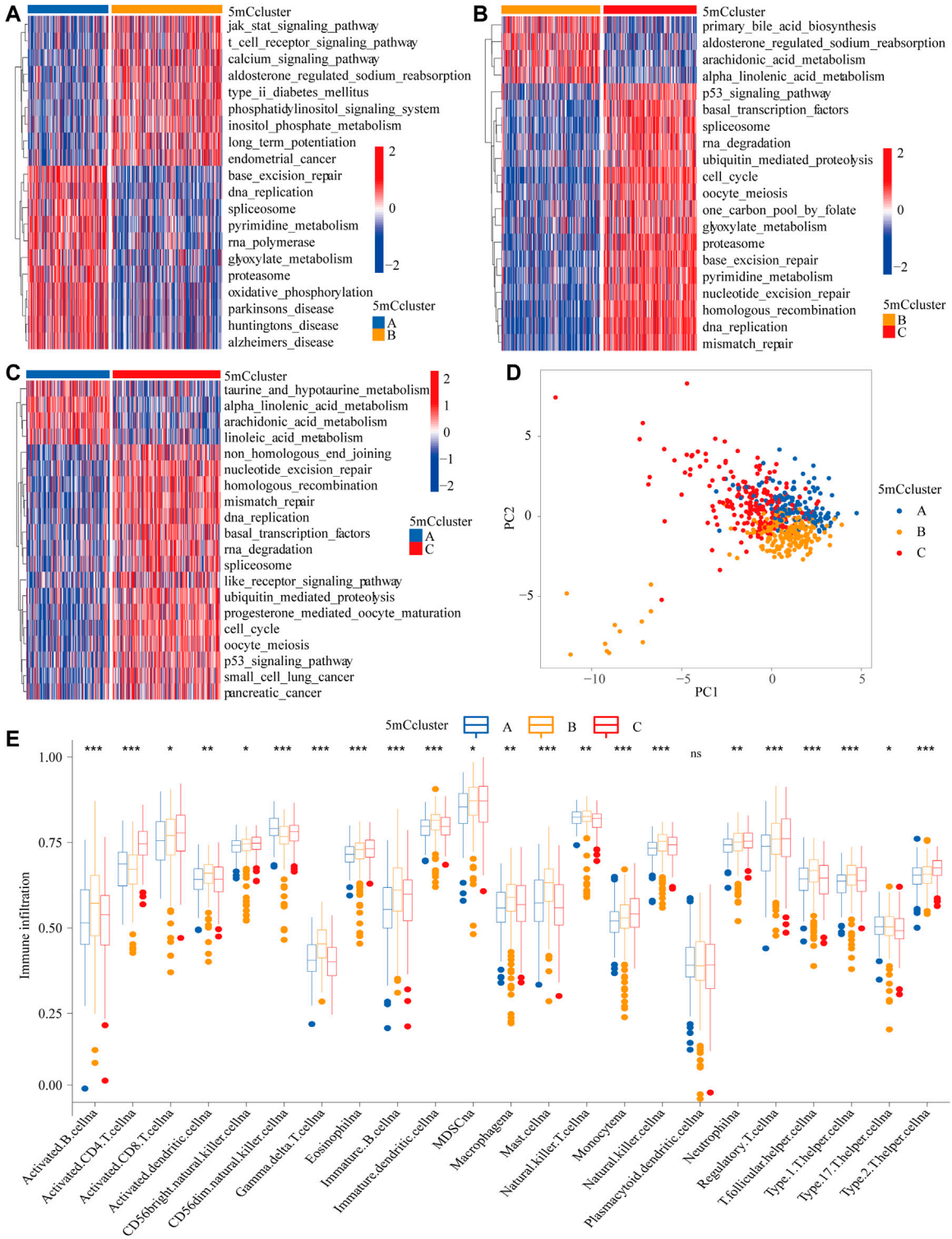


FIGURE 3 | GSVA enrichment analysis and TME cell infiltration characteristics of 5mC modification patterns. **(A–C)** The states of biological pathways among the three 5mC modification patterns enriched by GSVA analysis The general biological processes are shown as a heatmap, red represents activated pathways and blue represents inactivated pathways. **(A)** 5 mC cluster A *vs* 5mC cluster B; **(B)** 5mC cluster B *vs* 5mC cluster C; **(B)** 5 mC cluster A *vs* 5mC cluster C. **(D)** Principal component analysis of the 5mC modification patterns. **(E)** The abundance of TME infiltrating cells in the 5mC modification patterns (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ns means not significant).

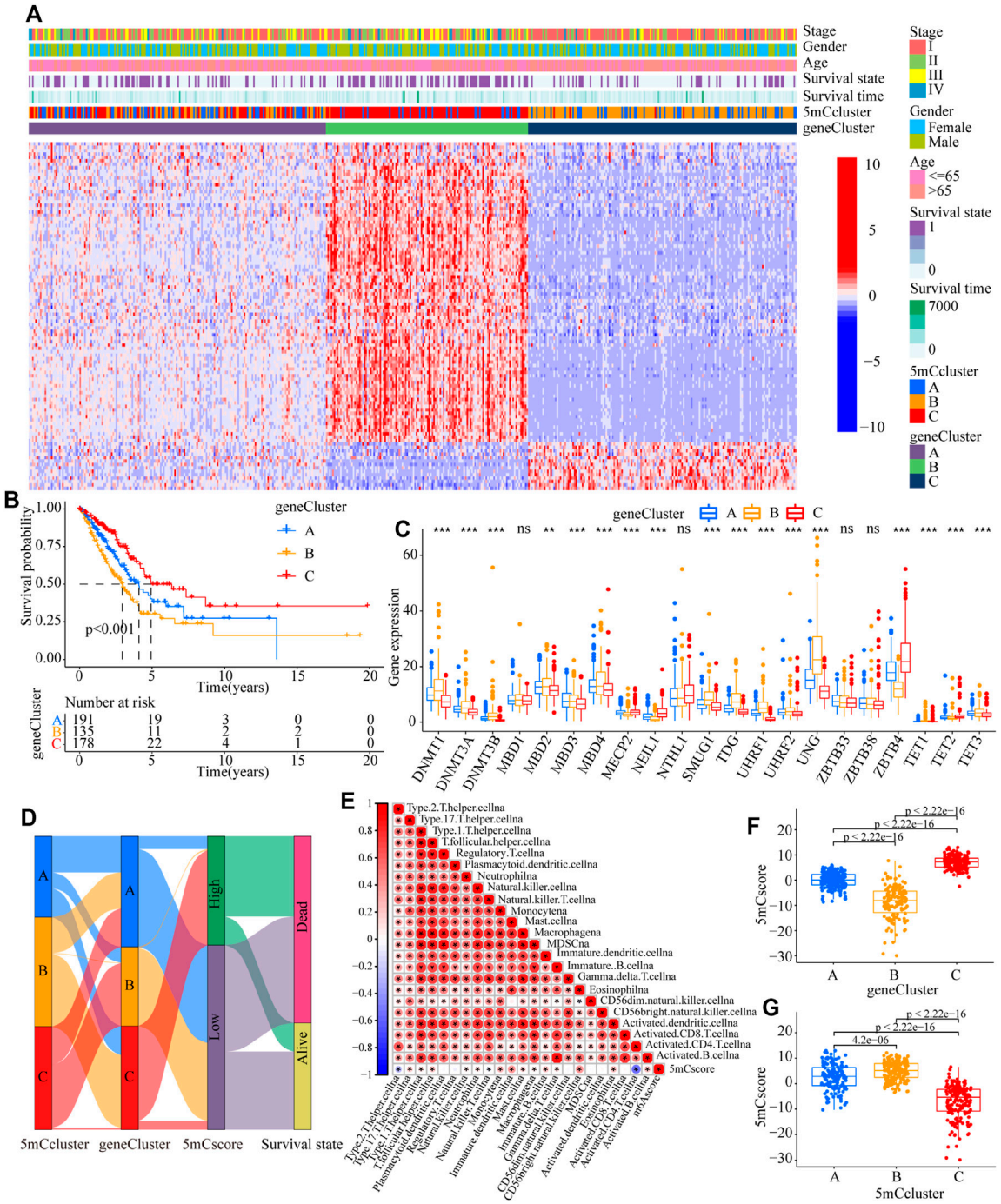


FIGURE 4 | Construction of 5mC gene signatures. **(A)** Unsupervised clustering of overlapping 5mC phenotype-related genes in the TCGA-LUAD cohort to classify patients into different genomic subtypes, termed as 5mC gene cluster **(A–C)**, respectively. The gene clusters, 5mC clusters, tumor stage, survival status, sex, and age were used as patient annotations. **(B)** Overall survival of patients with the three 5mC modification genomic clusters in the TCGA-LUAD cohort, including 504 cases in 5mC gene cluster A ($n = 191$), 5mC gene cluster B ($n = 135$), and 5mC gene cluster C ($n = 17$) ($p < 0.0001$, Log-rank test). **(C)** The expression of 21 5mC regulators in the three gene clusters (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ns means not significant). **(D)** Alluvial diagram showing the changes in 5mC clusters, 5mC gene cluster, 5mCscore, and survival. **(E)** Correlations between the 5mCscore and the known gene signatures in the TCGA-LUAD cohort using Spearman analysis. Negative correlations are marked in blue and positive correlation in red. **(F)** Differences in the 5mCscore among three gene clusters in the TCGA-LUAD cohort (***$p < 0.001$, Kruskal-Wallis test). **(G)** Differences in the 5mCscore among three the 5mC modification patterns in the TCGA-LUAD cohort (***$p < 0.001$, Kruskal-Wallis test).

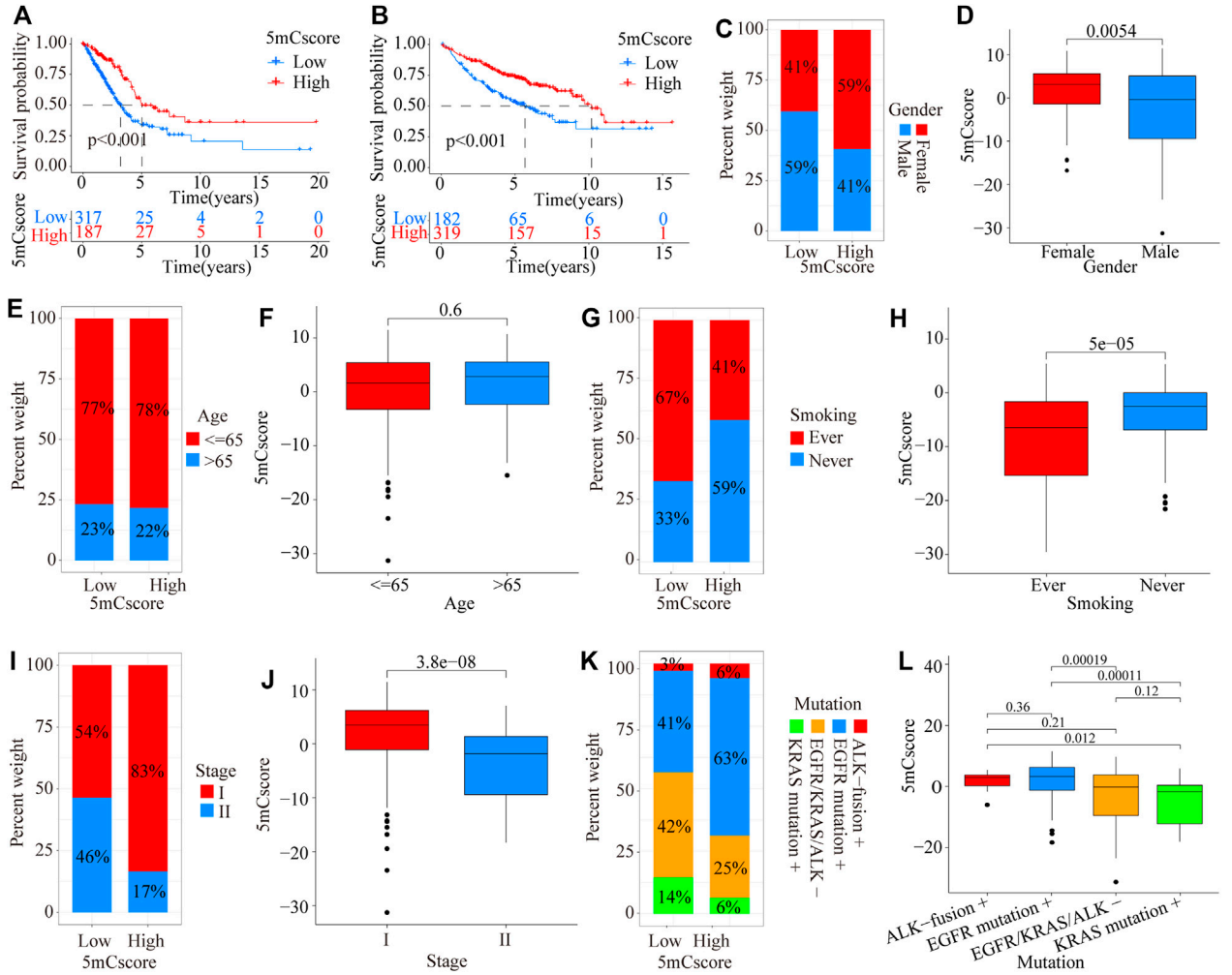


FIGURE 5 | Prognostic and genetic characteristics between high and low 5mCscore groups. **(A)** Survival analysis of the 5mCscore in the TCGA-LUAD cohort ($p < 0.0001$, Log-rank test). **(B)** Survival analysis of the 5mCscore in the GEO-meta cohort ($p < 0.0001$, Log-rank test). **(C)** Sex proportion between the high- and low-5mCscore groups. **(D)** The 5mCscore difference between females and males. **(E)** Age proportion between the high- and low-5mCscore groups. **(F)** The 5mCscore difference between age (≤65) and age (>65). **(G)** Smoking status proportion between the high- and low-5mCscore groups. **(H)** The 5mCscore difference between smoking status (ever) and smoking status (never). **(I)** Clinical stage status proportion between the high- and low-5mCscore groups. **(J)** The 5mCscore difference between stage I and stage II. **(K)** The 5mCscore difference between genetic mutations (–) and genetic mutations (+). **(L)** Genetic mutations status proportion between high- and low-5mCscore groups.

neutrophils (**Figure 3E** and **Supplementary Table S10**, $p < 0.001$). 5mC cluster C was remarkably rich in immune cell infiltration including myeloid-derived suppressor cells (MDSCs), regulatory T cells, type 1 T helper cells, type 2 T helper cells, and type 17 T helper cell (**Figure 3E** and **Supplementary Table S10**, $p < 0.001$). Prognosis analysis showed that patients with different 5mC modification patterns also had a matching survival advantage (**Figure 2B**, $p = 0.001$). Based on the above results, cluster A, characterized by innate immune cell infiltration, was defined as an immune-excluded phenotype; cluster B, characterized by adaptive immune cell infiltration and immune activation, was defined as an immune-inflamed phenotype; and cluster C, characterized by the inhibition of immunity, was defined as an immune-desert phenotype.

Identification of 5mC Methylation Gene Signature

To further identify the potential function of each m5C modification pattern, we determined 246 m5C phenotype-related DEGs (**Supplementary Table S11**). GO analysis showed that the 246 DEGs were associated with cell cycle, RNA transport, spliceosome, DNA replication, base excision, and human T-cell leukemia virus 1 infection (**Supplementary Figure S5A** and **Supplementary Table S12**). Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis indicated that the 5mC gene clusters were involved in DNA transcription and translation (**Supplementary Figure S5B** and **Supplementary Table S13**). To further determine the potential regulation mechanism, unsupervised clustering analyses was performed to identify the

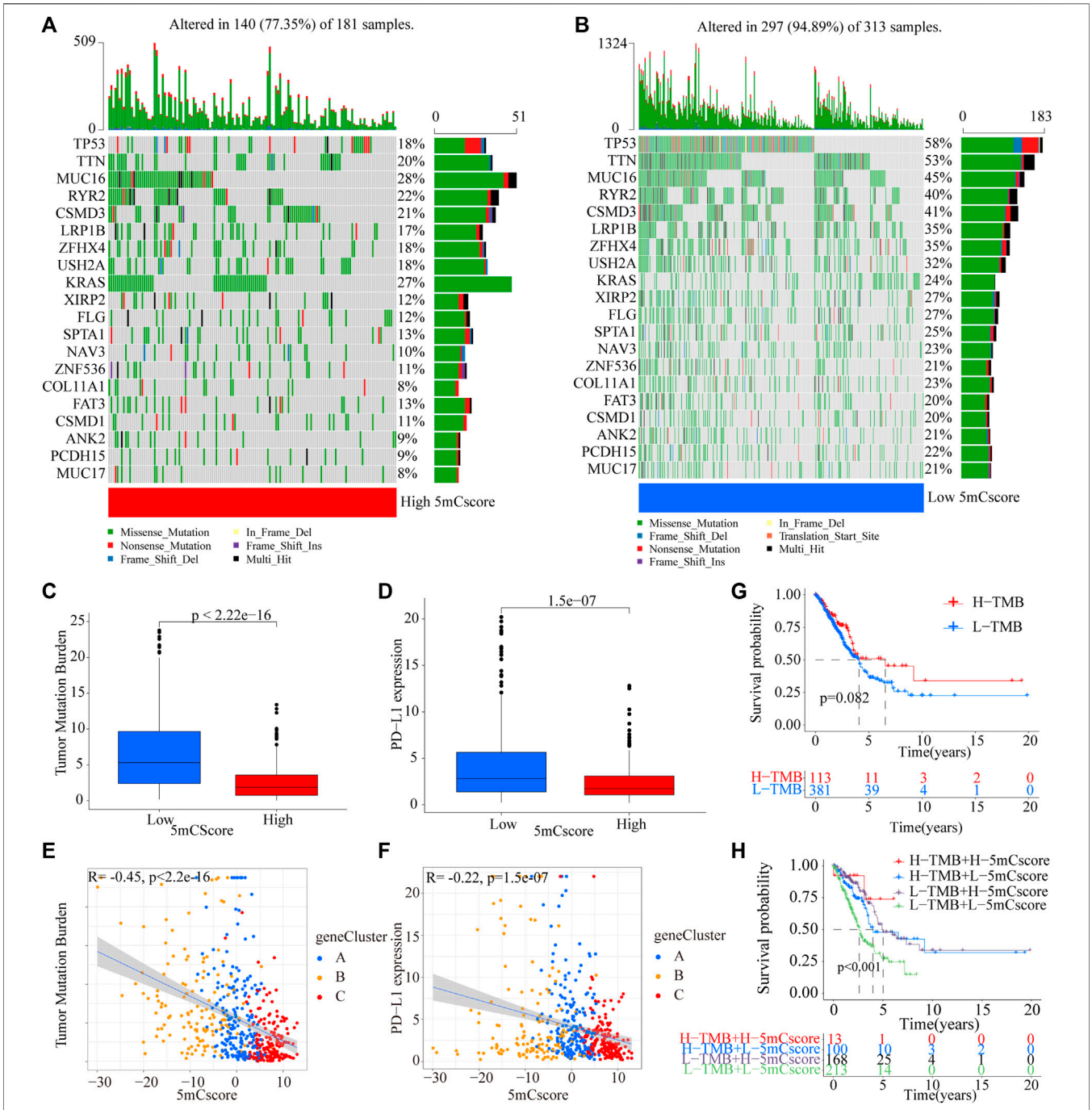


FIGURE 6 | Characteristics of 5mCscore in the TCGA molecular subtypes and tumor somatic mutations. **(A,B)** Waterfall plot of tumor somatic mutations established by those with a high 5mCscore **(A)** and a low 5mCscore **(B)**. Each column represents individual patients. The upper barplot shows the TMB, the number on the right indicates the mutation frequency in each gene. The right barplot shows the proportion of each variant type. **(C)** Tumor somatic mutation between high 5mCscore and low 5mCscore groups. **(D)** PD-L1 expression difference between high 5mCscore and low 5mCscore groups. **(E)** The correlation analysis between tumor somatic mutation and the 5mCscore. **(F)** The correlation analysis between PD-L1 expression and the 5mCscore. **(G)** Survival analysis of tumor somatic mutations in the TCGA-LUAD cohort ($p < 0.0001$, Log-rank test). **(H)** Survival analyses for patients stratified by both the 5mCscore and the tumor somatic mutation burden using Kaplan–Meier curves ($p < 0.0001$, Log-rank test).

genomic subtypes based on the 103 prognostic genes from the 246 5mC phenotype-related DEGs. The results showed that three distinct 5 mC genomic phenotypes (5mC gene Cluster A–C) could be identified (**Figure 4A** and **Supplementary Table S5C–J**). These results indicated that the 5mC methylation modification patterns did exist in LUAD and three distinct

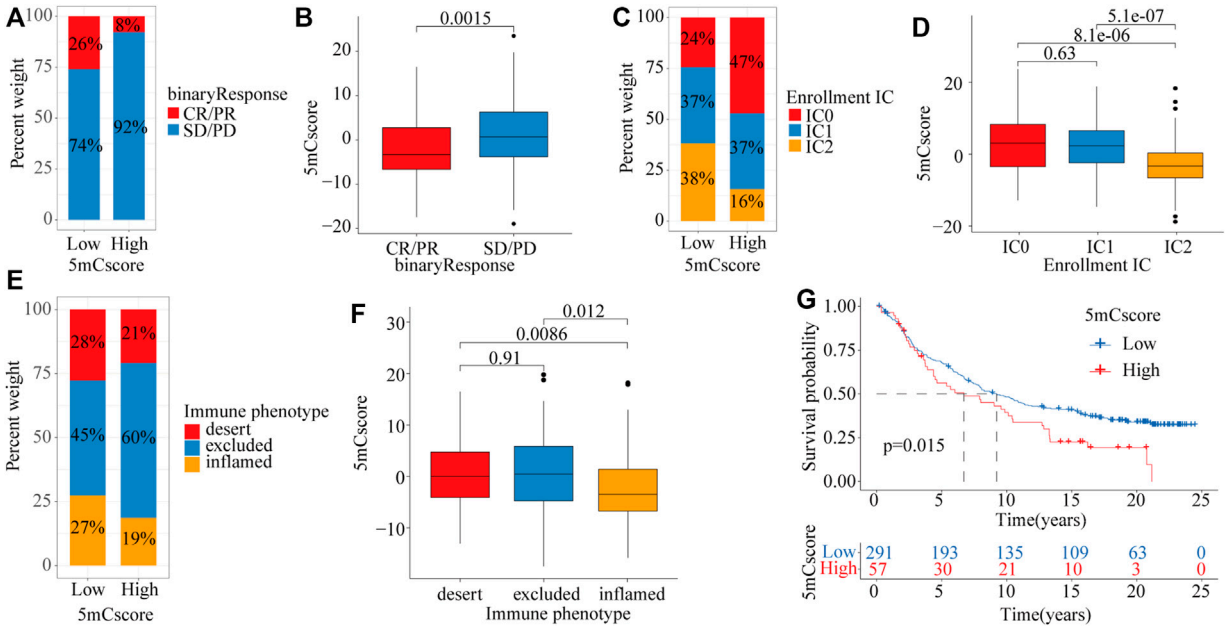


FIGURE 7 | The role of the 5mCscore in anti-PD-L1 immunotherapy. **(A)** The proportion of patients with a response to ICI in the low or high 5mCscore groups. Responder/Nonresponder: 26%/74% in the low 5mCscore groups and 8%/92% in the high 5mCscore groups. **(B)** 5mCscore differences between responders and nonresponders. **(C)** IC infiltration proportion between high 5mCscore and low 5mCscore groups. **(D)** 5mCscore differences between different IC subgroups. **(E)** Immune phenotype proportion between high 5mCscore and low 5mCscore groups. **(F)** 5mCscore differences among the immune-desert phenotype, immune-excluded phenotype, and immune-inflamed phenotype. **(G)** Survival analyses for low (n = 291) and high (n = 57) 5mCscore patient groups in the anti-PD-L1 immunotherapy cohort using Kaplan–Meier curves (IMvigor210 cohort; p = 0.015, Log-rank test). CR, complete response; IC, immune cell; ICI, immune check-point inhibitor; PD, progressive disease; PR, partial response; SD, stable disease.

5mC gene clusters were characterized by different signature genes. Cluster analysis showed that 178 of 504 patients with LUAD were clustered in 5mC gene cluster C, which was associated with better prognosis. Patients with LUAD with 5mC gene cluster B (n = 135) had poorer prognosis. 5mC gene cluster A, with 191 patients clustered, had an intermediate prognosis (**Figure 4B**, $p < 0.001$). The expression levels of the 5mC regulators among the 5mC gene clusters were distinctly different (**Figure 4C**).

Clinical Characteristics of 5mCscore Phenotypes

To better explore the pattern of 5mC modification in individual patients, based on the 5mC phenotype-related genes (**Supplementary Table S14**), the 5mCscore was used to quantify the 5mC modification patterns of individual patients with LUAD. An alluvial diagram was applied to clarify the attributed changes of the LUAD patients. As shown in **Figure 4D**, the 5mC modification patterns clusters were almost consistent with the 5mC gene clusters, i.e., the 5mC gene cluster B group patients mainly had a low 5mCscore, which was associated with poor survival. To determine the roles of 5mC-related phenotypes in immune regulation, correlation analysis showed that the 5mCscore was associated positively with most TME infiltrating cells (**Figure 4E**). The Kruskal–Wallis test revealed there was a significant difference in the 5mCscore among the 5mC gene clusters. 5mC gene cluster C showed the highest median 5mCscore, while 5mC gene cluster B had the lowest median 5mCscore, which indicated that a high 5mCscore was closely associated with immune activation-related signatures, whereas a low 5mCscore was associated with immune inactivation-related signatures (**Figure 4F**, $p < 0.001$). More importantly, compared with the other clusters, 5mC modification cluster C presented the lowest median 5mCscore, and 5mC modification cluster B showed the highest 5mCscore (**Figure 4G**, $p < 0.001$). These results indicated that a high 5mCscore correlated significantly with immune-activation and the 5mCscore could be used to identify the 5mC modification patterns in LUAD, and further assess the characteristics of TME cell infiltration of individual tumors.

To further validate the value of the 5mCscore, patients in the TCGA cohort were divided into low or high 5mCscore groups. Prognosis analysis showed that patients with a high 5mCscore showed a better survival benefit (**Figure 5A**, $p < 0.001$). Four GEO datasets (GSE19188, GSE31210, GSE37745, and GSE50081, **Supplementary Table S1**) were integrated into one meta-cohort. Survival analysis in the GEO meta-cohort also identified that a high 5mCscore was linked to a better clinical outcome (**Figure 5B**, $p < 0.001$). These results indicated that the 5mCscore could act as an independent prognostic biomarker to evaluate patient outcomes. To explore the effect of clinical characteristics on the 5mCscore, the subgroups of clinical

characteristics were further analyzed. A significant distribution difference of a high 5mCscore was observed for gender (59% in female *vs* 41% in male, $p = 0.0054$; **Figures 5C,D**), smoking status (41% *vs* 67% for ever smoking, P = 5e-05; **Figures 5G,H**), stage I–II (83% *vs* 54% for stage I, $p = 3.8e-08$; **Figures 5I,J**), and genetic mutations (63% *vs* 41% for EGFR mutations, $p = 0.00019$; 25% *vs* 42% in EGFR/KRAS/ALK mutations, $p < 0.001$; **Figures 5K-L**). However, there were no 5mCscore differences between age (≤65) and age (>65) (**Figures 5E,F**, $p = 0.6$). To assess the value of clinical characteristics, patients in the TCGA-LUAD cohort were further stratified by age (≤65/> 65), sex (female/male), T stage (T1–2/T3–4), N stage (N0–1/N2–3), M stage (M0/M1), and clinical stage (I–II/III–IV). We found that the clinical characteristics, particularly T1–2, N0–1, M0, and I–II clinical stages, could be clearly divided into high- and low-risk subgroups (**Supplementary Figure S6**). These results indicated that multiple clinical characteristics can have an effect on the 5mCscore, which led to the heterogeneity of 5mC regulators in LUAD.

The Potential of the 5mCscore to Predict the Response to anti-PD-L1 Immunotherapy

The above analyses demonstrated the impact of 5mCscore regulators on the TME, as well as on the prognosis in patients with LUAD. The genetic characteristics of the patients in different 5mCscore groups were further explored. As shown in **Figures 6A,B** and **Supplementary Table S15**, the somatic mutation landscapes in the high and low 5mCscore groups had a distinct difference. The mutation frequency was 77.35% in the high 5mCscore group and 94.89% in the low 5mCscore group. Specifically, except for *KRAS*, *TP53* (18% *vs* 58%), *TTN* (20% *vs* 53%), *MUC16* (28% *vs* 45%), and *RYR2* (22% *vs* 40%) had important differences between the high and low 5mCscore groups (**Figure 6B**). Besides, patients with a low 5mCscore showed a significantly higher tumor mutation burden (TMB) and PD-L1 expression than patients with a high 5mCscore (**Figures 6C,D** and **Supplementary Table S16**). 5mC gene cluster C showed lower PD-L1 expression and a lower TMB than 5mC gene cluster B. Correlation analysis further identified that the TMB and PD-L1 expression were related negatively with the 5 mCscore (**Figures 6E,F**, $p < 0.001$). These results revealed a significant association between the 5mCscore and the TMB and PD-L1 expression. These factors are important parameters in the assessment of immunotherapy outcomes. However, the survival analysis associated with the TMB found that there was no difference between the high and low TMB groups (**Figure 6G**, $p = 0.082$). Next, the crosstalk between the 5mCscore and TMB in terms of patient survival was investigated. The high 5mCscore and high TMB group had better survival than the low 5mCscore and high TMB group. The low 5mCscore and low TMB group was associated with poorer survival relative to those with a high 5mCscore and low TMB (**Figure 6H**, $p < 0.001$).

To explore the potential roles of the 5mCscore in clinical immune therapy of lung cancer, we investigated whether the 5mCscore could predict patients' response to PD-L1 (atezolizumab) therapy based on the PD-L1 immunotherapy cohort (IMvigor210). Compared with those with a high 5mCscore, patients with a low 5mCscore had significant therapeutic advantages and clinical responses to anti-PD-L1 immunotherapy (**Figures 7A,B** and **Supplementary Figure S7A–B**, $p = 0.0015$). The low 5mCscore group had a higher immune cells 2 (IC2) score (38% *vs* 16%) and a lower tumor cells 2+ (TC2+) score (77% *vs* 96%) than the high 5mCscore group, 5mCscore was significantly associated with the enrollment ICs and suppression of TCs (**Figures 7C,D** and **Supplementary Figure S7C–D**). These results identified that the 5mCscore played a non-negligible role in regulating TME immune cell infiltration. We further investigated different immune phenotypes among the high and low 5mCscore groups and found that a higher 5mCscore was markedly associated with exclusion and desert immune phenotypes, in which an antitumor effect is difficult to exert using ICI therapy (**Figures 7E,F**). Patients with a low 5mCscore exhibited significant clinical benefits and a markedly prolonged survival (**Figure 7G**, $p = 0.015$). These results clarified that 5mC modification patterns are significantly associated with immune phenotypes and PD-L1 expression, and that the 5mCscore could be a prominent biomarker to predict the response to ICI therapy.

DISCUSSION

DNA 5mC methylation is a dynamic and reversible post-transcriptional modification regulated by 5mC related regulators (Mayer et al., 2000; Oswald et al., 2000; Wu et al., 2020). Recent research highlighted the biological importance of 5mC modification on immune cell infiltration and tumor suppression (Schübeler, 2015; Dor and Cedar, 2018; Weng et al., 2021). However, most studies focused only on a single TME cell type or one 5mC related regulator, and the comprehensive roles of 5mC regulators on TME infiltration characteristics are not fully elaborated. Thus, further clarification of the potential roles of 5mC modification patterns in the infiltration of TME cells will raise our awareness of the effects of the heterogeneity and complexity of the TME on the response to ICI therapy and provide a novel biomarkers to evaluate the ICI response and predict prognosis.

Herein, three distinct 5mC methylation modification patterns were identified based on 21 5mC regulators. The patterns had significantly distinct TME cell infiltration characteristics. Based on the identified 246 5mC phenotype-related DEGs, three genomic clusters of 5mC-related genes were further identified, which were also validated for their association with transcription modification and immune infiltration. Recent studies had shown that DNA methylation can be involved in the maintenance and reinforcement of T cell exhaustion gene signatures (Pauken et al., 2016; Gate et al., 2018). In murine antigen-specific CD8 T cells, DNMT3A-mediated methylation impaired T cell expansion and led to immune cell exhaustion under treatment with anti-PD-1 *via* repression the expression of key genes (Ghoneim et al., 2017).

By contrast, in the context of T cell exhaustion, the involvement of DNA methylation in the reprogramming of the T cells has also been reported (Araki et al., 2013), such as demethylation of the *PD1* promoter resulting in permanent CD8+T cell exhaustion. Uhrf1-mediated tnf-α gene methylation controlled proinflammatory macrophages in experimental colitis resembling inflammatory bowel disease (Qi et al., 2019). In breast cancer, ZBTB33 subcellular partitioning functionally linked LC3A/B, the tumor microenvironment, and cancer survival (Singhal et al., 2021). These results indicated that the 5mC modification is intimately involved in shaping TME landscapes.

Epigenetic alterations are associated extensively with the immune response and tumore evasion. A DNA methylation signature (the EPIMMUNE signature) has been identified as an epigenetic biomarker of the response to ICI. The multicenter and retrospective analysis revealed that the EPIMMUNE signature could predict the response to anti-PD-1 treatment in non-small-cell lung cancer (Seremet et al., 2016). In metastatic melanoma treated with CTLA-4 blockers, responders and non-responders to ICI had a differential DNA methylation pattern (Chida et al., 2021). To better understand the individual heterogeneity of TME-meditated 5mC modification patterns, the 5mCscore was established to assess the 5mC modification pattern of individuals with LUAD. 5mC gene cluster C, characterized by an immune inflamed phenotype, exhibited a higher 5mCscore, and 5mC gene cluster B, characterized by an immune excluded phenotype, had a lower 5mCscore. These results revealed the 5mCscore was a useful biomarker to comprehensively assess individual tumor 5mC modification patterns, which could be used to evaluate TME immune cell infiltration patterns. Prognosis analyses also identified that the 5mCscore was an independent prognostic biomarker in LUAD.

Alterations in 5mC regulatory genes might also be associated with variations in LUAD. In this study, we identified twenty driver genes, including *TP53*, *TTN*, *MUC16*, *RYR2*, and CSMD3. Moreover, variations in *KRAS* were associated significantly with alterations in 5mC regulatory genes. As an oncogene, *KRAS* mutations were reported frequently in a variety of tumors, including colorectal cancer (Prior et al., 2012), pancreatic cancer (Arner et al., 2019), and bladder cancer (Santha et al., 2020). Recent studies identified that *KRAS* might have a critical role in the immunoregulation of NSCLC (Li et al., 2021; Wang et al., 2021). Our data also revealed that the 5mCscore had a markedly negative correlation with PD-L1 expression and the TMB. The 5mCscore integrating the TMB could be the more effective biomarker to predict ICI response. We also identified the predictive value of the 5mCscore in the IMvigor210 cohort. The 5mCscore between non-responders and responders was significantly different. These results provided new insights to clarify different tumor immune phenotypes and improve the clinical response to ICI therapy.

CONCLUSION

In summary, we comprehensively analyzed the potential mechanisms of 5mC methylation modification during the regulation of the TME. 5mC modification patterns contributed to the heterogeneity and complexity of the TME in LUAD, which was significantly associated with TMB, PD-L1 expression, and immune phenotypes. 5mCscore could act as a biomarker to predict a patient's response to ICI therapy.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

TL and JZ performed all experiments, prepared figures, and drafted the manuscript. LG, XH, XL, and JZ participated in data analysis and interpretation of the results. GL, JZ, XH, ZD, PY, MJ, and JW designed the study and participated in the data analysis. All authors have read and approved the manuscript.

FUNDING

This study was supported by grants from the National Natural Science Foundation of China (82003212), Discipline Construction Project of Guangzhou Medical University during the 14th Five-Year Plan (06-410-2107181), Guangzhou Key Medical Discipline Construction Project Fund (02-412-B205002-1004042), and the Medical and Health Technology Projects of Guangzhou (2015A011086).

ACKNOWLEDGMENTS

We thank the staff members of The Cancer Genome Atlas for their involvement in the cBioPortal for Cancer Genomics Program.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.779367/full#supplementary-material

REFERENCES

Alexander, M., Kim, S. Y., and Cheng, H. (2020). Update 2020: Management of Non-small Cell Lung Cancer. *Lung* 198 (6), 897–907. doi:10.1007/s00408-020-00407-5

Araki, K., Youngblood, B., and Ahmed, R. (2013). Programmed Cell Death 1-directed Immunotherapy for Enhancing T-Cell Function. *Cold Spring Harbor Symposia Quantitative Biol.* 78, 239–247. doi:10.1101/sqb.2013.78.019869

Arner, E. N., Du, W., and Brekken, R. A. (2019). Behind the Wheel of Epithelial Plasticity in KRAS-Driven Cancers. *Front. Oncol.* 9, 1049. doi:10.3389/fonc.2019.01049

Božić, T., Kuo, C.-C., Hapala, J., Franzen, J., Eipel, M., Platzbecker, U., et al. (2021). Investigation of Measurable Residual Disease in Acute Myeloid Leukemia by DNA Methylation Patterns. *Leukemia*. doi:10.1038/s41375-021-01316-z

Cavalcante, G. M., Borges, D. P., de Oliveira, R. T. G., Furtado, C. L. M., Alves, A. P. N. N., Sousa, A. M., et al. (2020). Tissue Methylation and Demethylation Influence Translesion Synthesis DNA Polymerases (TLS) Contributing to the Genesis of Chromosomal Abnormalities in Myelodysplastic Syndrome. *J. Clin. Pathol.*, jclinpath, 2020. doi:10.1136/jclinpath-2020-207131

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017b). Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cel Rep.* 18 (1), 248–262. doi:10.1016/j.celrep.2016.12.019

Chen, Y. T., Shen, J. Y., Chen, D. P., Wu, C. F., Guo, R., Zhang, P. P., et al. (2020). Identification of Cross-Talk between m6A and 5mC Regulators Associated with Onco-Immunogenic Features and Prognosis across 33 Cancer Types. *J. Hematol. Oncol.* 13 (1), 22. doi:10.1186/s13045-020-00854-w

Chida, K., Kotani, D., Masuishi, T., Kawakami, T., Kawamoto, Y., Kato, K., et al. (2021). The Prognostic Impact of KRAS G12C Mutation in Patients with Metastatic Colorectal Cancer: A Multicenter Retrospective Observational Study. *Oncol.* 26 (10), 845–853. doi:10.1002/onco.13870

Cristall, K., Bidard, F.-C., Pierga, J.-Y., Rauh, M. J., Popova, T., Sebbag, C., et al. (2021). A DNA Methylation-Based Liquid Biopsy for Triple-Negative Breast Cancer. *Npj Precis. Onc.* 5 (1), 53. doi:10.1038/s41698-021-00198-9

Curran, W. J., Paulus, R., Langer, C. J., Komaki, R., Lee, J. S., Hauser, S., et al. (2011). Sequential vs Concurrent Chemoradiation for Stage III Non-small Cell Lung Cancer: Randomized Phase III Trial RTOG 9410. *JNCI J. Natl. Cancer Inst.* 103 (19), 1452–1460. doi:10.1093/jnci/djr325

Dor, Y., and Cedar, H. (2018). Principles of DNA Methylation and Their Implications for Biology and Medicine. *The Lancet* 392 (10149), 777–786. doi:10.1016/S0140-6736(18)31268-6

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. Cel Biol* 16 (9), 519–532. doi:10.1038/nrm4043

Fountzilas, E., Lampaki, S., Koliou, G.-A., Koumarianou, A., Levva, S., Vagionas, A., et al. (2021). Real-world Safety and Efficacy Data of Immunotherapy in Patients with Cancer and Autoimmune Disease: The Experience of the Hellenic Cooperative Oncology Group. *Cancer Immunol. Immunother*. doi:10.1007/s00262-021-02985-6

Gate, R. E., Cheng, C. S., Aiden, A. P., Siba, A., Tabaka, M., Lituiev, D., et al. (2018). Genetic Determinants of Co-accessible Chromatin Regions in Activated T Cells across Humans. *Nat. Genet.* 50 (8), 1140–1150. doi:10.1038/s41588-018-0156-2

Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004). Affy--analysis of Affymetrix GeneChip Data at the Probe Level. *Bioinformatics* 20 (3), 307–315. doi:10.1093/bioinformatics/btg405

Ghoneim, H. E., Fan, Y., Moustaki, A., Abdelsamed, H. A., Dash, P., Dogra, P., et al. (2017). De Novo epigenetic Programs Inhibit PD-1 Blockade-Mediated T Cell Rejuvenation. *Cell* 170 (1), 142–157. doi:10.1016/j.cell.2017.06.007

Grant, M. J., Herbst, R. S., and Goldberg, S. B. (2021). Selecting the Optimal Immunotherapy Regimen in Driver-Negative Metastatic NSCLC. *Nat. Rev. Clin. Oncol.* 18 (10), 625–644. doi:10.1038/s41571-021-00520-1

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC Bioinformatics* 14, 7. doi:10.1186/1471-2105-14-7

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., et al. (2011). Tet Proteins Can Convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333 (6047), 1300–1303. doi:10.1126/science.1210597

Jazieh, A. R., Onal, H. C., Tan, D. S. W., Soo, R. A., Prabhash, K., Kumar, A., et al. (2021). Real-World Treatment Patterns and Clinical Outcomes in Patients with Stage III NSCLC: Results of KINDLE, a Multicountry Observational Study. *J. Thorac. Oncol.* 16 (10), 1733–1744. doi:10.1016/j.jtho.2021.05.003

Jiang, S. (2020). Tet2 at the Interface between Cancer and Immunity. *Commun. Biol.* 3 (1), 667. doi:10.1038/s42003-020-01391-5

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The Sva Package for Removing Batch Effects and Other Unwanted Variation in High-Throughput Experiments. *Bioinformatics* 28 (6), 882–883. doi:10.1093/bioinformatics/bts034

Li, B., Lu, Q., Song, Z. G., Yang, L., Jin, H., Li, Z. G., et al. (2013). Functional Analysis of DNA Methylation in Lung Cancer. *Eur. Rev. Med. Pharmacol. Sci.* 17 (9), 1191–1197.

Li, T., Pang, X., Wang, J., Wang, S., Guo, Y., He, N., et al. (2021). Exploration of the Tumor-Suppressive Immune Microenvironment by Integrated Analysis in EGFR-Mutant Lung Adenocarcinoma. *Front. Oncol.* 11, 591922. doi:10.3389/fonc.2021.591922

Lio, C. J., Yue, X., Lopez-Moyado, I. F., Tahiliani, M., Aravind, L., and Rao, A. (2020). TET Methylcytosine Oxidases: New Insights from a Decade of Research. *J. Biosci.* 45, 21. doi:10.1007/s12038-019-9973-4

Lussier, D. M., Alspach, E., Ward, J. P., Miceli, A. P., Runci, D., White, J. M., et al. (2021). Radiation-induced Neoantigens Broaden the Immunotherapeutic Window of Cancers with Low Mutational Loads. *Proc. Natl. Acad. Sci. USA* 118 (24), e2102611118. doi:10.1073/pnas.2102611118

Mariathasan, S., Turley, S. J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., et al. (2018). TGFβ Attenuates Tumour Response to PD-L1 Blockade by Contributing to Exclusion of T Cells. *Nature* 554 (7693), 544–548. doi:10.1038/nature25501

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the Zygotic Paternal Genome. *Nature* 403 (6769), 501–502. doi:10.1038/35000656

Miyakuni, K., Nishida, J., Koinuma, D., Nagae, G., Aburatani, H., Miyazono, K., et al. (2021). Genome-wide Analysis of DNA Methylation Identifies the Apoptosis-related Gene UQCRH as a Tumor Suppressor in Renal Cancer. *Mol. Oncol.* doi:10.1002/1878-0261.13040

Mo, Z., Cao, Z., Luo, S., Chen, Y., and Zhang, S. (2020). Novel Molecular Subtypes Associated with 5mC Methylation and Their Role in Hepatocellular Carcinoma Immunotherapy. *Front. Mol. Biosci.* 7, 562441. doi:10.3389/fmolb.2020.562441

Onodera, A., González-Avalos, E., Lio, C.-W. J., Georges, R. O., Bellacosa, A., Nakayama, T., et al. (2021). Roles of TET and TDG in DNA Demethylation in Proliferating and Non-proliferating Immune Cells. *Genome Biol.* 22 (1), 186. doi:10.1186/s13059-021-02384-1

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., et al. (2000). Active Demethylation of the Paternal Genome in the Mouse Zygote. *Curr. Biol.* 10 (8), 475–478. doi:10.1016/s0960-9822(00)00448-6

Pauken, K. E., Sammons, M. A., Odorizzi, P. M., Manne, S., Godec, J., Khan, O., et al. (2016). Epigenetic Stability of Exhausted T Cells Limits Durability of Reinvigoration by PD-1 Blockade. *Science* 354 (6316), 1160–1165. doi:10.1126/science.aaf2807

Prior, I. A., Lewis, P. D., and Mattos, C. (2012). A Comprehensive Survey of Ras Mutations in Cancer. *Cancer Res.* 72 (10), 2457–2467. doi:10.1158/0008-5472.CAN-11-2612

Qi, S., Li, Y., Dai, Z., Xiang, M., Wang, G., Wang, L., et al. (2019). Uhrf1-Mediated Tnf-α Gene Methylation Controls Proinflammatory Macrophages in Experimental Colitis Resembling Inflammatory Bowel Disease. *J.I.* 203 (11), 3045–3053. doi:10.4049/jimmunol.1900467

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

Santha, S., Ling, X., Aljahdali, I. A. M., Rasam, S. S., Wang, X., Liao, J., et al. (2020). Mutant Kras as a Biomarker Plays a Favorable Role in FL118-Induced Apoptosis, Reactive Oxygen Species (ROS) Production and Modulation of Survivin, Mcl-1 and XIAP in Human Bladder Cancer. *Cancers* 12 (11), 3413. doi:10.3390/cancers12113413

Schübeler, D. (2015). Function and Information Content of DNA Methylation. *Nature* 517 (7534), 321–326. doi:10.1038/nature14192

Seremet, T., Koch, A., Jansen, Y., Schreuer, M., Wilgenhof, S., Del Marmol, V., et al. (2016). Molecular and Epigenetic Features of Melanomas and Tumor Immune

Microenvironment Linked to Durable Remission to Ipilimumab-Based Immunotherapy in Metastatic Patients. *J. Transl. Med.* 14 (1), 232. doi:10.1186/s12967-016-0990-x

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer Statistics, 2020. *CA A. Cancer J. Clin.* 70 (1), 7–30. doi:10.3322/caac.21590

Singhal, S. K., Byun, J. S., Park, S., Yan, T., Yancey, R., Caban, A., et al. (2021). Kaiso (ZBTB33) Subcellular Partitioning Functionally Links LC3A/B, the Tumor Microenvironment, and Breast Cancer Survival. *Commun. Biol.* 4 (1), 150. doi:10.1038/s42003-021-01651-y

Slieker, R. C., Roost, M. S., Van Iperen, L., Suchiman, H. E. D., Tobi, E. W., Carlotti, F., et al. (2015). DNA Methylation Landscapes of Human Fetal Development. *Plos Genet.* 11 (10), e1005583. doi:10.1371/journal.pgen.1005583

Smith, Z. D., and Meissner, A. (2013). DNA Methylation: Roles in Mammalian Development. *Nat. Rev. Genet.* 14 (3), 204–220. doi:10.1038/nrg3354

Treat, J. (2005). Incorporating Novel Agents with Gemcitabine-Based Treatment of NSCLC. *Lung Cancer* 50 (Suppl. 1), S8–S9. doi:10.1016/s0169-5002(05)81551-x

Wang, H., Chen, S., Meng, D., Wu, C., Zhu, J., Jiang, M., et al. (2021). Tumor Mutation burden and Differentially Mutated Genes Among Immune Phenotypes in Patients with Lung Adenocarcinoma. *Ott.* Vol. 14, 2953–2965. doi:10.2147/OTT.S294993

Weng, R. R., Lu, H. H., Lin, C. T., Fan, C. C., Lin, R. S., Huang, T. C., et al. (2021). Epigenetic Modulation of Immune Synaptic-Cytoskeletal Networks Potentiates γδ T Cell-Mediated Cytotoxicity in Lung Cancer. *Nat. Commun.* 12 (1), 2163. doi:10.1038/s41467-021-22433-4

Wu, C.-Y., Zhang, B., Kim, H., Anderson, S. K., Miller, J. S., and Cichocki, F. (2020). Ascorbic Acid Promotes KIR Demethylation during Early NK Cell Differentiation. *J.I.* 205 (6), 1513–1523. doi:10.4049/jimmunol.2000212

Wu, S. C., and Zhang, Y. (2010). Active DNA Demethylation: Many Roads lead to Rome. *Nat. Rev. Mol. Cel Biol.* 11 (9), 607–620. doi:10.1038/nrm2950

Wyatt, G. R. (1951). Recognition and Estimation of 5-methylcytosine in Nucleic Acids. *Biochem. J.* 48 (5), 581–584. doi:10.1042/bj0480581

Yoda, S., Dagogo-Jack, I., and Hata, A. N. (2019). Targeting Oncogenic Drivers in Lung Cancer: Recent Progress, Current Challenges and Future Opportunities. *Pharmacol. Ther.* 193, 20–30. doi:10.1016/j.pharmthera.2018.08.007

Yuan, M., Huang, L.-L., Chen, J.-H., Wu, J., and Xu, Q. (2019). The Emerging Treatment Landscape of Targeted Therapy in Non-small-cell Lung Cancer. *Sig Transduct Target. Ther.* 4, 61. doi:10.1038/s41392-019-0099-9

Zhang, B., Wu, Q., Li, B., Wang, D., Wang, L., and Zhou, Y. L. (2020a). m6A Regulator-Mediated Methylation Modification Patterns and Tumor Microenvironment Infiltration Characterization in Gastric Cancer. *Mol. Cancer* 19 (1), 53. doi:10.1186/s12943-020-01170-0

Zhang, J., Zhong, X., Jiang, H., Jiang, H., Xie, T., Tian, Y., et al. (2020b). Comprehensive Characterization of the Tumor Microenvironment for Assessing Immunotherapy Outcome in Patients with Head and Neck Squamous Cell Carcinoma. *aging* 12 (22), 22509–22526. doi:10.18632/aging.103460

Zhao, Q., Wang, W., Li, J. X., Yuan, P., Liu, Y., Li, Y., et al. (2021). The DNA Cytosine-5-Methyltransferase 3 (DNMT3) Involved in Regulation of CgIL-17 Expression in the Immune Response of Oyster Crassostrea gigas. *Dev. Comp. Immunol.* 123, 104092. doi:10.1016/j.dci.2021.104092

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Cell and Developmental Biology

ORIGINAL RESEARCH
published: 08 December 2021
doi: 10.3389/fcell.2021.772436

Integrative 5-Methylcytosine Modification Immunologically Reprograms Tumor Microenvironment Characterizations and Phenotypes of Clear Cell Renal Cell Carcinoma

Wenhao Xu[1,2†], Wenkai Zhu[1,2†], Xi Tian[1,2†], Wangrui Liu[3†], Yuanyuan Wu[4†], Aihetaimujiang Anwaier[1,2], Jiaqi Su[1,2], Shiyin Wei[3†], Yuanyuan Qu[1,2]*, Hailiang Zhang[1,2]* and Dingwei Ye[1,2]**

[1]Department of Urology, Fudan University Shanghai Cancer Center, Shanghai, China, [2]Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China, [3]Department of Neurosurgery, Affiliated Hospital of Youjiang Medical University for Nationalities, Baise, China, [4]Department of Gastroenterology, Naval Medical Center of PLA, Naval Military Medical University, Shanghai, China

The tumor microenvironment (TME) affects the biologic malignancy of clear cell renal cell carcinoma (ccRCC). The influence of the 5-methylcytosine (m^5C) epigenetic modification on the TME is unknown. We comprehensively assessed m^5C modification patterns of 860 ccRCC samples (training, testing, and real-world validation cohorts) based on 17 m^5C regulators and systematically integrated the modification patterns with TME cell-infiltrating characterizations. Our results identified distinct m^5C modification clusters with gradual levels of immune cell infiltration. The distinct m^5C modification patterns differ in clinicopathological features, genetic heterogeneity, patient prognosis, and treatment responses of ccRCC. An elevated m^5C score, characterized by malignant biologic processes of tumor cells and suppression of immunity response, implies an immune-desert TME phenotype and is associated with dismal prognosis of ccRCC. Activation of exhausted T cells and effective immune infiltration were observed in the low m^5C score cluster, reflecting a noninflamed and immune-excluded TME phenotype with favorable survival and better responses to immunotherapy. Together, these findings provide insights into the regulation mechanisms of DNA m^5C methylation modification patterns on the tumor immune microenvironment. Comprehensive assessment of tumor m^5C modification patterns may enhance our understanding of TME cell-infiltrating characterizations and help establish precision immunotherapy strategies for individual ccRCC patients.

OPEN ACCESS

Edited by:
Chunjie Jiang, University of Pennsylvania, United States

Reviewed by:
Youyang Shi, Shanghai University of Traditional Chinese Medicine, China
Anli Zhang, University of Texas Southwestern Medical Center, United States
Huiyu Li, University of Texas Southwestern Medical Center, United States in collaboration with reviewer AZ.

***Correspondence:**
Shiyin Wei
yjweishiyin@163.com
Yuanyuan Qu
quyy1987@163.com
Hailiang Zhang
zhanghl918@163.com
Dingwei Ye
dwyelie@163.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular and Cellular Pathology, a section of the journal Frontiers in Cell and Developmental Biology

Received: *08 September 2021*
Accepted: *01 November 2021*
Published: *08 December 2021*

Citation:
Xu W, Zhu W, Tian X, Liu W, Wu Y, Anwaier A, Su J, Wei S, Qu Y, Zhang H and Ye D (2021) Integrative 5-Methylcytosine Modification Immunologically Reprograms Tumor Microenvironment Characterizations and Phenotypes of Clear Cell Renal Cell Carcinoma. Front. Cell Dev. Biol. 9:772436. doi: 10.3389/fcell.2021.772436

Abbreviations: ccRCC, clear cell renal cell carcinoma; CI, confidence interval; CNV, copy numbers variation; CPTAC, Clinical Proteomic Tumor Analysis Consortium; DEGs, differentially expressed genes; FUSCC, Fudan University Shanghai Cancer Center; GSEA, gene set enrichment analysis; GSVA, gene set variation analysis; HE, hematoxylin-eosin; HR, hazard ratio; ICTs, immune checkpoint therapies; ICGC, International Cancer Genome Consortium; IHC, immunohistochemistry; RCC, renal cell carcinoma; ROC, receiver operating characteristic curve; SNP, single nucleotide polypeptides; TCGA, The Cancer Genome Atlas; TME, tumor microenvironment.

Keywords: clear cell renal cell carcinoma, 5-methylcytosine, tumor microenvironment, renal cell carcinoma (RCC) clear cell renal cell carcinoma (CCRCC), immune checkpoint therapies, prognosis, machine learning algorithm

INTRODUCTION

Renal cell carcinoma (RCC) is the most common malignancy of the urinary system, accounting for approximately 3.8% of all newly diagnosed cancers. The incidence of RCC is increasing by 1.1% each year (Siegel et al., 2020). Clear cell RCC (ccRCC), which originates from proximal tubule epithelial cells, is the most common histology type of RCC, accounting for approximately 80% of all RCC cases (Capitanio and Montorsi, 2016; Linehan and Ricketts, 2019). The Von Hippel–Lindau (VHL) gene is frequently mutated in ccRCC, and mutations in *BAP1*, *PBRM1*, *SETD2*, and *PIK3CA* are also commonly observed in ccRCC. Studies show that mutations in these genes influence the prognosis and treatment response of ccRCC patients (Clark et al., 2019; Linehan and Ricketts, 2019). The standard first-line treatment strategy for metastatic or advanced ccRCC mainly involves tyrosine kinase inhibitors, such as sunitinib and sorafenib, that target vascular endothelial growth factor receptors. Over the past few decades, rapid progress has been made in immunotherapy as a new treatment strategy for cancer (Xu et al., 2019; Braun et al., 2020; Motzer et al., 2020).

DNA methylation is one of the most researched epigenetic modifications and is linked to the development of human malignancies (Qian et al., 2020). The main type of DNA methylation is the presence of an additional methyl group on the 5 position of cytosine (5-methylcytosine, m^5C) (Choi et al., 2021). The m^5C modification was the first discovered epigenetic marker and plays an important role in regulating the transcriptome profiles and carcinogenesis process of solid tumors, which often harbor aberrant DNA methylation (Martisova et al., 2021). The m^5C modification is frequently found in large clusters called CpG islands, which are present in gene-promoter regions and suppress gene transcription (Chen et al., 2019; Palei et al., 2020). A series of enzymes, called writers, readers, and erasers, is responsible for adding, recognizing, and removing the m^5C modifications, respectively (Rausch et al., 2020). Some tumor-suppressor genes are silenced as a consequence of hypermethylation in the promoter regions. Therefore, DNA methylation represents a potential signature and promising treatment target for human malignancies. Investigation of m^5C epigenetic modifications and their regulation of gene expression may, thus, provide insights into the mechanisms underlying cancer development.

The 5-methylcytidine modification occurs on both DNA and RNA. The major epigenetic mark in mammalian DNA is m^5C, which is associated with carcinogenesis and tumorigenesis of various cancers (Greenberg and Bourc'his, 2019). The phenotype of tumor microenvironment (TME) is dynamically regulated by cell signaling transduction and epigenetic drivers, which are critical factors influencing the efficacy of immunotherapy and both extrinsic and intrinsic resistance pathways. DNA methyltransferase enzymes (DNMTs) methylate CpG islands in gene promoters, and aberrant expression or activity of DNMTs can lead to tumorigenesis and aggressive progression (Zhang et al., 2018). Additionally, upregulated DNMT1 is shown to be necessary for maintaining cancer stemness and is associated with poor clinical outcome of cancers. DNMT1 is also shown to regulate the inhibitory function of $Foxp3^+$ T-regulated cells (Piperi et al., 2008; Wang et al., 2013; Zagorac et al., 2016). Therefore, comprehensively exploring the biological activities of epigenetic drivers in tumor phenotypes and TME is important (Xu et al., 2021a).

In this study, we examine the potential influence of DNA m^5C regulators on the clinical malignant characteristics and TME of ccRCC. We first constructed m^5C clusters using large-scale samples and algorithms and evaluated the relationship of m^5C clusters with immune cell infiltration, the DNA variation landscape, and immunotherapy in ccRCC.

MATERIALS AND METHODS

Sample Collection and Data Preprocessing

Gene expression, copy number variants, tumor somatic mutations, and matched clinical information of ccRCC from The Cancer Genome Atlas (TCGA) cohort were obtained. Gene expression data of 93 ccRCC tumors from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) were obtained at https://proteomics.cancer.gov/programs/cptac. In addition, RNA-seq and clinical data of 91 ccRCC patients from the RECA-EU cohort were available from the International Cancer Genome Consortium (ICGC, https://dcc.icgc.org/) database and included in this study. Patients without overall survival information were removed from further analysis. In addition, 232 ccRCC samples with proteomics sequencing data with available clinical and pathologic electronic records were enrolled from our institute, Fudan University Shanghai Cancer Center (FUSCC, Shanghai, China). In total, 860 ccRCC tumor samples were included for analysis. Batch effects from nonbiological technical biases were corrected using the "ComBat" algorithm of sva package and the fragments per kilobase of transcript per million values were transformed into transcripts per kilobase million values.

Unsupervised Clustering for 17 m^5C Regulators

A total of 17 m^5C regulators were extracted from the integrated gene expression profiles to identify different m^5C modification patterns. The 17 m^5C regulators included three writers (*DNMT1*, *DNMT3A*, and *DNMT3B*), three erasers (*TET1*, *TET2*, and *TET3*), one regulating factor (*DNMT3L*), and 10 readers (*MECP2*, *MBD1*, *MBD2*, *MBD3*, *MBD4*, *UHRF1*, *UHRF2*, *ZBTB4*, *ZBTB38*, and *ZFP57*). The "ConsensusClusterPlus" R package was used to classify patients for further analysis, and 1000 times repetitions were conducted to ensure the stability of the classification

(Wilkerson and Hayes, 2010). Overall survival was compared between patients with different m^5C modification patterns.

Gene Set Variation Analysis (GSVA) and TME Cell Infiltration Estimation

GSVA, a commonly employed method for estimating the variation in pathways (Hänzelmann et al., 2013), was used to evaluate the potential biological differences between the m^5C modification patterns using the "GSVA" R package. To estimate the TME cell infiltration, we applied the single-sample gene set enrichment analysis algorithm to evaluate the relative abundance of immune cells in the ccRCC TME. The reference gene sets for quantifying the immune cells were obtained from a previous study (Charoentong et al., 2017), and the examined immune cells included mast cells, monocyte, macrophage, activated $CD4^+$ T cells, and other types of immune cells. Immune cell abundance was compared between m^5C modification patterns, and the prognostic significance of each immune cell was also evaluated based on the overall survival information.

Differential Gene Expression Analysis and Functional Enrichment Analysis

To explore the potential biological differences between m^5C modification patterns, the limma package was used to identify differentially expressed genes (DEGs), and the threshold value was set as $p < .05$, $|\log FC| \geq 3$ (Ritchie et al., 2015). Functional enrichment analyses were carried out to explore the potential functions of the DEGs. The expression profiles of DEGs were extracted, and unsupervised clustering was applied again to identify the subgroups; the subgroups were defined as m^5C gene clusters.

Identifying m^5C Score as the m^5C Gene Signature

A scoring system was constructed to evaluate the m^5C modification patterns, and we termed it as m^5C score. Univariate Cox regression was used to evaluate the prognostic value for each gene, and the genes with prognostic significance were extracted for further analysis. Random forest analysis and principal component analysis were used to construct the m^5C relevant gene signature. Both principal components 1 and 2 were enrolled to calculate the signature scores, and the m^5C score was defined as follows: $m^5C \cdot score = \sum(PC2_i + PC2_i)$.

Copy Number Variant Analysis, Immunotherapy Response Prediction, and IC50 Evaluation

To explore potential associations between copy number variants and m^5C score, Genomic Identification of Significant Targets in Cancer (version 2.0) was used to identify significantly amplified or deleted regions using TCGA copy number data (Beroukhim et al., 2007; Mermel et al., 2011). $Q \leq 0.05$ was defined as significant, and the confidence interval was set to 0.95. Tumor immune dysfunction and exclusion (TIDE) was used to estimate the immunotherapy response based on the expression profiles (Jiang et al., 2018). Thus, the associations between m^5C score and immunotherapy response were evaluated. The pRRophetic package was used to predict the half-maximal inhibitory concentration (IC50) of chemotherapy drugs in the high and low m^5C score groups.

Immunohistochemistry (IHC)

IHC was performed to evaluate the expression levels of Ki-67 (ab15580; Abcam), CD4 (RMA-0620, Maxim, China), CD8 (RMA-0514, Maxim, China), Glut-1 (ab115730; Abcam), PD-L1 (ab205921; Abcam), CXCL13 (ab246518; Abcam), TGF-β (ab189778; Abcam), FASN (ab99359; Abcam), CK (Kit-0009, Maxim, China), and FoxP3 (98,377, CST) following previously described procedures (Xu et al., 2021a; Xu et al., 2021b). Opal multispectral was implemented to identify differential immune cell infiltration and PD-L1 expression in different groups on a multispectral imaging system (Vectra® Polaris™, Shanghai, China).

Statistical Analysis

A Wilcox test was used to compare differences between two groups. The Kaplan–Meier method was used to conduct survival analysis, and the cutoff value was defined *via* the survminer package. A log-rank test was used to detect the significance. The receiver operating characteristic curve (ROC) was drawn to evaluate the predictive ability for immunotherapy response.

RESULTS

The Overall Depiction of Genetic Variation of m^5C Regulators in ccRCC

A total of 17 m^5C regulators including three writers, three erasers, one regulating factor, and 10 readers were manually identified in this study. The dynamic reversible process of m^5C DNA methylation mediated by regulators as well as their potential biological functions for ccRCC are summarized in **Figure 1A**. We detected significant differences in the expressions of m^5C regulators between ccRCC and para-cancer tissues ($p < .05$) (**Figure 1B**). Analysis of CNV frequency indicated that CNV alterations were prevalent in the 17 m^5C regulators, and half of the m^5C regulators more frequently showed copy number amplification compared with copy number loss (**Figure 1C**). Besides this, in DNA variation profiles, we found 19 experienced samples of m^5C regulators with a frequency of 5.65% among 336 ccRCC samples from TCGA (**Figure 1D**). The location of CNV alterations of the m^5C regulators on chromosomes is shown in **Figure 1E**. Notably, ccRCC samples could be distinguished from normal samples completely based on the expression pattern of these m^5C regulators (**Figure 1F**). These findings suggest a high degree of m^5C modification–mediated intertumoral heterogeneity of genetic and expressional alteration landscape between ccRCC and adjacent normal samples, suggesting that the aberrant expression of m^5C regulators may play an essential role in ccRCC malignancy.

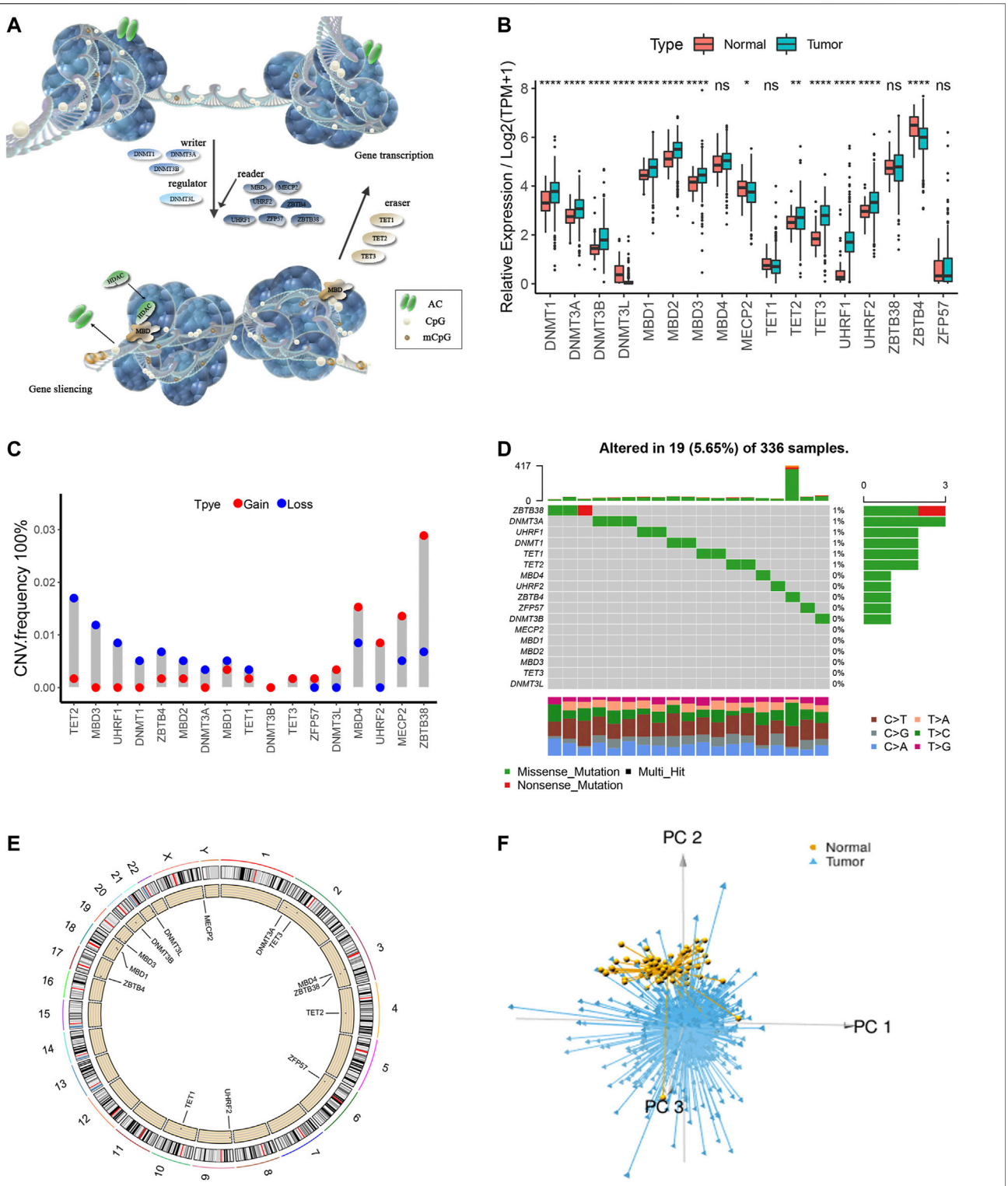


FIGURE 1 | The overall depiction of genetic variations of m^{5}C regulators in ccRCC. **(A)** The dynamic reversible process of m^{5}C DNA methylation mediated by regulators as well as their potential biological functions for ccRCC are summarized. **(B)** Comparison of the expression levels of 17 m^{5}C regulators in 530 ccRCC samples and >12,000 para-cancer tissues. **(C)** Copy number variations (CNVs) of the 17 m^{5}C regulators in ccRCC from TCGA cohort. **(D)** Somatic variant landscape of the 17 m^{5}C regulators in ccRCC from TCGA cohort. **(E)** The location of CNV alterations of m^{5}C regulators on chromosomes. **(F)** Principal component analysis of ccRCC samples from the TCGA cohort based on the expression of the 17 m^{5}C regulators.

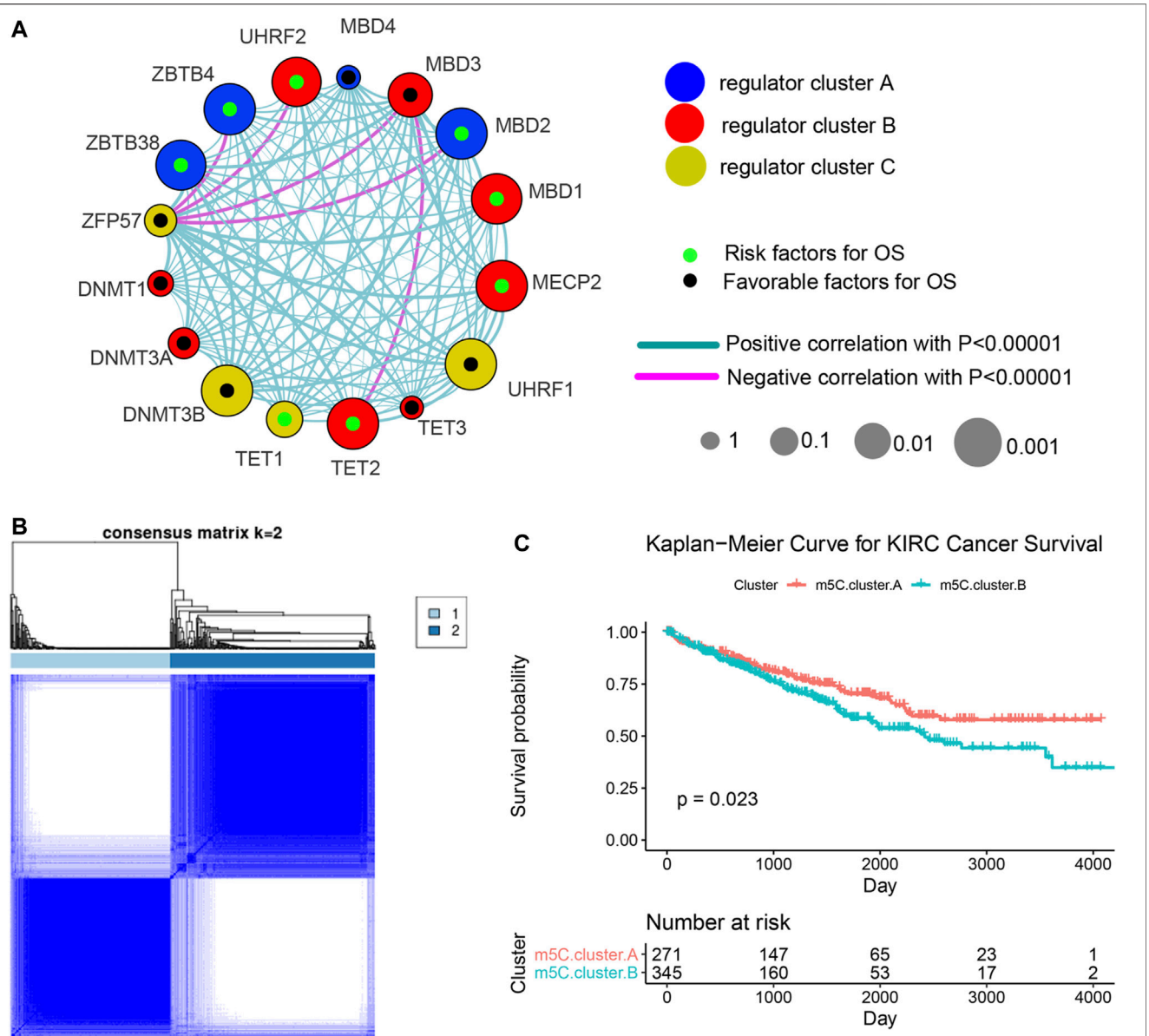


FIGURE 2 | Machine learning algorithms identify m^{5}C modification patterns mediated by the regulators. **(A)** Comprehensive landscape of the interactions of the 17 m^{5}C regulators, interaction network, and the prognostic implications for ccRCC patients. **(B)** Unsupervised clustering based on expression of 17 m^{5}C regulators. **(C)** Overall survival curves of ccRCC patients in the two m^{5}C modification pattern clusters.

Machine Learning Algorithms Identify m^{5}C Modification Patterns Mediated by the Regulators

Three data sets (both proteome and transcriptome) with available survival and clinicopathological information (TCGA, CPTAC, and RECA-EU) were included in one meta-cohort. **Figure 2A** shows the comprehensive landscape of the interaction of the 17 m^{5}C regulators, interaction network, and the prognostic implications for ccRCC patients. The results identified TET2, MBD1, MBD2, MECP2, ZBTB4, ZBTB38, and UHRF2 as protumorigenesis indicators for ccRCC, and MBD3, UHRF1, and DNMT3B were identified as significant favorable factors for ccRCC. We also found that expression of the m^{5}C regulators in the same functional category exhibited remarkable correlations, and a marked association was displayed among writers, regulators, erasers, and readers. For instance, whether ccRCC tumors with a high writer gene expression exhibit a high eraser gene expression normally depended on the different writer and eraser genes. However, we found that tumors with high expression of the m^{5}C reader gene *ZFP57* showed low expression of some reader genes (*ZBTB4*, *UHRF2*, *MBD3*, and *MBD2*) although the high expression of other reader genes was not affected. These

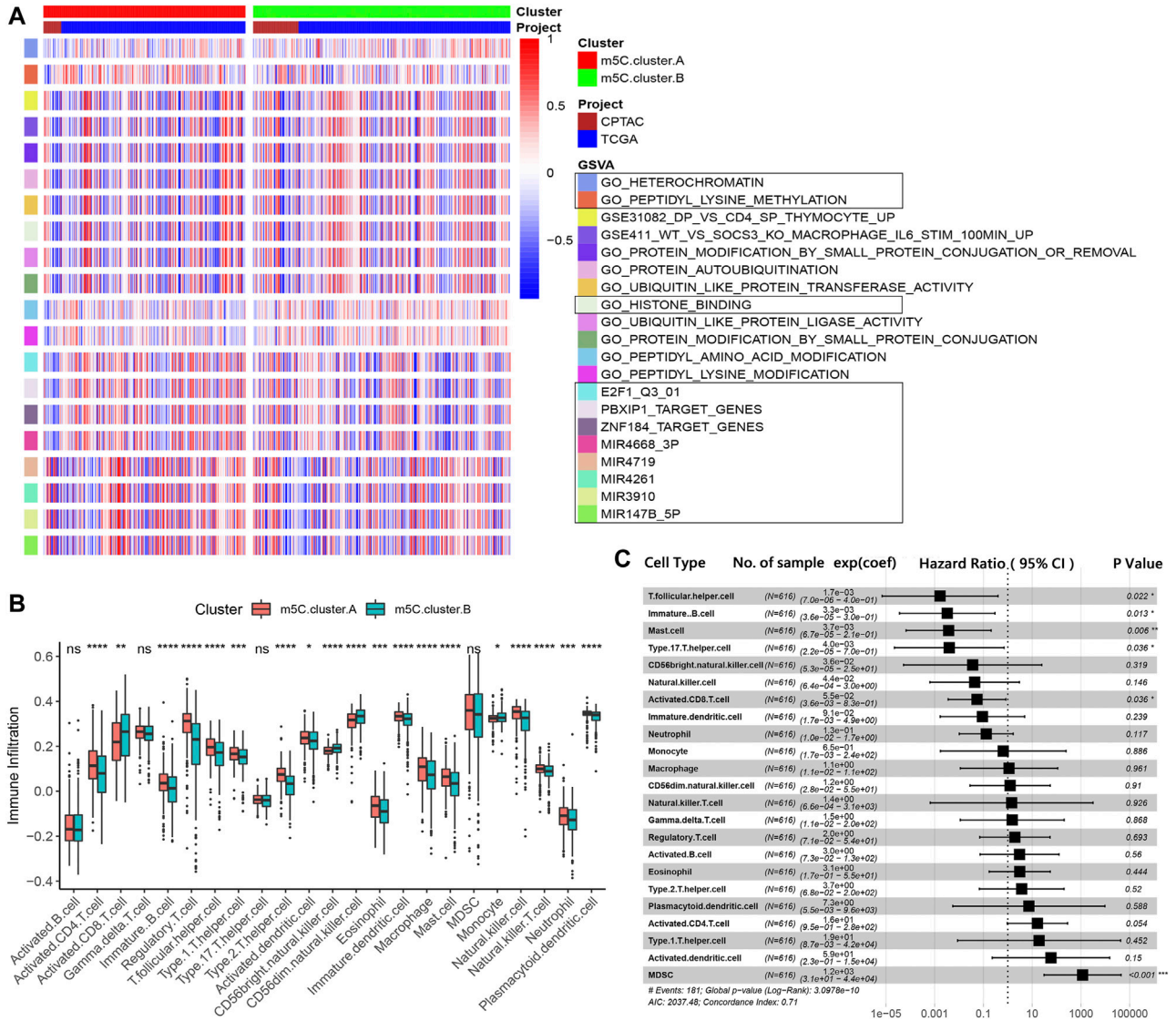


FIGURE 3 | TME characterizations and infiltrating immune cells in the two m^{5}C modification patterns. **(A)** GSVA results of the two m^{5}C modification patterns. **(B)** Estimation of immune cell infiltration in the two m^{5}C modification patterns. **(C)** Univariate regression analysis of various types of immune cells.

results imply that a cross-talk among the genes encoding m^{5}C writers, readers, regulators, and erasers could play essential roles in the malignancy of different m^{5}C modification patterns and tumor immune microenvironment characterization among individual ccRCC samples.

