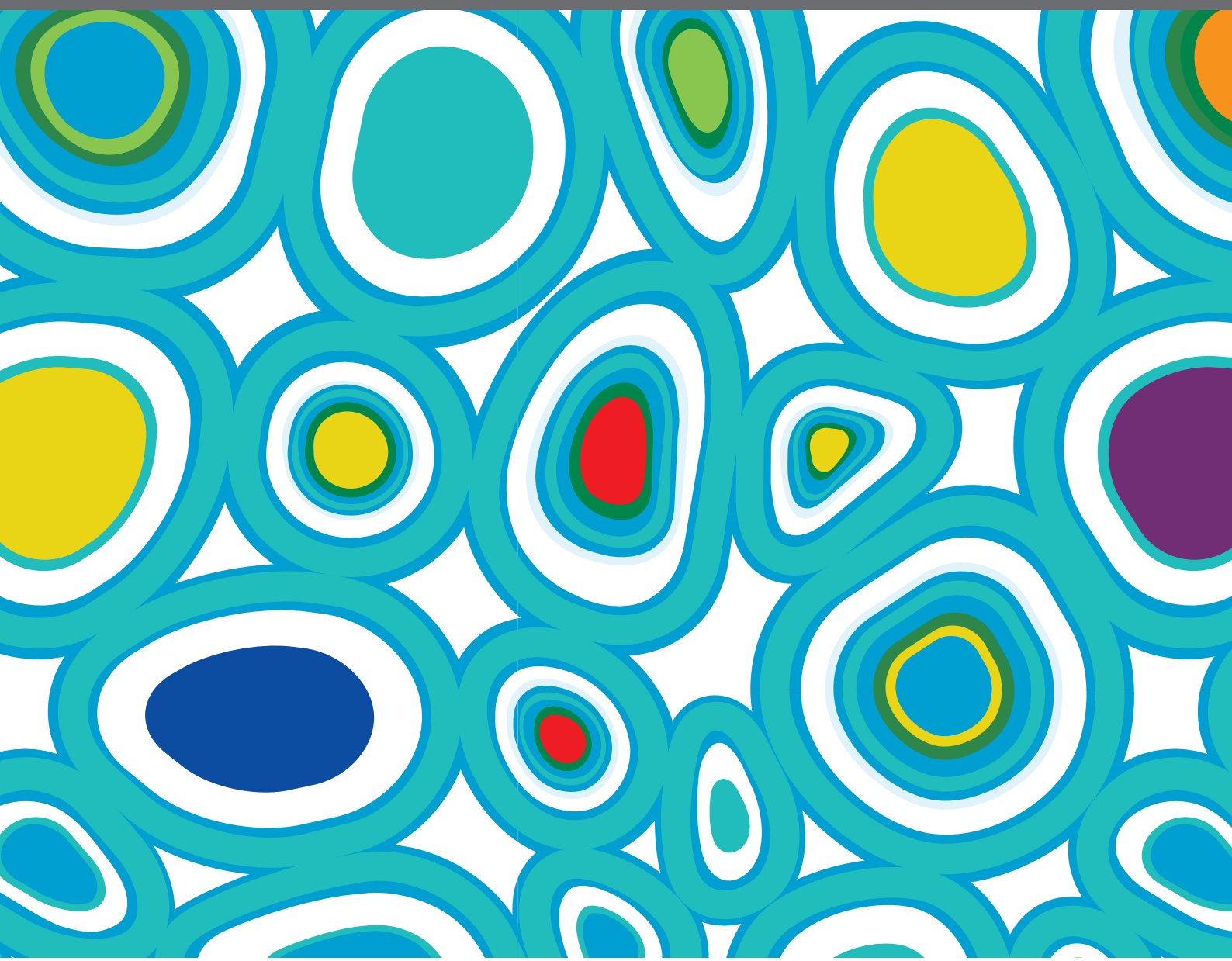
THE APPLICATION OF SEQUENCING TECHNOLOGIES AND BIOINFORMATICS METHODS IN CANCER BIOLOGY

EDITED BY: Geng Chen, Lu Xie, D. P. Kreil and Fangqing Zhao
PUBLISHED IN: Frontiers in Cell and Developmental Biology



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-83250-250-1
DOI 10.3389/978-2-83250-250-1

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

THE APPLICATION OF SEQUENCING TECHNOLOGIES AND BIOINFORMATICS METHODS IN CANCER BIOLOGY

Topic Editors:
Geng Chen, Stemirna Therapeutics Co., Ltd., China
Lu Xie, Shanghai Institute for Biomedical and Pharmaceutical Technologies, China
D. P. Kreil, Boku University Vienna
Fangqing Zhao, Beijing Institutes of Life Science (CAS), China

Citation: Chen, G., Xie, L., Kreil, D. P., Zhao, F., eds. (2022). The Application of Sequencing Technologies and Bioinformatics Methods in Cancer Biology. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-250-1

Table of Contents

frontiers | Frontiers in Cell and Developmental Biology

TYPE Editorial
PUBLISHED 06 September 2022
DOI 10.3389/fcell.2022.1002813

OPEN ACCESS

EDITED AND REVIEWED BY
Ramani Ramchandran,
Medical College of Wisconsin,
United States

*CORRESPONDENCE
Geng Chen,
chengeng66666@outlook.com

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

RECEIVED 25 July 2022
ACCEPTED 11 August 2022
PUBLISHED 06 September 2022

CITATION
Chen G, Xie L, Zhao F and Kreil DP (2022), Editorial: The application of sequencing technologies and bioinformatics methods in cancer biology.
Front. Cell Dev. Biol. 10:1002813.
doi: 10.3389/fcell.2022.1002813

Editorial: The application of sequencing technologies and bioinformatics methods in cancer biology

Geng Chen[1]*, Lu Xie[2], Fangqing Zhao[3] and David P. Kreil[4]

[1]Stemirna Therapeutics Co., Ltd., Shanghai, China, [2]Institute for Genome and Bioinformatics, Shanghai Institute for Biomedical and Pharmaceutical Technologies, Shanghai, China, [3]Computational Genomics Laboratory, Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing, China, [4]Department of Biotechnology, Boku University Vienna, Vienna, Austria

KEYWORDS

cancer biology, bulk sequencing, single-cell sequencing, bioinformatics analysis, machine learning

Editorial on the Research Topic
The application of sequencing technologies and bioinformatics methods in cancer biology

Sequencing technologies including bulk and single-cell approaches have been employed broadly to investigate molecular changes and the underlying mechanisms of cancer development, progression, and metastasis (Donoghue et al., 2020; Lei et al., 2021). For instance, whole-genome and whole-exome sequencing technologies allow the identification of diverse genetic alterations (e.g., point mutations, indels, and structural variations) that may contribute to tumorigenesis and tumor growth (Alexandrov et al., 2020; Consortium, 2020; Li et al., 2020). RNA-seq can enable the exploration of up/down-regulated genes or transcripts, the aberrant events of gene fusion, changed alternative splicing, and different RNA editing in cancers compared to corresponding controls (Dehghannasiri et al., 2019; Demircioglu et al., 2019; Group et al., 2020). The advent of single-cell sequencing approaches further provides unprecedented opportunities for dissecting tumor heterogeneity and cellular dynamics (Kinker et al., 2020; Li et al.; Leader et al., 2021). In the era of big cancer genomics data, bioinformatics approaches are crucial for effectively analyzing and interpreting the increasing amount of sequencing data from different perspectives, such as biomarker discovery (Cheong et al., 2022), outcome prediction (Zhou et al., 2022), and disease association elucidation (Xu et al., 2022). The combination of sequencing technologies and related bioinformatics approaches for cancer dissection is effectively deepening our understanding of tumor biology.

Although enormous progress and fruitful results have been achieved for diverse cancers, great efforts are needed to further promote the development of precision medicine. Specifically, the exploration of novel biomarkers and new computational

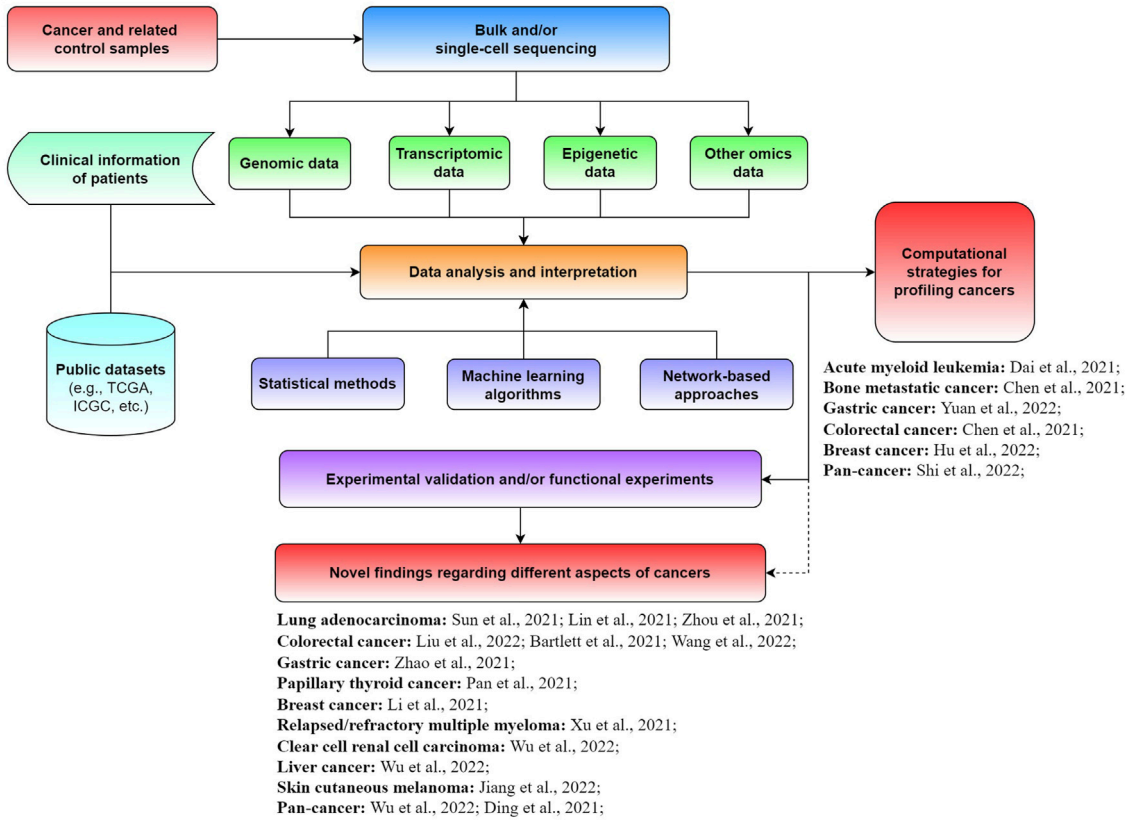


FIGURE 1
Schematic view of the application of sequencing technologies and bioinformatics approaches for cancer-related studies.

approaches is crucial for better diagnosis, prognosis, and treatment of cancer patients. In this Research Topic on *The Application of Sequencing Technologies and Bioinformatics Methods in Cancer Biology*, we aimed to collect novel findings and methods in the field of cancer biology related to mining bulk and single-cell sequencing data with bioinformatics approaches for illuminating different types of tumors (Figure 1). A total of 21 original research articles were published in this Research Topic, covering the investigation of new cancer-associated biomarkers and potential underlying mechanisms, as well as novel computational strategies for analyzing tumors. We summarize and discuss the main findings of these studies in this editorial.

Biomarker identification and characterization

Lung cancer

Three studies on this Research Topic dissected lung adenocarcinoma (LUAD) from different perspectives and identified a number of potential prognostic biomarkers. For example, Sun et al. uncovered three ferroptosis subtypes in the LUAD cohort based on expression profiling of ferroptosis driver and suppressor genes. They also detected remarkable differences in terms of the immune microenvironment and biological functions among these three subtypes. A set of ferroptosis-related signatures associated with the prognosis of LUAD patients were identified as well. Their results revealed that ferroptosis could play an important role in LUAD. In another study, Lin et al. investigated LUAD from the view of cellular senescence. They showed that the combination of senescence-related signature (SRS) and immune checkpoint expression or tumor mutation burden was an effective prognostic biomarker for immunotherapy. Their findings highlighted that SRS could regulate the immune microenvironment of LUAD.Zhou et al. explored the immune-oncology profile of LUAD and identified a set of immune genes significantly correlated with progression-free survival. They also found that the risk signature based on several of these immune genes was associated with neutrophil infiltration. Therefore, these three studies screened out different types of prognostic biomarkers for LUAD patients, which could be useful and promising for survival analysis.

Colorectal cancer

For colorectal cancer (CRC), three studies discovered potential prognostic biomarkers or treatment preditors by analyzing large-scale datasets. For instance, Liu et al. identified five CRC survival-related genes (SLC26A3, GUCA2A, CLCA4, CLCA1, and AQP8) based on hub gene analysis of co-expression and protein-protein interaction networks. They also constructed a promising risk model using three (CLCA1, CLCA4, and GUCA2A) of them for prognostic prediction. Their results indicated that these hub genes could be correlated with CRC development and the survival of patients. Wang et al. constructed a scoring model based on four necroptosis-related genes (FAS, CAMK2B, STAT4, and CYBB), which can effectively predict the prognosis and response of colon cancer patients to chemotherapy and immunotherapy. They also found that the activation of necroptosis-related genes could promote the metastasis of colon cancer. Unlike these two studies, Bartlett et al. investigated the association between miRNAs and immune checkpoint blockade for CRC. They detected a number of miRNAs that were correlated with mutation burden, microsatellite instability, or PD-L1 expression. Among them, three miRNAs (miRNA-146b, miRNA-155, and miRNA-22) were related to the M1 macrophage polarization state and their targets had the potential to impact the TGF-β pathway.

Other cancer types

In addition to biomarker exploration articles in lung and colorectal cancer, we received articles that explored biomarkers in gastric, thyroid, breast, and other tumor types. For example, Zhao et al. discovered that loss of ARID1A was strongly correlated with the high expression of CD47 in gastric cancer. They identified CD47 as a potential direct downstream target of ARID1A, while the combination of ARID1A and CD47 would be a promising prognostic biomarker. Pan et al. examined the landscape of papillary thyroid cancer (PTC) at single-cell resolution and detected molecular signatures correlated with the disease-free survival of patients. They also found that dendritic cells and B cells could play critical functions in preventing PTC progression. identified GPX1 as a promising prognostic biomarker for breast cancer (Li et al., 2021b). They observed that anhydroicaritin may suppress epithelial-mesenchymal transition by elevating GPX1 expression, which could provide potential guidance for breast cancer treatment. Xu et al. examined the dynamic genomic changes of relapsed/refractory multiple myeloma (RRMM) based on whole-exome sequencing and single-cell RNA-seq data. They found that RUNX3 could be a potential driver and therapeutic target for RRMM. Wu et al. detected KIF23 as an effective prognostic biomarker for clear cell renal cell carcinoma (ccRCC). They also revealed that KIF23 could promote the nuclear translocation of β-catenin, while the knockdown of KIF23 would reduce the proliferation, migration, and invasion of ccRCC. For liver cancer, Wu et al. showed that up-regulated expression of POSTN, LAYN, and HTRA3, and down-regulated expression of AANAT and AFM were associated with poorer overall survival of patients. They also found that the gut microbial metabolites of trimethylamine N-oxide TMAO and POSTN were potential targets for liver cancer treatment. Jiang et al. revealed that TMEM176B could be a diagnostic and prognostic biomarker for skin cutaneous melanoma. The expression of TMEM176B was shown to be correlated with tumor-infiltrating lymphocytes, pathological stage, therapy sensitivity to radiation, as well as tumor ulceration.