We next used the ConsensusClusterPlus R package to identify ccRCC patients with qualitatively different m^{5}C modification patterns based on the transcriptional expression of 17 m^{5}C regulators. Two distinct modification patterns were classified using unsupervised clustering: m^{5}C cluster A (including 271 cases) and m^{5}C cluster B (354 cases) (**Figure 2B**). Survival analysis of patients in the two clusters revealed that patients with the m^{5}C cluster A modification pattern showed improved survival compared with patients with the m^{5}C cluster B pattern (**Figure 2C**).

Evaluation of TME Characterizations and Immune Contexture Proportion in Distinct m^{5}C Modification Patterns

To investigate the clinical differences and biological processes between the two distinct m^{5}C modification patterns, we constructed a clustering heat map showing differentially expressed m^{5}C regulators and clinical information, including age, sex, stage, and survival status in the two m^{5}C modification patterns (**Supplementary Figure S1**). GSVA enrichment analysis indicated that m^{5}C clusters mainly differ in heterochromatin, peptidyl modification pathways, and microRNA post-transcriptional regulation (**Figure 3A**). The ccRCC samples in m^{5}C cluster A showed prominent upregulation in E2F1, miR-147B, miR-3910, miR-4261, miR-

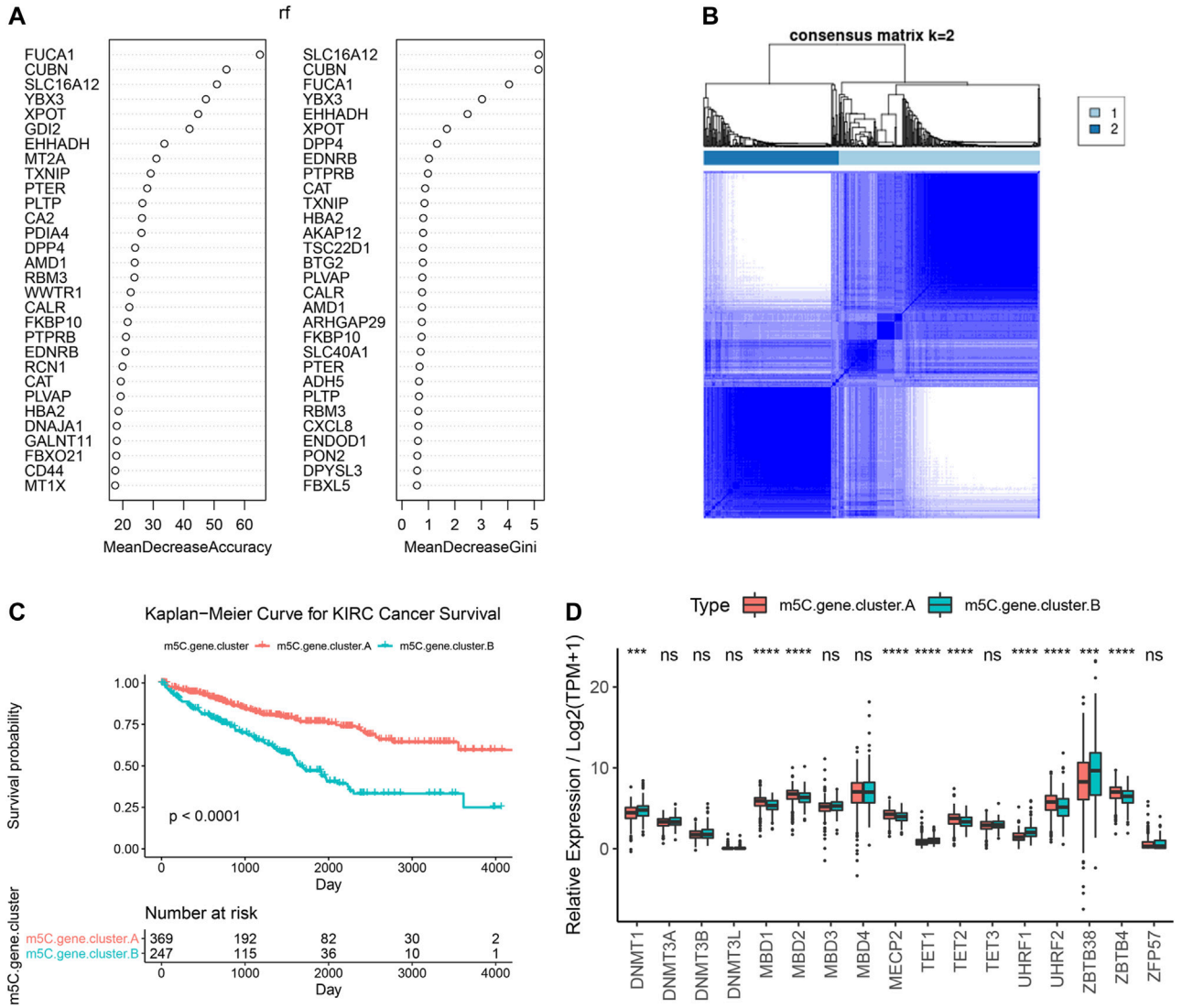


FIGURE 4 | Identification and prognostic implications of m^{5}C genotype signatures. **(A)** Random forest results for selecting the most important DEGs. **(B)** Unsupervised clustering based on the expression of selected DEGs. **(C)** Overall survival curves of ccRCC cases in the indicated subgroups. **(D)** Expression levels of 17 m^{5}C regulators in the indicated subgroups.

3689-3p, miR-4719, PBXIP1, and ZNF184 targeted regulation and downregulation in peptidyl modification processes, such as histone binding, peptide amino acid modification, protein autoubiquitination, ubiquitin-like protein ligase, and transferase activities (**Supplementary Figure S2A**).

We next examined immune cell infiltration to assess differences in the immune context of the TME between m^{5}C modification patterns. m^{5}C cluster A was remarkably rich in innate immune cell infiltration and the active immune response process with a high abundance of activated CD4 T cells, immature B cells, regulatory T cells, Tfh cells, dendritic cells, eosinophils, macrophages, mast cells, natural killer cells, and neutrophils (**Figure 3B**). The results from GSVA analyses demonstrate that the m^{5}C cluster A modification pattern, which predicts favorable clinical outcome, was significantly associated with antitumor immune responses. Therefore, we hypothesized that the peptidyl modification inactivation in m^{5}C cluster A may be involved in the antitumor effects of immune cell infiltration related to this cluster.

We further assessed the prognostic implications of immune cell infiltration in ccRCC (**Figure 3C**). Univariate Cox analysis indicated that T follicular helper cells (p = .022), immature B cells (p = .013), mast cells (p = .006), type 17 T helper cells (p = .036), and activated CD8 T cells (p = .036) could serve as independent prognostic protective factors in ccRCC, and MDSC (p < .001) was a remarkable risk indicator for 616 ccRCC patients from the TCGA and CPTAC cohorts (**Figure 3C**). When clinicopathological factors were analyzed, we found no significant differences in the pathology types and genetic variations between the patients in the two m^{5}C modification

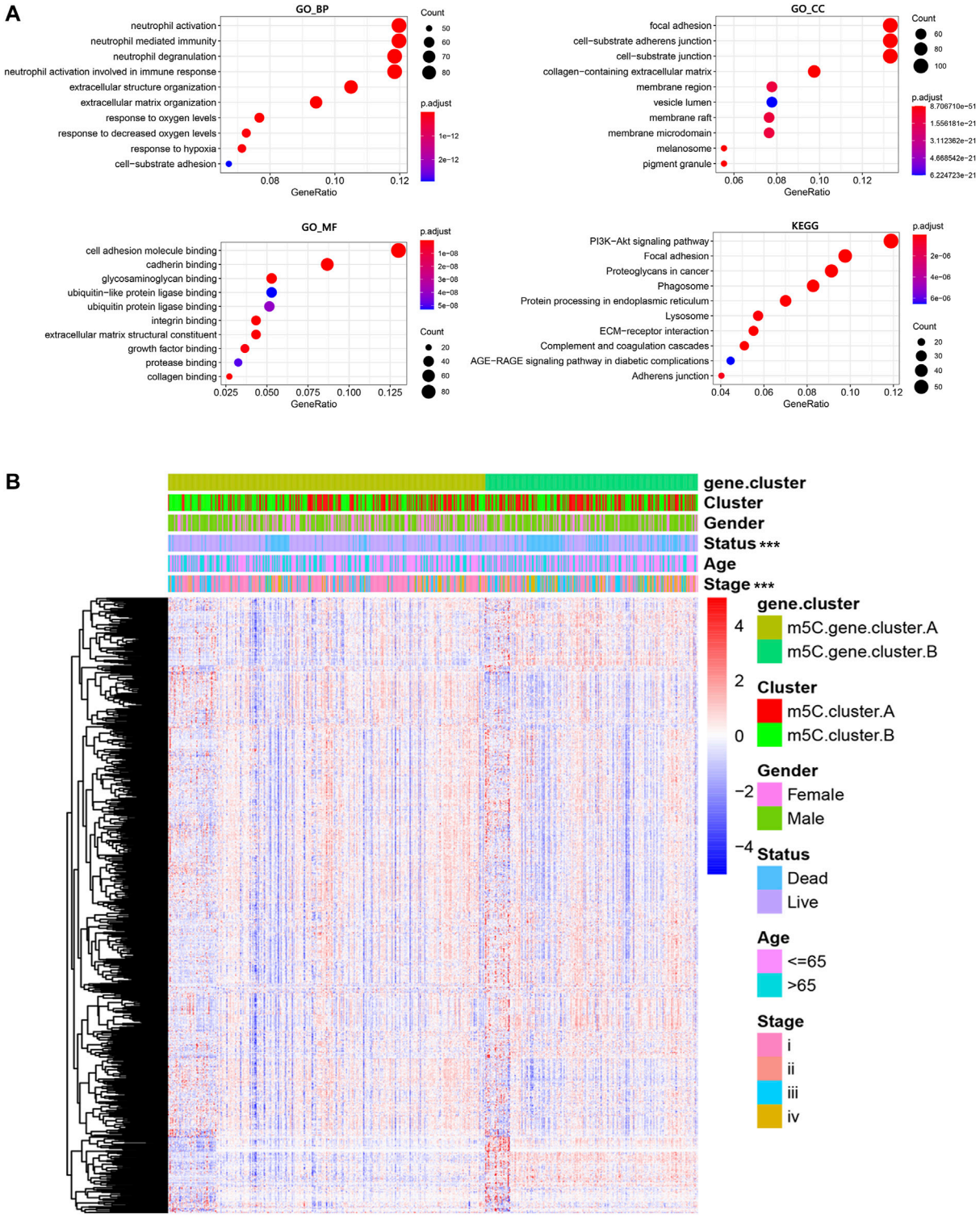


FIGURE 5 | Functional annotations of m^5C genotype signatures. **(A)** Functional enrichment results of the DEGs. **(B)** Integrative heat map including DEG expression, gender, age, clinical stage, and survival status in the two m^5C modification pattern groups.

pattern groups, which suggests that DNA m^5C methylation modification did not influence clinical and pathologic features of tumors (**Supplementary Figure S2B–G**).

Identification and Functional Annotations of m^5C Genotype Signatures

To further explore the biological consequences of the distinct m^5C modification patterns, we then investigated the genetic constitution of individual m^5C clusters patterns and identified 180 m^5C phenotype–related DEGs using the Limma package of R software. Random forest analysis was performed to determine the most important m^5C gene signatures in identifying m^5C modification patterns (**Figure 4A**). To investigate the regulation mechanism of DNA m^5C modifications on ccRCC, we then performed unsupervised clustering analyses based on the obtained 180 m^5C phenotype–related signatures to classify patients into different genotypes. Consistent with the clustering of m^5C modification patterns, the unsupervised clustering algorithm also revealed two distinct m^5C modification genomic subtypes, named as m^5C gene clusters A and B (**Figure 4B**). Kaplan–Meier analysis of ccRCC cases in the combined discovery TCGA and test CPTAC cohorts revealed that patients in the m^5C gene cluster B group (n = 247) showed significantly poor survival compared with cases in m^5C gene cluster A (n = 369) (**Figure 4C**). Prominent differences in the expression of m^5C regulators between the distinct m^5C gene clusters were confirmed using unpaired t test, and the results were in accordance with the results of DNA m^5C methylation modification patterns (**Figure 4D**). These results revealed the presence of distinct m^5C methylation modification patterns in ccRCC and showed that these patterns could distinguish aggressiveness in ccRCC.

Next, the clusterProfiler package was used to perform GO and KEGG functional enrichment analysis for the m^5C DEGs. The biological processes, cellular components, and molecular functions with significant enrichment are summarized in **Figure 5A**. Enriched terms in biological processes were related to m^5C modification, neutrophil activation–related immune response, and response to hypoxia, which provided a basis that m^5C modification may play an important role in the immune regulation of the ccRCC TME (**Figure 5A**). We further found that ccRCC samples in m^5C gene cluster B showed advanced clinical stages and exhibited higher mortality (**Figure 5B**). Older patients were concentrated in the m^5C gene cluster B, and the distinct genotype clusters were characterized by different m^5C signature genes.

Generation and Validation of the m^5C Score Model

The above findings demonstrate that the m^5C methylation modification plays a key regulatory role in reshaping different TME landscapes. Nevertheless, these results were determined on the patient population and might not provide accurate information on survivorship based on m^5C modification patterns in individual ccRCC patients. Considering the individual intratumor heterogeneity of m^5C methylation and using the phenotype-related genes, we establish a scoring system for easy quantification of the m^5C modification patterns for individual ccRCC patients and named this system m^5C score. The alluvial diagram was applied to visualize the alterations of individual patients (**Figure 6A**). The m^5C score clusters prominently classified the patients into two prognostic groups (good and poor) and enabled stratification of patients in both the discovery TCGA and validation real-world FUSCC cohorts. Survival analysis indicated that high m^5C score was significantly correlated with poor overall survival (HR = 0.3 with 95% CI from 0.22 to 0.41, $p < .0001$) in 516 patients with ccRCC from TCGA (**Figure 6B**) and correlated with worse overall survival in 266 patients with ccRCC from FUSCC (**Figure 6C**).

Relation of m^5C Modification with Clinicopathological Features and Tumor Somatic Mutation

We next investigated the relationship of m^5C score with clinical and pathologic characteristics in ccRCC patients from the training, testing, and validation cohorts. Consistent with its prognostic value, the m^5C score significantly increased with advancing clinical stages and aggressive ISUP grade and reached the highest level at stage IV or grade 4 (**Figures 7A,B**). There was no difference in age between the two clusters. The proportion of males in the high m^5C score group was markedly higher than that of females, which is consistent with the result that male patients have a worse prognosis than female patients with ccRCC (**Figures 7C,D**).

To reveal the role of the m^5C score phenotype in the comprehensive molecular landscape of ccRCC, we examined tumor somatic mutation and evaluated DNA variation in the m^5C score clusters. Patients with mutation in *PBRM1*, a gene frequently mutated in ccRCC, showed a prominently lower m^5C score compared with patients with wild-type *PBRM1* (**Figure 7E**). The m^5C score did not show a significant association with tumor mutation burden in patients with ccRCC (**Figure 7F**).

We next evaluated the differences in the DNA variation landscape in the two m^5C score clusters. The top 20 frequently mutated genes in the m^5C score clusters are shown in **Figure 7G,H**. *VHL* (mutation frequency, 40%), *BAP1* (13%), *SETD2* (13%), *TTN* (13%), and *MTOR* (11%) were the five most frequently mutated genes in the $m^5C\ score^{high}$ group (**Figure 7G**), whereas *VHL* (24%), *PBRM1* (19%), *TTN* (15%), *SETD2* (7%), and *MTOR* (6%) were the five most frequently mutated genes in the $m^5C\ score^{low}$ group (**Figure 7H**). Thus, we speculate that the significantly higher mutation frequency of *BAP1* in the high m5C score group may contribute to the poor prognosis for ccRCC patients and the low mutation frequency of *PBRM1* may reduce immunotherapy efficiency for ccRCC patients. Copy number variant features are depicted in **Figure 7I,J**. In addition to the common mutation site located in 5q35.3, copy number variant in m5C $score^{high}$ samples were generally located in 3q25.33, 2q10.53, and 9p12.3 loci.

Characteristics of TME and Immune Cell Distribution in m^5C-Related Phenotypes

To define the role of m^5C-related phenotypes in regulation of the TME, we first investigated cancer-related pathways

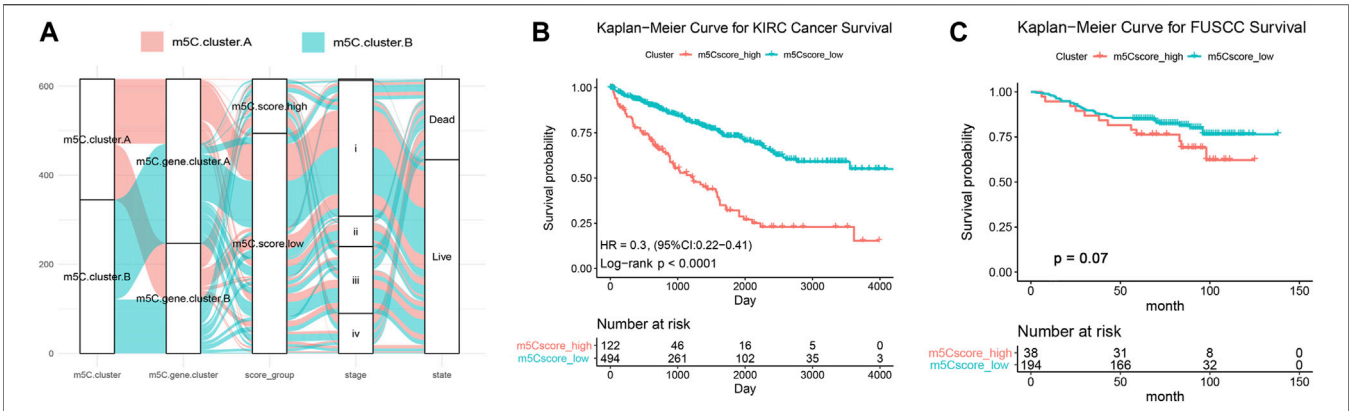


FIGURE 6 | Generation and validation of the m^{5}C score model. **(A)** Sanky diagram of the various clusters. **(B)** Overall survival curve of ccRCC patients from TCGA cohort stratified by m^{5}C score. **(C)** Overall survival curve of 233 ccRCC patients from FUSCC cohort stratified by m^{5}C score.

characterizing m^{5}C gene clusters based on training and testing cohorts. As shown in **Figure 8A**, TGF-β signaling, oxidative phosphorylation, and fatty acid metabolism were significantly downregulated in ccRCC samples in the m^{5}C score^{high} group compared with the m^{5}C score^{low} group, whereas pathways involved in protumorigenesis responses of the TME, such as hypoxia, glycolysis, epithelial-mesenchymal translation, and IL6-JAK/STAT3 signaling, were markedly upregulated in the m^{5}C score^{high} group. We next evaluated the immune cell infiltration in the TME in m^{5}C-related phenotype clusters. The results indicated that CD4$^+$ T cell memory resting, mast cell resting, and monocyte and NK cell infiltration significantly correlated with a high m^{5}C score, whereas plasma cell, M0 macrophage, Treg cell, and neutrophil infiltration were significantly associated with low m^{5}C score in ccRCC patients (**Figure 8B**). To evaluate the regulatory role of m^{5}C score in TME, we explored the expression of chemokine, cytokine, and immune checkpoints in m^{5}C score clusters (**Figure 8C**). We found the expression levels of immune checkpoint factors were significantly different in the m^{5}C score^{high} group, suggesting that the high m^{5}C score cluster may indicate an immune-suppressive microenvironment.

Influence of m^{5}C Modification Patterns on Chemotherapy and Immunotherapy Response

Immunotherapies, including anti-immune checkpoints, are revolutionizing the field of cancer therapy. RCC is resistant to traditional cytotoxic chemotherapy but can be responsive to immunotherapy. Therefore, we investigated whether the m^{5}C modification signature could predict the responses to chemotherapy and ICTs in the combined ccRCC cohorts (n = 860, TCGA, CPTAC, and FUSCC). Evaluation of the ICC50 of cisplatin showed that the low m^{5}C score group was significantly correlated with a higher IC50 value, which indicates that the low m^{5}C score group may be less sensitive to cisplatin (**Figure 9A**). However, no significant differences were observed in predicting IC50 values of gemcitabine between the m^{5}C modification groups (**Figure 9B**). The TIDE algorithm was used to predict intratumoral heterogeneity and responsiveness to immunotherapy. The findings indicate that a higher m^{5}C score was significantly correlated with an elevated TIDE score, suggesting that the high m^{5}C score group may show a reduced response to immunotherapy, such as PD-1 and PD-L1 blockade (**Figure 9C**). The ROC curve showed a relatively stable ability for predicting the immunotherapy response of m^{5}C score with an AUC of 0.676 (**Figure 9D**).

TME Characterization in the m^{5}C Modification Phenotypes

To further test the stability of m^{5}C score model, we applied the m^{5}C score signature established in the real-world FUSCC proteomics cohort and evaluated TME characteristics by IHC staining analysis of 30 consecutive ccRCC tissue sections. IHC staining revealed significantly decreased CD8, PD-L1, and GLUT-1 expression and elevated FoxP3, CXCL13, and FASN expression and Ki-67 staining in tumors from the FUSCC cohort (p < .05) in the m^{5}C score^{high} group (**Figure 10**), suggesting immune-suppressive characteristics of the TME. Furthermore, we found a significantly decreased number of infiltrated CD4$^+$ T cells and CD8$^+$FoxP3$^+$ Treg cells and downregulated PD-L1 expression in the immune-cold m^{5}C score^{high} group using opal multimarker IHC staining (**Figure 10**). In general, the data from multiomics bioinformatics to the real world demonstrate that lower m^{5}C score predicts better responses to immunotherapy for ccRCC patients.

DISCUSSION

Increasing evidence demonstrates that malignant biological behaviors of cancer cells are tightly regulated by the TME and genetic variations (Mehdi and Rabbani, 2021). DNA methylation plays an essential role in modulating the transcriptional regulation of genes and subsequent cell functions, including the infiltration and functional differentiation of immune cells

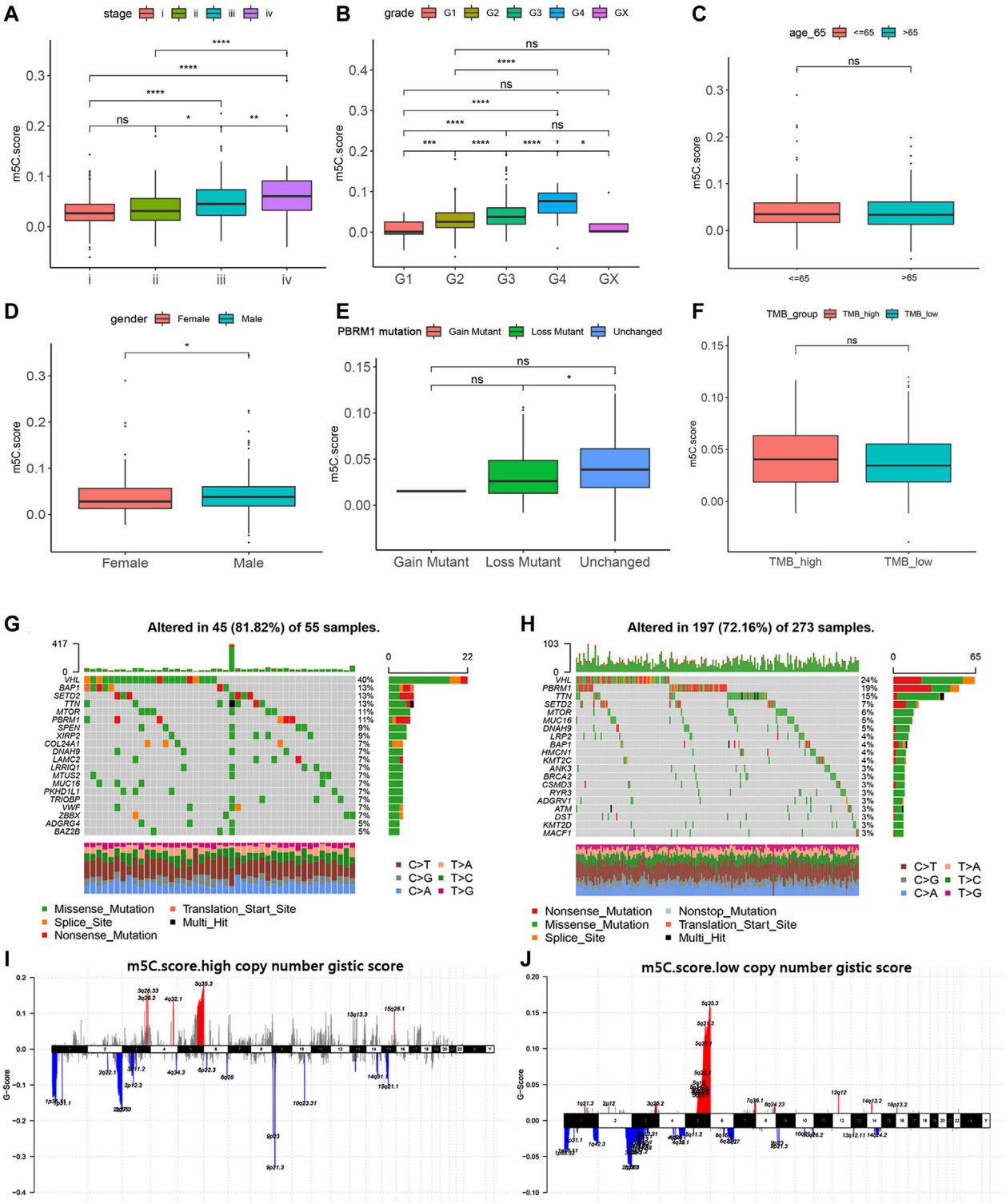


FIGURE 7 | Relation of m^{5}C modification with clinicopathological features and tumor somatic mutation. **(A–F)** Associations of m^{5}C score with stage, grade, age, gender, *PBRM1* mutation status, and TMB in TCGA cohort. **(G–H)** Landscapes of somatic variants of high and low m^{5}C score groups in TCGA cohort. **(I–J)** Copy number variants of high and low m^{5}C score groups in TCGA cohort.

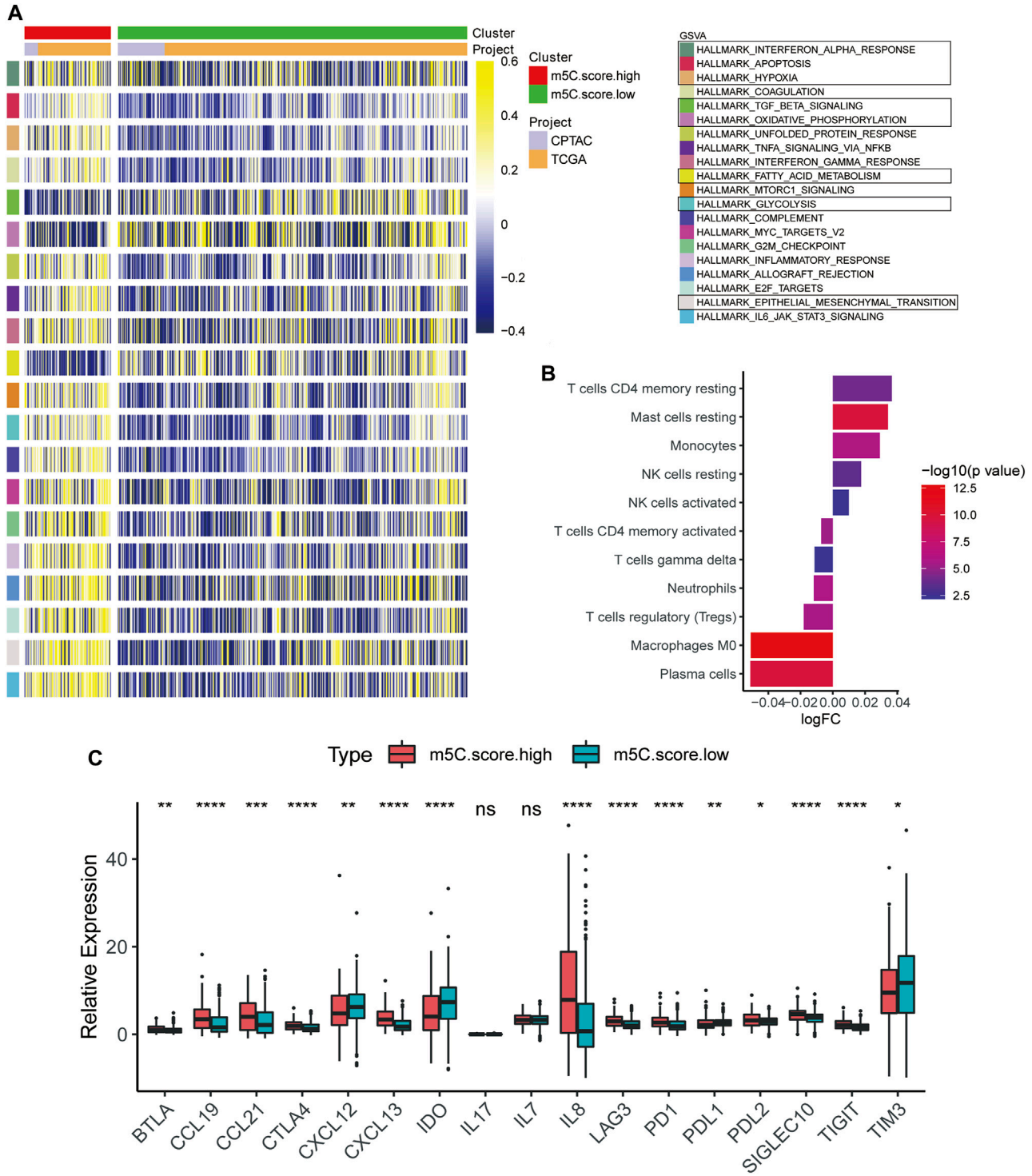


FIGURE 8 | Characteristics of TME and immune cell distribution in m^{5}C-related phenotypes. **(A)** GSVA results of high and low m^{5}C score groups based on expression profiles from CPTAC and TCGA cohorts. **(B)** Estimation of immune cell infiltration in high and low m^{5}C score groups. **(C)** Expression levels of chemokines, cytokines, and immune checkpoints between high and low m^{5}C score groups.

participating in protumor and antitumor immune responses (Saleh et al., 2020; Mehdi and Rabbani, 2021; Smiline Girija, 2021). Previous studies mainly focus on tumor-infiltrated lymphocytes or single signatures, and the influence of DNA m^{5}C epigenetic regulators on the TME was not comprehensively elucidated. Therefore, the overall

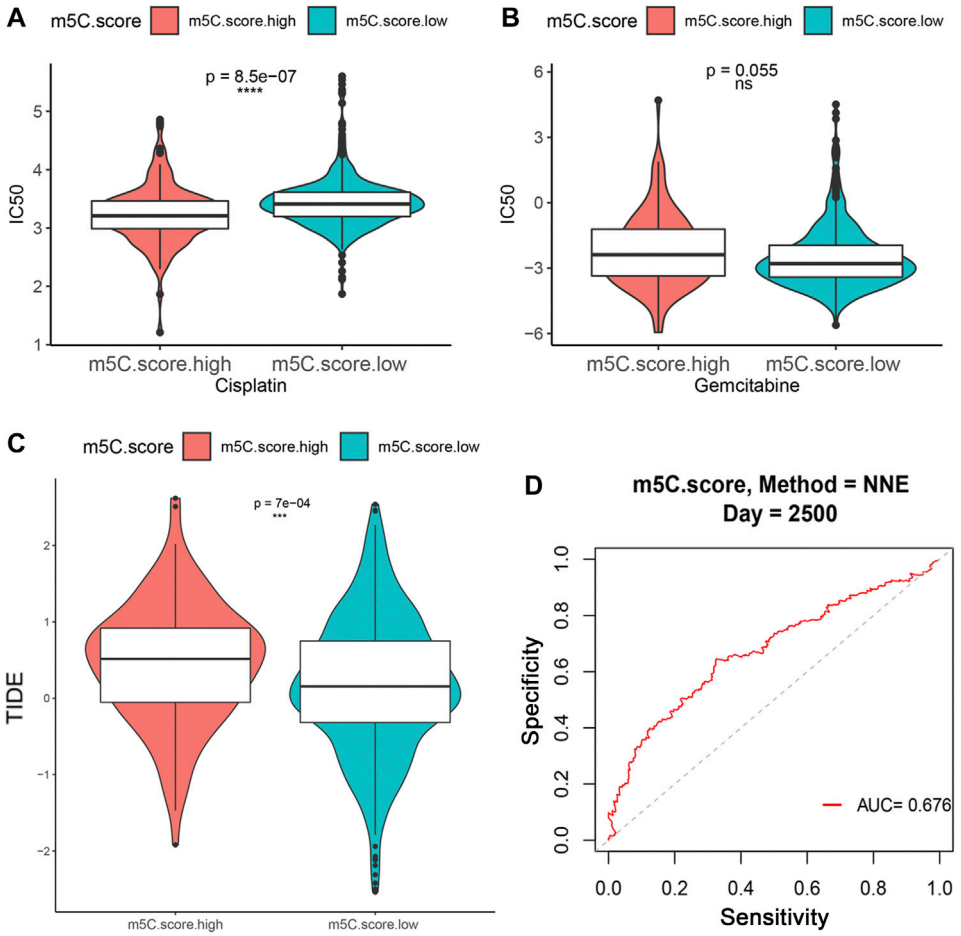


FIGURE 9 | Influence of m^5C modification patterns on chemotherapy and immunotherapy response. **(A–B)** IC50 value for cisplatin and gemcitabine in low and high m^5C score groups. **(C)** TIDE prediction score of low and high m^5C score groups. **(D)** ROC curve for evaluating the ability of m^5C score to predict immunotherapy response.

characteristics and implications of m^5C modification patterns on the tumor immune microenvironment in ccRCC warrant further study.

In the current study, we used transcriptome data of 17 DNA methylation regulators and identified two distinct m^5C methylation modification patterns that are associated with remarkable differences in molecular and clinical characteristics of TME in large-scale ccRCC samples in training, testing, and validation real-world cohorts. The m^5C $score^{high}$ cluster was characterized by poor prognosis and activation of innate immunity and metabolism, corresponding to the immune-desert phenotype. The m^5C $score^{low}$ cluster was characterized by the activation of antitumor immunity, corresponding to the immune-excluded phenotype. IHC analysis revealed that the immune-excluded phenotype showed the presence of abundant immune cell infiltrations retained in the parenchyma in ccRCC samples rather than being located in the stroma (Gajewski et al., 2013). This is consistent with our previous findings that, even in occasional cases of nested tertiary lymphatic structures in the immune-excluded phenotype, tumor-infiltrating lymphocytes rarely appear in the stromal component of ccRCC samples (Xu et al., 2021a). Moreover, the immune-desert phenotype, the m^5C $score^{high}$ cluster, prominently correlates with progressive malignancy, immune tolerance, and lack of T cell–mediated immune responses (Kim and Chen, 2016), guiding effectiveness of immune checkpoint therapy strategies for ccRCC patients.

Research has identified molecular features underlying the initiation and progression of ccRCC. VHL gene inactivation and copy number variation are shown to be involved in promoting the initiation and lethality of ccRCC (D'Avella et al., 2020). The development of sequencing technologies enables determination of the comprehensive DNA mutation landscape and intratumor heterogeneity in the carcinogenesis process (Wettersten et al., 2015; Young et al., 2018; Clark et al., 2019). These findings are extremely important contributions to the categorization and treatment guidance of ccRCC. However, DNA variation, tumor epigenomics, and TME characterizations

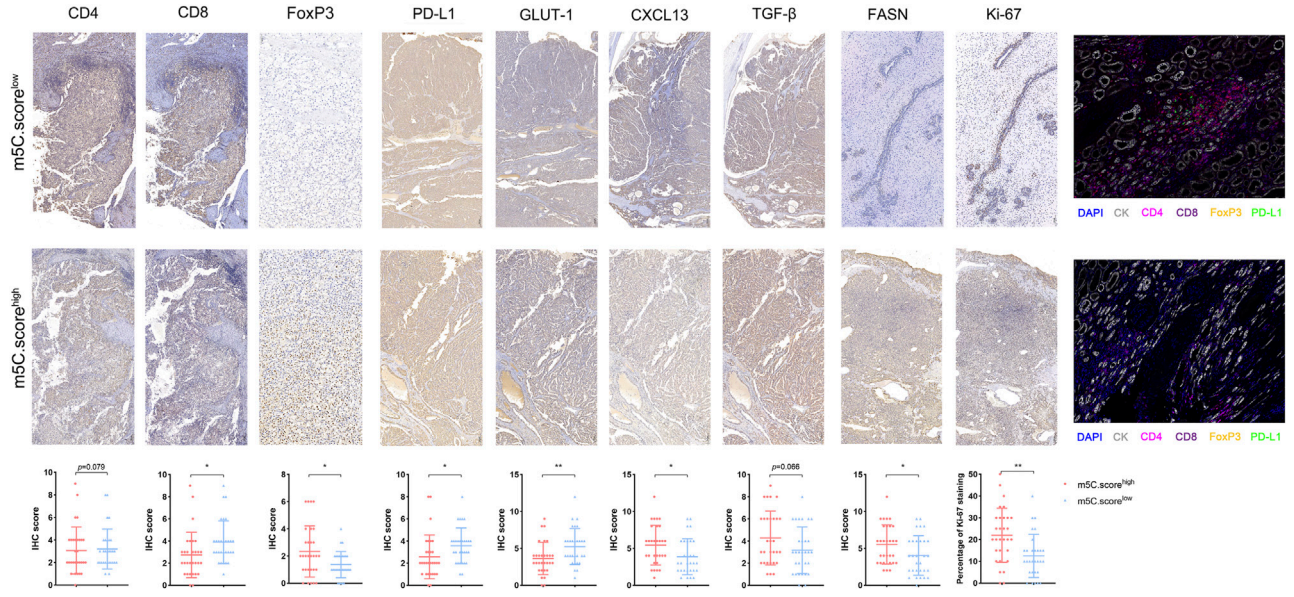


FIGURE 10 | TME characterization in the m^{5}C modification phenotypes. TME characterization assessment between high and low m^{5}C score groups based on IHC staining (CD4, CD8, FoxP3, PD-L1, GLUT-1, CXCL13, TGF-β, FASN, Ki-67) and opal multimarkers IHC staining (DAPI, CK, CD4, CD8, FoxP3, PD-L1).

of ccRCC remain unclear. Here, we find significantly decreased mutation frequency of *VHL* (40% vs. 24%) and *BAP1* (13% vs. 4%) and an elevated mutation frequency of *PBRM1* (11%. vs 19%) in the high m^{5}C score cluster compared with the low m^{5}C modification pattern. Currently, screening for germline mutations in *BAP1* and *PBRM1* is recommended as these genes may serve as promising targets to predict clinical outcomes and ICT treatment responses (Miao et al., 2018; Gallan et al., 2021; Jonasch et al., 2021). Therefore, we speculate that the significantly higher proportion of *BAP1* mutation in the m5C score^{high} cluster contributes to the poor prognosis for ccRCC patients, and the low proportion of *PBRM1* mutations in the immune-desert phenotype may reflect reduced immunotherapy efficiency of ccRCC patients.

DNA methylation has an important impact on tumor initiation and progression because of its critical role in transcriptional regulation (Bates, 2020). An overall decrease in methylated CpG content is typically observed in tumors, and this leads to genome instability and oncogene activation. CpG hypermethylation in the promoter region of specific genes is a hallmark of many tumors (Paz et al., 2003; Bai et al., 2021). DNA methylations have been identified in genes involved in immune modulation, inflammation, cell differentiation, and metabolic and development processes (Serena et al., 2020). Here, we show that m^{5}C methylation modification patterns may function to reshape different metabolism processes and the immune TME landscape, and our results suggest that m^{5}C modification may mediate the therapeutic efficacy of ICTs. The m^{5}C score together with integrated signatures, including tumor mutation load, PD-L1 expression, T cell infiltration, and immune TME based on multiomics large-scale samples data, may represent an effective predictive treatment strategy. In clinical practice, the m^{5}C score can be used to comprehensively assess the m^{5}C methylation modification patterns as well as distinct immune cell infiltration of the TME within individuals, allowing for determination of the genetic landscape and immunophenotypes and effective clinical treatment of ccRCC.

CONCLUSION

In summary, this work reveals the general regulation mechanisms of DNA m^{5}C methylation modification patterns on the tumor immune microenvironment. The m^{5}C modification patterns have marked influences on intratumoral heterogeneity and the complexity of the individual TME. Comprehensive assessment of tumor m^{5}C modification patterns enhances our understanding of TME cell-infiltrating characterizations and helps establish precision immunotherapy strategies for individual ccRCC patients.

MAIN FINDINGS

This work reveals the general regulation mechanisms of DNA m^{5}C methylation modification patterns on the tumor immune microenvironment. The different m^{5}C modification patterns have marked influences on intratumoral heterogeneity and complexity of the individual TME. Comprehensive assessment of tumor m^{5}C modification patterns may enhance our understanding of TME cell-infiltrating characterizations and help establish precision immunotherapy strategies for individual ccRCC patients.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**. The raw proteome data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

Conceptualization: WX, WZ, XT, and WL. Data curation and formal analysis: WX, WZ, XT, AA, YW, WL, and JS. Funding acquisition: SW, WX, YQ, HZ, and DY. Investigation and methodology: WX, WZ, AA, WZ, YW, and WL. Resources and software: WL, WS, YQ, HZ, YW, and DY. Supervision: YQ, SW, HZ, and DY. Validation and visualization: WX, WL, WZ, XT, and AA. Original draft: WX, WS, YW, and WL. Editing: WS, YQ, HZ, and DY.

FUNDING

This work is supported by Grants from the National Key Research and Development Project (No.2019YFC1316000), "Fuqing Scholar" Student Scientific Research Program of Shanghai Medical College, Fudan University (No. FQXZ202112B), the Natural Science Foundation of Shanghai (No.20ZR1413100) and Shanghai Municipal Health Bureau (No.2020CXJQ03).

ACKNOWLEDGMENTS

We are grateful to all patients for their dedicated participation in the current study. We expressed our sincere gratitude to Ms. ZOO for editing figures. We thank Gabrielle White Wolf, PhD, from Liwen Bianji (Edanz) (www.liwenbianji.cn/) for editing the English text of a draft of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.772436/full#supplementary-material

REFERENCES

Bai, L., Yang, G., Qin, Z., Lyu, J., Wang, Y., Feng, J., et al. (2021). Proteome-Wide Profiling of Readers for DNA Modification. *Adv. Sci.* 8, 2101426. doi:10.1002/advs.202101426

Bates, S. E. (2020). Epigenetic Therapies for Cancer. *N. Engl. J. Med.* 383, 650–663. doi:10.1056/NEJMra1805035

Beroukhim, R., Getz, G., Nghiemphu, L., Barretina, J., Hsueh, T., Linhart, D., et al. (2007). Assessing the Significance of Chromosomal Aberrations in Cancer: Methodology and Application to Glioma. *Proc. Natl. Acad. Sci.* 104, 20007–20012. doi:10.1073/pnas.0710052104

Braun, D. A., Hou, Y., Bakouny, Z., Ficial, M., Sant' Angelo, M., Forman, J., et al. (2020). Interplay of Somatic Alterations and Immune Infiltration Modulates Response to PD-1 Blockade in Advanced clear Cell Renal Cell Carcinoma. *Nat. Med.* 26, 909–918. doi:10.1038/s41591-020-0839-y

Capitanio, U., and Montorsi, F. (2016). Renal Cancer. *The Lancet* 387, 894–906. doi:10.1016/S0140-6736(15)00046-X

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017). Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cel Rep.* 18, 248–262. doi:10.1016/j.celrep.2016.12.019

Chen, X., Li, A., Sun, B.-F., Yang, Y., Han, Y.-N., Yuan, X., et al. (2019). 5-methylcytosine Promotes Pathogenesis of Bladder Cancer through Stabilizing mRNAs. *Nat. Cel Biol* 21, 978–990. doi:10.1038/s41556-019-0361-y

Choi, W. L., Mok, Y. G., and Huh, J. H. (2021). Application of 5-Methylcytosine DNA Glycosylase to the Quantitative Analysis of DNA Methylation. *Ijms* 22, 1072. doi:10.3390/ijms22031072

Clark, D. J., Dhanasekaran, S. M., Petralia, F., Pan, J., Song, X., Hu, Y., et al. (2019). Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. *Cell* 179, 964–e31. doi:10.1016/j.cell.2019.10.007

D'Avella, C., Abbosh, P., Pal, S. K., and Geynisman, D. M. (2020). Mutations in Renal Cell Carcinoma. *Urol. Oncol. Semin. Original Invest.* 38, 763–773. doi:10.1016/j.urolonc.2018.10.027

Gajewski, T. F., Woo, S.-R., Zha, Y., Spaapen, R., Zheng, Y., Corrales, L., et al. (2013). Cancer Immunotherapy Strategies Based on Overcoming Barriers within the Tumor Microenvironment. *Curr. Opin. Immunol.* 25, 268–276. doi:10.1016/j.coi.2013.02.009

Gallan, A. J., Parilla, M., Segal, J., Ritterhouse, L., and Antic, T. (2021). BAP1-Mutated Clear Cell Renal Cell Carcinoma. *Am. J. Clin. Pathol.* 155, 718–728. doi:10.1093/ajcp/aqaa176

Greenberg, M. V. C., and Bourc'his, D. (2019). The Diverse Roles of DNA Methylation in Mammalian Development and Disease. *Nat. Rev. Mol. Cel Biol* 20, 590–607. doi:10.1038/s41580-019-0159-6

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC bioinformatics* 14, 7. doi:10.1186/1471-2105-14-7

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T Cell Dysfunction and Exclusion Predict Cancer Immunotherapy Response. *Nat. Med.* 24, 1550–1558. doi:10.1038/s41591-018-0136-1

Jonasch, E., Walker, C. L., and Rathmell, W. K. (2021). Clear Cell Renal Cell Carcinoma Ontogeny and Mechanisms of Lethality. *Nat. Rev. Nephrol.* 17, 245–261. doi:10.1038/s41581-020-00359-2

Kim, J. M., and Chen, D. S. (2016). Immune Escape to PD-L1/pd-1 Blockade: Seven Steps to success (Or Failure). *Ann. Oncol.* 27, 1492–1504. doi:10.1093/annonc/mdw217

Linehan, W. M., and Ricketts, C. J. (2019). The Cancer Genome Atlas of Renal Cell Carcinoma: Findings and Clinical Implications. *Nat. Rev. Urol.* 16, 539–552. doi:10.1038/s41585-019-0211-5

Martisova, A., Holcakova, J., Izadi, N., Sebuyoya, R., Hrstka, R., and Bartosik, M. (2021). DNA Methylation in Solid Tumors: Functions and Methods of Detection. *Ijms* 22, 4247. doi:10.3390/ijms22084247

Mehdi, A., and Rabbani, S. A. (2021). Role of Methylation in Pro- and Anti-cancer Immunity. *Cancers* 13, 545. doi:10.3390/cancers13030545

Mermel, C. H., Schumacher, S. E., Hill, B., Meyerson, M. L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 Facilitates Sensitive and Confident Localization of the Targets of Focal Somatic Copy-Number Alteration in Human Cancers. *Genome Biol.* 12, R41. doi:10.1186/gb-2011-12-4-r41

Miao, D., Margolis, C. A., Gao, W., Voss, M. H., Li, W., Martini, D. J., et al. (2018). Genomic Correlates of Response to Immune Checkpoint Therapies in clear Cell Renal Cell Carcinoma. *Science* 359, 801–806. doi:10.1126/science.aan5951

Motzer, R. J., Jonasch, E., Boyle, S., Carlo, M. I., Manley, B., Agarwal, N., et al. (2020). NCCN Guidelines Insights: Kidney Cancer, Version 1.2021. *J. Natl. Compr. Canc Netw.* 18, 1160–1170. doi:10.6004/jnccn.2020.0043

Palei, S., Buchmuller, B., Wolffgramm, J., Muñoz-Lopez, Á., Jung, S., Czodrowski, P., et al. (2020). Light-Activatable TET-Dioxygenases Reveal Dynamics of 5-Methylcytosine Oxidation and Transcriptome Reorganization. *J. Am. Chem. Soc.* 142, 7289–7294. doi:10.1021/jacs.0c01193

Paz, M. F., Fraga, M. F., Avila, S., Guo, M., Pollan, M., Herman, J. G., et al. (2003). A Systematic Profile of DNA Methylation in Human Cancer Cell Lines. *Cancer Res.* 63, 1114–1121.

Piperi, C., Vlastos, F., Farmaki, E., Martinet, N., and Papavassiliou, A. G. (2008). Epigenetic Effects of Lung Cancer Predisposing Factors Impact on Clinical Diagnosis and Prognosis. *J. Cell. Mol. Med.* 12, 1495–1501. doi:10.1111/j.1582-4934.2008.00309.x

Qian, S., Sun, S., Zhang, L., Tian, S., Xu, K., Zhang, G., et al. (2020). Integrative Analysis of DNA Methylation Identified 12 Signature Genes Specific to Metastatic ccRCC. *Front. Oncol.* 10, 556018. doi:10.3389/fonc.2020.556018

Rausch, C., Hastert, F. D., and Cardoso, M. C. (2020). DNA Modification Readers and Writers and Their Interplay. *J. Mol. Biol.* 432, 1731–1746. doi:10.1016/j.jmb.2019.12.018

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

Saleh, R., Toor, S. M., Sasidharan Nair, V., and Elkord, E. (2020). Role of Epigenetic Modifications in Inhibitory Immune Checkpoints in Cancer Development and Progression. *Front. Immunol.* 11, 1469. doi:10.3389/fimmu.2020.01469

Serena, C., Millan, M., Ejarque, M., Saera-Vila, A., Maymó-Masip, E., Núñez-Roa, C., et al. (2020). Adipose Stem Cells from Patients with Crohn's Disease Show a Distinctive DNA Methylation Pattern. *Clin. Epigenet* 12, 53. doi:10.1186/s13148-020-00843-3

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer Statistics. *CA A. Cancer J. Clin.* 70, 7–30. doi:10.3322/caac.21590

Smiline Girija, A. S. (2021). Protean Role of Epigenetic Mechanisms and Their Impact in Regulating the Tregs in TME. *Cancer Gene Ther.* online ahead of print. doi:10.1038/s41417-021-00371-z

Wang, L., Liu, Y., Beier, U. H., Han, R., Bhatti, T. R., Akimova, T., et al. (2013). Foxp3+ T-Regulatory Cells Require DNA Methyltransferase 1 Expression to Prevent Development of Lethal Autoimmunity. *Blood* 121, 3631–3639. doi:10.1182/blood-2012-08-451765

Wettersten, H. I., Hakimi, A. A., Morin, D., Bianchi, C., Johnstone, M. E., Donohoe, D. R., et al. (2015). Grade-Dependent Metabolic Reprogramming in Kidney Cancer Revealed by Combined Proteomics and Metabolomics Analysis. *Cancer Res.* 75, 2541–2552. doi:10.1158/0008-5472.CAN-14-1703

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: a Class Discovery Tool with Confidence Assessments and Item Tracking. *Bioinformatics* 26, 1572–1573. doi:10.1093/bioinformatics/btq170

Xu, W.-H., Xu, Y., Wang, J., Wan, F.-N., Wang, H.-K., Cao, D.-L., et al. (2019). Prognostic Value and Immune Infiltration of Novel Signatures in clear Cell Renal Cell Carcinoma Microenvironment. *Aging* 11, 6999–7020. doi:10.18632/aging.102233

Xu, W., Anwaier, A., Ma, C., Liu, W., Tian, X., Palihati, M., et al. (2021). Multi-omics Reveals Novel Prognostic Implication of SRC Protein Expression in Bladder Cancer and its Correlation with Immunotherapy Response. *Ann. Med.* 53, 596–610. doi:10.1080/07853890.2021.1908588

Xu, W., Tian, X., Liu, W., Anwaier, A., Su, J., Zhu, W., et al. (2021). m6A Regulator-Mediated Methylation Modification Model Predicts Prognosis, Tumor Microenvironment Characterizations and Response to Immunotherapies of Clear Cell Renal Cell CarcinomaA Regulator-Mediated Methylation Modification Model Predicts Prognosis, Tumor Microenvironment Characterizations and Response to Immunotherapies of Clear Cell Renal Cell Carcinoma. *Front. Oncol.* 11, 709579. doi:10.3389/fonc.2021.709579

Young, M. D., Mitchell, T. J., Vieira Braga, F. A., Tran, M. G. B., Stewart, B. J., Ferdinand, J. R., et al. (2018). Single-cell Transcriptomes from Human Kidneys Reveal the Cellular Identity of Renal Tumors. *Science* 361, 594–599. doi:10.1126/science.aat1699

Zagorac, S., Alcala, S., Fernandez Bayon, G., Bou Kheir, T., Schoenhals, M., González-Neira, A., et al. (2016). DNMT1 Inhibition Reprograms Pancreatic Cancer Stem Cells via Upregulation of the miR-17-92 Cluster. *Cancer Res.* 76, 4546–4558. doi:10.1158/0008-5472.CAN-15-3268

Zhang, Z.-M., Lu, R., Wang, P., Yu, Y., Chen, D., Gao, L., et al. (2018). Structural Basis for DNMT3A-Mediated De Novo DNA Methylation. *Nature* 554, 387–391. doi:10.1038/nature25477

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Cell and Developmental Biology

ORIGINAL RESEARCH
published: 06 January 2022
doi: 10.3389/fcell.2021.781267

5-Hydroxymethylcytosine Signatures in Circulating Cell-Free DNA as Early Warning Biomarkers for COVID-19 Progression and Myocardial Injury

Hang-yu Chen[1,7†], Xiao-xiao Li[1†], Chao Li[1†], Hai-chuan Zhu[3†], Hong-yan Hou[4], Bo Zhang[4], Li-ming Cheng[4], Hui Hu[5], Zhong-xin Lu[5], Jia-xing Liu[3], Ze-ruo Yang[6], Lei Zhang[6], Nuo Xu[6], Long Chen[7], Chuan He[8], Chao-ran Dong[2], Qing-gang Ge[1]* and Jian Lin[1,7]**

[1]Department of Pharmacy and Department of Intensive Care Unit, Peking University Third Hospital, Beijing, China, [2]Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [3]Institute of Biology and Medicine, College of Life and Health Sciences, Wuhan University of Science and Technology, Hubei, China, [4]Department of Laboratory Medicine, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [5]Department of Medical Laboratory, the Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [6]Yang Sheng Tang Natural Medicine Research Institute, Hangzhou, China, [7]Synthetic and Functional Biomolecules Center, Beijing National Laboratory for Molecular Sciences, Key Laboratory of Bioorganic Chemistry and Molecular Engineering of Ministry of Education, College of Chemistry and Molecular Engineering, Innovation Center for Genomics, Peking University, Beijing, China, [8]Department of Chemistry, Department of Biochemistry and Molecular Biology, Howard Hughes Medical Institute, The University of Chicago, Chicago, IL, United States*

OPEN ACCESS

Edited by:
Li Tan,
Fudan University, China

Reviewed by:
Ruitu Lyu,
University of Chicago, United States
Chunxiao Song,
University of Oxford, United Kingdom
Jiahui Lee,
Fudan University, China

***Correspondence:**
Chao-ran Dong
dongcr@imm.ac.cn
Qing-gang Ge
qingganggelin@126.com
Jian Lin
linjian@pku.edu.cn

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Molecular and Cellular Pathology, a section of the journal Frontiers in Cell and Developmental Biology

Received: *22 September 2021*
Accepted: *16 November 2021*
Published: *06 January 2022*

Citation:
Chen H-y, Li X-x, Li C, Zhu H-c, Hou H-y, Zhang B, Cheng L-m, Hu H, Lu Z-x, Liu J-x, Yang Z-r, Zhang L, Xu N, Chen L, He C, Dong C-r, Ge Q-g and Lin J (2022) 5-Hydroxymethylcytosine Signatures in Circulating Cell-Free DNA as Early Warning Biomarkers for COVID-19 Progression and Myocardial Injury. Front. Cell Dev. Biol. 9:781267. doi: 10.3389/fcell.2021.781267

Background: The symptoms of coronavirus disease 2019 (COVID-19) range from moderate to critical conditions, leading to death in some patients, and the early warning indicators of the COVID-19 progression and the occurrence of its serious complications such as myocardial injury are limited.

Methods: We carried out a multi-center, prospective cohort study in three hospitals in Wuhan. Genome-wide 5-hydroxymethylcytosine (5hmC) profiles in plasma cell-free DNA (cfDNA) was used to identify risk factors for COVID-19 pneumonia and develop a machine learning model using samples from 53 healthy volunteers, 66 patients with moderate COVID-19, 99 patients with severe COVID-19, and 38 patients with critical COVID-19.

Results: Our warning model demonstrated that an area under the curve (AUC) for 5hmC warning moderate patients developed into severe status was 0.81 (95% CI 0.77–0.85) and for severe patients developed into critical status was 0.92 (95% CI 0.89–0.96). We further built a warning model on patients with and without myocardial injury with the AUC of 0.89 (95% CI 0.84–0.95).

Conclusion: This is the first study showing the utility of 5hmC as an accurate early warning marker for disease progression and myocardial injury in patients with COVID-19. Our

Abbreviations: 5hmC, 5-hydroxymethylcytosine; ARDS, acute respiratory distress syndrome; AUC, area under the curve; CD14, CD14 molecule; COPD, chronic obstructive pulmonary disease; COVID-19, coronavirus disease 2019; CRP C-reactive protein; CVA, cerebrovascular accident; DhMRs, differentially 5hMc-enriched regions; GS, gastrointestinal injury; hMRs, 5hMc-enriched Regions; MAP4K4, mitogen-activated protein kinase kinase kinase kinase 4; MI, myocardial injury; NLR, neutrophil-to-lymphocyte ratio; PDE4D, Phosphodiesterase 4D; RFECV, recursive feature elimination algorithm; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SIC, sepsis-induced coagulopathy; TET, ten-eleven translocation; WHO, World Health Organization.

results show that phosphodiesterase 4D and ten-eleven translocation 2 may be important markers in the progression of COVID-19 disease.

Keywords: COVID-19, 5hmC, myocardial injury, machine learning, PDE4D

INTRODUCTION

The pandemic of coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection, has become a global public health crisis (Pollard et al., 2020). According to the World Health Organization (WHO) latest numbers on July 21, 2021, over 192.8 million confirmed cases of COVID-19 and more than 4.0 million deaths worldwide. The clinical spectrum of COVID-19 pneumonia ranges from asymptomatic infection to critically ill cases (McArthur et al., 2020). Critical patients with a higher mortality rate suffered from organ failure, including cerebrovascular accident (CVA), myocardial injury (MI), and thrombotic events (Wang et al., 2020). MI has been the most reported cardiovascular complication with a significant risk of in-hospital mortality rate (51.2%) compared with those without MI (4.5%) (Bonow et al., 2020). These findings suggest that early identification of patients with COVID-19 at risk of critical illness could improve their outcomes. Recently, several studies demonstrated that the higher levels of inflammatory markers such as C-reactive protein (CRP) (Wang, 2020), ferritin (Lin et al., 2020), D-dimer (Zhang et al., 2020), high neutrophil-to-lymphocyte ratio (NLR) (Kong et al., 2020), and blood proteomic and metabolomic biomarkers (Wu D. et al., 2020; Shen et al., 2020) could be used to distinguish between moderate and severe cases. Unfortunately, so far there are no reliable indicators available to warn the COVID-19 progression and the occurrence of serious complications such as MI.

5-Hydroxymethylcytosine (5hmC) is an abundant epigenetic marker associated with gene expression and involves a wide range of biological processes ranging from development to pathogenesis (Han et al., 2016). It is derived from 5mC by ten-eleven translocation (TET) protein family and displays a tissue-specific mass distribution (Tan and Shi, 2012). Our laboratory and others have demonstrated that the 5hmC signatures from cell-free DNA (cfDNA) could serve as epigenetic biomarkers for several human diseases such as cancer, neurodegenerative disorders, and coronary heart disease (Szulwach et al., 2011; Li et al., 2017; Dong et al., 2020). These characteristics indicate that 5hmC may have potential value in COVID-19 warning and discovery of target therapy.

In this study, we performed 5hmC-Seal, a sensitive chemical labeling-based sequencing technology (Song et al., 2011; Dong et al., 2020) that allows rapid and reliable sequencing of whole-genome 5hmC in cfDNA from plasma of 256 patients with (n = 203) and without (n = 53) COVID-19 diagnosis. We found 5hmC characteristics detected in cfDNA could be used as early warning markers for the disease progression and MI of COVID-19.

MATERIALS AND METHODS

Data and Sample Source

We carried out a multi-center, prospective cohort study in three hospitals in Wuhan. From March to April 2020, we consecutively enrolled 203 patients aged at least 18 years and diagnosed with COVID-19 within 48 h after their hospitalization. Patients were classified into three groups according to the disease severity defined by the National Health Commission of the People's Republic of China (**Table 1**) (National Health Commission, 2020). Five-milliliter discarded plasma samples were collected from each patient as they were entering into the study cohort, and the results of clinical tests nearby were recorded. The plasma cfDNA was extracted using the Quick-cfDNA Serum and Plasma Kit (ZYMO) and then stored at −80°C. Each patient was prospectively followed up until hospital discharge or death. Complications during the first 28-day follow-up were evaluated and recorded: 1) MI, based on symptoms (if described), electrocardiogram/echocardiography (if any), and troponin I (TnI) (≥0.4 ng/ml) (Dong et al., 2020); 2) gastrointestinal injury (GS), defined as occurrence of gastrointestinal haemorrhage, gastroparesis, and severe/acute pancreatitis; 3) sepsis-induced coagulopathy (SIC), defined as blood platelet $<150*10^9$/L. The occurrence of all-cause death during the whole follow-up was confirmed by home page and final discussion in medical records. The Ethical Review Board of the Peking University Third Hospital approved the study protocol in March 2020 (IRB00006761-M2020083). On the basis of the consideration of discarded samples used, together with a complete set of information security system established, the informed consent of participants was exempted.

Study Design

We performed a prospective cohort study using 5hmC markers to distinguish patients with COVID-19 from healthy people and warn the disease progression and MI. The 256 samples were divided into four groups: healthy people [n = 53, aged 35 (IQR 29–40) years], moderate patients (n = 66), severe patients (n = 99), and critical patients (n = 38). 5hmC libraries for all samples were constructed with high-efficiency 5hmC-Seal technology, as previously described (Song et al., 2011). All 5hmC libraries were sequenced using Illumina Next500. Meanwhile, in data processing, we split patients with COVID-19 into a training cohort and a validation cohort. The objective of the first part of the study was to screen candidate genes with differential 5hmC modifications in these four groups from the training cohort. The objective of the second part of the study was to warn disease progression and MI using the model developed in the first part, in the validation cohort (**Figure 1**).

TABLE 1 | Statistical characteristics of baseline indicators in patients with COVID-19.

	Total (*n* = 203)	Moderate (*n* = 66)	Severe (*n* = 99)	Critical (*n* = 38)	*p*-value
Age, years	65 (54–73)	58 (37–67)	67 (57–77)	67 (56–73)	*** — §
≥65 years	74 (68–81)	72 (67–83)	77 (69–81)	72 (68–74)	—
Gender, female/male	98/105	40/26	43/56	15/23	—
Obesity (BMI ≥30)	4 of (22)	0 of 0	3 of 21	1 of 1	
Hypertension	84	16	47	21	
Coronary heart disease	18	2	6	10	
Heart failure	12	1	4	7	
Chronic liver disease	13	0	4	9	
Immunodeficiency	9	1	2	6	
Stroke history	12	2	3	7	
Diabetes	47	7	16	24	
Asthma, moderate severe	3	0	1	2	
COPD	14	0	6	8	
CKD	15	1	2	12	
Cancer	8	0	1	7	
Smoke history, naïve/ex-smoker/smoker	119/14/18	18/2/3	75/9/6	26/3/9	

[a]n *(§), median (p25-p75), mean ± SD.*
[b]p*-value: Moderate-Severe: *, Severe-Critical: #, Moderate-Critical: §.*
[c]*p < 0.05, **p < 0.01, and ***p < 0.001.*
COPD, chronic obstructive pulmonary disease; CKD, chronic kidney disease.

5hmC Library Construction, Sequencing, and Mapping

Briefly, because of the highly sensitive nature of the chemical labeling method, the input cfDNA can be as low as 1–10 ng. Paired-end 39–base pair (bp) high-throughput sequencing was performed on the NextSeq 500 platform. FASTQC (version 0.11.5) was used to assess the sequence quality. Raw reads were aligned to the human genome (version hg19) with bowtie2 (version 2.2.9) (Langmead and Salzberg, 2012) and further filtered with SAMtools (version 1.3.1) (Li et al., 2009) to retain unique non-duplicate matches to the genome. Pair-end reads were extended and converted into bedGraph format normalized to the total number of aligned reads using BEDTools (version 2.19.1) (Quinlan, 2014) and then converted to bigwig format using bedGraphToBigWig from the UCSC Genome Browser for visualization in the Integrated Genomics Viewer. Potential hMRs were identified using MACS (version 1.4.2), and the parameters used were macs 14-p 1e-3-f BAM-g hs (Consortium et al., 2007). Peak calls were merged using BEDTools merge, and only those peak regions that appeared in more than 10 samples and less than 1,000 bp were retained. Blacklisted genomic regions that tend to show artifact signals, according to ENCODE, were also filtered.

Feature Selection, Model Training, and Validation

Patients with COVID-19 were randomly divided into training and validation cohorts with a 3:2 ratio; using train_test_split in scikit-learn (version 0.22.1) package in *Python* (version 3.6.10), the logistic regression CV (LR) model was chosen to establish warning models. In the training cohort, we identified differentially 5hMc-enriched regions (DhMRs) using DESeq2 package (version 1.30.0) in R (version 3.5.0), with the filtering threshold (p-value< 0.01 and $|\log2FoldChange| \geq 0.5$). To avoid overfitting, five rounds of 10-fold cross-validation was performed. The details are as follows: The training cohort was randomly divided into fivefolds, four of which were selected as the training subset, and the remaining one was the test subset. Then, we performed 100 repeats to further filtered using the recursive feature elimination algorithm (RFECV) in scikit-learn [parameters used: estimator = LogisticRegressionCV (class_weight = "balanced", cv = 2, max_iter = 1,000), scoring = "accuracy"]. Meanwhile, 10-fold cross-validation was repeated 100 times in each round, and the final markers observed in at least three rounds were used to build the final warning model in the training cohort. Next, we trained the logistic regression CV model (LR) with the features selected from DhMRs (parameter used: maxiter = 100, method = "lbfgs"). Finally, the trained LR model was used to warn the progression and MI for patients with COVID-19 in the validation cohort. Receiver operating characteristic (ROC) analysis was used to evaluate model performance.

Clinical Indicators Prediction Model Construction

For clinical data, continuous variables are presented as mean (SD), and categorical variables are presented as count (percentages). To understand the relationship between categorical/continuous variables and treatment outcome, Kruskal–Wallis test by ranks and $\chi 2$ test were used, respectively. A two-sided p-value of < 0.05 was considered to be statistically significant. The warning power of clinical data was

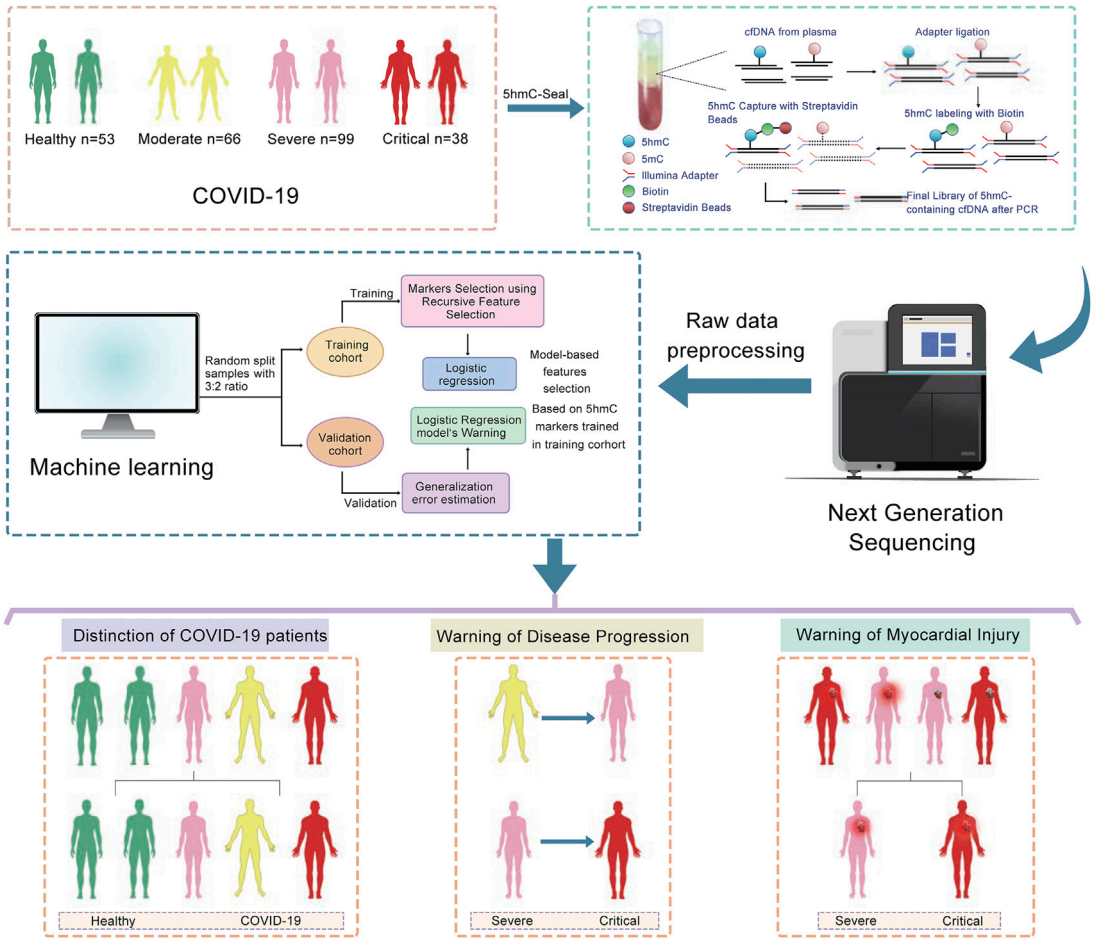


FIGURE 1 | Overview of study design. A total of 203 cfDNA samples were collected at the time of diagnosis from patients with COVID-19. 5hmC libraries for all samples were constructed with high-efficiency hmC-Seal technology. cfDNA is ligated with Illumina adapter and labeled with biotin on 5hmC for pull-down with streptavidin beads. The final library is completed by direct PCR from streptavidin beads. Next, high-throughput sequencing was performed on the NextSeq 500 platform. A logistic regression model was trained by the training cohort was used to warn COVID-19 progression and myocardial injury in the validation cohort.

estimated by generalized linear model function in R-base and pROC package (version 1.15.3) in R (version 3.6.2).

Exploring Functional Relevance of the 5hmC Markers

We used the ChIPseeker R Package (version 1.20.0) (Yu et al., 2015) to annotate the DhMRs, and the genes closest to the marker regions were used for the following functional analyses. The Gene Ontology (GO) enrichment analysis (Biological Process) was done by the ClueGO (version 2.5.5) and CluePedia (version 1.5.5) plug-in from Cytoscape software (version 3.7.2). We used the following parameters: medium network specificity, Bonferroni step down pV correction, and two-sided hypergeometric test.

GEO Datasets

For published RNA sequencing (RNA-seq) dataset, GSE150728 (Wilk et al., 2020) and GSE151879 (Chen et al., 2020), we downloaded the normalized expression values directly from Gene Expression Omnibus (GEO) database.

Statistical Analysis

Statistical analysis in **Table 1** was conducted in GraphPad Prism 5. We used two-tailed *t*-tests (paired or unpaired depending on the experiments) for normally distributed data. We used the percentile method to calculate 95% CIs, and p-value < 0.05 was considered statistically significant.

RESULTS

Sample Collected and Clinical Sample Characteristic

Table 1 shows the baseline characteristics of the patients. Of the 203 patients with COVID-19 (105 males and 98 females, the median age was 65 years), 66 patients were diagnosed with moderate symptoms, 99 patients with severe symptoms, and

TABLE 2 | Statistical characteristics of clinical indicators in patients with COVID-19.

	Total (n = 203)	Moderate (n = 66)	Severe (n = 99)	Critical (n = 38)	p-value
NEUT#, $\times10^9$/L	3.62 (3.05–6.95)	3.425 (2.83–5.41)	4.62 (3.01–6.48)	8.05 (4.79–11.86)	* ### §§§
LYMPH#, $\times10^9$/L	1.10 (0.57–1.51)	1.44 (1.14–1.83)	0.81 (0.53–1.35)	0.69 (0.37–1.22)	*** §§§
PLT, $\times10^9$/L	194 (146.25–247.5)	202.5 (169–254.75)	181 (141–239.25)	194.5 (136.25–260)	
($CD3^+CD19^-$) #,/ul	283 (77.19–708)	73.24 (67.12–84.75)	478.5 (222.25–901.5)	403.5 (192.75–597.75)	*** §§§
($CD3^+CD4^+$) #	41.8 (33.41–48.44)	39.04 (33.12–47.53)	46.84 (46.12–48.69)	44.19 (34.14–49.89)	
($CD3^+CD8^+$) #	23.89 (18.91–30.59)	24.8 (19.99–32.58)	19.34 (13.45–24.27)	21.34 (17.85–29.08)	
IL-6, pg/ml	9.44 (2.26–28.7)	2.92 (1.50–8.59)	11.99 (3.18–27.07)	29.54 (15.09–56.56)	** §§
TnI, pg/ml	0.02 (0.01–2.5)	0.005 (0.0017–0.01)	0.03 (0.01–7.60)	0.03 (0.012–0.065)	*** §§§
INR	1.02 (0.97–1.12)	0.99 (0.95–1.05)	1.02 (0.97–1.12)	1.12 (0.125–1.215)	* ## §§§
Mechanical ventilation	45	0	25	20	
If yes, mechanical ventilation, hours	494 (392–696)	NA	456 (254–684)	504 (360–696)	§§§
Hospital length of stay, days	26 (17–36)	20.5 (11–29)	31 (21–41)	26.5 (22–39)	*** §§§
Survival/non-survival	184/15	66/0	90/8	28/7	

38 patients with critical symptoms. Those severe to critical ones had a heavy burden of comorbidities such as hypertension, diabetes, and coronary heart disease. The levels of lymphocyte count (1.1×10^9/L) and neutrophil count (3.62×10^9/L) at the time of hospital admission were in the lower limit. In this study, compared with patients with moderate COVID-19, those severe to critical patients showed elder age, active cellular immunity (lower lymphocyte but higher $CD3^+CD19^-$level), and increased inflammatory response [higher neutrophil and interleukin-6 (IL-6) level)] at the baseline (**Table 2**). In addition, there were more likely to have poor outcomes presented as more extended mechanical ventilation, more days of hospital, stay, and even death. During the first 28-day follow-up, MI, SIC, and GS occurred in 40, 10, and 9 patients, respectively. A total of 14 patients died in the hospital, with the median time of 28 (8, 32) days from entering into cohort to death occurring.

The Landscape of 5hmC Profile Between the Healthy Sample and Patients with COVID-19

According to the clinical presentation, the patients with COVID-19 were classified into three disease groups (**Figure 1**; **Table 1**). First, we perform quality control (QC) analysis for 5hmC-Seal data in four groups and each sample such as the unique mapping rate and number of unique reads (**Supplementary Figure S1A-F**, **Supplementary Table S1**). Then, we identified the 5hmC-enriched peaks among the four groups and found that the groups of patients with COVID-19 enriched more peaks than the healthy control group. This was more manifest in the critical group, which showed the highest 5-hmC in different genomic characteristic regions, such as promoters and exon (**Figure 2A**). In addition, we found that the groups of patients with COVID-19 have more peaks enriched in the enhancers (**Supplementary Figure S1**). Next, we conducted differential analysis ($|\log2FoldChange| \geq 0.5$, $p < 0.01$) and observed 10,585 DhMRs (differentially 5hmC enriched regions), including upregulated (n = 7,801) and downregulated (n = 2,784) regions in the patients with COVID-19 compared with the healthy group (**Figure 2B**, **Supplementary Table S2**). We clustered the top 200 DhMRs (190 up and 10 down) detected by hierarchical clustering method. The results showed that the COVID-19 groups were well separated from the healthy people group. Meanwhile, moderate, severe, and critical groups tended to differentiate from each other (**Figure 2C**).

Next, we did GO biological pathway analysis to explore the function pathway of signature 5hmC genes in patients with COVID-19. The result showed that genes with upregulated 5hmC signal were enriched in neutrophil cells mediated immune response pathway, such as neutrophil activation (**Figure 2D**); meanwhile, the hubs of the GO functional interaction networks showed that these 5hmC-associated differential genes (n = 52), including phosphodiesterase 4D (PDE4D), CD14 molecule (CD14), and mitogen-activated protein kinase kinase kinase 4 (MAP4K4) participated in regulating neutrophil activation pathway (**Figure 2E**). The pathway enrichment in the patients with COVID-19 was consistent with the previous studies, which reported the higher level of neutrophil-to-lymphocyte is associated with severe COVID-19 (Kong et al., 2020). Furthermore, the downregulated gene enriched pathways included cell development signaling pathways (**Figure 2F**). Strikingly, consistent with our findings, 5hmC-enriched genes involved with immune response signaling pathways had higher mRNA

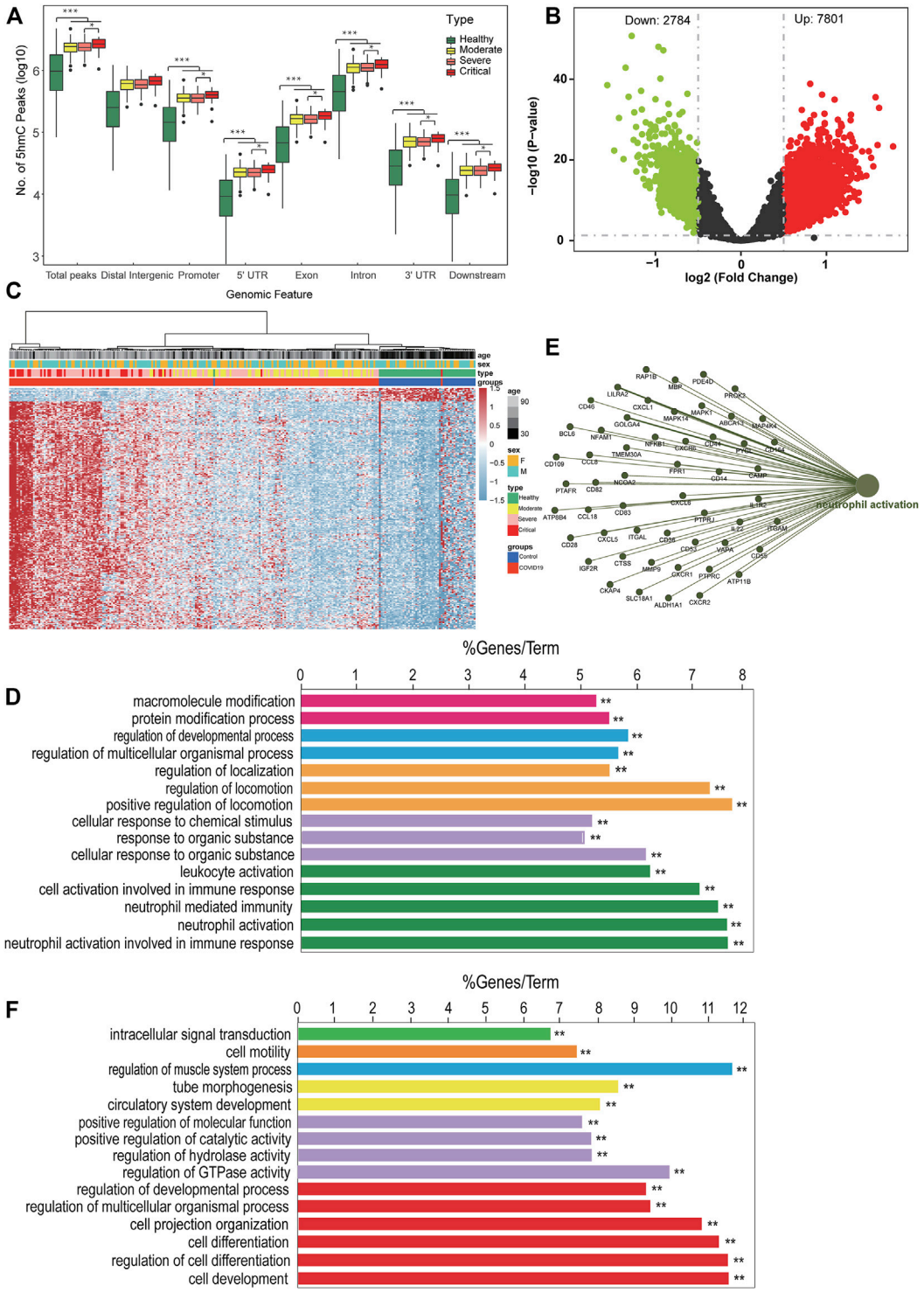


FIGURE 2 | The landscape of 5hmC in circulating cell-free DNA of patients with COVID-19. **(A)** Genome-wide 5hmC distribution in different genomic features grouped by patients with COVID-19 (**$p < 0.01$). **(B)** Volcano plot (patients with COVID-19 vs. healthy control). Significantly altered hMRs [abs (log2FoldChange) ≥0.5; p-value <0.01] are highlighted in red (up) or green (down) using the COVID-19 groups as the reference. Black dots represent the hMRs that are not differences. **(C)** Heatmap of 203 patients with COVID-19 and 53 healthy control based on top 200 DhMRs (|log2FoldChange| ≥ 0.5 and $p < 0.01$). **(D–F)** GO enrichment analysis and function exploration of 5hmC markers using Cytoscape software ($p < 0.01$). GO enrichment with 5hmC increase **(D)** or decrease **(F)** in patients with COVID-19. **(E)** GO enrichment and Gene-Concept Network. hMRs, 5 hMc-enriched regions; DhMRs, differentially 5hMc-enriched regions.

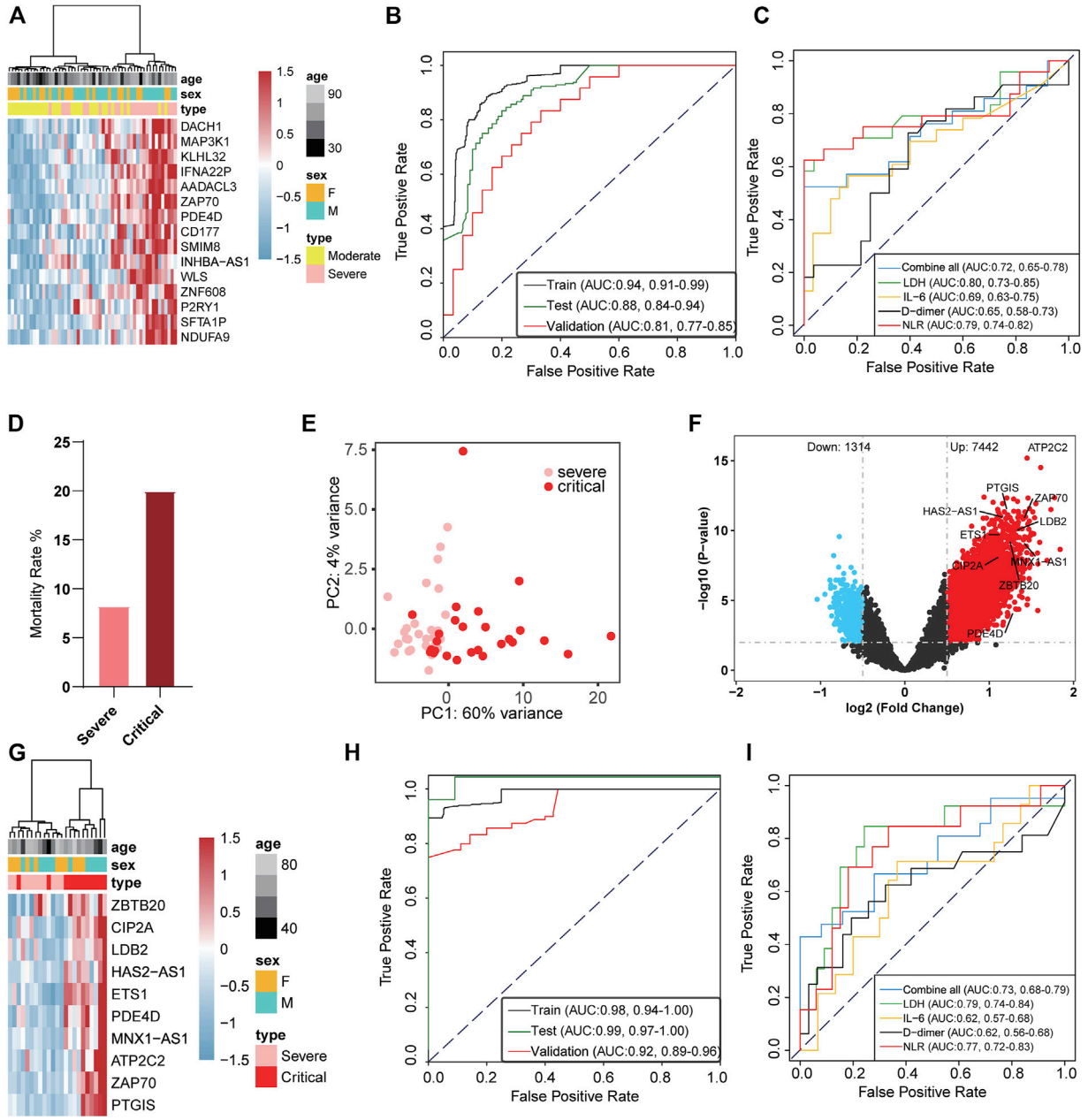


FIGURE 3 | 5-Hydroxymethylcytosine signatures in circulating cell-free DNA as a novel early warning biomarker for COVID-19 progression. **(A)** Heatmap of validation cohort based on 15 DhMRs-associated genes selected in the warning model. **(B)** Receiver operating characteristic (ROC) curve of the warning model with DhMRs in training and validation cohorts for COVID-19 progression. **(C)** ROC curve of the prediction model with clinical indicators in patients with COVID-19. **(D)** Mortality ratio in severe patients and critical patients. **(E)** Principal component analysis plot of normalized 5hmC reads from 27 severe patients and 26 critical patients. **(F)** Volcano plot (critical patients vs. severe patients). Significantly altered hMRs [abs (log2FoldChange) ≥ 0.5; *p*-value < 0.01] are highlighted in red (up) or blue (down) using the critical patients group as the reference (n = 8,756). Black dots represent the hMRs that are not differences. **(G)** Heatmap of DhMRs-associated genes selected for use in the warning model in the validation cohort. **(H)** ROC curve of the warning model with DhMRs in training and validation cohorts for COVID-19 progression. **(I)** ROC curve of the warning model with clinical indicators in COVID-19 progression. hMRs, 5hMc-enriched regions; DhMRs, differentially 5hMc-enriched regions.

expression levels in neutrophils of patients with COVID-19 from the small conditional RNA-seq dataset (GSE150728); see **Supplementary Figure S2A**. In addition, CD14 (**Supplementary Figure S2B**) and MAP4K4 (**Supplementary Figure S2C**) were highly enriched in hydroxymethylation for patients with COVID-19 (p = 0.00033 and 0.044), and the levels of hydroxymethylation increased gradually in groups moderate, severe, and critical patients and MI. All these results suggest that

differentially regulated 5hmC modified genes may have the potential to distinguish patients with COVID-19 from healthy people, and a unique combination of 5hmC modified genes would warn the disease progress.

5hmC as Early Warning Biomarkers for COVID-19 Progression

We further analyzed whether 5hmC characteristics detected in cfDNA could be used as early biomarkers for COVID-19 progression. First, we investigated whether the candidate DhMRs were associated with the severity of the disease. A total of 132 patients with COVID-19 (66 with moderate and 66 with severe) were randomly divided into training (n = 78) and validation cohorts (n = 54). Using the RFECV based on the logistic regression CV estimator, we reduced the number of DhMRs (15 DhMRs, |log2FoldChange| ≥ 0.5, p < 0.01, **Supplementary Table S3**) in the training cohort, which achieved the best cross-validation score. We found the 15 DhMRs (**Supplementary Table S9**), selected by the LR model in the training cohort, that could distinguish severe patients from moderate patients in the training (**Supplementary Figure S3A**) and validation cohorts (**Figure 3A**). Fifteen DhMRs could effectively warn moderate patients and severe patients in the training cohort [area under the curve (AUC) = 0.94, 95% CI: 0.91–0.99] and the validation cohort (AUC = 0.81, 95% CI: 0.77–0.85); see **Figure 3B**. Recent studies demonstrated that uncontrolled inflammation contributes to disease severity in COVID-19 (Wu C. et al., 2020). By developing and validating our model, we also confirmed that certain inflammatory markers such as IL-6, D-dimer, NLR, and lactate dehydrogenase (LDH) could be used as predictors of COVID-19 disease severity. Notably, the combination of IL-6, D-dimer, NLR, and LDH as a warning indicator of disease progression achieved an AUC of 0.72 (95% CI: 0.65–0.78), which is lower than the 5 hmC indicators (**Figure 3C**).

A total of 18.7% (38 cases) of patients with COVID-19 rapidly developed a critical illness and had a higher mortality rate than severe illness (**Figure 3D**). Therefore, the warning of patients who become critically ill has clear significance for the early treatment of COVID-19. To investigate whether the candidate DhMRs were associated with critical patients, we randomly separated 38 severe patients and 38 critical patients into training (n = 53) and validation cohorts (n = 23). Then, we examined the difference in 5hmC regions between severe and critical patients in the training cohort. Principal component analysis (PCA) based on top variance genes showed that severe patients could separate from the critical patients based on the 5hmC patterns (**Figure 3E**). Meanwhile, we conducted differential analysis (| log2FoldChange| ≥ 0.5, p < 0.01) and observed 8,756 DhMRs, including upregulated (n = 7,442) and downregulated (n = 1,314) regions in severe compared to critical patients (**Figure 3F** and **Supplementary Table S4**). Using the RFECV based on the logistic regression CV estimator, we further reduced the number of top 200 DhMRs from the training cohort to 10 DhMRs (**Supplementary Table S10**), which achieved the best cross-validation score. In the validation cohort, 10 DhMRs differentiated severe patients from critical patients in the training (**Supplementary Figure S3D**) and validation cohorts (**Figure 3G**). The AUC value of this model for warning severe patients and critical patients was 0.92 (95% CI: 0.89 to 0.96, **Figure 3H**), which was much higher than the clinical indicators showed, such as LDH (AUC = 0.79, 95% CI: 0.74–0.84), IL-6 (AUC = 0.62, 95% CI: 0.57–0.68), D-dimer (AUC = 0.62, 95% CI: 0.56–0.68), and NLR (AUC = 0.77, 95% CI: 0.72–0.83); see **Figure 3I**. These results implied that the 5hmC-based biomarkers of circulating cfDNA were highly indicating of COVID-19 progression. Gene functional enrichment analysis showed that upregulated 5hmC modified genes in the severe and critical patients were mainly enriched in neutrophil degranulation, neutrophil-mediated immunity, and neutrophil activation involved in immune response (see **Supplementary Figures S3B and S3C**), which are associated with development and progression of acute respiratory distress syndrome (ARDS) (Wu C. et al., 2020). Interestingly, PDE4D was one of the warning biomarkers that might indicate moderate to severe illness and severe to critical illness and mediate cell chemotaxis signaling pathways to affect the neutrophil-related immune system in critical patients (**Supplementary Figure S3E**).

5hmC as Warning Biomarkers for Myocardial Injury

In our studies, about 19.7% (40 cases) of patients with COVID-19 had MI who required ICU admission and ended up with a higher mortality rate than those without MI (**Supplementary Figure S4A**). Although the pathogenesis and biomarkers of COVID-19 have been reported, few studies addressed complications of COVID-19, especially MI. A total of 40 MI patients and 40 patients without MI were randomly selected from severe and critical patients and utilized in the warning model. Eighty patients were randomly divided into training (n = 49) and validation cohorts (n = 31). We observed MI patients separated from the without MI patients based on the DhMRs (**Figure 4A**). Meanwhile, we identified the 5hmC-enriched peaks in MI patients and found that MI patients enrich less peaks than without MI patients in different genomic characteristic regions, such as promoters and exon (**Supplementary Figure S4B**). Similarly, we identified 3,068 DhMRs (**Figure 4B**) from the training set and generated a warning model using 12 DhMRs (**Supplementary Table S11**) from all DhMRs (| log2FoldChange| ≥0.5, p-value < 0.01, **Supplementary Table S5**) was able to effectively differentiate MI patients in the training cohort (**Supplementary Figure S4C**) and the validation cohort (**Figure 4C**). In the validation group, the AUC value of this model for early warning the patients whom potentially at risk of MI was 0.89 (95% CI: 0.84 to 0.95, **Figure 4D**). This indicates that DhMRs could be early warning signs for complications of COVID-19. Besides that, we found that the 5hmC characteristics from cfDNA could be separated from other complications, such as GS and SIC (**Supplementary Figures S4D and S4E**, **Supplementary Table S6 and S7**), and separate between the non-survival patients and survival patients (**Supplementary Figure S4F**, **Supplementary Table S8**). These results suggest that the 5hmC is a potential tool for the warning of COVID-19 progression and its complications.

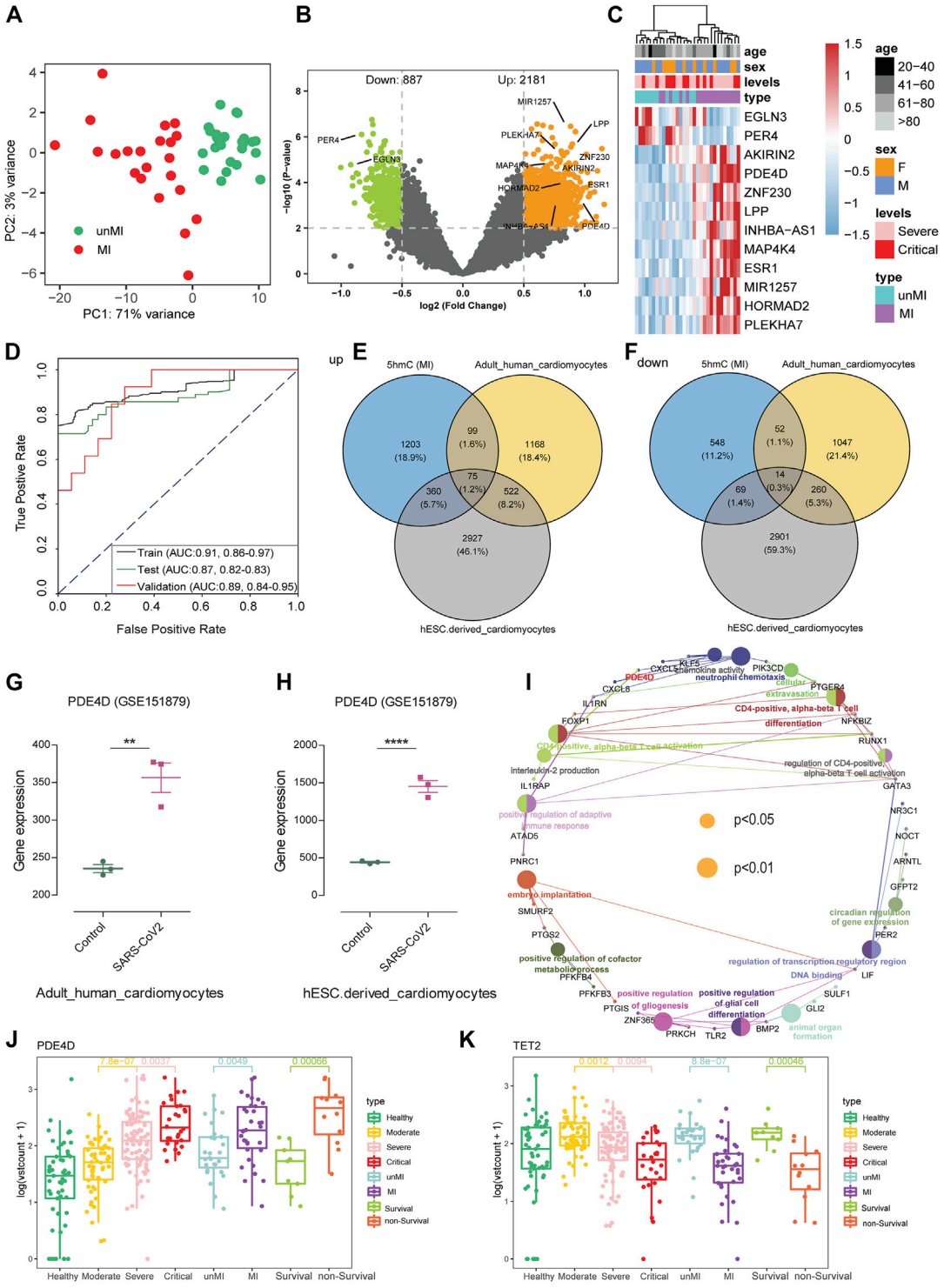


FIGURE 4 | 5-Hydroxymethylcytosine signatures as early warning biomarkers for myocardial injury. **(A)** Principal component analysis (PCA) using normalized read counts from patients with myocardial injury (MI) and those without MI (unMI). **(B)** Volcano plot of hMRs (MI patients vs. unMI patients). **(C)** Heatmaps of 12 5hmC markers with myocardial injury patients, levels, sex, and age information labeled in the validation cohort. Hierarchical clustering was performed across DhMRs-associated genes and samples. **(D)** ROC curve of the warning model with DhMRs in training and validation cohorts for myocardial injury. **(E,F)** Venn diagram indicating overlap and specificity of increase **(E)** or decrease **(F)** in myocardial injury patients from 5hmC-seq and RNA-seq (GSE151879) dataset. **(G,H)** The expression data are from the GSE151879 dataset. Each dot represents one healthy person or myocardial injury patients; error bars represent mean values. Statistically significant *p* values are indicated with asterisks (**$p < 0.01$ and ****$p < 0.0001$, by *t*-test). **(I)** GO enrichment and Gene-Concept Network with overlapping genes in myocardial injury patients. (*Continued*)

FIGURE 4 | The node size is proportional to the *p*-value calculated from the network ($p < 0.05$ and $p < 0.01$). **(J,K)** Boxplots of *PDE4D* and *TET2* grouped by healthy people, patients with COVID-19, myocardial injury (MI), and death. Log2 transformation of TMM normalized 5hmC enrichment values were plotted, and Wilcoxon *t*-test was used. 5hmC, 5-hydroxymethylcytosine; hMRs, 5hMc-enriched regions; DhMRs, differentially 5hMc-enriched regions; PDE4D, phosphodiesterase 4D; TET2, ten-eleven translocation 2.

Potential Associations Between 5hmC Markers and Myocardial Injury in patients with COVID-19

To explore the correlation of DhMRs and tissue-specific genes, we initially evaluated the tissue-specific transcriptome profiles of heart tissue from autopsies of healthy and patients with COVID-19, and human embryonic stem cell (hESC)–derived cardiomyocytes in GSE151879 datasets. Comparison between our 5hmC-seq and RNA-seq data, we found that, in addition to the unique genes between three groups, there are 75 and 14 overlapping genes in upregulated genes and downregulated genes, respectively (**Figures 4E,F**). In particular, we found a common gene, PDE4D, in all three warning models, and PDE4D are highly expressed in the heart tissue from patients with COVID-19, especially in hESC-derived cardiomyocytes (**Figures 4G,H**). Recent studies showed that uncontrolled inflammation contributes to the disease severity (Del Valle et al., 2020). Our study found the 5hmC modified genes of critical patients enriched in neutrophil-mediated immunity pathways consistent with previous research indicating that the neutrophil elastase inhibitor (Sivelestat) is a promising therapeutic option in COVID-19 with ARDS (Sahebnasagh et al., 2020). To investigate whether there are other targets for the MI or patients with a high risk of death, we examined the pathway enrichment in 75 overlapping genes. We found several immune-related signaling pathways, including chemokine activity, neutrophil chemotaxis, and CD4-positive, alpha-beta T cell activation pathways (**Figure 4I**). Among them, PDE4D plays an important role in the immune signaling pathway and influences the immune system by activating chemokines and mediating neutrophil chemotaxis (**Figure 4I**). Interestingly, we found that the 5hmC modification level of PDE4D, a drug target for chronic obstructive pulmonary disease (COPD) (Yuan et al., 2016), was significantly increased in the death group and MI group (**Figure 4J**). In addition, hydroxymethylation levels were significantly reduced in MI patients (**Supplementary figure S4B**), and the TET2 enzyme, which can catalyze the conversion of 5-methylcytosine (5mC) to 5hmC, had a lower 5hmC level in the death group and MI patients' group (**Figure 4K**).

DISCUSSION

Recent studies have reported that 5hmC plays a critical role in gene expression regulation and is also a novel tool to identify biomarkers for disease diagnosis and prognosis (Cui et al., 2020). In this study, we profiled genome-wide 5hmC in cfDNA from blood plasma, investigated its association with COVID-19 disease progression, and identified the prognostic factors associated with disease progression, MI, and mortality risk. Our primary analysis found the patients' groups enrich more peaks than the healthy group, and 5hmC marker genes differed by clinical characteristics of patients at diagnosis. We have identified COVID-19–associated 5hmC signature peaks enriched in the gene bodies and promoter regions (**Figure 2A**). The COVID-19–associated 5hmC signature gene was enriched in the neutrophil migration pathway, consistent with the previous studies that reported neutrophils and neutrophil extracellular traps drive necroinflammation in COVID-19 (Tomar et al., 2020).

Moreover, we developed a machine learning model based on 5hmC data from patients with COVID-19 at different disease severity classes (moderate, severe, and critical) to warn the disease progression. The 5hmC indicators model improved accuracy compared to the clinical markers such as LDH, IL-6, D-dimer, and NLR (**Figures 3B,C,H,I**). Overall, these findings suggest the profiled genome-wide 5hmC in cfDNA from blood plasma can be regarded as an early warning of critical illness in COVID-19. Immune phenotyping based on the LDH, IL-6, D-dimer, and NLR is a well-established marker in predicting disease severity and ICU-mortality outcomes in patients with COVID-19 (Yan et al., 2020). However, few studies were focused on the complications, diagnosis, and warning. COVID-19 is regarded as a systemic disease involving multiple systems, including cardiovascular, respiratory, gastrointestinal, and immune system (Iddir et al., 2020; Liu et al., 2020). Our results confirmed that patients who suffer from MI had higher mortality (**Supplementary figure S4A**), and 5hmC was a potential warning biomarkers of occurring MI in the COVID-19 (**Figure 4C**).

Our research tried to expand the application of 5hmC markers in the disease, especially for exploring the potential therapeutic targets (**Supplementary figure S3E**). There are two primary reasons that we think this strategy is reliable. First, Cui et al. performed the 5hmC-Seal and RNA-seq in 19 human tissues derived from 10 organ systems and found that gene-level 5hmC modifications can reflect the gene expression status in different human tissues (Cui et al., 2020). This indicates that 5hmC level is associated with gene expression, consistent with previous studies. Second, several targets used for COVID-19 treatment were also found in our results, including TET2 (Zhang et al., 2021) and neutrophil-mediated immunity pathway (Sahebnasagh et al., 2020). In addition, for the MI, the 12 5hmC-enriched regions (hMRs) were able to differentiate and predict effectively. Interestingly, in a previous study, MAP4K4 (from the 12hMRs) has been to promote cardiac muscle cell death (Virbasius and Czech, 2016). Moreover, MAP4K4 carried a higher 5hmC modification, which positively regulated the gene transcription. MAP4K4 is a key kinase in the mating pathway and is involved in many aspects of cell functions and pathological processes (Yue et al., 2014; Gao et al., 2016). Several studies found MAP4K4 as a therapeutic target in cancer (Gao et al., 2016). Thus, whether

the MAP4K4 is a potential therapeutic target for MI needs further study.

We found a potential target for the COVID-19 besides the known targets, such as the PDE4D. In our study, PDE4D was one of the 10 DhMRs that warn severe to critical illness and may mediate cell chemotaxis signaling pathways to affect the neutrophil-related immune system (**Supplementary figure S3E**). We speculated that the PDE4D could be a potential drug therapeutic target of COVID-19, especially for critical patients. GSE151879 data showed that PDE4D was highly expressed in the heart tissues of patients with COVID-19, especially in hESC-derived cardiomyocytes. It also implicated that PDE4D might play an important role in the immune system, for influencing the immune system by activating chemokines and mediating neutrophil chemotaxis (**Figure 4I**). In addition, TET2 had a lower 5hmC level in the death group and MI patients' group (**Figure 4K**), and vitamin C restoring TET2 function could provide therapy for patients with COVID-19. Recent studies consistently demonstrated that a high dose of intravenous vitamin C could improve outcomes and reduce mortality for patients with COVID-19 (Zhang et al., 2021). We believe that the combination of PDE4D inhibitor and vitamin C is a potential drug combination for the treatment of COVID-19, especially in severely ill patients.

This cohort study has several limitations. First, the number of cases was small, and small cases were used in the machine learning model generation, but our study was the first study using 5hmC as warning biomarkers in patients with COVID-19 and the exploratory study found the relevant targets may have some far-reaching significance. Second, although some clinical information is missing, it does not involve important variables. For example, we used the TnI instead of the electrocardiogram and/or echocardiography for the MI diagnosis. Third, all of the targets should be confirmed by further validation. In current pandemic, we believe that showing this important discovery in advance maybe attract more attention to understanding this new disease. Of course, we are already working on validation studies.

CONCLUSION

In conclusion, we identified potential 5hmC markers for patients with COVID-19. This is the first study using 5hmC as early warning biomarkers in patients with COVID-19, and we showed that 5hmC has advanced advantages for COVID-19 progression and MI warning.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethical Review Board of the Peking University Third Hospital. The ethics committee waived the requirement of written informed consent for participation.

AUTHOR CONTRIBUTIONS

X-xL, CL, H-yH, BZ, L-mC, HH, and Z-xL were responsible for the sample collection and summary of clinical information. H-yC, LZ, and NX participated in library construct and 5hmC sequence. H-yC, Z-rY, J-xL, LC, and C-rD contributed to bioinformatics analysis. JL, Q-gG, Cr-D, H-cZ, X-xL, CL, and H-yC were responsible for conceptual design, data analysis, manuscript writing, and submission. CH revised the article and gave suggestions.

FUNDING

This work was supported in part by National Science and Technology Major Projects for "Major New Drugs Innovation and Development", no. 2018ZX09711003, receiver: JL, 2018.

ACKNOWLEDGMENTS

We thank the patients whose samples were used in this study and thank the doctors, nurses, and public health workers for fighting against SARS-CoV-2. We appreciate for the help of study design from the Ethical Review Board of the Peking University Third Hospital. We are very grateful to Prof. Chen Jun (Department of Laboratory Medicine, Wuhan pulmonary hospital) for the help of sample collection. We would like to express our gratitude to EditSprings (https://www.editsprings.com/) for the expert linguistic services provided.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.781267/full#supplementary-material

REFERENCES

Bonow, R. O., Fonarow, G. C., O'Gara, P. T., and Yancy, C. W. (2020). Association of Coronavirus Disease 2019 (COVID-19) with Myocardial Injury and Mortality. *JAMA Cardiol.* 5 (7), 751–753. doi:10.1001/jamacardio.2020.1105

Chen, S., Yang, L., Nilsson-Payant, B., Han, Y., Jaffré, F., Zhu, J., et al. (2020). SARS-CoV-2 Infected Cardiomyocytes Recruit Monocytes by Secreting CCL2. *Res. Sqrs.3.rs-94634.* doi:10.21203/rs.3.rs-94634/v1

Consortium, E. P., Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigó, R., Gingeras, T. R., et al. (2007). Identification and Analysis of Functional Elements in 1% of the Human Genome by the ENCODE Pilot Project. *Nature* 447 (7146), 799–816. doi:10.1038/nature05874

Cui, X.-L., Nie, J., Ku, J., Dougherty, U., West-Szymanski, D. C., Collin, F., et al. (2020). A Human Tissue Map of 5-hydroxymethylcytosines Exhibits Tissue Specificity through Gene and Enhancer Modulation. *Nat. Commun.* 11 (1), 6161. doi:10.1038/s41467-020-20001-w

Del Valle, D. M., Kim-Schulze, S., Huang, H.-H., Beckmann, N. D., Nirenberg, S., Wang, B., et al. (2020). An Inflammatory Cytokine Signature Predicts COVID-19 Severity and Survival. *Nat. Med.* 26 (10), 1636–1643. doi:10.1038/s41591-020-1051-9

Dong, C., Chen, J., Zheng, J., Liang, Y., Yu, T., Liu, Y., et al. (2020). 5-Hydroxymethylcytosine Signatures in Circulating Cell-free DNA as Diagnostic and Predictive Biomarkers for Coronary Artery Disease. *Clin. Epigenet* 12 (1), 17. doi:10.1186/s13148-020-0810-2

Gao, X., Gao, C., Liu, G., and Hu, J. (2016). MAP4K4: an Emerging Therapeutic Target in Cancer. *Cell Biosci* 6, 56. doi:10.1186/s13578-016-0121-7

Han, D., Lu, X., Shih, A. H., Nie, J., You, Q., Xu, M. M., et al. (2016). A Highly Sensitive and Robust Method for Genome-wide 5hmC Profiling of Rare Cell Populations. *Mol. Cel.* 63 (4), 711–719. doi:10.1016/j.molcel.2016.06.028

Iddir, M., Brito, A., Dingeo, G., Fernandez Del Campo, S. S., Samouda, H., La Frano, M. R., et al. (2020). Strengthening the Immune System and Reducing Inflammation and Oxidative Stress through Diet and Nutrition: Considerations during the COVID-19 Crisis. *Nutrients* 12 (6), 1562. doi:10.3390/nu12061562

Kong, M., Zhang, H., Cao, X., Mao, X., and Lu, Z. (2020). Higher Level of Neutrophil-To-Lymphocyte Is Associated with Severe COVID-19. *Epidemiol. Infect.* 148, e139. doi:10.1017/S0950268820001557

Langmead, B., and Salzberg, S. L. (2012). Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* 9 (4), 357–359. doi:10.1038/nmeth.1923

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp352

Li, W., Zhang, X., Lu, X., You, L., Song, Y., Luo, Z., et al. (2017). 5-Hydroxymethylcytosine Signatures in Circulating Cell-free DNA as Diagnostic Biomarkers for Human Cancers. *Cell Res* 27 (10), 1243–1257. doi:10.1038/cr.2017.121

Lin, Z., Long, F., Yang, Y., Chen, X., Xu, L., and Yang, M. (2020). Serum Ferritin as an Independent Risk Factor for Severity in COVID-19 Patients. *J. Infect.* 81 (4), 647–679. doi:10.1016/j.jinf.2020.06.053

Liu, P. P., Blet, A., Smyth, D., and Li, H. (2020). The Science Underlying COVID-19. *Circulation* 142 (1), 68–78. doi:10.1161/CIRCULATIONAHA.120.047549

McArthur, L., Sakthivel, D., Ataide, R., Chan, F., Richards, J. S., and Narh, C. A. (2020). Review of Burden, Clinical Definitions, and Management of COVID-19 Cases. *Am. J. Trop. Med. Hyg.* 103 (2), 625–638. doi:10.4269/ajtmh.20-0564

National Health Commission & National Administration of Traditional Chinese Medicine (2020). Clinical Management of Human Infection with Novel Coronavirus (2019-nCoV) (Trial Guidance V8). Available at: http://www.gov.cn/zhengce/zhengceku/2020-08/19/5535757/files/da89edf7cc9244fbb34ecf6c61df40bf.pdf (Accessed 08 2020, 18).

Pollard, C. A., Morran, M. P., and Nestor-Kalinoski, A. L. (2020). The COVID-19 Pandemic: a Global Health Crisis. *Physiol. Genomics* 52 (11), 549–557. doi:10.1152/physiolgenomics.00089.2020

Quinlan, A. R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr. Protoc. Bioinformatics* 47, 1111–1234. doi:10.1002/0471250953.bi1112s47

Sahebnasagh, A., Saghafi, F., Safdari, M., Khataminia, M., Sadremomtaz, A., Talaei, Z., et al. (2020). Neutrophil Elastase Inhibitor (Sivelestat) May Be a Promising Therapeutic Option for Management of Acute Lung Injury/acute Respiratory Distress Syndrome or Disseminated Intravascular Coagulation in COVID-19. *J. Clin. Pharm. Ther.* 45 (6), 1515–1519. doi:10.1111/jcpt.13251

Shen, B., Yi, X., Sun, Y., Bi, X., Du, J., Zhang, C., et al. (2020). Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell* 182 (1), 59–72. doi:10.1016/j.cell.2020.05.032

Song, C.-X., Szulwach, K. E., Fu, Y., Dai, Q., Yi, C., Li, X., et al. (2011). Selective Chemical Labeling Reveals the Genome-wide Distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* 29 (1), 68–72. doi:10.1038/nbt.1732

Szulwach, K. E., Li, X., Li, Y., Song, C.-X., Wu, H., Dai, Q., et al. (2011). 5-hmC-mediated Epigenetic Dynamics during Postnatal Neurodevelopment and Aging. *Nat. Neurosci.* 14 (12), 1607–1616. doi:10.1038/nn.2959

Tan, L., and Shi, Y. G. (2012). Tet Family Proteins and 5-hydroxymethylcytosine in Development and Disease. *Development* 139 (11), 1895–1902. doi:10.1242/dev.070771

Tomar, B., Anders, H.-J., Desai, J., and Mulay, S. R. (2020). Neutrophils and Neutrophil Extracellular Traps Drive Necroinflammation in COVID-19. *Cells* 9 (6), 1383. doi:10.3390/cells9061383

Virbasius, J. V., and Czech, M. P. (2016). MAP4K4 Signaling Nodes in Metabolic and Cardiovascular Diseases. *Trends Endocrinol. Metab.* 27 (7), 484–492. doi:10.1016/j.tem.2016.04.006

Wang, L. (2020). C-reactive Protein Levels in the Early Stage of COVID-19. *Médecine et Maladies Infectieuses* 50 (4), 332–334. doi:10.1016/j.medmal.2020.03.007

Wang, T., Du, Z., Zhu, F., Cao, Z., An, Y., Gao, Y., et al. (2020). Comorbidities and Multi-Organ Injuries in the Treatment of COVID-19. *The Lancet* 395 (10228), e52. doi:10.1016/S0140-6736(20)30558-4

Wilk, A. J., Rustagi, A., Zhao, N. Q., Roque, J., Martínez-Colón, G. J., McKechnie, J. L., et al. (2020). A Single-Cell Atlas of the Peripheral Immune Response in Patients with Severe COVID-19. *Nat. Med.* 26 (7), 1070–1076. doi:10.1038/s41591-020-0944-y

Wu, C., Chen, X., Cai, Y., Xia, J. a., Zhou, X., Xu, S., et al. (2020a). Risk Factors Associated with Acute Respiratory Distress Syndrome and Death in Patients with Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA Intern. Med.* 180 (7), 934–943. doi:10.1001/jamainternmed.2020.0994

Wu, D., Shu, T., Yang, X., Song, J. X., Zhang, M., Yao, C., et al. (2020b). Plasma Metabolomic and Lipidomic Alterations Associated with COVID-19. *Natl. Sci. Rev.* 7 (7), 1157–1168. doi:10.1093/nsr/nwaa086

Yan, L., Zhang, H.-T., Goncalves, J., Xiao, Y., Wang, M., Guo, Y., et al. (2020). An Interpretable Mortality Prediction Model for COVID-19 Patients. *Nat. Mach. Intell.* 2 (5), 283–288. doi:10.1038/s42256-020-0180-7

Yu, G., Wang, L.-G., and He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor Package for ChIP Peak Annotation, Comparison and Visualization. *Bioinformatics* 31 (14), 2382–2383. doi:10.1093/bioinformatics/btv145

Yuan, L., Dai, X., Yang, M., Cai, Q., and Shao, N. (2016). Potential Treatment Benefits and Safety of Roflumilast in COPD: a Systematic Review and Meta-Analysis. *Copd* 11, 1477–1483. doi:10.2147/COPD.S106370

Yue, J., Xie, M., Gou, X., Lee, P., Schneider, M. D., and Wu, X. (2014). Microtubules Regulate Focal Adhesion Dynamics through MAP4K4. *Dev. Cel.* 31 (5), 572–585. doi:10.1016/j.devcel.2014.10.025

Zhang, J., Rao, X., Li, Y., Zhu, Y., Liu, F., Guo, G., et al. (2021). Pilot Trial of High-Dose Vitamin C in Critically Ill COVID-19 Patients. *Ann. Intensive Care* 11 (1), 5. doi:10.1186/s13613-020-00792-3

Zhang, L., Yan, X., Fan, Q., Liu, H., Liu, X., Liu, Z., et al. (2020). D-dimer Levels on Admission to Predict In-hospital Mortality in Patients with Covid-19. *J. Thromb. Haemost.* 18 (6), 1324–1329. doi:10.1111/jth.14859

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.

frontiers in Cell and Developmental Biology

ORIGINAL RESEARCH
published: 07 March 2022
doi: 10.3389/fcell.2022.827391

Comprehensive Analysis of DNA 5-Methylcytosine and N6-Adenine Methylation by Nanopore Sequencing in Hepatocellular Carcinoma

*Lili Zhang[1,2†], Weiqi Rong[3†], Jie Ma[4†], Hexin Li[1], Xiaokun Tang[1], Siyuan Xu[1], Luyao Wang[1], Li Wan[1], Qing Zhu[4], Boyue Jiang[2], Fei Su[1] and Hongyuan Cui[2,4]**

[1]Clinical Biobank, Beijing Hospital, National Center of Gerontology, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing, China, [2]Department of General Surgery, Department of Hepato-Bilio-Pancreatic Surgery, Beijing Hospital, National Center of Gerontology, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing, China, [3]Department of Hepatobiliary Surgery, National Cancer Center, National Clinical Research Center for Cancer, Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [4]Department of Hepatopancreatobiliary Surgery, Affiliated Hospital of Qinghai University, Qinghai, China

DNA methylation is a widespread epigenetic signal in human genome. With Nanopore technology, differential methylation modifications including 5-methylcytosine (5mC) and 6-methyladenine (6mA) can be identified. 5mC is the most important modification in mammals, although 6mA may also function in growth and development as well as in pathogenesis. While the role of 5mC at CpG islands in promoter regions associated with transcriptional regulation has been well studied, but the relationship between 6mA and transcription is still unclear. Thus, we collected two pairs of tumor tissues and adjacent normal tissues from hepatocellular carcinoma (HCC) surgical samples for Nanopore sequencing and transcriptome sequencing. It was found that 2,373 genes had both 5mC and 6mA, along with up- and down-regulated methylation sites. These genes were regarded as unstable methylation genes. Compared with 6mA, 5mC had more inclined distribution of unstable methylation sites. Chi-square test showed that the levels of 5mC were consistent with both up- and down-regulated genes, but 6mA was not significant. Moreover, the top three unstable methylation genes, TBC1D3H, CSMD1, and ROBO2, were all related to cancer. Transcriptome and survival analyses revealed four potential tumor suppressor genes including KCNIP4, CACNA1C, PACRG, and ST6GALNAC3. In this study, we firstly proposed to combine 5mC and 6mA methylation sites to explore functional genes, and further research found top of these unstable methylation genes might be functional and some of them could serve as potential tumor suppressor genes. Our study provided a new solution for epigenetic regulation research and therapy of HCC.

Keywords: 5-methylcytosine, N6-adenine methylation, nanopore sequencing, hepatocellular carcinoma, unstable methylation genes

OPEN ACCESS

Edited by:
Chunjie Jiang,
University of Pennsylvania,
United States

Reviewed by:
Tao P. Wu,
Baylor College of Medicine,
United States
Fei Peng,
Baylor College of Medicine,
United States
Hao Zhang,
University of Pennsylvania,
United States

****Correspondence:***
Hongyuan Cui
cuihongyuan3921@bjhmoh.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular and Cellular Pathology, a section of the journal Frontiers in Cell and Developmental Biology

Received: *02 December 2021*
Accepted: *10 February 2022*
Published: *07 March 2022*

Citation:
Zhang L, Rong W, Ma J, Li H, Tang X, Xu S, Wang L, Wan L, Zhu Q, Jiang B, Su F and Cui H (2022) Comprehensive Analysis of DNA 5-Methylcytosine and N6-Adenine Methylation by Nanopore Sequencing in Hepatocellular Carcinoma. Front. Cell Dev. Biol. 10:827391. doi: 10.3389/fcell.2022.827391

1 INTRODUCTION

Hepatocellular carcinoma (HCC) is a common malignant tumor and also the fourth leading cause of cancer-related death worldwide (Kole et al., 2020; Sung et al., 2021). Although the exact etiologies of HCC remain unclear, both acute and chronic infections with hepatitis B virus (HBV) and hepatitis C virus (HCV) can be the major causes. Hepatitis can lead to liver cirrhosis and HCC (Ringelhan et al., 2017). Some HCC patients may benefit from current treatment strategies including radiofrequency ablation, surgery, liver transplantation, and immunotherapy (Kole et al., 2020). Traditionally, genetic instability is regarded as one of the most common events in HCC. Recent studies have revealed that HCC can be triggered by epigenetic modifications (Sceusi et al., 2011).

In mammal genomes, cytosine DNA methylation (5-methylcytosine, 5mC) is a widespread form of methylation and can function by directly regulating the occurrence and development of cancers (Mo et al., 2020; Lowe et al., 2021). In HCC, 5mC is closely associated with survival rates and prognosis (Hlady et al., 2019; Mo et al., 2020). Recent evidence has accumulated that the regulation of 5mC-mediated gene silencing usually operate via CpG islands methylation, while most CpG islands associated with gene promoters are rarely methylated (de Mendoza et al., 2019). It is inferred that gene body 5mC also positively influences gene expression, which is targeted by DNMT3 in PWWP domain, and the PWWP domain also attracts an active transcription factor H3K36me3 (Baubec et al., 2015; Morselli et al., 2015). Thus, gene body 5mC may prime for active transcription but are not predictive. The 5mC derived from cell-free DNA can be markers in HCC (Hlady et al., 2019), reveals potential advantages of methylation to the development of liquid biopsy.

N6-adenine methylation (6mA), initially a marker of DNA modification in prokaryotes, has been identified in eukaryotes, especially in mammalian and plant genomes. 6mA plays a role in growth, development, and tumor progression (Zhou et al., 2018). The whole 6mA modification map in human has been obtained via SMRT sequencing, confirming that 6mA is widespread in nuclear genome and mitochondrial genome (Xiao et al., 2018). Further studies have shown the amount of 6mA is extremely low in eukaryote, while the modification of 6mA in mammals is mainly concentrated on the mitochondrial DNA (Kawarada et al., 2017). However, the small number of sites does not mean they are not functional. For example, in c-kit gene, a single 5mC methylation site is enough to affect the binding of the transcription factor GATA-1 to the gene body and regulate the normal development of hematopoietic system (Yang et al., 2020). During the development of mouse trophoblast stem cells, the 6mA-mediated repression of stress-induced DNA double helix destabilization-SATB1 interactions is essential for gene regulation. 6mA can balance the boundaries between euchromatin and heterochromatin (Li et al., 2020), and knockdown of 6mA methyltransferase also affects transcriptome-wide fluctuation of gene expression (Luo et al., 2018). Thus, there are connections between DNA modification and gene expression, although the specific regulatory mechanisms are still uncertain.

Differential methylation can be identified by using long-sequencing nanopore technology. We selected methylation sites that are different between tumor and adjacent normal tissues. Chi-square test revealed that the levels of 5mC were significantly correlated with gene transcription, but 6mA didn't show much relevance. In HCC, 2,373 genes had both 5mC and 6mA, with up- and down-regulated methylation sites. We considered these genes as unstable methylation genes. Since 5mC and 6mA can influence gene expressions, unstable methylation genes with top amount of methylation sites were selected. Based on transcriptome information, we found eleven genes in top 100 unstable methylation genes could affect the prognosis of patients and four of them could be potential tumor suppressor genes.

2 MATERIALS AND METHODS

2.1 Patients and Samples

As sequencing samples, two pairs of tumor tissues and adjacent normal tissues were collected from HCC surgical samples at the Cancer Hospital of Chinese Academy of Medical Sciences. The specimens were immediately stored at −80°C. The tumor tissues comprised >80% malignant cells, and the normal tissues comprised a mixture of normal epithelial cells and stromal cells. The constructed Nanopore sequencing library and RNA library were same as in our previous research (Zhuo et al., 2021). All the patients signed informed consent forms, and the research was approved by the Ethics Committee of Beijing Hospital.

2.2 Nanopore Sequencing and Illumina Sequencing

For Nanopore sequencing, DNA in tissues was extracted by using MagAttract HMW DNA kit (Qiagen). The double-stranded DNA was quantified by Nanodrop 2000 and Qubit dsDNA HS analysis kits (Thermo Fisher Scientific). AMPure XP (Beckman Coulter) and Qubit®3.0 fluorometer (Life Technologies) were used to purify and concentrate DNA. The Nanopore sequencing platform was GridION, R9.4.1 chip (ONT). After the quality of chip was checked in accordance with the manufacturer's instructions, samples were sequenced. MinKNOW (v3.5.10) and Guppy (v3.2.6) software were used to collect raw electronic signals and convert the files to FASTQ format.

For Illumina sequencing, the DynabeadsTM mRNA Purification Kit (Invitrogen) was used to extract mRNA from total RNA. Ribo-Zero Gold Kits were utilized to remove rRNA. According to the instructions of the NEB-Next Ultra Directional RNA Library Prep Kit for Illumina (NEB, Ispawich, USA), different index tags were selected. The constructed libraries were sequenced using Illumina with 150 bp paired-end reads.

2.3 Methylation Calling for Nanopore Reads

For 5mC, Minimap2 (Li, 2018) was used to align sequencing reads to human genome (GRCh37). Nanopolish call-methylation

(Simpson et al., 2017) was used to determine the methylation status of CpG site. The "-s" option in Nanopolish script was used to split group into individual sites.

For 6mA, we aligned data by using the re-squiggle algorithm in Tombo (version 1.4) (https://nanoporetech.github.io/tombo/). The alternative DNA models were available in Tombo and the 6mA model was used. The method identified 6mA methylation sites better than the canonical expected levels, which signal matches a specific alternative base expected signal levels.

After the methylation signal scores of each genome sites were obtained, we removed sites with all zero scores. The scores ranged from 0 to 1, in which 0 means completely unmethylated and 1 means completely methylated. Only sites whose methylation score changed by greater than 2 folds were considered. When scores in part of tissues were zero, the sites with scores greater than 0.6 in the remaining tissues were defined as different sites. The sites with significantly different scores were regarded as unstable methylation sites. We used Bedtools intersect (Quinlan and Hall, 2010) to annotate genes corresponding to unstable methylation sites. The distribution of methylation sites in genome was drawn by package CMplot (Yin et al., 2021) in R language. The methylation intersection map used VennDetail-Shiny (http://hurlab.med.und.edu:3838/VennDetail/).

In order to verify the accuracy of methylation sites, Megalodon (https://github.com/nanoporetech/megalodon) was applied to reanalyze the sequencing data based on the previous studies (Yuen et al., 2021). The intersection of two data sets from different methods was obtained. The methylation model was downloaded from Rerio (https://github.com/nanoporetech/rerio) as recommended "res_dna_r941_min_modbases-all-context_v001". In order to verify the accuracy of methylation analysis, the PCR amplified fragments of top unstable methylation genes and tumor suppressor genes were performed on Nanopore (**Supplementary Figure S1**), and the sequencing data were analyzed parallelly. The sequences generated by PCR amplification did not have any methylation sites (**Supplementary Figure S2**) which proved the reliability of our data processes.

2.4 Differential Gene Expression

FastQC (Andrews, 2010) was used to evaluated the quality of sequencing data, which revealed some reads mixed with adapters. In order to filter reads, we used Trimmomatic (Bolger et al., 2014). The software dropped reads less than 28 nt, and the average quality was less than 15 through a four-base sliding window. After the data was qualified, the reads were mapped to genome by STAR (Dobin et al., 2013). For each gene, we chose featureCounts (Liao et al., 2014) to quantify read counts. The parameter "requireBothEndsMapped" and "isPairedEnd" were set TRUE. By DGEList and rpkm function in package edgeR (Robinson et al., 2010), we calculated the normalized expression levels of genes. We chose genes with fold-change greater than 2 as differential expressed genes.

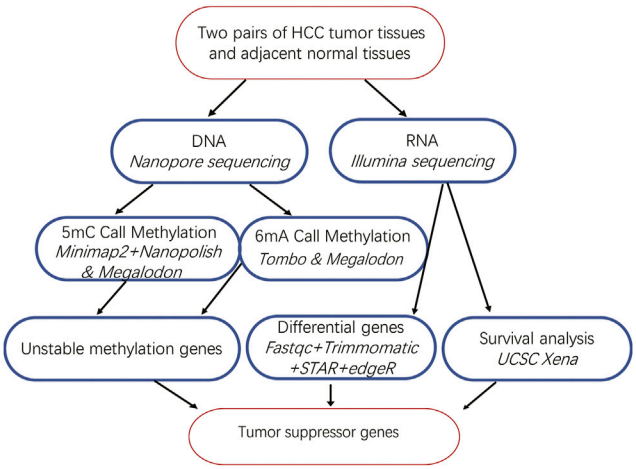


FIGURE 1 | Bioinformatics analysis pipeline of our study.

2.5 Functional Enrichment Analysis and Survival Analysis

The R package ClusterProfiler (Wu et al., 2021) was used to analyse gene set enrichment. It queried the latest online database to perform functional analysis and allowed the output up-to-date. The pathways were drawn by using ggplot2 (Wickham, 2016).

For survival analysis, we used TCGA liver cancer (LIHC) in UCSC Xena (Goldman et al., 2020) with overall survival to get Kaplan-Meier plots of unstable methylation genes. The "p-value < 0.05" genes in survival analysis were selected.

3 RESULT

3.1 The Distribution of Methylation Sites in Genome

After obtaining tumor and adjacent normal tissues, we constructed sequencing libraries and developed analysis pipeline (**Figure 1**). Output of sequencing data showed the wide distribution of 5mC and 6mA sites in the genome (**Figure 2**). Except for centromere region which full of repetitive sequences affect detection, both 5mC and 6mA were widely distributed in the genome. For the comparison of the methylation signals between HCC tumor and adjacent normal liver tissues, we extracted 5mC and 6mA sites that were significantly different and compared the methylation transition of these sites and their genes (**Figure 3**). Since the change of a single site may affect gene expression (Yang et al., 2020), we recalled all the sites with methylation score changed at the gene level. The filtered up-regulated or down-regulated sites were the methylation signals with scores changed by more than two folds. For genes that might be affected, the amount of 5mC changed was significantly higher than that of 6mA (**Figures 3A,B**). Notably, there were 2,373 genes having both own 5mC and

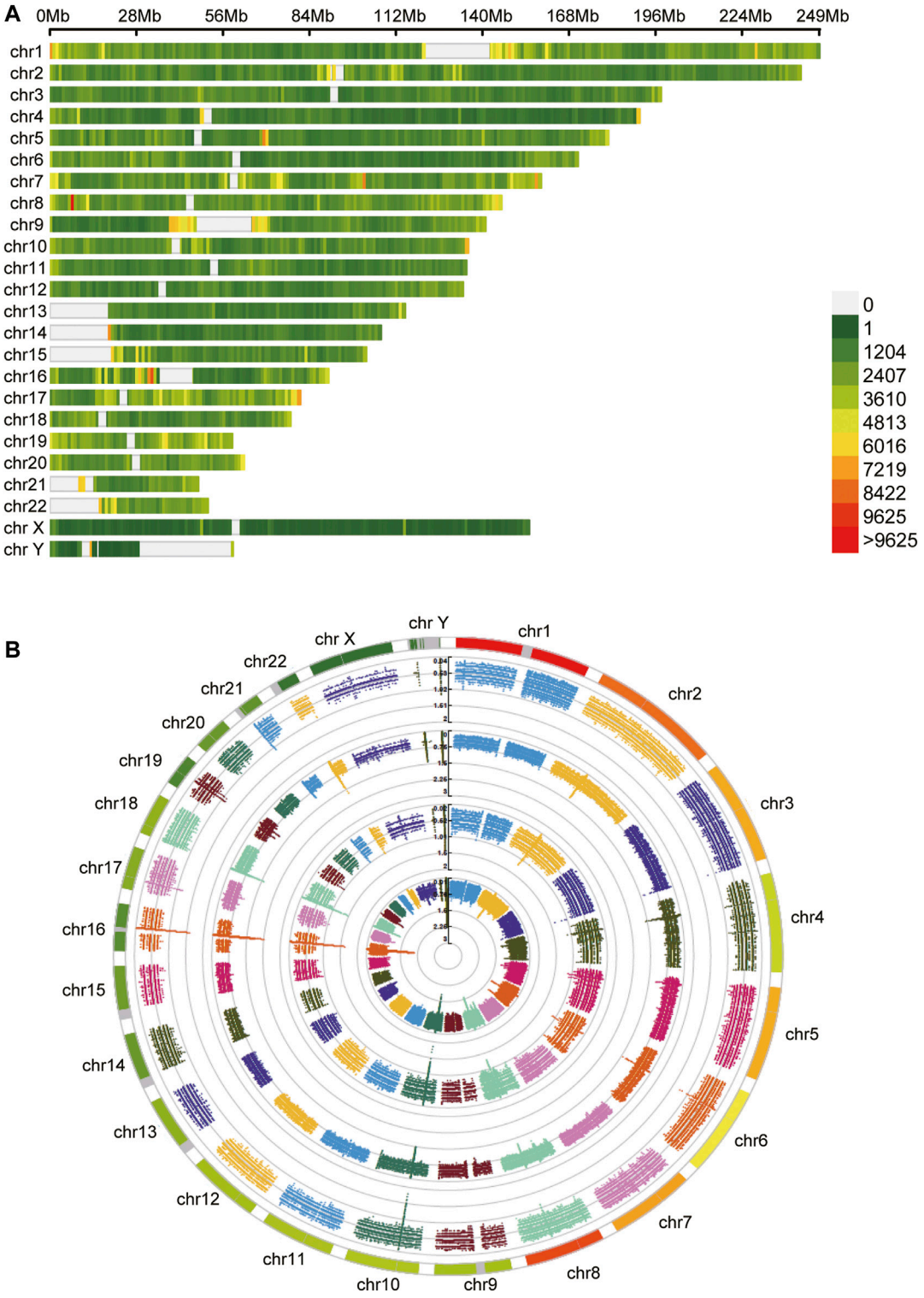


FIGURE 2 | The distribution of methylation sites. **(A)** The 5mC unstable methylation sites in genome. The gap regions belong to genome centromere region. **(B)** The 6mA unstable methylation sites in genome.

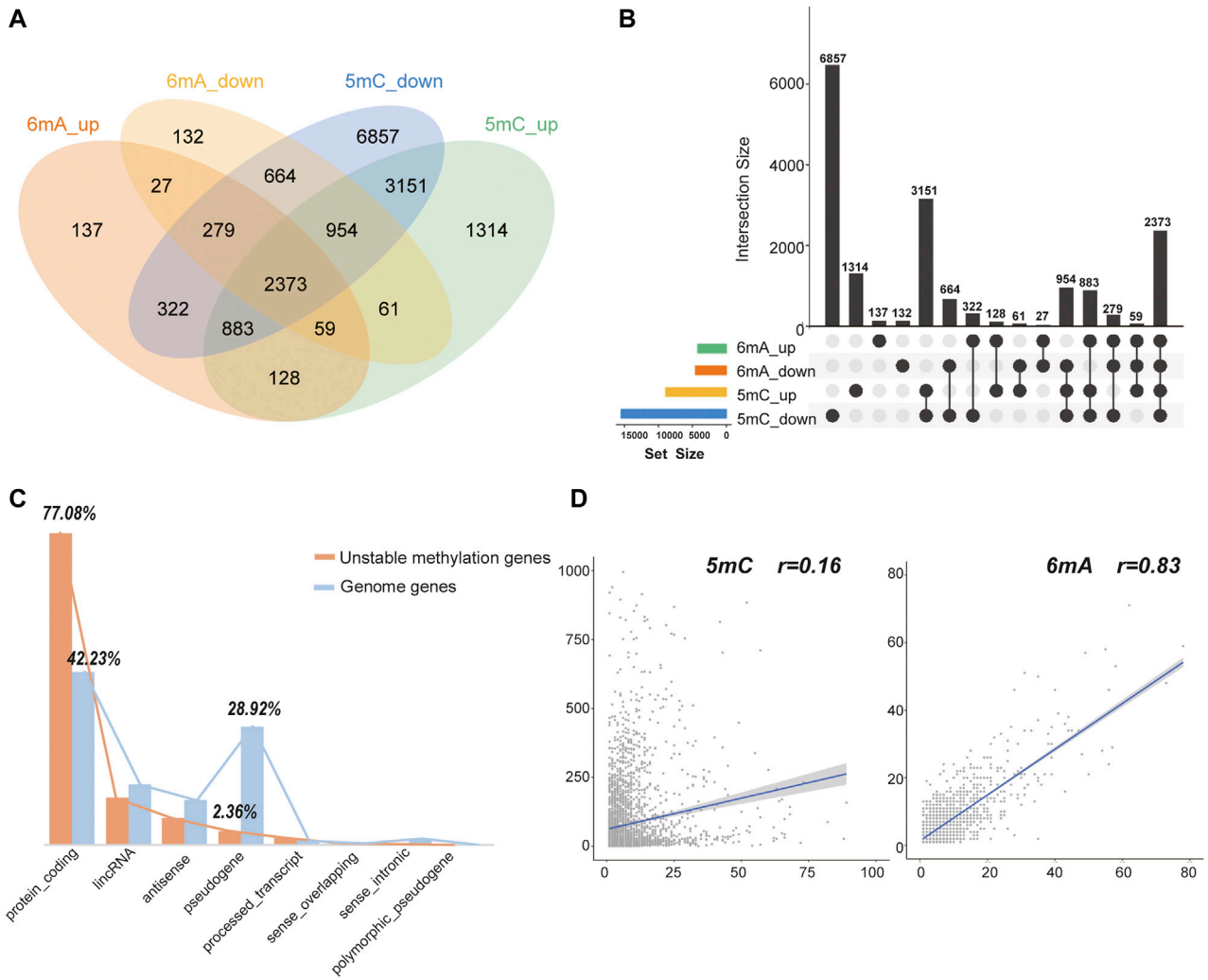


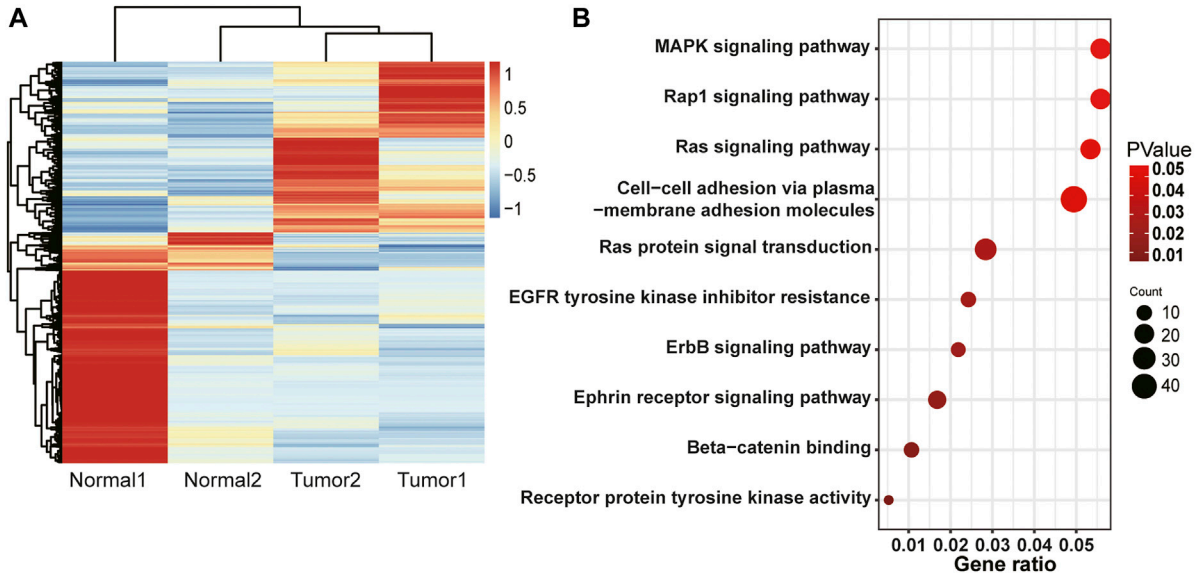
FIGURE 3 | The statistics of 5mC and 6mA methylation changed sites. **(A,B)** The intersection number of genes with 5mC and 6mA, up- and down-regulated methylation sites. **(C)** The gene types of unstable methylation genes. The orange column represented unstable methylation genes, and the blue column represented genes in human genome. **(D)** The Pearson correlation coefficient between the number of up-regulated and down-regulated sites in 5mC and 6mA. It is only 0.16 for 5mC but 0.83 for 6mA.

6mA, with up- and down-regulated methylation sites (**Supplementary File S1**). These genes were considered as unstable methylation genes in HCC.

Further analysis showed that 77.08% of these unstable methylation genes belonged to protein-coding genes, which was much higher than 42.23% of protein coding genes in genome. While 28.92% of pseudogenes were found in genome, the proportion of 5mC and 6mA changed was only 2.36% (**Figure 3C**). Thus, genes with unstable methylation were more likely to be protein-coding genes, and these unstable methylation genes might have regulatory functions in transcription and translation. In addition, the Pearson correlation coefficient between the number of up-regulated and down-regulated sites in 6mA was 0.83, while the correlation coefficient in 5mC was only 0.16 (**Figure 3D**). It demonstrates the distribution of 5mC was more specific than mA. In addition, 45.47% of these unstable methylation genes had altered transcriptome suggesting the methylation in 5mC site was more specific than 6mA.

3.2 Relationship Between Methylation and Transcription

After acquiring the sequencing data, we focused on transcriptome (**Figure 4A**) and found the differentially expressed genes were indeed enriched in pathways related to liver cancer (**Figure 4B**). Furthermore, Chi-square test was used to explore whether 5mC, 6mA were related to differentially expressed genes. It showed that the amount of differentially expressed genes was highly correlated with 5mC-modifications instead of 6mA which is also consistent with the correlation of up- and down-regulated sites in **Figure 3D**. In HCC, 1,470 genes had specifically up-regulated



C

Chi-square Test P-value=3.265e-04

5mC / Transcriptome	5mC up	5mC down	SUM
Low expression	174	1207	1381
High expression	173	785	958
SUM	347	1992	2339

Chi-square Test P-value=0.2199

6mA / Transcriptome	6mA up	6mA down	SUM
Low expression	239	284	523
High expression	159	157	316
SUM	398	441	839

D

Chi-square Test P-value=2.574e-02

5mC / Transcriptome	up>=down	up<down	SUM
Low expression	442	1411	1853
High expression	290	758	1048
SUM	732	2169	2901

Chi-square Test P-value=0.3274

6mA / Transcriptome	up>=down	up<down	SUM
Low expression	484	336	820
High expression	240	146	386
SUM	724	482	1206

FIGURE 4 | The association between differentially expressed genes and methylation related genes. **(A)** Heatmap of two pairs transcriptome sequencing data. **(B)** The enriched pathways of differentially expressed genes in HCC. They were closely associated with HCC. **(C)** The number of 5mC, 6mA genes and differentially expressed genes. The Chi-square test indicated that 5mC were significantly associated with differentially expressed genes, but 6mA were not. **(D)** The number of genes both own 5mC or 6mA up and down methylation sites, as well as differentially expressed genes. The Chi-square test also indicated that 5mC were more related with gene transcription than 6mA.

6mA sites. The amount of the specific down-regulated 6mA genes was 1,811. Totally, there were 1,417 differentially expressed genes. After classification, Chi-square test indicated that the specifically regulated 5mC genes instead of 6mA were significantly associated with differentially expressed genes (**Figure 4C**). Similarly, the significant difference of 5mC was also observed with the *p*-value of 3.265e-04 (**Figure 4C**). The results indicated that in HCC patients, 5mC methylation plays more important role than 6mA in differential transcription.

While genes possess both up and downregulated methylation signals, it is worth to note that whether more methylation signals have more influence on differentially expressed genes, such as

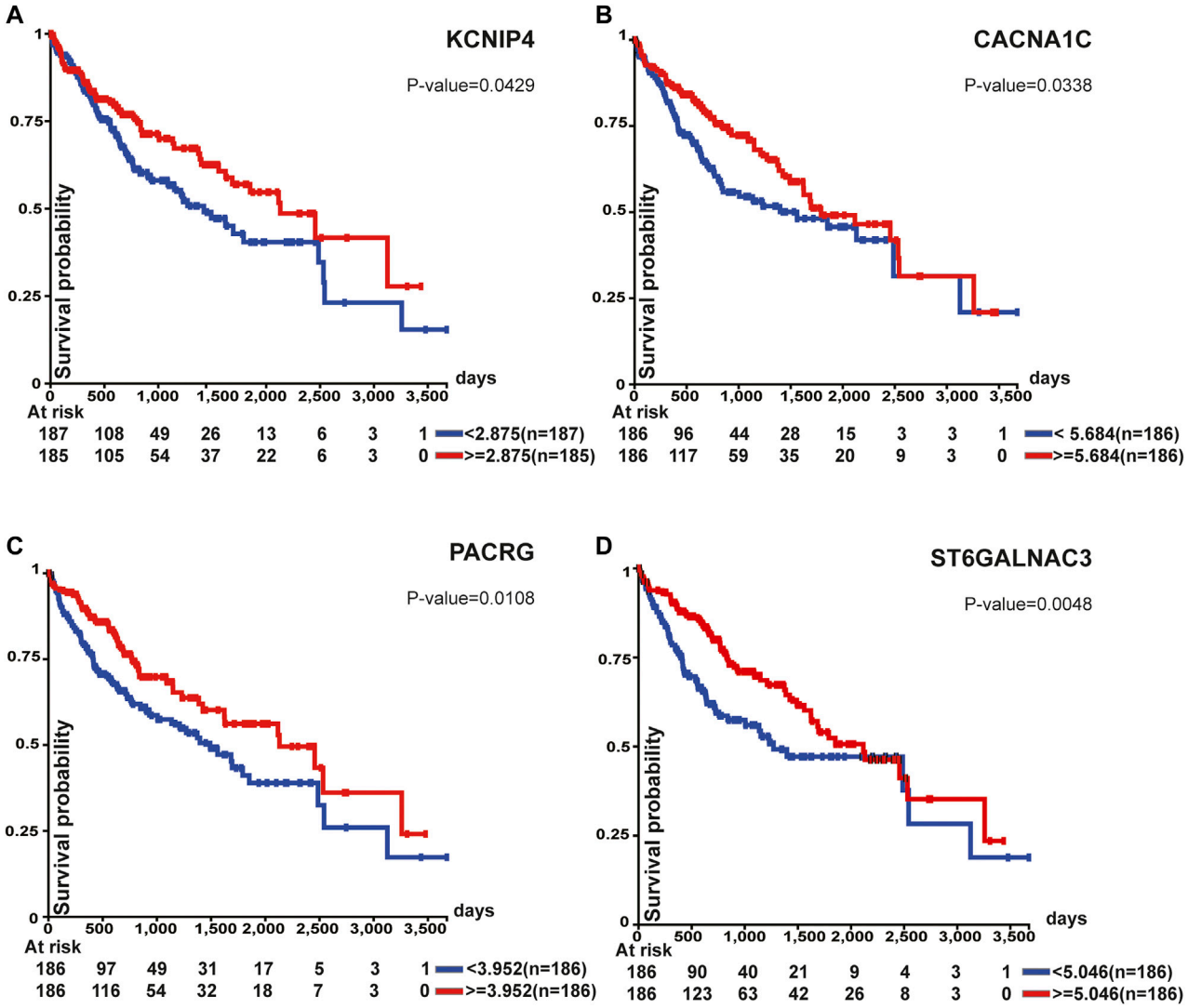


FIGURE 5 | Survival analysis of tumor suppressor genes. All of these genes were associated with overall survival.

genes with more upregulated methylation sites tend to be low-expressed in transcriptome. Thus, we counted 2,901 genes with 5mC up and downregulated methylation signals and 1,206 genes with 6mA (**Figure 4D**). Chi-square test also demonstrated that the levels of 5mC were consistent with both up and downregulated genes, but 6mA was not significant. Although the relationship between the level of 5mC and the mRNA was not linear, the influence of 5mC on transcription is more remarkable than 6mA. It revealed that DNA methylation on the 5mC plays a more important role than 6mA in transcriptional regulation.

3.3 Unstable Methylation Genes and Their Relationships With Survival

DNA methylation accumulation and the epigenotype formation on the genome may occur in the early stages of carcinogenesis and can predict the future cancer type (Kaneda et al., 2014). Therefore, genes with more accumulated methylation changed sites may play roles in the occurence and development of cancer. In order to explore HCC-related unstable methylation genes, we combined HCC transcriptome data and TCGA survival analysis.

In HCC, the genes at top list of unstable methylation sites are TBC1D3H, CSMD1 and ROBO2. TBC1D3H belongs to TBC1 domain family, which can act as a GTPase activating protein for RAB5 (Itoh et al., 2006; UniProt, 2021). Its dysregulation can lead to tumorigenesis (Jian et al., 2020). Other proteins with TBC1 domain also function in cancer; for instance, TBC1 domain family member 23 can interact with Ras-related protein Rab-11A to promote poor prognosis in lung cancer (Zhang et al., 2021). The second unstable methylation gene is CSMD1. Deregulation of CSMD1 can link inflammation to carcinogenesis via activating NF-κB pathway. The process subsequently leads to the upregulation of c-Myc and epithelial-mesenchymal transition markers (Chen et al., 2019).

It has also been reported that combined identification of ARID1A, SENP3, and CSMD1 are effective prognostic biomarkers for HCC patients (Zhao et al., 2021). The third gene RoBo2 can suppress cancer development through TGF-β signalling and stroma activation (Pinho et al., 2018). The evidence showed that the top unstable methylation genes are involved in the occurrence and development of cancer.

Survival analysis of top 100 unstable methylation genes was performed. As shown by the Kaplan-Meier plot (p-value<0.05), 11 genes were significantly associated with overall survival of patients (**Supplementary File S1**). Downregulation of tumor suppressor genes function importantly in cancer formation and progression, which involves alteration of epigenetic modifications in genome-wide (Davenport et al., 2021). Combined with transcriptome, six genes could be regarded as tumor suppressor genes (**Figure 5**). Among them, CTNNA3 was reported as a tumor suppressor in HCC before (He et al., 2016) and the same as DLG2 in osteosarcoma (Shao et al., 2019). The other four identified genes were KCNIP4, CACNA1C, PACRG, and ST6GALNAC3 (**Figure 5**). For KCNIP4, its related pathways were regulation of Wnt-mediated β-catenin signaling and target gene transcription (Kitagawa et al., 2007), in which the elevated Wnt-mediated β-catenin signaling could enhance the proliferation of liver cells in HCC (Wang et al., 2019). CACNA1C could be a prognostic predictor in ovarian cancer (Chang and Dong, 2021), and its overall survival was equally significant in HCC (**Figure 5B**). Abnormal promoter methylation of PACRG (Agirre et al., 2006), and ST6GALNAC3 (Haldrup et al., 2018) were associated with downregulation of gene expression in cancers. However, methylation was not limited to gene promoter. We found the unstable methylation sites in their gene body also changed markedly. Therefore, some of the unstable methylation genes may also serve as tumor suppressor genes, providing new ideas for mining tumor-related genes in future.

4 DISCUSSION

Compared with the well elucidated mechanism of 5mC on transcriptional regulation, other DNA methylation modification remains uncertain. It has been reported that 6mA is complementary to 5mC as an epigenomic mark in rice (Zhou et al., 2018). Thus, we analyzed Nanopore sequencing data to observe whether there is a clearly transcriptional regulatory mechanism for 5mC and 6mA. Our study found that 6mA had less influence on gene expression than 5mC. The Pearson correlation coefficient of the number of up and downregulated sites of 5mC was 0.16, and that of 6mA was 0.83. The distribution of 5mC up and downregulated sites was more inclined and the distribution of 6mA was more "uniform". There were 2,373 unstable methylation genes having both 5mC and 6mA, with up and downregulated methylation sites. The expressions of 1371 genes were different in tumor tissues and adjacent normal tissues. The statistics for 1371 genes did not prove obviously complementarity. It might be due to the species different. The number of genes with 5mC showed significant correlation with the number of differentially expressed genes, although such correlation was not linear. Meanwhile, 6mA had less effect on transcription, which requires further studies.

These 2,373 genes were regarded as unstable methylation genes, and their expressions were related to the number of up and downregulated methylation sites within them. In order to explore HCC related genes, we combined the transcriptome and survival data of TCGA liver cancer. Among the top 100 unstable methylation genes, we found eleven genes significantly affected the prognosis, and four of them can be defined as tumor suppressor genes. Normally, tumor suppressor genes can inhibit tumor cell proliferation and development. A typical tumor suppressor gene often occurs genetic alterations or epigenetic abnormality that reduce gene expression (Guo et al., 2010). While they are expressed at low level or inactivated in tumors, cell growth may lose control and facilitate tumor progression. Decreased expression of tumor suppressor gene also correlates with poor prognosis and reduced survival. Thus, the newly found tumor suppressor genes in this study can be prognostic predictor in HCC.

The top three unstable methylation genes in HCC are TBC1D3H, CSMD1, and ROBO2, which are closely related to the occurrence and development of HCC. TBC1D3H possesses TBC1 domain, and the proteins of this family were reported to regulate GTPase activation, which is related to tumorigenesis. CSMD1 is involved in such well-known tumor pathways as NF-κB pathway and epithelial-mesenchymal transition pathway. ROBO2 can suppress cancer development. Thus, genes with more unstable methylation sites in HCC are closely related to tumors. For the four newly discovered tumor suppressor genes, i.e., KCNIP4, CACNA1C, PACRG, and ST6GALNAC3, previous studies have proved their regulatory effect in tumors. These four genes were firstly considered to be tumor suppressor genes through methylation site screening. DNA methylation plays a critical interaction between tumor and immune cells. Tumor cells can escape immune restriction by various epigenetic mechanisms including DNA methylation (Cao and Yan, 2020). The top unstable methylation genes in HCC could be the pharmaceutical candidates like epigenetic regulators. The repair of aberrant methylation may trigger antitumor immune responses and further improve immunological surveillance. The targeting agents of unstable methylation genes will have major impacts in tumor-related treatment.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://ngdc.cncb.ac.cn/gsa-human/browse/HRA001037, HRA001037.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Beijing Hospital. The

patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LZ and HC conceived the study. LZ and WR designed the detail analysis pipeline. LZ and JM did the bioinformatics analysis. HL, XT, SX, LYW, and LW performed the experiments. LZ and FS wrote the manuscript and HC, QZ, BJ participated in revising the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by Transformation project of scientific and technological achievements in Qinghai Province (2020-SF-162), National Natural Science Foundation of China (Grants 81902618 & 81870048), Beijing Hospital Nova Project (No. BJ-2020-083) and Beijing Hospital Project (No. BJ-2019-153), Beijing Youth Talent Project (2018000032600G394).

ACKNOWLEDGMENTS

The authors would like to express their deep gratitude to the Clinical Biobank, Beijing Hospital for the biological sample collection, processing, storage, and information management.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2022.827391/full#supplementary-material

REFERENCES

Agirre, X., Román-Gómez, J., Vázquez, I., Jiménez-Velasco, A., Garate, L., Montiel-Duarte, C., et al. (2006). Abnormal Methylation of the commonPARK2andPACRGpromoter Is Associated with Downregulation of Gene Expression in Acute Lymphoblastic Leukemia and Chronic Myeloid Leukemia. *Int. J. Cancer* 118 (8), 1945–1953. doi:10.1002/ijc.21584

Andrews, S. (2010). Fastqc: A Quality Control Tool for High Throughput Sequence Data. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Accessed January 10, 2022).