Another two studies conducted pan-cancer analyses and detected potential biomarkers for the clinical management of cancers. For example, Wu et al. systematically analyzed AP3S1 in diverse tumor types and found that its expression was widely associated with the immunosuppressive microenvironment. They also demonstrated that AP3S1 could be a pan-cancer biomarker for prognosis and immunotherapy. The other study explored the gene expression profiles of 988 cell lines from 20 distinct cancers by employing several computational methods Ding et al. They identified robust pan-cancer biomarkers for differentiating a variety of cancer types. Overall, the aforementioned studies discovered different types of potential biomarkers for diverse tumors, which could be useful for the diagnosis, prognosis, or treatment of corresponding cancers.

Computational strategies for profiling cancers

Besides biomarker investigation, several studies on this Research Topic developed new computational strategies for profiling tumors from different aspects. For instance, Dai et al. developed a novel model based on the score of cell type compositions (CTCs) for improving the prognostic analysis of acute myeloid leukemia patients (AML). They further showed that the CTC score could potentially benefit the individualized treatment of AML patients. Chen et al. systematically investigated the chromosome instability (CIN) profile of 280 patients with bone metastatic cancer based on the copy number variations inferred from cell-free DNA (cfDNA) sequencing data. They revealed that CIN quantification with cfDNA provided an effective and non-invasive method for predicting the survival of spine metastasis patients. Yuan et al. developed a scoring approach based on 15-DNA repair gene signatures for effectively predicting the prognosis of gastric cancer patients who received immunotherapies. The scoring system developed by them may benefit the tailored immunotherapy of gastric cancers. Chen et al. proposed an efficient strategy for identifying the circular RNAs (circRNAs)

with protein-coding potential in CRC. They also suggested that those circRNAs might be functional in promoting proliferation and invasion ability, while the peptides derived from circRNAs could be potential targets for CRC therapy or diagnosis. Hu et al. designed a panel based on 28 breast cancer-related genes for long-read sequencing (e.g., Oxford Nanopore and Pacific Biosciences platforms). They demonstrated that this approach can effectively detect structural variations in breast cancer patients, which could be used in related clinical investigations. Shi et al. systematically evaluated a gene panel containing ~1,300 key immuno-oncology genes designed for characterizing tumor microenvironments. Based on the analysis of >1,200 formalin-fixed paraffin-embedded tumor samples, they showed that this panel was comparable with orthogonal platforms (e.g., RNA-seq, hematoxylin and eosin staining, and immunohistochemistry). The computational strategies developed in these studies were useful and promising for exploring and characterizing various aspects of different cancers.

Summary and perspective

Taken together, the studies published in this Research Topic presented a diversity of interesting and meaningful results for a range of different cancers, which could facilitate our understanding of cancer biology. It is well known that bulk and single-cell sequencing technologies can respectively obtain whole-system and cellular views of tumors. Multi-omics and multimodal strategies are superior to single-omics methods for dissecting cancers since different types of data could be complementary (Li et al., 2021c). Third-generation sequencing technologies are gradually maturing, enabling the production of much longer reads than that of the currently abundant next-generation sequencing protocols (Logsdon et al., 2020; De Coster et al., 2021). Novel computational tools based on advanced machine learning algorithms may further help researchers to process the growing amount of sequencing data more efficiently (Park et al., 2021). These advancements are transforming data analysis and interpretation for better cancer management. Overall, innovation both in sequencing technologies and bioinformatics approaches will continue to facilitate the translation of big cancer genomic data into clinical practice and benefit precision medicine. We hope that this Research Topic will inspire researchers to further investigate cancers by integrative analysis of different omics data in a systematic way.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

Funding

This work was supported by the Science and Technology Commission of Shanghai Municipality, China (20S11909100); Shanghai strategic emerging industry development special fund (ZJ640070216); the Science and Technology Project of Guizhou Province [(2020)4Y217]; and the project of mRNA Innovation and Translation Center, Shanghai, China.

Conflict of interest

Author GC is employed by Stemirna Therapeutics Co., Ltd.

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

Publisher's note

References

Alexandrov, L. B., Kim, J., Kim, N. J., Haradhvala, M. N., Huang, A. W., Tian Ng, Y., et al. (2020). The repertoire of mutational signatures in human cancer. *Nature* 578, 94–101. doi:10.1038/s41586-020-1943-3

Cheong, J. H., Wang, S. C., Park, S., Porembka, M. R., Christie, A. L., Kim, H., et al. (2022). Development and validation of a prognostic and predictive 32-gene signature for gastric cancer. *Nat. Commun.* 13, 774. doi:10.1038/s41467-022-28437-y

Consortium, I. T. P.-C. A. o. W. G. (2020). Pan-cancer analysis of whole genomes. *Nature* 578, 82–93. doi:10.1038/s41586-020-1969-6

De Coster, W., Weissensteiner, M. H., and Sedlazeck, F. J. (2021). Towards population-scale long-read sequencing. *Nat. Rev. Genet.* 22, 572–587. doi:10.1038/s41576-021-00367-3

Dehghannasiri, R., Freeman, D. E., Jordanski, M., Hsieh, G. L., Damljanovic, A., Lehnert, E., et al. (2019). Improved detection of gene fusions by applying statistical methods reveals oncogenic RNA cancer drivers. *Proc. Natl. Acad. Sci. U. S. A.* 116, 15524–15533. doi:10.1073/pnas.1900391116

Demircioglu, D., Cukuroglu, E., Kindermans, M., Nandi, T., Calabrese, C., Fonseca, N. A., et al. (2019). A pan-cancer transcriptome analysis reveals pervasive regulation through alternative promoters. *Cell* 178, 1465–1477 e1417. doi:10.1016/j.cell.2019.08.018

Donoghue, M. T. A., Schram, A. M., Hyman, D. M., and Taylor, B. S. (2020). Discovery through clinical sequencing in oncology. *Nat. Cancer* 1, 774–783. doi:10.1038/s43018-020-0100-0

Group, P. T. C., Calabrese, C., Davidson, N. R., Demircioglu, D., Fonseca, N. A., He, Y., et al. (2020). Genomic basis for RNA alterations in cancer. *Nature* 578, 129–136. doi:10.1038/s41586-020-1970-0

Kinker, G. S., Greenwald, A. C., Tal, R., Orlova, Z., Cuoco, M. S., McFarland, J. M., et al. (2020). Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nat. Genet.* 52, 1208–1218. doi:10.1038/s41588-020-00726-6

Leader, A. M., Grout, J. A., Maier, B. B., Nabet, B. Y., Park, M. D., Tabachnikova, A., et al. (2021). Single-cell analysis of human non-small cell lung cancer lesions refines tumor classification and patient stratification. *Cancer Cell* 39, 1594–1609. e1512. doi:10.1016/j.ccell.2021.10.009

Lei, Y., Tang, R., Xu, J., Wang, W., Zhang, B., Liu, J., et al. (2021). Applications of single-cell sequencing in cancer research: Progress and perspectives. *J. Hematol. Oncol.* 14, 91. doi:10.1186/s13045-021-01105-2

Li, Y., Chen, J., Xu, Q., Han, Z., Tan, F., Shi, T., et al. (2021b). Single-cell transcriptomic analysis reveals dynamic alternative splicing and gene regulatory networks among pancreatic islets. *Sci. China Life Sci.* 64, 174–176. doi:10.1007/s11427-020-1711-x

Li, Y., Ma, L., Wu, D., and Chen, G. (2021c). Advances in bulk and single-cell multi-omics approaches for systems biology and precision medicine. *Brief. Bioinform* 22. doi:10.1093/bib/bbab024

Li, Y., Roberts, N. D., Wala, J. A., Shapira, O., Schumacher, S. E., Kumar, K., et al. (2020). Patterns of somatic structural variation in human cancer genomes. *Nature* 578, 112–121. doi:10.1038/s41586-019-1913-9

Logsdon, G. A., Vollger, M. R., and Eichler, E. E. (2020). Long-read human genome sequencing and its applications. *Nat. Rev. Genet.* 21, 597–614. doi:10.1038/s41576-020-0236-x

Park, Y., Heider, D., and Hauschild, A. C. (2021). Integrative analysis of next-generation sequencing for next-generation cancer research toward artificial intelligence. *Cancers (Basel)* 13. doi:10.3390/cancers13133148

Xu, N., Yao, Z., Shang, G., Ye, D., Wang, H., Zhang, H., et al. (2022). Integrated proteogenomic characterization of urothelial carcinoma of the bladder. *J. Hematol. Oncol.* 15, 76. doi:10.1186/s13045-022-01291-7

Zhou, Z., Ding, Z., Yuan, J., Shen, S., Jian, H., Tan, Q., et al. (2022). Homologous recombination deficiency (HRD) can predict the therapeutic outcomes of immuno-neoadjuvant therapy in NSCLC patients. *J. Hematol. Oncol.* 15, 62. doi:10.1186/s13045-022-01283-7

frontiers in Cell and Developmental Biology

ORIGINAL RESEARCH
published: 14 October 2021
doi: 10.3389/fcell.2021.754507

The miRNA Profile of Inflammatory Colorectal Tumors Identify TGF-β as a Companion Target for Checkpoint Blockade Immunotherapy

Bjarne Bartlett[1,2†], Zitong Gao[1,2†], Monique Schukking[2,3], Mark Menor[1], Vedbar S. Khadka[1], Muller Fabbri[3], Peiwen Fei[3] and Youping Deng[1,2]*

[1] Bioinformatics Core, Department of Quantitative Health Sciences, University of Hawaii, Honolulu, HI, United States,
[2] Department of Molecular Biosciences and Bioengineering, University of Hawai'i at Mānoa, Honolulu, HI, United States,
[3] Cancer Biology Program, University of Hawai'i Cancer Center, Honolulu, HI, United States

Extrinsic factors such as expression of PD-L1 (programmed dealth-ligand 1) in the tumor microenvironment (TME) have been shown to correlate with responses to checkpoint blockade therapy. More recently two intrinsic factors related to tumor genetics, microsatellite instability (MSI), and tumor mutation burden (TMB), have been linked to high response rates to checkpoint blockade drugs. These response rates led to the first tissue-agnostic approval of any cancer therapy by the FDA for the treatment of metastatic, MSI-H tumors with anti-PD-1 immunotherapy. But there are still very few studies focusing on the association of miRNAs with immune therapy through checkpoint inhibitors. Our team sought to explore the biology of such tumors further and suggest potential companion therapeutics to current checkpoint inhibitors. Analysis by Pearson Correlation revealed 41 total miRNAs correlated with mutation burden, 62 miRNAs correlated with MSI, and 17 miRNAs correlated with PD-L1 expression. Three miRNAs were correlated with all three of these tumor features as well as M1 macrophage polarization. No miRNAs in any group were associated with overall survival. TGF-β was predicted to be influenced by these three miRNAs (p = 0.008). Exploring miRNA targets as companions to treatment by immune checkpoint blockade revealed three potential miRNA targets predicted to impact TGF-β. M1 macrophage polarization state was also associated with tumors predicted to respond to therapy by immune checkpoint blockade.

Keywords: PD-L1, immunotherapy, checkpoint blockade, MSI, mutation burden

OPEN ACCESS

Edited by:
Geng Chen, GeneCast Biotechnology Co., Ltd., China

Reviewed by:
Chengming Liu, Chinese Academy of Medical Sciences and Peking Union Medical College, China
Theofano Orfanelli, Stony Brook University, United States

****Correspondence:***
Youping Deng
dengy@hawaii.edu

† 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: *06 August 2021*
Accepted: *13 September 2021*
Published: *14 October 2021*

Citation:
Bartlett B, Gao Z, Schukking M, Menor M, Khadka VS, Fabbri M, Fei P and Deng Y (2021) The miRNA Profile of Inflammatory Colorectal Tumors Identify TGF-β as a Companion Target for Checkpoint Blockade Immunotherapy. Front. Cell Dev. Biol. 9:754507. doi: 10.3389/fcell.2021.754507

INTRODUCTION

Despite therapeutic advances and declining mortality since 1990, an estimated 50,630 patients in the United States die annually from colorectal adenocarcinomas (Siegel et al., 2018). New tools for precision medicine are necessary to build upon decades of progress in diagnosing and treating colon cancer. Immune Checkpoint inhibition (ICI) therapies, which block interactions between

Abbreviations: ORR, Overall Response Rate; MSI, Microsatellite Instability; MSS, Microsatellite Stable; COAD, Colon adenocarcinoma; READ, Rectal Adenocarcinoma; MOSAIC, MicrOSAtellite Instability Classifier; TCGA, The Cancer Genome Atlas; TGF-β, Transforming growth factor beta.

ligands and receptors, are one such innovation that have shown durable anti-tumor response. A combination of both intrinsic and extrinsic tumor features has been shown to correlate with response to checkpoint blockade therapy. Extrinsic factors, such as programmed cell-death ligand 1 (PD-L1) expression in the tumor microenvironment have been shown to correlate with responses to checkpoint blockade therapy (Topalian et al., 2012). More recently, two intrinsic factors related to tumor genetics, microsatellite instability (MSI), and tumor somatic mutation burden (TMB), have been linked to high ICI response rates (Snyder et al., 2014; Le et al., 2015). The high overall response rate (ORR) of solid tumors that are MSI-high (MSI-H) has led to the first tissue agnostic approval for a cancer therapy by the FDA in MSI-H metastatic tumors (Le et al., 2017; U.S. Food and Drug Administration, 2017). However, individual tumors continue to display a range of responses to checkpoint inhibition, highlighting the need for additional research to improve biomarkers and therapeutic approaches.

microRNAs (miRNA) are small, non-coding RNAs that usually function to regulate the expression of a particular gene by depleting the cellular protein contents. This is achieved post-transcriptionally through binding of miRNA to a complementary part of the mRNA transcript for a specific protein. The binding of miRNA to mRNA largely takes place in the 3′ untranslated region and results in either a particular mRNA not being translated or its degradation by the RNA interference effector complex (RISC) (Catalanotto et al., 2016). Because of their importance in many cellular processes, the discovery of miRNAs has led to major advances in understanding and treatment of diseases including pharmocologic approaches. In the first pharmacologic use of miRNAs, Krutzfeldt et al. (2005) showed that a 23-nucleotide RNA molecule, complementary to the miR-122 target, could be delivered to liver tissue ablating endogenous miR-122.

Dysregulated miRNAs are common feature of tumor cells that target oncogenes, tumor suppressor genes, and key immunologic pathways for tumorigenesis (Zhou et al., 2014; Chen et al., 2016; Fang et al., 2018; Vannini et al., 2018). miRNAs have been identified as important aspects of the molecular circuitry underlying cancer—miR-155, for example, has been found to be upregulated in many cancers. Van Roosbroeck et al. (2017) demonstrated that miR-155 directly targets *TP53*, thus functioning as an oncogene. Up till now, there have been several publications concerned with miRNA-based signatures in CRC screening programs. For example, miR-320d is found to be a promising non-invasive diagnostic biomarker that can significantly distinguish the metastatic from non-metastatic CRC patients (Tang et al., 2019). miR-378a-3p were identified as a potential circulating marker to differentiate the CRC patients from healthy subjects (Zanutto et al., 2020). Decreased exosomal miR-139-3p expression may take a role as a novel biomarker for early diagnosis monitoring in CRC patients (Liu et al., 2020). miRNAs have also been found to play an important role in regulating the immune environment. In addition to functioning as an oncogene, miR-155 was found by Lu et al. (2016) to promote M1 polarization along with miR-147-3p, and miR-9-5p. But there are still very few studies focusing on the association of miRNAs with immune therapy through checkpoint inhibitors in CRC.