Baubec, T., Colombo, D. F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A. R., et al. (2015). Genomic Profiling of DNA Methyltransferases Reveals a Role for DNMT3B in Genic Methylation. *Nature* 520 (7546), 243–247. doi:10.1038/nature14176

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30 (15), 2114–2120. doi:10.1093/bioinformatics/btu170

Cao, J., and Yan, Q. (2020). Cancer Epigenetics, Tumor Immunity, and Immunotherapy. *Trends Cancer* 6 (7), 580–592. doi:10.1016/j.trecan.2020.02.003

Chang, X., and Dong, Y. (2021). CACNA1C Is a Prognostic Predictor for Patients with Ovarian Cancer. *J. Ovarian Res.* 14 (1), 88. doi:10.1186/s13048-021-00830-z

Chen, X.-L., Hong, L.-L., Wang, K.-L., Liu, X., Wang, J.-L., Lei, L., et al. (2019). Deregulation of CSMD1 Targeted by microRNA-10b Drives Gastric Cancer Progression through the NF-Kb Pathway. *Int. J. Biol. Sci.* 15 (10), 2075–2086. doi:10.7150/ijbs.23802

Davenport, C. F., Scheithauer, T., Dunst, A., Bahr, F. S., Dorda, M., Wiehlmann, L., et al. (2021). Genome-Wide Methylation Mapping Using Nanopore Sequencing Technology Identifies Novel Tumor Suppressor Genes in Hepatocellular Carcinoma. *Ijms* 22 (8), 3937. doi:10.3390/ijms22083937

de Mendoza, A., Lister, R., and Bogdanovic, O. (2020). Evolution of DNA Methylome Diversity in Eukaryotes. *J. Mol. Biol.* 432, 1687–1705. doi:10.1016/j.jmb.2019.11.003

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* 29 (1), 15–21. doi:10.1093/bioinformatics/bts635

Goldman, M. J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., et al. (2020). Visualizing and Interpreting Cancer Genomics Data via the Xena Platform. *Nat. Biotechnol.* 38 (6), 675–678. doi:10.1038/s41587-020-0546-8

Guo, X., Liu, W., Pan, Y., Ni, P., Ji, J., Guo, L., et al. (2010). Homeobox Gene IRX1 Is a Tumor Suppressor Gene in Gastric Carcinoma. *Oncogene* 29 (27), 3908–3920. doi:10.1038/onc.2010.143

Haldrup, C., Pedersen, A. L., Øgaard, N., Strand, S. H., Høyer, S., Borre, M., et al. (2018). Biomarker Potential ofST6GALNAC3andZNF660promoter Hypermethylation in Prostate Cancer Tissue and Liquid Biopsies. *Mol. Oncol.* 12 (4), 545–560. doi:10.1002/1878-0261.12183

He, B., Li, T., Guan, L., Liu, F.-E., Chen, X.-M., Zhao, J., et al. (2016). CTNNA3 Is a Tumor Suppressor in Hepatocellular Carcinomas and Is Inhibited by miR-425. *Oncotarget* 7 (7), 8078–8089. doi:10.18632/oncotarget.6978

Hlady, R. A., Zhao, X., Pan, X., Yang, J. D., Ahmed, F., Antwi, S. O., et al. (2019). Genome-wide Discovery and Validation of Diagnostic DNA Methylation-Based Biomarkers for Hepatocellular Cancer Detection in Circulating Cell Free DNA. *Theranostics* 9 (24), 7239–7250. doi:10.7150/thno.35573

Itoh, T., Satoh, M., Kanno, E., and Fukuda, M. (2006). Screening for Target Rabs of TBC (Tre-2/Bub2/Cdc16) Domain-Containing Proteins Based on Their Rab-Binding Activity. *Genes Cells* 11 (9), 1023–1037. doi:10.1111/j.1365-2443.2006.00997.x

Jian, Z., Zhang, L., Jin, L., Lan, W., Zhang, W., and Gao, G. (2020). Rab5 Regulates the Proliferation, Migration and Invasion of Glioma Cells via Cyclin E. *Oncol. Lett.* 20 (2), 1055–1062. doi:10.3892/ol.2020.11660

Kaneda, A., Matsusaka, K., Sakai, E., and Funata, S. (2014). DNA Methylation Accumulation and its Predetermination of Future Cancer Phenotypes. *J. Biochem.* 156 (2), 63–72. doi:10.1093/jb/mvu038

Kawarada, L., Suzuki, T., Ohira, T., Hirata, S., Miyauchi, K., and Suzuki, T. (2017). ALKBH1 Is an RNA Dioxygenase Responsible for Cytoplasmic and Mitochondrial tRNA Modifications. *Nucleic Acids Res.* 45 (12), 7401–7415. doi:10.1093/nar/gkx354

Kitagawa, H., Ray, W. J., Glantschnig, H., Nantermet, P. V., Yu, Y., Leu, C.-T., et al. (2007). A Regulatory Circuit Mediating Convergence between Nurr1 Transcriptional Regulation and Wnt Signaling. *Mol. Cell Biol* 27 (21), 7486–7496. doi:10.1128/MCB.00409-07

Kole, C., Charalampakis, N., Tsakatikas, S., Vailas, M., Moris, D., Gkotsis, E., et al. (2020). Immunotherapy for Hepatocellular Carcinoma: A 2021 Update. *Cancers* 12 (10), 2859. doi:10.3390/cancers12102859

Li, H. (2018). Minimap2: Pairwise Alignment for Nucleotide Sequences. *Bioinformatics* 34 (18), 3094–3100. doi:10.1093/bioinformatics/bty191

Li, Z., Zhao, S., Nelakanti, R. V., Lin, K., Wu, T. P., Alderman, M. H., 3rd, et al. (2020). N6-methyladenine in DNA Antagonizes SATB1 in Early Development. *Nature* 583 (7817), 625–630. doi:10.1038/s41586-020-2500-9

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features. *Bioinformatics* 30 (7), 923–930. doi:10.1093/bioinformatics/btt656

Lowe, P., Olinski, R., and Ruzov, A. (2021). Evidence for Noncytosine Epigenetic DNA Modifications in Multicellular Eukaryotes: An Overview. *Methods Mol. Biol.* 2198, 15–25. doi:10.1007/978-1-0716-0876-0_2

Luo, G.-Z., Hao, Z., Luo, L., Shen, M., Sparvoli, D., Zheng, Y., et al. (2018). N6-methyldeoxyadenosine Directs Nucleosome Positioning in Tetrahymena DNA. *Genome Biol.* 19 (1), 200. doi:10.1186/s13059-018-1573-3

Mo, Z., Cao, Z., Luo, S., Chen, Y., and Zhang, S. (2020). Novel Molecular Subtypes Associated with 5mC Methylation and Their Role in Hepatocellular Carcinoma Immunotherapy. *Front. Mol. Biosci.* 7, 562441. doi:10.3389/fmolb.2020.562441

Morselli, M., Pastor, W. A., Montanini, B., Nee, K., Ferrari, R., Fu, K., et al. (2015). *In Vivo* targeting of De Novo DNA Methylation by Histone Modifications in Yeast and Mouse. *Elife* 4, e06205. doi:10.7554/eLife.06205

Pinho, A. V., Van Bulck, M., Chantrill, L., Arshi, M., Sklyarova, T., Herrmann, D., et al. (2018). ROBO2 Is a Stroma Suppressor Gene in the Pancreas and Acts via TGF-β Signalling. *Nat. Commun.* 9 (1), 5083. doi:10.1038/s41467-018-07497-z

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a Flexible Suite of Utilities for Comparing Genomic Features. *Bioinformatics* 26 (6), 841–842. doi:10.1093/bioinformatics/btq033

Ringehan, M., McKeating, J. A., and Protzer, U. (2017). Viral Hepatitis and Liver Cancer. *Phil. Trans. R. Soc. B* 372 (1732), 20160274. doi:10.1098/rstb.2016.0274

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* 26 (1), 139–140. doi:10.1093/bioinformatics/btp616

Sceusi, E. L., Loose, D. S., and Wray, C. J. (2011). Clinical Implications of DNA Methylation in Hepatocellular Carcinoma. *Hpb* 13 (6), 369–376. doi:10.1111/j.1477-2574.2011.00303.x

Shao, Y. W., Wood, G. A., Lu, J., Tang, Q.-L., Liu, J., Molyneux, S., et al. (2019). Cross-species Genomics Identifies DLG2 as a Tumor Suppressor in Osteosarcoma. *Oncogene* 38 (2), 291–298. doi:10.1038/s41388-018-0444-4

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

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A. Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

UniProt, C. (2021). UniProt: the Universal Protein Knowledgebase in 2021. *Nucleic Acids Res.* 49 (D1), D480–D489. doi:10.1093/nar/gkaa1100

Wang, W., Smits, R., Hao, H., and He, C. (2019). Wnt/β-Catenin Signaling in Liver Cancers. *Cancers* 11 (7), 926. doi:10.3390/cancers11070926

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., et al. (2021). clusterProfiler 4.0: A Universal Enrichment Tool for Interpreting Omics Data. *The Innovation* 2 (3), 100141. doi:10.1016/j.xinn.2021.100141

Xiao, C.-L., Zhu, S., He, M., ChenZhang, D., Zhang, Q., Chen, Y., et al. (2018). N6-Methyladenine DNA Modification in the Human Genome. *Mol. Cell* 71 (2), 306–318. doi:10.1016/j.molcel.2018.06.015

Yang, L., Chen, Z., Stout, E. S., Delerue, F., Ittner, L. M., Wilkins, M. R., et al. (2020). Methylation of a CGATA Element Inhibits Binding and Regulation by GATA-1. *Nat. Commun.* 11 (1), 2560. doi:10.1038/s41467-020-16388-1

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., et al. (2021). rMVP: A Memory-Efficient, Visualization-Enhanced, and Parallel-Accelerated Tool for Genome-wide Association Study. *Genomics, Proteomics & Bioinformatics*. doi:10.1016/j.gpb.2020.10.007

Yuen, Z. W., Srivastava, A., Daniel, R., McNevin, D., Jack, C., and Eyras, E. (2021). Systematic Benchmarking of Tools for CpG Methylation Detection from Nanopore Sequencing. *Nat. Commun.* 12 (1), 3438. doi:10.1038/s41467-021-23778-6

Zhang, Y., Su, H., Wudu, M., Ren, H., Xu, Y., Zhang, Q., et al. (2021). TBC1 Domain Family Member 23 Interacts with Ras-related Protein Rab-11A to Promote Poor Prognosis of Non-small-cell Lung Cancer via β1-integrin. *J. Cell Mol Med* 25 (18), 8821–8835. doi:10.1111/jcmm.16841

Zhao, Y., Yang, B., Chen, D., Zhou, X., Wang, M., Jiang, J., et al. (2021). Combined Identification of ARID1A, CSMD1, and SENP3 as Effective Prognostic Biomarkers for Hepatocellular Carcinoma. *Aging* 13 (3), 4696–4712. doi:10.18632/aging.202586

Zhou, C., Wang, C., Liu, H., Zhou, Q., Liu, Q., Guo, Y., et al. (2018). Identification and Analysis of Adenine N6-Methylation Sites in the rice Genome. *Nat. Plants* 4 (8), 554–563. doi:10.1038/s41477-018-0214-x

Zhuo, Z., Rong, W., Li, H., Li, Y., Luo, X., Liu, Y., et al. (2021). Long-read Sequencing Reveals the Structural Complexity of Genomic Integration of HBV DNA in Hepatocellular Carcinoma. *Npj Genom. Med.* 6 (1), 84. doi:10.1038/s41525-021-00245-1

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Cell and Developmental Biology

ORIGINAL RESEARCH
published: 25 March 2022
doi: 10.3389/fcell.2022.845641

Integrative Analysis of 5-Hydroxymethylcytosine and Transcriptional Profiling Identified 5hmC-Modified lncRNA Panel as Non-Invasive Biomarkers for Diagnosis and Prognosis of Pancreatic Cancer

Shuangquan Li[1†], *Yiran Wang*[1†], *Caiyun Wen*[1], *Mingxi Zhu*[2*], *Meihao Wang*[1*] and *Guoquan Cao*[1*]

[1]The First School of Medicine, School of Information and Engineering, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China, [2]Department of Anatomy, School of Basic Medicine and Life Science, Hainan Medical University, Haikou, China

OPEN ACCESS

Edited by:
Chunjie Jiang,
University of Pennsylvania,
United States

Reviewed by:
Jie Zhang,
Tongji University, China
Hao Lin,
University of Electronic Science and Technology of China, China

***Correspondence:**
Mingxi Zhu
hy0206175@hainmc.edu.cn
Meihao Wang
wzwmh@wmu.edu.cn
Guoquan Cao
caoguoquan@wmu.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular and Cellular Pathology, a section of the journal Frontiers in Cell and Developmental Biology

Received: 30 December 2021
Accepted: 07 February 2022
Published: 25 March 2022

Citation:
Li S, Wang Y, Wen C, Zhu M, Wang M and Cao G (2022) Integrative Analysis of 5-Hydroxymethylcytosine and Transcriptional Profiling Identified 5hmC-Modified lncRNA Panel as Non-Invasive Biomarkers for Diagnosis and Prognosis of Pancreatic Cancer. Front. Cell Dev. Biol. 10:845641. doi: 10.3389/fcell.2022.845641

Pancreatic adenocarcinoma (PAAD) is the fourth leading cause of cancer-related deaths worldwide. 5-Hydroxymethylcytosine (5hmC)-mediated epigenetic regulation has been reported to be involved in cancer pathobiology and has emerged to be promising biomarkers for cancer diagnosis and prognosis. However, 5hmC alterations at long non-coding RNA (lncRNA) genes and their clinical significance remained unknown. In this study, we performed the genome-wide investigation of lncRNA-associated plasma cfDNA 5hmC changes in PAAD by plotting 5hmC reads against lncRNA genes, and identified six PAAD-specific lncRNAs with abnormal 5hmC modifications compared with healthy individuals. Then we applied machine-learning and Cox regression approaches to develop predictive diagnostic (5hLRS) and prognostic (5hLPS) models using the 5hmC-modified lncRNAs. The 5hLRS demonstrated excellent performance in discriminating PAAD from healthy controls with an area under the curve (AUC) of 0.833 in the training cohort and 0.719 in the independent testing cohort. The 5hLPS also effectively divides PAAD patients into high-risk and low-risk groups with significantly different clinical outcomes in the training cohort (log-rank test p = 0.04) and independent testing cohort (log-rank test p = 0.0035). Functional analysis based on competitive endogenous RNA (ceRNA) and enrichment analysis suggested that these differentially regulated 5hmC modified lncRNAs were associated with angiogenesis, circulatory system process, leukocyte differentiation and metal ion homeostasis that are known important events in the development and progression of PAAD. In conclusion, our study indicated the potential clinical utility of 5hmC profiles at lncRNA loci as valuable biomarkers for non-invasive diagnosis and prognostication of cancers.

Keywords: 5-hydroxymethylcytosine, pancreatic cancer, machining learning, long non-coding RNA, non-invasive biomarker

INTRODUCTION

Pancreatic adenocarcinoma (PAAD) is the fourth leading cause of cancer-related deaths worldwide (Siegel et al., 2020). The lack of early-stage diagnostics has hindered the development of therapeutics that can slow down or reverse PAAD (Moutinho-Ribeiro et al., 2017; Sohal et al., 2017). Carbohydrate antigen 19–9 (CA19-9) is the biomarker currently used for PAAD diagnosis (Poruk et al., 2013). However, CA19-9 has a pooled sensitivity of 75.4% (95% CI: 73.4–77.4%) and a specificity of 77.6% (95% CI: 75.4–79.7%) for differentiation between malignant and non-malignant forms of cancer (Zhang et al., 2015). Moreover, the specificity of distinction between PAAD and CP often does not exceed 60% (Schultz et al., 2014), which has prompted a search for alternative biomarkers.

Circulating cell-free DNA (cfDNA) originates from cell death in different tissues, which has attracted massive interest as a non-invasive biomarker for cancer detection (Wan et al., 2017). Tumor cells release small nucleic acid fragments into the blood *via* multiple mechanisms, allowing cancer-associated genetic alterations to be detected (Diaz and Bardelli, 2014; Wan et al., 2017). Non-invasive biomarkers offer substantial advantages over tissue biopsy as their easily accessible characteristics make them ideal candidates for cancer diagnosis and progression monitoring (Xu et al., 2017).

5-methylcytosine (5mC) modifications potentially characterized various health conditions (Lehmann-Werman et al., 2016; Guo et al., 2017). Nonetheless, there has been no investigation to identify and sequence alternative modifications in circulating cfDNA because the DNA samples are low-input. 5-Hydroxymethylcytosine (5hmC) is a novel identified epigenetic mark generated from 5mC by the ten-eleven translocation proteins (Tahiliani et al., 2009). Increasing evidence showed that low levels of 5hmC are observed in many tumors frequently compared to corresponding normal tissues (Jin et al., 2011). 5hmC is a stable intermediate of cytosine demethylation. Active gene modification was associated with the levels of 5hmC accumulation in promoters, gene bodies and gene regulatory elements (Han et al., 2016). 5hmC modifications play a curial role in cell development, differentiation, maturation and self-renewal (Wang et al., 2014). These characteristics suggested 5hmC signaling changes in cfDNA may have potential values in cancer diagnosis and progression monitoring using highly robust and sensitive 5hmC sequencing technologies (Song et al., 2017).

Long non-coding RNA (lncRNA) expression is the most pervasive transcriptional change in cancer, which is demonstrated by the recent genome-wide characterization of the human cancer transcriptome (Marchese et al., 2017; Zhou et al., 2019a; Bao et al., 2020; Zhou et al., 2021b). Multiple studies have shown that lncRNAs play a critical role in the mechanism of occurrence, evolution, invasion and metastasis of pancreatic cancer (Fu et al., 2017; Zhang et al., 2020; Tang et al., 2021). Several studies have indicated the impact of lncRNAs on the prognosis of pancreatic cancer (Huang et al., 2016; Zhou et al., 2016). There is increasing evidence that lncRNA expression could also be regulated by epigenetic DNA modifications, such as DNA methylation, histone modifications and 5hmC modification (Guttman et al., 2009; White et al., 2014; Yan et al., 2015). Epigenetic DNA modifications in lncRNA genes have been revealed to be valuable non-invasive biomarkers for cancer diagnosis, prognosis and surveillance (Hu et al., 2017; Zhou et al., 2021a).

In the present study, we performed machine learning-based integrative analysis of 5-hydroxymethylcytosine and transcriptional profiling to identify plasma-derived lncRNAs with PAAD-specific abnormal 5hmC modifications and conducted a two-phase discovery-validation experiment to explore the clinical potential of 5hmC-modified lncRNAs as non-invasive biomarkers for diagnosis and prognosis of pancreatic cancer.

MATERIALS AND METHODS

Sample Datasets

A total of 130 publicly available genome-wide nano-hmC-Seal profiles from plasma cfDNA samples were used in this study, including 34 PAAD patients and 96 healthy individuals from Sequence Read Archive (SRA, SRP080977) of the National Center for Biotechnology Information (NCBI) (Li et al., 2017). Clinical and transcriptomic data (RNA-Seq, v.2019-07-20) of 182 PAAD cases, including 176 primary tumors and four solid tissue normal, generated by The Cancer Genome Atlas (TCGA) were downloaded from UCSC Xena (https://xena.ucsc.edu/). In addition, stem-loop expression (miRNA, v.2019.7.20) of 183 PAAD cases, including 1881 miRNAs, was also obtained from UCSC Xena. Detailed information on the study population is shown in **Table 1**. The workflow diagram of the study design is shown in **Figure 1**

Data Preprocessing and Mapping of 5hmC-Modified lncRNAs

Read sequences were extracted in FASTQ files using the SRA toolkit (https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view= software, version 2.9.2), and then were aligned to the human genome GRCh37 using Bowtie2 (version 2.3.4.2) with default parameters (Langmead and Salzberg, 2012). To convert and sort the alignment SAM files into BAM files, SAMtools (version 1.9) was used to generate the files (Li et al., 2009). The picard-2.18.4 was used to retain unique non-duplicate matches to the genome (http://broadinstitute.github.io/picard/). The released version of the lncRNA reference gene annotation file (GRCh38 version 34) was downloaded from the GENCODE database (https://www.gencodegenes.org/). LiftOver was used to transfer the mapping information from the GRCh38 version of the lncRNA reference gene annotation file to the GRCh37 version. Genes encoding lncRNAs were extracted based on GRCh37 annotation. Read counts of 5hmC-modified lncRNAs were calculated using the fragment counts in each RefSeq lncRNA obtained by BEDtools (version 2.27.1) (Quinlan and Hall, 2010). The read counts were converted into Transcripts Per kilobase of 5hmC in lncRNA per

TABLE 1 | Baseline clinical and pathological characteristics of the study population.

Cohorts	Sample type	Number (%)	Platform	Data source
Li's cohort	Pancreatic Cancer	130	nano-hmC-Seal-seq	SRP080977
		34		
	Healthy	96		
TCGA-PAAD		180	RNA-seq	TCGA
	Primary Tumor	176		
	Solid Tissue Normal	4		

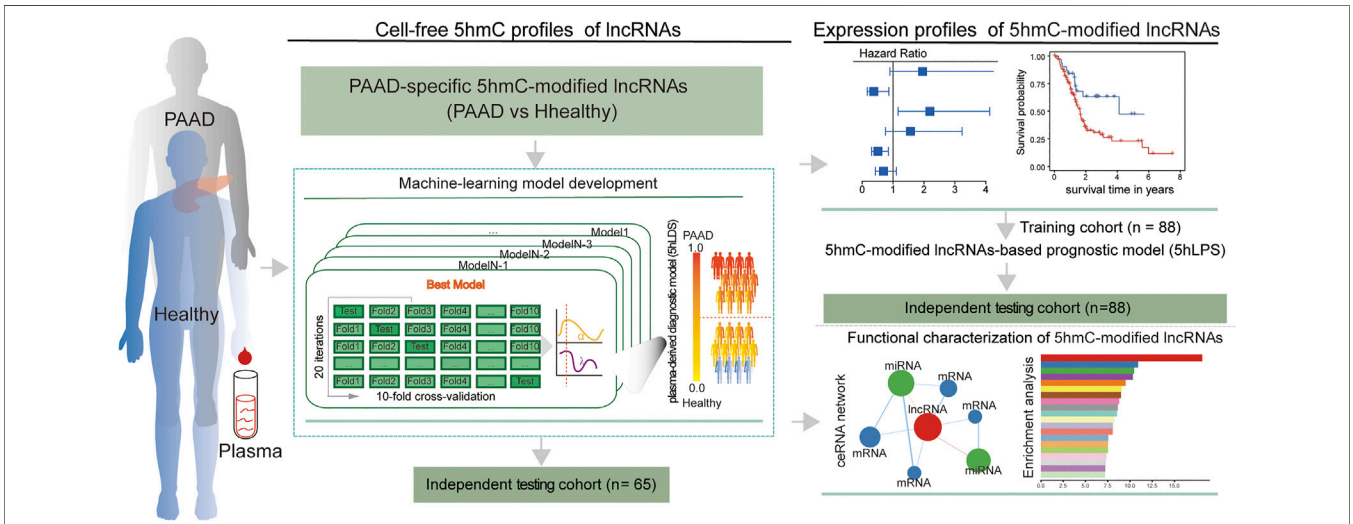


FIGURE 1 | Workflow diagram of the study design. A two-phase discovery-validation study was conducted. The 5hmC-modified lncRNA-based predictive and prognostic models for PAAD were developed using the machine-learning methods in the training cohort and validated in the independent testing cohort. PAAD, Pancreatic adenocarcinoma; lncRNA, long non-coding RNAs; 5hmC, 5-Hydroxymethylcytosine.

million mapped reads. Finally, 5hmC profiles of 16,827 lncRNAs were obtained for further analysis.

Machining Learning-Based Establishment of a Non-Invasive Diagnostic Model Based on 5hmc Modified lncRNAs

The 5hmC profiles of lncRNAs were compared between PAAD and healthy control samples. The lncRNAs with differential 5hmC modification were identified using the DESeq2 package (version 1.22.2) (Love et al., 2014) with a |log2foldchange|>0.58 and false discovery rate adjusted $p < 0.05$, and were selected as PAAD-specific 5hmC-modified lncRNAs. Then a 5hmC-modified lncRNA-based risk scoring model (termed 5hLDS) was constructed using the elastic net regularization on a multivariable logistic regression approaches to distinguish between PAAD and healthy individuals. The 5hLDS was trained with 10-fold cross-validation and optimized using a ROC curve for a grid of parameter values for α and λ (α range, 0.05 to 1.00 with a length = 10; λ range: from 10–1 to 5*10–1 with a 0.1 increment), and this selection process was repeated 20 times. Finally, the 5hLDS was established based on PAAD-specific 5hmC-modified lncRNA markers. The 5hLDS range was 0–1.0 and represented a final probability of PAAD for each sample. The 5hLDS was established using the "trainControl" and "train" functions from the Caret (version 6.0–86) R package.

Functional Analysis of PAAD-Specific 5hmC-Modified lncRNA Markers

Pearson correlation coefficient was used to measure the expression relevance among PAAD-specific 5hmC-modified lncRNA markers, miRNAs and mRNAs in the TCGA cohort. Human experimentally validated miRNA-mRNA interactions were downloaded from ENCORI (https://starbase.sysu.edu.cn/index.php), the updated version of the StarBase database providing the most comprehensive network of miRNA-mRNA interactions supported by CLIP-Seq data sets (Li et al., 2014). The collected data contain 9,664 experimentally validated miRNA-mRNA interactions, including 276 miRNAs and 14,837 mRNAs. Finally, PAAD-specific 5hmC-modified lncRNA marker-related competitive endogenous RNA (ceRNA) networks were constructed based on ceRNA mechanism as follows: 1) there was a significantly high negative co-expression relationship ($p < 0.05$ and $r < -0.2$)

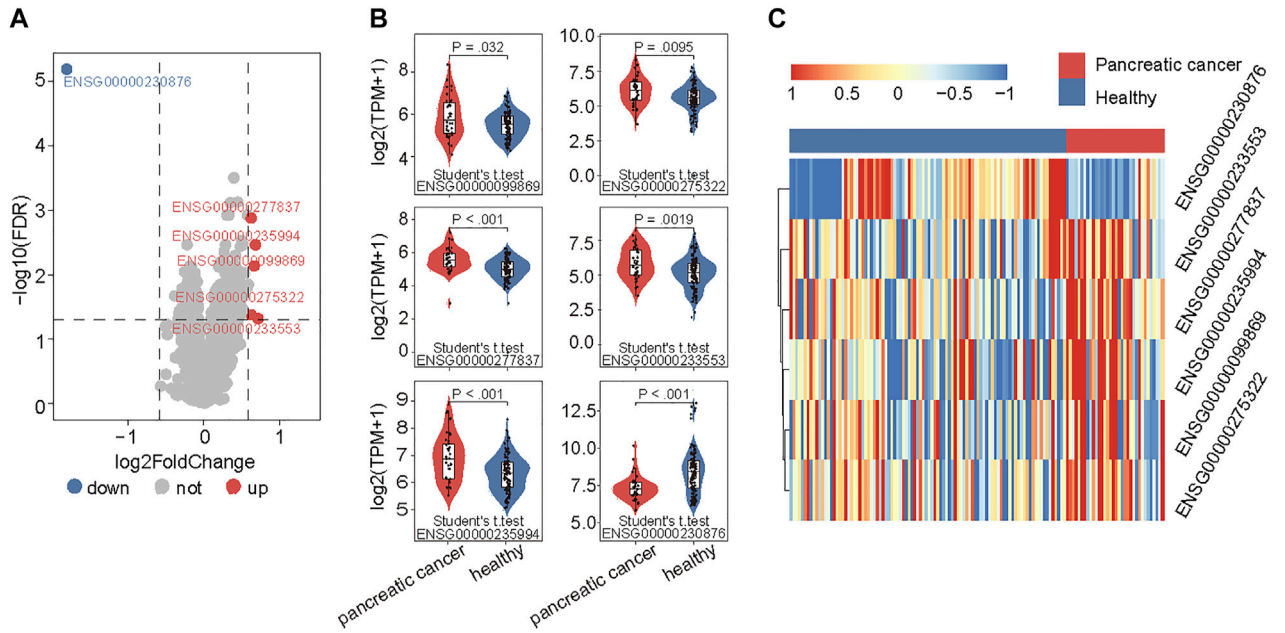


FIGURE 2 | Identification of PAAD-specific 5hmC-modified lncRNAs. **(A)** Volcano plot visualizing the intersections of 5hmC-modified up- and down-enriched lncRNAs in PAAD samples compared with healthy controls. **(B)** Boxplots showing 5hmC modification levels of lncRNAs between PAAD and healthy samples. Statistical significance was determined using the Student's t-test. **(C)** Heatmap of unsupervised hierarchical clustering of PAAD-specific 5hmC modification levels in lncRNAs. PAAD, Pancreatic adenocarcinoma; lncRNA, long non-coding RNAs; 5hmC, 5-Hydroxymethylcytosine.

between miRNA and mRNA, and between miRNA and lncRNAs; 2) there was significantly high positive co-expression relationship ($p < 0.05$ and $r > 0.4$) between mRNA and lncRNAs; 3) there were experimentally validated miRNA-mRNA interactions; This ceRNA network was visualized using the Cytoscape software (version 3.8.2). Functional enrichment analysis of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) for mRNAs in the ceRNA network was performed with Metascape (Zhou et al., 2019b).

Statistical Analysis

A diagnostic model was developed using the machining learning approach, and a prognostic model was developed using the Cox regression model. The diagnostic model was evaluated with cross-validation and ROC methods. The conventional univariate Cox proportional hazards regression model for overall survival data was implemented to identify variables associated with overall survival. Significant factors in univariate analysis were further subjected to a multivariate Cox regression analysis. Hazard ratios (HR) and corresponding 95% confidence interval (CI) were calculated in the Cox models. The optimal risk cut-off value of the prognostic model was calculated using the R package "survminer". The PAAD patients were stratified into high- and low-risk groups according to the risk cut-off. Kaplan-Meier curve analysis and Log-rank test were performed to visualize and compare the survival difference between the two risk groups. A Multivariate Cox regression model was used to examine whether the prognostic model remained significant after adjusting clinical variables. Statistical significance was performed for categorical variables using the Wilcoxon signed-rank test for two-group comparisons unless otherwise specified in the figure legend. All statistical analysis was conducted in R software, version 3.6.1.

RESULTS

Identification of Altered Plasma 5hmC Modifications in lncRNAs Genes Involved in PAAD

To examine the 5hmC modification of lncRNAs genes in PAAD, we compared the 5hmC profiles of lncRNAs between PAAD ($n = 34$) and healthy ($n = 96$) samples and found a total of six lncRNAs that showed differential 5hmC modification patterns (**Figure 2A**). Among six lncRNAs, five lncRNAs (*RP11-262A16.1*, *IGF2-AS1*, *AC108462.1*, *RP11-714M23.2* and *RP3-470B24.5*) showed increasing 5hmC modification and one lncRNAs (*LINC00486*) showed decreasing 5hmC modification in PAAD compared to healthy samples (**Figure 2B**). These six

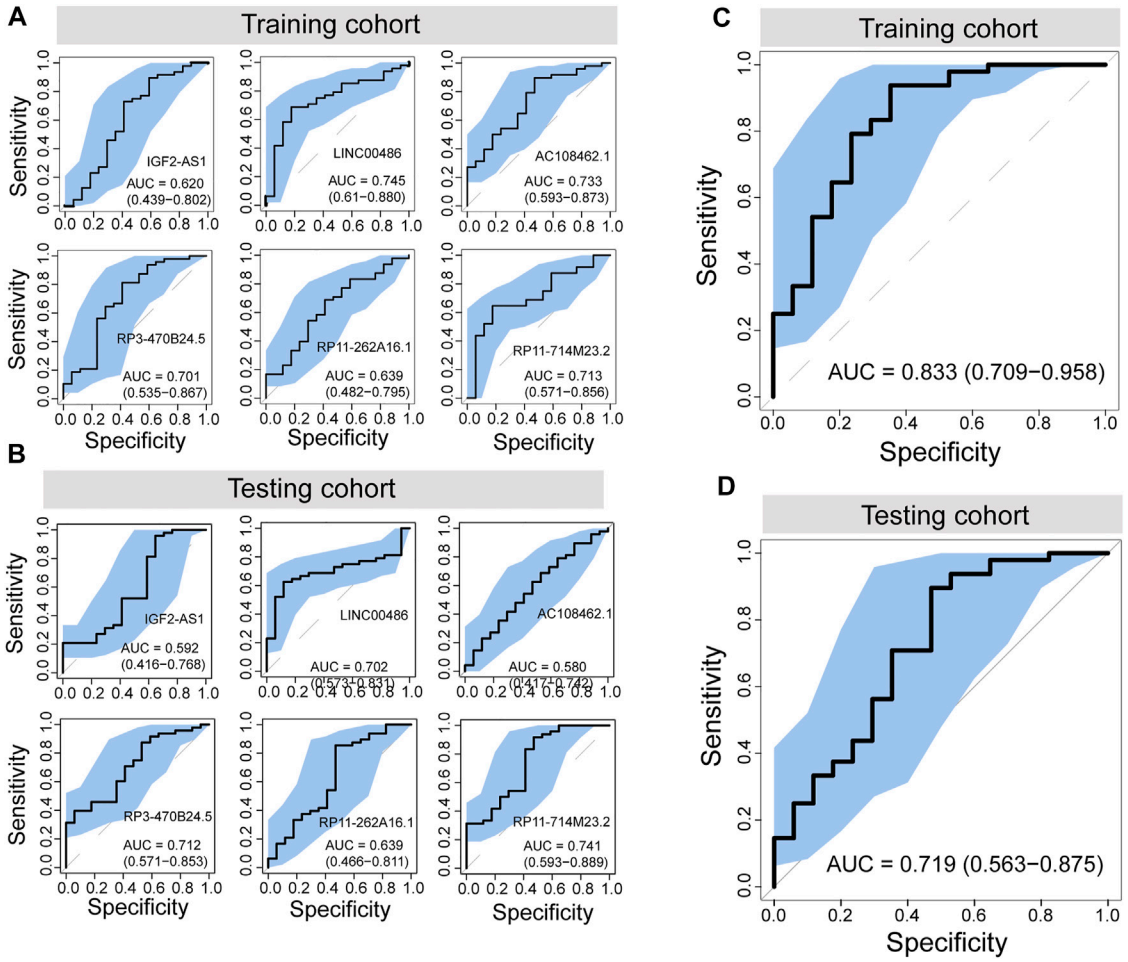


FIGURE 3 | Development and independent validation of a plasma-derived diagnostic model (5hLDS) based on 5hmC-modified lncRNAs. ROC curve and AUC value of each PAAD-specific 5hmC-modified lncRNA for cancer diagnosis in the training cohort **(A)** and testing cohort **(B)**. ROC curve and AUC value of the 5hLDS for cancer diagnosis in the training cohort **(C)** and testing cohort **(D)**. AUC, Area under the ROC curve; PAAD, Pancreatic adenocarcinoma; p5hmC-score, PAAD 5hmC-LncRNA diagnosis score. PAAD, pancreatic adenocarcinoma.

lncRNAs with abnormal 5hmC modifications were defined as 5hmC-modified lncRNAs. As shown in **Figure 2C**, the 5hmC levels of six 5hmC-modified lncRNAs were significantly different and revealed clinical potential as diagnostic biomarkers in distinguishing PAAD from healthy controls (**Figure 2C**).

Development and Validation of a Plasma-Derived Diagnostic Model Based on 5hmC-Modified lncRNAs

To further evaluate the potential of differentially regulated 5hmC modified lncRNAs as diagnostic biomarkers for PAAD, we conducted a discovery-validation study, in which 130 plasma cfDNA samples were split randomly into equally sized training cohort and testing cohort, respectively. We estimated the contribution of each differentially regulated 5hmC modified lncRNAs to disease diagnosis using the 5hmC-modified levels as the contributed score in the training and testing cohorts. As shown in **Figures 3A,B**, each of the differentially regulated 5hmC modified lncRNAs exhibited a diagnostic performance AUC of 0.62–0.745 in the training cohort and 0.58 to 0.741, respectively.

Then, we developed a machining learning-based non-invasive diagnostic model based on six 5hmC modified lncRNAs using the elastic net regularization on a multivariable logistic regression approach (termed 5hLDS) in the training cohort. Using the ten-cross-validation, the 5hLDS could discriminate PAAD from healthy individuals with an overall AUC of 0.833 (95% CI 0.709-0.958) (**Figure 3C**). Then the 5hLDS was validated in the independent testing cohort and yielded a classification performance AUC of 0.719 (95% CI 0.563-0.875) for PAAD detection (**Figure 3D**). At a disease risk cut-off of 0.5, the 5hLDS achieved 81.54% (95% CI 69.97%-90.08%) and 78.46% (95% CI 66.51%-87.69%) of accuracy to distinguish PAAD patients from healthy individuals in the training cohort and testing cohort, respectively. These results demonstrated AUC improvement

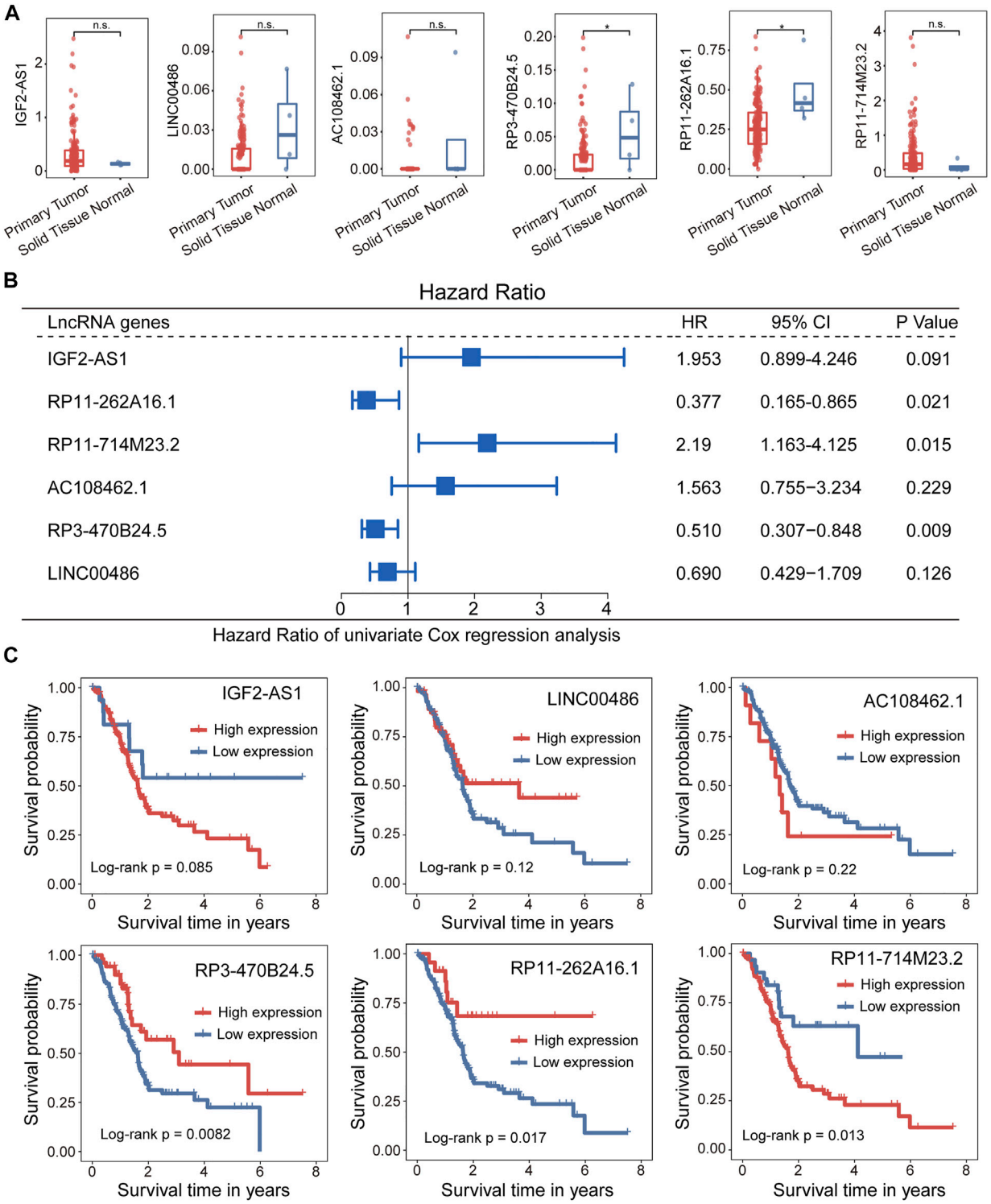


FIGURE 4 | Association between the 5hLD-score and disease outcome. **(A)** Boxplots showing the distribution of expression levels of lncRNAs between PAAD patients and healthy controls. Statistical significance was determined using Wilcoxon signed-rank test. **(B)** Metaplot of univariate analysis of 5hmC-modified lncRNA between high-expression and low-expression group. **(C)**. Kaplan-Meier survival curves of each PAAD-specific lncRNA between high-expression and low-expression groups. PAAD, pancreatic adenocarcinoma.

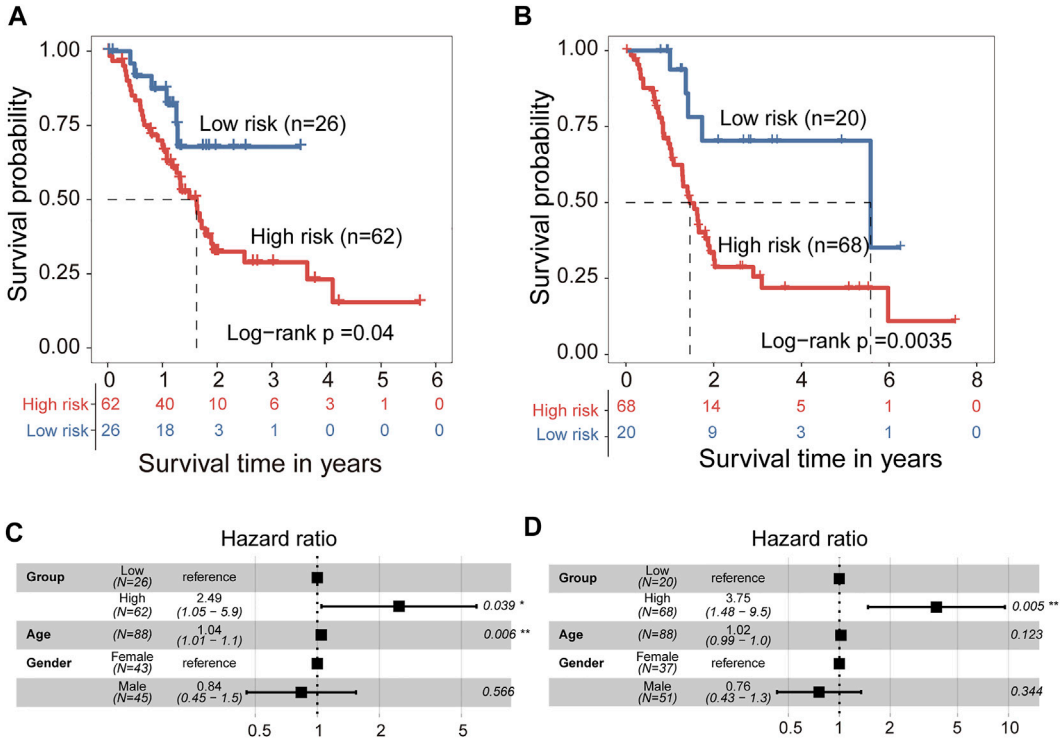


FIGURE 5 | Development and independent validation of 5hmC-modified lncRNAs-based prognostic prediction model (5hLPS). Kaplan-Meier survival curves of patients between the high-risk and low-risk groups in training cohort **(A)** and testing cohort **(B)**. Forest plot of HR deriving from multivariate Cox regression analysis of 5hLPS with other clinical characteristics in training cohort **(C)** and testing cohort **(D)**. HR, Hazard ratios.

when incorporated six differentially regulated 5hmC modified lncRNAs into the diagnostic model.

Association Between 5hmC-Modified lncRNAs and Prognosis

We further examined whether expression levels of 5hmC-modified lncRNAs were dysregulated in PAAD compared to controls. Expression profiles of 5hmC-modified lncRNAs were obtained from TCGA and were compared between PAAD primary tumors and normal solid tissues. Two of six 5hmC-modified lncRNAs exhibited differential expression patterns between PAAD primary tumors and normal solid tissues. As shown in **Figure 4A**, lncRNAs RP3-470B24.5 and RP11-262A16.1 revealed significant or marginally significant downregulated expression levels in PAAD compared to normal solid tissues. Furthermore, three of six 5hmC-modified lncRNAs showed significant association correlation between expression levels and OS, including *RP11-262A16.1* (HR = 0.377, 95% CI 0.165–0.865, p = 0.021), *RP11-714M23.2* (HR = 2.190, 95% CI 1.163–4.125, p = 0.015) and *RP3-470B24.5* (HR = 0.51, 95% CI 0.307–0.848, p = 0.009) (**Figure 4B**). Kaplan-Meier survival analysis showed that PAAD patients with high expression of *RP3-470B24.5* and *RP11-262A16.1* have significant improved OS compared to those with low expression (log-rank test p < 0.0082 for *RP3-470B24.5* and p < 0.017 for *RP11-262A16.1*), while PAAD patients with high expression of *RP11-714M23.2* tended to have poor OS compared to those with low expression (log-rank test p < 0.013) (**Figure 4C**). These results demonstrated the potential of three 5hmC-modified lncRNAs as prognostic biomarkers.

5hmC-Modified lncRNAs-Based Prognostic Prediction Model for PAAD

We developed a 5hmC-modified lncRNAs-based prognostic risk score model (5hLPS) for the prognostication of PAAD using the linear combination of the expression of three 5hmC-modified lncRNAs and weighted by relative coefficients in the multivariate Cox regression as follows: 5hLPS = (-0.24378700) × expression value of *RP11-262A16.1* + 0.02015193 × expression value of *RP11-714M23.2* + (-1.97034856) × expression value of *RP3-470B24.5*. Then we also conducted a discovery-validation study by splitting PAAD samples from TCGA into equally sized training (n = 88) and testing (n = 88) cohort. The optimal risk cut-off value of the 5hLPS stratified 88 patients of the training cohort into the high-risk group (n = 62) and low-risk group (n = 26) with significantly different OS. As shown in **Figure 5A**, patients in the high-risk group had significantly shorter OS time than those in the low-risk group (log-rank test p = 0.04). When validated in the testing cohort, the 5hLPS separated the 88 patients of the testing cohort into high-risk (n = 68) and low-

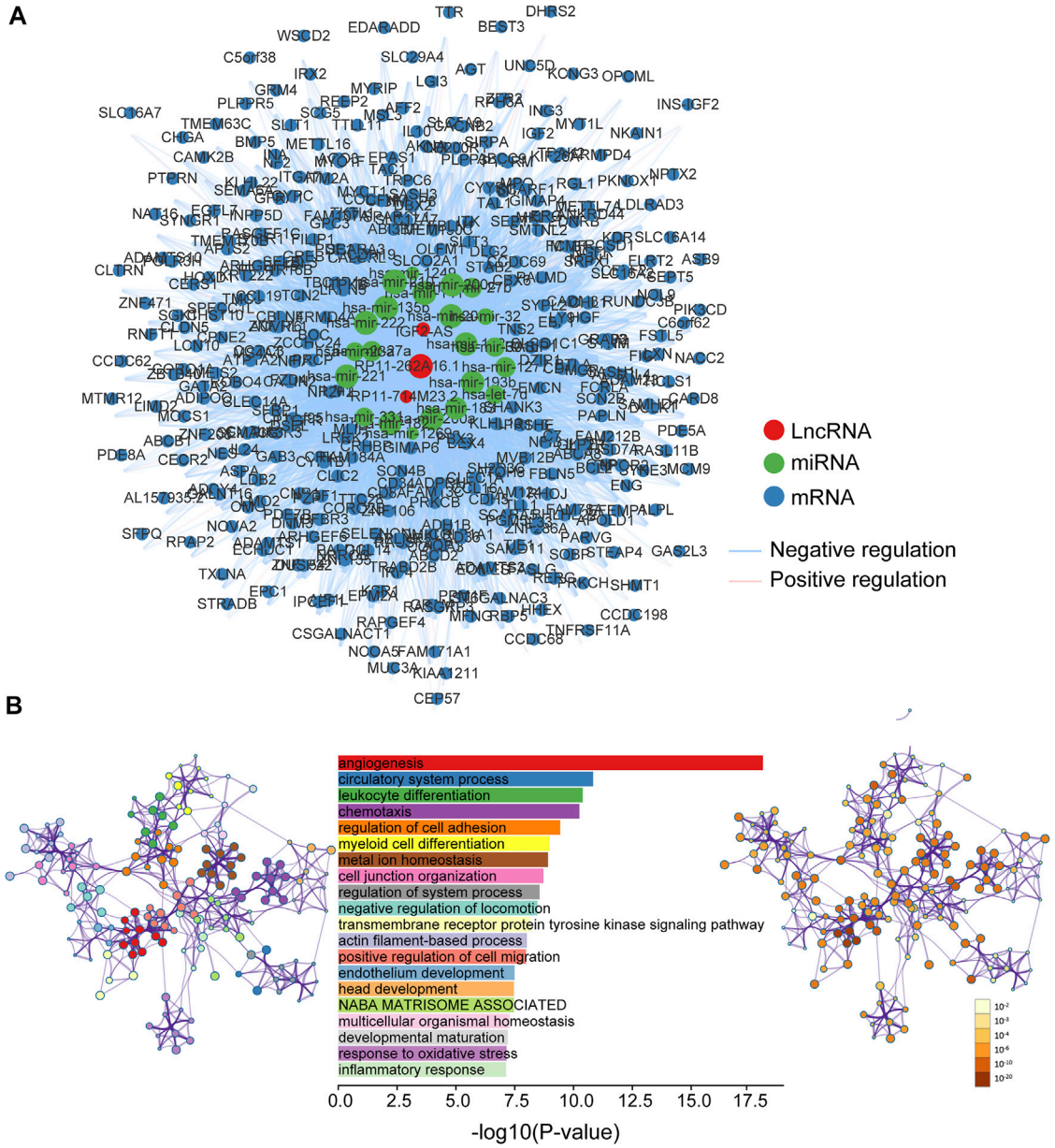


FIGURE 6 | Functional characterization of PAAD-specific 5hmC-modified lncRNAs. **(A)** PAAD-specific 5hmC-modified lncRNA-associated ceRNA network. Blue balls represent mRNAs; red balls represent lncRNA; green balls represent miRNA. Blue edges indicate lncRNA-miRNA-mRNA negative interactions; pink edges indicate lncRNA-miRNA-mRNA positive interactions. The size of the shapes represents the degrees that are involved. The bigger the shape, the higher the degree. **(B)** Bar graph of enriched terms across mRNA gene lists, length by -log10(P-value). Network graph of the enriched term (left pannel) colored by cluster-ID, where nodes that share the same cluster-ID are typically close to each other; and enriched *p*-value (right panel) colored by *p*-value, where terms containing more genes tend to have a more significant *p*-value. miRNA, microRNA; lncRNA, long non-coding RNA; ceRNA, competitive endogenous RNA; PAAD, pancreatic adenocarcinoma.

risk groups ($n = 20$). The median survival time in the low-risk group was significantly better than that in the high-risk group (log-rank test $p = 0.0035$) (**Figure 5B**). The univariate Cox analysis also showed that the 5hLPS has a significant association with OS both in the training (HR = 2.4, 95% CI: 1.0–5.7, $p < 0.05$) and testing (HR = 3.6, 95% CI: 1.4–9.2, $p < 0.01$) cohorts. To further examine whether the 5hLPS was independent of other clinical and pathological factors. We performed multivariable Cox proportional hazards analysis, including individual clinical variables with the 5hLPS in each cohort. As shown in **Figure 5C**, in the training cohort, the 5hLPS (HR = 2.49, 95% CI 1.05–5.9, $p < 0.039$) and age (HR = 1.04, 95% CI 1.01–1.1, $p = 0.006$) were significantly associated with OS in the multivariate analysis. In the testing cohort, only the 5hLPS maintained a significant association with OS (HR = 3.75, 95% CI: 1.48–9.5, $p = 0.005$) (**Figure 5D**). These

results suggested that 5hLPS is an independent predictive factor for patients with PAAD.

Functional Characterization of 5hmC-Modified lncRNA Markers

To explore the functional roles of 5hmC-modified lncRNAs, we constructed a 5hmC-modified lncRNA marker-associated ceRNA network which included 5,604 interactions among 358 mRNAs, three lncRNAs and 22 miRNAs (**Supplementary File S1**), as shown in **Figure 6A**. Then we performed pathway functional enrichment analysis for 358 mRNAs in this ceRNA *via* Metascape to infer potential functional roles of 5hmC-modified lncRNA markers. As shown in **Figure 6B**, 5hmC-modified lncRNA marker-associated mRNAs were significantly enriched in angiogenesis, circulatory system process, leukocyte differentiation, and metal ion homeostasis, which are known important events in the development and progression of the development and progression of PAAD.

DISCUSSION

Pancreatic adenocarcinoma (PAAD) has an inferior prognosis and remains a lethal malignancy (Schizas et al., 2020). Recent studies suggest that aberrant expression of lncRNA drives the initiation and progression of malignancies (Bhan et al., 2017). Liquid biopsies provide a non-invasive approach to detect tumors (Shen et al., 2018) and a novel Nano-hmC-Seal technology to generate the genome-wide profiles of 5hmC in cfDNA from blood plasma for multiple cancer types (Li et al., 2017). Whether plasma-derived 5hmC-modified lncRNA is a conclusive biomarker for distinguishing the type of cancer and diagnosing cancer is unclear. In this study, we explored the potential application of the plasma-derived 5hmC modification level in lncRNA being used as an alternative biomarker for PAAD diagnosis and monitoring.

In this study, by redefining 5hmC sequencing reads to lncRNA genes, 5hmC alterations of lncRNAs were characterized in PAAD. Several altered 5hmC modifications were distributed at lncRNAs in patients with PAAD compared with healthy subjects. Herein, these differentially regulated 5hmC modified lncRNAs were considered as PAAD-specific markers. We trained machine-learning algorithms with 10-fold cross-validation using the differentially regulated 5hmC modified lncRNAs as features (termed 5hLDS), and evaluated the prediction performance in the training and testing cohorts. The 5hLDS achieved superior diagnostic performance in distinguishing PAAD from healthy controls both in the training and testing cohorts.

Further exploring of expression relevance of the 5hmC-modified lncRNAs in PAAD identified two of six 5hmC-modified lncRNAs that were dysregulated expression in PAAD compared to normal tissues, suggesting that these 5hmC modifications changes might lead to dysregulated lncRNAs expression that involved in the development of PAAD. Furthermore, we further investigated the effect of expression of 5hmC-modified lncRNAs on survival and found that 5hmC-modified lncRNAs were also able to make a distinction between each PAAD patient into distinct groups with better or worse survival outcomes. These results also support the idea that 5hmC-modified lncRNAs can serve as potential biomarkers for the prognosis of PAAD patients. Therefore, we trained a regression model with Cox analysis using the differentially regulated 5hmC modified lncRNAs as features (termed 5hLPS), and evaluated the prognostic performance in the training and testing cohorts. The 5hLPS exhibited superior performance in classifying the patients into two groups with significantly different overall survival independent of clinical variables.

Although many lncRNAs have been discovered and recorded in a vast amount of literature, only a few have been well-functionally studied and characterized. It has been reported that lncRNAs can act as ceRNA for miRNA and emerge as important regulators involved in diverse biological and physiopathological contexts. Among six 5hmC-modified lncRNAs identified in this study, lncRNA *IGF2-AS1* is expressed in antisense to the insulin-like growth factor 2 (IGF2) gene and is imprinted and paternally expressed (Vu et al., 2003). Many studies suggested that transcripts from *IGF2-AS1* are produced in tumors and may suppress cell growth (Unger et al., 2017; Kasprzak and Adamek, 2019; Contractor et al., 2020). Another 5hmC-modified lncRNAs, *LINC00486*, has been reported to be a hot spot of breakpoints and has a high rearrangement rate in non-small cell lung cancer cells (Wang et al., 2018). A recent study also found that *LINC00486* may act as a tumor suppressor gene and its overexpression can inhibit the proliferation and promote the apoptosis of breast cancer tissues by suppressing miR-182–5p expression (Yuan et al., 2020). To further gain a novel functional insight for 5hmC-modified lncRNAs, we construed a 5hmC-modified lncRNAs-associated ceRNA network according to the ceRNA hypothesis and found that mRNA in this ceRNA network were involved in the known crucial event in the development and progression of PAAD, such as angiogenesis, circulatory system process, leukocyte differentiation and metal ion homeostasis.

The present study still had its limitations. First, this study did not obtain significant clinical variables of 5hmC-modified lncRNAs, such as follow-up time. Further independent validation studies should be performed to help address problems such as the potential selection bias for model construction. Secondly, the regulatory mechanism of 5hmC and lncRNA genes remains unclear due to the lack of paired 5hmC profiles and lncRNAs expression profiles. Finally, although the 5hmC-modified lncRNAs have been validated in expression levels from the TCGA cohort, further validation is required in other retrospective or prospective cohorts to demonstrate the generalizability of 5hmC-modified lncRNA for PAAD diagnosis and prognosis.

CONCLUSION

This study characterized the genome-wide pattern of lncRNA-associated plasma cfDNA 5hmC changes in PAAD, and demonstrated potential roles for 5hmC in the transcriptional regulation of lncRNAs contributing to the development and progression of PAAD. Finally, we identified and validated 5hmC-modified lncRNA panels for the diagnosis and prognosis of PAAD with high and robust performance, which presented potential clinical utility of 5hmC-modified lncRNAs as valuable biomarkers for non-invasive diagnosis and prognostication of cancers.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The datasets of this article are available in the Sequence Read Archive repository (https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP080977) and UCSC Xena (https://xena.ucsc.edu/).

AUTHOR CONTRIBUTIONS

GC, MW, and MZ designed the study; SL and YW developed the methodology; SL, YW, and CW performed data analysis. SL and GC drafted the manuscript. All authors read and approved the final manuscript.

FUNDING

This study was supported by the Key Laboratory of Intelligent Medical Imaging of Wenzhou (No. 2021HZSY0057).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2022.845641/full#supplementary-material

REFERENCES

Bao, S., Zhao, H., Yuan, J., Fan, D., Zhang, Z., Su, J., et al. (2020). Computational Identification of Mutator-Derived lncRNA Signatures of Genome Instability for Improving the Clinical Outcome of Cancers: a Case Study in Breast Cancer. *Brief Bioinform* 21, 1742–1755. doi:10.1093/bib/bbz118

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

Contractor, T., Clausen, R., Harris, G. R., Rosenfeld, J. A., Carpizo, D. R., Tang, L., et al. (2020). IGF2 Drives Formation of Ileal Neuroendocrine Tumors in Patients and Mice. *Endocr. Relat. Cancer* 27, 175–186. doi:10.1530/erc-19-0505

Diaz, L. A., Jr., and Bardelli, A. (2014). Liquid Biopsies: Genotyping Circulating Tumor DNA. *Jco* 32, 579–586. doi:10.1200/jco.2012.45.2011

E. Poruk, K., Z. Gay, D., Brown, K., D. Mulvihill, J., M. Boucher, K., L. Scaife, C., et al. (2013). The Clinical Utility of CA 19-9 in Pancreatic Adenocarcinoma: Diagnostic and Prognostic Updates. *Cmm* 13, 340–351. doi:10.2174/1566524011313030003

Fu, Z., Chen, C., Zhou, Q., Wang, Y., Zhao, Y., Zhao, X., et al. (2017). LncRNA HOTTIP Modulates Cancer Stem Cell Properties in Human Pancreatic Cancer by Regulating HOXA9. *Cancer Lett.* 410, 68–81. doi:10.1016/j.canlet.2017.09.019

Guo, S., Diep, D., Plongthongkum, N., Fung, H.-L., Zhang, K., and Zhang, K. (2017). Identification of Methylation Haplotype Blocks Aids in Deconvolution of Heterogeneous Tissue Samples and Tumor Tissue-Of-Origin Mapping from Plasma DNA. *Nat. Genet.* 49, 635–642. doi:10.1038/ng.3805

Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., et al. (2009). Chromatin Signature Reveals over a Thousand Highly Conserved Large Non-coding RNAs in Mammals. *Nature* 458, 223–227. doi:10.1038/nature07672

Han, D., Lu, X., Shih, A. H., Nie, J., You, Q., Xu, M. M., et al. (2016). A Highly Sensitive and Robust Method for Genome-wide 5hmC Profiling of Rare Cell Populations. *Mol. Cel* 63, 711–719. doi:10.1016/j.molcel.2016.06.028

Hu, H., Shu, M., He, L., Yu, X., Liu, X., Lu, Y., et al. (2017). Epigenomic Landscape of 5-hydroxymethylcytosine Reveals its Transcriptional Regulation of lncRNAs in Colorectal Cancer. *Br. J. Cancer* 116, 658–668. doi:10.1038/bjc.2016.457

Huang, X., Zhi, X., Gao, Y., Ta, N., Jiang, H., and Zheng, J. (2016). LncRNAs in Pancreatic Cancer. *Oncotarget* 7, 57379–57390. doi:10.18632/oncotarget.10545

Jin, S.-G., Jiang, Y., Qiu, R., Rauch, T. A., Wang, Y., Schackert, G., et al. (2011). 5-Hydroxymethylcytosine Is Strongly Depleted in Human Cancers but its Levels Do Not Correlate with IDH1 Mutations. *Cancer Res.* 71, 7360–7365. doi:10.1158/0008-5472.can-11-2023

Kasprzak, A., and Adamek, A. (2019). Insulin-Like Growth Factor 2 (IGF2) Signaling in Colorectal Cancer-From Basic Research to Potential Clinical Applications. *Int. J. Mol. Sci.* 20, 4915. doi:10.3390/ijms20194915

Langmead, B., and Salzberg, S. L. (2012). Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi:10.1038/nmeth.1923

Lehmann-Werman, R., Neiman, D., Zemmour, H., Moss, J., Magenheim, J., Vaknin-Dembinsky, A., et al. (2016). Identification of Tissue-specific Cell Death Using Methylation Patterns of Circulating DNA. *Proc. Natl. Acad. Sci. USA* 113, E1826–E1834. doi:10.1073/pnas.1519286113

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352

Li, J.-H., Liu, S., Zhou, H., Qu, L.-H., and Yang, J.-H. (2014). starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and Protein-RNA Interaction Networks from Large-Scale CLIP-Seq Data. *Nucl. Acids Res.* 42, D92–D97. doi:10.1093/nar/gkt1248

Li, W., Zhang, X., Lu, X., You, L., Song, Y., Luo, Z., et al. (2017). 5-Hydroxymethylcytosine Signatures in Circulating Cell-free DNA as Diagnostic Biomarkers for Human Cancers. *Cell Res* 27, 1243–1257. doi:10.1038/cr.2017.121

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, 550. doi:10.1186/s13059-014-0550-8

Marchese, F. P., Raimondi, I., and Huarte, M. (2017). The Multidimensional Mechanisms of Long Noncoding RNA Function. *Genome Biol.* 18, 206. doi:10.1186/s13059-017-1348-2

Moutinho-Ribeiro, P., Coelho, R., Giovannini, M., and Macedo, G. (2017). Pancreatic Cancer Screening: Still a Delusion? *Pancreatology* 17, 754–765. doi:10.1016/j.pan.2017.07.001

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a Flexible Suite of Utilities for Comparing Genomic Features. *Bioinformatics* 26, 841–842. doi:10.1093/bioinformatics/btq033

Schizas, D., Charalampakis, N., Kole, C., Economopoulou, P., Koustas, E., Gkotsis, E., et al. (2020). Immunotherapy for Pancreatic Cancer: A 2020 Update. *Cancer Treat. Rev.* 86, 102016. doi:10.1016/j.ctrv.2020.102016

Schultz, N. A., Dehlendorff, C., Jensen, B. V., Bjerregaard, J. K., Nielsen, K. R., Bojesen, S. E., et al. (2014). MicroRNA Biomarkers in Whole Blood for Detection of Pancreatic Cancer. *JAMA* 311, 392–404. doi:10.1001/jama.2013.284664

Shen, S. Y., Singhania, R., Fehringer, G., Chakravarthy, A., Roehrl, M. H. A., Chadwick, D., et al. (2018). Sensitive Tumour Detection and Classification Using Plasma Cell-free DNA Methylomes. *Nature* 563, 579–583. doi:10.1038/s41586-018-0703-0

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer Statistics, 2020. *CA A. Cancer J. Clin.* 70, 7–30. doi:10.3322/caac.21590

Sohal, D. P. S., Willingham, F. F., Falconi, M., Raphael, K. L., and Crippa, S. (2017). Pancreatic Adenocarcinoma: Improving Prevention and Survivorship. *Am. Soc. Clin. Oncol. Educ. Book* 37, 301–310. doi:10.1200/edbk_175222

Song, C.-X., Yin, S., Ma, L., Wheeler, A., Chen, Y., Zhang, Y., et al. (2017). 5-Hydroxymethylcytosine Signatures in Cell-free DNA Provide Information about Tumor Types and Stages. *Cel Res* 27, 1231–1242. doi:10.1038/cr.2017.106

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* 324, 930–935. doi:10.1126/science.1170116

Tang, X., Zhang, M., Sun, L., Xu, F., Peng, X., Zhang, Y., et al. (2021). The Biological Function Delineated across Pan-Cancer Levels through lncRNA-Based Prognostic Risk Assessment Factors for Pancreatic Cancer. *Front. Cel Dev. Biol.* 9, 694652. doi:10.3389/fcell.2021.694652

Unger, C., Kramer, N., Unterleuthner, D., Scherzer, M., Burian, A., Rudisch, A., et al. (2017). Stromal-derived IGF2 Promotes colon Cancer Progression via Paracrine and Autocrine Mechanisms. *Oncogene* 36, 5341–5355. doi:10.1038/onc.2017.116

Vu, T. H., Chuyen, N. V., Li, T., and Hoffman, A. R. (2003). Loss of Imprinting of IGF2 Sense and Antisense Transcripts in Wilms' Tumor. *Cancer Res.* 63, 1900–1905.

Wan, J. C. M., Massie, C., Garcia-Corbacho, J., Mouliere, F., Brenton, J. D., Caldas, C., et al. (2017). Liquid Biopsies Come of Age: towards Implementation of Circulating Tumour DNA. *Nat. Rev. Cancer* 17, 223–238. doi:10.1038/nrc.2017.7

Wang, C., Yin, R., Dai, J., Gu, Y., Cui, S., Ma, H., et al. (2018). Whole-genome Sequencing Reveals Genomic Signatures Associated with the Inflammatory Microenvironments in Chinese NSCLC Patients. *Nat. Commun.* 9, 2054. doi:10.1038/s41467-018-04492-2

Wang, J., Tang, J., Lai, M., and Zhang, H. (2014). 5-Hydroxymethylcytosine and Disease. *Mutat. Research/Reviews Mutat. Res.* 762, 167–175. doi:10.1016/j.mrrev.2014.09.003

White, N. M., Cabanski, C. R., Silva-Fisher, J. M., Dang, H. X., Govindan, R., and Maher, C. A. (2014). Transcriptome Sequencing Reveals Altered Long Intergenic Non-coding RNAs in Lung Cancer. *Genome Biol.* 15, 429. doi:10.1186/s13059-014-0429-8

Xu, R.-h., Wei, W., Krawczyk, M., Wang, W., Luo, H., Flagg, K., et al. (2017). Circulating Tumour DNA Methylation Markers for Diagnosis and Prognosis of Hepatocellular Carcinoma. *Nat. Mater* 16, 1155–1161. doi:10.1038/nmat4997

Yan, X., Hu, Z., Feng, Y., Hu, X., Yuan, J., Zhao, S. D., et al. (2015). Comprehensive Genomic Characterization of Long Non-coding RNAs across Human Cancers. *Cancer Cell* 28, 529–540. doi:10.1016/j.ccell.2015.09.006

Yuan, X., Li, Y., and Li, P. (2020). Regulatory Effects of Linc00486 on Biological Characteristics of Breast Cancer Cells. *Revista argentina de clínica psicológica* 29, 356–363.

Zhang, H., Zhu, C., He, Z., Chen, S., Li, L., and Sun, C. (2020). LncRNA PSMB8-AS1 Contributes to Pancreatic Cancer Progression via Modulating miR-382-3p/STAT1/PD-L1 axis. *J. Exp. Clin. Cancer Res.* 39, 179. doi:10.1186/s13046-020-01687-8

Zhang, Y., Yang, J., Li, H., Wu, Y., Zhang, H., and Chen, W. (2015). Tumor Markers CA19-9, CA242 and CEA in the Diagnosis of Pancreatic Cancer: a Meta-Analysis. *Int. J. Clin. Exp. Med.* 8, 11683–11691.

Zhou, M., Diao, Z., Yue, X., Chen, Y., Zhao, H., Cheng, L., et al. (2016). Construction and Analysis of Dysregulated lncRNA-Associated ceRNA Network Identified Novel lncRNA Biomarkers for Early Diagnosis of Human Pancreatic Cancer. *Oncotarget* 7, 56383–56394. doi:10.18632/oncotarget.10891

Zhou, M., Hou, P., Yan, C., Chen, L., Li, K., Wang, Y., et al. (2021a). Cell-free DNA 5-hydroxymethylcytosine Profiles of Long Non-coding RNA Genes Enable Early Detection and Progression Monitoring of Human Cancers. *Clin. Epigenet* 13, 197. doi:10.1186/s13148-021-01183-6

Zhou, M., Zhang, Z., Bao, S., Hou, P., Yan, C., Su, J., et al. (2021b). Computational Recognition of lncRNA Signature of Tumor-Infiltrating B Lymphocytes with Potential Implications in Prognosis and Immunotherapy of Bladder Cancer. *Brief Bioinform* 22, bbaa047. doi:10.1093/bib/bbaa047

Zhou, M., Zhao, H., Wang, X., Sun, J., and Su, J. (2019a). Analysis of Long Noncoding RNAs Highlights Region-specific Altered Expression Patterns and Diagnostic Roles in Alzheimer's Disease. *Brief Bioinform* 20, 598–608. doi:10.1093/bib/bby021

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019b). Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets. *Nat. Commun.* 10, 1523. doi:10.1038/s41467-019-09234-6

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, orclaim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Cell and Developmental Biology

ORIGINAL RESEARCH
published: 11 April 2022
doi: 10.3389/fcell.2022.788799

miR-29b-3p Inhibitor Alleviates Hypomethylation-Related Aberrations Through a Feedback Loop Between miR-29b-3p and DNA Methylation in Cardiomyocytes

Fang Wu[1,2,3†], Qian Yang[1,2†], Yaping Mi[2], Feng Wang[1,2], Ke Cai[2], Yawen Zhang[1,2], Youhua Wang[4], Xu Wang[5], Yonghao Gui[1,2] and Qiang Li[1*]*

[1]Translational Medical Center for Development and Disease, Shanghai Key Laboratory of Birth Defect Prevention and Control, NHC Key Laboratory of Neonatal Diseases, Institute of Pediatrics, Children's Hospital of Fudan University, National Children's Medical Center, Shanghai, China, [2]Cardiovascular Center, NHC Key Laboratory of Neonatal Diseases, Children's Hospital of Fudan University, National Children's Medical Center, Shanghai, China, [3]Department of Neonatology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [4]Department of Cardiology, Longhua Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China, [5]Cancer Institute, Fudan University Shanghai Cancer Center, Shanghai, China

OPEN ACCESS

Edited by:
Chunjie Jiang,
University of Pennsylvania,
United States

Reviewed by:
Cristina Tufarelli,
University of Leicester,
United Kingdom
Xiaolei Liu,
University of Pennsylvania,
United States

***Correspondence:**
Qiang Li
liq@fudan.edu.cn
Yonghao Gui
yhgui@shmu.edu.cn

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Molecular and Cellular Pathology, a section of the journal Frontiers in Cell and Developmental Biology

Received: 03 October 2021
Accepted: 18 March 2022
Published: 11 April 2022

Citation:
Wu F, Yang Q, Mi Y, Wang F, Cai K, Zhang Y, Wang Y, Wang X, Gui Y and Li Q (2022) miR-29b-3p Inhibitor Alleviates Hypomethylation-Related Aberrations Through a Feedback Loop Between miR-29b-3p and DNA Methylation in Cardiomyocytes. Front. Cell Dev. Biol. 10:788799. doi: 10.3389/fcell.2022.788799

As a member of the miR-29 family, miR-29b regulates global DNA methylation through target DNA methyltransferases (DNMTs) and acts as both a target and a key effector in DNA methylation. In this study, we found that miR-29b-3p expression was inversely correlated with DNMT expression in the heart tissues of patients with congenital heart disease (CHD), but whether it interacts with DNMTs in cardiomyocytes remains unknown. Further results revealed a feedback loop between miR-29b-3p and DNMTs in cardiomyocytes. Moreover, miR-29b-3p inhibitor relieved the deformity of hypomethylated zebrafish and restored the DNA methylation patterns in cardiomyocytes, resulting in increased proliferation and renormalization of gene expression. These results suggest mutual regulation between miR-29b-3p and DNMTs in cardiomyocytes and support the epigenetic normalization of miRNA-based therapy in cardiomyocytes.

Keywords: DNA methylation, miR-29b-3p, DNA methyltransferases, congenital heart disease, zebrafish, proliferation

INTRODUCTION

Dynamic DNA methylation orchestrates cardiomyocyte development, postnatal maturation and cardiovascular diseases (CVDs) (Gilsbach, et al., 2014). Various studies have suggested that DNA methylation aberrations contribute to the development of CVDs, such as congenital heart disease (CHD), atherosclerosis, hypertension, and cardiac hypertrophy (Fernández-Sanlés, et al., 2017; Zhong, et al., 2016). The global DNA methylation of heart tissues in patients with tetralogy of Fallot (TOF) is lower than that of control tissues, while the *NKX2.5*, *HAND1*, *RXRA* and *TBX5* promoters were present under high methylation conditions (Sheng, et al., 2013; Sheng, et al., 2014; Sheng, et al., 2012). The global DNA methylation level decreased in patients with ischemic cardiac disease, atherosclerosis or essential hypertension, while the methylation level of the estrogen receptor (ER)-α

or 11 β-hydroxysteroid dehydrogenase 2 (*11βHSD2*) promoter increased (Friso, et al., 2008; Huang, et al., 2009; Smolarek, et al., 2010; Ying, 2000). These studies revealed that DNA methylation in the global genome and specific genes present their own specific methylation patterns in CVDs. DNA methyltransferases (*DNMTs*), including *DNMT1*, *DNMT3A* and *DNMT3B*, participate in the process of DNA methylation, where *DNMT1* functions in the maintenance of DNA methylation and *DNMT3A* and *DNMT3B* mediate *de novo* DNA methylation (Rideout, et al., 2001). *DNMT* siRNA disrupted the assembly of sarcomeres and reduced the beating frequency, contraction movement, field action potential amplitude and cytosolic calcium signal of cardiomyocytes (Fang, et al., 2016).

An increasing number of studies have focused on the role of miRNAs, which contain 21–25 nucleotides. They regulate posttranscriptional gene expression through mRNA cleavage and degradation or translational inhibition, which depends on the degree of complementarity between miRNA and target mRNA sequence (Kasinski and Slack 2011). Aberrant miRNA expression patterns have been reported in various CVDs, including cardiac hypertrophy, fibrosis, heart failure, arrhythmia, atherosclerosis and TOF (Bruneau 2008; Liu and Olson 2010). A number of clinical studies have also shown that the miR-29 family plays a role in the occurrence and development of CHD. Among 21 patients with TOF, the expression patterns of 18 miRNAs were significantly different, and the expression of miR-29c, which belongs to the miR-29c family, was downregulated (Zhang, et al., 2013a). Maternal blood tests of 30 CHD fetuses showed that miR-29c is significantly elevated in pregnant women with fetal VSD, ASD, and TOF (Nagy, et al., 2019; Zhu, et al., 2013). In patients with persistent atrial fibrillation (AF) after rheumatic heart disease, miR-29b-1-5p and miR-29b-2-5p interact with 24 downregulated circRNAs to participate in the remodeling of heart structure in patients with AF (Hu, et al., 2019). It has been reported that miRNA expression patterns are often disrupted by aberrant DNA methylation in many diseases. Many miRNAs were downregulated or upregulated by DNA hypermethylation or hypomethylation, respectively. Saito et al. (Saito, et al., 2006) first found that miR-127 upregulation was associated with its DNA methylation status. It has been reported that CpG island hypermethylation occurs in the miR-124a (Lujambio, et al., 2007), miR-34 (Chen, et al., 2012; Suzuki, et al., 2010) and miR-9 families (Roman-Gomez, et al., 2009), which is related to the transcriptional inactivation of these miRNAs in human tumors. It has been widely reported that hypermethylation of specific CpG islands in gene promoter regions is a common mechanism of miRNA silencing.

miRNAs are novel regulators of DNA methylation and act by targeting methylation-related proteins, including *DNMTs*, *MBD2*, *MBD4* and *MeCP2* (Wang, et al., 2017). miR-101 inhibited the expression of *DNMT3A*, resulting in a decrease in global DNA methylation in lung cancer (Yan, et al., 2014). The expression of *DNMT3A* and *DNMT3B* was high in lung cancer, and the expression of the miR-29 family was negatively correlated with them. Further study showed that the miR-29 family directly targets the 3′ untranslated regions (3′UTRs) of *DNMT3A* and *DNMT 3B* (Fabbri, et al., 2007). The expression of miR-17–92 was decreased in patients with idiopathic pulmonary fibrosis (IPF), while the expression of *DNMT1* increased. Further study identified that several miRNAs from the miR-17–92 cluster targeted the *DNMT1* gene (Dakhlallah, et al., 2013). It was shown that miR-212 repressed the expression of *MeCP2* in gastric cancer (Wada, et al., 2010) and miR-373 suppressed the expression of *MBD2* in hilar cholangiocarcinoma (Chen, et al., 2011).

miRNAs regulate DNA methylation by modulating methylation-related proteins and expand strategies based on this to treat diseases with aberrant DNA methylation. The miR-29 family, the most widely studied epigenetic factor, reverted aberrant methylation by targeting *DNMT3A* and *DNMT3B* (Garzon, et al., 2009; Morita, et al., 2013; Qiu, et al., 2018; Zhang, et al., 2018b) in leukemia, in porcine early embryo development or in lung cancer cells, while the interaction between miR-29b-3p and *DNMTs* in CHD remains unknown. Our previous results showed that in patients with CHD, *DNMT1*, *DNMT3A* and *DNMT3B* showed a statistically significant negative correlation with miR-29b-3p expression. In this study, it was also revealed that there was a feedback loop between miR-29b-3p and *DNMTs* in cardiomyocytes. It is worth continuing to explore whether upregulation of DNMT by suppressing miR-29b-3p expression is sufficient to induce effective DNA hypermethylation in cardiomyocytes. Therefore, in the present study, we will explore the interaction between miR-29b-3p and DNA methylation and the efficacy of miRNA in the treatment of hypomethylated zebrafish and cardiomyocytes.

MATERIALS AND METHODS

Patients With CHD

Heart tissues were obtained from 17 patients with CHD (mean age: 24.5 ± 22.0 months; 47.1% female and 52.9% male) between 2014 and 2016 from the Children's Hospital of Fudan University, Shanghai, China. The residual tissues were trimmed from the right ventricular outflow tract (RVOT) during surgery and immediately stored in liquid nitrogen. The Ethics Committee of Children's Hospital of Fudan University approved this study.

RNA Extraction and Quantitative RT-PCR Analysis

Total RNA was isolated from frozen heart tissues, cardiomyocytes or zebrafish using TRIzol reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol. Reverse transcription was conducted with the PrimeScript RT reagent kit (TaKaRa, Shiga, Japan). RNA expression was quantified with SYBR Premix Ex Taq™ (TaKaRa, Shiga, Japan). The primer sequences are listed in **Supplementary Table S4**. MiRNAs were reverse-transcribed by the miRcute Plus miRNA First-Strand cDNA Synthesis Kit (TIANGEN, Beijing, China) and quantified by the miRcute Plus miRNA qPCR Detection Kit (TIANGEN, Beijing, China) with specific primers (TIANGEN, Beijing, China). The relative miRNA and mRNA quantification were determined using the

comparative CT method and were normalized against U6 for miRNA or β-actin for mRNA.

Bioinformatics Analysis

The prediction of TF binding sites was performed via TFSEARCH (http://www.cbrc.jp/research/db/TFSEARCH.html). The CpG enrichment regions were analyzed, and the BSP primers were designed by the online MethPrimer software (http://www.urogene.org/methprimer/index1.html). The primers for targeted bisulfite sequencing (MethylTarget) were designed using the online primer3 software (http://primer3.ut.ee/). The TargetScan and PicTar algorithms were used to predict the target genes of miR-29b-3p.

Cell Culture

The HL1 cell line was provided by Professor Duan Ma (Fudan University, Shanghai, China). The HEK293 cell line was purchased from the Cell Bank of the Chinese Academy of Sciences (Shanghai, China). Both cell lines were cultured in Dulbecco's modified Eagle medium (DMEM, Gibco, Waltham, MA) containing 10% FBS with 1% penicillin-streptomycin. Cells were cultured in a humid environment with 5% CO_2 and at 37°C.

Plasmid Constructs

The promoter regions of the hsa-miR-29b-1 and hsa-miR-29b-2 genes were amplified and inserted into the KpnI and SacI sites of the pGL3-promoter vector (Promega, Madison, Wisconsin) to generate the pGL3-hsa-miR-29b-1/2-promoter plasmid. The primers for PCR amplification are listed in **Supplementary Table S5**. The *DNMT3A* and *DNMT3B* 3′UTRs from human/rat/mouse genomic DNA were cloned into XhoI and NotI sites downstream of Renilla luciferase in the psiCHECK-2 vector (Promega, Madison, Wisconsin), while the firefly luciferase gene was used as an internal control. Mutation of the *DNMT3A* and *DNMT3B* 3′ UTRs was performed using the Fast Mutagenesis System (TransGen Biotech, Beijing, China). The PCR primers are listed in **Supplementary Table S6**.

Transfection and Luciferase Assay

HEK293T cells plated in 96-well plates were transfected with 100 ng of pGL3-basic, pGL3-promoter, pGL3-hsa-miR-29b-1/2-promoter (unmethylated), or mpGL3-hsa-miR-29b-1/2-promoter (methylated) plasmid. The pGL3-basic vector without the promoter sequences was used as a negative control. The pRL-TK plasmid (Promega, Madison, Wisconsin) containing the Renilla luciferase gene was cotransfected with the above plasmid to standardize the luciferase activity.

HL1 cells were cotransfected with psiCHECK-2 vector (100 ng) containing the 3′UTR of *DNMT3A* or *DNMT3B* (WT or MUT) and miRNA mimic (20 pmol) in 96-well plates. The four groups were psiCHECK-2-*DNMT3A*/3B-WT + miR-NC mimic, psiCHECK-2-*DNMT3A*/3B-WT + miR-29b-3p mimic, psiCHECK-2-*DNMT3A*/3B-MUT + miR-NC mimic and psiCHECK-2-*DNMT3A*/3B-MUT + miR-29b-3p mimic. All transfections were performed with Lipofectamine 3000 transfection reagent (Invitrogen, Carlsbad, CA).

Luciferase analysis was performed 24 h later by a dual-luciferase reporter assay (Promega, Madison, Wisconsin) according to the manufacturer's instructions. After lysis in passive lysis buffer at room temperature for 15 min, the relative Renilla luciferase activity of cultured cells was obtained after normalization to firefly luciferase gene activity through reaction with Luciferase Assay Reagent II and Stop & Glo Reagent.

BSP and Cloning-Based Sequencing

DNA was subjected to bisulfite modification by the EpiTect bisulfite kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. The BSP primers for the promoter regions of the miR-29b-3p gene were designed by the online MethPrimer software (http://www.urogene.org/methprimer/) (**Supplementary Table S7**). The purified PCR products were used to ligate the pMD™18-T vector (TaKaRa, Shiga, Japan) and were then transformed into DH5α competent cells (TIANGEN, Beijing, China). After 12 h of incubation at 37°C, blue/white and ampicillin selection was performed. Ten different positive clones were randomly selected for sequencing. The BSP sequencing data were analyzed by BIQ Analyzer software (Max Planck Institute for Informatics, Saarbrücken, Germany).

CpG Methyltransferase (M. SssI) Treatment

M. SssI (New England BioLabs, Beverly, MA) was incubated with 1 μg of pGL3-hsa-miR-29b-1/2-promoter plasmid in 20 μl of 1X NEBuffer 2, 10 mM $MgCl_2$, 1 mM dithiothreitol, and 160 μM S-adenosylmethionine for 3 h at 37°C.

Targeted Bisulfite Sequencing

MethylTarget performed by Genesky Biotechnologies Inc. (Shanghai, China) was used to detect the miR-29 methylation density. The primers used for miR-29 are shown in **Supplementary Table S8**. A detailed description of the MethylTarget assay was reported previously (Zhang, et al., 2020).

Western Blot Analysis

Total protein was extracted using RIPA lysis buffer (Beyotime, Shanghai, China) and protease inhibitor cocktail (Thermo Fisher Scientific, Waltham, MA). The protein concentration was determined by a Pierce BCA protein assay kit (Thermo Fisher Scientific, Waltham, MA). Anti-*DNMT1*, anti-*DNMT3A*, anti-*DNMT3B*, and anti-β-actin antibodies and HRP-labeled goat anti-rabbit IgG secondary antibodies were purchased from Cell Signaling Technology (Danvers, MA) and Abcam (Cambridge, MA), respectively. ECL reagents (Merck Millipore, Darmstadt, Germany) were used to visualize specific protein bands.

Cell Proliferation Assay

Cell viability was measured by the CCK-8 assay (Dojindo Laboratories, Kumamoto, Japan). At 24, 48, and 72 h, the CCK-8 solution was prepared with medium to a concentration of 10%, 100 μl of the mixed solution was added to each 96-well plate, and the operation was protected from light. The plates were incubated at 37°C for 2.5 h, and the absorbance at 450 nm was then measured.

EdU Incorporation Assay

The EdU incorporation assay was performed according to the manufacturer's protocol (Life Technologies, Waltham, MA). The cell proliferation rate was calculated as the proportion of nucleated cells incorporated into EdU to the total number of cells by randomly selecting 10 high-power fields per well.

Zebrafish Embryology and Microinjection

Zebrafish breeding, embryo collection and maintenance were carried out in accordance with recognized standard operating procedures. The injection concentration of miRNA inhibitor was 5 μM, and that of 5-azacytidine was 25 μM. At the 1–4 cell stage, 3 nl of miRNA inhibitor or 5-azacytidine was injected into the yolk of each zebrafish embryo. A Leica M205 FA digital camera was used to photograph the embryos, and Adobe Photoshop CS5 software was used to process the digital images.

5-Azacytidine and 5-aza-2′-Deoxycytidine (Decitabine) Working Solution

5-azacytidine powder (Sigma-Aldrich, St. Louis, MO) was dissolved in an appropriate amount of DMEM (cardiomyocyte treatment) or blue egg water (zebrafish embryo treatment). The concentration of the stock solution was 500 μM. The working solution was diluted to 25 μM. Five milligrams of decitabine powder (Sigma-Aldrich, St. Louis, MO) was dissolved in 1 ml of DMSO. The working solution was diluted to 20 μM. The stock solutions were stored at −80°C.

General Morphology Score System

The general morphology score (GMS) system is used as a quantitative assessment method to evaluate the development of zebrafish embryos, which displays the development scores of zebrafish embryos at 24 hpf, 48 hpf and up to 72 hpf, with different scores assigned to specific developmental endpoints (Hermsen, et al., 2011). It included evaluation indicators such as tail detachment, somite formation, eye development, heartbeat and blood flow speed. The full score was 7 at 24 hpf, 12 at 48 hpf and 15 at 72 hpf.

Shortening Fraction Quantification

The maximum systolic and diastolic frames of the video were saved as JPEGs, and the width of the maximum systolic and diastolic hearts of the ventricles was measured from the image by ImageJ. The ventricular shortening fraction (%) was calculated as follows: ×100 (diastolic width-systolic width)/(diastolic width)%.

Statistical Analysis

Statistical analysis was carried out by Stata. Values are expressed as the means ± SEM. Spearman's rank correlation was used to examine the correlation between two continuous variables. One-way analysis of variance (ANOVA) was used to analyze differences among multiple groups. Two-way ANOVA was used to evaluate the expression of miR-29b-3p at different time points after 5-azacytidine or decitabine treatment and the effects of miR-29b-3p and time variables on the proliferation of HL1 cells treated with 5-azacytidine. Student's t-test was used to determine the statistical significance. Significance was defined as follows: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and $^{****}p < 0.0001$.

RESULTS

Negative Correlation Between the Expression of *DNMTs* and miR-29b-3p in Patients With CHD

To determine whether a correlation exists between the expression of *DNMTs* and miR-29b-3p in patients with CHD, we analyzed the qPCR data from 17 patients with CHD (**Supplementary Tables S1, S2**). *DNMT1*, *DNMT3A* and *DNMT3B* showed a statistically significant negative correlation with the expression of miR-29b-3p ($r = -0.5137$, $p = 0.0349$; $r = -0.5123$, $p = 0.0355$; and $r = -0.6012$, $p = 0.0107$, respectively) (**Figures 1A–C**).

Transcriptional Regulatory Activity of the miR-29b-1 and miR-29b-2 Gene Promoters

As there was a negative correlation between *DNMT* and miR-29b-3p expression in patients with CHD, we further performed experiments to explore whether mutual regulation existed between them. We first detected the DNA methylation status of the miR-29b gene promoter and analyzed its correlation to miR-29b-3p expression in patients with CHD. Hsa-miR-29b-3p is the mature form of premiR-29b-1 and premiR-29b-2. The gene encoding premiR-29b-1 is located on Chr. 7q32.3, while the gene encoding premiR-29b-2 is located on Chr. 1q32.2. A fragment from −1,530 bp to +165 bp relative to the transcription start site (TSS) of the miR-29b-1 gene was shown to have promoter activity and include binding sites for transcription factors Gli, Myc and NF-κB (Mott, et al., 2010). Based on the information analyzed by MethPrimer software, we found a CpG-enriched area from −873 bp to +158 bp relative to the TSS of the miR-29b-1 gene, which contained 20 CpG units (**Figures 2A,B**, **Supplementary Figure S1A**). The promoter region of the miR-29b-2 gene has rarely been reported, so we focused on the fragment from −2000 bp to +200 bp relative to the TSS of the miR-29b-2 gene. With the online MethPrimer software, we found a CpG-enriched area containing 9 CpG units, which was located in the −1495 bp to −1077 bp region relative to the TSS of the miR-29b-2 gene (**Figures 2C,D**, **Supplementary Figure S1B**).

To identify the transcriptional regulatory activity of the two fragments, the −873 bp to +158 bp region of the miR-29b-1 gene and the −1495 bp to −1077 bp region of the miR-29b-2 gene were cloned into the pGL3-promoter plasmid. The relative luciferase activities of the pGL3-miR-29b-1-promoter and pGL3-miR-29b-2-promoter increased by 3.1 times and 2.2 times, respectively, compared with that of the pGL3-promoter (**Figures 2E,F**). The pGL3-basic plasmid was the negative control.

Negative Correlation Between miR-29b-3p Expression and its Promoter Methylation Status in Patients With CHD

To explore the relationship between miR-29b-3p expression and its promoter methylation status in patients with CHD, we performed Spearman's correlation tests by Stata. The methylation status of the promoter regions of the miR-29b-1

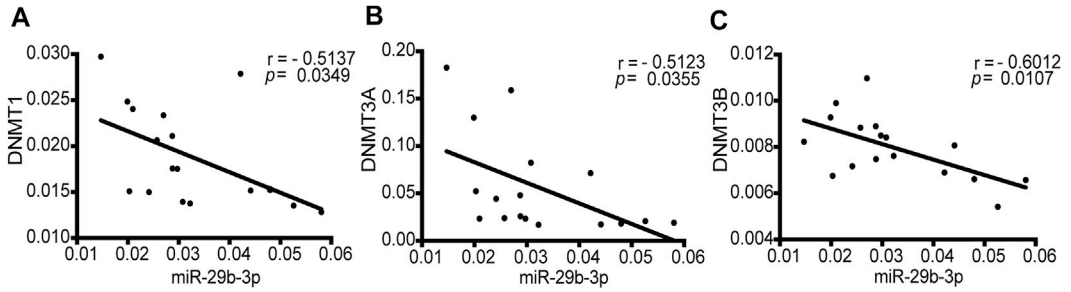


FIGURE 1 | Correlations between the mRNA expression of *DNMTs* and miR-29b-3p in patients with CHD. **(A)** The correlation between *DNMT1* and miR-29b-3p expression (r = –0.5137, p = 0.0349). **(B)** The correlation between *DNMT3A* and miR-29b-3p expression (r = –0.5123, p = 0.0355). **(C)** The correlation between *DNMT3B* and miR-29b-3p expression (r = –0.6012, p = 0.0107). Spearman's correlation tests were used.

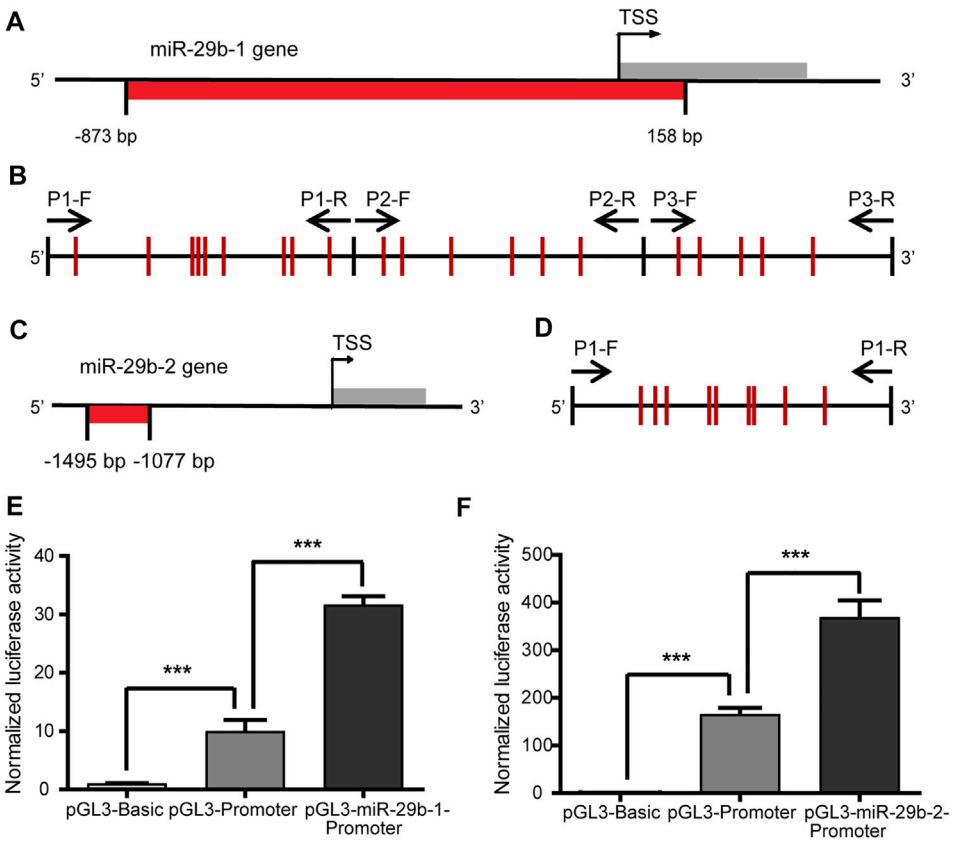


FIGURE 2 | Transcriptional regulatory activity of the miR-29b-1 and miR-29b-2 gene promoters. **(A)** The location of the CpG sites in the promoter region of the miR-29b-1 gene (–873 bp to +158 bp). **(B)** The distribution of the CpG sites and three pairs of primers designed for bisulfite sequencing PCR in the promoter region of the miR-29b-1 gene. **(C)** The location of the CpG sites in the promoter region of the miR-29b-2 gene (–1495 bp to –1077 bp). **(D)** The distribution of the CpG sites and the primers designed for bisulfite sequencing PCR in the promoter region of the miR-29b-2 gene. **(E)** The effect of the –873 bp to +158 bp region on the regulation of miR-29b-1 gene promoter transcriptional activity. **(F)** The effect of the –1407 bp to -1173 bp region on the regulation of miR-29b-2 gene promoter transcriptional activity (***p < 0.001).

gene and miR-29b-2 gene was measured in heart tissues obtained from eight patients with CHD by bisulfite sequencing PCR (BSP) sequencing (**Figures 3A,B**). Three pairs of primers were designed to measure the methylation levels of the miR-29b-1 promoter (P1: –597 bp ~ –257 bp, P2: –226 bp ~ +108 bp and P3: +103 bp ~ +403 bp) (**Figure 2B**), and 1 pair of primers was designed to measure the methylation levels of the miR-29b-2 promoter (P1: –1407 bp ~ –1173 bp) (**Figure 2D**).

As shown in **Figure 3C** a significant negative correlation was observed between miR-29b-3p expression and the

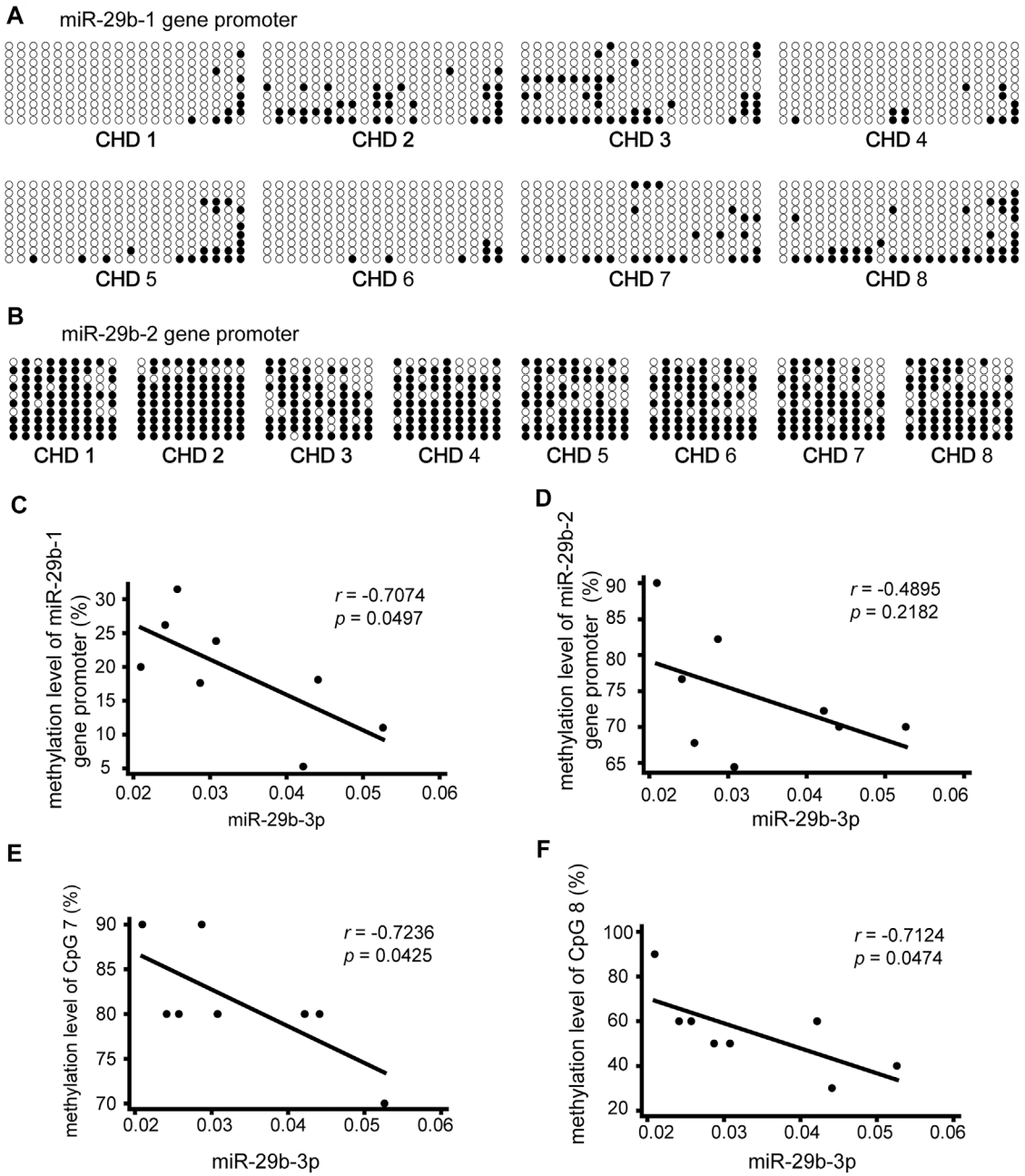


FIGURE 3 | Association of miR-29b-3p expression with its methylation status in eight patients with CHD. **(A** and **B)** Methylation status of the promoter region of the miR-29b-1 gene and miR-29b-2 gene. The black and white circles represent methylated and unmethylated CpG dinucleotides, respectively. **(C)** Correlations between miR-29b-3p expression and the methylation status of the miR-29b-1 gene (r = –0.7074, p = 0.0497, and N = 8). **(D)** Correlations between miR-29b-3p expression and the methylation status of the miR-29b-2 gene (r = –0.4895, p = 0.2182, and N = 8). **(E)** Correlations between miR-29b-3p expression and the methylation status of CpG 7 located in the miR-29b-2 gene (r = –0.7236, p = 0.0425, and N = 8). **(F)** Correlations between miR-29b-3p expression and the methylation status of CpG 8 located in the miR-29b-2 gene (r = –0.7124, p = 0.0474, and N = 8). Spearman's correlation tests were used.

methylation status of the miR-29b-1 gene (r = −0.7074, p = 0.0497, and N = 8). The association between miR-29b-3p expression and the methylation status of the miR-29b-2 gene was negative but not significant (r = −0.4895, p = 0.2182, and N = 8) (**Figure 3D**).

We further analyzed the association between miR-29b-3p expression and the methylation status of each CpG unit. No significant correlations were observed between miR-29b-3p expression and the methylation status of each CpG site located in the miR-29b-1 gene. The associations between miR-29b-3p

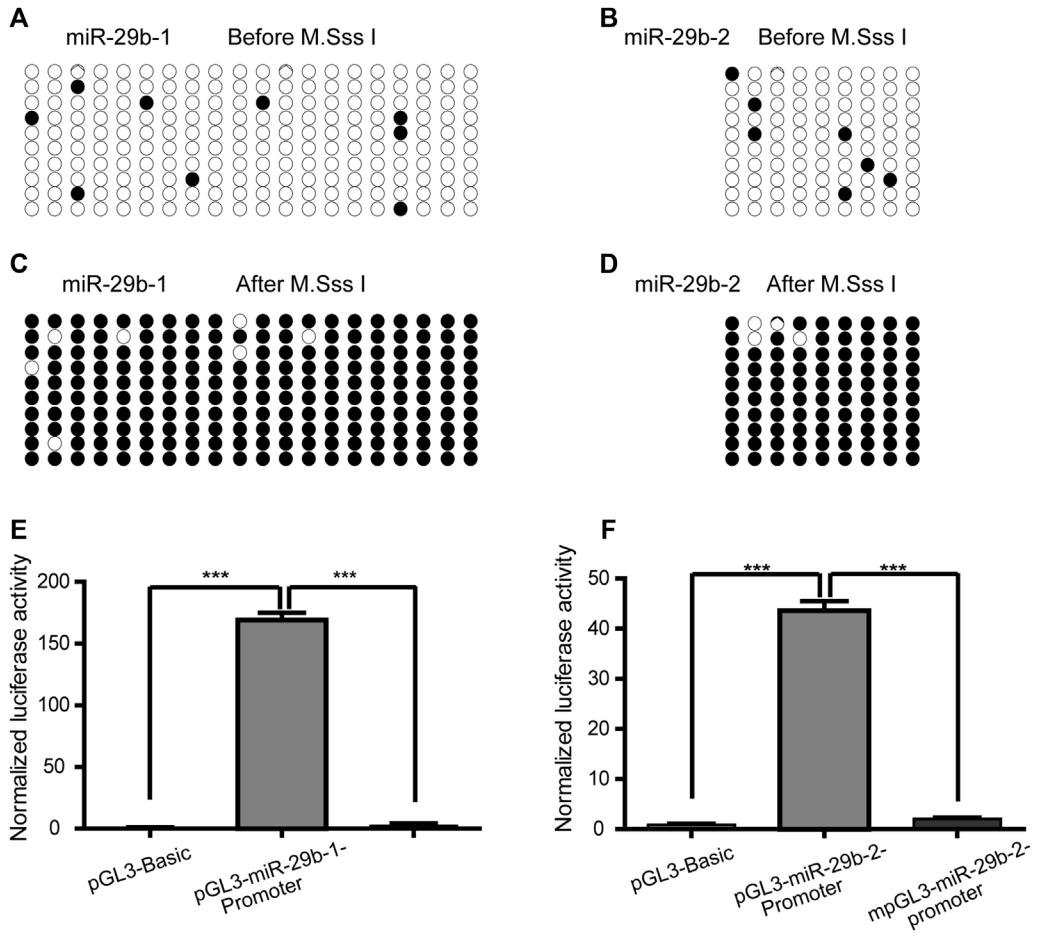


FIGURE 4 | The effect of promoter hypermethylation on the expression of miR-29b-3p. **(A** and **B)** The methylation levels of the pGL3-miR-29b-1 promoter and pGL3-miR-29b-2 promoter before M. SssI treatment. For each plasmid, the methylation status of CpG units is shown for ten clones. Black and white circles indicate the methylated and unmethylated CpG units, respectively. **(C** and **D)** The methylation levels of the pGL3-miR-29b-1 promoter and pGL3-miR-29b-2 promoter after M. SssI treatment. **(E)** The effect of methylation on the transcriptional activity of the −873 bp to +158 bp region of the miR-29b-1 gene. **(F)** The effect of methylation on the transcriptional activity of the −1407 bp to −1173 bp region of the miR-29b-2 gene. (^{ns} not significant, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and $^{****}p < 0.0001$).

expression and the methylation status of CpG 7 and CpG eight located in the miR-29b-2 gene were statistically significant (r = −0.7236, $p = 0.0425$; r = −0.7124, $p = 0.0474$) (**Figures 3E,F**).

Promoter Hypermethylation Decreased the Expression of miR-29b-3p

To explore the impact of methylation on the transcriptional regulatory activity of the two fragments in the miR-29b-1 and miR-29b-2 gene promoters, the pGL3-miR-29b-1-promoter plasmid and pGL3-miR-29b-2-promoter plasmid were methylated by M. SssI methylase. Before M. SssI treatment, the total methylation levels of the 20 inserted CpG units and 9 CpG units were 4.5 and 7.8%, respectively (**Figures 4A,B**). After M. SssI treatment, the overall methylation status was 96.5 and 95.5%, respectively (**Figures 4C,D**). Methylated pGL3-miR-29b-1/2-promoter plasmids (mpGL3-miR-29b-1/2-promoter) or unmethylated pGL3-miR-29b-1/2-promoter plasmids were then transfected into HEK293T cells and assayed for dual-luciferase activities. The relative luciferase activity of the mpGL3-miR-29b-1 promoter decreased 58-fold compared with that of the pGL3-miR-29b-1 promoter, while the relative luciferase activity of the mpGL3-miR-29b-2 promoter decreased nearly 20-fold compared with that of the pGL3-miR-29b-2 promoter (**Figures 4E,F**).

Gene Hypomethylation Increased the Expression of miR-29b-3p

To explore the role of hypomethylation in the regulation of miR-29b-3p expression *in vitro*, we treated HL1 cells with 5-azacytidine or decitabine, which functions as a potent DNA methyltransferase inhibitor. The results showed that the mir-29b gene was hypomethylated in HL1 cells treated with 5-azacytidine or decitabine, as analyzed by the MethylTarget assay (**Figures 5A,B**). Moreover, the expression of miR-29b-3p

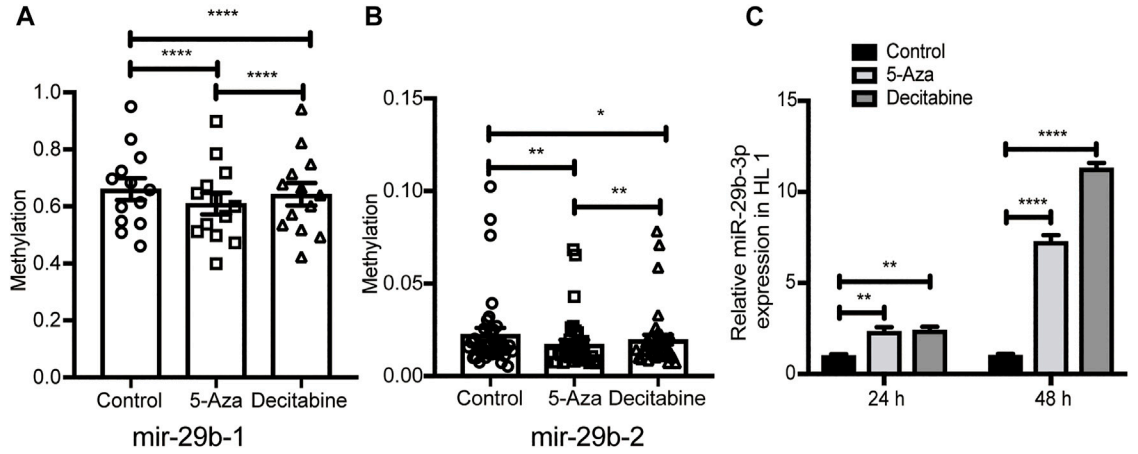


FIGURE 5 | Identification of miR-29b methylation level and mRNA expression after exposure to 5-azacytidine or decitabine. **(A** and **B)** The methylation level of miR-29b-1 and miR-29b-2 in additional samples from the control (n = 3), 5-Aza (n = 5) and decitabine (n = 3) groups analyzed by targeted bisulfite sequencing. **(C)** miR-29b-3p expression in HL1 cells after exposure to 5-azacytidine at 25 μM or decitabine at 20 μM. *p*-values were calculated by one-way ANOVA. ^{ns} not significant; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; and $^{****}p < 0.0001$.

in the 5-azacytidine or decitabine-treated groups was higher than that in the control group at different time points, and the highest expression level was detected at 48 h (**Figure 5C**).

miR-29b-3p Directly Targeted the 3′UTRs of *DNMT3A* and *DNMT3B*

We identified that DNA methylation regulated the expression of miR-29b-3p, and further study aimed to explore the regulatory effect of miR-29b-3p on the expression of *DNMTs*. miR-29b-3p and *DNMTs* are highly conserved in humans, rats, mice and zebrafish (**Supplementary Figure S2**). We used bioinformatic tools (PicTar and TargetScan algorithms) to predict the targets of miR-29b-3p, and *DNMT3A* and *3B* were the putative targets of miR-29b-3p, while the 3′UTR of *DNMT1* showed no complementary binding site with miR-29b-3p (**Figures 6A,B**).

To ascertain the function of miR-29b-3p on *DNMT3A* and *DNMT3B*, a fragment containing the 3′UTRs of *DNMT3A* of *DNMT3B* was spliced to the 3′-end of the synthetic Renilla luciferase reporter gene in the psiCHECK™-2 vector. HL1 cells were cotransfected with miR-29b-3p mimic and psiCHECK™-2-*DNMT3A*-3′UTR plasmid or psiCHECK™-2-*DNMT3B*-3′UTR plasmid and cultured for 48 h. The results showed that miR-29b-3p mimic inhibited the relative luciferase activity of the psiCHECK™-2-*DNMT3A*-3′UTR plasmid and psiCHECK™-2-*DNMT3B*-3′UTR plasmid. We further mutated the 3′UTRs of *DNMT3A* and *DNMT3B* and cotransfected them with miR-29b-3p mimic into HL1 cells. The inhibitory effect of the miR-29b-3p mimic on relative luciferase activity was abrogated after cotransfection with the psiCHECK™-2-*DNMT3A*-3′UTR-MUT plasmid and psiCHECK™-2-*DNMT3B*-3′UTR-MUT plasmid (**Figures 6C,D**). The mRNA expression of *DNMTs* in HL1 cells transfected with the miR-29b-3p mimic decreased, while the expression increased in the miR-29b-3p inhibitor group (**Figure 6E**). Furthermore, the protein expression of *DNMTs* was significantly altered by the miR-29b-3p mimic and its inhibitor (**Figures 6F,G**).

miR-29b-3p Inhibitor Relieved the Aberration of Zebrafish Embryos Treated With 5-azacytidine

On the one hand, miR-29b-3p regulates DNA methylation by targeting *DNMTs*; on the other hand, the miR-29b gene promoter can be hypermethylated or hypomethylated due to its transcriptional ability. Understanding the crosstalk between miR-29b-3p and DNA methylation may promote the discovery of novel therapeutic targets.

To evaluate the impact of the miR-29b-3p inhibitor on the overall development of hypomethylated zebrafish embryos, we assessed their survival, malformation rate and general morphology score by coinjecting 5-azacytidine and miR-29b-3p inhibitor into the yolk of zebrafish embryos at the 1–4-cell stage. The miR-29b-3p expression in zebrafish embryos injected with 5-azacytidine and/or miR-29b-3p inhibitor is shown in **Supplementary Figure S3A**. The mRNA and protein expression of *DNMTs* in zebrafish injected with 5-azacytidine and/or miR-29b-3p inhibitor is shown in **Supplementary Figures S3B–D**. The results showed that 5-azacytidine exposure led to an increased mortality rate and deformity rate in a time-dependent manner, while coinjection with the miR-29b-3p inhibitor partially reduced the mortality and deformity rates (**Figure 7D**). General development assessed by the GMS system was significantly delayed by 5-azacytidine exposure (Yang, et al., 2019). The miR-29b-3p inhibitor partially promoted the developmental status at 48 h postfertilization (hpf) and 72 hpf (**Figures 7G–I**). Zebrafish embryos coinjected with 5-azacytidine

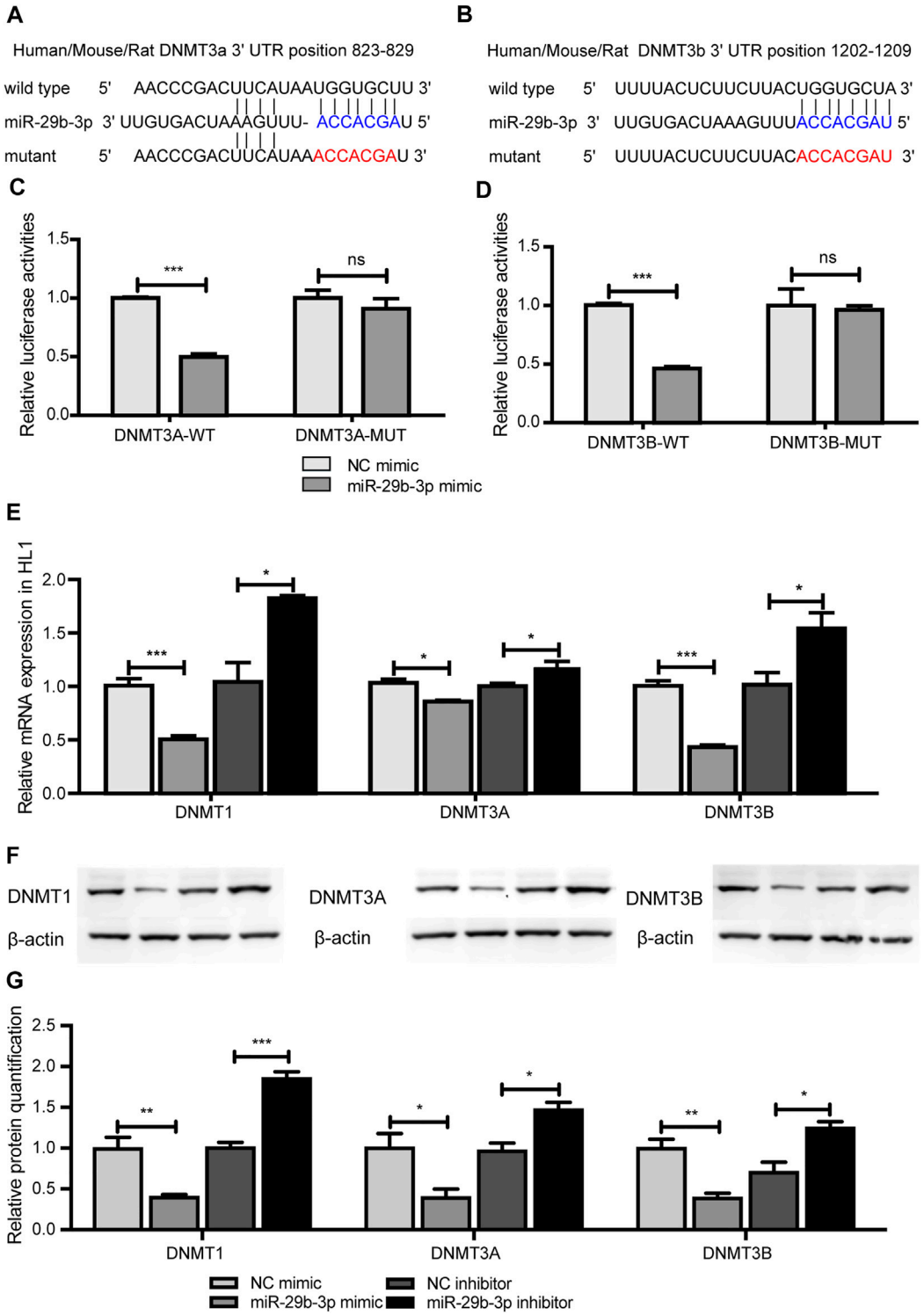


FIGURE 6 | The effect of miR-29b-3p on the expression of *DNMTs*. **(A** and **B)** The predicted binding site of miR-29b-3p in the 3′ untranslated regions of *DNMT3A* and *DNMT3B*. **(C** and **D)** The relative luciferase activity of HL1 cells cotransfected with miR-29b-3p mimic or NC mimic and plasmid containing DNMT3A or 3B wild-type or mutated 3′UTRs. (miR-29b-3p mimic + psiCHECK™-2-DNMT3A/3B-3′UTR vs. NC mimic + psiCHECK™-2-DNMT3A/3B-3′UTR, ***$p < 0.001$; miR-29b-3p mimic + psiCHECK™-2-DNMT3A/3B-3′UTR-MUT vs. NC mimic + psiCHECK™-2-DNMT3A/3B-3′UTR-MUT, p = ns). **(E)** mRNA expression of *DNMTs* in HL1 cells transfected with miR-29b-3p mimic or its inhibitor (miR-29b-3p mimic vs. NC, *$p < 0.05$, miR-29b-3p inhibitor vs. NC inhibitor, ***$p < 0.001$). **(F)** Protein expression of *DNMTs* in HL1 cells transfected with miR-29b-3p mimic or its inhibitor. **(G)** Relative quantification of *DNMT* proteins (ns not significant, *$p < 0.05$, **$p < 0.01$ and ***$p < 0.001$).

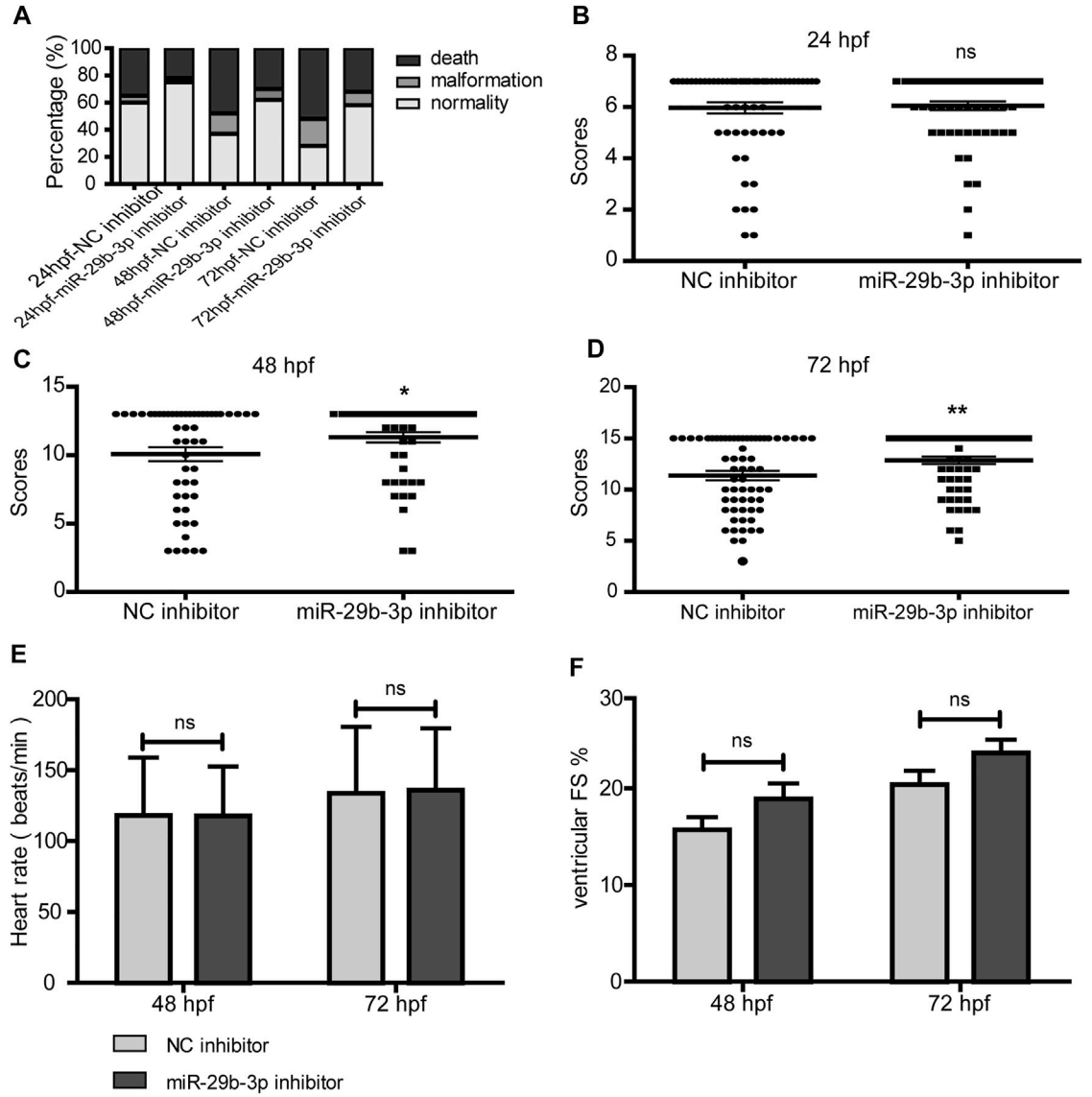


FIGURE 7 | The impact of miR-29b-3p inhibitor on the overall development of hypomethylated zebrafish embryos. **(A)** miR-29b-3p inhibitor partially reduced the mortality and deformity rates of zebrafish that resulted from 5-azacytidine exposure (n > 100). **(B–D)** The zebrafish general development score of the miR-29b-3p inhibitor group was significantly better than that of the NC inhibitor group at 48 hpf and 72 hpf (n > 50). **(E** and **F)** The heart rate and fractional shortening of zebrafish embryos coinjected with 5-azacytidine and miR-29b-3p inhibitor displayed no significant difference from those of the 5-azacytidine and NC inhibitor groups (n = 20).

and NC inhibitor exhibited obvious deformities, including body curvature, yolk sac edema and blood congestion at the cardiac inflow tract at 48 hpf, while coinjection with miR-29b-3p inhibitor relieved the degree of deformity. Representative images of the overview are shown in **Supplementary Figure S4**.

5-Azacytidine exposure induced a decrease in heart rate and fractional shortening (Yang, et al., 2019). The heart rate of zebrafish embryos coinjected with miR-29b-3p inhibitor displayed no significant difference from that of the control group (**Figure 7E**). The fractional shortening was 15.7 ± 5.8% at 48 hpf and 20.4 ± 6.5% at 72 hpf in the NC inhibitor group, which showed no significant differences from those of the miR-29b-3p inhibitor group (18.9 ± 7.3% at 48 hpf and 23.7 ± 6.2% at 72 hpf) (**Figure 7F**).

miR-29b-3p Inhibitor Increased the Proliferation of Hypomethylated Cardiomyocytes

Our previous study showed that a miR-29b-3p inhibitor significantly promoted HL1 cell proliferation (Yang, et al., 2020) (**Figures 8A,D**). After treatment with 5-azacytidine, the proliferation ability of cardiomyocytes decreased as the

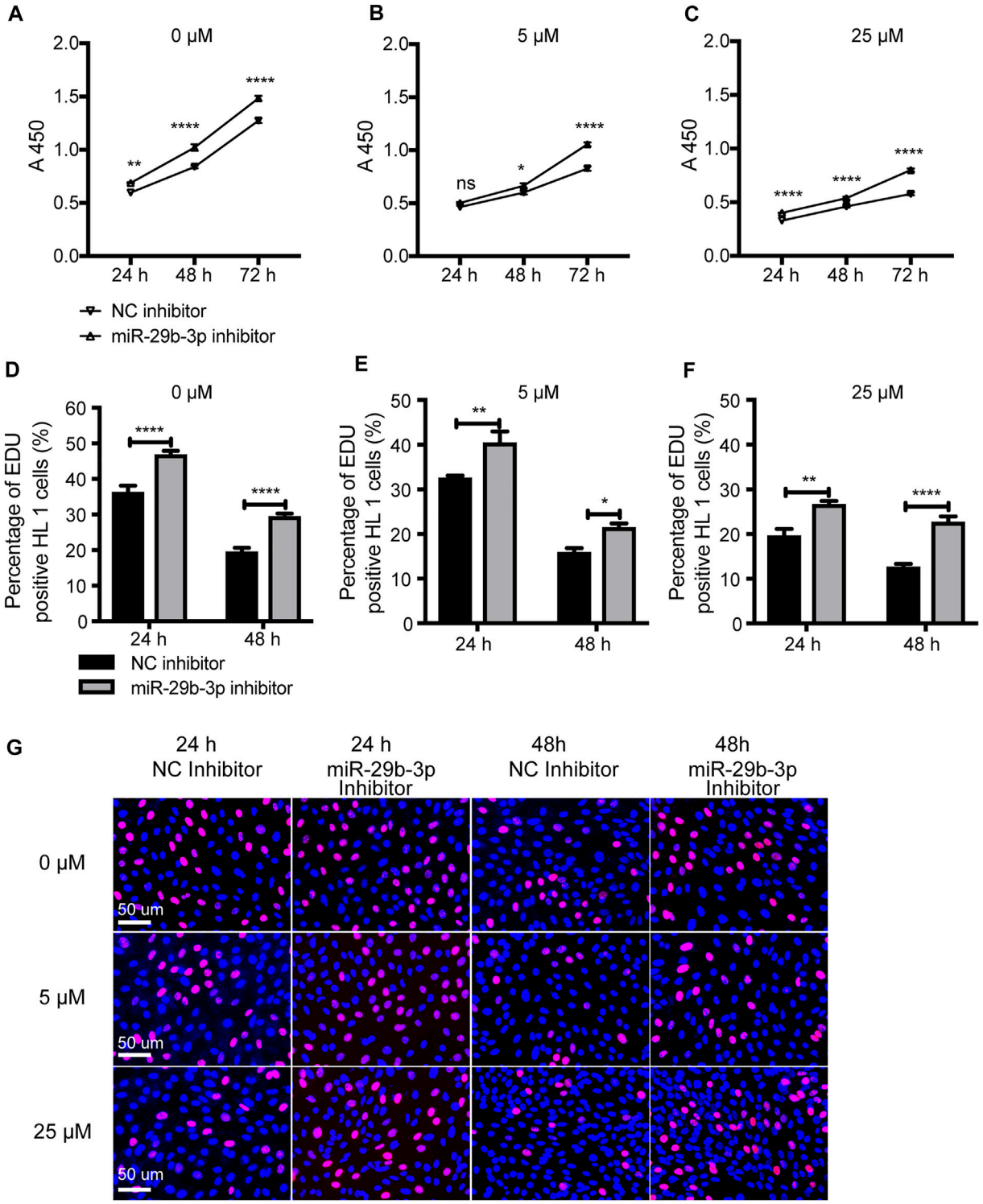


FIGURE 8 | The proliferation ability of hypomethylated cardiomyocytes transfected with miR-29b-3p inhibitor. **(A–C)** Cell proliferation ability detected by a CCK-8 assay at 24, 48 and 72 h **(D–F)** The cell proliferation ability detected by an EdU incorporation assay at 24, 48 and 72 h. **(G)** Representative images of HL1 cells stained with EdU and Hoechst (0 μM, 5 μM, and 25 μM represent 3 concentrations of 5-azacytidine; miR-29b-3p inhibitor vs. NC inhibitor, ^{ns} not significant, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and $^{****}p < 0.0001$).

concentration increased (Yang, et al., 2019) (**Figures 8A–C**). Next, we further evaluated the effect of the miR-29b-3p inhibitor on the proliferation of hypomethylated cardiomyocytes *in vitro*. The mRNA and protein expression of *DNMTs* in HL1 cells treated with 5-azacytidine or decitabine or transfected with miR-29b-3p inhibitor is shown in **Supplementary Figure S5**. The proliferation ability of hypomethylated cardiomyocytes transfected with miR-29b-3p inhibitor or NC inhibitor was detected by CCK-8 and EdU incorporation assays. The proliferation ability of cardiomyocytes exposed to 5-azacytidine at 5 μM was increased after transfection with miR-29b-3p inhibitor compared with that in the control group at 24 h (ns), 48 h ($p < 0.05$) and 72 h ($p < 0.0001$) (**Figure 8B**). The proliferation ability of cardiomyocytes exposed to 5-azacytidine at 25 μM was significantly increased after transfection with miR-29b-3p inhibitor compared with that in the control group at 24, 48 and 72 h ($p < 0.0001$) (**Figure 8C**). The EdU assay results also showed that the miR-29b-3p inhibitor promoted hypomethylated HL1 cell proliferation (**Figures 8D–F**). Representative images of the EdU assay are shown in **Figure 8G**.

Effect of miR-29b-3p Inhibitor on the Gene Expression of Hypomethylated Cardiomyocytes

We found that global hypomethylation resulted in increased expression of 2 genes and decreased expression of 20 genes among 45 candidate genes (Yang, et al., 2019). Here, we analyzed the effect of a miR-29b-3p inhibitor on the expression of heart-related genes in hypomethylated cardiomyocytes. The results showed that the expression of *FGF10*, *TNNT2*, *SSB*, *MYH6* and *ERBB3* was decreased in hypomethylated cardiomyocytes (Yang, et al., 2019). was upregulated when transfected with miR-29b-3p inhibitor (**Supplementary Table S3**).

DISCUSSION

In this study, we found that the expression of miR-29b-3p was negatively correlated with the expression of *DNMTs* in CHD patients. Further results revealed that there was a feedback loop between miR-29b-3p and *DNMTs* in cardiomyocytes. The reduction in miR-29b-3p expression alleviated the deformity of hypomethylated zebrafish and increased the proliferation and renormalization of gene expression by activating DNMT-dependent DNA methylation in cardiomyocytes.

The miR-29 family is a classic effector of epi-miRNAs, which regulate DNA methylation and demethylation (Garzon, et al., 2009; Zhang, et al., 2013b). In our experiments, the expression of *DNMT1* in HL1 cells treated with miR-29b-3p mimic were downregulated, similar to the results found for *DNMT3A* and *DNMT3B*. The results were consistent with those observed in K562, MV4-11, and Kasumi-1 cells (Garzon, et al., 2009) and in the GC-1 germ cell line (Meunier, et al., 2012). Garzon, R et al. confirmed that miR-29 indirectly downregulates *DNMT1* by directly targeting its transactivator Sp1, a zinc finger transcription factor (Garzon, et al., 2009). Compared with *DNMT3A* and *DNMT3B*, *DNMT1* does not show a complementary binding site with miR-29b-3p and is thus not be directly targeted by miR-29b. Garzon, R et al. confirmed that miR-29 indirectly downregulates *DNMT1* by directly targeting its transactivator Sp1, a zinc finger transcription factor. Our data suggested that miR-29b-3p directly targeted the 3′UTRs of *DNMT3A* and *DNMT3B* while indirectly regulating the expression of *DNMT1*.

Promoter hypermethylation is usually associated with gene silencing, and the higher the methylation of the gene promoter is, the lower the gene expression. Our study suggested that miR-29b-3p expression was negatively related to the methylation status of CpG 7 and CpG 8 located in the miR-29b-2 gene. Several key transcription factors (TFs), including *C/EBP*, *SRF*, *Nrf2* and *HES-1*, were predicted to bind to CpG 7 and CpG 8. The hypermethylation of CpG 7 and CpG 8 may block the binding of TFs to the promoter of the miR-29b-2 gene, resulting in reduced expression of miR-29b-3p. Aberrant DNA methylation may also disrupt the expression of TFs that are essential to the transcription of miRNA, indirectly leading to decreased expression of miRNA. Altered methylation of miRNA-encoding genes may also contribute to aberrant miRNA expression.

The negative correlation between miR-29s and *DNMTs* has been explored in many diseases, such as cholangiocarcinoma (Cao, et al., 2021), osteoarthritis (Dou, et al., 2020), and leukemia (Qiu, et al., 2018). The enforced expression of miR-29s regulated downstream genes mediated by *DNMT3B* (Cao, et al., 2021; Qiu, et al., 2018). *DNMT3B* regulates the miR-29b/PTHLH/CDK4/RUNX2 axis by inducing hypermethylation of specific CpG sites in the miR-29b promoter region, preventing chondrocyte loss due to osteoarthritis (Dou, et al., 2020). Similar to the "DNMT-miR-29" epigenetic circuit, negative feedback regulatory loops between "DNMT1-miR-148/152" in esophageal squamous cell carcinoma and "DNMT1-miR-126" in breast cancer have been reported (Liu, et al., 2015; Zhao, et al., 2011). The feedback loop between miR-29b-3p and *DNMTs* represents a new level of complexity in gene regulation. Exogenous miR-29b-3p inhibitor increased the expression of *DNMTs*, which in turn resulted in a decreased expression of endogenous miR-29b-3p. In our study, an exogenous miR-29b-3p inhibitor relieved the degree of demethylated zebrafish deformity, including body curvature, yolk sac edema and blood congestion at the cardiac inflow tract. Furthermore, the miR-29b-3p inhibitor promoted the proliferation and renormalized the gene expression of hypomethylated cardiomyocytes. These findings illustrate the regulatory role of miR-29 in the normalization of disease epigenetics and provide a theoretical basis for the development of miRNA-based therapeutic strategies.

Epigenetic changes are often reversible, which makes miRNAs attractive in the development of new treatment

approaches. Many miRNAs act as biomarkers and prognostic factors for diseases, while only a few are available as therapeutic strategies. This phenomenon may occur for several reasons: one of these reasons may be the absence of pathways to certain physiological organs or tissues (Gallas, et al., 2013; Gavrilov and Saltzman 2012; Pereira, et al., 2017). In our study, a miR-29b-3p inhibitor relieved the deformity of demethylated zebrafish but had no significant effect on heart rate or fractional shortening. However, miR-29b-3p inhibitor increased the proliferation of hypomethylated cardiomyocytes, and this finding is consistent with some research results. Ginkgolide B inhibits hypoxic H9c2 cell apoptosis through miR-29-based inhibition (Ren, et al., 2020). In the ischemia/reperfusion (I/R) injury under PM2.5 exposure, the lncRNA PEAMIR can inactivate the PI3K(p85a)/Akt/GSK3b/p53 cascade pathway that mediates inflammation and apoptosis by downregulating miR-29b-3p (Pei, et al., 2020). The lncRNA TUG1 inhibits apoptosis in H9c2 cells treated with LPS by downregulating miR-29b (Zhang, et al., 2018a). The ineffectivity of the treatment on the cardiovascular system of zebrafish may be due to the injection method, which lacks transmission to specific physiological organs and tissues, resulting in insufficient cellular uptake and processing. Therefore, it is necessary to solve the problem of low cell uptake and processing efficiency. In addition, it is worth noting that some research results are contrary to ours and the above-described studies. *In vivo* experiments of rats induced by endotoxin, doxorubicin or ischemia-reperfusion showed that the upregulation of miR-29b can reduce cardiomyocyte apoptosis, whereas the inhibition of miR-29b exerts the opposite effect (Jing, et al., 2018; Li, et al., 2021; Li, et al., 2020). We speculate that the inconsistent results may be due to different cell types, different experimental models, different expression levels of target mRNA, different degree of cell injury, different stages of tissue development and different doses of miRNA mimic or inhibitor.

Several results from the present study could be serve as motivation for future study. First, the correlations between miR-29b-3p expression and the total methylation level of the miR-29b-2 gene or the methylation status of each CpG site located in the miR-29b-1 gene in patients with CHD were not clear in our study. This may be due to the small sample size, which we will expand to further analyze the relationship between miR-29b-3p expression and the total methylation level of the miR-29b-2 gene. Second, several TFs, including *C/EBP*, *SRF*, *Nrf2* and *HES-1*, were predicted to bind to differentially methylated CpG sites located in the miR-29b-2 gene promoter, and further experiments will focus on the specific regulatory mechanism. Third, mutations in any *DNMT* can cause embryonic lethality in mice (Li, et al., 1992; Okano, et al., 1999). To investigate the specific impact of miRNAs on heart development, a heart-targeted miRNA delivery system needs to be developed for in-depth research.

CONCLUSION

These results suggest mutual regulation between miR-29b-3p and *DNMTs* in cardiomyocytes and provide evidence that miRNA-based therapy can normalize the epigenome of cardiomyocytes.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Ethics Committee of Children's Hospital of Fudan University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. The animal study was reviewed and approved by The Ethics Committee of Children's Hospital of Fudan University.

AUTHOR CONTRIBUTIONS

YG and QL designed and supervised the experiments. FaW and QY performed most of the experiments and wrote the manuscript. YM collected and processed the tissue samples. FeW performed the experiments *in vivo* experiments with zebrafish. YZ performed other experiments. XW and YW provided expert technical assistance and analyzed the data. All the authors read and approved the final version of the paper.

FUNDING

This study was supported by grants from the Natural Science Foundation of Shanghai (21ZR1410100), the National Natural Science Foundations of China (81771632) and the National Key Research and Development Program (2016YFC1000500) awarded to QL; grants from the National Natural Science Foundations of China (81741081, 81470442 and 81873481) awarded to YG; and grants from the National Natural Science Foundations of China (82172884) and the Natural Science Foundation of Shanghai (18ZR1404500) awarded to XW.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2022.788799/full#supplementary-material

REFERENCES

Bruneau, B. G. (2008). The Developmental Genetics of Congenital Heart Disease. *Nature* 451 (7181), 943–948. doi:10.1038/nature06801

Cao, K., Li, B., Zhang, Y.-W., Song, H., Chen, Y.-G., Gong, Y.-J., et al. (2021). miR-29b Restrains Cholangiocarcinoma Progression by Relieving DNMT3B-Mediated Repression of CDKN2B Expression. *Aging* 13 (4), 6055–6065. doi:10.18632/aging.202549

Chen, X., Hu, H., Guan, X., Xiong, G., Wang, Y., Wang, K., et al. (2012). CpG Island Methylation Status of miRNAs in Esophageal Squamous Cell Carcinoma. *Int. J. Cancer* 130 (7), 1607–1613. doi:10.1002/ijc.26171

Chen, Y., Luo, J., Tian, R., Sun, H., and Zou, S. (2011). miR-373 Negatively Regulates Methyl-CpG-Binding Domain Protein 2 (MBD2) in Hilar Cholangiocarcinoma. *Dig. Dis. Sci.* 56 (6), 1693–1701. doi:10.1007/s10620-010-1481-1

Dakhlallah, D., Batte, K., Wang, Y., Cantemir-Stone, C. Z., Yan, P., Nuovo, G., et al. (2013). Epigenetic Regulation ofmiR-17~92Contributes to the Pathogenesis of Pulmonary Fibrosis. *Am. J. Respir. Crit. Care Med.* 187 (4), 397–405. doi:10.1164/rccm.201205-0888oc

Dou, P., He, Y., Yu, B., and Duan, J. (2020). Downregulation of microRNA-29b by DNMT3B Decelerates Chondrocyte Apoptosis and the Progression of Osteoarthritis via PTHLH/CDK4/RUNX2 axis. *Aging* 13 (5), 7676–7690. doi:10.18632/aging.103778

Fabbri, M., Garzon, R., Cimmino, A., Liu, Z., Zanesi, N., Callegari, E., et al. (2007). MicroRNA-29 Family Reverts Aberrant Methylation in Lung Cancer by Targeting DNA Methyltransferases 3A and 3B. *Proc. Natl. Acad. Sci. U.S.A.* 104 (40), 15805–15810. doi:10.1073/pnas.0707628104

Fang, X., Poulsen, R. R., Wang-Hu, J., Shi, O., Calvo, N. S., Simmons, C. S., et al. (2016). Knockdown of DNA Methyltransferase 3a Alters Gene Expression and Inhibits Function of Embryonic Cardiomyocytes. *FASEB j.* 30 (9), 3238–3255. doi:10.1096/fj.201600346r

Fernández-Sanlés, A., Sayols-Baixeras, S., Subirana, I., Degano, I. R., and Elosua, R. (2017). Association between DNA Methylation and Coronary Heart Disease or Other Atherosclerotic Events: A Systematic Review. *Atherosclerosis* 263, 325–333. doi:10.1016/j.atherosclerosis.2017.05.022

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 (2), 323–327. doi:10.1016/j.atherosclerosis.2007.11.029

Gallas, A., Alexander, C., Davies, M. C., Puri, S., and Allen, S. (2013). Chemistry and Formulations for siRNA Therapeutics. *Chem. Soc. Rev.* 42 (20), 7983–7997. doi:10.1039/c3cs35520a

Garzon, R., Liu, S., Fabbri, M., Liu, Z., Heaphy, C. E. A., Callegari, E., et al. (2009). MicroRNA-29b Induces Global DNA Hypomethylation and Tumor Suppressor Gene Reexpression in Acute Myeloid Leukemia by Targeting Directly DNMT3A and 3B and Indirectly DNMT1. *Blood* 113 (25), 6411–6418. doi:10.1182/blood-2008-07-170589

Gavrilov, K., and Saltzman, W. M. (2012). Therapeutic siRNA: Principles, Challenges, and Strategies. *Yale J. Biol. Med.* 85 (2), 187–200.

Gilsbach, R., Preissl, S., Grüning, B. A., Schnick, T., Burger, L., Benes, V., et al. (2014). Dynamic DNA Methylation Orchestrates Cardiomyocyte Development, Maturation and Disease. *Nat. Commun.* 5, 5288. doi:10.1038/ncomms6288

Hermsen, S. A. B., van den Brandhof, E.-J., van der Ven, L. T. M., and Piersma, A. H. (2011). Relative Embryotoxicity of Two Classes of Chemicals in a Modified Zebrafish Embryotoxicity Test and Comparison with Their *In Vivo* Potencies. *Toxicol. Vitro* 25 (3), 745–753. doi:10.1016/j.tiv.2011.01.005

Hu, M., Wei, X., Li, M., Tao, L., Wei, L., Zhang, M., et al. (2019). Circular RNA Expression Profiles of Persistent Atrial Fibrillation in Patients with Rheumatic Heart Disease. *Anatol J. Cardiol.* 21 (1), 2–10. doi:10.14744/AnatolJCardiol.2018.35902

Huang, Y.-S., Zhi, Y.-F., and Wang, S.-R. (2009). Hypermethylation of Estrogen Receptor-α Gene in Atheromatosis Patients and its Correlation with Homocysteine. *Pathophysiology* 16 (4), 259–265. doi:10.1016/j.pathophys.2009.02.010

Jing, X., Yang, J., Jiang, L., Chen, J., and Wang, H. (2018). MicroRNA-29b Regulates the Mitochondria-dependent Apoptotic Pathway by Targeting Bax in Doxorubicin Cardiotoxicity. *Cell Physiol Biochem* 48 (2), 692–704. doi:10.1159/000491896

Kasinski, A. L., and Slack, F. J. (2011). MicroRNAs en route to the clinic: progress in validating and targeting microRNAs for cancer therapy. *Nat. Rev. Cancer* 11 (12), 849–864. doi:10.1038/nrc3166

Li, E., Bestor, T. H., and Jaenisch, R. (1992). Targeted Mutation of the DNA Methyltransferase Gene Results in Embryonic Lethality. *Cell* 69 (6), 915–926. doi:10.1016/0092-8674(92)90611-f

Li, K., Zhou, P., Li, S., Zheng, S., and Wang, D. (2021). MicroRNA-29b Reduces Myocardial Ischemia-Reperfusion Injury in Rats via Down-Regulating PTEN and Activating the Akt/eNOS Signaling Pathway. *J. Thromb. Thrombolysis* 53, 123–135. doi:10.1007/s11239-021-02535-y

Li, Z., Yi, N., Chen, R., Meng, Y., Wang, Y., Liu, H., et al. (2020). miR-29b-3p Protects Cardiomyocytes against Endotoxin-Induced Apoptosis and Inflammatory Response through Targeting FOXO3A. *Cell Signal.* 74, 109716. doi:10.1016/j.cellsig.2020.109716

Liu, N., and Olson, E. N. (2010). MicroRNA Regulatory Networks in Cardiovascular Development. *Developmental Cel* 18 (4), 510–525. doi:10.1016/j.devcel.2010.03.010

Liu, R., Gu, J., Jiang, P., Zheng, Y., Liu, X., Jiang, X., et al. (2015). DNMT1-microRNA126 Epigenetic Circuit Contributes to Esophageal Squamous Cell Carcinoma Growth via ADAM9-EGFR-AKT Signaling. *Clin. Cancer Res.* 21 (4), 854–863. doi:10.1158/1078-0432.ccr-14-1740

Lujambio, A., Ropero, S., Ballestar, E., Fraga, M. F., Cerrato, C., Setién, F., et al. (2007). Genetic Unmasking of an Epigenetically Silenced microRNA in Human Cancer Cells. *Cancer Res.* 67 (4), 1424–1429. doi:10.1158/0008-5472.can-06-4218

Meunier, L., Siddeek, B., Vega, A., Lakhdari, N., Inoubli, L., Bellon, R. P., et al. (2012). Perinatal Programming of Adult Rat Germ Cell Death after Exposure to Xenoestrogens: Role of microRNA miR-29 Family in the Down-Regulation of DNA Methyltransferases and Mcl-1. *Endocrinology* 153 (4), 1936–1947. doi:10.1210/en.2011-1109

Morita, S., Horii, T., Kimura, M., Ochiya, T., Tajima, S., and Hatada, I. (2013). miR-29 Represses the Activities of DNA Methyltransferases and DNA Demethylases. *Ijms* 14 (7), 14647–14658. doi:10.3390/ijms140714647

Mott, J. L., Kurita, S., Cazanave, S. C., Bronk, S. F., Werneburg, N. W., and Fernandez-Zapico, M. E. (2010). Transcriptional Suppression of Mir-29b-1/mir-29a Promoter by C-Myc, Hedgehog, and NF-kappaB. *J. Cel. Biochem.* 110 (5), 1155–1164. doi:10.1002/jcb.22630

Nagy, O., Baráth, S., and Ujfalusi, A. (2019). The Role of microRNAs in Congenital Heart Disease. *Ejifcc* 30 (2), 165–178.

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 (3), 247–257. doi:10.1016/s0092-8674(00)81656-6

Pei, Y.-H., Chen, J., Wu, X., He, Y., Qin, W., He, S.-Y., et al. (2020). LncRNA PEAMIR Inhibits Apoptosis and Inflammatory Response in PM2.5 Exposure Aggravated Myocardial Ischemia/reperfusion Injury as a Competing Endogenous RNA of miR-29b-3p. *Nanotoxicology* 14 (5), 638–653. doi:10.1080/17435390.2020.1731857

Pereira, P., Queiroz, J. A., Figueiras, A., and Sousa, F. (2017). Current Progress on microRNAs-Based Therapeutics in Neurodegenerative Diseases. *Wiley Interdiscip. Rev. RNA* 8 (3). doi:10.1002/wrna.1409

Qiu, F., Tong, H., Wang, Y., Tao, J., Wang, H., and Chen, L. (2018). Inhibition of miR-21-5p Suppresses High Glucose-Induced Proliferation and Angiogenesis of Human Retinal Microvascular Endothelial Cells by the Regulation of AKT and ERK Pathways via Maspin. *Biosci. Biotechnol. Biochem.* 82 (8), 1366–1376. doi:10.1080/09168451.2018.1459179

Ren, D., Li, F., Gao, A., Cao, Q., Liu, Y., and Zhang, J. (2020). RETRACTED ARTICLE: Hypoxia-Induced Apoptosis of Cardiomyocytes Is Restricted by Ginkgolide B-Downregulated microRNA-29. *Cell Cycle* 19 (10), 1067–1076. doi:10.1080/15384101.2020.1731651

Rideout, W. M., 3rd, Eggan, K., and Jaenisch, R. (2001). Nuclear Cloning and Epigenetic Reprogramming of the Genome. *Science* 293 (5532), 1093–1098. doi:10.1126/science.1063206

Roman-Gomez, J., Agirre, X., Jiménez-Velasco, A., Arqueros, V., Vilas-Zornoza, A., Rodriguez-Otero, P., et al. (2009). Epigenetic Regulation of microRNAs in Acute Lymphoblastic Leukemia. *Jco* 27 (8), 1316–1322. doi:10.1200/jco.2008.19.3441

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

Sheng, W., Qian, Y., Wang, H., Ma, X., Zhang, P., Diao, L., et al. (2013). DNA Methylation Status of NKX2-5, GATA4 and HAND1in Patients with Tetralogy of Fallot. *BMC Med. Genomics* 6, 46. doi:10.1186/1755-8794-6-46

Sheng, W., Qian, Y., Zhang, P., Wu, Y., Wang, H., Ma, X., et al. (2014). Association of Promoter Methylation Statuses of Congenital Heart Defect Candidate Genes with Tetralogy of Fallot. *J. Transl Med.* 12, 31. doi:10.1186/1479-5876-12-31

Sheng, W., Wang, H., Ma, X., Qian, Y., Zhang, P., Wu, Y., et al. (2012). LINE-1 Methylation Status and its Association with Tetralogy of Fallot in Infants. *BMC Med. Genomics* 5, 20. doi:10.1186/1755-8794-5-20

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 (3), Cr149–155.

Suzuki, H., Yamamoto, E., Nojima, M., Kai, M., Yamano, H.-o., Yoshikawa, K., et al. (2010). Methylation-associated Silencing of microRNA-34b/c in Gastric Cancer and its Involvement in an Epigenetic Field Defect. *Carcinogenesis* 31 (12), 2066–2073. doi:10.1093/carcin/bgq203

Wada, R., Akiyama, Y., Hashimoto, Y., Fukamachi, H., and Yuasa, Y. (2010). miR-212 Is Downregulated and Suppresses Methyl-CpG-Binding Protein MeCP2 in Human Gastric Cancer. *Int. J. Cancer* 127 (5), 1106–1114. doi:10.1002/ijc.25126

Wang, S., Wu, W., and Claret, F. X. (2017). Mutual Regulation of microRNAs and DNA Methylation in Human Cancers. *Epigenetics* 12 (3), 187–197. doi:10.1080/15592294.2016.1273308

Yan, F., Shen, N., Pang, J., Xie, D., Deng, B., Molina, J. R., et al. (2014). Restoration of miR-101 Suppresses Lung Tumorigenesis through Inhibition of DNMT3a-dependent DNA Methylation. *Cell Death Dis* 5 (9), e1413. doi:10.1038/cddis.2014.380

Yang, Q., Wu, F., Mi, Y., Wang, F., Cai, K., Yang, X., et al. (2020). Aberrant Expression of miR-29b-3p Influences Heart Development and Cardiomyocyte Proliferation by Targeting NOTCH2. *Cell Prolif* 53 (3), e12764. doi:10.1111/cpr.12764

Yang, Q., Wu, F., Wang, F., Cai, K., Zhang, Y., Sun, Q., et al. (2019). Impact of DNA Methyltransferase Inhibitor 5-azacytidine on Cardiac Development of Zebrafish *In Vivo* and Cardiomyocyte Proliferation, Apoptosis, and the Homeostasis of Gene Expression *In Vitro*. *J. Cell. Biochem.* 120 (10), 17459–17471. doi:10.1002/jcb.29010

Ying, A. (2000). Methylation of the Estrogen Receptor-α Gene Promoter Is Selectively Increased in Proliferating Human Aortic Smooth Muscle Cells. *Cardiovasc. Res.* 46 (1), 172–179. doi:10.1016/s0008-6363(00)00004-3

Zhang, H., Li, H., Ge, A., Guo, E., Liu, S., and Zhang, L. (2018a). RETRACTED: Long Non-coding RNA TUG1 Inhibits Apoptosis and Inflammatory Response in LPS-Treated H9c2 Cells by Down-Regulation of miR-29b. *Biomed. Pharmacother.* 101, 663–669. doi:10.1016/j.biopha.2018.02.129

Zhang, J., Chang, J.-J., Xu, F., Ma, X.-J., Wu, Y., Li, W.-C., et al. (2013a). MicroRNA Deregulation in Right Ventricular Outflow Tract Myocardium in Nonsyndromic Tetralogy of Fallot. *Can. J. Cardiol.* 29 (12), 1695–1703. doi:10.1016/j.cjca.2013.07.002

Zhang, P., Huang, B., Xu, X., and Sessa, W. C. (2013b). Ten-eleven Translocation (Tet) and Thymine DNA Glycosylase (TDG), Components of the Demethylation Pathway, Are Direct Targets of miRNA-29a. *Biochem. Biophysical Res. Commun.* 437 (3), 368–373. doi:10.1016/j.bbrc.2013.06.082

Zhang, T. J., Xu, Z. J., Gu, Y., Wen, X. M., Ma, J. C., Zhang, W., et al. (2020). Identification and Validation of Prognosis-Related DLX5 Methylation as an Epigenetic Driver in Myeloid Neoplasms. *Clin. Transl Med.* 10 (2), e29. doi:10.1002/ctm2.29

Zhang, Z., Cao, Y., Zhai, Y., Ma, X., An, X., Zhang, S., et al. (2018b). MicroRNA-29b Regulates DNA Methylation by Targeting Dnmt3a/3b and Tet1/2/3 in Porcine Early Embryo Development. *Develop. Growth Differ.* 60 (4), 197–204. doi:10.1111/dgd.12537

Zhao, S., Wang, Y., Liang, Y., Zhao, M., Long, H., Ding, S., et al. (2011). MicroRNA-126 Regulates DNA Methylation in CD4+ T Cells and Contributes to Systemic Lupus Erythematosus by Targeting DNA Methyltransferase 1. *Arthritis Rheum.* 63 (5), 1376–1386. doi:10.1002/art.30196

Zhong, J., Agha, G., and Baccarelli, A. A. (2016). The Role of DNA Methylation in Cardiovascular Risk and Disease. *Circ. Res.* 118 (1), 119–131. doi:10.1161/circresaha.115.305206

Zhu, S., Cao, L., Zhu, J., Kong, L., Jin, J., Qian, L., et al. (2013). Identification of Maternal Serum microRNAs as Novel Non-invasive Biomarkers for Prenatal Detection of Fetal Congenital Heart Defects. *Clinica Chim. Acta* 424, 66–72. doi:10.1016/j.cca.2013.05.010

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Cell and Developmental Biology

ORIGINAL RESEARCH
published: 03 May 2022
doi: 10.3389/fcell.2022.885568

M5C-Related lncRNA Predicts Lung Adenocarcinoma and Tumor Microenvironment Remodeling: Computational Biology and Basic Science

Ming Bai[1] *and Chen Sun*[2]*

[1]Department of Medical Oncology, The First Hospital of China Medical University, Shenyang, China, [2]Department of Radiology, Shengjing Hospital of China Medical University, Shenyang, China

Purpose: Epigenetic RNA modification regulates gene expression post-transcriptionally. The aim of this study was to construct a prognostic risk model for lung adenocarcinoma (LUAD) using long non-coding RNAs (lncRNAs) related to m5C RNA methylation.

Method: The lncRNAs regulated by m5C methyltransferase were identified in TCGA-LUAD dataset using Pearson correlation analysis (coefficient > 0.4), and clustered using non-negative matrix decomposition. The co-expressing gene modules were identified by WGCNA and functionally annotated. The prognostically relevant lncRNAs were screened by LASSO regression and a risk model was constructed. LINC00628 was silenced in the NCI-H460 and NCI-H1299 cell lines using siRNA constructs, and migration and invasion were assessed by the Transwell and wound healing assays respectively.

Results: We identified 185 m5C methyltransferase-related lncRNAs in LUAD, of which 16 were significantly associated with overall survival. The lncRNAs were grouped into two clusters on the basis of m5C pattern, and were associated with significant differences in overall and disease-free survival. GSVA revealed a close relationship among m5C score, ribosomes, endolysosomes and lymphocyte migration. Using LASSO regression, we constructed a prognostic signature consisting of LINC00628, LINC02147, and MIR34AHG. The m5C-lncRNA signature score was closely related to overall survival, and the accuracy of the predictive model was verified by the receiver operating characteristic curve and decision curve analysis. Knocking down LINC00628 in NCI-H460 and NCI-H1299 cells significantly reduced their migration and invasion compared to that of control cells.

Conclusion: We constructed a prognostic risk model of LUAD using three lncRNAs regulated by m5C methyltransferase, which has potential clinical value.