In addition, the development of therapeutic targets that utilize RNA interference is an active area of pharmacologic research. Our team also sought to further explore the biology of MSI-H tumors and suggest potential companion therapeutics to current checkpoint inhibitors. To do this, we initiated an *in silico* study to look at all three molecular phenotypes indicative of response to ICI therapeutics in the colon and rectal adenocarcinoma (CRC) cohorts from The Cancer Genome Atlas (TCGA) and further characterized changes in both the miRNA and transcriptomes.

MATERIALS AND METHODS

Gathering Data

COAD data from The Cancer Genome Atlas was selected for analysis because many different types of analysis were available for the same patient cohort including: somatic mutation burden, MSI status, mRNA analysis, and miRNA analysis. For our TCGA cohort, miRNA and mRNA expression data were procured from the Broad Firehose (Firehose, 2016). Somatic mutation calls were obtained from the Genomic Data Commons for all CRC patients in TCGA (Grossman et al., 2016).

Obtaining Tumor Features

We chose tumor pathologies previously associated with response to checkpoint blockade immunotherapy for assessment in our CRC patient cohort from TCGA (**Figure 1B**). To compare and contrast miRNA expression between these tumor features, we also compared tumor phenotypes where one would expect a great deal of overlap, for example, MSI and TMB. MSI was assessed with the MicrOSAtellite Instability Classifier (MOSAIC) from Hause et al. (2016) to predict MSI status based on Whole Exome Sequencing (WES) data. The proportion of unstable microsatellite loci across the exome was correlated with the expression of miRNA. TMB was assessed using Mutect2 and a 5% cutoff for allele frequency (Cibulskis et al., 2013). Expression of PD-L1 was assessed by quantifying gene expression—FPKM values from TCGA were used for this.

Statistical Analysis

To assess whether each tumor feature was correlated with the presence of a particular miRNA, a Pearson correlation coefficient was used. miRNAs were individually assessed for correlation with each tumor feature. Once correlations were assessed for the different tumor features, miRNAs were pooled to look for miRNAs that were correlated with all 3 tumor features.

Immune Cell Deconvolution

In order to alleviate bias from any one algorithm, three separate tools were used to assess immune cell populations: xCell, TIMER, and CIBERSORT. CIBERSORT reports the fraction of 22 different immune cell lineages that are present in a particular RNA-Seq sample (Chen et al., 2018). xCell, similar to CIBESORT, is a gene signature-based method used to infer 64 immune and stromal cell types (Aran et al., 2017). The Tumor Immune Estimation Resource (TIMER) allows the calculation of six tumor-infiltrating immune subsets from gene expression

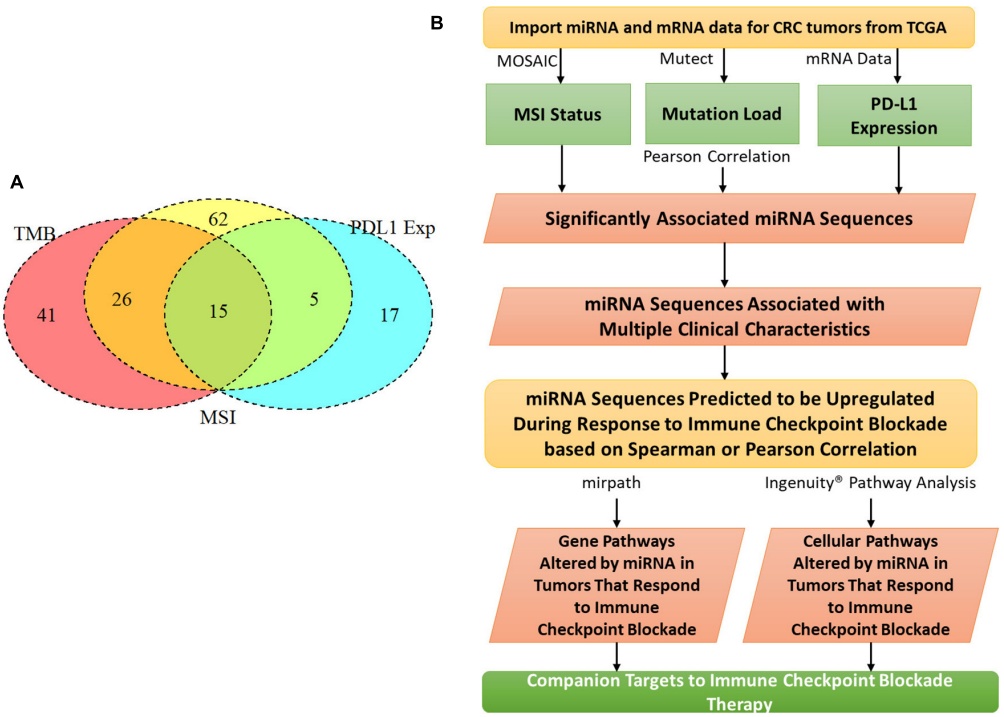


FIGURE 1 | MiRNAs correlated with clinical features related to immunotherapy. **(A)** Tumor mutation burden, programmed death ligand 1 expression, CD8 fraction, and microsatellite instability were analyzed for a cohort of 549 colorectal cancer patients in The Cancer Genome Atlas. 15 miRNAs were identified that correlated with all 3 clinical features. **(B)** The whole analysis pipeline of the whole project.

data (Li et al., 2017). *T*-tests were used for each algorithm to determine whether the fraction of immune cells differed between phenotypic classifications of tumors. Once the group of miRNAs was determined to influence macrophage polarization in aggregate, each miRNA was individually assessed to determine whether it was correlated with macrophage polarization.

Pathway Analysis

mirPath (v3), a tool for predicting gene targets of miRNA sequences, was used to analyze which pathways the selected group of miRNA would preferentially (Vlachos et al., 2015). Once miRNA's were identified that correlated with macrophage polarization, these miRNAs were analyzed with mirPath to see which genes and pathways were targeted. TargetScan was queried using a conservation score of 0.1 to find genes and pathways intersected by miR-22, miR-155, or miR-146b (Karagkouni et al., 2018). Cancer-related genes and pathways were selected from those targeted by these miRNAs.

RESULTS

Patient Cohort

The CRC patient cohort (n = 549) from TCGA was made up of 406 colon adenocarcinoma (COAD) patients and 143 rectal adenocarcinoma (READ) patients (**Supplementary Table 1**). More detailed information is shown in **Supplementary Table 2**. Typical immunotherapy recipients have late-stage cancers—we looked at stage in order to ensure a patient population representative of current immunotherapy recipients. We found

TABLE 1 | miRNAs associated with 3 tumor phenotypes.

miRNA	Association MMR- tumor features*	Associated with survival? (Z score)	Associated with macrophage polarization
let-7i	Up	No	No
mir-1266	Up	No	No
mir-132	Up	No	No
mir-146b	Up	No	Yes
mir-155	Up	No	Yes
mir-212	Up	No	No
mir-22	Up	No	Yes
mir-223	Up	No	No
mir-511 (3p/5p)	Up	No	No
mir-625	Up	No	No
mir-629	Up	No	No
mir-335	Down	No	No
mir-552	Down	No	No
mir-92a	Down	No	No

MiRNAs found to be associated with 3 tumor phenotypes through the analysis described in **Figure 1B. 15 miRNAs were found to have a common association with all 3 tumor phenotypes using Pearson's correlation. Whether the association was positive (Yes) or negative (No) was determined from the correlation coefficient, association with survival was determined using the R survival package, and association with macrophage polarization was determined using CIBERSORT and Pearson's correlation.*

that 14% of the total CRC patient cohort was advanced stage (IV). CRC patients were MSI-H at a rate of 18% in our CRC cohort, consistent with the literature. We chose three clinical features previously found to influence response to immunotherapy: MSI, tumor mutation burden (TMB), and PD-L1 expression. By aggregating these features, we aimed to predict an immunogenic subset of tumors from TCGA.

miRNAs Associated With Clinical Features Related to Immunotherapy

To characterize the relationship between miRNAs and the clinical features analyzed, a Pearson correlation was chosen. We measured linear correlations between each clinical variable and miRNA expression. Our Pearson correlation analysis resulted in 41 miRNAs significantly correlated with mutation burden, 62 miRNAs significantly correlated with MSI, and 17 miRNAs significantly correlated with PD-L1 expression. Of these three lists, 15 miRNA were overlapped and 12 of them were consistently positively correlated with the 3 tumor features and three of them were negatively correlated with the three tumor features (**Figure 1A**). 15 of these miRNAs were used for further analysis because they were correlated with all three tumor features (**Table 1**). To further characterize the 15 miRNA that were correlated with our clinical features, we conducted pathway analysis revealing 2 immune-related pathways for further exploration: Colorectal cancer and TGF-β.

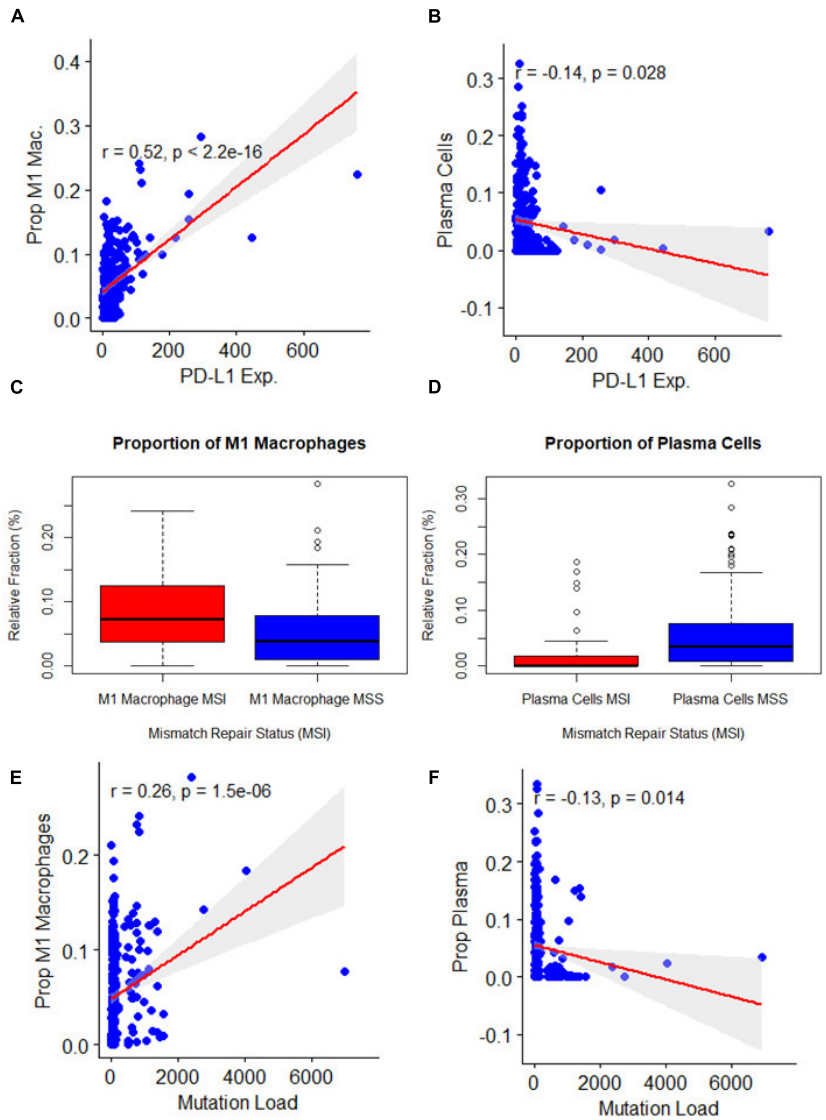


FIGURE 2 | Association of clinical features related to immunotherapy with immune cell types **(A,B)**. Association of microsatellite instability status with: M1 macrophage polarization (p = 0.00) and plasma cells (p = 0.00). **(C,D)** Association of programmed death-ligand 1 expression with: M1 macrophage polarization (p = 0) and plasma cells (p = 0.03), y axis represented the immune cell deconvolution results as fraction relative to the immune-cell content: M1 macrophage and plasma cells. **(E,F)**: Association of mutation burden with: M1 macrophage polarization (p = 0.00) and plasma cells (p = 0.01).

Association of Clinical Feature Related to Immunotherapy With Immune Cell Types

MSI, TMB, and PD-L1 expression were all separately assessed for Pearson correlations with the proportions of different immune cell types as reported by three separate immune cell deconvolution algorithms. Out of many cell types, only the proportions of plasma cells and M1 macrophages were significantly correlated with all three tumor features. The proportion of M1 macrophages was highly positively correlated with PD-L1 expression (**Figure 2A**, $p < 0.001$), MSI (**Figure 2C**, $p = 0.001$), and TMB (**Figure 2E**, $P < 0.001$). However, the proportion of plasma cells was negatively correlated with PD-L1 expression (**Figure 2B**, $p < 0.001$), MSI (**Figure 2D**, $p = 0.001$), and TMB (**Figure 2F**, $P < 0.001$) To further characterize the relationship between the proportion of M1 macrophages and the three tumors analyzed, we looked at correlations between

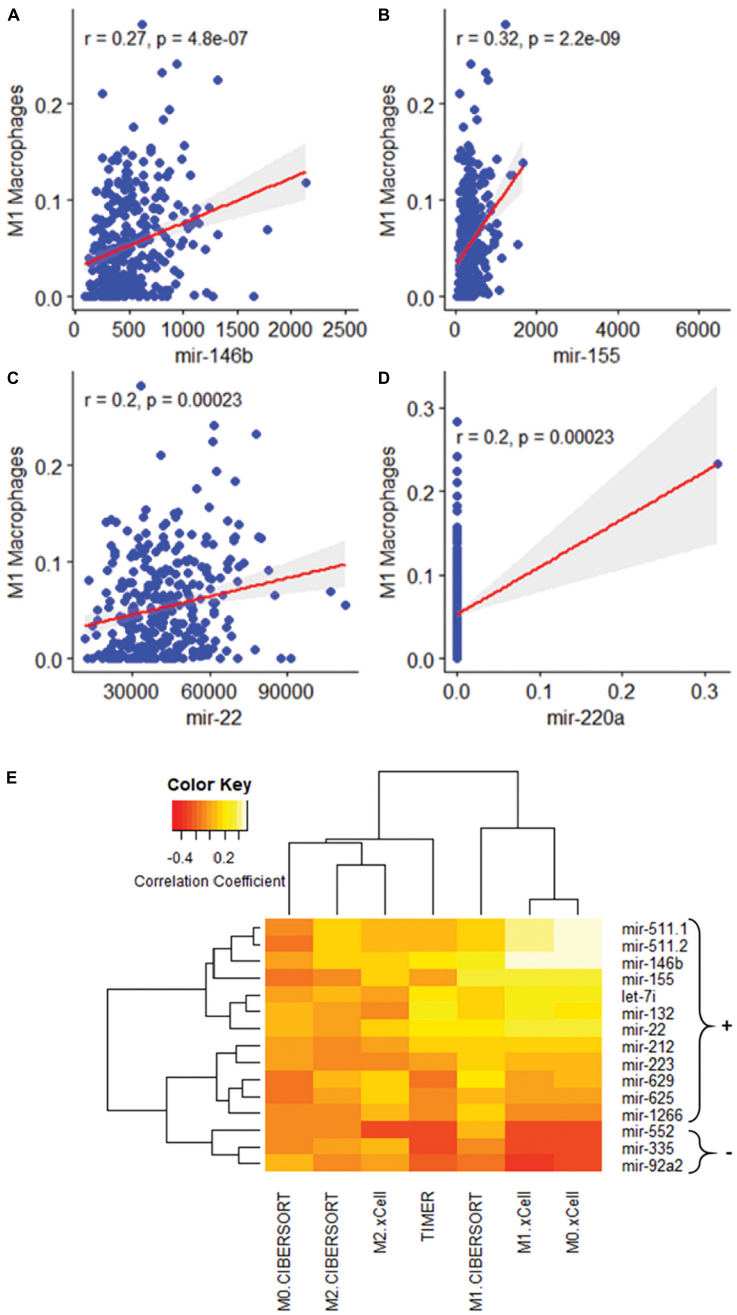


FIGURE 3 | Association of M1 macrophage polarization with miRNAs. **(A–D)** Association of M1 macrophage polarization with: mir-146b ($p = 0.00$), mir-155 ($p = 0.00$), and mir-22 ($p = 0.00$). mir-220a was excluded as the correlation was the result of a single outlier. **(E)** A heatmap of 15 miRNA sequences correlated with macrophage polarization.

TABLE 2 | Genes targeted by miRNAs interact with in the TGF-β pathway and colorectal cancer pathway.

Gene*	Ensembl ID	TGF-β	CRC	hsa-miR-155-5p	hsa-miR-22-3p
SMAD2	ENSG00000175387	Yes	Yes	Yes	
ACVR1B	ENSG00000135503	Yes	No		Yes
SKP1	ENSG00000113558	Yes	No		Yes
ACVR2B	ENSG00000114739	Yes	No	Yes	Yes
SMAD4	ENSG00000141646	Yes	Yes		Yes
ZFYVE9	ENSG00000157077	Yes	No		Yes
ACVR2A	ENSG00000121989	Yes	No	Yes	
SP1	ENSG00000185591	Yes	No		Yes
EP300	ENSG00000100393	Yes	No		Yes
TGFBR2	ENSG00000163513	Yes	Yes	Yes	
FOS	ENSG00000170345	No	Yes	Yes	
GSK3B	ENSG00000082701	No	Yes	Yes	
PIK3CB	ENSG00000051382	No	Yes		Yes
KRAS	ENSG00000133703	No	Yes	Yes	
TP53	ENSG00000141510	No	Yes		Yes
PIK3CD	ENSG00000171608	No	Yes		Yes
CCND1	ENSG00000110092	No	Yes	Yes	
PIK3R1	ENSG00000145675	No	Yes	Yes	
AKT3	ENSG00000117020	No	Yes		Yes
PIK3CA	ENSG00000121879	No	Yes	Yes	
MAPK10	ENSG00000109339	No	Yes		Yes

**MiRNA associated with microsatellite instability status, somatic tumor mutation burden, PD-L1 expression, M1 macrophage polarization that interact with the TGF-β signaling pathway (p = 0.008) and CRC pathways (p = 0.0001). Most of these genes interact with 2 miRNA sequences: hsa-miR-155-5p (p = 0.004) and hsa-miR-22-3p (p = 0.038). miRNA associations with genes were predicted by TargetScan (Conservation Score = 0.1). Results for TGF-β were merged by pathway union and results for CRC were merged by gene union.*

M1 macrophage proportion and the expression of individual miRNAs. Among the 15 miRNAs, we found three miRNAs were significantly correlated with both the three clinical characteristics and M1 macrophage polarization: miR-22, miR-146b, and miR-155 (**Figure 3**). One miRNA, miR-220a, was excluded from further analysis because the correlation was based entirely on a single outlier.

Pathway Analysis

To characterize the crucial pathways for modulating the tumor immune environment, we predicted pathways that would be influenced by the three miRNAs related to both macrophage polarization and three tumor features. Unsurprisingly, these three miRNAs were predicted to influence the expression of genes in key immune and cancer-related pathways (**Table 2**). As a group, miR-155 and miR-22 were predicted to strongly influence pathways related to Colorectal Cancer ($p = 0.0001$) and TGF-β signaling ($p = 0.008$). Out of 21 genes predicted to be influenced by these miRNAs, three genes were shared between pathways related to COAD and TGF-β signaling: SMAD2, SMAD4, and TGFBR2.

DISCUSSION

The aim of this study was to explore new targets for checkpoint blockade immunotherapy by exploring the unique biology of tumors known to respond to these drugs. Three features common to such tumors including high mutation burden, MSI, and PD-L1 expression were added into analysis. These features had 15 miRNAs in common, however, none of the 15 miRNAs predicted survival. M1 macrophage were found correlated with all three features through Pearson Correlation analysis. As a group, these 15 miRNAs predicted macrophage polarization. Individually assessing each of the 15 miRNAs for a correlation with macrophage polarization revealed three miRNAs that were strongly correlated with macrophage polarization: miRNA-146b, miRNA-155, and miRNA-22. Subsequent pathway analysis revealed these three miRNAs as important components of the TGF-β and Colorectal Cancer pathways.

In this study, we searched all possible datasets from TCGA and GEO, TCGA is the only dataset that both contains miRNA and transcriptome, and our study analyzed TCGA data as controlling for bias using a randomly selected testing/training dataset. MicroRNA has been regarded as important promising molecular biomarkers in several tumor types (Zhang et al., 2013; Chou et al., 2017). TGF-β has been identified as inhibiting the expansion and function of many components of the immune system (Batlle and Massague, 2019). A recent pair of papers has shown TGF-β to be an important modulator of the tumor microenvironment (Mariathasan et al., 2018; Tauriello et al., 2018). These experiments identify TGF-β signaling as an important aspect of response to PD-1-PD-L1 immunotherapy, connecting it to lower proportions of T cells in the tumor and poorer responses. This research supports the discovery of miRNAs targeting TGF-β in immunogenic tumors. TGF-β has also been shown to modulate the proportion of

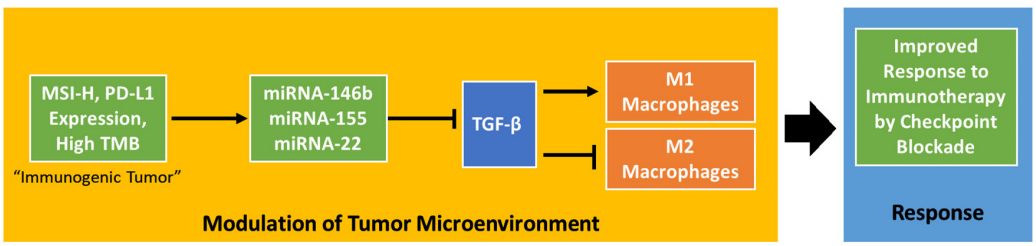


FIGURE 4 | A potential mechanism by which MSI status, PD-L1 expression, and tumor mutation burden influence the tumor microenvironment. The proposed mechanism shows that these tumor phenotypes influence the tumor microenvironment in a TGF-β-dependent way to improve response to checkpoint blockade immunotherapy.

macrophages in the tumor microenvironment, promoting their polarization to an M2-like phenotype (Gong et al., 2012). Both ideas support a key role for suppressing TGF-β in immunogenic tumors that respond to checkpoint blockade therapy. Our research further characterizes this interaction by suggesting dysregulation of miRNA in immunogenic tumors as part of the biological system enabling responses to checkpoint blockade drugs (**Figure 4**).

Although, we didn't have validation of miRNA panel, TGF-β has been widely identified in many biological experiments which can be solid support evidence for the hypothesis we presented in this research. For example, miR-146b has been found to inhibit TGF-β by binding to the 3′ untranslated region (UTR) of SMAD4, an important member of the signaling pathway. Increased SMAD 4 levels and decreased cellular proliferation was observed by Geraldo et al. (2012) in human papillary carcinoma cells. Another study found the overexpression of SMAD4 in BCPAP cells, which is a validated target of miR-146b-5p and key protein in the TGF-β signaling pathway, significantly decreased migration and invasion to a degree very similar to that observed with the antagomir-146b-5p (Lima et al., 2016). miR-155 is one of the most extensively studied miRNAs and was the first miRNA shown to be oncogenic. An extensive body of research has established an important role for miR-155 throughout cellular process related to human cancer (Costinean et al., 2006; Volinia et al., 2006). Geraldo et al. (2012) showed that miR-22 is significantly downregulated in TGF-β treated HT-29, a commonly used human colorectal cancer cell line (Cai et al., 2013).

In this study, we identified three miRNAs common in three immunotherapy-related clinical characteristics as well as M1 macrophage polarization, and function prediction of miRNAs showed SMAD2, SMAD4, and TGFBR2 were in common from COAD and TGF-β signaling pathways. miR-155 and miR-22 could influence pathways related to Colorectal Cancer and TGF-β signaling. Previous studies have already proved the regulation function of SMAD2, SMAD4, and TGFBR2 in cancers (Matsuzaki et al., 2009; Zhu et al., 2020), and these genes were also found related with miRNAs that strongly correlated with tumor features, indicating the potential function and clinical utility in immunotherapy.

CONCLUSION

Our comprehensive, integrated analysis of three miRNAs in colorectal cancer revealed a crucial component of TGF-β that modulate tumor immune environment and significantly correlated with macrophage polarization. The work highlights the important clinical implications of miRNAs functions in checkpoint blockade immunotherapy and helps develop potential therapeutical strategies for CRC 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**.

ETHICS STATEMENT

Existing data was used with permission and requested from dbGap and the Genomic Data Commons. No human subjects' data was collected.

AUTHOR CONTRIBUTIONS

BB, ZG, and YD conducted the study and prepared the manuscript. BB, ZG, PF, and YD revised the manuscript. VK and MM provided statistical and informatics support and helped prepare the figures. MS and MF provided biological expertise in analyzing miRNA data. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was partially supported by the NIH grants and Hawaii community foundation: 1R01CA223490, 2U54 CA143727, 5P30GM114737, 5P20GM103466, 5U54MD007601, 5P30CA071789, 1R01CA230514, and 21ADVC109520.

ACKNOWLEDGMENTS

JaNay Wyss for outstanding administrative and technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.754507/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.

Batlle, E., and Massague, J. (2019). Transforming growth factor-beta signaling in immunity and cancer. *Immunity* 50, 924–940. doi: 10.1016/j.immuni.2019.03.024

Cai, Z. G., Zhang, S. M., Zhang, H., Zhou, Y. Y., Wu, H. B., and Xu, X. P. (2013). Aberrant expression of microRNAs involved in epithelial-mesenchymal transition of HT-29 cell line. *Cell Biol. Int.* 37, 669–674.

Catalanotto, C., Cogoni, C., and Zardo, G. (2016). MicroRNA in control of gene expression: an overview of nuclear functions. *Int. J. Mol. Sci.* 17:1712. doi: 10.3390/ijms17101712

Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., and Alizadeh, A. A. (2018). Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol Biol.* 1711, 243–259.

Chen, H., Liu, H., Zou, H., Chen, R., Dou, Y., Sheng, S., et al. (2016). Evaluation of plasma miR-21 and miR-152 as diagnostic biomarkers for common types of human cancers. *J. Cancer* 7, 490–499.

Chou, C. K., Liu, R. T., and Kang, H. Y. (2017). MicroRNA-146b: a novel biomarker and therapeutic target for human papillary thyroid cancer. *Int. J. Mol. Sci.* 18:636. doi: 10.3390/ijms18030636

Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., et al. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* 31, 213–219. doi: 10.1038/nbt.2514

Costinean, S., Zanesi, N., Pekarsky, Y., Tili, E., Volinia, S., Heerema, N., et al. (2006). Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E(mu)-miR155 transgenic mice. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7024–7029. doi: 10.1073/pnas.0602266103

Fang, R., Zhu, Y., Hu, L., Khadka, V. S., Ai, J., Zou, H., et al. (2018). Plasma microRNA pair panels as novel biomarkers for detection of early stage breast cancer. *Front. Physiol.* 9:1879. doi: 10.3389/fphys.2018.01879

Firehose, B. G. (2016). *Broad Institute TCGA Genome Data Analysis Center (2016): Analysis-Ready Standardized TCGA Data From Broad GDAC Firehose, 2016_01_28 run. Broad Institute of MIT and Harvard. Dataset*. Available online at: http://gdac.broadinstitute.org/runs/stddata__2016_01_28/ (accessed January 28, 2016).

Geraldo, M. V., Yamashita, A. S., and Kimura, E. T. (2012). MicroRNA miR-146b-5p regulates signal transduction of TGF-beta by repressing SMAD4 in thyroid cancer. *Oncogene* 31, 1910–1922. doi: 10.1038/onc.2011.381

Gong, D., Shi, W., Yi, S. J., Chen, H., Groffen, J., and Heisterkamp, N. (2012). TGFbeta signaling plays a critical role in promoting alternative macrophage activation. *BMC Immunol.* 13:31. doi: 10.1186/1471-2172-13-31

Grossman, R. L., Heath, A. P., Ferretti, V., Varmus, H. E., Lowy, D. R., Kibbe, W. A., et al. (2016). Toward a shared vision for cancer genomic data. *N. Engl. J. Med.* 375, 1109–1112. doi: 10.1056/NEJMp1607591

Hause, R. J., Pritchard, C. C., Shendure, J., and Salipante, S. J. (2016). Classification and characterization of microsatellite instability across 18 cancer types. *Nat. Med.* 22, 1342–1350. doi: 10.1038/nm.4191

Karagkouni, D., Paraskevopoulou, M. D., Chatzopoulos, S., Vlachos, I. S., Tastsoglou, S., Kanellos, I., et al. (2018). DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA-gene interactions. *Nucleic Acids Res.* 46, D239–D245. doi: 10.1093/nar/gkx1141

Krutzfeldt, J., Rajewsky, N., Braich, R., Rajeev, K. G., Tuschl, T., Manoharan, M., et al. (2005). Silencing of microRNAs in vivo with 'antagomirs'. *Nature* 438, 685–689. doi: 10.1038/nature04303

Le, D. T., Durham, J. N., Smith, K. N., Wang, H., Bartlett, B. R., Aulakh, L. K., et al. (2017). Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* 357, 409–413. doi: 10.1126/science.aan6733

Le, D. T., Uram, J. N., Wang, H., Bartlett, B. R., Kemberling, H., Eyring, A. D., et al. (2015). PD-1 blockade in tumors with mismatch-repair deficiency. *N. Engl. J. Med.* 372, 2509–2520. doi: 10.1056/NEJMoa1500596

Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. S., et al. (2017). TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res.* 77, e108–e110. doi: 10.1158/0008-5472.CAN-17-0307

Lima, C. R., Geraldo, M. V., Fuziwara, C. S., Kimura, E. T., and Santos, M. F. (2016). MiRNA-146b-5p upregulates migration and invasion of different papillary thyroid carcinoma cells. *BMC Cancer* 16:108. doi: 10.1186/s12885-016-2146-z

Liu, W., Yang, D., Chen, L., Liu, Q., Wang, W., Yang, Z., et al. (2020). Plasma Exosomal miRNA-139-3p is a novel biomarker of colorectal cancer. *J. Cancer* 11, 4899–4906. doi: 10.7150/jca.45548