Keywords: TCGA-LUAD, M5C methylation, long non-coding RNA, tumor microenvironment remodeling, cell migration

OPEN ACCESS

Edited by:
Chunjie Jiang,
University of Pennsylvania,
United States

Reviewed by:
Qinglan Li,
University of Pennsylvania,
United States
Qifei Wang,
Dalian Medical University, China

***Correspondence:**
Chen Sun
sunchen@cmu.edu.cn

Specialty section:
This article was submitted to Molecular and Cellular Pathology, a section of the journal Frontiers in Cell and Developmental Biology

Received: *28 February 2022*
Accepted: *18 April 2022*
Published: *03 May 2022*

Citation:
Bai M and Sun C (2022) M5C-Related lncRNA Predicts Lung Adenocarcinoma and Tumor Microenvironment Remodeling: Computational Biology and Basic Science. Front. Cell Dev. Biol. 10:885568. doi: 10.3389/fcell.2022.885568

Abbreviations: GSEA, gene set enrichment analysis; LASSO, least absolute shrinkage and selection operator; LUAD, lung adenocarcinoma; m5C, 5-methylcytosine RNA methyltransferases; NMF, non-negative matrix factorization; WGCNA, weighted correlation network analysis.

INTRODUCTION

Each year, 1.8 million people are diagnosed with lung cancer worldwide and about 1.6 million eventually die, indicating a very high mortality rate. Lung cancer is associated with the highest morbidity and mortality rates in China as well, with more than 800,000 new cases every year (Bray et al., 2018). Despite advances in radical surgery, radiotherapy, chemotherapy, targeted therapy, immunotherapy etc., local recurrence and distant metastasis still cannot be achieved (Bray et al., 2018). Lung adenocarcinoma (LUAD) and lung squamous cell carcinoma are the most common types of non-small cell lung carcinoma (NSCLC), of which LUAD accounts for about 70% of all NSCLC cases (Siegel et al., 2020), and is associated with high mortality and recurrence rates (Li et al., 2014). Given the limited understanding of the pathophysiology of LUAD, there is a paucity of effective prognostic indicators (Müller et al., 2016). Although therapies targeting EGFR, TP53, AKT1, KRAS, and PTEN, which frequently undergo mutations and copy number changes in LUAD, have been widely applied in patients with advanced lung cancer (Brose et al., 2002; Bean et al., 2007; Bleeker et al., 2008; Jin et al., 2010), their clinical potential is still limited (Murayama et al., 2016; Schneider et al., 2016). Therefore, there is an urgent need to identify more effective therapeutic targets in order to improve patient outcomes. At present, with the development of high-throughput sequencing and next-generation sequencing, we can get the human genome by gene chip 19000 protein-coding genes of somatic mutation data and copy number amplification; this gives us a more comprehensive understanding of the pathogenesis of lung adenocarcinoma, development disease-related biomarkers provide technical support (Bejjani and Shaffer, 2008).

Epigenetic modifications, including DNA and RNA methylation, genomic imprinting, gene silencing and non-coding RNA activities, regulate gene expression at the transcriptional level (Kaliman, 2019), and are thus involved in multiple pathological processes, including tumorigenesis (Kanwal et al., 2015). Studies show that N6-methyladenine (m6A) and 5-methylcytosine (m5C) RNA methylation play crucial roles in tumor development and progression (He et al., 2019; Ma et al., 2019). High-throughput sequencing has revealed that RNA m5C methylation can modify the sequences of both coding and non-coding RNAs (Chellamuthu and Gray, 2020). The methyltransferase complex that catalyzes DNA/RNA methylation consists of a methyltransferase ("writer"), a methylase demethylase ("eraser"), and an m5C binding protein ("reader ") (Nombela et al., 2021). There is evidence that the expression levels of m5C-related genes are correlated to the prognosis of lung and pancreatic cancers, indicating that m5C methylation influences tumor growth (Pan et al., 2021; Yu et al., 2021).

Long non-coding RNAs (lncRNAs) are more than 200 nucleotides in length (Qian et al., 2019), and are involved in epigenetic processes such as gene silencing, histone processing, transcriptional regulation and transcriptional interference. Several lncRNAs have been identified in recent years that are involved tumor formation and progression (Bridges et al., 2021). In addition, methylation-related genes affect tumor cell proliferation by regulating the methylation level of specific lncRNAs. For instance, the methyltransferase METTL14 promotes breast tumor development by regulating LINC00942 and its downstream targets (Sun et al., 2020). Nevertheless, little is known regarding the correlation between non-coding RNAs and m5C methylation in LUAD.

In the present study, we used computational biology to identify the lncRNAs regulated by m5C methyltransferase in LUAD, and analyzed the biological functions and pathways associated with the prognostically relevant lncRNAs.

METHODS

Data Collection

The clinical and transcriptomic data of 576 LUAD patients were obtained from TCGA (https://cancergenome.nih.gov/). The clinical data included gender, survival status, survival time, tumor stage and TNM stage.

Negative Matrix Factorization Clustering of m5C-lncRNA Gene Set

Thirteen m5C-related genes encoding for lncRNAs were retrieved from literature mining, including NOP2, NSUN2, NSUN3, NSUN4, NSUN5, NSUN7, TRDMT1, TET1, TET2, TET3, ALKBH1, YBX1, and ALYREF (Archer et al., 2016; Blanco et al., 2016; Müller et al., 2016; Cheng et al., 2018; García et al., 2018; Li et al., 2018; Chen et al., 2019; Gao et al., 2019; Janin et al., 2019; Carella et al., 2020; Mei et al., 2020; Sato et al., 2020). After excluding those with median absolute difference <0.5, the correlation of the remaining candidate genes with overall survival was analyzed by the Cox regression model using the "survival" package. The genes with an absolute median >0.5 and $p < 0.05$ were used for NMF dimensionality reduction using the "NMF" package in R (Gaujoux and Seoighe, 2010).

Weighted Gene Co-expression Network Analysis

A weighted co-expression network was constructed using the WGCNA package in R (Langfelder and Horvath, 2008). PickSoftThreshold was applied to calculate the optimal value of the adjacent function weighting parameter, which was then used as a soft threshold for subsequent network construction. Following construction of a weighted adjacency matrix, the modules of related genes were identified based on hierarchical clustering of the dissimilarity measure (1-TOM) of the topological overlap matrix (Ravasz et al., 2002). The significance of the modules and the mean gene significance within each module were calculated. Finally, the correlation between the co-expression modules and the expression

patterns of the resulting subtypes of NMF clustering were calculated.

Functional Enrichment

The co-expressed module genes were functionally annotated by the GO enrichment analysis and KEGG signaling pathway analysis using the "cluster profile" (Wu et al., 2021).

Construction of m5C lncRNA Risk Model

The genes significantly associated with the overall survival of LUAD patients were identified by univariate Cox regression

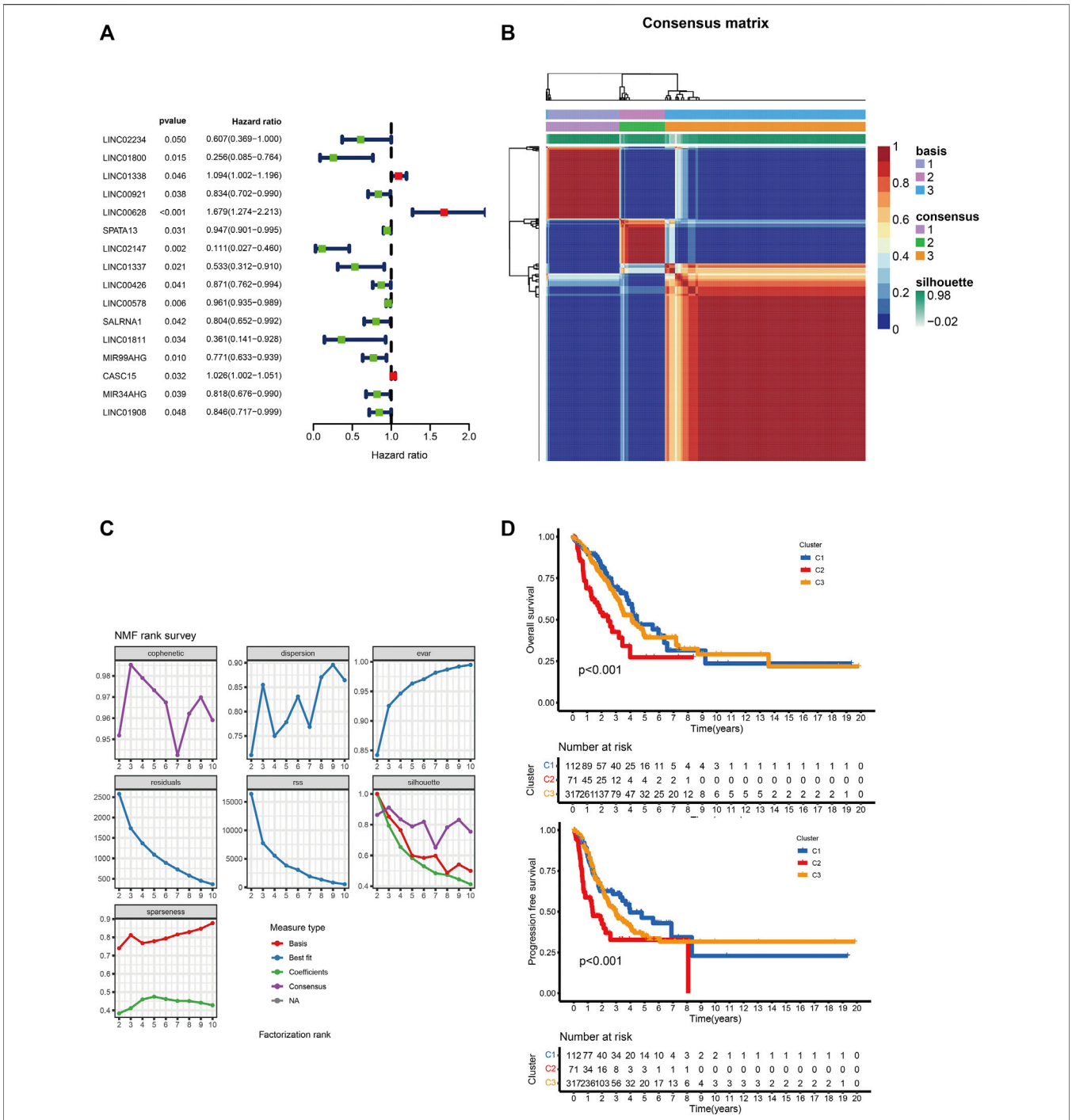


FIGURE 1 | (A) Univariate Cox regression analysis of 16 lncRNAs and overall survival. **(B)** Consensus map of NMF clustering. **(C)**Consensus clustering parameters. **(D)** Overall and disease-free survival of C1, C2 and C3 clusters.

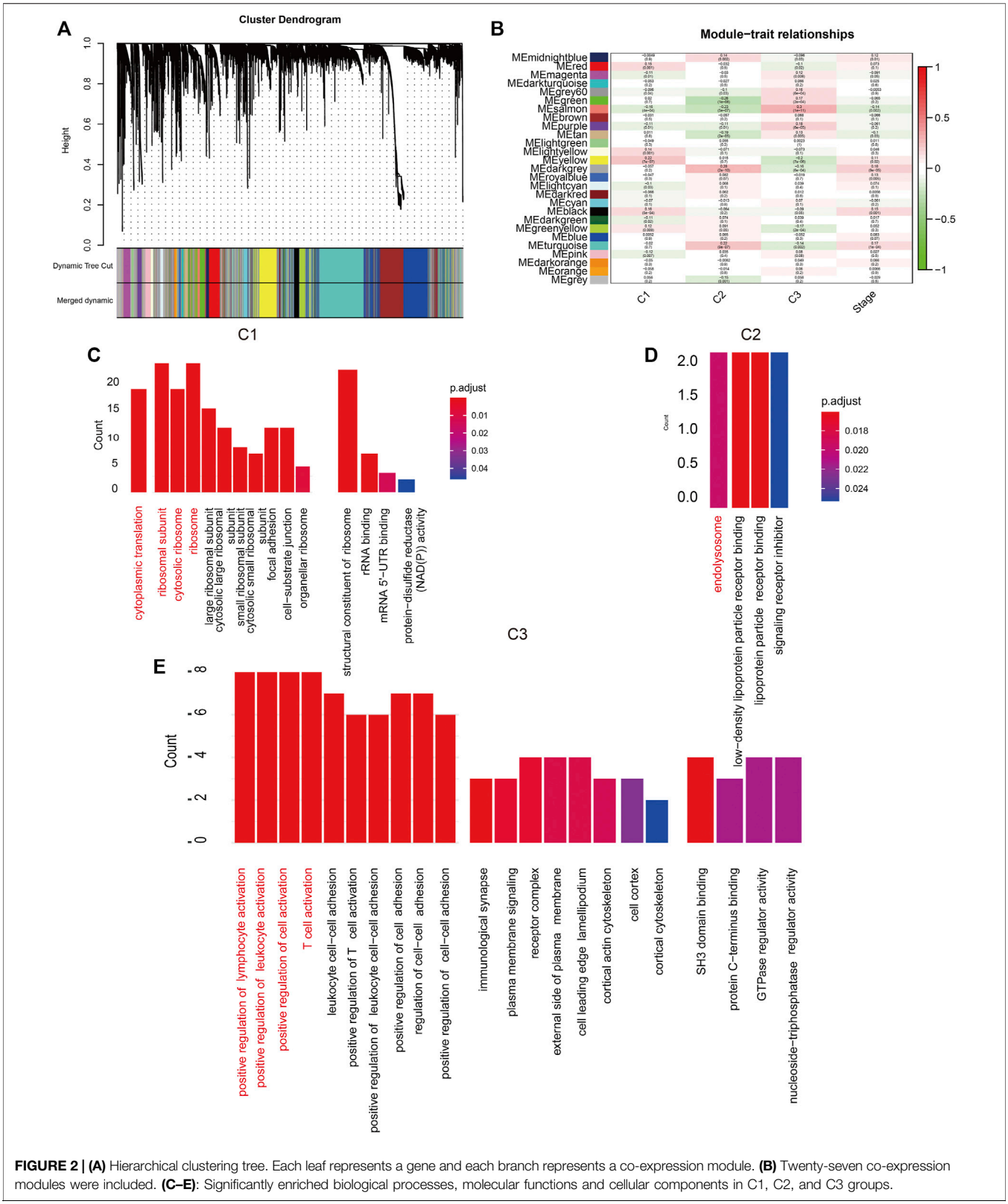


FIGURE 2 | (A) Hierarchical clustering tree. Each leaf represents a gene and each branch represents a co-expression module. **(B)** Twenty-seven co-expression modules were included. **(C–E)**: Significantly enriched biological processes, molecular functions and cellular components in C1, C2, and C3 groups.

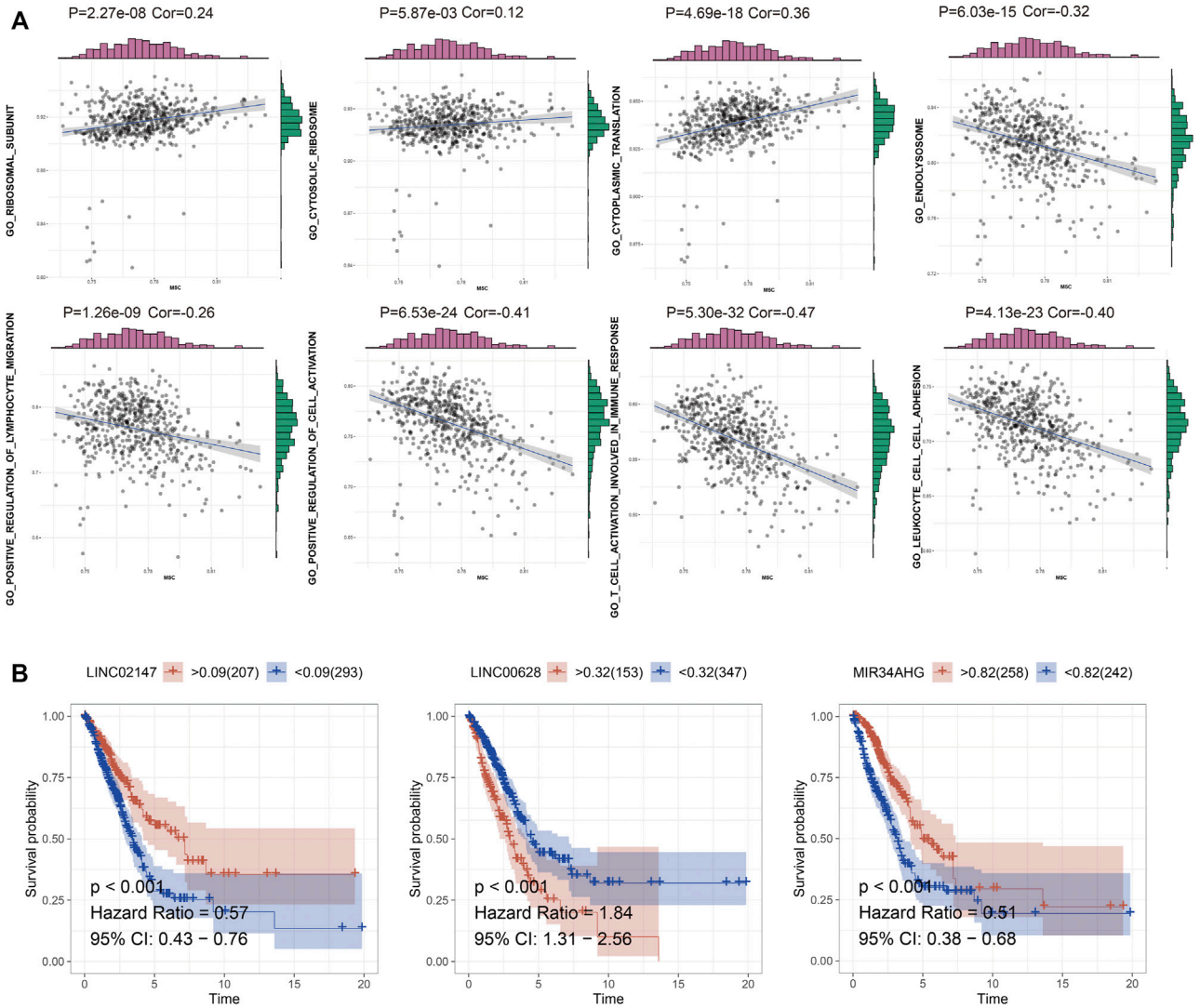


FIGURE 3 | (A) GSVA of m5C RNA methylation dataset and enriched functions in C1, C2, and C3. **(B)** The survival curve of the LINC02147, LINC00628, and MIR34AHG.

analysis ($p < 0.05$) using the LASSO regression algorithm and a risk score model was constructed (Tibshirani, 1997).

GSVA

GSVA was used to assess the correlation of different gene set scores with m5C methyltransferase-related scores.

Immune Infiltration Analysis

The relative proportion of immune-infiltrating cells in the two risk groups was analyzed using the CIBERSORT (Newman et al., 2019; Steen et al., 2020), EPIC (Racle et al., 2017), quanTIseq (Finotello et al., 2019), MCPcounter (Becht et al., 2016), XCELL (Aran et al., 2017), and 和TIMER (Li et al., 2016) algorithms. The differences in immune responses were visualized through heat maps.

Cell Culture

NCI-H460 and NCI-H1299 cell lines were provided by the Shanghai Cell Bank of the Chinese Academy of Sciences. The cells were cultured in DMEM supplemented with 10% fetal bovine serum at 37°C in a 5% CO_2 incubator. The cells were seeded in a 6-well culture plate at the density of 4×10^5 cells per well and cultured overnight for subsequent experiments.

Cell Transduction

NCI-H460 and NCI-H1299 cells in the logarithmic growth phase were harvested, seeded in 6-well plates, and transduced with lentiviruses expressing si-NC, si1-LINC00628 or si2-LINC00628 at the multiplicity of infection of 10. After 24 h of culture, 2 μL polybrene was added to a final concentration of 5 μg/ml for 1–2 weeks, and the medium was changed every

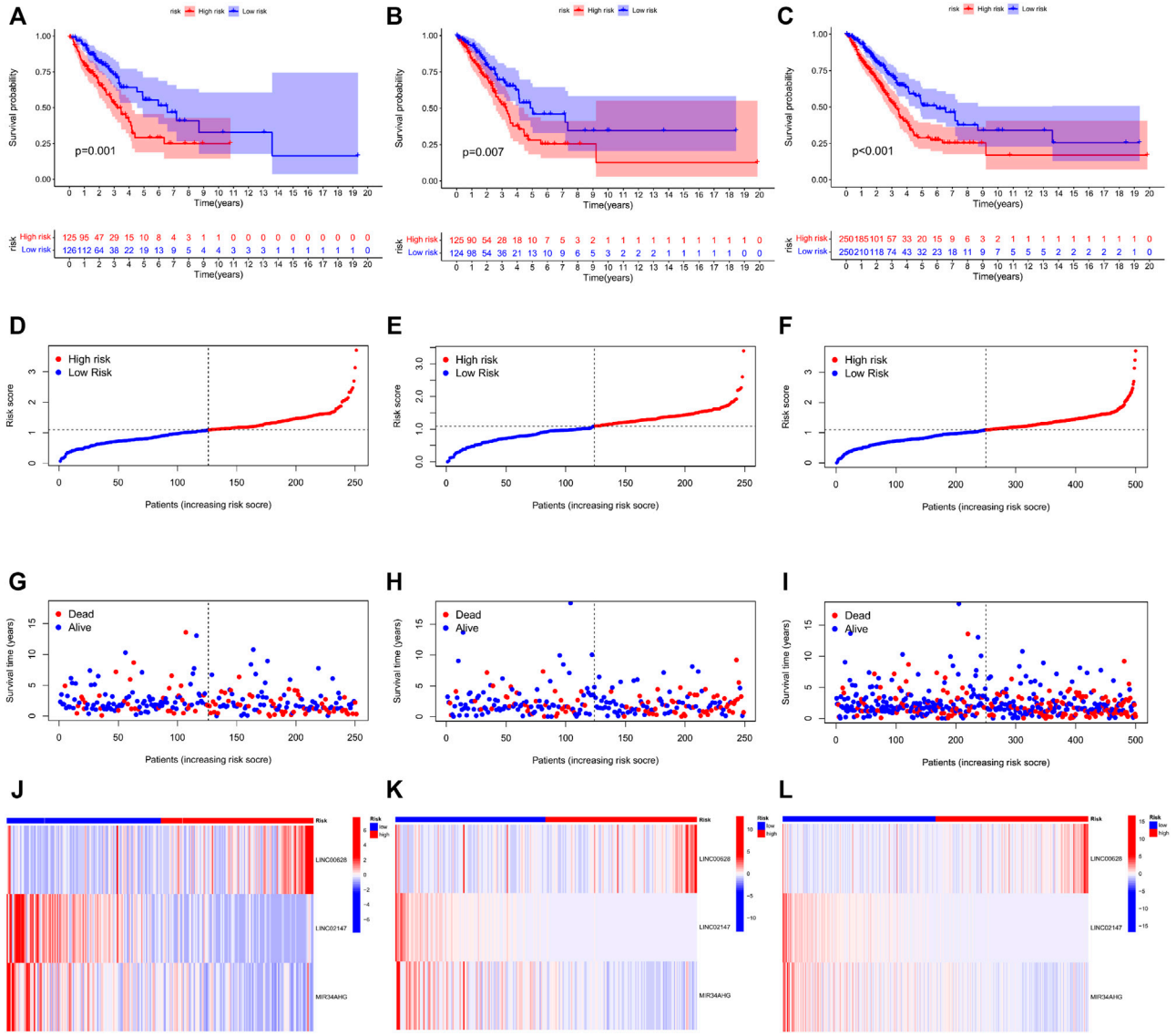


FIGURE 4 | Risk score distribution and survival status in the training, validation and total TCGA-LUAD sets. The survival status, risk score, gene expression level in the risk score and follow-up time of each sample are shown.

8–12 h. The stably transduced cells expressing GFP were detected 72–96 later under a fluorescence microscope, and expanded further.

RT-PCR

LINC00628 silencing in the transduced cells was analyzed by RT-PCR using the following primers: Forward—5′-CAGTGGGGAACTCTGACTCG-3′ and Reverse—5′-GTGCCTGGTGCTCTCTTACC-3′.

Wound Healing and Transwell Assays

The *in vitro* migration of the control and LINC00628-knockdown cell lines were determined by the wound healing and Transwell assays respectively.

RESULTS

Identification of M5C-Related lncRNA Molecular Subtypes in LUAD Based on NMF Classification

We identified 185 M5C-related lncRNAs with Pearson correlation coefficients greater than 0.4 (**Supplementary Table S1**), of which 16 were significantly associated with LUAD prognosis ($p < 0.05$; **Figure 1A**). NMF clustering was then performed on these lncRNA-related genes with 50 iterations, which identified nine clusters. The number of collections (k) was 2–10, and the minimum sample of each group was set to 10. According to cophenetic, dispersion and

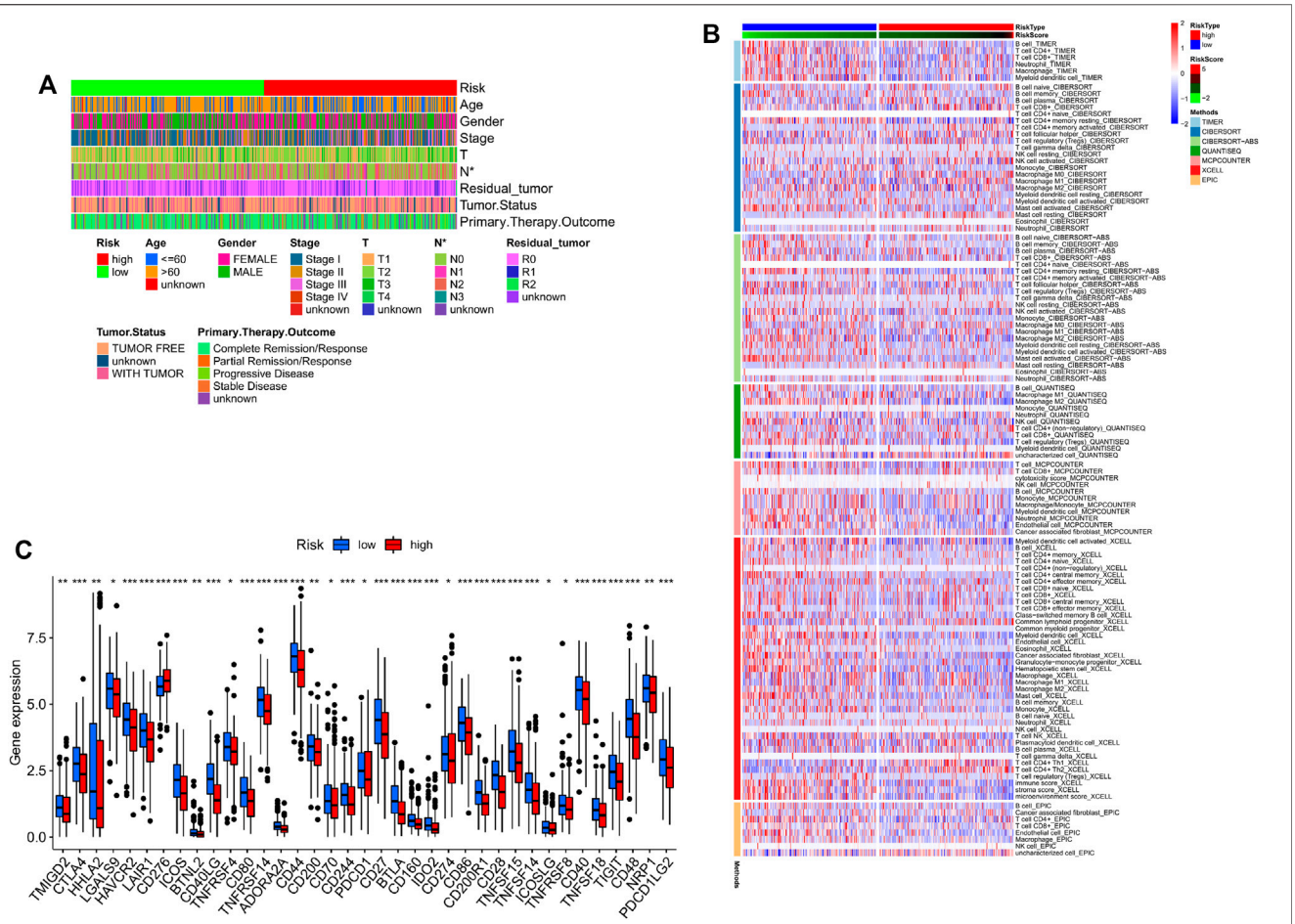


FIGURE 5 | (A) Clinical phenotypes of patients in the two risk groups. **(B)** Immune infiltration in the high- and low-risk groups. **(C)** Immune checkpoint expression levels in the two risk score groups.

silhouette, we selected the ideal clustering group as 3 (**Figures 1B,C**). The grouping details are shown in **Supplementary Table S2**. The molecular subtypes of lncRNAs based on m5C methylation patterns was associated with significant differences in overall and disease-free survival (**Figure 1D**; log-rank $p < 0.05$).

Biological Characteristics of the lncRNA-Related Genes

WCGNA of the lncRNA-related genes revealed 27 co-expression modules with β value set to 3, and genes in the smallest module set to 30 (**Figure 2A**). The C1 cluster had the strongest correlation with the yellow module (**Figure 2B**; Cor = 0.2). Functional annotation of the top 20 genes most closely related to C1 in the yellow module indicated that these genes are associated with ribosomal subunits (**Figure 2C**). The C2 cluster was strongly correlated with the black-gray module (**Figure 2B**; Cor = 0.28), and the top 20 genes were enriched in lysosome-related functions (**Figure 2D**). The C3 cluster showed significant correlation with the salmon color module (**Figure 2B**; Cor = 0.3), and was mainly associated with lymphocyte regulation (**Figure 2E**). To directly demonstrate the relationship between m5C and the biological function of each expression pattern signature, we performed GSVA for each gene set and obtained the scores for m5C methyltransferase. As shown in **Figure 3A**, m5C RNA methyltransferase, ribosome subunits, lysosomes, and lymphocyte migration were closely associated.

Construction of a LUAD Predictive Model Based on lncRNA-Related Predictive Genes

TCGA-LUAD cohort was randomly divided into the training and validation sets. Briefly, the data samples were sorted in the ascending order by ID and random numbers were assigned to each sample using SPSS. Both sets were similar in terms of age, clinical-stage, follow-up time and survival, as well as the gene expression profiles (**Supplementary Table S3**). LASSO regression was used to construct the predictive model by incorporating the 16 prognostic genes and overall survival rates. LINC00628, LINC02147, and MIR34AHG were identified as the independent prognostic factors and used for the final predictive risk model. The risk score for each sample was

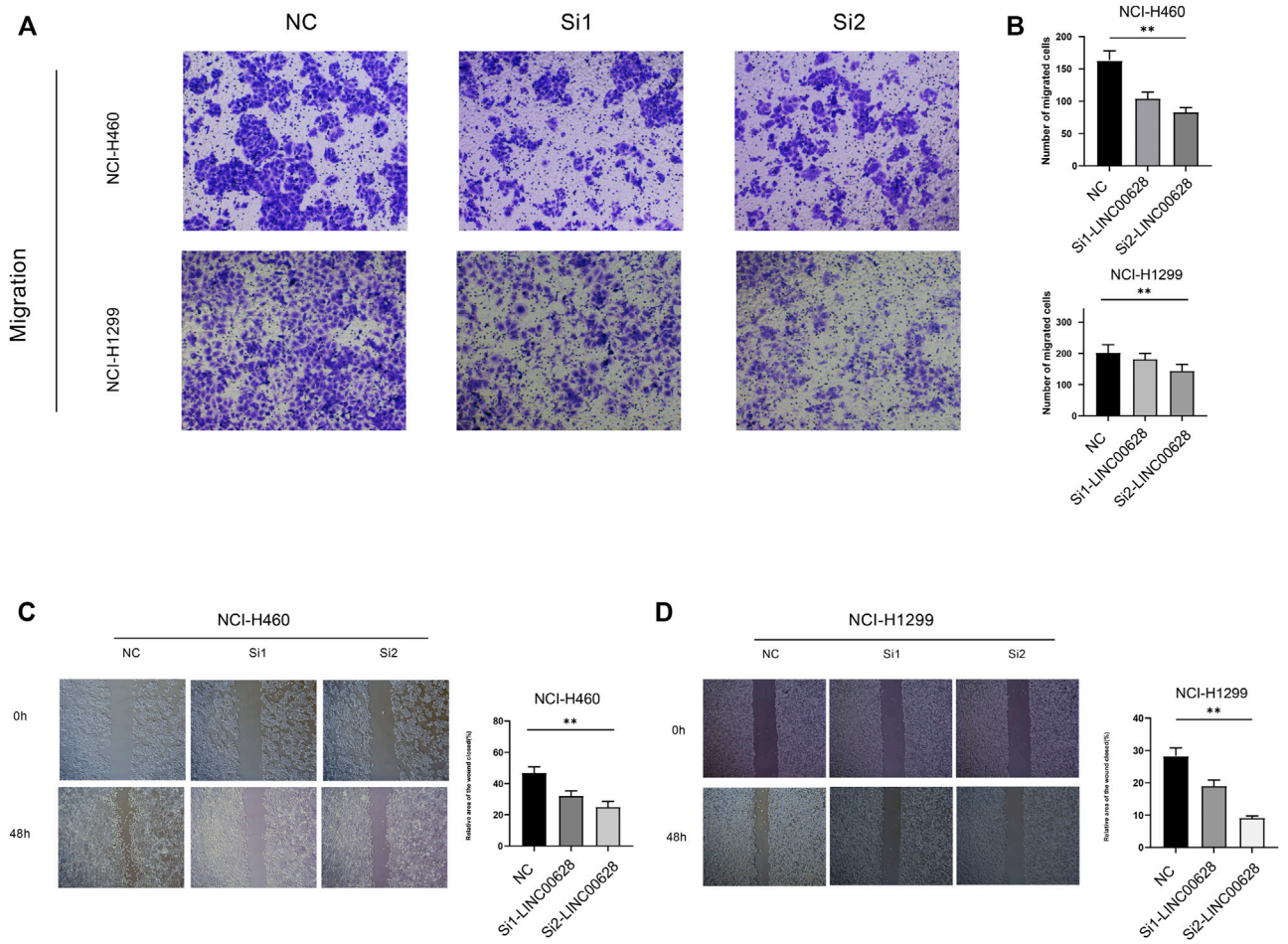


FIGURE 6 | (A,B) Transwell assay of si1-LINC00628 and si2- LINC00628 groups. **(C,D)** Wound healing assay of si1-LINC00628 and si2-LINC00628 groups. *p < 0.5, **p < 0.01, ***p < 0.001, ****p < 0.0001, ns, no significance.

calculated as 0.45 * exp LINC00628—2.67*expLINC02147—0.30 * exp MIR34AHG. The survival curves for each independent prognostic factor are shown in **Figure 3B**. The patients in the training and validation sets were classified into the high- and low-risk groups based on the risk scores, which showed significant differences in survival in both sets (**Figure 4**; $p < 0.05$).

Correlation of Risk Scores With Immune Checkpoints and Immune Infiltration

The LUAD patients were divided into high- and low-risk groups according to the median risk score, and their clinical characteristics are summarized in **Figure 5A**. Immune cell infiltration was assessed using CIBERSORT, EPIC, quanTIseq, MCPcounter, XCELL and TIMER programs. As shown in the heatmap in **Figure 5B**, the proportion of infiltrating macrophages was lower in the high-risk group. In addition, the infiltration levels of $CD8^+$ T cells and M1/M2 macrophages were also significantly different between the two risk groups, as were the expression levels of immune checkpoints such as CTAL4 and CD276. This finding suggests that risk scores can be used as immunotherapy biomarkers for LUAD (**Figure 5C**).

LINC006328 Regulates Migration of LUAD Cells

To further determine the biological significance of LINC00628 in LUAD, the gene was silenced in the NCI-H460 and NCI-H1299 cell lines using two specific siRNA constructs. The LINC006328-knockdown cells showed significantly reduced invasion ($p < 0.05$; **Figures 6A,B**) and migration ($p < 0.05$; **Figures 6C,D**) rates compared to the control groups in the Transwell and wound healing assays respectively. Thus, LINC00628 likely functions as an oncogene in LUAD and promotes tumor cell migration and invasion.

DISCUSSION

Although surgery, radiotherapy and chemotherapy have prolonged the survival of LUAD patients, the prognosis is far from satisfactory. Early diagnostic markers for LUAD are lacking due to incomplete understanding of its pathological basis. Therefore, it is crucial to explore the genetic and epigenetic factors involved in LUAD in order to identify novel therapeutic targets and diagnostic biomarkers (Sung et al., 2021).

RNA methylation is a critical epigenetic modification involved in post-transcriptional gene regulation, and includes m6A, m1A, m5C, m7G, and other types. Methylation of the fifth cytosine (m5C) is particularly widespread (Traube and Carell, 2017; Mongan et al., 2019; Xie et al., 2020), and participates in various physiological and pathological processes (Li et al., 2017; Zhao et al., 2017; Bohnsack et al., 2019). The m5C modification in tRNA and rRNA regulates translation and the quality of ribosome biosynthesis respectively. In addition, methylation of 5C in mRNA affects its structure, stability and translation.

LncRNAs consist of more than 200 nucleotides and were initially considered "junk sequences" with no specific biological functions. However, recent studies show that lncRNAs are widely expressed in human cells and are associated with tumor development. In fact, several lncRNAs have been identified as potential prognostic markers and therapeutic targets for multiple tumors (Nandwani et al., 2021). In addition to post-transcriptional regulation of protein-coding RNAs, the lncRNAs can also bind to proteins and molecular scaffolds and affect tumor growth through *in situ* regulation and molecular convergence (Kopp and Mendell, 2018; Nandwani et al., 2021). In the present study, we used computational biology to identify lncRNAs that are regulated by m5C methyltransferase, and determined their prognostic relevance in LUAD. Bioinformatics analysis further indicated that the m5C scores of these lncRNAs were correlated with ribosome subunit, cytosolic ribosome, cytoplasmic translation, endolysosome and lymphocyte migration.

We identified 185 lncRNAs related to genes encoding m5C methyltransferase, of which 16 were prognostically relevant, including LINC00628 that was significantly correlated with multiple m5C methyltransferases. Therefore, we hypothesized that LINC00628 is regulated by m5C methylation during LUAD progression. Indeed, knocking down LINC00628 in two LIAD cell lines significantly reduced their migration and invasion rates *in vitro*, which is suggestive of an oncogenic role in LUAD. However, studies show that LINC00628 can function as an oncogene or tumor suppressor in different cancers (Zhang et al., 2016; He et al., 2018). For example, overexpression of LINC00628 inhibited the proliferation and migration of osteosarcoma cells (He et al., 2018), whereas its knockdown in gastric cells had a similar inhibitory effect (Zhang et al., 2016). Consistent with our results, Xu et al. found that LINC00628 promoted LUAD progression by targeting the LAMA3 promoter region (Xu et al., 2019). In addition, we also identified LNC02147 as a protective factor in LUAD.

However, a major limitation of our study is that our findings are based on bioinformatics analysis, and will have to be validated on cross-cohort samples. In addition, the mechanisms underlying the function of m5C-related lncRNAs in LUAD also need to be explored.

CONCLUSION

We identified prognostically relevant LUAD-related lncRNAs that are regulated by m5C methyltransferase, and constructed a predictive model based on these lncRNAs. Our findings provide a basis for further research on the role of m5C modification in LUAD.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The datasets TCGA-LUAD for this study can be found at http://cancergenome.nih.gov/.

AUTHOR CONTRIBUTIONS

MB wrote the manuscript. CS revised the manuscript.

FUNDING

This work was supported by grants from the Doctor Research Initiation Program of Liaoning Province (2020-BS- 094).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2022.885568/full#supplementary-material

REFERENCES

Aran, D., Hu, Z., and Butte, A. J. (2017). xCell: Digitally Portraying the Tissue Cellular Heterogeneity Landscape. *Genome Biol.* 18, 220. doi:10.1186/s13059-017-1349-1

Archer, N. P., Perez-Andreu, V., Scheurer, M. E., Rabin, K. R., Peckham-Gregory, E. C., Plon, S. E., et al. (2016). Family-based Exome-wide Assessment of Maternal Genetic Effects on Susceptibility to Childhood B-Cell Acute Lymphoblastic Leukemia in Hispanics. *Cancer* 122, 3697–3704. doi:10.1002/cncr.30241

Bean, J., Brennan, C., Shih, J.-Y., Riely, G., Viale, A., Wang, L., et al. (2007). MET Amplification Occurs with or without T790M Mutations in EGFR Mutant Lung Tumors with Acquired Resistance to Gefitinib or Erlotinib. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20932–20937. doi:10.1073/pnas.0710370104

Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the Population Abundance of Tissue-Infiltrating Immune

and Stromal Cell Populations Using Gene Expression. *Genome Biol.* 17, 218. doi:10.1186/s13059-016-1070-5

Bejjani, B. A., and Shaffer, L. G. (2008). Clinical Utility of Contemporary Molecular Cytogenetics. *Annu. Rev. Genom. Hum. Genet.* 9, 71–86. doi:10.1146/annurev.genom.9.081307.164207

Blanco, S., Bandiera, R., Popis, M., Hussain, S., Lombard, P., Aleksic, J., et al. (2016). Stem Cell Function and Stress Response Are Controlled by Protein Synthesis. *Nature* 534, 335–340. doi:10.1038/nature18282

Bleeker, F. E., Felicioni, L., Buttitta, F., Lamba, S., Cardone, L., Rodolfo, M., et al. (2008). AKT1E17K in Human Solid Tumours. *Oncogene* 27, 5648–5650. doi:10.1038/onc.2008.170

Bohnsack, K., Höbartner, C., and Bohnsack, M. (2019). Eukaryotic 5-methylcytosine (m5C) RNA Methyltransferases: Mechanisms, Cellular Functions, and Links to Disease. *Genes* 10, 102. doi:10.3390/genes10020102

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 A Cancer J. Clin.* 68, 394–424. doi:10.3322/caac.21492

Bridges, M. C., Daulagala, A. C., and Kourtidis, A. (2021). LNCcation: lncRNA Localization and Function. *J. Cell Biol.* 220, 220. doi:10.1083/jcb.202009045

Brose, M. S., Volpe, P., Feldman, M., Kumar, M., Rishi, I., Gerrero, R., et al. (2002). BRAF and RAS Mutations in Human Lung Cancer and Melanoma. *Cancer Res.* 62, 6997–7000.

Carella, A., Tejedor, J. R., García, M. G., Urdinguio, R. G., Bayón, G. F., Sierra, M., et al. (2020). Epigenetic Downregulation of TET3 Reduces Genome-wide 5hmC Levels and Promotes Glioblastoma Tumorigenesis. *Int. J. Cancer* 146, 373–387. doi:10.1002/ijc.32520

Chellamuthu, A., and Gray, S. G. (2020). The RNA Methyltransferase NSUN2 and its Potential Roles in Cancer. *Cells* 9, 1758. doi:10.3390/cells9081758

Chen, X., Li, A., Sun, B.-F., Yang, Y., Han, Y.-N., Yuan, X., et al. (2019). 5-methylcytosine Promotes Pathogenesis of Bladder Cancer through Stabilizing mRNAs. *Nat. Cell Biol.* 21, 978–990. doi:10.1038/s41556-019-0361-y

Cheng, J. X., Chen, L., Li, Y., Cloe, A., Yue, M., Wei, J., et al. (2018). RNA Cytosine Methylation and Methyltransferases Mediate Chromatin Organization and 5-azacytidine Response and Resistance in Leukaemia. *Nat. Commun.* 9, 1163. doi:10.1038/s41467-018-03513-4

Finotello, F., Mayer, C., Plattner, C., Laschober, G., Rieder, D., Hackl, H., et al. (2019). Molecular and Pharmacological Modulators of the Tumor Immune Contexture Revealed by Deconvolution of RNA-Seq Data. *Genome Med.* 11, 34. doi:10.1186/s13073-019-0638-6

Gao, Y., Wang, Z., Zhu, Y., Zhu, Q., Yang, Y., Jin, Y., et al. (2019). NOP 2/Sun RNA Methyltransferase 2 Promotes Tumor Progression *via* its Interacting Partner RPL 6 in Gallbladder Carcinoma. *Cancer Sci.* 110, 3510–3519. doi:10.1111/cas.14190

García, M. G., Carella, A., Urdinguio, R. G., Bayón, G. F., Lopez, V., Tejedor, J. R., et al. (2018). Epigenetic Dysregulation of TET2 in Human Glioblastoma. *Oncotarget* 9, 25922–25934. doi:10.18632/oncotarget.25406

Gaujoux, R., and Seoighe, C. (2010). A Flexible R Package for Nonnegative Matrix Factorization. *BMC Bioinforma.* 11, 367. doi:10.1186/1471-2105-11-367

He, L., Li, H., Wu, A., Peng, Y., Shu, G., and Yin, G. (2019). Functions of N6-Methyladenosine and its Role in Cancer. *Mol. Cancer* 18, 176. doi:10.1186/s12943-019-1109-9

He, R., Wu, J. X., Zhang, Y., Che, H., and Yang, L. (2018). LncRNA LINC00628 Overexpression Inhibits the Growth and Invasion through Regulating PI3K/Akt Signaling Pathway in Osteosarcoma. *Eur. Rev. Med. Pharmacol. Sci.* 22, 5857–5866. doi:10.26355/eurrev_201809_15915

Janin, M., Ortiz-Barahona, V., de Moura, M. C., Martínez-Cardús, A., Llinàs-Arias, P., Soler, M., et al. (2019). Epigenetic Loss of RNA-Methyltransferase NSUN5 in Glioma Targets Ribosomes to Drive a Stress Adaptive Translational Program. *Acta Neuropathol.* 138, 1053–1074. doi:10.1007/s00401-019-02062-4

Jin, G., Kim, M. J., Jeon, H.-S., Choi, J. E., Kim, D. S., Lee, E. B., et al. (2010). PTEN Mutations and Relationship to EGFR, ERBB2, KRAS, and TP53 Mutations in Non-small Cell Lung Cancers. *Lung Cancer* 69, 279–283. doi:10.1016/j.lungcan.2009.11.012

Kaliman, P. (2019). Epigenetics and Meditation. *Curr. Opin. Psychol.* 28, 76–80. doi:10.1016/j.copsyc.2018.11.010

Kanwal, R., Gupta, K., and Gupta, S. (2015). Cancer Epigenetics: an Introduction. *Methods Mol. Biol.* 1238, 3–25. doi:10.1007/978-1-4939-1804-1_1

Kopp, F., and Mendell, J. T. (2018). Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* 172, 393–407. doi:10.1016/j.cell.2018.01.011

Langfelder, P., and Horvath, S. (2008). WGCNA: an R Package for Weighted Correlation Network Analysis. *BMC Bioinforma.* 9, 559. doi:10.1186/1471-2105-9-559

Li, B., Severson, E., Pignon, J.-C., Zhao, H., Li, T., Novak, J., et al. (2016). Comprehensive Analyses of Tumor Immunity: Implications for Cancer Immunotherapy. *Genome Biol.* 17, 174. doi:10.1186/s13059-016-1028-7

Li, Q., Li, X., Tang, H., Jiang, B., Dou, Y., Gorospe, M., et al. (2017). NSUN2-Mediated m5C Methylation and METTL3/METTL14-Mediated m6A Methylation Cooperatively Enhance P21 Translation. *J. Cell. Biochem.* 118, 2587–2598. doi:10.1002/jcb.25957

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

Li, Y., Li, J., Luo, M., Zhou, C., Shi, X., Yang, W., et al. (2018). Novel Long Noncoding RNA NMR Promotes Tumor Progression *via* NSUN2 and BPTF in Esophageal Squamous Cell Carcinoma. *Cancer Lett.* 430, 57–66. doi:10.1016/j.canlet.2018.05.013

Ma, S., Chen, C., Ji, X., Liu, J., Zhou, Q., Wang, G., et al. (2019). The Interplay between m6A RNA Methylation and Noncoding RNA in Cancer. *J. Hematol. Oncol.* 12, 121. doi:10.1186/s13045-019-0805-7

Mei, L., Shen, C., Miao, R., Wang, J.-Z., Cao, M.-D., Zhang, Y.-S., et al. (2020). RNA Methyltransferase NSUN2 Promotes Gastric Cancer Cell Proliferation by Repressing p57Kip2 by an m5C-dependent Manner. *Cell Death Dis.* 11, 270. doi:10.1038/s41419-020-2487-z

Mongan, N. P., Emes, R. D., and Archer, N. (2019). Detection and Analysis of RNA Methylation. *F1000Res* 8, 559. doi:10.12688/f1000research.17956.1

Müller, M. F., Ibrahim, A. E. K., and Arends, M. J. (2016). Molecular Pathological Classification of Colorectal Cancer. *Virchows Arch.* 469, 125–134. doi:10.1007/s00428-016-1956-3

Murayama, T., Nakaoku, T., Enari, M., Nishimura, T., Tominaga, K., Nakata, A., et al. (2016). Oncogenic Fusion Gene CD74-NRG1 Confers Cancer Stem Cell-like Properties in Lung Cancer through a IGF2 Autocrine/Paracrine Circuit. *Cancer Res.* 76, 974–983. doi:10.1158/0008-5472.CAN-15-2135

Nandwani, A., Rathore, S., and Datta, M. (2021). LncRNAs in Cancer: Regulatory and Therapeutic Implications. *Cancer Lett.* 501, 162–171. doi:10.1016/j.canlet.2020.11.048

Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., et al. (2019). Determining Cell Type Abundance and Expression from Bulk Tissues with Digital Cytometry. *Nat. Biotechnol.* 37, 773–782. doi:10.1038/s41587-019-0114-2

Nombela, P., Miguel-López, B., and Blanco, S. (2021). The Role of m6A, m5C and Ψ RNA Modifications in Cancer: Novel Therapeutic Opportunities. *Mol. Cancer* 20, 18. doi:10.1186/s12943-020-01263-w

Pan, J., Huang, Z., and Xu, Y. (2021). m5C-Related lncRNAs Predict Overall Survival of Patients and Regulate the Tumor Immune Microenvironment in Lung Adenocarcinoma. *Front. Cell Dev. Biol.* 9, 671821. doi:10.3389/fcell.2021.671821

Qian, X., Zhao, J., Yeung, P. Y., Zhang, Q. C., and Kwok, C. K. (2019). Revealing lncRNA Structures and Interactions by Sequencing-Based Approaches. *Trends Biochem. Sci.* 44, 33–52. doi:10.1016/j.tibs.2018.09.012

Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E., and Gfeller, D. (2017). Simultaneous Enumeration of Cancer and Immune Cell Types from Bulk Tumor Gene Expression Data. *Elife* 6, 6. doi:10.7554/eLife.26476

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., and Baraba´si, A.-L. (2002). Hierarchical Organization of Modularity in Metabolic Networks. *Science* 297, 1551–1555. doi:10.1126/science.1073374

Sato, K., Tahata, K., and Akimoto, K. (2020). Five Genes Associated with Survival in Patients with Lower-Grade Gliomas Were Identified by Information-Theoretical Analysis. *Anticancer Res.* 40, 2777–2785. doi:10.21873/anticanres.14250

Schneider, F., Derrick, V., Davison, J. M., Strollo, D., Incharoen, P., and Dacic, S. (2016). Morphological and Molecular Approach to Synchronous Non-small Cell Lung Carcinomas: Impact on Staging. *Mod. Pathol.* 29, 735–742. doi:10.1038/modpathol.2016.66

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer Statistics, 2020. *CA A Cancer J. Clin.* 70, 7–30. doi:10.3322/caac.21590

Steen, C. B., Liu, C. L., Alizadeh, A. A., and Newman, A. M. (2020). Profiling Cell Type Abundance and Expression in Bulk Tissues with CIBERSORTx. *Methods Mol. Biol.* 2117, 135–157. doi:10.1007/978-1-0716-0301-7_7

Sun, T., Wu, Z., Wang, X., Wang, Y., Hu, X., Qin, W., et al. (2020). LNC942 Promoting METTL14-Mediated m6A Methylation in Breast Cancer Cell Proliferation and Progression. *Oncogene* 39, 5358–5372. doi:10.1038/s41388-020-1338-9

Sung, H., Ferlay, J., and Siegel Rebecca, L. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* 71, 209–249.

Tibshirani, R. (1997). The Lasso Method for Variable Selection in the Cox Model. *Stat. Med.* 16, 385–395. doi:10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3

Traube, F. R., and Carell, T. (2017). The Chemistries and Consequences of DNA and RNA Methylation and Demethylation. *RNA Biol.* 14, 1099–1107. doi:10.1080/15476286.2017.1318241

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., et al. (2021). clusterProfiler 4.0: A Universal Enrichment Tool for Interpreting Omics Data. *Innovation* 2, 100141. doi:10.1016/j.xinn.2021.100141

Xie, S., Chen, W., Chen, K., Chang, Y., Yang, F., Lin, A., et al. (2020). Emerging Roles of RNA Methylation in Gastrointestinal Cancers. *Cancer Cell Int.* 20, 585. doi:10.1186/s12935-020-01679-w

Xu, S.-F., Zheng, Y., Zhang, L., Wang, P., Niu, C.-M., Wu, T., et al. (2019). Long Non-coding RNA LINC00628 Interacts Epigenetically with the LAMA3 Promoter and Contributes to Lung Adenocarcinoma. *Mol. Ther. - Nucleic Acids* 18, 166–182. doi:10.1016/j.omtn.2019.08.005

Yu, X., Zhang, Q., Gao, F., Zhang, M., Zheng, Q., He, Y., et al. (2021). Predictive Value of m5C Regulatory Gene Expression in Pancreatic Adenocarcinoma. *Sci. Rep.* 11, 17529. doi:10.1038/s41598-021-96470-w

Zhang, Z.-Z., Zhao, G., Zhuang, C., Shen, Y.-Y., Zhao, W.-Y., Xu, J., et al. (2016). Long Non-coding RNA LINC00628 Functions as a Gastric Cancer Suppressor *via* Long-Range Modulating the Expression of Cell Cycle Related Genes. *Sci. Rep.* 6, 27435. doi:10.1038/srep27435

Zhao, B. S., Roundtree, I. A., and He, C. (2017). Post-transcriptional Gene Regulation by mRNA Modifications. *Nat. Rev. Mol. Cell Biol.* 18, 31–42. doi:10.1038/nrm.2016.132

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Cell and Developmental Biology

ORIGINAL RESEARCH
published: 17 May 2022
doi: 10.3389/fcell.2022.897096

Analysis of DNA Repair-Related Prognostic Function and Mechanism in Gastric Cancer

Liqiang Wang[1,2†], *Jianping Lu*[3†], *Ying Song*[1,2], *Jing Bai*[1,2], *Wenjing Sun*[1,2], *Jingcui Yu*[1,4], *Mengdi Cai*[1,2*] *and Songbin Fu*[1,2*]

[1]Key Laboratory of Preservation of Human Genetic Resources and Disease Control in China (Harbin Medical University), Ministry of Education, Harbin, China, [2]Laboratory of Medical Genetics, Harbin Medical University, Harbin, China, [3]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China, [4]Scientific Research Centre, The Second Affiliated Hospital of Harbin Medical University, Harbin, China

OPEN ACCESS

Edited by:
Chunjie Jiang,
University of Pennsylvania,
United States

Reviewed by:
Hong Chen,
University of Texas MD Anderson Cancer Center, United States
Juan Chen,
Hefei University of Technology, China
Wei Jiang,
Nanjing University of Aeronautics and Astronautics, China

***Correspondence:**
Songbin Fu
fusb@ems.hrbmu.edu.cn
Mengdi Cai
caimengdi@ems.hrbmu.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular and Cellular Pathology, a section of the journal Frontiers in Cell and Developmental Biology

Received: 15 March 2022
Accepted: 18 April 2022
Published: 17 May 2022

Citation:
Wang L, Lu J, Song Y, Bai J, Sun W, Yu J, Cai M and Fu S (2022) Analysis of DNA Repair-Related Prognostic Function and Mechanism in Gastric Cancer. Front. Cell Dev. Biol. 10:897096. doi: 10.3389/fcell.2022.897096

DNA repair mechanisms have been proven to be essential for cells, and abnormalities in DNA repair could cause various diseases, such as cancer. However, the diversity and complexity of DNA repair mechanisms obscure the functions of DNA repair in cancers. In addition, the relationships between DNA repair, the tumor mutational burden (TMB), and immune infiltration are still ambiguous. In the present study, we evaluated the prognostic values of various types of DNA repair mechanisms and found that double-strand break repair through single-strand annealing (SSA) and nonhomologous end-joining (NHEJ) was the most prognostic DNA repair processes in gastric cancer (GC) patients. Based on the activity of these two approaches and expression profiles, we constructed a HR-LR model, which could accurately divide patients into high-risk and low-risk groups with different probabilities of survival and recurrence. Similarly, we also constructed a cancer-normal model to estimate whether an individual had GC or normal health status. The prognostic value of the HR-LR model and the accuracy of the cancer-normal model were validated in several independent datasets. Notably, low-risk samples, which had higher SSA and NHEJ activities, had more somatic mutations and less immune infiltration. Furthermore, the analysis found that low-risk samples had higher and lower methylation levels in CpG islands (CGIs) and open sea regions respectively, and had higher expression levels of programmed death-ligand 1 (PD-L1) and lower methylation levels in the promoter of the gene encoding PD-L1. Moreover, low-risk samples were characterized primarily by higher levels of $CD4^+$ memory T cells, $CD8^+$ naive T cells, and $CD8^+$ TEM cells than those in high-risk samples. Finally, we proposed a decision tree and nomogram to help predict the clinical outcome of an individual. These results provide an improved understanding of the complexity of DNA repair, the TMB, and immune infiltration in GC, and present an accurate prognostic model for use in GC patients.

Keywords: gastric cancer, DNA repair, tumor mutational burden, immune infiltration, DNA methylation, prognostic model

INTRODUCTION

Gastric cancer (GC) is the sixth most prevalent cancer in the world and the third leading cause of cancer-related deaths (Bray et al., 2018). Despite remarkable progress in diagnostic and therapeutic methods, GC remains a refractory malignancy. In recent years, significant progress has been made in understanding cancer-associated molecular genetics as a result of the development of molecular biology methods. Several studies have found that mutations in the genome play an indispensable role in genomic maintenance and evolution. Furthermore, studies have demonstrated that when DNA repair mechanisms are disrupted or deregulated this may increase rates of mutagenesis and genomic instability and thereby mediate cancer progression (Bouwman and Jonkers, 2012; Wolters and Schumacher, 2013). The main DNA repair mechanisms include direct repair, base excision repair, and double-strand break repair, among which DNA double-strand break (DSB) repair plays a crucial role in maintaining genomic integrity (Gillyard and Davis, 2021). In addition, DSB affects the prognosis of cancer by preventing disadvantageous mutations (Stok et al., 2021). However, it remains unknown whether DSB is valuable for predicting the clinical outcomes of GC patients.

The tumor mutational burden (TMB) refers to the number of somatic mutations per DNA megabase in tumor cells. TMB is considered as the primary driver of antitumor adaptive immune responses and serves as a positive predictive biomarker for immune checkpoint inhibitors (Castle et al., 2012). Recently, various studies have illustrated that the genomic instability resulting from the inadequacy of DNA repair mechanisms is associated with high TMBs. Preclinical studies identified that tumor cells with higher TMB could produce more neoantigens, which are more easily recognized by T cells, and thus activate stronger immune killing activity (Parikh et al., 2019; Klempner et al., 2020). However, there are some limitations to the use of the TMB as a biomarker for clinical utilization. This is mainly because heterogeneity among intratumoral neoantigens may also be important for immunotherapy response (McGranahan et al., 2016). Furthermore, it is also important to acknowledge that the cut-off values of TMB lack standardization and consistency because they are defined differently across studies, testing platforms, and patient populations (Gibney et al., 2016). Thus, the development of predictive biomarkers is urgently needed to benefit patients. In this situation, an increasing number of oncologists have begun to focus their studies on PD-L1 expression of tumor cells and found that PD-L1 positivity has emerged as a major predictive marker (Borghaei et al., 2015; Larkin et al., 2015). However, the accuracy is not satisfactory based on PD-L1 as a single molecule for GC patients (Patel and Kurzrock, 2015). This is mainly because the expression of PD-L1 is also affected by other factors. For example, its expression is associated with global hypomethylation (Emran et al., 2019). To improve clinical outcomes in GC, it is essential to explore emerging biomarkers through a comprehensive multifaceted analysis involving DNA repair mechanisms, tumor-infiltrating lymphocytes, TMB, mutational signatures, immune microenvironments of tumors, and immune checkpoints.

In the present study, we constructed a HR-LR model to improve the performance of the prognosis of overall survival and recurrence of GC patients by integrating the DNA repair-related GO processes from the MsigDB database and the clinical data from TCGA stomach adenocarcinoma (GC) patients. Moreover, we estimated the expression level of PD-L1 as an immune checkpoint and methylation level of CpG sites in the PD-L1 promoter region, TMB, and the systemic immune status. To a certain extent, attempting to exploit a novel model based on DNA repair mechanisms will significantly help select GC patients who would benefit from predictions of clinical outcomes and improve the accuracy of prognostic assessments.

MATERIALS AND METHODS

Transcriptomic, Genomic, and Clinical Datasets of the Cancer Genome Atlas Cohort

Transcriptional profiles of cancer and normal tissues of patients with GC were obtained from stomach adenocarcinoma (STAD) patients of The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov), including 375 cancer samples and 32 normal samples (**Table 1**). For cancer or normal samples, genes with FPKM expression values of 0 in >70% of samples were removed, and the remaining 0 values were imputed with K-Nearest Neighbors. Then, expression values were log2 transformed for subsequent analysis.

Mutational data of GC patients was also downloaded from the TCGA database. After removing the synonymous variants, we calculated the tumor mutational burden (TMB), which was defined as the number of somatic mutations per megabase of interrogated genomic sequence. Mutation profiles were analyzed and visualized by the R package "maftools" (Mayakonda et al., 2018).

We obtained clinical information of GC patients from the TCGA database, including survival state, survival time, recurrence state, recurrence time, disease stage, therapeutic response, age, gender, and other clinical characteristics.

DNA Repair Related GO Terms and Pathways

The genes of GO terms and pathways related to DNA repair mechanism were downloaded from the Molecular Signatures Database (MsigDB) (Liberzon et al., 2015). Based on the transcriptional data of these genes in the TCGA cohort, we calculated the score of each GO term or pathway using a single-sample gene set enrichment analysis (ssGSEA). A univariate Cox proportional-hazards regression model was used to evaluate the prognostic significance of each GO term or pathway in GC by R package "survival". Taken the results of DNA repair-related GO terms, KEGG pathways, and Reactome pathways, double-strand break repair through single-strand annealing (SSA) or nonhomologous end-joining (NHEJ) were the most prognostic approaches. The result was validated in GSE66254 (GPL570) dataset.

TABLE 1 | Patient cohorts from TCGA and GEO databases.

Cohort	Cancer samples	Normal samples	Recurrence	GPL
TCGA	375[a] (242 training +106 test)	32	r[b]	-
GSE62254	300	-	R	GPL570
GSE26253	432	-	R	GPL8432
GSE84437	433	-	-	GPL6947
GSE26899	96	-	R	GPL6947
GSE15460	248[c]	-	-	GPL570
GSE13861	65	25	r	GPL6884
GSE13911	38	31	-	GPL570
GSE33335	25	25	-	GPL5175
GSE66229	300 (GSE62254)	100	-	GPL570

[a]*In total, we obtained 375 cancer and 32 normal samples from TCGA, database. Screening samples with survival data and deleting those samples died in 10 days, we finally obtained 348 samples. Then we selected 242 samples as the training set randomly while the remaining 106 samples were as the test set.*
[b]*"r" in the table represented the datasets with recurrence information.*
[c]*The dataset GSE15460 included GSE15455, GSE15456, GSE15459, GSE15537, GSE22183, GSE34942 datasets. Deleting those cell line datasets and the datasets sequenced by GPL96, we finally obtained 248 samples, including GSE15459 and GSE34942.*

TABLE 2 | Positive and negative genes used in the HR-LR model and Cancer-Normal model.

	Gene symbols
Positive genes	BRIP1, CDC45, CDC7, CDCA2, CENPK, CLSPN, DDIAS, DLGAP5, DTL, E2F7, EZH2, FANCA, HELLS, HIST1H2AH, KIF11, KIF15, KIF18A, KIF23, KIF2C, KNTC1, LMNB1, MCM10, MND1, NCAPG, ORC1, PCNA, PLK4, POLE2, POLQ, POLR3G, RAD51AP1, RAD54L, RFC4, RRM2, TYMS, UHRF1, XRCC2
Negative genes	ADCY5, APOD, C15orf59, C16orf89, C1QTNF2, C1QTNF7, CGNL1, CRYAB, DAAM2, DACT3, DCN, ELN, FAM110B, FMOD, GHR, GREM2, GSTM5, HSPA2, HSPB8, KCNK3, LRRN4CL, MFAP4, NDNF, NEGR1, NFATC4, PDE2A, PPP1R14A, PPP1R3C, SAMD11, SCN4B, SCUBE2, SLC22A17, SMARCD3, SRPX, TCEAL7, TMEM100, TMOD1, TNFAIP8L3, ZCCHC24

The ssGSEA score of SSA or NHEJ GO term in normal samples and good or poor outcome samples were also calculated by the ssGSEA algorithm. Those samples that lived more than 1 year were defined as good outcome samples while dead samples in 1 year were defined as poor outcome samples.

Construction of HR-LR Model

We proposed a computational method to establish the HR-LR model, which involved three steps. First, we randomly divided the STAD TCGA cohort into training and test sets, including 242 and 106 samples, respectively (**Table 1**). The HR-LR model was established based on the training set. Second, we selected those genes with significant associations with a score of SSA and NHEJ GO terms as the DNA repair-related marker genes ($|R| > 0.4$, *BH-FDR* < 0.001). In total, we obtained 37 positive genes and 39 negative genes. Third, an SSA-NHEJ score was defined by the *T* statistic of a two-sided *t*-test for each tumor sample by comparing the expression values of the 37 positively correlated genes with the expression values of the 39 negatively correlated genes (**Table 2** and **Supplementary Table S1**).

The median SSA-NHEJ score of training samples was defined as the cutoff (cutoff = 3.46). An SSA-NHEJ score >3.46 represented that those 37 positive genes were overexpressed while the 39 negative genes were underexpressed. An SSA-NHEJ score <3.46 meant the opposite. As SSA and NHEJ GO terms were protective factors in GC, patients with higher SSA-NHEJ scores were considered with a better outcome. Therefore, the samples were divided into high-risk and low-risk groups, with low and high SSA-NHEJ scores, respectively. The survival hypothesis was validated in training, test sets of the TCGA cohort, and other GEO cohorts. In addition, we found that those high-risk samples were more likely to be recurrent, compared with those low-risk samples.

Construction of Cancer-Normal Model

The model was established exactly like the HR-LR model. However, the cutoff was changed. Using the R package "pROC" (Robin et al., 2011), we selected 0.008 as the cutoff. Samples with SSA-NHEJ score ≥0.008 were predicted as tumor samples while other samples were predicted as normal samples.

Transcriptomic and Clinical Datasets of GEO Validation Cohorts

The independent validation datasets were downloaded from the Gene Expression Omnibus database, including nine datasets. The detailed information is shown in **Table 1**. We chose six datasets with expression profiles and survival information to validate our HR-LR model, including GSE62254, GSE26253, GSE84437, GSE26899, GSE15460, and GSE13861. For validation of GC recurrence, we selected four datasets with recurrence information, including GSE62254, GSE26253, GSE13861, and GSE26899. Finally, we chose the datasets with cancer and normal samples to validate the cancer-normal model, including GSE13861, GSE13911, GSE33335, and GSE66229.

Generation of ImmuneScore, StromalScore, EstimateScore, and MicroenvironmentScore

For each patient sample, ImmuneScore, StromalScore, EstimateScore, and MicroenvironmentScore were generated by R package "estimate" (Yoshihara et al., 2013) and R package "xCell" (Aran et al., 2017). The higher the respective score, the larger the ratio of the corresponding component in the tumor microenvironment (TME). The infiltration of immune and stromal cell types in an individual sample was evaluated by R package "xCell".

Calculation of DNA Hyper- and Hypomethylation Scores in Tumor Samples

DNA methylation dataset of STAD patients detected by Illumina Infinium HumanMethylation450 BeadChip array was also downloaded from the TCGA data portal. After selecting the promoter CpG islands (CGIs) and open sea regional clusters on the genome, we calculated the aberrant hypermethylation (over CGI probes) and hypomethylation (over open sea probes) values for each tumor sample compared with normal samples according to one of the previous studies (Yang et al., 2015). The scores of hyper- and hypomethylation were calculated as follows: 1) all genome CpG sites were classified into different regional classes, then CGI and open sea regions were selected to be grouped into regional clusters by boundedClusterMaker in R-package bumphunter (maximum cluster width of 1500bp and maximum gap of 500bp between any two neighboring regional classes), respectively; 2) the methylation in each cluster was defined as the mean beta value of the sites within the cluster; 3) for each cluster in a certain cancer sample, the relative methylation was calculated as the beta value of this cluster in single cancer sample subtracting the mean value and further dividing the standard value of beta value of this cluster in all normal samples; 4) since, cluster regions in promoter CGIs and open sea usually show hypermethylated and hypomethylated in cancer samples, we calculated the hyper- and hypomethylation for a cancer sample as the mean of positive and negative relative methylation value in all region clusters, respectively.

Survival Analysis

Kaplan–Meier survival plots and log-rank tests were used to evaluate the survival differences between groups of patients. The univariate Cox proportional-hazards regression model and multivariate Cox proportional-hazards regression model were used to evaluate the prognostic significance of factors. This process was performed using the R package "survival".

Decision Tree

Combining the HR-LR model and patient stage information, a decision tree was established to predict the single sample into low-risk, moderate-risk, and high-risk groups. The relationship between predicted results and the sample real tags was plotted as a Sankey diagram, using the R package "networkD3".

Nomogram Plot

A nomogram was built with SSA-NHEJ score and other clinical features to quantify the risk assessment for the individual patient, using the R package "rms" (Zhang and Kattan, 2017).

RESULTS

The Single-Strand Annealing and Nonhomologous End-Joining DNA Repair Approaches ere Identified as the Primary Predictive Factors for Overall Survival in Gastric Cancer Patients

On the basis of the expression profiles of the TCGA STAD cohort and the gene lists extracted from the MSigDB, we calculated the performance score (ssGSEA score) for each DNA repair-related GO term, KEGG pathway, and Reactome pathway (**Supplementary Table S2**). The univariate Cox coefficient and significance of each GO term and pathway were determined. Summarizing the results, we found that DSB repair was the most effective prognostic factor (**Supplementary Figures S1A–C**). Among the DSB repair mechanisms, SSA and NHEJ were found to be the primary predictive factors for outcomes in the TCGA STAD cohort (**Figure 1A**). Because DNA repair mechanisms are described in more detail by GO terms, we then selected "double-strand break repair *via* single-strand annealing," "positive regulation of double-strand break repair *via* nonhomologous end joining," and "regulation of double-strand break repair *via* nonhomologous end joining," for subsequent analysis (HR = 0.009, 0.002, and 0.004, respectively; p = 8.01e−03, 1.05e−02 and 4.33e−02, respectively). Kaplan–Meier survival plots showed that patients with higher ssGSEA scores have better outcomes (**Figure 1B**; log-rank p = 4.3e−04, 4.35e−02, and 9.98e−03, respectively). These results were validated in the independent GSE62254 dataset (**Figure 1C**). This dataset was selected because there are relatively more genes corresponding to the three abovementioned GO terms in data from the GPL570 platform.

To examine the performance of these three prognostic DNA repair-related GO terms, we then compared their ssGSEA scores between three groups from the TCGA cohort, namely, normal samples, good-outcome samples, and poor-outcome samples. Here, we defined the good-outcome samples as those patients who had a survival time of greater than 1 year, while those patients who died within 1 year were regarded as poor-outcome samples. The results showed that the ssGSEA scores of SSA and NHEJ were the lowest in normal samples (**Figure 1D**). This suggests that the SSA and NHEJ repair patterns have higher activity in cancer cells in comparison with normal cells, which may be caused by the fact that more DNA replication and more errors occur in cancer cells. On the other hand, among cancer samples, good-outcome samples had significantly higher ssGSEA scores in comparison with poor-outcome samples (**Figure 1D**). This result was consistent with the result of univariate Cox regression as shown in **Figure 1A**. Because repair by SSA and NHEJ could lead to an accumulation of mutations, we then

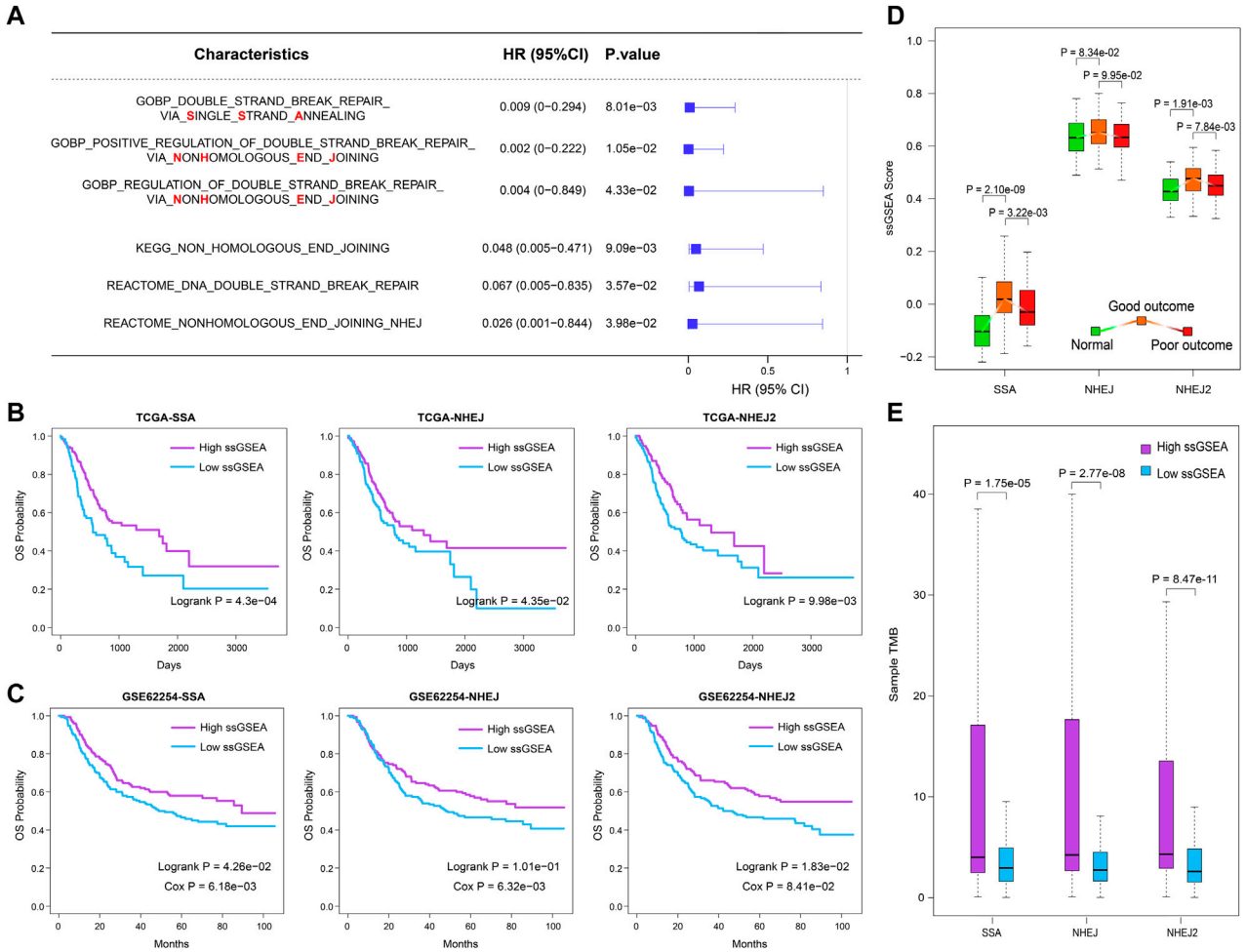


FIGURE 1 | SSA and NHEJ DNA repair approaches are primary protection factors for overall survival in GC patients. **(A)** The prognostic value of SSA and NHEJ DNA repair process evaluated by univariate Cox proportional-hazards regression model. **(B,C)** Samples with higher ssGSEA scores of SSA or NHEJ have better clinical outcomes in the TCGA cohort and GSE62254 dataset. **(D)** ssGESA scores of SSA and NHEJ processes increase first and then decline in normal, good outcome, and poor outcome samples. **(E)** Samples with a higher ssGSEA score of SSA or NHEJ are with more somatic mutations.

calculated the TMB for each cancer sample and compared the TMBs of samples with high and low ssGSEA scores. As was expected, samples with higher ssGSEA scores had significantly higher TMBs (p = 1.75e−05, 2.77e−08, and 8.47e−11 for SSA, NHEJ, and NHEJ2, respectively; **Figure 1E**).

Combining these results, we suggest that the SSA and NHEJ repair mechanisms were enhanced when normal cells were converted into cancer cells and that those patients with stronger SSA and NHEJ repair mechanisms had more mutations, which may have resulted in the production of more new antigens, and thus finally had better outcomes.

The HR-LR Model Is a Highly Effective Prognostic Factor for Gastric Cancer Patients

We developed the HR-LR model in three steps, as described in the Materials and Methods section. Using the HR-LR model, we calculated the SSA-NHEJ score for each sample in the training and test sets. Kaplan–Meier survival plots were generated and log-rank tests were executed for samples with high- and low-SSA-NHEJ scores (with the median score as the cut-off) in the training set (n = 242). As a result, we found that samples with higher SSA-NHEJ scores had better clinical outcomes (log-rank p = 5.06e−03; **Figure 2A**). Then, we validated the predictive effect of the SSA-NHEJ score in the test set (n = 106, log-rank p = 4.07e−02; **Figure 2B**). The cut-off remained unchanged in the test set.