Lu, L., McCurdy, S., Huang, S., Zhu, X., Peplowska, K., Tiirikainen, M., et al. (2016). Time Series miRNA-mRNA integrated analysis reveals critical miRNAs and targets in macrophage polarization. *Sci. Rep.* 6:37446. doi: 10.1038/srep37446

Mariathasan, S., Turley, S. J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., et al. (2018). TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* 554, 544–548. doi: 10.1038/nature25501

Matsuzaki, K., Kitano, C., Murata, M., Sekimoto, G., Yoshida, K., Uemura, Y., et al. (2009). Smad2 and Smad3 phosphorylated at both linker and COOH-terminal regions transmit malignant TGF-beta signal in later stages of human colorectal cancer. *Cancer Res.* 69, 5321–5330. doi: 10.1158/0008-5472.CAN-08-4203

Siegel, R. L., Miller, K. D., and Jemal, A. (2018). Cancer statistics, 2018. *CA Cancer J. Clin.* 68, 7–30. doi: 10.3322/caac.21442

Snyder, A., Makarov, V., Merghoub, T., Yuan, J., Zaretsky, J. M., Desrichard, A., et al. (2014). Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* 371, 2189–2199. doi: 10.1056/NEJMoa1406498

Tang, Y., Zhao, Y., Song, X., Song, X., Niu, L., and Xie, L. (2019). Tumor-derived exosomal miRNA-320d as a biomarker for metastatic colorectal cancer. *J. Clin. Lab. Anal.* 33:e23004. doi: 10.1002/jcla.23004

Tauriello, D. V. F., Palomo-Ponce, S., Stork, D., Berenguer-Llergo, A., Badia-Ramentol, J., Iglesias, M., et al. (2018). TGFbeta drives immune evasion in genetically reconstituted colon cancer metastasis. *Nature* 554, 538–543. doi: 10.1038/nature25492

Topalian, S. L., Hodi, F. S., Brahmer, J. R., Gettinger, S. N., Smith, D. C., McDermott, D. F., et al. (2012). Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N. Engl. J. Med.* 366, 2443–2454. doi: 10.1056/NEJMoa1200690

U.S. Food and Drug Administration (2017). *FDA Approves First Cancer Treatment For Any Aolid Tumor With a Specific Genetic Feature*. Available online at: https://www.fda.gov/news-events/press-announcements/fda-approves-first-cancer-treatment-any-solid-tumor-specific-genetic-feature (accessed May 23, 2017).

Van Roosbroeck, K., Fanini, F., Setoyama, T., Ivan, C., Rodriguez-Aguayo, C., Fuentes-Mattei, E., et al. (2017). Combining anti-Mir-155 with chemotherapy for the treatment of lung cancers. *Clin. Cancer Res.* 23, 2891–2904. doi: 10.1158/1078-0432.CCR-16-1025

Vannini, I., Fanini, F., and Fabbri, M. (2018). Emerging roles of microRNAs in cancer. *Curr. Opin. Genet. Dev.* 48, 128–133. doi: 10.1016/j.gde.2018.01.001

Vlachos, I. S., Zagganas, K., Paraskevopoulou, M. D., Georgakilas, G., Karagkouni, D., Vergoulis, T., et al. (2015). DIANA-miRPath v3.0: deciphering microRNA function with experimental support. *Nucleic Acids Res.* 43, W460–W466. doi: 10.1093/nar/gkv403

Volinia, S., Calin, G. A., Liu, C. G., Ambs, S., Cimmino, A., Petrocca, F., et al. (2006). A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl. Acad. Sci. U.S.A.* 103, 2257–2261. doi: 10.1073/pnas.0510565103

Zanutto, S., Ciniselli, C. M., Belfiore, A., Lecchi, M., Masci, E., Delconte, G., et al. (2020). Plasma miRNA-based signatures in CRC screening programs. *Int. J. Cancer* 146, 1164–1173. doi: 10.1002/ijc.32573

Zhang, X., Li, M., Zuo, K., Li, D., Ye, M., Ding, L., et al. (2013). Upregulated miR-155 in papillary thyroid carcinoma promotes tumor growth by targeting APC and activating Wnt/beta-catenin signaling. *J. Clin. Endocrinol. Metab.* 98, E1305–E1313. doi: 10.1210/jc.2012-3602

Zhou, H., Chen, J. X., Yang, C. S., Yang, M. Q., Deng, Y., and Wang, H. (2014). Gene regulation mediated by microRNAs in response to green tea polyphenol EGCG in mouse lung cancer. *BMC Genomics* 15(Suppl. 11):S3. doi: 10.1186/1471-2164-15-S11-S3

Zhu, X., Zhang, T., Zhang, Y., Chen, H., Shen, J., Jin, X., et al. (2020). A super-enhancer controls TGF-beta signaling in pancreatic cancer through downregulation of TGFBR2. *Cell Signal.* 66:109470. doi: 10.1016/j.cellsig.2019.109470

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: 20 October 2021
doi: 10.3389/fcell.2021.743724

Ferroptosis Characterization in Lung Adenocarcinomas Reveals Prognostic Signature With Immunotherapeutic Implication

Sijin Sun[1†], Yannan Yang[1†], Zhenlin Yang[1], Juhong Wang[1], Renda Li[1], He Tian[1], Fengwei Tan[1], Qi Xue[1,2], Yibo Gao[1,2] and Jie He[1,2*]*

[1] Department of Thoracic Surgery, National Cancer Center, National Clinical Research Center for Cancer, Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [2] State Key Laboratory of Molecular Oncology, National Cancer Center, National Clinical Research Center for Cancer, Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

OPEN ACCESS

Edited by:
Geng Chen, GeneCast Biotechnology Co., Ltd., China

Reviewed by:
Zitong Gao, University of Hawai'i at Mānoa, United States
Ahreum Kim, CHA University, South Korea

***Correspondence:**
Yibo Gao
gaoyibo@cicams.ac.cn
Jie He
prof.jiehe@gmail.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: *19 July 2021*
Accepted: *22 September 2021*
Published: *20 October 2021*

Citation:
Sun S, Yang Y, Yang Z, Wang J, Li R, Tian H, Tan F, Xue Q, Gao Y and He J (2021) Ferroptosis Characterization in Lung Adenocarcinomas Reveals Prognostic Signature With Immunotherapeutic Implication. Front. Cell Dev. Biol. 9:743724. doi: 10.3389/fcell.2021.743724

The iron-dependent cell death named ferroptosis has been implicated in the progression and therapeutic response of several tumors. However, potential role of ferroptosis in lung adenocarcinomas (LUAD) remained less well understood. In TCGA-LUAD cohort, unsupervised clustering was first conducted based on ferroptosis regulators extracted from FerrDb database. Comprehensive correlation analysis and comparisons were performed among ferroptosis subtypes. The ferroptosis-related prognostic (FRP) signature was identified based on filtered features and repeated LASSO and was validated in five independent cohorts. The clinical relevance between the risk score and therapeutic response was further explored by multiple algorithms. qPCR was implemented to verify gene expression. A total of 1,168 LUAD patients and 161 ferroptosis regulators were included in this study. Three ferroptosis subtypes were identified and patients in subtype B had the best prognosis among the three subtypes. Significant differences in immune microenvironment and biological function enrichment were illustrated in distinct subtypes. The Boruta algorithm was conducted on 308 common differentially expressed genes for dimensionality reduction. A total of 56 genes served as input for model construction and a six-gene signature with the highest frequencies of 881 was chosen as FRP. The prognostic significance of FRP was validated in five independent cohorts. High FRP risk score was also linked to increased tumor mutation burden, PD-L1 protein expression and number of neoantigens. Of the FRP genes, 83.3% was abnormally expressed in LUAD cell lines. In conclusion, ferroptosis plays a non-negligible role in LUAD. Exploration of the ferroptosis pattern will enhance the prognostic stratification of individual patients and move toward the purpose of personalized treatment.

Keywords: lung adenocarcinoma, ferroptosis, classification, immunotherapy, machine learning

INTRODUCTION

Ferroptosis is a non-apoptotic modality of regulated cell death (RCD) that is induced by excessive phospholipid peroxidation that progresses in an iron-dependent manner (Dixon et al., 2012). The biology of ferroptotic cells is morphologically and mechanistically distinct from that of apoptotic cells, as indicated by ferroptosis induction in *BAX/BAK*-deficient cells and the lack of detected caspase-3 cleavage during ferroptosis (Stockwell et al., 2017). Extrinsic and intrinsic pathways that result in unrestricted phospholipid peroxidation can trigger ferroptosis (Tang and Kroemer, 2020). The extrinsic pathway includes blockade of the system x_c^- cystine/glutamate antiporter, which is the main transporter of cystine, the rate-limiting precursor of glutathione, and activator of the iron transporters serotransferrin and lactotransferrin, while the intrinsic pathway is initiated by the inhibition of intracellular antioxidant enzymes, such as glutathione peroxidase *GPX4*. Current studies have identified ferroptosis as a critical tumor suppression mechanism, and inhibition of ferroptosis, similar to inhibition of apoptosis, promotes tumor development (Gao et al., 2019; Liu et al., 2019). In addition, multiple tumor suppressors can sensitize cancer cells to ferroptosis, and among these suppressors, p53 has been thoroughly investigated (Zhang et al., 2018, 2019; Jiang et al., 2021). p53 can inhibit the transcription of the systemic x_c^- subunit *SLC7A11* (Jiang et al., 2015; Wang et al., 2016), and *P47S* amino acid mutation of p53 was reported to confer resistance to ferroptosis, thus promoting tumor formation (Jennis et al., 2016).

A number of immune checkpoint inhibitors (ICIs) have been developed and utilized in cancer patients in recent decades (He and Xu, 2020), among which anti-PD-1/PD-L1 therapy has led to great success and has been approved to treat a wide variety of cancer types and has led to intermediate response rates (Ribas and Wolchok, 2018). Anti-PD-1/PD-L1 therapy reactivates the immune system to exert antitumor effects by relieving the suppressive functions of PD-1 signaling (Ribas, 2015). Mechanistically, the antitumor effect of ICIs is strongly associated with the host's immune system and tumor immune microenvironment (TIME). The interaction between cancer cells and the TIME is complicated due to the coexistence of both protumorigenic and antitumorigenic factors. Induction of the T cell-inflamed TIME induced by JHU083 (Leone et al., 2019), CDK4/6 (Schaer et al., 2018), DCR-BCAT (Ganesh et al., 2018), or other treatments has been demonstrated to enhance the efficacy of ICIs. Thus, deeper insight into the TIME is needed to facilitate the identification of novel predictive biomarkers and the development of novel therapeutic strategies.

Recently, several studies have explored the specific association between ferroptosis and TIME infiltrating immune cells. It has been reported that anti-PD-L1 and ferroptosis inducers [such as erastin, *RSL3*, and cyst(e)inase] synergistically inhibit tumor development both *in vivo* and *in vitro* (Wang et al., 2019). Mechanistically, interferon-γ (IFNγ) released from activated cytotoxic T cells downregulates the expression of *SCL7A11* and *SLC3A2* through activation of the JAK-STAT1 pathway and thus induces ferroptosis of cancer cells. However, the cellular components released by ferroptotic cancer cells play dual roles in antitumor immunity. KRAS-G12D within exosomes secreted by ferroptotic pancreatic cancer cells can be absorbed by macrophages through *AGER*, which results in the polarization of macrophages to the M2 phenotype and an immunosuppressive microenvironment for tumor growth (Dai et al., 2020a). In addition, ferroptosis induced by conditional depletion of *Gpx4* in the pancreas promotes mutant Kras-driven tumorigenesis in mice through secretion of DNA and activation of the TMEM173/STING-dependent DNA sensor pathway in macrophages (Dai et al., 2020b).

Although success in revealing the molecular mechanisms between ferroptosis and TIME has been achieved, the abovementioned studies focused on limited ferroptosis regulators and were mainly conducted *in vitro*. Therefore, comprehensive analysis of ferroptosis regulators and TIME in real-world cohorts of patients is urgently needed. In this study, we integrated transcriptome datasets of 500 lung adenocarcinoma (LUAD) samples from The Cancer Genome Atlas (TCGA) and found three distinct ferroptosis subtypes, among which significant survival differences, diverse biological pathways, and different infiltrated immune cells were observed. Furthermore, we constructed a robust ferroptosis-related prognostic (FRP) signature composed of 6 genes (*EIF5A*, *CACYBP*, *CYCS*, *ANLN*, *ARNTL2*, and*PPM1M*) based on defined ferroptosis subtypes and confirmed the FRP signature as an independent prognostic factor for LUAD patients. Notably, patients in the high-risk FRP signature group may potentially benefit from immunotherapy due to the elevation of tumor mutation burden (TMB) and PD-L1 and neoantigen expression. In conclusion, our results revealed that ferroptosis plays critical roles in LUAD and that the FRP signature may be a promising predictive biomarker for prognosis and immunotherapy response.

MATERIALS AND METHODS

Lung Adenocarcinoma Dataset Acquisition and Preprocessing

Public lung adenocarcinoma transcriptome datasets were obtained from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO). Patients without prognostic information were excluded from the study. A total of six datasets (TCGA-LUAD, GSE13213, GSE30219, GSE31210, GSE3141, and GSE41271) were gathered for further analysis (Bild et al., 2006; Tomida et al., 2009; Okayama et al., 2012; Rousseaux et al., 2013; Sato et al., 2013). For RNA-Seq data from the TCGA, FPKM values were downloaded from the TCGA Data Portal[1] and transformed into TPM values before analysis. For datasets in the GEO, normalized probe expression values generated from the microarray were acquired using the R package GEOquery (Davis and Meltzer, 2007). The probes in each dataset were annotated according to the appropriate platform annotation file. When multiple probes mapped to the same gene, the average expression values of these genes were used as the final expression value. To calculate tumor mutation burden (TMB) in TCGA-LUAD, gene-level mutation (VarScan) data were extracted from the TCGA Data Portal. In the present study, TMB was defined as

[1]https://portal.gdc.cancer.gov

the total number of non-synonymous mutations divided by the size of the exome.

Consensus Clustering of Ferroptosis Regulators

A comprehensive list of ferroptosis drivers and suppressors was acquired from the FerrDb database[2] (Zhou and Bao, 2020). After removing duplicated entries and performing univariate Cox regression for each gene, 37 unique prognosis-associated genes were obtained for unsupervised clustering. The identification of ferroptosis subtypes was conducted with the TCGA cohort by applying the partitioning around medoids cluster algorithm in the R package ConsensusClusterPlus (Wilkerson and Hayes, 2010). The number of repetitions was set to 100 to ensure the stability of the result. The other parameters of the cluster algorithm were set to the default values.

Gene Set Variation Analysis and Functional Comparison Between Subtypes

To determine the biological function associated with different ferroptosis subtypes, GSVA was performed to detect subtle pathway activity in each sample using the R package GSVA (Hänzelmann et al., 2013). GSVA is a non-parametric, unsupervised method for calculating gene set enrichment through expression profiles. The gene set file named "c2.cp.kegg.v7.4.symbols.gmt," including 186 canonical pathways, was downloaded from MSigDB and was chosen for pathway enrichment (Liberzon et al., 2011). A pathway comparison was conducted using the R package limma (Smyth, 2005), and an adjusted p value of less than 0.05 was considered statistically significant.

Exploration of Tumor-Infiltrating Immune Cells in the TCGA-Lung Adenocarcinomas Cohort

Two widely validated algorithms, CIBERSORT and ImmuCellAI (Newman et al., 2015; Miao et al., 2020), were implemented to estimate the abundance of specific immune components. CIBERSORT is a computational approach reinforced by support vector regression to quantify 22 immune cell fractions. ImmuCellAI was used for abundance estimation of 24 immune cell types, including 18 T cell subsets, from gene expression matrices. The absolute fraction and abundance of immune cells were estimated by CIBERSORT and ImmuCellAI, respectively. Only intergroup comparisons were conducted and the differences in immune cell abundance were depicted between risk groups and among subtypes.

Construction of a Ferroptosis-Related Prognostic Model

To identify common DEGs among the three ferroptosis subtypes, gene expression was quantified by TPM value and empirical Bayes statistics was applied for differential expression analysis using limma package (Smyth, 2005). The obtained p-values were further adjusted by the Benjamini–Hochberg method. The DEGs were defined as genes with an adjusted p-value less than 0.001. By taking the intersection of the outputs, common DEGs of the three subtypes were obtained and illustrated using a Venn diagram. Next, the Boruta feature selection algorithm was applied to narrow the genes for model construction using the R package Boruta (Kursa and Rudnicki, 2010). Principal component analysis (PCA) was p to verify the ability to distinguish subtypes using selected genes. Furthermore, genes related to prognosis were screened out through univariate Cox regression and served as input for model training. To generate a robust ferroptosis-related prognostic (FRP) signature, LASSO Cox regression with 10-fold cross-validation was performed using a random seed, and the whole procedure was repeated 1,000 times (Tibshirani, 1997). The model with the highest frequency was selected as the final signature. Patients were subsequently stratified into high-risk and low-risk groups according to the median of the risk score. Both Kaplan–Meier curves and multivariate Cox regression were performed in the training cohort.

Evaluation of Ferroptosis Activity in the Different Risk Groups

To quantify the activity of ferroptosis, an established indicator named the ferroptosis potential index (FPI) was obtained from a previous study and calculated for each sample based on gene expression data (Liu et al., 2020). During computation, key ferroptosis regulator genes were first divided into positive and negative groups according to their potential function. Then, the enrichment score of these two groups was calculated separately using the R package GSVA (Hänzelmann et al., 2013). The FPI was defined as the enrichment score of the positive group minus the negative group and was considered a reliable indicator of ferroptosis level.

Validation of the Ferroptosis-Related Prognostic Signature in Multiple Validation Cohorts

To further validate the prognostic value of the FRP signature, the risk score was first calculated for each sample based on the formula derived from the training cohort. Then, samples were divided into low-risk and high-risk groups according to the median risk score in each cohort. Kaplan–Meier curves were generated, and univariate Cox regression was performed, and the concordance index (C-index) was applied to evaluate the performance of the FRP signature.

Exploring the Implications of the Ferroptosis-Related Prognostic Signature on Immunotherapeutic and Chemotherapeutic Responses

To infer the potential benefit of immunotherapy for each patient, the number of neoantigens and the PD-L1 protein expression of patients in the TCGA-LUAD cohort were extracted from The

[2]http://www.zhounan.org/ferrdb

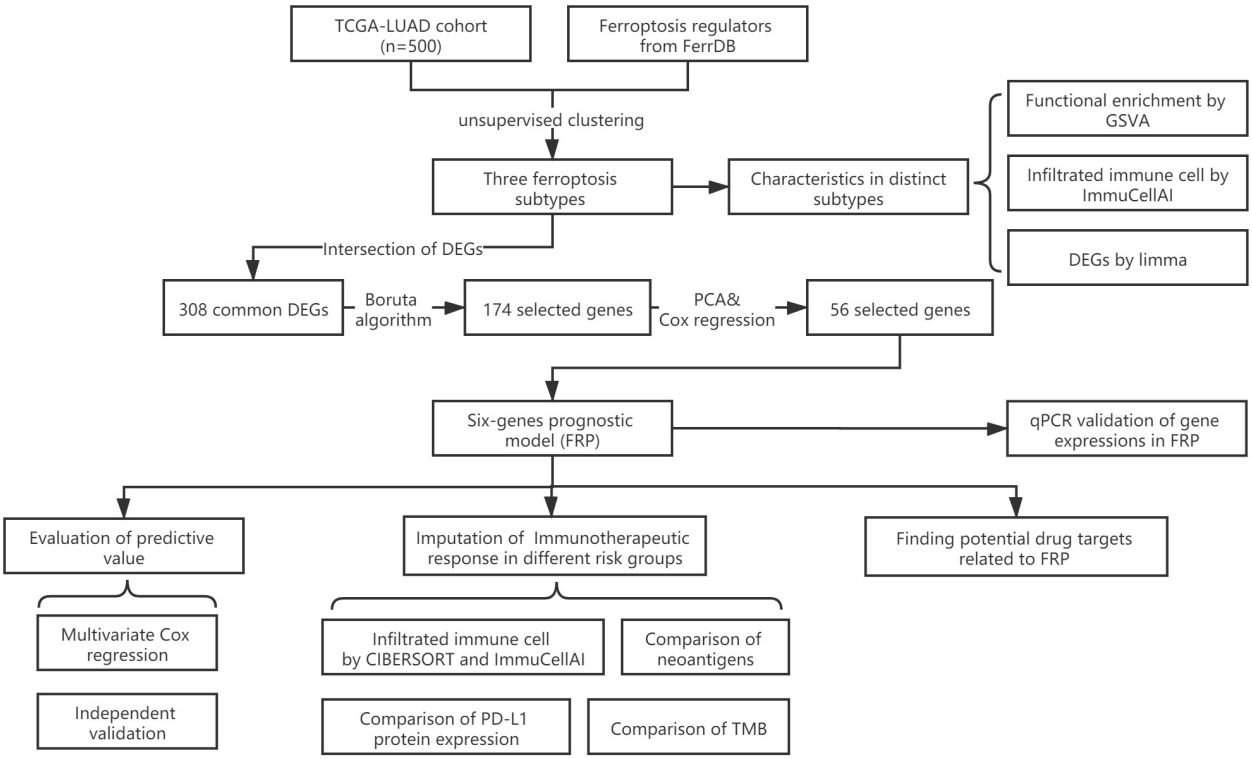


FIGURE 1 | Flowchart of the present study.

Cancer Immunome Atlas[3] and cBioPortal for Cancer Genomics[4], respectively (Gao et al., 2013; Charoentong et al., 2017). To determine the chemotherapeutic implications, experimental data for cancer cell drug sensitivity were obtained from the Genomics of Drug Sensitivity in Cancer (GDSC), and the whole prediction process was conducted with the R package pRRophetic (Geeleher et al., 2014). "allSoldTumours" was selected for the tissue type parameter, and the remaining parameters were set to the default values.

Cell Lines and Cell Culture Conditions

BEAS-2B and HCC827 cell lines were purchased from ATCC. BEAS-2B cells were cultured in BEBM basal medium supplemented with 10% fetal bovine serum (FBS), 1,000 U ml^{-1} penicillin and 100 μg ml^{-1} streptomycin. HCC827 cells were maintained in RPMI 1640 supplemented with 10% FBS, 1,000 U ml^{-1} penicillin and 100 μg ml^{-1} streptomycin. The cells were cultured with 5% CO_2 at 37°C in a humidified incubator.

RNA Extraction and Quantitative RT-PCR Analysis

Total RNA was extracted from cells using TRIzol reagent (Invitrogen, United States) according to the manufacturer's instructions. Equal amounts of RNA samples were used for cDNA synthesis with a RevertAid First Strand cDNA synthesis kit (Thermo). SYBR Green-based qRT-PCR on a 7900HT fast real-time PCR system (Applied Biosystems/Life Technologies, Waltham, United States) was performed. The relative mRNA expression levels were calculated by the $2^{-\Delta\Delta Ct}$ method with normalization to *ACTB*; the PCR primers are listed in **Supplementary Table 1**.

Statistical Analysis

All bioinformatic analyses were conducted in R software (Version 3.6.3). Wilcoxon test and Kruskal–Wallis test were performed to statistically test the difference in continuous data for defined subtypes. The Spearman coefficient was used to evaluate the association between variables. Survival analyses, including Kaplan–Meier curves and Cox regression, were performed using the R package survival. The C-index was calculated by R package survcomp. A *P*-value less than 0.05 was considered statistically significant.

RESULTS

Identification of Distinct Ferroptosis Subtypes on the Basis of Ferroptosis Regulators

The flow chart of this study is shown in **Figure 1**. A comprehensive list of genes including 108 that encode

[3]https://tcia.at/home
[4]http://www.cbioportal.org/

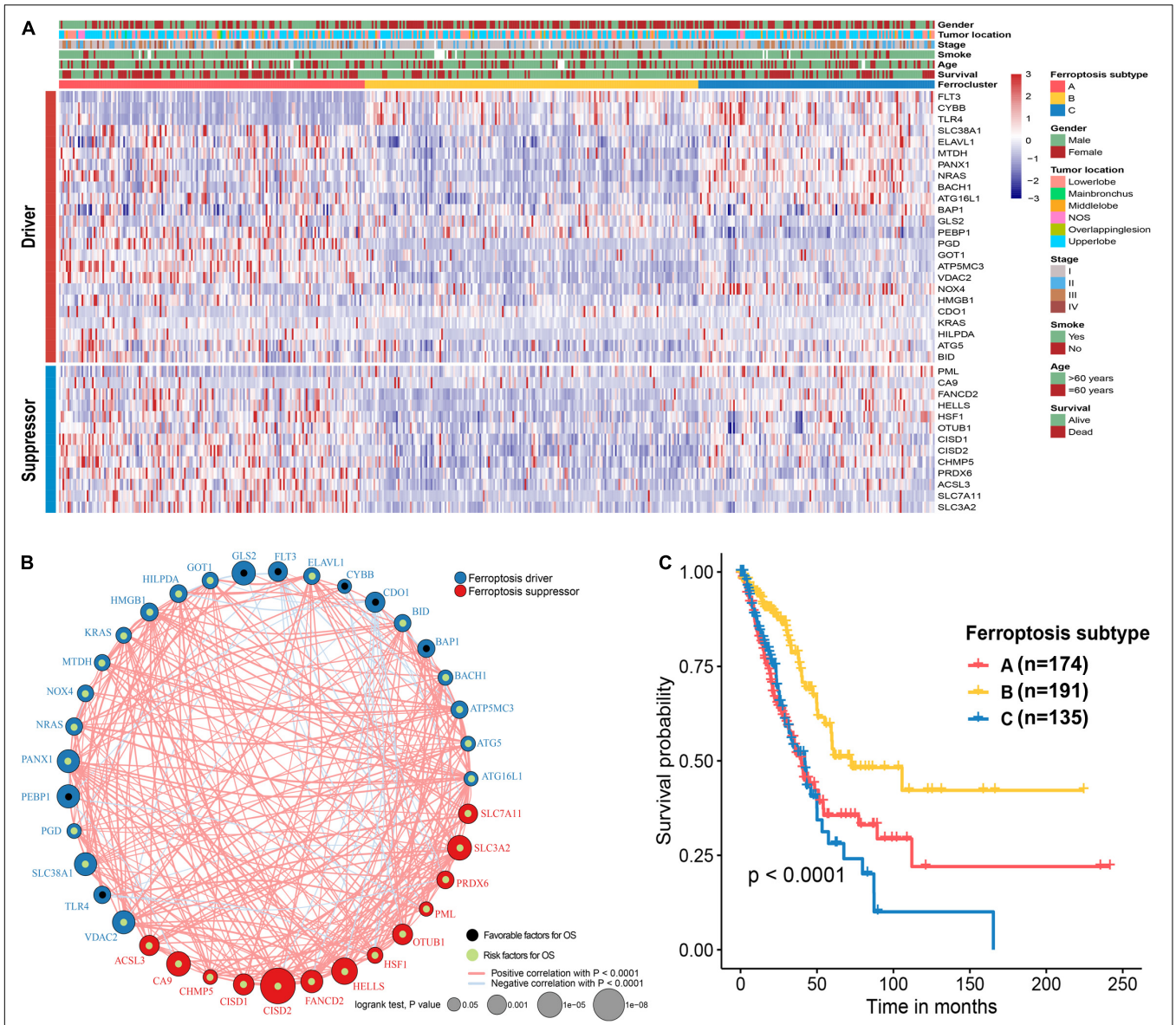


FIGURE 2 | Unsupervised clustering of ferroptosis regulators and characteristics of ferroptosis subtypes. **(A)** Heat map showing the expression patterns of 37 ferroptosis regulator genes in the TCGA-LUAD cohort. Sex, tumor location, stage, smoking history, age, and survival status were the annotations. **(B)** Interactions of 37 ferroptosis regulator genes. The size of each circle indicates the survival significance of each gene. Ferroptosis drivers are shown in blue, and ferroptosis suppressors are shown in red. Favorable prognostic factors are indicated in black, and unfavorable prognostic factors are indicated in green. The strength of the relationship between two genes is symbolized by the thickness of the line. **(C)** Kaplan–Meier curves of the overall survival in the TCGA-LUAD cohort on the basis of the ferroptosis subtypes.

ferroptosis drivers and 69 that encode suppressors was extracted from the FerrDb. After removing multirole regulators that would be repeatedly counted, a total of 161 genes remained for screening. Univariate Cox regression was performed, and 37 genes with prognostic significance ($p < 0.05$) were chosen from the list. Based on the expression level of these genes, three subtypes were identified in the TCGA-LUAD cohort. The clinical information and gene expression patterns are displayed in **Figure 1**. In subtype A, most of the genes, including both drivers and suppressors, were highly expressed, and in contrast, a substantial portion of ferroptosis regulators was expressed at low levels in subtype B (**Figure 2A**). The expression pattern in subtype C was somewhere in between the other two subtypes. An association analysis suggested that there was a close connection between the expression levels of the 37 ferroptosis regulators (**Figure 2B**). Of 24 ferroptosis drivers, nearly one-third (7/24) of the genes was revealed to be favorable prognostic factors, and the remaining genes (17/24) were risk factors for overall survival (OS; **Figure 2B**). All 13 ferroptosis suppressors were unfavorable prognostic factors (**Figure 2B**). These results suggested that

TABLE 1 | Correlations between ferroptosis subtypes and clinicopathological parameters in TCGA-LUAD cohort.