To characterize the functions of genes in the HR-LR model, a functional statistical analysis was performed on them. Among the 37 positively correlated genes, 8, 15, and 20 genes were associated with DSB repair, DNA repair, and cancer hallmarks, respectively. Of the 39 negatively correlated genes, 12 were associated with cancer hallmarks (**Figure 2C**, **Table 2**, and **Supplementary Table S1**). Furthermore, we have listed the associated cancer hallmarks in detail in **Figure 2D**, including several well-known examples

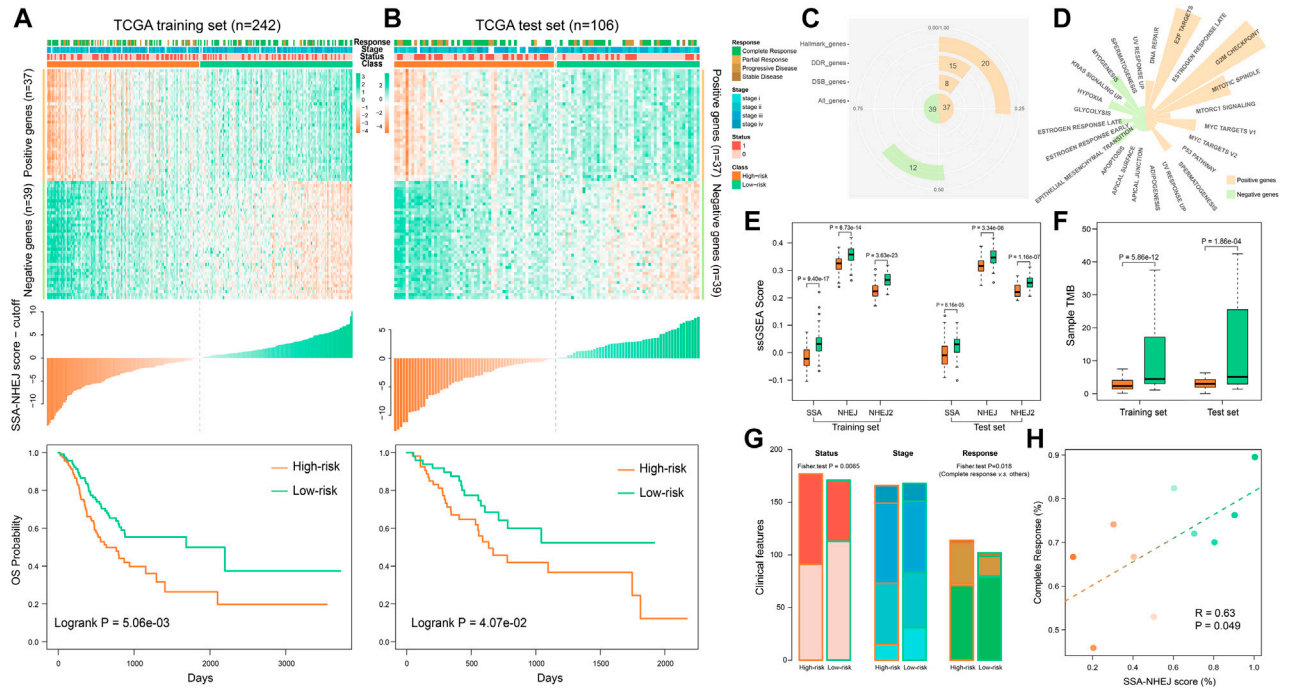


FIGURE 2 | Using HR-LR model to predict survival risk of samples. **(A)** Top panel shows the expression profile of the SSA and NHEJ-related marker genes in the TCGA training set. The middle panel shows the score of each training sample by the HR-LR model. The bottom panel represents the Kaplan–Meier survival plot of high-risk and low-risk samples in the training set. **(B)** Expression profile, sample score, and Kaplan–Meier survival plot for TCGA test samples. **(C)** The statistic of 76 marker genes. **(D)** Marker genes related to cancer hallmarks. **(E)** Low-risk samples are with significantly higher SSA and NHEJ ssGSEA scores in training and test sets. **(F)** Low-risk samples have significantly more somatic mutations. **(G)** Low-risk samples are with more alive, low stage, and complete response samples. **(H)** Correlation of SSA-NHEJ score with complete response. Samples were divided into ten groups according to their scores. Samples with higher scores are with a higher probability of complete response.

such as DNA repair, apoptosis, hypoxia, glycolysis, the epithelial–mesenchymal transition, the p53 pathway, and KRAS signaling (**Supplementary Table S1**). These results suggested that these 76 genes were associated with not only DNA repair by SSA and NHEJ but also significant biological mechanisms in the body and, thus, ultimately affected the development of GC and clinical outcomes in patients.

Comparing the performance of the SSA and NHEJ GO terms, we found that high-risk samples had significantly lower ssGSEA scores in both the training set and the test set (**Figure 2E**). Furthermore, we found that high-risk samples had significantly lower TMB or fewer somatic mutations (**Figure 2F**). These results were consistent with previous studies, as samples with higher TMBs had better outcomes owing to the generation of more new antigens (Parikh et al., 2019; Klempner et al., 2020). Analyzing the clinical features of samples, we found that high-risk samples had more deaths, higher disease stages, and poor responses to therapy (**Figure 2G**). The probability of patients achieving complete response was significantly correlated with the SSA-NHEJ score (**Figure 2H**; Pearson correlation coefficient r = 0.63, p = 0.049).

To validate the prognostic effect of the SSA-NHEJ score, we first performed a multivariate Cox regression modeling analysis involving the SSA-NHEJ score, age, gender, disease stage and grade, lymph node count, and family history of stomach cancer. The results showed that the SSA-NHEJ score was the most effective prognostic factor (HR = 0.50 and p = 0.002 for low-risk samples), which indicated that the SSA-NHEJ score was an independent prognostic factor for GC patients (**Figure 3A**).In addition, patients with stage IV disease and older ages were found to have poor outcomes. To investigate the applicability of the HR-LR model and validate its prognostic effect, we collected several independent datasets involving GC patients from the GEO database (**Table 1**). Using the expression of genes in the HR-LR model and the cut-off derived from the TCGA training set in each GEO dataset, we calculated the SSA-NHEJ score for each sample and further divided the samples into high-risk and low-risk groups. The results showed that in six independent GEO datasets from different GEO platforms the HR-LR model had a significant prognostic effect (**Figure 3B**). These results confirmed the prognostic value of the HR-LR model and established that this model could be applied in different datasets, including RNA-Seq data and microarray data from different platforms.

The HR-LR Model Can Also Predict Recurrence in Gastric Cancer Patients

As the HR-LR model had a robust prognostic effect with regard to overall survival in GC patients, we then tested its predictive effect on recurrence in GC patients. Integrating the training and test sets, we found that the HR-LR model could divide samples into

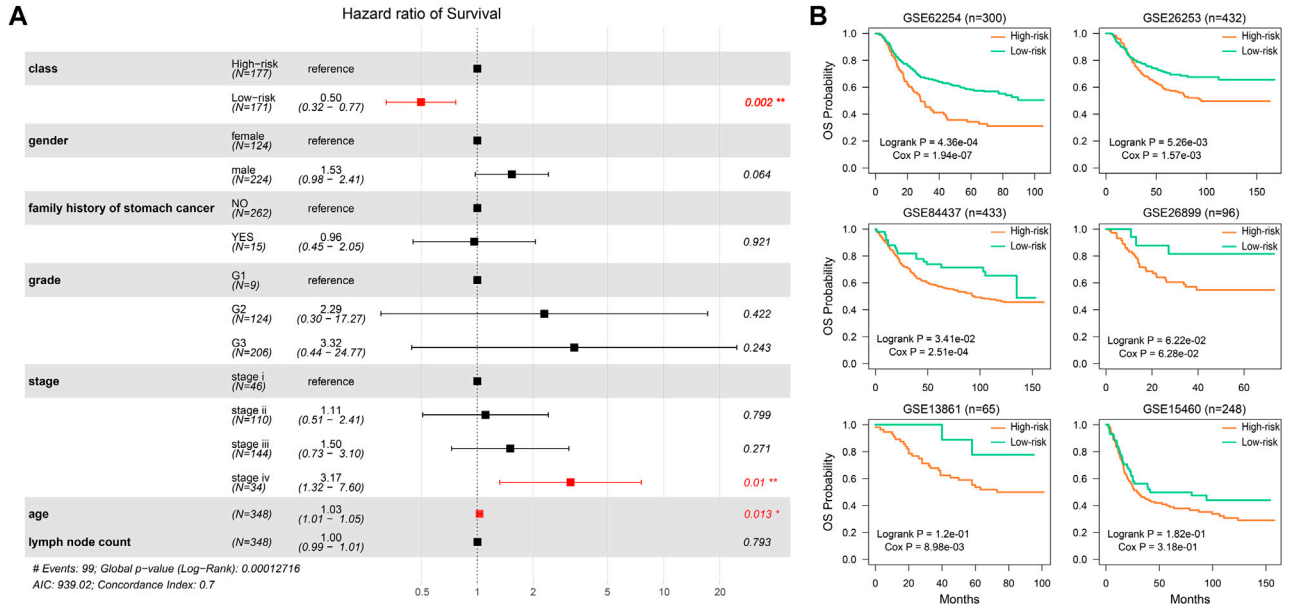


FIGURE 3 | The prognostic effect of the HR-LR model. **(A)** Multivariate Cox proportional-hazards regression result of SSA-NHEJ score and other clinical characteristics. **(B)** Validation of the prognostic effect of the SSA-NHEJ score in six GEO cohorts. *n* represents the number of samples in each GEO dataset.

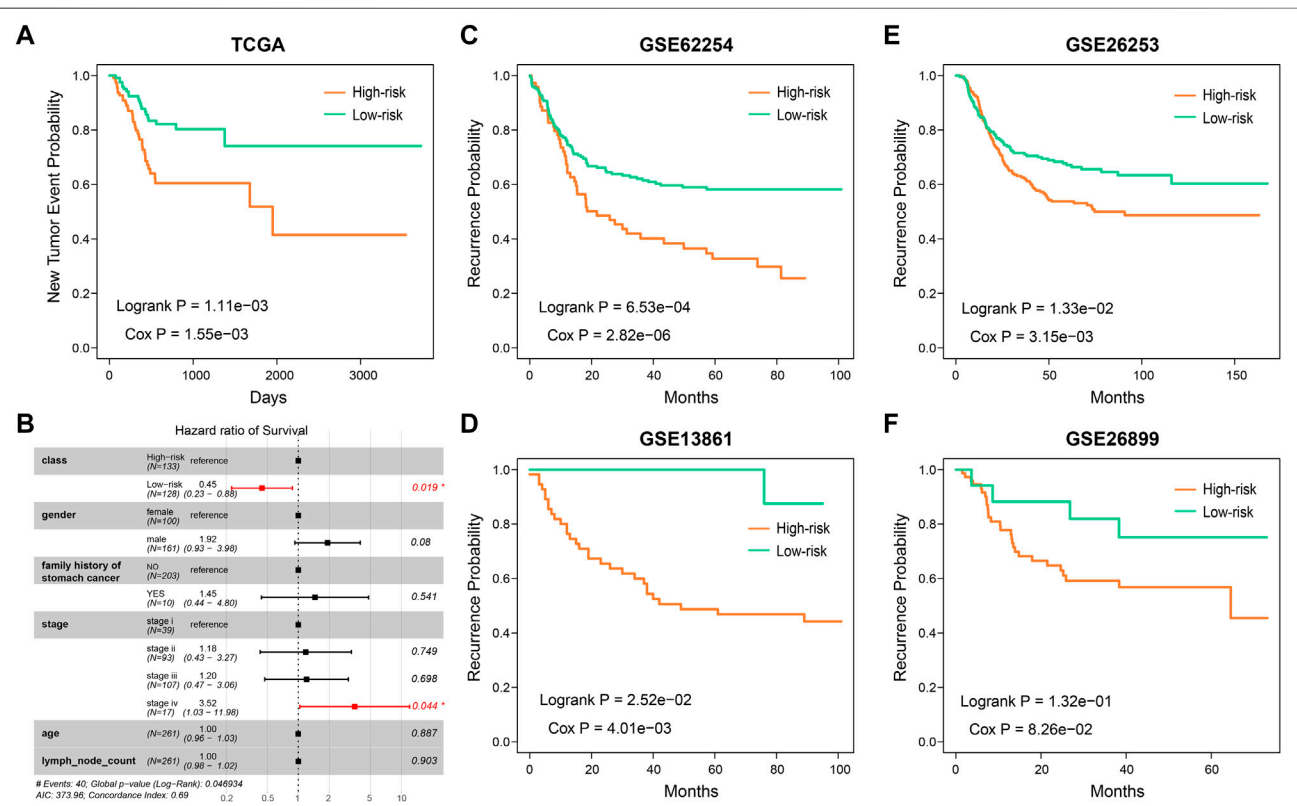


FIGURE 4 | The predictive ability of SSA-NHEJ score for recurrence of GC patients. **(A)** New Tumor Event probability of patients in TCGA cohort. **(B)** Multivariate Cox proportional-hazards regression result for recurrence prognosis of the SSA-NHEJ score and other clinical characteristics. **(C–F)** Recurrence probability of patients in four GEO cohorts.

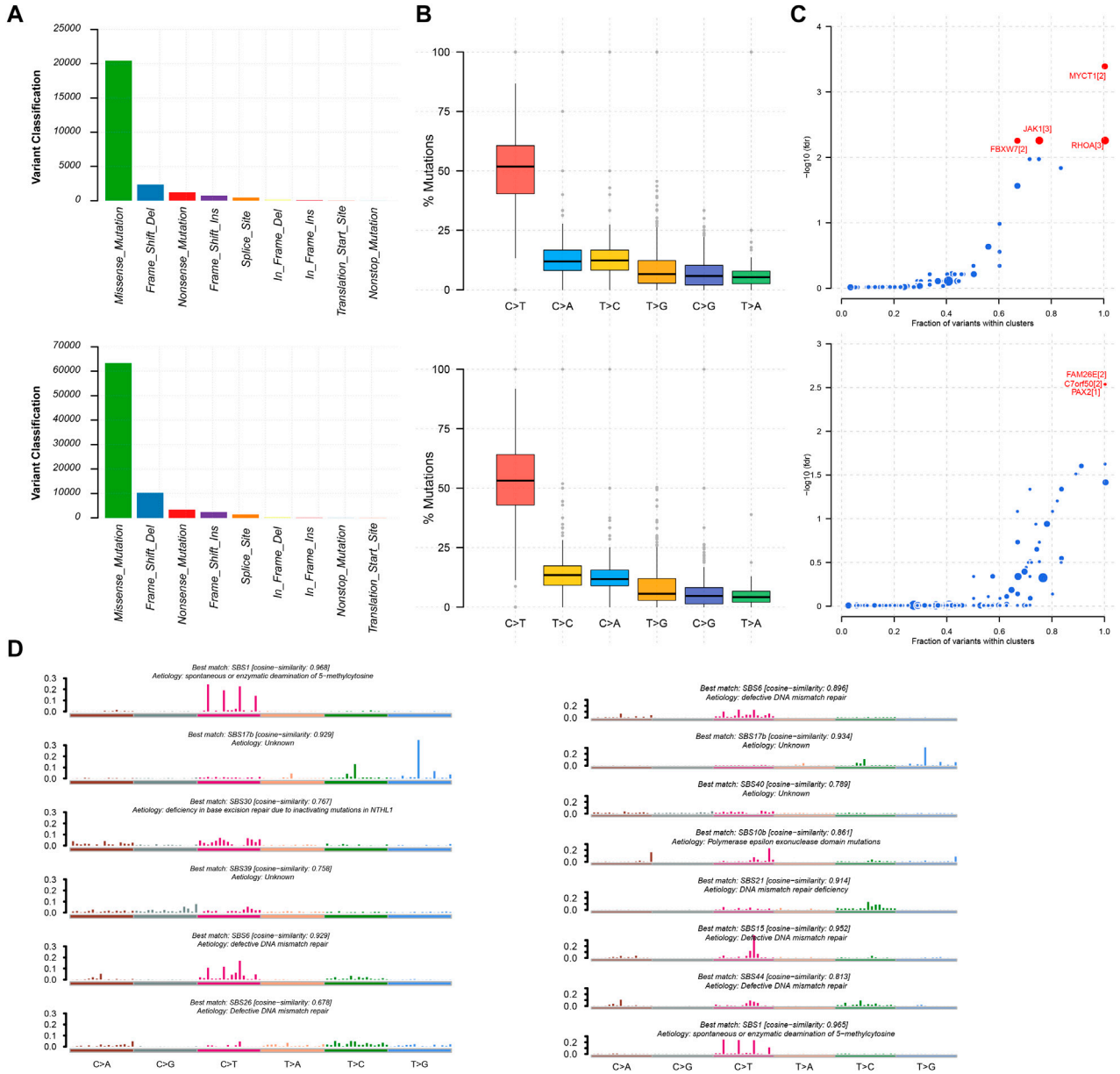


FIGURE 5 | Mutation information of high-risk samples and low-risk samples. **(A,B)** Statistic of variant classification and mutation type of samples with high-risk (top panel) and low-risk (bottom panel) survival. **(C)** Driver mutation genes in high-risk and low-risk samples (top and bottom panel respectively). **(D)** Mutational signatures were identified in high-risk and low-risk samples, respectively. The plot title indicates the best match against validated COSMIC signatures (left and right panel respectively).

high recurrence risk and low recurrence risk groups in the TCGA cohort (**Figure 4A**; log-rank p = 1.11e−03, Cox p = 1.55e−03). Multivariate Cox regression modeling analysis showed that the SSA-NHEJ score was a prognostic factor for recurrence of GC that was independent of clinical features (**Figure 4B**; HR = 0.45 and Cox p = 0.019 for low-risk samples). In addition, the prognostic effect of the HR-LR model with regard to recurrence was also validated in four other GEO datasets (**Figures 4C–F**; log-rank p = 6.53e−04, 1.33e−02, 2.52e−02 and 1.32e−01 and Cox p = 2.82e−06, 3.15e−03, 4.01e−03 and 8.26e−02 for the GSE62254, GSE26253, GSE13861, and GSE26899 datasets, respectively).

Low-Risk Samples Had Higher TMBs

We utilized the maftools package to visualize the results on the basis of somatic mutation data from the TCGA STAD cohort. These somatic mutations included point mutations and insertions/deletions. An oncoplot plot showed that missense mutations occupied an absolute position among total mutations and that the number of mutations in low-risk

samples was higher than that in high-risk samples (**Figure 5A**). Then, a transition and transversion plot was used to classify single-nucleotide variants into six categories (i.e., T > G, T > C, T > A, C > T, C > G, and C > A; **Figure 5B**). Moreover, it was found that C > T had the highest frequency among single-nucleotide variants in both low-risk and high-risk samples.

Cancer driver genes could provide an advantage for the selective growth of cancer cells. Therefore, we applied the OncodriveCLUST algorithm in the TCGA STAD cohort, which detected a majority of the activating mutations and identified seven well-known oncogenes as significantly mutated in 348 samples (false discovery rate <0.01). MYCT1, RHOA, JAK1, and FBXW7 were identified in high-risk samples. Likewise, we identified FAM26E, C7orf50, and PAX2 in low-risk samples (**Figure 5C**). Whereas the Oncodrive algorithm was more advantageous in terms of sensitivity in identifying oncogenes with mutational hotspots, in contrast, the Oncodrive algorithm underperformed in identifying potential tumor suppressor genes (Mayakonda et al., 2018). In consequence, we did not identify potential tumor suppressor genes, such as *TP53*.

As cancer progresses, a characteristic mutational pattern may be left behind at various points in time, which may reveal its underlying mutagenic process (Alexandrov et al., 2013). Therefore, we further analyzed mutational signatures on the STAD TCGA cohort by performing signature enrichment. In total, we identified six and eight signatures in high-risk and low-risk samples, respectively (**Figure 5D**). Interestingly, we identified the common signature associated with DNA mismatch repair in both types of samples was associated with high numbers of small (shorter than 3bp) insertions and deletions at mono/polynucleotide repeats. The deficiency in base excision repair due to inactivating mutations in NTHL1 was specific to the mutational signature of high-risk samples and primarily caused predominantly C > G mutations. This may be due to the generation of abasic sites after removal of uracil by base excision repair. In addition, polymerase epsilon exonuclease domain mutations were specific to the mutational signature of low-risk samples and exhibited strand bias for C > A mutations in the TpCpT context and T > G mutations in the TpTpT context. The mutational process underlying this signature was altered activity of the error-prone polymerase POLE. It has been proposed that the presence of large numbers of this signature was associated with recurrent POLE somatic mutations. The observations from signature enrichment suggest that DNA mismatch repair mechanisms play a crucial role in the development of malignant tumors, in accordance with previous reports (Belfield et al., 2018; McCarthy et al., 2019).

Low-Risk Samples Were Associated With Less Immune Cell Infiltration

To further clarify the intrinsic biological differences between high-risk and low-risk samples, the ESTIMATE algorithm was used for the estimation of stromal cells and immune cells in malignant tumors by calculating the corresponding scores. A higher immune score or stromal score represents a larger amount of immune or stromal components in the tumor microenvironment. To investigate correlations between the stromal score, immune score, and ESTIMATE score and the SSA-NHEJ score, Pearson's correlation coefficient were used to measure the strength of the respective correlations. The results indicated that the stromal score, immune score, and ESTIMATE score were negatively correlated with the SSA-NHEJ score and decreased significantly with an increase in the SSA-NHEJ score (**Figures 6A–C**). We then determined the differences in the immune score, stromal score, and ESTIMATE Score between high-risk and low-risk samples. The results showed that the average immune score (**Figure 6D**), stromal score (**Figure 6E**), and ESTIMATE score (**Figure 6F**) were significantly higher in high-risk samples than in low-risk samples.

We further compared the enrichment levels of different types of cells from gene expression data for 64 types of immune and stromal cells in the two types of samples (**Figure 7A**, **Supplementary Table S3**). The results showed that the low-risk samples had more infiltrating adaptive immune cells including $CD8^+$ TEM cells, $CD8^+$ naive T cells, $CD4^+$ TCM cells, Th1 cells, and Th2 cells, and the scores of infiltration immune cells were also higher, which might be a key factor for favorable prognosis of the low-risk samples. Recent studies have shown that $CD8^+$ T cells are regarded as the main driver of anti-tumor immunity (Reiser and Banerjee, 2016; Fang et al., 2020). Additionally, we found infiltration scores of stromal cells, such as endothelial cells, fibroblasts, and pericytes, were relatively high in high-risk samples compared to low-risk samples. Subsequently, on the basis of data for the high-risk and low-risk samples, we generated a heatmap of immune cells with significant differences and performed a differential analysis of gene expression of the immune checkpoint PD-L1 (**Figures 7B–D**). Using the HR-LR model, the low-risk samples were associated with a favorable prognosis with a median survival time of 2,197 days, while the high-risk samples had a median survival time of 635 days. The low-risk samples were characterized primarily by infiltration of high levels of $CD4^+$ memory T cells, $CD8^+$ naive T cells, $CD8^+$ TEM cells, natural killer cells, plasma cells, pro-B cells, Th1 cells, and Th2 cells. Subjects in the high-risk samples had shorter overall survival times and exhibited significant increases in the infiltration of $CD4^+$ naive T cells, $CD4^+$ T cells, $CD4^+$ TCM cells, megakaryocytes, dendritic cells, and eosinophils. PD-L1 has been found to be an immune checkpoint. A previous study found that global hypomethylation of DNA could contribute to the upregulation of PD-L1 in melanoma cells and had an impact on DNA repair pathways (Emran et al., 2019). We selected CpG islands and open-sea regions to measure methylation levels in each STAD sample by the method described in the Materials and Methods section. The results showed that high-risk samples had lower methylation levels in CpG islands and higher methylation levels in open-sea regions in comparison with low-risk samples (Wilcoxon p = 2.37e−11 and 1.66e−10, respectively; **Figure 7C**). In addition, high-risk samples were characterized by a significantly lower expression level of PD-L1 (Wilcoxon p = 5.65e−04; **Figure 7D**). This might have been caused by the higher methylation level of CpG sites in the promoter region of the gene encoding PD-L1 (Wilcoxon p = 4.19e−05; **Figure 7D**). The relationships between global methylation, promoter methylation, and expression of PD-L1 and DNA repair processes need to be further studied. Moreover, we also

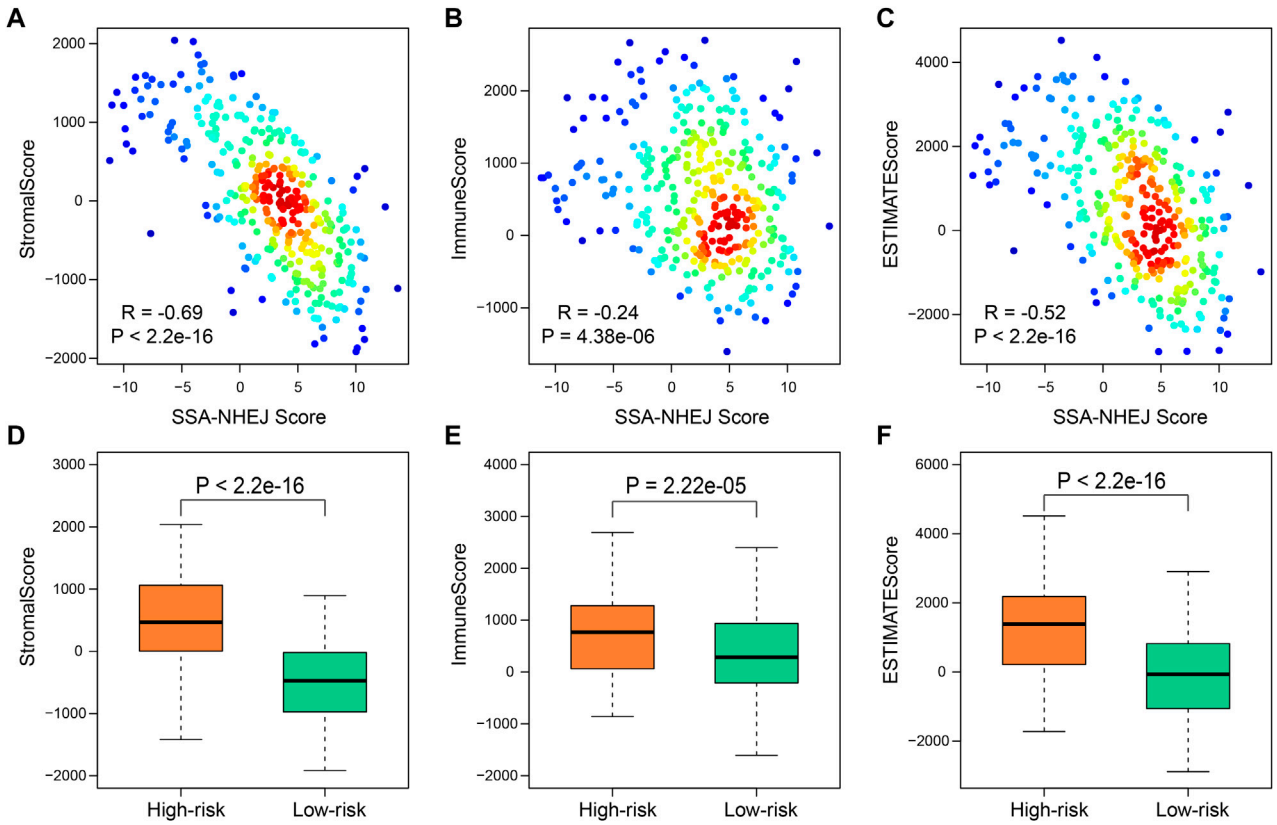


FIGURE 6 | Immune-related score by "estimate". **(A–C)** Correlation between SSA-NHEJ score and StromalScore **(A)**, ImmuneScore **(B)**, and ESTIMATEScore **(C)** in TCGA samples. R and P were calculated by the Pearson test. **(D–F)** Low-risk samples have lower StromalScore **(D)**, ImmuneScore **(E)**, and ESTIMATEScore **(F)** compared with high-risk samples. *p* values were calculated by Wilcoxon test.

calculated the expression of other immune checkpoints. We found that high-risk samples were characterized by a significantly higher expression level of immune checkpoint genes including CD27, CD40, and CD160. A previous study found that tumor samples with high levels of $CD8^+$ tumor-infiltrating lymphocytes were associated with good outcomes in bladder cancer and had significantly higher levels of genes encoding immune checkpoints, such as PD-L1 (Vidotto et al., 2019). Several studies have shown that tumors with a higher TMB can produce more neoantigens, which are more easily recognized by T cells and induce greater immune cytotoxic activity (Rooney et al., 2015). In addition, multiple studies have shown that patients with positive expression of PD-L1 in tumors and infiltration of $CD8^+$ T cells have longer overall survival times (Wang et al., 2018; Wang et al., 2021), as was observed in our analysis.

Cancer-Normal Model Was Developed to Predict the State of an Individual

As described in the first Results subsection, we found that SSA and NHEJ have low activity in normal cells because less replication occurs in normal cells in comparison with cancer cells. As the SSA-NHEJ score is related to the activity of the SSA and NHEJ repair mechanisms, we inferred that normal samples should have lower SSA-NHEJ scores in comparison with cancer samples. This hypothesis was validated, as shown in **Figure 8A**. Normal samples had the lowest SSA-NHEJ scores, while cancer samples associated with good outcomes had the highest SSA-NHEJ scores (**Figure 8A**; all $p < 0.05$). From the receiver operating characteristic plot, we selected 0.008 as the cut-off value for distinguishing cancer samples from normal samples, as 0.008 corresponded to high sensitivity and specificity (**Figure 8B**; AUC = 0.918). The forecasting performance of the Cancer–Normal model was validated in four other independent datasets obtained from the GEO database (**Figure 8C**; AUC = 0.855, 0.902, 0.917, and 0.949 for the GSE13861, GSE139911, GSE33335, and GSE66229 datasets, respectively). Five measures were utilized to evaluate the performance of the Cancer–Normal model, namely, the true positive rate (or sensitivity), 1—the false positive rate (or specificity), accuracy, precision, and the F-measure. All these measures showed that the Cancer–Normal model gives good predictions of the status of individuals (**Figure 8D**).

Evaluation of the HR-LR and Cancer-Normal Models

The prognostic effect of the HR-LR model and the predictive accuracy of the Cancer–Normal model with regard to cancer or

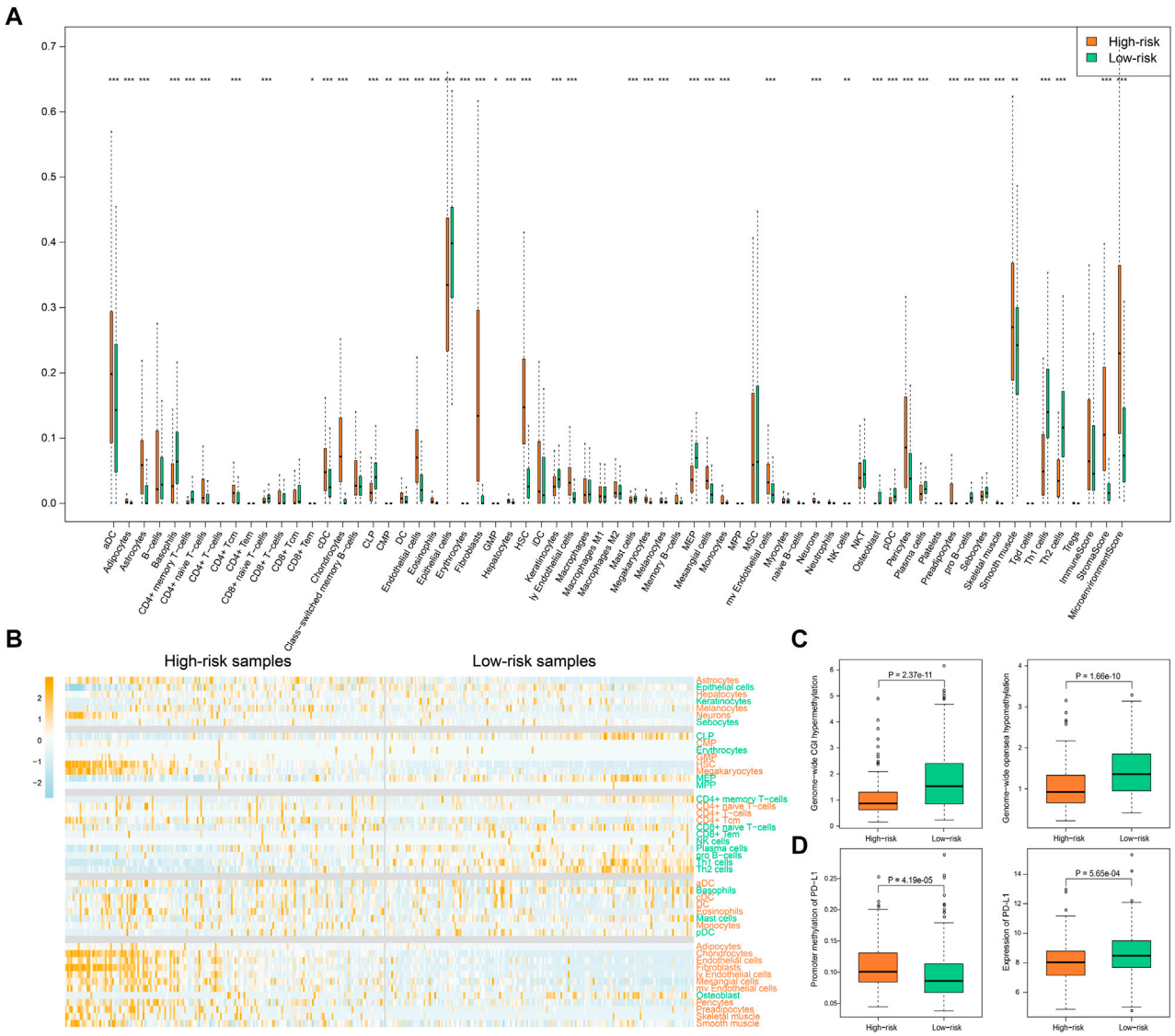


FIGURE 7 | Immune-related score by "xCell". **(A)** The infiltration of immune and stromal cell types as well as the immune-related scores in high-risk and low-risk samples. *p* values less than 0.05, 0.01, and 0.001 are marked with "*", "**", and "***". **(B)** The heatmap of infiltration degree of those immune and stromal cell types with significantly different infiltration in high- and low-risk samples. Cell types marked with orange color represent higher infiltration in high-risk samples while cell types marked with green color represent higher infiltration in low-risk samples. Samples are sorted by SSA-NHEJ score. **(C)** Genome-wide hypermethylation in CGI regions and hypomethylation in open sea regions. **(D)** Promoter methylation and expression of PD-L1 in high-risk and low-risk samples. *p* values were evaluated by Wilcoxon test.

normal health status were validated in the TCGA cohort and various GEO cohorts. We then evaluated these two models by comparing their predictive accuracy with that of random models. By randomly selecting 76 genes from the expression profiles, computing random scores, and selecting a cut-off as in the case of the real models, we constructed a random HR-LR model and Cancer–Normal model. By repeating this process 1000 times, we obtained 1000 random HR-LR models and Cancer–Normal models. We compared the prognostic effects of the random HR-LR models with that of the real HR-LR model. The results showed that the SSA-NHEJ score and the derived class (high-risk vs. low-risk class) were significantly more accurate prognostic factors (**Figure 9A**; $p = 0.03$, $p < 0.001$, and $p < 0.001$ for univariate Cox p of SSA-NHEJ score, univariate Cox p of class and log-rank p of class, respectively). On the other hand, we compared the AUC values of the 1000 random Cancer–Normal models with that of the real Cancer–Normal model. The results showed that the SSA-NHEJ-related Cancer–Normal model had a significantly higher AUC value (**Figure 9B**; $p = 0.003$). Together, these results validated the effectiveness of our HR-LR model and Cancer–Normal model and confirmed that their predictive accuracy was not randomly achieved.

From previous results, we found that the patient disease stage also had a significant prognostic value as well as the SSA-NHEJ

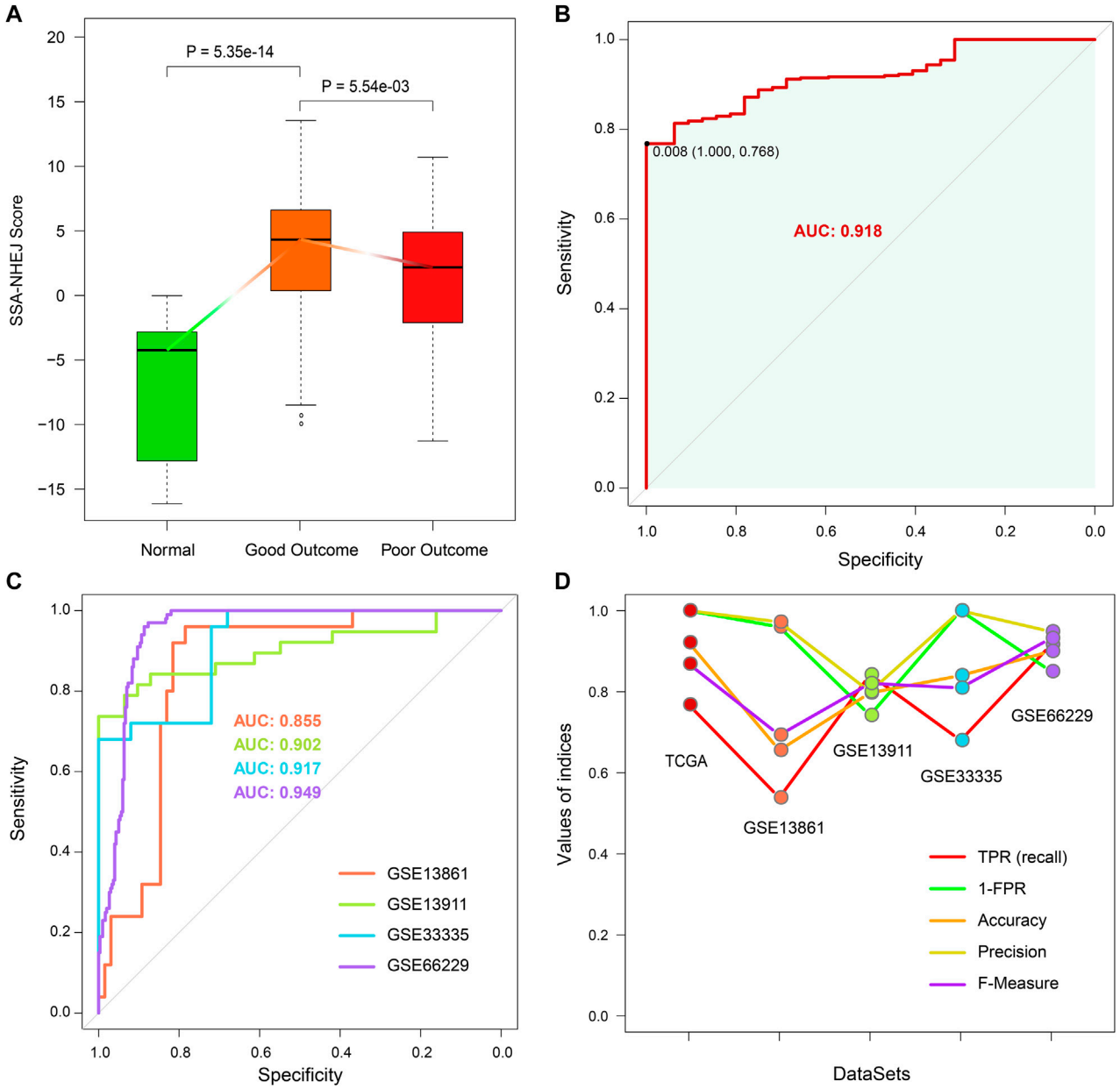


FIGURE 8 | Constructing Cancer-Normal model and predicting the status of samples. **(A)** SSA-NHEJ score of a normal, good outcome, and poor outcome samples. *p* values were evaluated by Wilcoxon test. **(B)** ROC curve of predicting TCGA STAD cancer and normal samples. 0.008 is selected to be the cutoff to predict a sample as normal or cancer status. **(C)** ROC curves of predicting sample status in four independent GEO cohorts. AUC values are listed. **(D)** Five measures of predictive effect for TCGA and four GEO cohorts are listed, including true positive rate, 1—false-positive rate, accuracy, precision, and F-measure.

score (**Figure 3A**). We, therefore, combined the HR-LR model and patient disease stage and obtained three classes of samples with more significantly different clinical outcomes (**Figure 9C**; log-rank $p = 2.29e{-}04$). On the basis of this result, we constructed a decision tree to help predict the clinical outcome for an individual (**Figure 9D**). In addition, a nomogram was constructed to quantify the survival probability for individual GC patients (**Figure 9E**). The SSA-NHEJ-derived class (high-risk or low-risk) and several clinicopathological features were included in the nomogram, such as age, gender, disease stage, and disease grade. The C-index reached 0.702 after 1000 bootstrap iterations in the TCGA cohort (0.734 and 0.728 for the training and test sets, respectively). The calibration curve also indicated good agreement between the estimates and observations, which suggested that our nomogram had a high level of accuracy (**Supplementary Figure S2**). The decision tree and nomogram could contribute to the prognosis in the case of an individual.

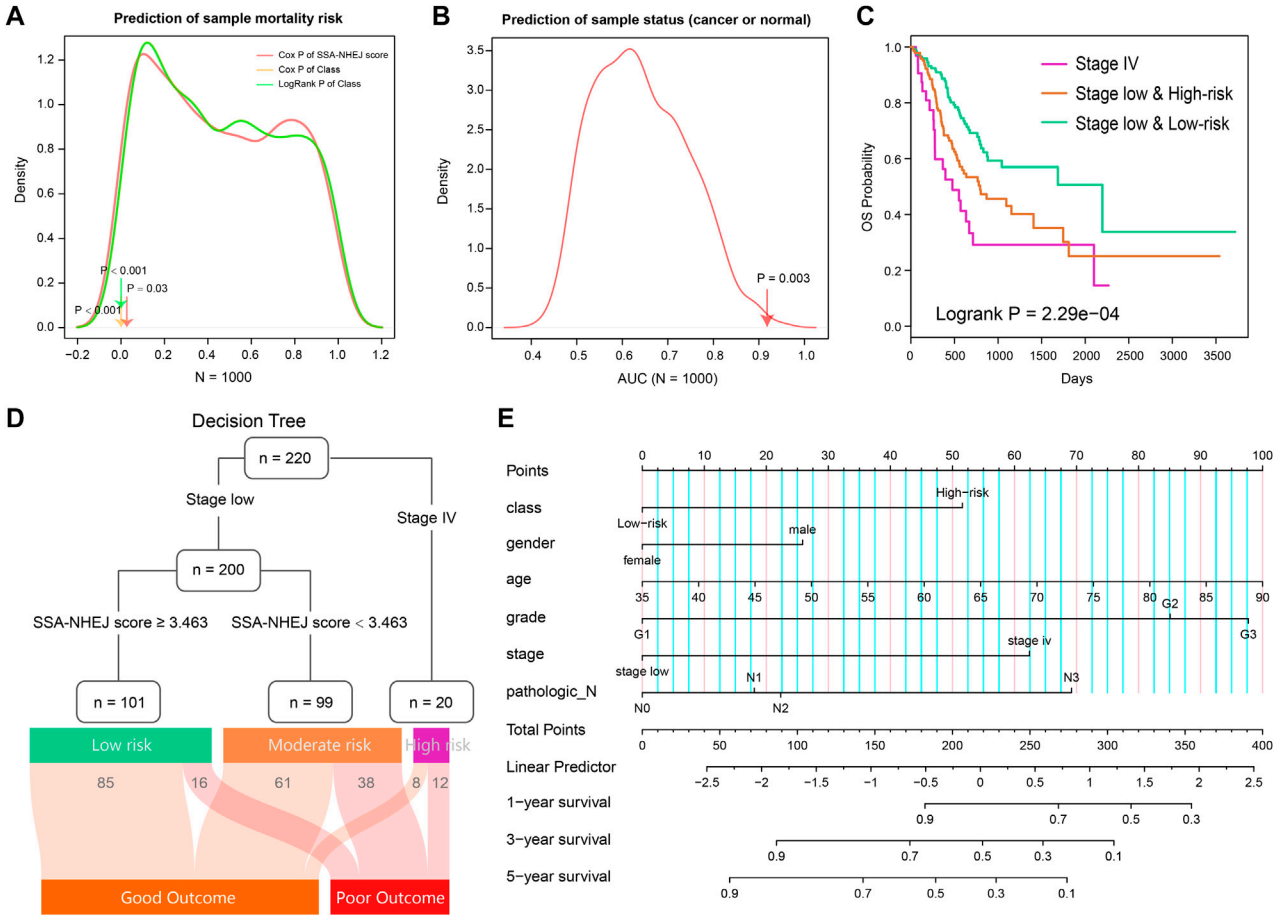


FIGURE 9 | Evaluation of two models and combing clinical features to predict risk assessment for individuals. **(A)** The prognosis of 1000 random HR-LR models based on the expression of random gene sets. **(B)** AUC values of 1000 random Cancer-Normal models based on the expression of random gene sets. **(C)** Survival plot of TCGA STAD patients classified by stage and SSA-NHEJ score. **(D)** Decision tree to predict the patient's clinical outcome. The result of samples in the TCGA cohort is listed at to bottom. **(E)** A nomogram plot is constructed to quantify risk assessment for an individual patient.

DISCUSSION

DNA repair is a vital biological process in normal physiological conditions and includes various types of repair approaches, such as base excision repair, mismatch repair, and DSB repair. Some repair methods could reduce the number of mutations in individuals, while other repair approaches such as SSA and NHEJ might result in fairly large errors and lead to the accumulation of many somatic mutations (Negrini et al., 2010; Lu et al., 2014). In this study, we extracted GO processes and pathways related to DNA repair from the MSigDB. By combining clinical data from patients in the TCGA STAD cohort, we found that clinical outcomes in GC patients were significantly associated with DSB repair and were especially strongly correlated with the SSA and NHEJ approaches. By the Pearson correlation test, we selected 76 genes with significant correlations with the ssGSEA scores of SSA and NHEJ. These 76 genes comprised 37 positively correlated genes and 39 negatively correlated genes, which were further subjected to the *t*-test to obtain the value of the *t*-statistic for each sample. We referred to the value of the *t*-statistic as the SSA-NHEJ score. Follow-up analyses showed that the SSA-NHEJ score was a valuable prognostic factor for overall survival and recurrence in GC patients. In addition, the SSA-NHEJ score could also predict whether an individual had GC.

Numerous studies have demonstrated that the SSA and NHEJ approach in DSB repair would lead to the accumulation of somatic mutations in comparison with homologous recombination repair (Turner et al., 2004; Yang et al., 2011). In our study, we further confirmed this conclusion. Samples with higher SSA and NHEJ activities (or ssGSEA scores) were found to have higher TMBs. On the other hand, we found that SSA and NHEJ activities were significantly higher in cancer samples with good outcomes in comparison with normal samples, but the increase was significantly less in cancer samples with poor outcomes. On the basis of this phenomenon, we suggest that in normal physiological conditions biological activities such as DNA replication are maintained within normal limits, and hence the SSA and NHEJ approaches are

not much needed. However, in cancer cells, DNA replication and other activities increase rapidly, which leads to increases in the activities of the SSA and NHEJ processes. On the other hand, the higher activities of SSA and NHEJ and the higher TMB would lead to the apoptosis of cancer cells (Lu et al., 2014; Hu et al., 2019) or the production of new antigens and finally result in good outcomes in patients (Klempner et al., 2020). In contrast, in other GC patients lower activities of SSA and NHEJ finally, lead to poor outcomes.

There have also been many studies that focused on the relationship between TMB and immune infiltration. Some studies found that a higher TMB is associated with more immune infiltration, while some studies obtained the opposite result. In this study, low-risk samples were found to have a higher TMB but less infiltration of general immune cells. However, significantly higher abundances of $CD8^+$ T cells and $CD4^+$ memory T cells were observed in low-risk samples. In addition, we found that PD-L1 had higher expression levels in low-risk samples. Similarly, other investigators also found that CD8 positivity is significantly associated with PD-L1 expression (Vidotto et al., 2019). The relationship between the TMB and immune infiltration and their roles in GC needs to be further investigated.

The HR-LR model was constructed on the basis of the training set of sequencing data from the Illumina platform. The predictive accuracy of the model was validated in the test set, which also comprised sequencing data. Furthermore, the prognostic value of the HR-LR model was validated in several GEO datasets, which contained microarray data from different platforms. Similarly, the Cancer–Normal model was also constructed on the basis of Illumina sequencing data and validated in microarray data from different platforms. In addition, we validated the accuracy of the two models by comparing them with random models. These results proved that our HR-LR model and Cancer–Normal model had stable accuracy and could be used universally on different platforms. Combining the models with clinical features will contribute to the prognosis in GC patients.

In summary, we found that SSA and NHEJ are vital prognostic factors in GC, proposed two models to help predict clinical outcomes in GC patients, and investigated the relationships among the SSA and NHEJ approaches, the TMB and immune infiltration, and their roles in GC. The present study aims to provide an improved understanding of the complexity of DNA repair, the TMB, and immune infiltration in GC and to contribute to the development of clinical diagnosis and treatment.

CONCLUSION

In summary, we found that SSA and NHEJ were the most prognostically effective DNA repair processes in GC patients. On the basis of the activities of these two approaches and expression profiles, in this study, we proposed two models to help predict clinical outcomes in GC patients and investigated the relationships among the SSA and NHEJ approaches, the TMB and immune infiltration, and their roles in GC. Moreover, we estimated methylation levels in each STAD sample. The present study aims to provide novel insights for the understanding of the complexity of DNA repair, the TMB, and immune infiltration in GC and for further investigation of their diagnostic value.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

SF and MC conceived and designed the experiments. LW, JL, YS, JB, WS, and JY collected and analyzed data. SF, LW, and JL wrote this manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the research foundation of a key laboratory for the preservation of human genetic resources and disease control in China (Harbin Medical University), Ministry of Education, China, HMU Marshal Initiative Funding (No. HMUMIF-21007), and China Postdoctoral Science Foundation (2020T130163).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2022.897096/full#supplementary-material

REFERENCES

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., et al. (2013). Signatures of Mutational Processes in Human Cancer. *Nature* 500 (7463), 415–421. doi:10.1038/nature12477

Aran, D., Hu, Z., and Butte, A. J. (2017). xCell: Digitally Portraying the Tissue Cellular Heterogeneity Landscape. *Genome Biol.* 18 (1), 220. doi:10.1186/s13059-017-1349-1

Belfield, E. J., Ding, Z. J., Jamieson, F. J. C., Visscher, A. M., Zheng, S. J., Mithani, A., et al. (2018). DNA Mismatch Repair Preferentially Protects Genes from Mutation. *Genome Res.* 28 (1), 66–74. doi:10.1101/gr.219303.116

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

Bouwman, P., and Jonkers, J. (2012). The Effects of Deregulated DNA Damage Signalling on Cancer Chemotherapy Response and Resistance. *Nat. Rev. Cancer* 12 (9), 587–598. doi:10.1038/nrc3342

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

Castle, J. C., Kreiter, S., Diekmann, J., Löwer, M., van de Roemer, N., de Graaf, J., et al. (2012). Exploiting the Mutanome for Tumor Vaccination. *Cancer Res.* 72 (5), 1081–1091. doi:10.1158/0008-5472.CAN-11-3722

Emran, A. A., Chatterjee, A., Rodger, E. J., Tiffen, J. C., Gallagher, S. J., Eccles, M. R., et al. (2019). Targeting DNA Methylation and EZH2 Activity to Overcome Melanoma Resistance to Immunotherapy. *Trends Immunol.* 40 (4), 328–344. doi:10.1016/j.it.2019.02.004

Fang, Z., Jiang, C., and Li, S. (2020). The Potential Regulatory Roles of Circular RNAs in Tumor Immunology and Immunotherapy. *Front. Immunol.* 11, 617583. doi:10.3389/fimmu.2020.617583

Gibney, G. T., Weiner, L. M., and Atkins, M. B. (2016). Predictive Biomarkers for Checkpoint Inhibitor-Based Immunotherapy. *Lancet Oncol.* 17 (12), e542–e551. doi:10.1016/S1470-2045(16)30406-5

Gillyard, T., and Davis, J. (2021). DNA Double-Strand Break Repair in Cancer: A Path to Achieving Precision Medicine. *Int. Rev. Cel Mol Biol* 364, 111–137. doi:10.1016/bs.ircmb.2021.06.003

Hu, W., Jiang, C., Guan, D., Dierickx, P., Zhang, R., Moscati, A., et al. (2019). Patient Adipose Stem Cell-Derived Adipocytes Reveal Genetic Variation that Predicts Antidiabetic Drug Response. *Cell Stem Cell* 24 (2), 299–308. e296. doi:10.1016/j.stem.2018.11.018

Klempner, S. J., Fabrizio, D., Bane, S., Reinhart, M., Peoples, T., Ali, S. M., et al. (2020). Tumor Mutational Burden as a Predictive Biomarker for Response to Immune Checkpoint Inhibitors: A Review of Current Evidence. *Oncologist* 25 (1), e147–e159. doi:10.1634/theoncologist.2019-0244

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

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) Hallmark Gene Set Collection. *Cel Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004

Lu, J., Wu, D., Li, C., Zhou, M., and Hao, D. (2014). Correlation between Gene Expression and Mutator Phenotype Predicts Homologous Recombination Deficiency and Outcome in Ovarian Cancer. *J. Mol. Med.* 92 (11), 1159–1168. doi:10.1007/s00109-014-1191-9

Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: Efficient and Comprehensive Analysis of Somatic Variants in Cancer. *Genome Res.* 28 (11), 1747–1756. doi:10.1101/gr.239244.118

McCarthy, A. J., Capo-Chichi, J. M., Spence, T., Grenier, S., Stockley, T., Kamel-Reid, S., et al. (2019). Heterogenous Loss of Mismatch Repair (MMR) Protein Expression: A Challenge for Immunohistochemical Interpretation and Microsatellite Instability (MSI) Evaluation. *J. Pathol. Clin. Res.* 5 (2), 115–129. doi:10.1002/cjp2.120

McGranahan, N., Furness, A. J. S., Rosenthal, R., Ramskov, S., Lyngaa, R., Saini, S. K., et al. (2016). Clonal Neoantigens Elicit T Cell Immunoreactivity and Sensitivity to Immune Checkpoint Blockade. *Science* 351 (6280), 1463–1469. doi:10.1126/science.aaf1490

Negrini, S., Gorgoulis, V. G., and Halazonetis, T. D. (2010). Genomic Instability-Aan Evolving Hallmark of Cancer. *Nat. Rev. Mol. Cel Biol* 11 (3), 220–228. doi:10.1038/nrm2858

Parikh, A. R., He, Y., Hong, T. S., Corcoran, R. B., Clark, J. W., Ryan, D. P., et al. (2019). Analysis of DNA Damage Response Gene Alterations and Tumor Mutational Burden across 17,486 Tubular Gastrointestinal Carcinomas: Implications for Therapy. *Oncologist* 24 (10), 1340–1347. doi:10.1634/theoncologist.2019-0034

Patel, S. P., and Kurzrock, R. (2015). PD-L1 Expression as a Predictive Biomarker in Cancer Immunotherapy. *Mol. Cancer Ther.* 14 (4), 847–856. doi:10.1158/1535-7163.MCT-14-0983

Reiser, J., and Banerjee, A. (2016). Effector, Memory, and Dysfunctional CD8(+) T Cell Fates in the Antitumor Immune Response. *J. Immunol. Res.* 2016, 8941260. doi:10.1155/2016/8941260

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., et al. (2011). pROC: An Open-Source Package for R and S+ to Analyze and Compare ROC Curves. *BMC Bioinform.* 12, 77. doi:10.1186/1471-2105-12-77

Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G., and Hacohen, N. (2015). Molecular and Genetic Properties of Tumors Associated with Local Immune Cytolytic Activity. *Cell* 160 (1-2), 48–61. doi:10.1016/j.cell.2014.12.033

Stok, C., Kok, Y. P., van den Tempel, N., and van Vugt, M. A. T. M. (2021). Shaping the BRCAness Mutational Landscape by Alternative Double-Strand Break Repair, Replication Stress and Mitotic Aberrancies. *Nucleic Acids Res.* 49 (8), 4239–4257. doi:10.1093/nar/gkab151

Turner, N., Tutt, A., and Ashworth, A. (2004). Hallmarks of 'BRCAness' in Sporadic Cancers. *Nat. Rev. Cancer* 4 (10), 814–819. doi:10.1038/nrc1457

Vidotto, T., Nersesian, S., Graham, C., Siemens, D. R., and Koti, M. (2019). DNA Damage Repair Gene Mutations and Their Association with Tumor Immune Regulatory Gene Expression in Muscle Invasive Bladder Cancer Subtypes. *J. Immunother. Cancer* 7 (1), 148. doi:10.1186/s40425-019-0619-8

Wang, J., Li, R., Cao, Y., Gu, Y., Fang, H., Fei, Y., et al. (2021). Intratumoral CXCR5(+)CD8(+)T Associates with Favorable Clinical Outcomes and Immunogenic Contexture in Gastric Cancer. *Nat. Commun.* 12 (1), 3080. doi:10.1038/s41467-021-23356-w

Wang, Y., Zhu, C., Song, W., Li, J., Zhao, G., and Cao, H. (2018). PD-L1 Expression and CD8(+) T Cell Infiltration Predict a Favorable Prognosis in Advanced Gastric Cancer. *J. Immunol. Res.* 2018, 4180517. doi:10.1155/2018/4180517

Wolters, S., and Schumacher, B. (2013). Genome Maintenance and Transcription Integrity in Aging and Disease. *Front. Genet.* 4, 19. doi:10.3389/fgene.2013.00019

Yang, D., Khan, S., Sun, Y., Hess, K., Shmulevich, I., Sood, A. K., et al. (2011). Association of BRCA1 and BRCA2 Mutations with Survival, Chemotherapy Sensitivity, and Gene Mutator Phenotype in Patients with Ovarian Cancer. *JAMA* 306 (14), 1557–1565. doi:10.1001/jama.2011.1456

Yang, Z., Jones, A., Widschwendter, M., and Teschendorff, A. E. (2015). An Integrative Pan-Cancer-Wide Analysis of Epigenetic Enzymes Reveals Universal Patterns of Epigenomic Deregulation in Cancer. *Genome Biol.* 16, 140. doi:10.1186/s13059-015-0699-9

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring Tumour Purity and Stromal and Immune Cell Admixture from Expression Data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612

Zhang, Z., and Kattan, M. W. (2017). Drawing Nomograms with R: Applications to Categorical Outcome and Survival Data. *Ann. Transl. Med.* 5 (10), 211. doi:10.21037/atm.2017.04.01

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Cell and Developmental Biology

ORIGINAL RESEARCH
published: 27 May 2022
doi: 10.3389/fcell.2022.882698

Pan-Cancer Methylated Dysregulation of Long Non-coding RNAs Reveals Epigenetic Biomarkers

Ning Zhao[1], *Maozu Guo*[2]*, *Chunlong Zhang*[3], *Chunyu Wang*[4] *and Kuanquan Wang*[1,4]

[1]School of Life Science and Technology, Harbin Institute of Technology, Harbin, China, [2]School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing, China, [3]College of Information and Computer Engineering, Northeast Forest University, Harbin, China, [4]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China

Different cancer types not only have common characteristics but also have their own characteristics respectively. The mechanism of these specific and common characteristics is still unclear. Pan-cancer analysis can help understand the similarities and differences among cancer types by systematically describing different patterns in cancers and identifying cancer-specific and cancer-common molecular biomarkers. While long non-coding RNAs (lncRNAs) are key cancer modulators, there is still a lack of pan-cancer analysis for lncRNA methylation dysregulation. In this study, we integrated lncRNA methylation, lncRNA expression and mRNA expression data to illuminate specific and common lncRNA methylation patterns in 23 cancer types. Then, we screened aberrantly methylated lncRNAs that negatively regulated lncRNA expression and mapped them to the ceRNA relationship for further validation. 29 lncRNAs were identified as diagnostic biomarkers for their corresponding cancer types, with lncRNA *AC027601* was identified as a new KIRC-associated biomarker, and lncRNA *ACTA2-AS1* was regarded as a carcinogenic factor of KIRP. Two lncRNAs *HOXA-AS2* and *AC007228* were identified as pan-cancer biomarkers. In general, the cancer-specific and cancer-common lncRNA biomarkers identified in this study may aid in cancer diagnosis and treatment.

Keywords: lncRNA, pan-cancer, DNA methylation, biomarker, ceRNA

OPEN ACCESS

Edited by:
Chunjie Jiang,
University of Pennsylvania,
United States

Reviewed by:
Qinglan Li,
University of Pennsylvania,
United States
Xiaowen Chen,
Harbin Medical University, China
Meng Zhou,
Wenzhou Medical University, China

***Correspondence:**
Maozu Guo
guomaozu@bucea.edu.cn

Specialty section:
This article was submitted to Molecular and Cellular Pathology, a section of the journal Frontiers in Cell and Developmental Biology

Received: *24 February 2022*
Accepted: *28 April 2022*
Published: *27 May 2022*

Citation:
Zhao N, Guo M, Zhang C, Wang C and Wang K (2022) Pan-Cancer Methylated Dysregulation of Long Non-coding RNAs Reveals Epigenetic Biomarkers. Front. Cell Dev. Biol. 10:882698. doi: 10.3389/fcell.2022.882698

INTRODUCTION

Cancer is a general term that refers to malignant tumors, and its world-wide incidence and mortality have been high for many years. In 2020, there were 19.29 million new cancer cases and 9.96 million cancer deaths world-wide (Sung et al., 2021). There is no cure for cancer at the moment. Not only do different types of cancer have common biological characteristics such as abnormal cell differentiation and proliferation, lack of growth control, invasion and metastasis, but they also have many specifically biological characteristics, respectively. Therefore, it is necessary to conduct pan-cancer research on a variety of cancer types to ascertain the similarities and differences in molecular characteristics across different cancers.

DNA methylation is an important epigenetic modification that plays an important role in many physiological processes (Martin-Subero, 2011; Li et al., 2013; Neri et al., 2017), and its aberrant behavior can result in gene instability, proto-oncogene activation and tumor suppressor gene inactivation (Li et al., 2012; Hou et al., 2021). Aberrant DNA methylation occurs in almost all

cancers, where unmethylated promoters become methylated or methylated sequences lose their methylation. Focusing on the pan-cancer analysis of DNA methylation can inform research and therapy. Saghafinia et al. (Saghafinia et al., 2018) described an algorithmic strategy for identifying pan-cancer-related DNA methylation alteration affecting gene expression. Numerous DNA methylation dysregulation were discovered to be associated with patient prognosis and therapeutic response. Methylated research in cancer patients has been shown to improve and maintain the efficiency of cancer treatment (Pauken et al., 2016; Sen et al., 2016; Marwitz et al., 2017).

Long non-coding RNAs (lncRNA) are non-coding RNAs with a length of more than 200 nucleotides. In comparison to protein-coding genes, they have a high degree of tissue specificity in their expression (Cabili et al., 2011). LncRNAs play an important role in the occurrence and development of cancer and many other complex diseases (Zhou et al., 2019). LncRNAs are associated with practically every major cancer type and contribute to all ten hallmarks of cancer (Huarte, 2015; Bartonicek et al., 2016; Esposito et al., 2019; Bao et al., 2020). Additionally, lncRNAs play an important role in immune regulation (Chen et al., 2017; Zhang et al., 2021b). Lin et al. (Lin and Yang, 2018) explored the mechanisms by which lncRNAs regulated cellular responses to extracellular signals and their clinical potential as diagnostic indicators, stratification markers, and therapeutic targets for combinatorial treatments. Zhang et al. (Zhang et al., 2018) identified lncRNA *MT1JP* as a ceRNA for the tumor suppressor *FBXW7* in gastric cancer by demonstrating competitive binding with MiR-92A-3p. Xu et al. (Xu et al., 2019) discovered that lncRNA *SNHG6* regulated the expression of the oncogene *EZH2* in colorectal cancer *via* the ceRNA sponge-associated with MiR-26a/b and MiR-214. Chen et al. (Chen et al., 2019) discovered that the lncRNA *PVT1* promoted tumor development in gallbladder cancer by regulating the miR-143/HK2 axis. Wang et al. (Wang et al., 2017) identified the lncRNA *HOXD-AS1* as a ceRNA that regulated *SOX4* and promoted liver cancer metastasis. The above studies established the role of lncRNAs in corresponding cancer types, and identified the cancer-associated lncRNAs for each cancer type (Dong et al., 2019). Pan-cancer analysis of lncRNAs can help in identifying the similarities and differences between distinct cancer types and identifying potential therapeutic targets for cancer treatment.

Numerous pan-cancer studies have been carried out on lncRNAs. Li et al. (Yongsheng Li et al., 2020) identified multiple pan-cancer immune-associated lncRNAs as potential oncogenic biomarkers. Martens-Uzunova et al. (Martens-Uzunova et al., 2014) summarized the role of lncRNAs in the diagnosis and treatment of urinary tumors, and concluded that lncRNAs could be used as new biomarkers for prostate cancer, kidney cancer and bladder cancer. Zhang et al. (Zhang et al., 2021) identified clinically distinct tumor subtypes by characterizing pan-cancer lncRNA modifiers of the immune microenvironment. Bao et al. (Bao et al., 2021) proposed a framework for identifying lncRNA signatures associated with pan-cancer prognosis.

Previous studies have established a correlation between lncRNAs and epigenetic regulation (Spizzo et al., 2012; Xu et al., 2018), suggesting that they regulates chromatin state and epigenetic inheritance (Tsai et al., 2010). Lu et al. (Lu et al., 2020) found that DNA methylation-mediated lncRNA activation improved temozolomide resistance in glioblastoma, implying that *SNHG12* could be a therapeutic target for overcoming temozolomide tolerance.

Detecting the dynamic pattern of lncRNA methylation during cancer development across pan-cancer may help highlight epigenetic changes and aid in cancer diagnosis and treatment. Yang et al. (Yang et al., 2021) presented a novel integrative analysis framework, termed MeLncTRN for integrating data on gene expression, copy number variation, methylation and lncRNA expression. They identified epigenetically-driven lncRNA-gene regulation circuits across 18 cancer types. Wei et al. (Wei et al., 2019) constructed a systematic biological framework to evaluate the co-methylation events between two lincRNAs in nine cancer types. The lincRNA prognostic signatures were identified to significantly correlate with overall survival in cancers. Wang et al. (Wang et al., 2018) characterized the epigenetic landscape of genes encoding lncRNAs associated with pan-cancer and identified *EPIC1* as an oncogenic lncRNA. Xu et al. (Xu et al., 2021) constructed networks of lncRNA-associated dysregulated ceRNA across eight cancer types. They screened nine pan-cancer epigenetically related lncRNAs.

However, no research has been conducted to systematically compare methylation changes of lncRNAs in pan-cancer to identify the specific methylation-related lncRNAs. In this study, we used pan-cancer lncRNA methylation data from the TCGA to examine the lncRNA methylation patterns of 23 cancer types and identified differentially methylated lncRNAs (DMlncs). Subsequently, we examined differentially methylated lncRNAs

TABLE 1 | The number of samples and lncRNAs for each cancer in methylation data.

Cancer	No. of tumor samples	No. of normal samples	No. of lncRNAs
BLCA	412	21	4,317
BRCA	783	96	4,314
CESC	307	3	4,314
CHOL	36	9	4,309
COAD	296	38	4,317
ESCA	185	16	4,317
GBM	141	2	4,317
HNSC	528	50	4,317
KIRC	319	160	4,317
KIRP	275	45	4,317
LIHC	377	50	4,317
LUAD	458	32	4,316
LUSC	370	42	4,317
PAAD	184	10	4,317
PCPG	179	3	4,315
PRAD	498	50	4,317
READ	98	7	4,316
SARC	261	4	4,310
SKCM	470	2	4,317
STAD	395	2	4,314
THCA	507	56	4,314
THYM	124	2	4,316
UCEC	431	46	4,314

from different cancer types to identify cancer-specific and cancer-common differentially methylated lncRNAs. Further, combining lncRNA expression data with survival data, lncRNAs with a negative correlation between methylation and expression dysregulation were found as diagnostic biomarkers for each cancer. Finally, the lncRNAs were mapped into the ceRNA network to establish ceRNA relationships with mRNAs confirming their important roles in cancer.

MATERIALS AND METHODS

Data

Pan-cancer DNA methylation data for lncRNAs from Infinium 450k arrays were downloaded from the TCGA database. The cancer types with normal samples were selected, and a total of 7,634 tumor samples and 746 normal samples from 23 cancer types were retained. **Table 1** shows the number of tumor and normal samples, as well as the number of lncRNAs. The methylation status of each probe in each sample was measured using the β-value (Aryee et al., 2014). The β-value denoted the ratio of methylation intensity of the probe to total intensity, with a range of 0 (low methylation) to 1 (high methylation). The probes with β-values greater than 0 in more than 50% of the samples were retained in the methylation profile for each cancer, and the missing values were filled with the average of all non-zero values on the probes. The average β-value of the promoter region was used to determine the methylation level of each lncRNA.

Expression data of lncRNAs and mRNAs for 13 cancer types were downloaded from the TANRIC and the TCGA databases, respectively (Li et al., 2015). The number of tumor and normal samples and the number of lncRNAs and mRNAs are shown in **Supplementary Tables S1, S2**. Each lncRNA and mRNA expression value was defined as its reads per kilobase per million mapped reads (RPKM) (Mortazavi et al., 2008). Subsequently, we transformed the expression data by log2 (RPKM+1), reserved the lncRNAs and mRNAs with expression values in more than 70% of the samples, and filled their missing values using the average expression values of these RNAs in the samples.

The expression data of mRNA and lncRNA was downloaded from different databases, so we got the human gene annotation files from the GENCODE database (https://www.gencodegenes.org/) to obtain the corresponding relations of ENSG IDs and gene symbols. Then, using Entrez IDs as the main reference, different versions of human gene names (Entrez IDs, gene symbols and ENSG IDs) were converted to standard human gene names.

Identification of DMlncs

The DMlncs for each cancer were first screened using the following formula Δβ:

$$\Delta\beta = \left|\bar{\beta}_t - \bar{\beta}_n\right| \quad (1)$$

Where, $\bar{\beta}_t$ and $\bar{\beta}_n$ denoted the average level of methylation in tumor and normal samples, respectively. Δβ was the subtraction difference in average methylation levels between tumor and normal samples. lncRNAs with Δβ>0.1 were selected as candidates for DMlncs.

Additionally, the "limma" package (Ritchie et al., 2015) in R language was used to measure the degree of difference between tumor and normal samples. The lncRNAs with $|log_2(fold\ change)|>1$ (PCAWG Transcriptome Core Group et al., 2020) and FDR<0.05 were identified as DMlncs.

Identification of Differentially Expressed lncRNAs

The "limma" package was used to calculate differential expression between tumor and normal samples for lncRNA expression data. We took the lncRNAs with $|log_2(fold\ change)|>1$ and FDR<0.05 as differentially expressed lncRNAs.

Functional Enrichment Analysis

The GREAT software (McLean et al., 2010) was used to conduct functional enrichment analysis on lncRNAs. We took the lncRNA BED data as input. The lncRNA BED information included the chromosome, start site and end site extracted from GENCODE database. Gene ontology (GO) functions of the output results were selected for subsequent analysis.

Recognition of ceRNAs

LncRNA-miRNA and mRNA-miRNA targeted relationships were downloaded from the ENCORI platform (Li et al., 2014). For each pair of lncRNA and mRNA, the intersection of their target miRNAs should be more than two (Zhang et al., 2021a). The intersections of their miRNA lists were subjected to the hypergeometric test, and the lncRNA-mRNA pairs with a *p*-value less than 0.05 were considered. A total of 4,739,668 pairs were obtained.

Subsequently, the Pearson correlation coefficient between lncRNA expression and mRNA expression was calculated. LncRNA-mRNA pairs with Pearson correlation coefficient >0.3 and *p*-value<0.05 were selected as ceRNAs. **Supplementary Table S3** shows the number of lncRNA-mRNA pairs, lncRNAs and mRNAs in each cancer.

Survival Analysis

Survival analysis of patients was carried out using the "survival" package in R language, in which the maxstat model was used to evaluate the best cut-off point to divide high-risk and low-risk groups. Kaplan-Meier curves were then drawn to depict the survival of patients of high-risk and low-risk groups.

Immunological Score

Three scores were used to assess the immunological effect of lncRNAs: Major Histocompatibility Complex (MHC), Cytolytic Activity (CYT) and Cytotoxic T Lymphocyte (CTL). The MHC score of each sample was calculated as:

$$Score_{MHC_i} = \left(\sum exp_n\right)/9 \quad (2)$$

Where *i* denoted the sample, exp_n represented the expression of gene *n*, and *n* was one of the nine genes (*HLA-A, PSMB9, HLA-B,*

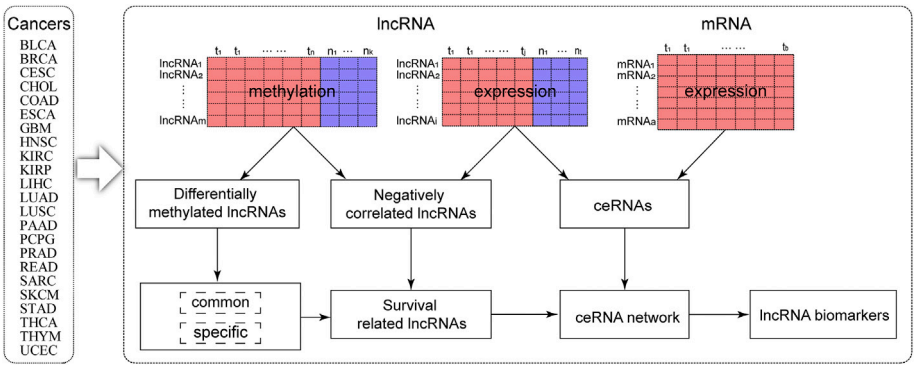


FIGURE 1 | The framework for identifying the lncRNA methylation biomarkers.

PSMB8, HLA-C, B2M, TAP2, NLRC5, and *TAP1*). These nine genes had a strong correlation and were the core gene set of MHC-I (Rooney et al., 2015; Lauss et al., 2017).

The CYT score of each sample was calculated as follows:

$$Score_{CYT_i} = \left(exp_{GZMA} + exp_{PRF1}\right)/2 \quad (3)$$

In which, i represented the sample, exp_{GZMA} and exp_{PRF1} represented the expression of *GZMA* and *PRF1*, respectively. These two genes were key factors of cytolysis and were upregulated in activated CD8+T cells and strongly responded to *CTLA4* and *PDCD1* immunotherapy (Narayanan et al., 2018).

The CTL score of each sample was calculated as follows:

$$Score_{CTL_i} = \left(exp_{GZMA} + exp_{PRF1} + exp_{GZMB}\right)/3. \quad (4)$$

In which, i represented the sample, exp_{GZMA}, exp_{PRF1} and exp_{GZMB} represented the expression of *GZMA*, *PRF1*, and *GZMB*, respectively. These three genes were important factors to measure T cell toxicity and immune cell effector function (Basu et al., 2016).

RESULTS

The Workflow of DMlncs Identification

In this study, the identification flow of methylation-related lncRNA biomarkers for pan-cancer is shown in **Figure 1**. We conducted a study on a total of 23 cancer types. Firstly, DMlncs between tumor and normal samples were obtained using lncRNA methylation data. Specific and common lncRNAs were identified for pan-cancer. Subsequently, lncRNA methylation data and lncRNA expression data were integrated to identify DMlncs whose methylation changes were negatively correlated with expression changes. Following that, prognostic lncRNAs were screened and mapped into the ceRNA network. The ceRNA network was constructed by combining lncRNA and mRNA expression data. Finally, pan-cancer lncRNA biomarkers were identified by analyzing the lncRNAs of the ceRNA network.

Identification of Pan-Cancer DMlncs

To identify DMlncs, we downloaded cancer methylation profiles of 23 cancer types from the TCGA, including both tumor and normal samples. The data of 7,634 tumor samples and 746 normal samples was downloaded. The proportion of tumor samples for each cancer is shown in **Figure 2A**. BRCA had the largest number of samples, which was nearly double that of other cancer types, while CHOL had the lowest number of samples. A total of 7,542 tumor samples had corresponding clinical data of the patients. As shown in **Figure 2B**, the age, sex, and survival status of patients were analyzed. The male to female ratio of the patients was 1:1, and the age was concentrated among the elderly, which was consistent with the law of the general onset age of cancer. The majority of patients survived following treatment. **Figure 2C** shows the mortality rate of tumor patients, three-quarters of the patients survived after surgery. GBM had the highest mortality rate, followed by CHOL. PRAD had the lowest mortality rate.

After standardizing the data, DMlncs for each cancer were identified. The number of DMlncs is shown in **Table 2**. A total of 2,286 DMlncs were obtained. The majority of the cancer types had a large number of DMlncs, and only a few cancer types had a small number of DMlncs.

Additionally, DMlncs were divided into up-methylated and down-methylated groups. Up-methylated lncRNAs were those whose methylation level was elevated in cancer samples compared with normal samples, while down-methylated lncRNAs were the inverse. **Table 2** shows the number of patients in each of the two groups. There were overlaps in the genes of different cancers. In total, there were 1,229 DMlncs in the up-methylated group and 1,654 DMlncs in the down-methylated group. Among them, 597 lncRNAs were up-methylated in some cancers but down-methylated in some other cancers, demonstrating uneven regulation tendencies across different cancers. As shown in **Figure 3**, the proportion of up-methylated and down-methylated lncRNAs varied between cancer types. There were 13 types of cancer had a higher number of down-methylated lncRNAs and ten types of cancer had a higher number of up-methylated lncRNAs. The first few cancer types of most DMlncs had obvious higher number of down-methylated lncRNAs and most of the other cancer types had more up-methylated lncRNAs. Certain cancer types contained only a single type

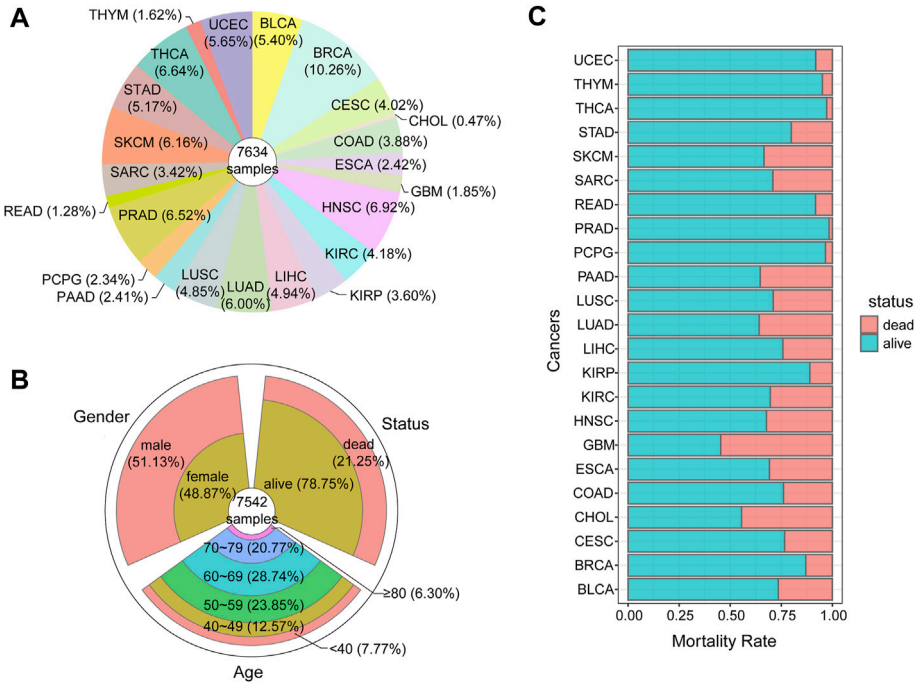


FIGURE 2 | Characteristics of pan-cancer tumor samples **(A)**The proportion of samples collected for each cancer **(B)**The clinical information of tumor samples **(C)** The mortality rate of each cancer.

of DMlncs. For example, all 74 DMlncs in SKCM were up-methylated, and all three DMlncs in THYM were up-methylated as well.