Category	Ferroptosis subtypes		
	Subtype A (*n* = 163)	**Subtype B (*n* = 174)**	**Subtype C (*n* = 127)**
Age [mean (*SD*)]	63.3 (10.3)	67.0 (9.6)	65.3 (9.6)
Gender			
Male	97 (59.5%)	58 (33.3%)	56 (44.1%)
Female	66 (40.5%)	116 (66.7%)	71 (55.9%)
Smoking			
Yes	147 (90.2%)	142 (81.6%)	110 (86.6%)
No	16(9.8%)	32(18.4%)	17 (13.4%)
T stage			
T1	38 (23.3%)	87 (50.0%)	36 (28.3%)
T2	96 (58.9%)	71 (40.8%)	77 (60.6%)
T3	22 (13.5%)	13 (7.5%)	9 (7.1%)
T4	7 (4.3%)	3 (1.7%)	5 (3.9%)
N stage			
N0	99 (60.7%)	140 (80.5%)	74 (58.3%)
N1	37 (22.7%)	20 (11.5%)	27 (21.3%)
N2	27 (16.6%)	14 (8.0%)	24 (18.9%)
N3	0 (0.0%)	0 (0.0%)	2 (1.6%)
M stage			
M0	150 (92.0%)	171 (98.3%)	122 (96.1%)
M1	13 (8.0%)	3 (1.7%)	5 (3.9%)
TNM stage			
I	75 (46.0%)	121 (69.5%)	60 (47.2%)
II	44 (27.0%)	34 (19.5%)	34 (26.8%)
III	31 (19.0%)	16 (9.2%)	28 (22.0%)
IV	13 (8.0%)	3 (1.7%)	5 (3.9%)
Status			
Survival	87 (53.4%)	137 (78.7%)	75 (59.1%)
Death	76 (46.6%)	37 (21.3%)	52 (40.9%)
Tumor location			
Lower lobe	53 (32.5%)	65 (37.4%)	36 (28.3%)
Middle lobe	8 (4.9%)	7 (4.0%)	5 (3.9%)
Upper lobe	90 (55.2%)	98 (56.3%)	84 (66.1%)
Main bronchus	2 (1.2%)	0 (0.0%)	0 (0.0%)
Overlapping lesion	2 (1.2%)	2 (1.1%)	0 (0.0%)
Lung NOS	8 (4.9%)	2 (1.1%)	2 (1.6%)

inhibition of ferroptosis suppressors may improve the survival of LUAD patients. In addition, a significant survival difference ($p < 0.0001$) was observed among the three ferroptosis subtypes (**Figure 2C**). Patients with subtype B had the best prognosis compared with patients with the other subtypes, while those with subtype C had the worst prognosis (**Figure 2C**). The clinicopathological characteristics of the three subtypes were summarized in **Table 1**.

Functional Enrichment and Immune Microenvironment Assessment of the Ferroptosis Subtypes

To further explore the biological pathway underlying each distinct ferroptosis subtype, GSVA was performed to determine pathway enrichment (Subtype A vs. Subtype B; Subtype B vs. Subtype C). As shown in **Figure 3A**, ferroptosis subtype A was significantly enriched in genes involved with RNA polymerase, pyrimidine metabolism, purine metabolism, cell cycle, and mismatch repair. Ferroptosis subtype B was markedly enriched in genes related to fatty acid and drug metabolism, including bile acid biosynthesis, alpha-linolenic acid metabolism, arachidonic acid metabolism, linoleic acid metabolism, and cytochrome P450 drug metabolism (**Figure 3B**). In addition, subtype B genes are also involved in cell adhesion and molecules (CAMS), intestinal immune network, and JAK-STAT signaling pathway (**Figure 3A**). Subtype C genes were predominantly related to pathogenic *Escherichia coli* infection, the p53 signaling pathway, ubiquitin-mediated proteolysis, and basal transcription factors (**Figure 3B**). Subsequent evaluation of infiltrated immune cells revealed conspicuous differences in the immune microenvironment among subtypes (**Figure 4A**). Particularly, as the subtype

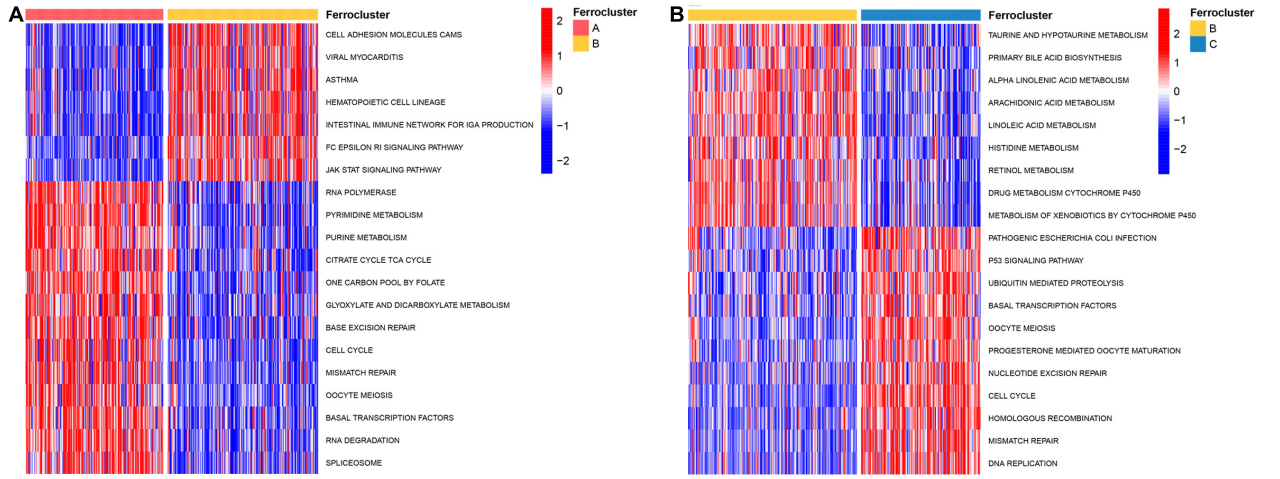


FIGURE 3 | Biological characteristics of different ferroptosis subtypes. GSVA analysis revealed highly enriched KEGG pathways in distinct ferroptosis subtypes. The 20 most enriched pathways as identified in each comparison are visualized by heat map. **(A)** Ferroptosis subtype A vs. subtype B. **(B)** Ferroptosis subtype B vs. subtype C.

with the best prognosis, the ferroptosis subtype B genes were prominently enriched in Tfh, NKT, NK, $CD4^+$ T, and $CD8^+$ T cells (**Figure 4A**).

Construction of an Ferroptosis-Related Prognostic Signature for the TCGA-Lung Adenocarcinomas Cohort

To construct a prognostic model related to the defined ferroptosis subtypes, the intersecting DEGs identified in the pairwise comparisons were identified, and a total of 308 genes were selected as initial features (**Figure 4B**). After the dimension reduction by the Boruta algorithm, PCA was performed on the residual 174 genes, and the results indicated that the expression of these genes had great discriminatory power for the different subtypes (**Figure 4C**). Thus, these 174 genes constituted a representative collection of genes in the defined ferroptosis subtypes. These demonstrated that the dimensionality reduction process successfully removed unuseful features (n = 134) and retained relevant genes (n = 174). By applying univariate Cox regression, 56 genes with prognostic significance ($p < 0.05$) were retained for repeated LASSO Cox regression to generate the best model (**Supplementary Table 2**). A total of 1,000 iterations were conducted, and the distribution of model occurrences is shown in **Figure 4D**. As presented in **Figure 4D**, a prognostic signature with 6 genes (*EIF5A*, *CACYBP*, *CYCS*, *ANLN*, *ARNTL2*, and *PPM1M*) had the highest frequencies, with 881, which greatly exceeded those of the other four models and was, therefore, chosen as the FRP signature. The risk score of the FRP signature was calculated as follows: risk score = $(0.0413 \times \mathrm{EXP}_{EIF5A}) + (0.0068 \times \mathrm{EXP}_{CACYBP}) + (0.0254 \times \mathrm{EXP}_{CYCS}) + (0.1300 \times \mathrm{EXP}_{ANLN}) + (0.1503 \times \mathrm{EXP}_{ARNTL2}) - (0.0848 \times \mathrm{EXP}_{PPM1M})$. After calculating the risk score for each sample in the training cohort, samples were stratified into low-risk and high-risk groups based on the median risk score, and the survival analysis demonstrated that patients in the high-risk group had a significantly shorter overall survival time (log-rank $p < 0.001$, HR = 2.33, 95% CI = 1.72–3.17) than patients in the low-risk group (**Figure 4E**). The FPI was applied to quantify the activity of ferroptosis in each group, and the results suggested that the FPI in the high-risk group was significantly ($p = 0.044$) higher than that in the low-risk group (**Figure 4F**), which indicated a higher ferroptosis activity in samples that with higher risk. The association of attributes including ferroptosis subtype, risk group, FPI, and survival status of individual patients is presented in an alluvial diagram (**Figure 4G**).

Identification of the Ferroptosis-Related Prognostic Signature as an Independent Prognostic Factor

To decrease the effects of other confounding factors on the relationship between the FRP risk score and OS time, multivariate Cox regression analysis including age, sex, stage, and smoking history was conducted with the TCGA-LUAD cohort. The results proved that age, stage, and risk score ($p < 0.001$, HR = 2.17, 95% CI = 1.54–3.04) were independent prognostic factors of the OS time for LUAD patients (**Figure 5**).

Validation and Evaluation of the Predictive Performance of the Ferroptosis-Related Prognostic Signature in Five Validation Cohorts

The basic information related to the five validation cohorts is described in **Table 2**. Briefly, a total of 668 cases made up our validation datasets. Consistent and significant differences were observed in validation cohort 1 (log-rank $p = 0.001$, HR = 2.62, 95% CI = 1.44–4.76), validation cohort 2 (log-rank $p = 0.002$, HR = 2.62, 95% CI = 1.40–4.89), validation cohort 3 (log-rank $p = 0.005$, HR = 2.75, 95% CI = 1.32–5.74), validation

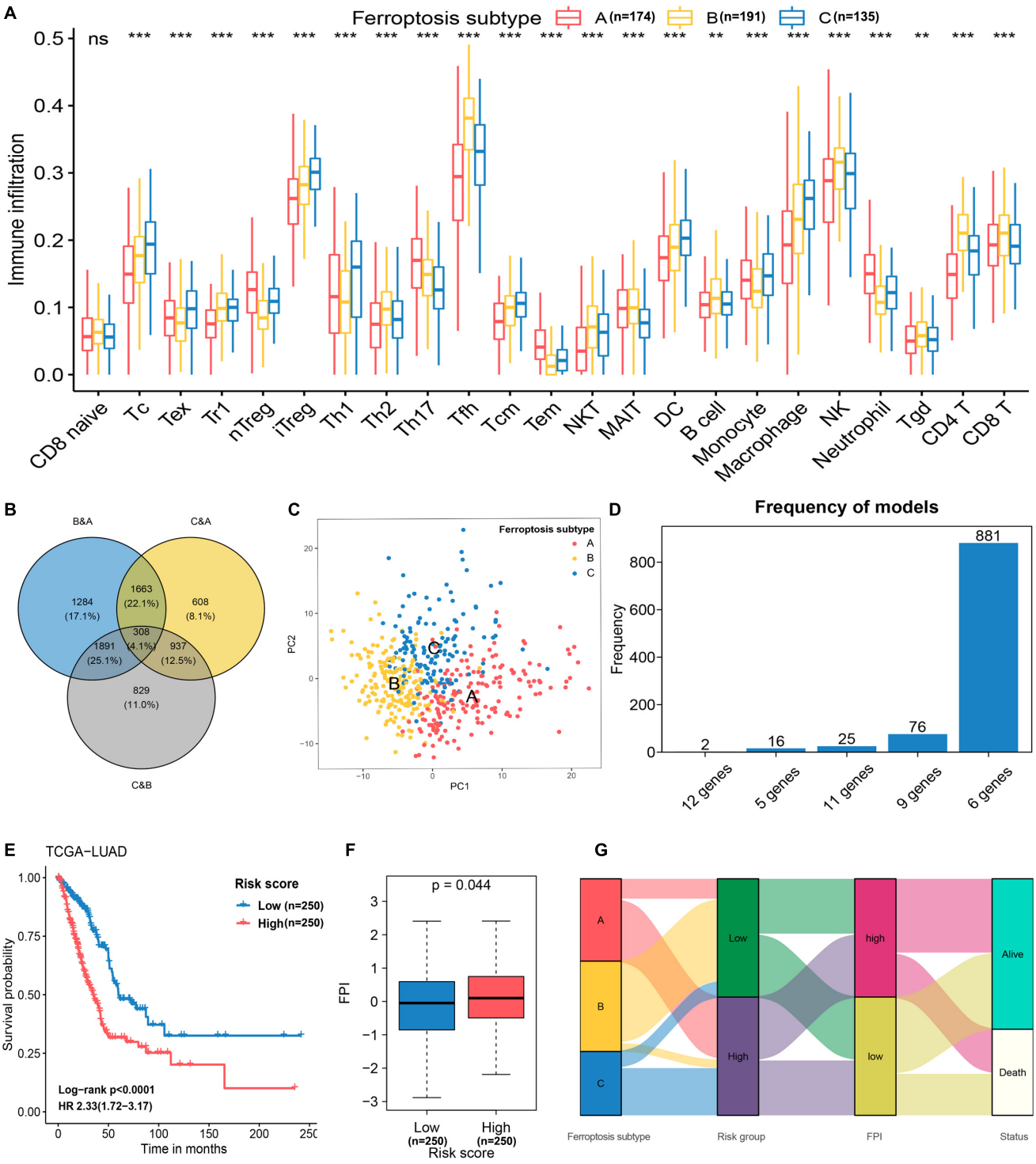


FIGURE 4 | Feature selection and signature construction process. **(A)** Assessing the abundance of infiltrated immune cells in distinct ferroptosis subtypes. $^{**}p < 0.01$, $^{***}p < 0.001$; ns, non-significant. **(B)** The number of unique and shared DEGs from different comparisons as revealed in a Venn diagram. **(C)** Principal component analysis of 174 genes selected with the Boruta feature selection algorithm. These genes can effectively discriminate the different ferroptosis subtypes. **(D)** The distribution of model occurrences after 1,000 iterations of LASSO regression. The six-gene prognostic signature was selected because it had a higher frequency than the other four models. **(E)** Kaplan–Meier curves and univariate Cox regression revealed that the FRP risk score was markedly related to overall survival in the TCGA-LUAD cohort. **(F)** Differences in FPI scores between the high-risk and low-risk groups. **(G)** Alluvial diagram showing the relations of ferroptosis subtype, FRP risk group, FPI, and status.

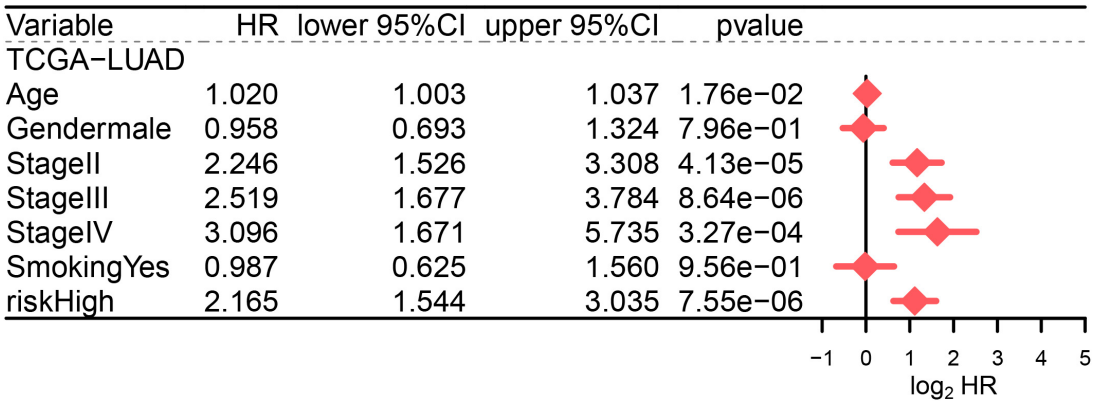


FIGURE 5 | Multivariate Cox regression analysis including age, sex, stage, smoking history, and FRP risk score in the TCGA-LUAD cohort.

cohort 4 (log-rank $p = 0.014$, HR = 2.39, 95% CI = 1.17–4.92), and validation cohort 5 (log-rank $p < 0.001$, HR = 2.59, 95% CI = 1.57–4.26; **Figures 6A–E**). In addition, the C-index was calculated to assess predictive performance and the C-index was 0.663 for the training cohort, 0.641 for validation cohort 1, 0.694 for validation cohort 2, 0.666 for validation cohort 3, 0.647 for validation cohort 4, and 0.685 for validation cohort 5 (**Figure 6F**). Statistical significance ($p < 0.05$) was observed between the actual value of the C-index and the null hypothesis (C-index = 0.5) in all datasets (**Figure 6F**).