Cancer-Specific lncRNA Biomarkers

As a complex disease, cancer has a high heterogeneity and distinct pathogenesis. In this study, we searched for specific DMlncs for

TABLE 2 | The number of DMlncs of each cancer.

Cancer	Total lncRNAs	Down-methylated lncRNAs	Up-methylated lncRNAs	Cancer-specific up-methylated	Cancer-specific down-methylated
BLCA	1,126	931	195	2	45
BRCA	753	411	342	23	9
CESC	395	99	296	20	8
CHOL	433	80	353	65	23
COAD	762	464	298	28	14
ESCA	569	263	306	28	1
GBM	48	7	41	13	2
HNSC	854	581	273	9	12
KIRC	435	286	149	5	24
KIRP	406	154	252	57	4
LIHC	1,074	936	138	10	114
LUAD	583	349	234	5	1
LUSC	932	674	258	11	18
PAAD	344	150	194	11	9
PCPG	174	112	62	26	47
PRAD	604	236	368	77	13
READ	575	375	200	5	6
SARC	7	1	6	2	1
SKCM	74	0	74	22	0
STAD	8	7	1	0	7
THCA	118	99	19	3	5
THYM	3	0	3	2	0
UCEC	1,015	651	364	32	47

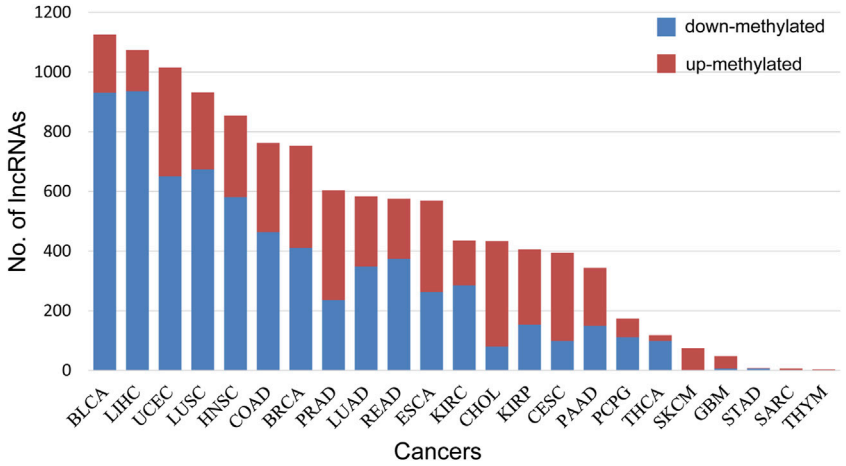


FIGURE 3 | The percentage of up-methylated and down-methylated lncRNAs for each cancer. Blue represents the number of down-methylated lncRNAs, and red represents the number of up-methylated lncRNAs.

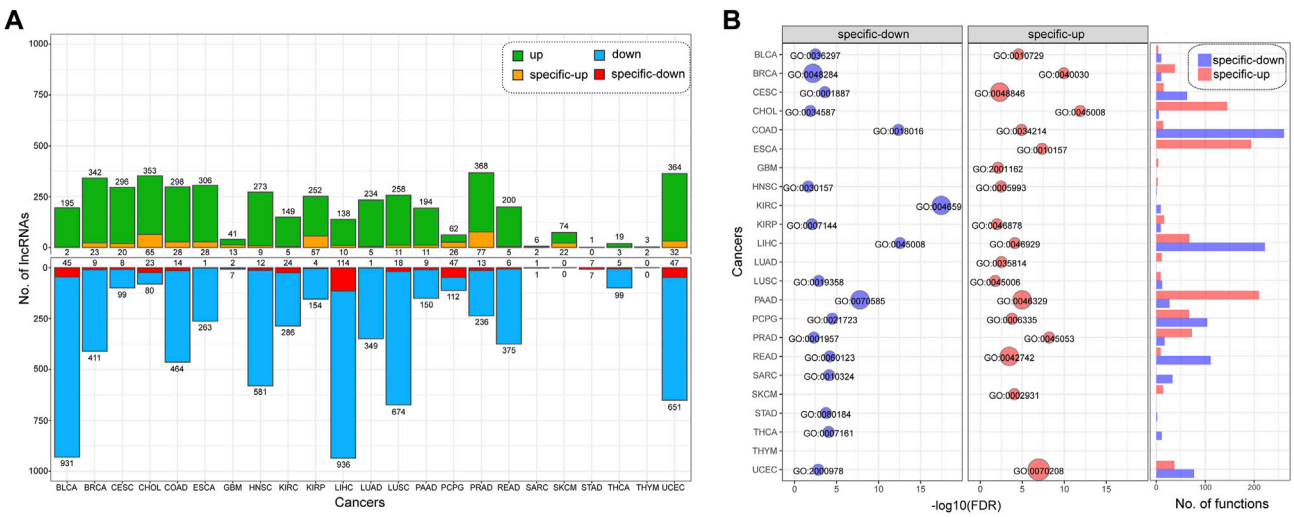


FIGURE 4 | The cancer-specific lncRNAs with differential methylation **(A)**The number and proportion of cancer-specific lncRNAs. Green represents up-methylated lncRNAs, blue represents down-methylated lncRNAs, yellow represents specifically up-methylated lncRNAs, and red represents specifically down-methylated lncRNAs **(B)** The functions of specific lncRNAs in each cancer type. Red represents the functions of up-methylated lncRNAs, while blue represents the functions of down-methylated lncRNAs.

each cancer. **Figure 4A** shows the proportion of cancer-specific lncRNAs, and only a fraction of the DMlncs were cancer-specific. LIHC had the most specifically down-methylated lncRNAs, whereas PRAD had the most specifically up-methylated lncRNAs. The proportion of both specifically up and down lncRNAs in PCPG was about 42%. SARC, STAD, and THCA had DMlncs less than ten, preventing them from being compared with other cancers in terms of lncRNA proportion. Except for SARC, STAD, and THCA, PCPG had the highest proportion of specific lncRNAs. STAD had no up-methylated lncRNAs, while all its seven down-methylated lncRNAs were specific. SARC possessed a single down-methylated lncRNA, and it was specific. Two of the three DMlncs of THYM were specific. **Supplementary Table S4** shows the specific DMlncs of each cancer.

Subsequently, we performed functional enrichment analysis of specifically up-methylated and down-methylated lncRNAs in each cancer. For each cancer, the functions enriched by specifically up-methylated and down-methylated lncRNAs were analyzed separately, and the number of functions enriched by each cancer is shown on the right side of **Figure 4B**. There were significant differences in the number of enriched functions for the cancers. LIHC, COAD, and PAAD enriched in more than 200 functions, while the enriched

TABLE 3 | The functions of specific lncRNAs for each cancer.

Group	Cancer	ID	Function	FDR *q*-value
down	BLCA	GO:0,036,297	interstrand cross-link repair	3.34E-03
up	BLCA	GO:0,010,729	positive regulation of hydrogen peroxide biosynthetic process	2.77E-05
down	BRCA	GO:0,048,284	organelle fusion	6.43E-03
up	BRCA	GO:0,040,030	regulation of molecular function, epigenetic	1.15E-10
down	CESC	GO:0,001,887	selenium compound metabolic process	2.47E-04
up	CESC	GO:0,048,846	axon extension involved in axon guidance	4.69E-03
down	CHOL	GO:0,034,587	piRNA metabolic process	1.32E-02
up	CHOL	GO:0,045,008	Depyrimidination	1.29E-12
down	COAD	GO:0,018,016	N-terminal peptidyl-proline dimethylation	4.41E-13
up	COAD	GO:0,034,214	protein hexamerization	1.23E-05
up	ESCA	GO:0,010,157	response to chlorate	4.36E-08
up	GBM	GO:2,001,162	positive regulation of histone H3-K79 methylation	7.71E-03
down	HNSC	GO:0,030,157	pancreatic juice secretion	2.35E-02
up	HNSC	GO:0,005,993	trehalose catabolic process	3.32E-03
down	KIRC	GO:0,046,597	negative regulation of viral entry into host cell	3.49E-18
down	KIRP	GO:0,007,144	female meiosis I	8.51E-03
up	KIRP	GO:0,046,878	positive regulation of saliva secretion	1.03E-02
down	LIHC	GO:0,045,008	Depyrimidination	2.81E-13
up	LIHC	GO:0,021,897	forebrain astrocyte development	1.48E-05
up	LUAD	GO:0,035,814	negative regulation of renal sodium excretion	3.03E-03
down	LUSC	GO:0,019,358	nicotinate nucleotide salvage	1.34E-03
up	LUSC	GO:0,045,006	DNA deamination	1.65E-02
down	PAAD	GO:0,070,585	protein localization to mitochondrion	1.63E-08
up	PAAD	GO:0,046,329	negative regulation of JNK cascade	1.04E-05
down	PCPG	GO:0,021,723	medullary reticular formation development	3.25E-05
up	PCPG	GO:0,006,335	DNA replication-dependent nucleosome assembly	1.72E-04
down	PRAD	GO:0,001,957	intramembranous ossification	5.15E-03
up	PRAD	GO:0,045,053	protein retention in Golgi apparatus	6.23E-09
down	READ	GO:0,060,123	regulation of growth hormone secretion	6.57E-05
up	READ	GO:0,042,742	defense response to bacterium	3.49E-04
down	SARC	GO:0,010,324	membrane invagination	7.69E-05
up	SKCM	GO:0,002,931	response to ischemia	8.74E-05
down	STAD	GO:0,080,184	response to phenylpropanoid	1.85E-04
down	THCA	GO:0,007,161	calcium-independent cell-matrix adhesion	8.69E-05
down	UCEC	GO:2,000,978	negative regulation of forebrain neuron differentiation	1.50E-03
up	UCEC	GO:0,070,208	protein heterotrimerization	1.10E-07

functions of THYM, STAD, and HNSC were less than five. There also a difference between the number of functions of up-methylated and down-methylated lncRNAs for each cancer. The specifically down-methylated lncRNAs of COAD were enriched in the majority of functions (261), whereas up-methylated lncRNAs of COAD were enriched in only a few functions (14), and LIHC demonstrated a similar pattern. The specifically up-methylated lncRNAs of ESCA were found to be enriched in a variety of functions (194), but the down-methylated lncRNAs were enriched in no functions. The details of enriched functions are shown in **Supplementary Table S5**.

Among the functions, we selected the most significant function for each group and displayed them on the left side of **Figure 4B**. The function names, enrichment *p*-values and other information are shown in **Table 3**. The lncRNAs of most cancer types were enriched in the "regulation" or "response" functions. Both the up-methylated group of CHOL and the down-methylated group of LIHC were enriched in "depyrimidine". The specifically up-methylated lncRNAs of GBM were enriched in "methylation", the specifically down-methylated lncRNAs of COAD were enriched in "dimethylation", and the specifically up-methylated lncRNAs of BRCA were enriched in "epigenetic". These results established the important role of lncRNAs in the epigenetic process.

In this study, we hypothesized that up-methylated lncRNAs would exhibit decreased expression and vice versa. That is, there is a negative correlation between changes in methylation and expression levels. In order to screen out negative correlated lncRNAs in DMlncs, we used lncRNA expression data to identify differentially expressed lncRNAs in various cancers.

LncRNA expression data of 13 cancer types were downloaded, and the number of differentially expressed lncRNAs in each cancer is shown in **Supplementary Table S6**. We identified 2,887 over-expressed lncRNAs and 2,375 low-expressed lncRNAs in different cancers. The total number of these lncRNAs was 4,155, and several genes were overlapped between two groups and displayed conflicting regulation patterns across different cancers. The numerical distribution of differentially expressed lncRNAs in various cancers is shown in **Figure 5A**. Although the number of over-expressed and low-expressed lncRNAs was similar in the majority of cancer types, there were significantly more low-expressed lncRNAs in BRCA and THCA, and significantly more over-expressed lncRNAs in LIHC and STAD.

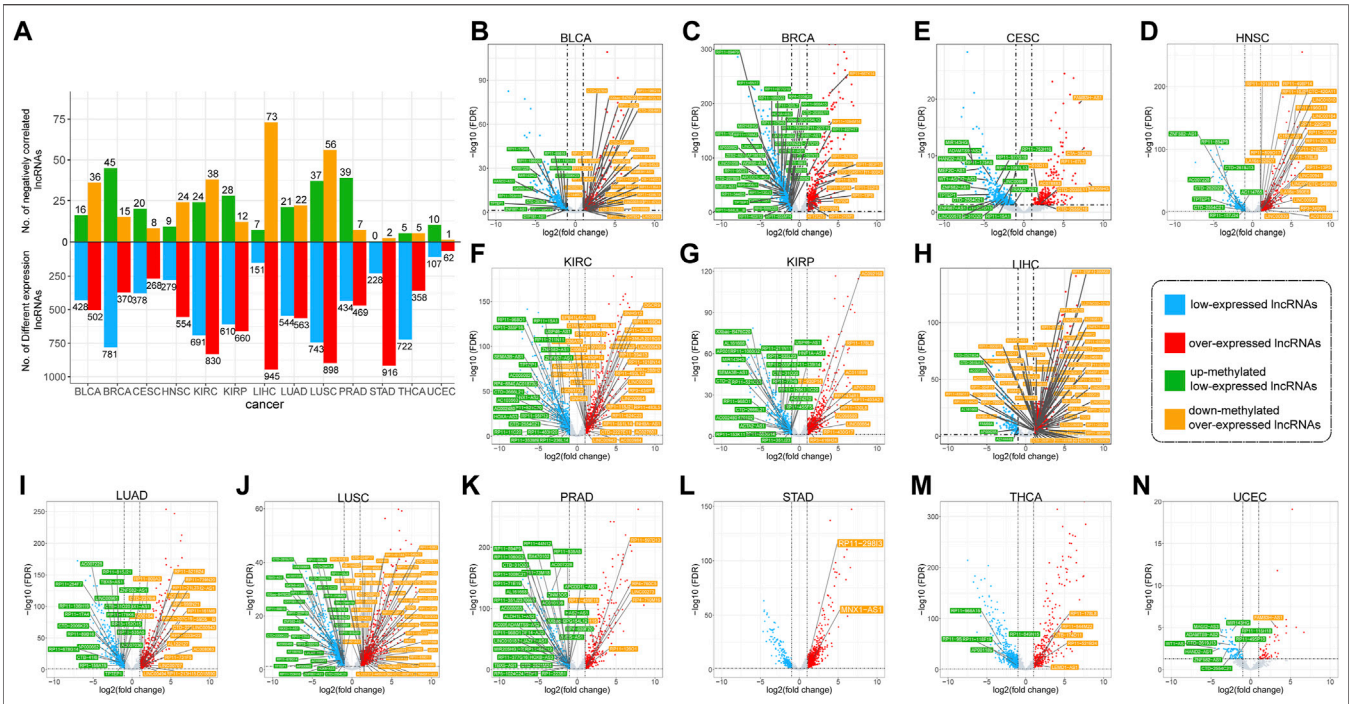


FIGURE 5 | The negatively correlated lncRNAs for cancers **(A)** Quantitative analysis of lncRNAs that are differently expressed and negatively correlated **(B–N)** Distribution and names of negatively correlated lncRNAs for cancers. Red and blue dots represent differentially expressed lncRNAs. In which, blue represents down-expressed lncRNAs, and red represents up-expressed lncRNAs. Green and yellow boxes represent the names of negatively correlated lncRNAs. Where green indicates lncRNAs with up-methylated and low-expressed, and yellow represents lncRNAs with down-methylated and over-expressed.

TABLE 4 | Cancer specific negatively correlated lncRNAs.

Cancer	Negatively correlated lncRNAs
BLCA	*XXbac-B476C20, RP5-943J3, RP11-390P2, AC006116, AC073046, RP11-514P8*
BRCA	*RP11-667K14, RP11-1094M14, RP11-497H17, LINC00619*
CESC	*MEF2C-AS1, CTA-384D8, CTD-2035E11*
HNSC	*LA16c-390E6, CTC-548K16, LA16c-325D7*
KIRC	*RP11-488L18, AC027601, RP1-118J21, HLA-F-AS1, SNHG12, EPB41L4A-AS1*
KIRP	*ACTA2-AS1, RP11-77H9, RP11-126K1*
LIHC	*AC007879, AC025335, CTC-246B18, HULC, LINC00665, RP11-215P8, RP11-890B15, RP11-968A15, RP11-973H7, RP3-395M20 and TEX41*
LUSC	*Z83851, RP11-311F12, RP11-757G1, RP11-12L8*
PRAD	*MIR205HG, LINC01018, JAZF1-AS1, RP4-639F20, RP1-223B1, RP11-597D13, LINC00115*
STAD	*MNX1-AS1, RP11-298I3*

Subsequently, negatively correlated lncRNAs (NClncs) were screened from differentially expressed lncRNAs. The number of NClncs in each cancer is shown in **Figure 5A**. NClncs were divided into two types: up-methylated-low-expressed lncRNAs (UMLElncs) and down-methylated-over-expressed lncRNAs (DMOElncs) based on the changes in expression and methylation levels. The two types of lncRNAs in each cancer are shown in **Figure 5B–N**. As seen from the figure, the number of NClncs was proportional to the number of differentially expressed lncRNAs. However, the proportion of differential lncRNAs was not balanced in several cancers. For example, STAD and THCA had many differentially expressed lncRNAs, but a small number of DMlncs, implying a small number of NClncs. The proportion of UMLElncs and DMOElncs was different in each cancer. For example, both types of lncRNAs for LUSC were abundant. HNSC and LIHC had significantly more DMOElncs, while BRCA and PRAD had significantly more UMLElncs.

The cancer-specific negatively correlated lncRNAs (CSNClncs) were then identified, and the results are shown in **Table 4**. Among the cancer-specific DMlncs, 49 NClncs were identified, with LIHC having the highest CSNClncs (11), followed by PRAD (7), BLCA (6) and KIRC (6). LUAD, THCA, and UCEC did not contain any CSNClnc.

TABLE 5 | Significantly survival associated lncRNAs in negatively correlated lncRNAs for each cancer.

Cancer	Count	lncRNAs
BLCA	5	*RP11-390P2, XXbac-B476C20, RP5-943J3, AC073046, AC006116*
BRCA	3	*LINC00619, RP11-497H17, RP11-667K14*
CESC	2	*MEF2C-AS1, CTD-2035E11*
HNSC	3	*LA16c-390E6, LA16c-325D7, CTC-548K16*
KIRC	5	*SNHG12, EPB41L4A-AS1, RP11-488L18, AC027601, RP1-118J21*
KIRP	2	*ACTA2-AS1, RP11-77H9*
LIHC	4	*RP11-215P8, RP11-968A15, RP11-973H7, CTC-246B18*
LUSC	2	*RP11-757G1, RP11-311F12*
PRAD	2	*RP1-223B1, LINC01018*
STAD	1	*MNX1-AS1*

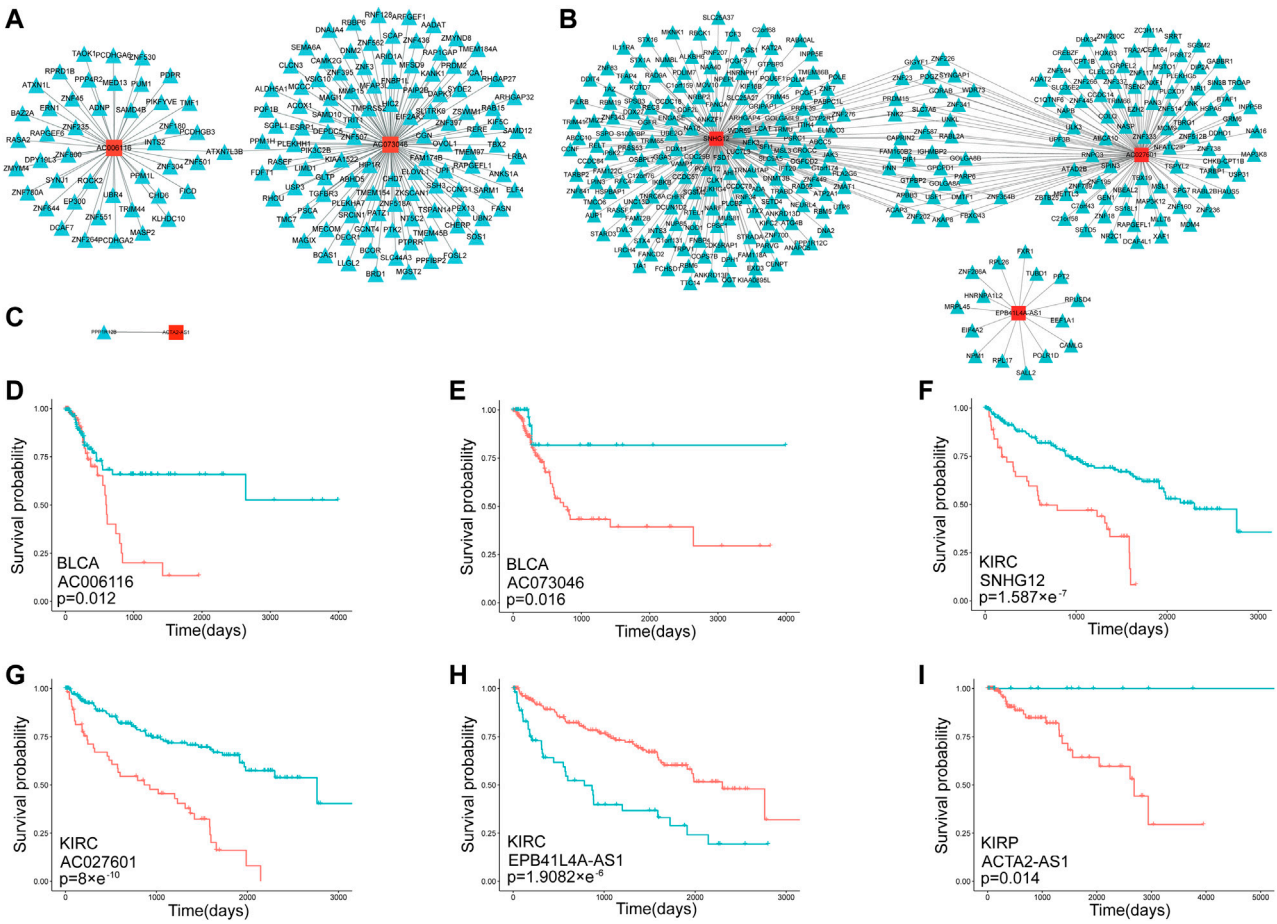


FIGURE 6 | The negatively correlated lncRNAs which specific to each cancer in the ceRNA network **(A)**The ceRNAs of BLCA-specific negatively correlated lncRNAs **(B)** The ceRNAs of KIRC-specific negatively correlated lncRNAs **(C)** The ceRNAs of KIRP-specific negatively correlated lncRNAs **(D–I)** Kaplan-Meier curves for the lncRNAs associated with each cancer types. The red line represents the group with a high level of expression, while the blue line represents the group with a low level of expression. Additionally, the "+" on the lines represents patients who were lost to follow-up. At this point, the number of patients decreases but the overall survival rate remains stable.

Additionally, survival-correlated lncRNAs were identified by analyzing the CSNClncs of each cancer. As shown in **Table 5**, 29 lncRNAs were identified to associated with survival in ten cancers. These lncRNAs may be used as specifically diagnostic markers for corresponding cancers. They not only exhibited synergistic alterations in expression and methylation, but were also closely associated with the survival of cancer patients.

The main function of lncRNAs was to combine miRNAs competing with mRNAs, thereby increasing mRNA expression. As the methylation of lncRNAs increased, their expression decreased, which indirectly led to the decrease in mRNA expression, and vice versa. Therefore, we screened the ceRNAs of DMlncs. First, the lncRNA-miRNA and mRNA-miRNA regulatory relationships were integrated. A lncRNA and a mRNA sharing more than two miRNAs were considered to have a ceRNA relationship. The correlation between lncRNA and mRNA expression was calculated for each cancer to confirm the ceRNA relationship. We maintained the ceRNA relationships that showed a positive correlation between lncRNA and mRNA expression. Each survival-related CSNClnc was put into the ceRNA network to search for associated mRNAs. The visualized results of ceRNAs for BLCA, KIRC and KIRP are shown in **Figures 6A–C**.

BLCA had two lncRNAs mapped into the ceRNA network (*AC006116* and *AC073046*). Both lncRNAs formed a ceRNA relationship with the mRNA *ATXN7L3B* and formed ceRNA relationships with some other mRNAs, respectively. *ATXN7L3B* have been confirmed to be associated with an increased risk of colorectal cancer (Leberfarb et al., 2020), and cytoplasmic *ATXN7L3B* interfered with the nuclear functions of the SAGA deubiquitinase module (Li et al., 2016). The SAGA complex was composed of two enzymatic modules, which house histone acetyltransferase (HAT) and deubiquitinase (DUB) activities. The DUB module was important for normal embryonic development (Glinsky, 2006; Lin et al., 2012), and alterations in the expression or structure of component proteins were linked to cancer (Lan et al., 2015). Therefore, *AC006116* and *AC073046*, as its ceRNAs, could regulate the expression of *ATXN7L3B* and were also closely related to cancer. The survival correlation of *AC006116* and *AC073046* in BLCA is shown in **Figures 6D,E**. Both lncRNAs were significantly associated with the survival of BLCA patients.

Three lncRNAs in KIRC were mapped into the ceRNA network, among which *SNHG12* and *AC027601* shared more than 30 mRNAs, while *EPB41L4A-AS1* did not share any mRNAs with the other two lncRNAs. The survival analysis results for the three lncRNAs in KIRC are shown in **Figures 6F–H**. The high expression of *SNHG12* and *AC027601* both showed a worse prognosis. However, the low expression group of *EPB41L4A-AS1* showed a worse prognosis. Therefore, we inferred that in KIRC, lncRNAs *SNHG12* and *AC027601* had a carcinogenic effect, whereas *EPB41L4A-AS1* had a tumor-suppressive effect, which explains why it shared no mRNAs with the other two lncRNAs. Numerous studies have established that *SNHG12* is associated with cancers (Zhang et al., 2020), and could be used as a potential therapeutic target and biomarker for human cancers (Tamang et al., 2019). DNA-methylation-mediated activation of *SNHG12* promoted temozolomide resistance in glioblastoma (Lu et al., 2020). *SNHG12* promoted tumor progression and sunitinib resistance by upregulating *CDCA3* in renal cell carcinoma (Liu et al., 2020). *EPB41L4A-AS1* was a repressor of the Warburg effect and played an important role in the metabolic reprogramming of cancer (Liao et al., 2019), and *EPB41L4A-AS1* has been identified as a potential biomarker in non-small cell lung cancer (Wang et al., 2020). At present, *AC027601* has been identified as a survival signature in renal clear cell carcinoma (Qi-Dong et al., 2020), but has not been reported in other cancers. Given that the three lncRNAs were simultaneously identified as KIRC-related lncRNAs in this study, we believed that *AC027601* should be closely associated with the occurrence and development of KIRC, and it is a newly identified cancer-related lncRNA.

Only *ACTA2-AS1* was mapped into the ceRNA network in KIRP, where it established a ceRNA relationship with *PPP1R12B*. *PPP1R12B* has been shown to inhibit tumor growth and metastasis by regulating Grb2/PI3K/Akt signaling in colorectal cancer (Ding et al., 2019). *ACTA2-AS1* plays an important role in a variety of cancers, for example, *ACTA2-AS1* is significantly associated with overall survival in ovarian cancer patients (Li and Zhan, 2019). *ACTA2-AS1* plays different roles in different cancers. *ACTA2-AS1* knockdown promotes liver cancer cell proliferation, migration and invasion (Zhou and Lv, 2019), while *ACTA2-AS1* suppresses lung adenocarcinoma progression (Ying et al., 2020), implying an inhibitory effect on the two cancers. However, *ACTA2-AS1* promotes cervical cancer progression (Luo et al., 2020), suggesting its carcinogenic role in cancer. **Figure 6I** shows the survival analysis result for *ACTA2-AS1* in KIRP. We believed that *ACTA2-AS1* had a carcinogenic effect in KIRP.

Common lncRNA Biomarkers in Cancers

All cancer types exhibited infinite proliferation, transformation and ease of metastasis. Therefore, we sought to identify DMlncs common to different cancers to help understand the mechanisms underlying the occurrence of common features in cancers. First, the intersection of DMlncs were searched in cancers, and the results are shown in **Figure 7**. The upper triangle and lower triangle reflected the intersection of up-methylated lncRNAs and the intersection of down-methylated lncRNAs, respectively. The findings were consistent with the hypothesis that the larger the lncRNA set, the greater the overlap with other cancers. The intersections of DMlncs of SARC, STAD, and THYM with other cancers were small. The down-methylated lncRNAs of GBM and SKCM had small intersections with other cancers, whereas the up-methylated lncRNAs of THCA had small intersections with other cancers. There were amount of up-methylated lncRNAs in PCPG (62), but the overlaps with other cancers were small. Among the down-methylated lncRNAs, the intersection of KIRC and KIRP was the largest of KIRP, but only ranked 10th of the KIRC. Among the up-methylated lncRNAs, the intersection of KIRC and KIRP was the largest of KIRC, while was the second largest of KIRP.

Subsequently, we extracted DMlncs which were common in various cancers. The findings indicated that there were 19 common DMlncs in more than 15 cancers (*RP4-792G4, RP5-855F14, OTX2-AS1, RP11-52L5, CYP1B1-AS1, RP11-175E9, RP11-552E20, HCCAT3, RP11-718O11, HOXA-AS2, RP3-326L13, AC007228, RP11-297B11, CTC-523E23, LINC01010, RP11-227D2, EVX1-AS, AC018730,* and *RP11-465L10*). Fourteen of them were up-methylated in all the cancers, indicating their carcinogenic potential, whereas three lncRNAs were down-methylated in the majority of cancers and may act as

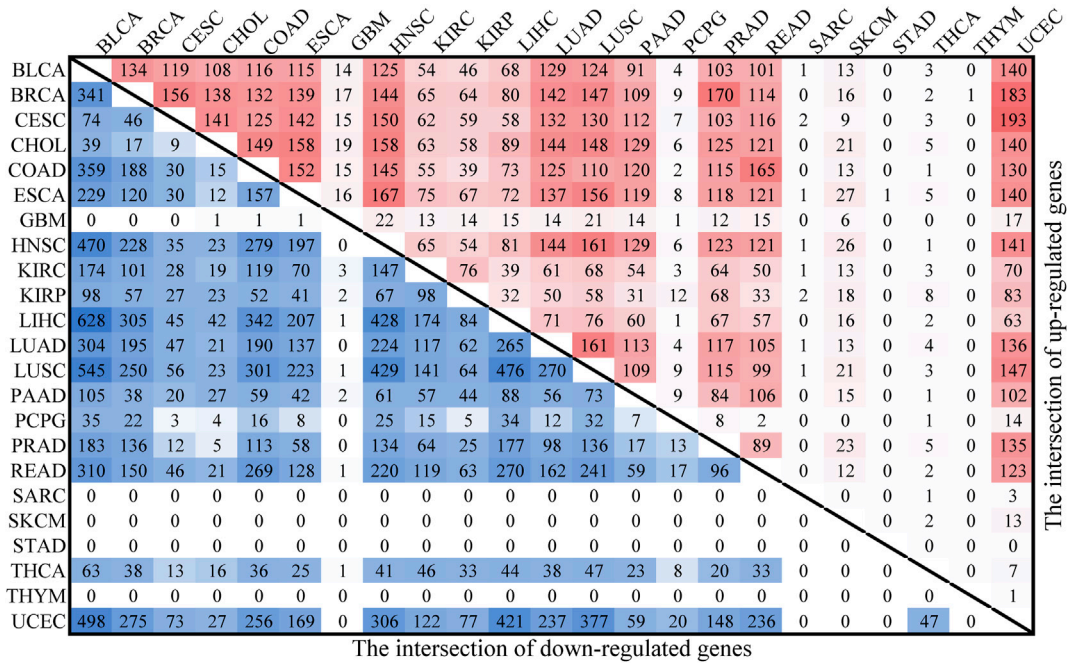


FIGURE 7 | The intersection of DMlncs for cancers. The upper red triangle represents the intersection of up-methylated lncRNAs, while the lower blue triangle represents the intersection of down-methylated lncRNAs.

potential tumor suppressors (**Figure 8A**). **Figure 8B** shows the comparison of methylation levels of the 19 lncRNAs in tumor and normal samples. It is intuitive to conclude that there were significant differences in lncRNAs methylation levels between tumor and normal samples. Except for *LINC01010*, *RP11-552E20,* and *RP5-855F14*, the lncRNAs had a higher methylation level in tumor samples.

Among the 19 lncRNAs, O*XT2-AS1* was shown to be significantly down-methylated in lung squamous cell carcinoma and was closely associated with poor prognosis of cancer (Zheng et al., 2021). *CYP1B1-AS1* has been confirmed to play an important role in triple-negative breast cancer, lung adenocarcinoma and acute myeloid leukemia, and was associated with the prognosis of these cancers (Cheng et al., 2021; Ren et al., 2021; Vishnubalaji and Alajez, 2021). Abnormal methylation and low expression of *CTC-523E23* led to poor prognosis in patients with lung squamous cell carcinoma (Rui Li et al., 2020). Inhibition of *LINC01010* may promote the migration and invasion of lung cancer cells (Cao et al., 2020) and help in the prediction of neuroblastoma prognosis (Gao et al., 2020). *EVX1-AS* is closely associated with the prognosis of colon cancer and has been predicted to be potentially associated with the development of multiple cancers by LncRNADisease V2.0 (Bao et al., 2019; Gao et al., 2021). The findings revealed that the abnormalities of lncRNAs played an important role in the occurrence and development of cancer, and the unconfirmed lncRNAs could serve as entry points for future research. Additionally, numerous lncRNAs were identified as abnormal in lung cancer, indicating that lung cancer may be influenced by a variety of pathogenic factors.

Then, we investigated the functions of the 19 lncRNAs and performed functional enrichment analysis on these lncRNAs using the GREAT software. **Figure 8C** shows that these lncRNAs were enriched in processes required for organisms such as metabolism and biosynthesis, as well as those closely associated with the occurrence and development of cancer, such as gene expression and transcription.

NClncs in common lncRNAs were investigated in combination with the results of differentially expressed lncRNAs. Seven of the 19 lncRNAs were found to be differentially expressed, and four (*CYP1B1-AS1* (BLCA, BRCA, and LUSC), *AC007228* (BLCA, HNSC, LIHC, LUAD, LUSC and PRAD), *HOXA-AS2* (BRCA) and *LINC01010* (HNSC)) of them had a negative correlation in multiple cancers. Additionally, we analyzed the four NClncs shared by cancers for their correlation with survival, and survival-related lncRNAs were identified. The four lncRNAs were associated with patients' survival in a variety of cancers, including *CYP1B1-AS1* in eight types of cancers (BRCA, HNSC, KIRC, KIRP, LIHC, PUAD, PRAD and STAD), *AC007228* in seven types of cancers (BLCA, HNSC, KIRC, KIRP, LIHC, LUAD and THCA), *LINC01010* in five types of cancers (BRCA, LIHC, LUAD, LUSC and THCA) and *HOXA-AS2* in four types of cancers (BRCA, KIRC, KIRP and THCA) (**Supplementary Figures S1–S4**).

To further validate the cancer-common lncRNAs identified in this study, significantly survival-related lncRNAs were mapped to the ceRNA network, and only mRNAs shared by more than four types of cancers were selected for further study. Finally, a subnetwork comprising 33 mRNAs and three lncRNAs

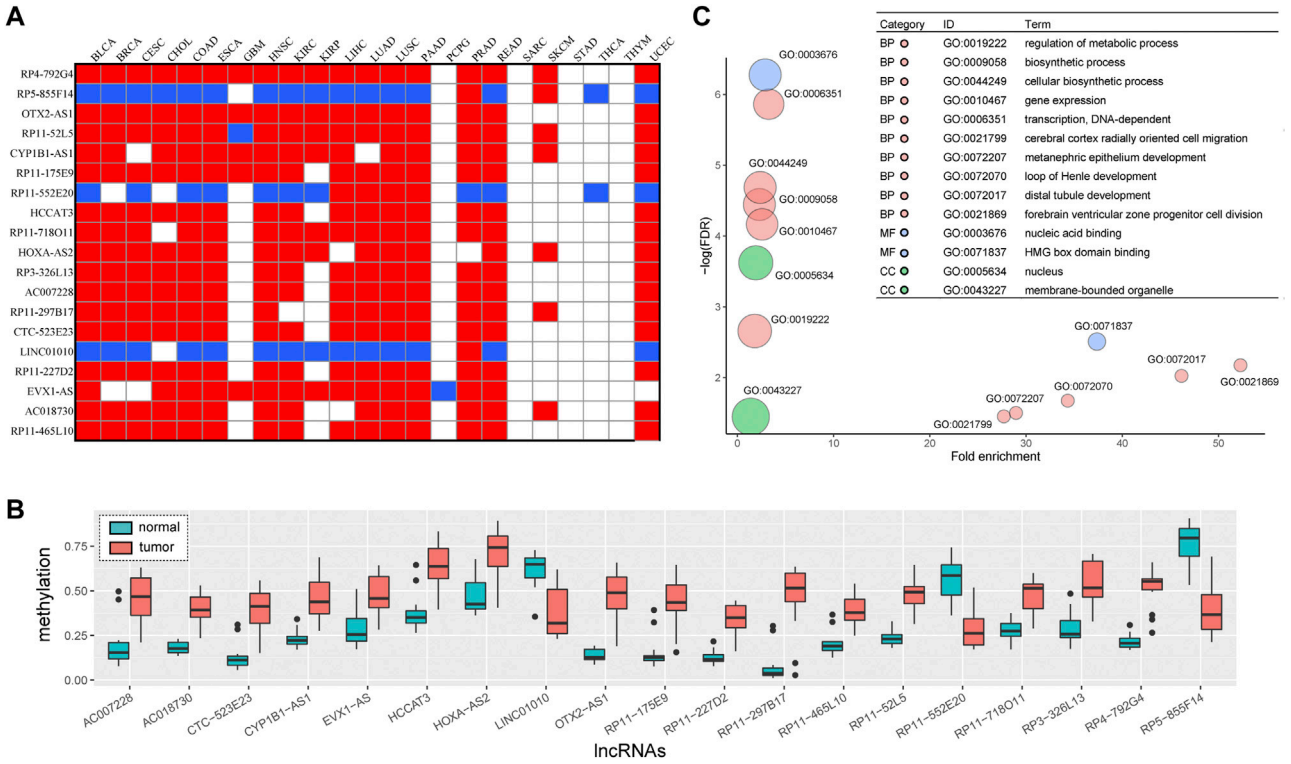


FIGURE 8 | Common DMlncs of cancers **(A)** Methylation changes of common DMlncs in cancers. Red indicates up-methylation of the lncRNA, while blue indicates down-methylation in corresponding cancer **(B)** The functions of common DMlncs in cancer. Red represents a biological process (BP), blue represents a molecular function (MF), green represents cell component (CC), and the size of the bubble represents the amount of lncRNA enriched **(C)** A comparison of the methylation states of common DMlncs. Red represents tumor samples, while blue represents normal samples.

(*AC007228, CYP1B1-AS1,* and *HOXA-AS2*) was identified in ten cancers. The ceRNA distribution of the 33 mRNAs in cancers is shown in **Figure 9A**. Although these mRNAs were shared by multiple cancers, they generally formed ceRNA relationships with a greater number of lncRNAs and had higher correlation coefficients in KIRP. *HOXA3* had ceRNA lncRNAs in ten cancers and showed high correlation coefficients in BRCA, CESC, HNSC and LUSC. **Figure 9B** shows the ceRNA subnetwork, in which *HOXA-AS2* and *AC007228* were shared by ten cancers, and the two lncRNAs shared some ceRNA relationships with some mRNAs (*DMTF1, HOXB3, NOD1, RABL2A, TRIOBP, ZNF443,* and *ZNF789*) but formed ceRNA relationships with some other mRNAs, respectively (*AC007228: ZNF10, ZNF211, ZNF229, ZNF471, ZNF583, ZNF614, ZNF649, ZNF763, ZNF793,* and *ZNF879*; *HOXA-AS2*: *DHRS3* and *HOXA3*).

The genomic locations of *HOXA-AS2* and *AC007228* were checked. As shown in **Figure 10A**, the genomic locations of lncRNAs belonging to the *AC007228* family and mRNAs belonging to the *ZNF* family were extremely similar. As shown in **Figure 10B**, *HOXA-AS2* (Zhao et al., 2013) and its ceRNA *HOXA3* were both located in the same genomic region. Their sequences were similar, and the ceRNA relationships were generated by the lncRNAs' cis-regulatory interactions with mRNAs.

The major histocompatibility complex (MHC) region was one of the regions with the highest gene density and polymorphism. High-throughput sequencing and other technologies confirmed the role of MHC in disease and showed that MHC was associated with cancer and neurological diseases in addition to infection and autoimmune diseases. MHC is involved in antigen recognition during the immune response and is capable of inducing immune cells to participate in immune response (Trowsdale and Knight, 2013). Studies have also reported that immune cytolytic activity (CYT) is positively correlated with the presence of inhibitory receptors (*PDCD1, PDL1, CTLA4, LAG3, TIM3,* and *IDO1*), and the presence of CYT is more responsive to immune checkpoint inhibition, suggesting that it can be used as a key marker for immune checkpoint therapy (Narayanan et al., 2018; Wang et al., 2019). Immune control of tumor lesions requires local antigen recognition, activation and amplification of tumor-specific cytotoxic T lymphocytes (CTL). The activated CTL infiltrate the tumor microenvironment and scan the tumor tissue, where they directly interact with the target cells, inducing tumor cell apoptosis and atrophy (Basu et al., 2016). First, effector T lymphocytes are required to migrate to the tumor foci, a process referred to as immune cell infiltration. Following that, they have to make physical contact with the tumor

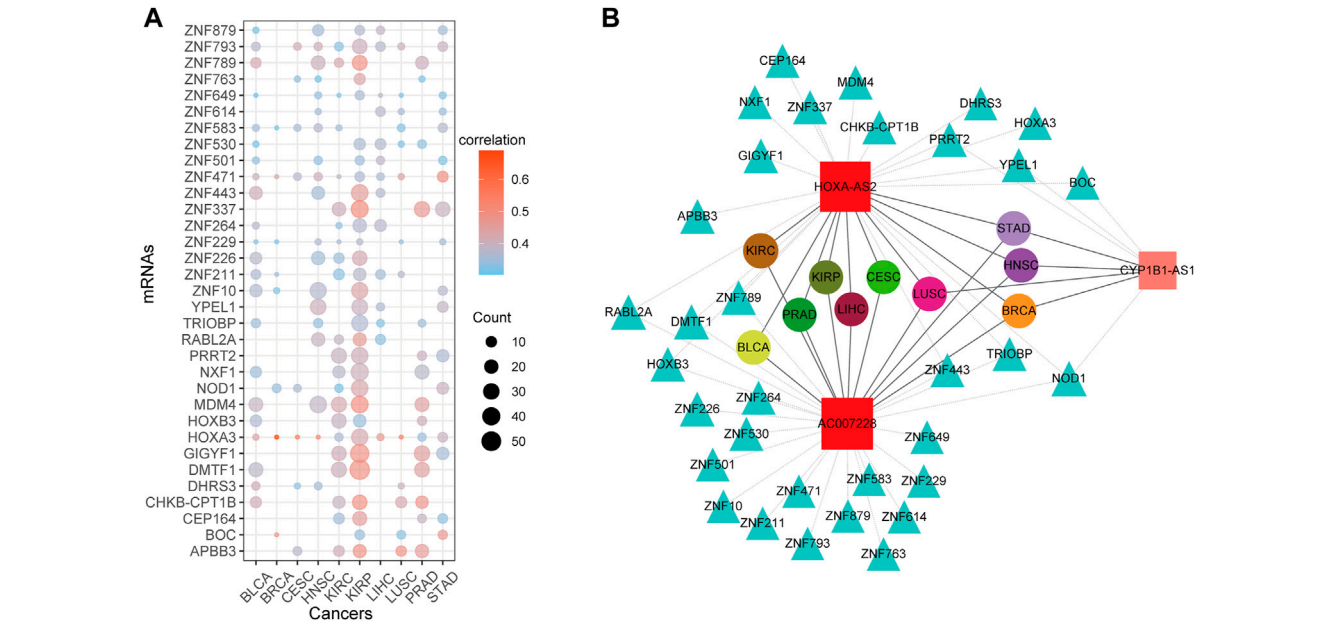


FIGURE 9 | The ceRNA network of common DMlncs of cancers **(A)** The ceRNA distribution of mRNAs in each cancer type. The color of the dots represents the mean of the correlation coefficient for the lncRNAs that form ceRNA relationships in corresponding cancer. The size of the dots represents the number of lncRNAs that have a ceRNA relationship with the corresponding mRNA in cancer **(B)** The ceRNA network. Square nodes represent lncRNAs, triangle nodes represent mRNAs, and round nodes represent cancer types. The size of lncRNA nodes is proportional to the number of cancer types sharing it, and nodes of different cancer types are distinguished by different colors.

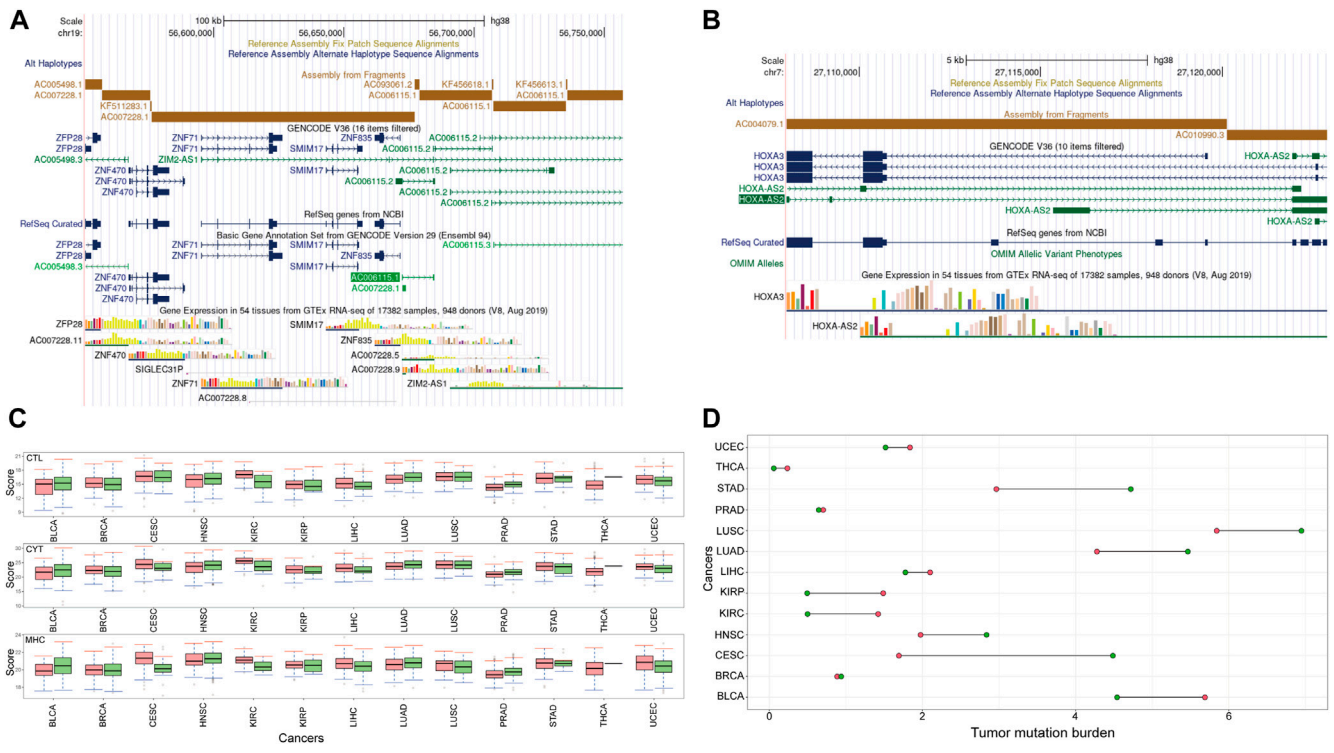


FIGURE 10 | The characteristics of *HOXA-AS2* and *AC007228* **(A)** The genomic location of *AC007228* **(B)** The genomic location of *HOXA-AS2* **(C)** Immune response scores of *HOXA-AS2* and *AC007228* **(D)** TMB of *HOXA-AS2* and *AC007228*.

cells and scan their MHC. Finally, by releasing perforin or fas/fasl to bind target cells, CTL activates and induces apoptosis (Weigelin et al., 2011). Therefore, we assess the immunological effects of *HOXA-AS2* and *AC007228* using MHC, CYT, and CTL scores.

To improve the evaluation of the effect of lncRNAs on immunity, the R package "ConsensusClusterPlus" (Wilkerson and Hayes, 2010) was used to cluster samples of each cancer based on their expression profiles of *HOXA-AS2* and *AC007228*. We varied the parameter k from two to six, and then selected the optimal subtypes for subsequent immune score evaluation. The score comparison shown in **Figure 10C** demonstrates that all three scores were consistent across cancer subtypes, indicating that the identified key lncRNAs may aid in predicting the immunological status of cancer patients and provide a basis for tumor treatment. Subsequently, the Wilcoxon rank-sum test was used to compare the immunity scores of different subtypes, and significant differences were observed in the immunity scores of different subtypes in KIRC, LIHC, LUAD, and PRAD.

Finally, we assessed the tumor mutation burden (TMB) among subtypes. TMB was a novel biological target for which therapeutic impact may be predicted. Previous research has demonstrated that the more somatic mutations a cancer patient possesses, the more likely it is that new antigens are produced. Antigen peptides could be loaded onto the MHC and displayed on the cell surface, aiding in their recognition by T cells (Jiang et al., 2018). Therefore, cancer patients with high TMB levels responded better to immune checkpoint blockade therapy. As shown in **Figure 10D**, the TMB of KIRC, KIRP, LIHC, LUAD, and UCEC corresponded to the immunity score. Subtypes of key lncRNAs played an important role in the immune effect in KIRC, LIHC, and LUAD.

DISCUSSION

LncRNAs have been implicated in the occurrence and development of cancer. This study aimed to identify the specific and common lncRNAs with aberrant methylation in pan-cancer. After searching for DMlncs in a variety of cancers, the pan-cancer results were compared. Subsequently, data on lncRNA expression, lncRNA methylation and mRNA expression was integrated to identify lncRNAs with a negative correlation between methylation and expression changes, and survival analysis was performed to further verify the results. Following that, survival-related lncRNAs were mapped to the ceRNA network, and pan-cancer biomarkers were identified by examining the connection characteristics of the network. Finally, the immune effect of the lncRNAs was verified.

In this study, DMlncs for 23 cancers were acquired, and cancer-specific lncRNAs and cancer-common lncRNAs were identified. The NClncs were further screened for ten cancers, and the correlation between the lncRNAs and mRNAs, as well as the association with survival were verified. Cancer-specific lncRNAs may be used as diagnostic biomarkers for corresponding cancers. In clinical application, these lncRNAs could apply to make detection kits of corresponding cancers. Common lncRNAs in pan-cancers could be used to understand the mechanism underlying common features in cancers. These lncRNAs can be used for the development of targeting drugs for the remission of general symptoms and treatment of cancer.

This study yielded significant results. Not only did we validate several previously known cancer-associated lncRNAs, but we also identified new cancer-related lncRNAs. *AC027601* was identified as a novel KIRC-associated lncRNA, and *ACTA2-AS1* was discovered to be carcinogenic in KIRP. Additionally, two lncRNAs, *HOXA-AS2,* and *AC007228* were identified as pan-cancer lncRNAs.

However, there are some limitations to this study. Because the number of DMlncs for SARC, STAD, and THYM was less than ten, the more systematic comparisons for these cancers were impossible. The number of normal samples with methylation data for these three cancers was less than ten. The sample proportion was skewed when differences were calculated, resulting in less statistically significant results. Additionally, the DNA methylation data only included 450k arrays and did not cover the entire genome, which might have contributed to the study's insufficiency outcomes. We only identified the epigenetically dysregulated lncRNAs based on DNA methylation, although N6-methyladenosine (m6A) as the RNA post-transcriptional modification has been shown to influence the function of RNAs as well. We did not evaluate the relationship between m6A and lncRNA due to the lack of data.

In the future, more complete data sets on cancers may be collected to allow for more rigorous comparisons. We may use copy number data to analyze the change in copy number of lncRNA and conduct a more comprehensive investigation of the change and function of lncRNA in cancer. Additionally, other types of omics data may be integrated, and factors affecting lncRNAs and gene expression could be evaluated more comprehensively bringing the research process closer to the way molecules interact in the human body. With the development of sequencing techniques, additional methylation data sets such as HM850K, whole-genome bisulfite sequencing (WGBS), and reduced representation bisulfite sequencing (RRBS), as well as other types of methylation data sets, will be used to analyze the function of DNA methylation for lncRNAs.

In general, this study screened cancer-related lncRNA biomarkers based on their methylation alterations and their competing mRNAs. This study considered multiple omics data more comprehensively and used more stringent screening criteria, which effectively eliminated of data deviation errors. The lncRNA biomarkers identified in this study may aid in the investigation of cancer mechanisms.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The datasets ANALYZED for this study can be found in the TCGA (https://portal.gdc.cancer.gov/) and TANRIC (https://www.tanric.org/).

AUTHOR CONTRIBUTIONS

NZ collected data, carried out the initial analyses and drafted the manuscript. MG conceived of the study, and participated in its design and coordination and helped to draft the manuscript. CW and KW coordinated and supervised data collection, and critically commented on the important intellectual content of the manuscript. CZ participated in the design of the study and performed the statistical analysis. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant No. 62031003), and the High Level Innovation Team Construction Project of Beijing Municipal Universities (Grant No. IDHT20190506).

ACKNOWLEDGMENTS

We are grateful to all patients for their dedicated participation in the current study. We thank the TCGA database for sharing the multi-omics data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2022.882698/full#supplementary-material

REFERENCES

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), 1363–1369. doi:10.1093/bioinformatics/btu049

Bao, Z., Yang, Z., Huang, Z., Zhou, Y., Cui, Q., and Dong, D. (2019). LncRNADisease 2.0: an Updated Database of Long Non-coding RNA-Associated Diseases. *Nucleic Acids Res.* 47 (D1), D1034–D1037. doi:10.1093/nar/gky905

Bao, S., Zhao, H., Yuan, J., Fan, D., Zhang, Z., Su, J., et al. (2020). Computational Identification of Mutator-Derived Lncrna Signatures of Genome Instability For Improving the Clinical Outcome of Cancers: A Case Study in Breast Cancer. *Brief Bioinform* 21 (5), 1742–1755. doi:10.1093/bib/bbz118

Bao, G., Xu, R., Wang, X., Ji, J., Wang, L., Li, W., et al. (2021). Identification of lncRNA Signature Associated with Pan-Cancer Prognosis. *IEEE J. Biomed. Health Inf.* 25 (6), 2317–2328. doi:10.1109/JBHI.2020.3027680

Bartonicek, N., Maag, J. L. V., and Dinger, M. E. (2016). Long Noncoding RNAs in Cancer: Mechanisms of Action and Technological Advancements. *Mol. Cancer* 15 (1), 43. doi:10.1186/s12943-016-0530-6

Basu, R., Whitlock, B. M., Husson, J., Le Floc'h, A., Jin, W., Oyler-Yaniv, A., et al. (2016). Cytotoxic T Cells Use Mechanical Force to Potentiate Target Cell Killing. *Cell* 165 (1), 100–110. doi:10.1016/j.cell.2016.01.021

Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., et al. (2011). Integrative Annotation of Human Large Intergenic Noncoding RNAs Reveals Global Properties and Specific Subclasses. *Genes Dev.* 25 (18), 1915–1927. doi:10.1101/gad.17446611

Cao, Q., Dong, Z., Liu, S., An, G., Yan, B., and Lei, L. (2020). Construction of a Metastasis-Associated ceRNA Network Reveals a Prognostic Signature in Lung Cancer. *Cancer Cell Int.* 20, 208. doi:10.1186/s12935-020-01295-8

Chen, Y. G., Satpathy, A. T., and Chang, H. Y. (2017). Gene Regulation in the Immune System by Long Noncoding RNAs. *Nat. Immunol.* 18 (9), 962–972. doi:10.1038/ni.3771

Chen, J., Yu, Y., Li, H., Hu, Q., Chen, X., He, Y., et al. (2019). Long Non-coding RNA PVT1 Promotes Tumor Progression by Regulating the miR-143/HK2 axis in Gallbladder Cancer. *Mol. Cancer* 18 (1), 33. doi:10.1186/s12943-019-0947-9

Cheng, Y., Wang, X., Qi, P., Liu, C., Wang, S., Wan, Q., et al. (2021). Tumor Microenvironmental Competitive Endogenous RNA Network and Immune Cells Act as Robust Prognostic Predictor of Acute Myeloid Leukemia. *Front. Oncol.* 11, 584884. doi:10.3389/fonc.2021.584884

Ding, C., Tang, W., Wu, H., Fan, X., Luo, J., Feng, J., et al. (2019). The PEAK1-Ppp1r12b axis Inhibits Tumor Growth and Metastasis by Regulating Grb2/PI3K/Akt Signalling in Colorectal Cancer. *Cancer Lett.* 442, 383–395. doi:10.1016/j.canlet.2018.11.014

Dong, Y., Xiao, Y., Shi, Q., and Jiang, C. (2019). Dysregulated lncRNA-miRNA-mRNA Network Reveals Patient Survival-Associated Modules and RNA Binding Proteins in Invasive Breast Carcinoma. *Front. Genet.* 10, 1284. doi:10.3389/fgene.2019.01284

Esposito, R., Bosch, N., Lanzós, A., Polidori, T., Pulido-Quetglas, C., and Johnson, R. (2019). Hacking the Cancer Genome: Profiling Therapeutically Actionable Long Non-coding RNAs Using CRISPR-Cas9 Screening. *Cancer Cell* 35 (4), 545–557. doi:10.1016/j.ccell.2019.01.019

Gao, L., Lin, P., Chen, P., Gao, R. Z., Yang, H., He, Y., et al. (2020). A Novel Risk Signature that Combines 10 Long Noncoding RNAs to Predict Neuroblastoma Prognosis. *J. Cell Physiol.* 235 (4), 3823–3834. doi:10.1002/jcp.29277

Gao, M., Guo, Y., Xiao, Y., and Shang, X. (2021). Comprehensive Analyses of Correlation and Survival Reveal Informative lncRNA Prognostic Signatures in Colon Cancer. *World J. Surg. Onc.* 19 (1), 104. doi:10.1186/s12957-021-02196-4

Glinsky, G. V. (2006). Genomic Models of Metastatic Cancer: Functional Analysis of Death-From-Cancer Signature Genes Reveals Aneuploid, Anoikis-Resistant, Metastasis-Enabling Phenotype with Altered Cell Cycle Control and Activated PcG Protein Chromatin Silencing Pathway. *Cell Cycle* 5 (11), 1208–1216. doi:10.4161/cc.5.11.2796

Hou, P., Bao, S., Fan, D., Yan, C., Su, J., Qu, J., et al. (2021). Machine Learning-Based Integrative Analysis of Methylome and Transcriptome Identifies Novel Prognostic DNA Methylation Signature in Uveal Melanoma. *Brief Bioinform.* 22 (4). doi:10.1093/bib/bbaa371

Huarte, M. (2015). The Emerging Role of lncRNAs in Cancer. *Nat. Med.* 21 (11), 1253–1261. doi:10.1038/nm.3981

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T Cell Dysfunction and Exclusion Predict Cancer Immunotherapy Response. *Nat. Med.* 24 (10), 1550–1558. doi:10.1038/s41591-018-0136-1

Lan, X., Koutelou, E., Schibler, A. C., Chen, Y. C., Grant, P. A., and Dent, S. Y. R. (2015). Poly(Q) Expansions in ATXN7 Affect Solubility but Not Activity of the SAGA Deubiquitinating Module. *Mol. Cell Biol.* 35 (10), 1777–1787. doi:10.1128/MCB.01454-14

Lauss, M., Donia, M., Harbst, K., Andersen, R., Mitra, S., Rosengren, F., et al. (2017). Mutational and Putative Neoantigen Load Predict Clinical Benefit of Adoptive T Cell Therapy in Melanoma. *Nat. Commun.* 8 (1), 1738. doi:10.1038/s41467-017-01460-0

Leberfarb, E. Y., Degtyareva, A. O., BrusentsovII, Maximov, V. N., Voevoda, M. I., Autenshlus, A. I., et al. (2020). Potential Regulatory SNPs in the ATXN7L3B

and KRT15 Genes Are Associated with Gender-specific Colorectal Cancer Risk. *Pers. Med.* 17 (1), 43–54. doi:10.2217/pme-2019-0059

Li, N., and Zhan, X. (2019). Identification of Clinical Trait-Related lncRNA and mRNA Biomarkers with Weighted Gene Co-expression Network Analysis as Useful Tool for Personalized Medicine in Ovarian Cancer. *EPMA J.* 10 (3), 273–290. doi:10.1007/s13167-019-00175-0

Li, X.-Q., Guo, Y. Y., and De, W. (2012). DNA Methylation and microRNAs in Cancer. *Wjg* 18 (9), 882–888. doi:10.3748/wjg.v18.i9.882

Li, Y., Xu, J., Chen, H., Zhao, Z., Li, S., Bai , J., et al. (2013). Characterizing Genes With Distinct Methylation Patterns in the Context of Protein-Protein Interaction Network: Application to Human Brain Tissues. *PLoS One* 8 (6), e65871 . doi:10.1371/journal.pone.0065871

Li, J.-H., Liu, S., Zhou, H., Qu, L.-H., and Yang, J.-H. (2014). starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and Protein-RNA Interaction Networks from Large-Scale CLIP-Seq Data. *Nucl. Acids Res.* 42 (Database issue), D92–D97. doi:10.1093/nar/gkt1248

Li, J., Han, L., Roebuck, P., Diao, L., Liu, L., Yuan, Y., et al. (2015). TANRIC: An Interactive Open Platform to Explore the Function of lncRNAs in Cancer. *Cancer Res.* 75 (18), 3728–3737. doi:10.1158/0008-5472.CAN-15-0273

Li, W., Atanassov, B. S., Lan, X., Mohan, R. D., Swanson, S. K., Farria, A. T., et al. (2016). Cytoplasmic ATXN7L3B Interferes with Nuclear Functions of the SAGA Deubiquitinase Module. *Mol. Cell Biol.* 36 (22), 2855–2866. doi:10.1128/MCB.00193-16

Liao, M., Liao, W., Xu, N., Li, B., Liu, F., Zhang, S., et al. (2019). LncRNA EPB41L4A-AS1 Regulates Glycolysis and Glutaminolysis by Mediating Nucleolar Translocation of HDAC2. *EBioMedicine* 41, 200–213. doi:10.1016/j.ebiom.2019.01.035

Lin, C., and Yang, L. (2018). Long Noncoding RNA in Cancer: Wiring Signaling Circuitry. *Trends Cell Biol.* 28 (4), 287–301. doi:10.1016/j.tcb.2017.11.008

Lin, Z., Yang, H., Kong, Q., Li, J., Lee, S.-M., Gao, B., et al. (2012). USP22 Antagonizes P53 Transcriptional Activation by Deubiquitinating Sirt1 to Suppress Cell Apoptosis and Is Required for Mouse Embryonic Development. *Mol. Cell* 46 (4), 484–494. doi:10.1016/j.molcel.2012.03.024

Liu, Y., Cheng, G., Huang, Z., Bao, L., Liu, J., Wang, C., et al. (2020). Long Noncoding RNA SNHG12 Promotes Tumour Progression and Sunitinib Resistance by Upregulating CDCA3 in Renal Cell Carcinoma. *Cell Death Dis.* 11 (7), 515. doi:10.1038/s41419-020-2713-8

Lu, C., Wei, Y., Wang, X., Zhang, Z., Yin, J., Li, W., et al. (2020). DNA-methylation-mediated Activating of lncRNA SNHG12 Promotes Temozolomide Resistance in Glioblastoma. *Mol. Cancer* 19 (1), 28. doi:10.1186/s12943-020-1137-5

Luo, L., Wang, M., Li, X., Luo, C., Tan, S., Yin, S., et al. (2020). A Novel Mechanism by Which ACTA2-AS1 Promotes Cervical Cancer Progression: Acting as a ceRNA of miR-143-3p to Regulate SMAD3 Expression. *Cancer Cell Int.* 20, 372. doi:10.1186/s12935-020-01471-w

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 (6), 1140–1151. doi:10.1016/j.eururo.2013.12.003

Martín-Subero, J. I. (2011). How Epigenomics Brings Phenotype into Being. *Pediatr. Endocrinol. Rev.* 9 Suppl 1, 506–510.

Marwitz, S., Scheufele, S., Perner, S., Reck, M., Ammerpohl, O., and Goldmann, T. (2017). Epigenetic Modifications of the Immune-Checkpoint Genes CTLA4 and PDCD1 in Non-small Cell Lung Cancer Results in Increased Expression. *Clin. Epigenet.* 9, 51. doi:10.1186/s13148-017-0354-2

McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., et al. (2010). GREAT Improves Functional Interpretation of Cis-Regulatory Regions. *Nat. Biotechnol.* 28 (5), 495–501. doi:10.1038/nbt.1630

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and Quantifying Mammalian Transcriptomes by RNA-Seq. *Nat. Methods* 5 (7), 621–628. doi:10.1038/nmeth.1226

Narayanan, S., Kawaguchi, T., Yan, L., Peng, X., Qi, Q., and Takabe, K. (2018). Cytolytic Activity Score to Assess Anticancer Immunity in Colorectal Cancer. *Ann. Surg. Oncol.* 25 (8), 2323–2331. doi:10.1245/s10434-018-6506-6

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., et al. (2017). Intragenic DNA Methylation Prevents Spurious Transcription Initiation. *Nature* 543 (7643), 72–77. doi:10.1038/nature21373

Pauken, K. E., Sammons, M. A., Odorizzi, P. M., Manne, S., Godec, J., Khan, O., et al. (2016). Epigenetic Stability of Exhausted T Cells Limits Durability of Reinvigoration by PD-1 Blockade. *Science* 354 (6316), 1160–1165. doi:10.1126/science.aaf2807

PCAWG Transcriptome Core GroupCalabrese, C., Davidson, N. R., Demircioğlu, D., Fonseca, N. A., He, Y., et al. (2020). Genomic Basis for RNA Alterations in Cancer. *Nature* 578 (7793), 129–136. doi:10.1038/s41586-020-1970-0

Qi-Dong, X., Yang, X., Lu, J.-L., Liu, C.-Q., Sun, J.-X., Li, C., et al. (2020). Development and Validation of a Nine-Redox-Related Long Noncoding RNA Signature in Renal Clear Cell Carcinoma. *Oxid. Med. Cell. Longev.* 2020, 1–30. doi:10.1155/2020/6634247

Ren, J., Wang, A., Liu, J., and Yuan, Q. (2021). Identification and Validation of a Novel Redox-Related lncRNA Prognostic Signature in Lung Adenocarcinoma. *Bioengineered* 12 (1), 4331–4348. doi:10.1080/21655979.2021.1951522

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G., and Hacohen, N. (2015). Molecular and Genetic Properties of Tumors Associated with Local Immune Cytolytic Activity. *Cell* 160 (1-2), 48–61. doi:10.1016/j.cell.2014.12.033

Li, R., Yin, Y.-H., Jin, J., Liu, X., Zhang, M.-Y., Yang, Y.-E., et al. (2020). Integrative Analysis of DNA Methylation-Driven Genes for the Prognosis of Lung Squamous Cell Carcinoma Using MethylMix. *Int. J. Med. Sci.* 17 (6), 773–786. doi:10.7150/ijms.43272

Saghafinia, S., Mina, M., Riggi, N., Hanahan, D., and Ciriello, G. (2018). Pan-Cancer Landscape of Aberrant DNA Methylation across Human Tumors. *Cell Rep.* 25 (4), 1066–1080.e8. doi:10.1016/j.celrep.2018.09.082

Sen, D. R., Kaminski, J., Barnitz, R. A., Kurachi, M., Gerdemann, U., Yates, K. B., et al. (2016). The Epigenetic Landscape of T Cell Exhaustion. *Science* 354 (6316), 1165–1169. doi:10.1126/science.aae0491

Spizzo, R., Almeida, M. I., Colombatti, A., and Calin, G. A. (2012). Long Non-coding RNAs and Cancer: a New Frontier of Translational Research? *Oncogene* 31 (43), 4577–4587. doi:10.1038/onc.2011.621

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Tamang, S., Acharya, V., Roy, D., Sharma, R., Aryaa, A., Sharma, U., et al. (2019). SNHG12: An LncRNA as a Potential Therapeutic Target and Biomarker for Human Cancer. *Front. Oncol.* 9, 901. doi:10.3389/fonc.2019.00901

Trowsdale, J., and Knight, J. C. (2013). Major Histocompatibility Complex Genomics and Human Disease. *Annu. Rev. Genom. Hum. Genet.* 14, 301–323. doi:10.1146/annurev-genom-091212-153455

Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., et al. (2010). Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science* 329 (5992), 689–693. doi:10.1126/science.1192002

Vishnubalaji, R., and Alajez, N. M. (2021). Epigenetic Regulation of Triple Negative Breast Cancer (TNBC) by TGF-β Signaling. *Sci. Rep.* 11 (1), 15410. doi:10.1038/s41598-021-94514-9

Wang, H., Huo, X., Yang, X.-R., He, J., Cheng, L., Wang, N., et al. (2017). STAT3-mediated Upregulation of lncRNA HOXD-AS1 as a ceRNA Facilitates Liver Cancer Metastasis by Regulating SOX4. *Mol. Cancer* 16 (1), 136. doi:10.1186/s12943-017-0680-1

Wang, Z., Yang, B., Zhang, M., Guo, W., Wu, Z., Wang, Y., et al. (2018). lncRNA Epigenetic Landscape Analysis Identifies EPIC1 as an Oncogenic lncRNA that Interacts with MYC and Promotes Cell-Cycle Progression in Cancer. *Cancer Cell* 33 (4), 706–720.e9. doi:10.1016/j.ccell.2018.03.006

Wang, Z.-l., Wang, Z., Li, G.-z., Wang, Q.-w., Bao, Z.-s., Zhang, C.-b., et al. (2019). Immune Cytolytic Activity Is Associated with Genetic and Clinical

Properties of Glioma. *Front. Immunol.* 10, 1756. doi:10.3389/fimmu.2019.01756

Wang, M., Zheng, S., Li, X., Ding, Y., Zhang, M., Lin, L., et al. (2020). Integrated Analysis of lncRNA-miRNA-mRNA ceRNA Network Identified lncRNA EPB41L4A-AS1 as a Potential Biomarker in Non-small Cell Lung Cancer. *Front. Genet.* 11, 511676. doi:10.3389/fgene.2020.511676

Wei, Y., Dong, S., Zhu, Y., Zhao, Y., Wu, C., Zhu, Y., et al. (2019). DNA Co-methylation Analysis of lincRNAs across Nine Cancer Types Reveals Novel Potential Epigenetic Biomarkers in Cancer. *Epigenomics* 11 (10), 1177–1190. doi:10.2217/epi-2018-0138

Weigelin, B., Krause, M., and Friedl, P. (2011). Cytotoxic T Lymphocyte Migration and Effector Function in the Tumor Microenvironment. *Immunol. Lett.* 138 (1), 19–21. doi:10.1016/j.imlet.2011.02.016

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: a Class Discovery Tool with Confidence Assessments and Item Tracking. *Bioinformatics* 26 (12), 1572–1573. doi:10.1093/bioinformatics/btq170

Xu, M., Chen, X., Lin, K., Zeng, K., Liu, X., Xu, X., et al. (2019). lncRNA SNHG6 Regulates EZH2 Expression by Sponging miR-26a/b and miR-214 in Colorectal Cancer. *J. Hematol. Oncol.* 12 (1), 3. doi:10.1186/s13045-018-0690-5

Xu, D., Wang, L., Pang, S., Cao, M., Wang, W., Yu, X., et al. (2021). The Functional Characterization of Epigenetically Related lncRNAs Involved in Dysregulated CeRNA-CeRNA Networks across Eight Cancer Types. *Front. Cell Dev. Biol.* 9, 649755. doi:10.3389/fcell.2021.649755

Xu, J., Wang, Z., Li, S., Chen, J., Zhang, J., Jiang, C., et al. (2018). Combinatorial Epigenetic Regulation of Non-coding RNAs has Profound Effects on Oncogenic Pathways in Breast Cancer Subtypes. *Brief Bioinform.* 19 (1), 52–64. doi:10.1093/bib/bbw099

Yang, Z., Xu, F., Wang, H., Teschendorff, A. E., Xie, F., and He, Y. (2021). Pan-cancer Characterization of Long Non-coding RNA and DNA Methylation Mediated Transcriptional Dysregulation. *EBioMedicine* 68, 103399. doi:10.1016/j.ebiom.2021.103399

Ying, K., Wang, L., Long, G., Lian, C., Chen, Z., and Lin, W. (2020). ACTA2-AS1 Suppresses Lung Adenocarcinoma Progression via Sequestering miR-378a-3p and miR-4428 to Elevate SOX7 Expression. *Cell Biol. Int.* 44 (12), 2438–2449. doi:10.1002/cbin.11451

Li, Y., Jiang, T., Zhou, W., Li, J., Li, X., Wang, Q., et al. (2020). Pan-Cancer Characterization of Immune-Related lncRNAs Identifies Potential Oncogenic Biomarkers. *Nat. Commun.* 11 (1), 1000. doi:10.1038/s41467-020-14802-2

Zhang, G., Li, S., Lu, J., Ge, Y., Wang, Q., Ma, G., et al. (2018). LncRNA MT1JP Functions as a ceRNA in Regulating FBXW7 through Competitively Binding to miR-92a-3p in Gastric Cancer. *Mol. Cancer* 17 (1), 87. doi:10.1186/s12943-018-0829-6

Zhang, C., Ren, X., Zhang, W., He, L., Qi, L., Chen, R., et al. (2020). Prognostic and Clinical Significance of Long Non-coding RNA SNHG12 Expression in Various Cancers. *Bioengineered* 11 (1), 1112–1123. doi:10.1080/21655979.2020.1831361

Zhang, H., Bian, C., Tu, S., Yin, F., Guo, P., Zhang, J., et al. (2021a). Integrated Analysis of lncRNA-miRNA-mRNA ceRNA Network in Human Aortic Dissection. *BMC Genomics* 22 (1), 724. doi:10.1186/s12864-021-08012-3

Zhang, Z., Yan, C., Li, K., Bao, S., Li, L., Chen, L., et al. (2021b). Pan-cancer Characterization of lncRNA Modifiers of Immune Microenvironment Reveals Clinically Distinct De Novo Tumor Subtypes. *npj Genom. Med.* 6 (1), 52. doi:10.1038/s41525-021-00215-7

Zhao, H., Zhang, X., Frazão, J. B., Condino-Neto, A., and Newburger, P. E. (2013). HOX Antisense lincRNA HOXA-AS2 Is an Apoptosis Repressor in allTransretinoic Acid Treated NB4 Promyelocytic Leukemia Cells. *J. Cell. Biochem.* 114 (10), 2375–2383. doi:10.1002/jcb.24586

Zheng, R., Zheng, M., Wang, M., Lu, F., and Hu, M. (2021). Identification of a Prognostic Long Noncoding RNA Signature in Lung Squamous Cell Carcinoma: a Population-Based Study with a Mean Follow-Up of 3.5 Years. *Arch. Public Health* 79 (1), 61. doi:10.1186/s13690-021-00588-2

Zhou, R. J., and Lv, H. Z. (2019). Knockdown of ACTA2AS1 Promotes Liver Cancer Cell Proliferation, Migration and Invasion. *Mol. Med. Rep.* 19 (3), 2263–2270. doi:10.3892/mmr.2019.9856

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

GLOSSARY

BLCA Bladder Urothelial Carcinoma

BRCA Breast Invasive Carcinoma

CESC Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma

CHOL Cholangiocarcinoma

COAD Colon Adenocarcinoma

ESCA Esophageal Carcinoma

GBM Glioblastoma Multiforme

HNSC Head and Neck Squamous Cell Carcinoma

KIRC Kidney Renal Clear Cell Carcinoma

KIRP Kidney Renal Papillary Cell Carcinoma

LIHC Liver Hepatocellular Carcinoma

LUAD Lung Adenocarcinoma

LUSC Lung Squamous Cell Carcinoma

PAAD Pancreatic Adenocarcinoma

PCPG Pheochromocytoma and Paraganglioma

PRAD Prostate Adenocarcinoma

READ Rectum Adenocarcinoma

SARC Sarcoma

SKCM Skin Cutaneous Melanoma

STAD Stomach Adenocarcinoma

THCA Thyroid Carcinoma

THYM Thymoma

UCEC Uterine Corpus Endometrial Carcinoma

lncRNA Long non-coding RNAs

DMlnc differentially methylated lncRNA

NClnc negatively correlated lncRNA

UMLElnc up-methylated-low-expressed lncRNA

DMOElnc down-methylated-over-expressed lncRNA

CSNClncs cancer-specific negatively correlated lncRNAs

MHC major histocompatibility complex

CYT immunolytic cell activity

CTL cytotoxic T cell

TMB tumor mutation burden

lncRNA Long non-coding RNA

DMlnc Differentially Methylated lncRNA

NClnc Negatively Correlated lncRNA

UMLElnc Up-Methylated-Low-Expressed lncRNA

DMOElnc Down-Methylated-Over-Expressed lncRNA

CSNClnc Cancer-Specific Negatively Correlated lncRNA

MHC Major Histocompatibility Complex

CYT Cytolytic Activity

CTL Cytotoxic T Lymphocyte

TMB Tumor Mutation Burden

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: frontiersin.org/about/contact

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