Depiction of the Immune Microenvironment and Implications of the Immunotherapeutic Response Between Risk Groups

The abundance of infiltrated immune cells was estimated using CIBERSORT and ImmuCellAI for each sample, and a heat map was generated to illustrate the association between immune cell infiltration and the FRP risk score (**Figure 7A**). The top five immune cell types with significant changes in abundance between risk groups were $CD4^+$ T cells ($p < 0.001$), nTregs ($p < 0.001$), Tfhs ($p < 0.001$), Tems ($p < 0.001$), and monocytes ($p < 0.001$). The detailed results of these differential analyses are presented in **Supplementary Table 3**. Furthermore, an association analysis of the FRP risk score and potential immunotherapeutic biomarkers, including TMB, PD-L1 protein, and neoantigen, was performed. The results suggested that TMB in the high-risk group was significantly ($p < 0.001$) higher than that in the low-risk group (**Figure 7B**). A correlation analysis revealed that the FRP risk score was significantly associated ($R = 0.365$, $p < 0.0001$) with TMB (**Figure 7C**). In addition, the expression level of the PD-L1 protein ($p < 0.001$) and the numbers of subclonal neoantigens ($p = 0.027$), clonal neoantigens ($p < 0.001$), and total neoantigens ($p < 0.001$) were all significantly higher in the high-risk group than in the low-risk group (**Figures 7D–G**). These results implied that patients in the high-risk group may potentially benefit from immunotherapy.

TABLE 2 | Description of the training and validation cohorts used in this study.

Dataset	Source	Accession	Platform	Number of samples
Training cohort	TCGA	TCGA-LUAD	Illumina HiSeq	500
Validation cohort 1	GEO	GSE13213	GPL6480	117
Validation cohort 2	GEO	GSE30219	GPL570	85
Validation cohort 3	GEO	GSE31210	GPL570	226
Validation cohort 4	GEO	GSE3141	GPL570	58
Validation cohort 5	GEO	GSE41271	GPL6884	182

Extrapolating the Relationship Between the Ferroptosis-Related Prognostic Signature and Chemotherapy Response

Considering that chemotherapy constitutes an important component of LAUD treatment, drug susceptibility data including 138 compounds from the GDSC were applied to the training group to predict drug sensitivities for each sample in the TCGA-LUAD cohort. Based on the significance level, the top 10 compounds that might benefit patients in the high-risk group were identified: A.443654, obatoclax mesylate, epothilone B, GW843682X, BI.2536, TW.37, thapsigargin, ZM.447439, PF.562271, and *S*-Trityl-L-cysteine (**Figures 8A–J**). These results provide a novel perspective for exploring new drug candidates. A detailed list showing the significance value for all 138 compounds is available in **Supplementary Table 4**.

Quantification of Gene Expression in Normal and Lung Adenocarcinoma Cell Lines

To further investigate the expression pattern *in vitro*, six genes that constituted the FRP signature were quantified by qPCR and those in normal bronchial epithelial cells (BEAS-2B cells) and LUAD cells (HCC827 cells) were compared. Of these genes, *EIF5A* and *ARNTL2* were significantly overexpressed in tumor cells ($p < 0.05$; **Figures 9A,E**). In contrast, the expression levels of *CYCS*, *ANLN*, and *PPM1M* were significantly elevated in normal lung cells ($p < 0.05$; **Figures 9C,D,F**). No significant

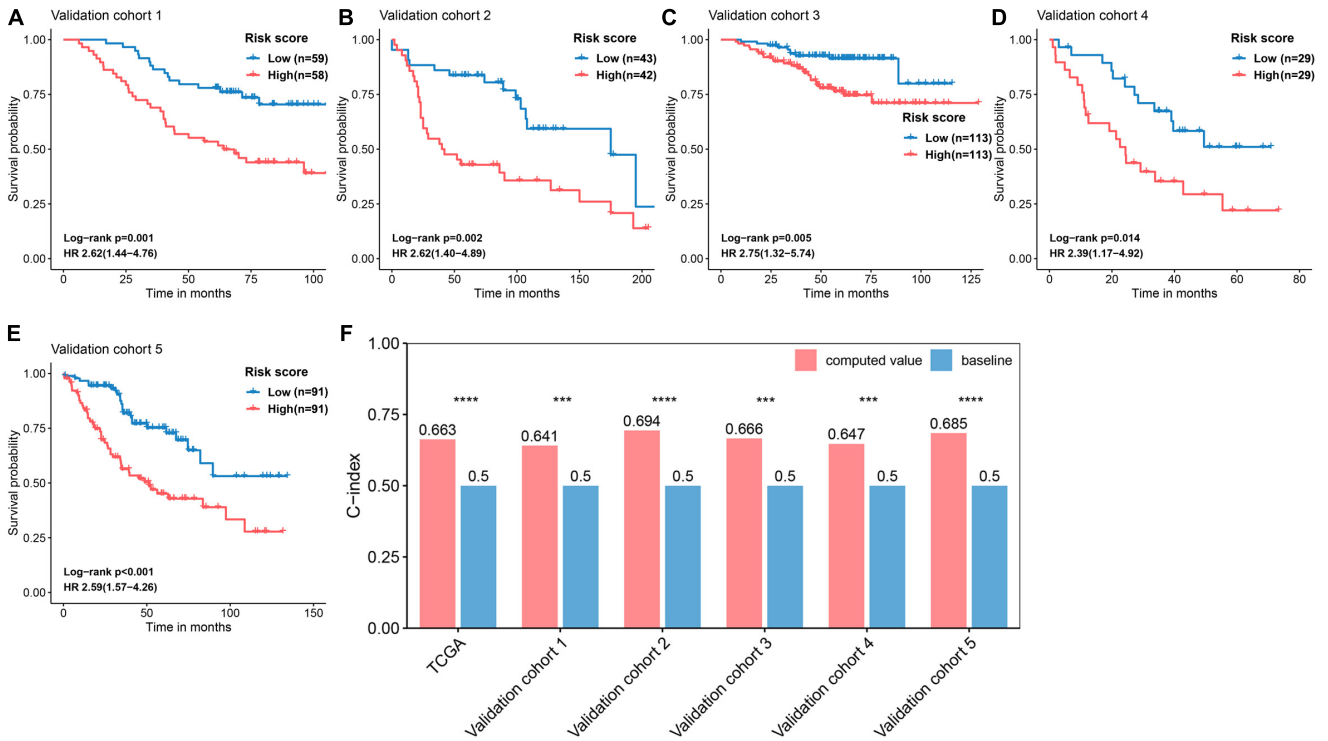


FIGURE 6 | Validation and evaluation of the performance of the FRP risk score. Kaplan–Meier curves and univariate Cox regression were performed with respect to the different risk groups in **(A)** validation cohort 1 (GSE13213), **(B)** validation cohort 2 (GSE30219), **(C)** validation cohort 3 (GSE31210), **(D)** validation cohort 4 (GSE3141), and **(E)** validation cohort 5 (GSE41271). **(F)** Assessing the predictive performance of the FRS by the C-index. The C-index was calculated for each cohort and compared to the null hypothesis. ***$p < 0.001$, ****$p < 0.0001$.

difference in *CACYBP* expression was observed between the two cell lines (**Figure 9B**).

DISCUSSION

Increasing studies about the association between ferroptosis and the TIME have been performed since the term ferroptosis was coined in 2012, and most of these studies demonstrated that ferroptosis might contribute to the immune surveillance and cytotoxic effects of $CD8^+$ T cells (Wang et al., 2019), while others reported that ferroptosis facilitated the formation of a tumor-promoting immune microenvironment (Wen et al., 2019). As the roles of ferroptosis in shaping the tumor suppressive or promoting microenvironment remain elusive and most studies focus on a single ferroptotic regulator and a single cell type populating the TIME, the overall characteristics of the TIME linked to a specific ferroptotic signature in LUAD need to be elaborated.

Here, we identified three distinct ferroptosis subtypes based on the expression level of 37 ferroptosis driver and suppressor genes with prognostic significance ($p < 0.05$) in the TCGA-LUAD cohort. Patients with ferroptosis subtype B, in which most of the ferroptosis regulators were expressed at low levels, had the best prognosis. In addition, we evaluated the infiltrated immune cells in these three ferroptosis subtypes and found that subtype B was significantly enriched in Tfh, NKT, NK, $CD4^+$ T, and $CD8^+$ T cells, compared with the other two subtypes, suggesting that the TIME in association with ferroptosis subtype B was enriched in tumor-suppressive immune cells, which can partially explain the reason that patients with subtype B have the best prognosis.

Then, we constructed an FRP signature consisting of 6 genes (*EIF5A*, *CACYBP*, *CYCS*, *ANLN*, *ARNTL2*, and *PPM1M*) based on the defined ferroptosis subtypes and identified the FRP signature as an independent prognostic factor ($p < 0.001$, HR = 2.17, 95% CI = 1.54–3.04) in the training TCGA-LUAD cohort and five validation cohorts. Patients in the high-risk group had a significantly shorter overall survival time than patients in the low-risk group. In addition, we applied the FPI to quantify the overall activity of ferroptosis in each group and unsurprisingly found that the FPI in the low-risk group patients, who had a better prognosis, was significantly lower than that in the high-risk group, which was in accordance with the abovementioned results indicating that most of the ferroptosis regulators are expressed at low levels in patients with ferroptosis subtype B, who have the best prognosis, and a previous study showing that a high FPI predicts a poor prognosis (Liu et al., 2020).

Among the six genes in the FRP signature, *ANLN* and *ARNTL2* were mechanistically explored in LUAD in previous studies. *ANLN*, a homolog of anillin whose stability and nuclear localization are regulated by the PI3K/AKT pathway, can bind to and activate *RHOA*, thus promoting cancer

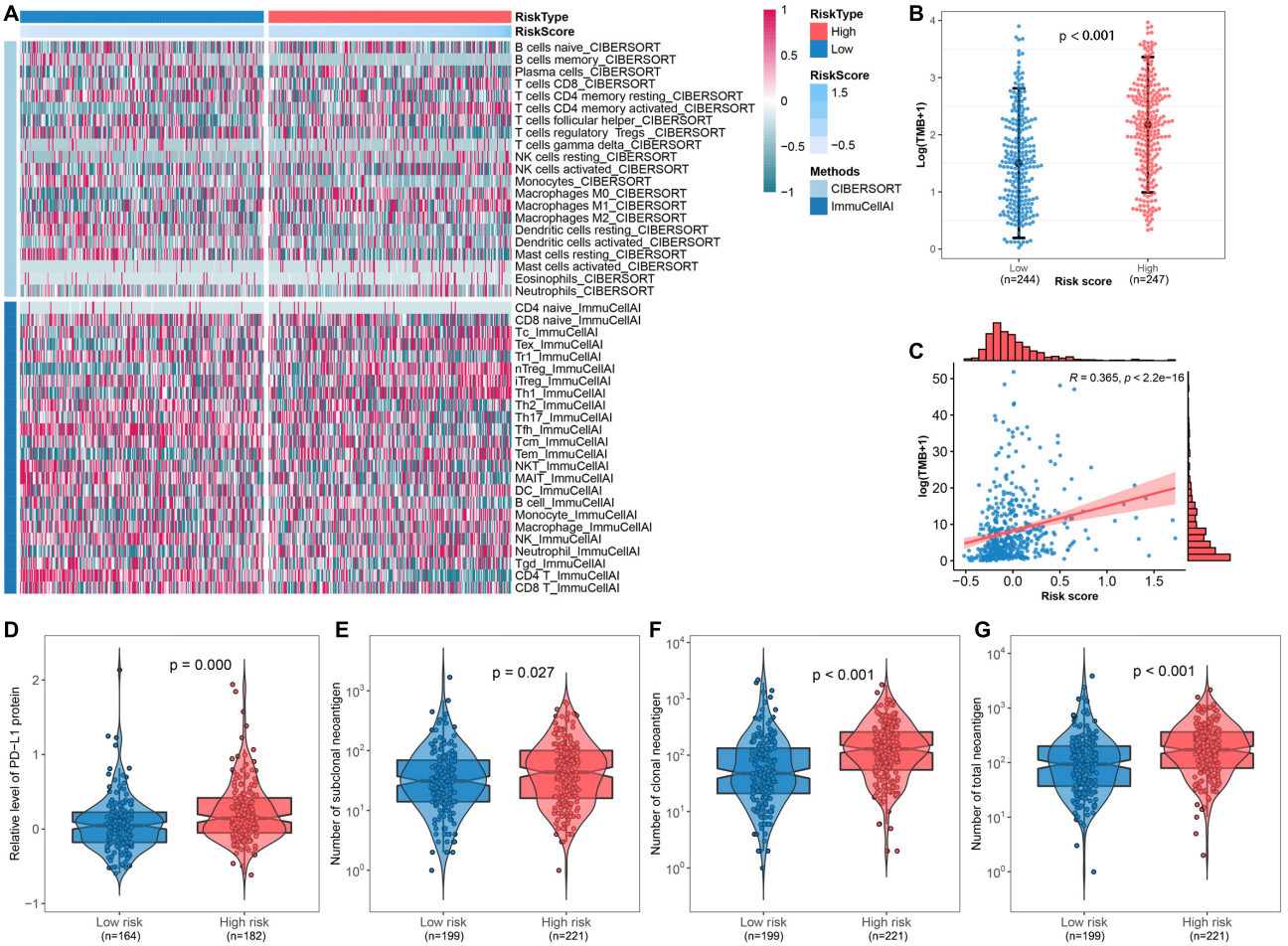


FIGURE 7 | Depiction of the immune microenvironment and implications of the immunotherapeutic response. **(A)** Heat map showing the immune microenvironment of the TCGA-LUAD cohort as estimated by CIBERSORT and ImmuCellAI. **(B)** Comparison of TMB between high-risk and low-risk groups. **(C)** Scatter plot presenting the correlation between TMB and risk score. Comparisons of **(D)** the expression of PD-L1 protein, **(E)** number of subclonal neoantigens, **(F)** number of clonal neoantigens, and **(G)** number of total neoantigens between different risk groups.

growth and development (Suzuki et al., 2005). In addition, Xu et al. (2019) reported that *ANLN* is involved in the metastasis of LUAD by promoting epithelial mesenchymal transformation of tumor cells, and Long et al. (2018) pointed out that higher expression of *ANLN* predicted a relatively poor prognosis. *ARNTL2* is a transcription factor that is involved in promoting LUAD metastasis by orchestrating the expression of a complex prometastatic secretome (Brady et al., 2016). With respect to the other genes, *EIF5A*, a eukaryotic translation initiation factor, enhances the expression of a series of proliferation-associated proteins essential for cancer cells and is upregulated in several cancers, including pancreatic ductal adenocarcinoma, glioblastoma, leukemia, liver, colon, lung, cervical, and ovarian cancer (Mathews and Hershey, 2015; Strnadel et al., 2017). *CACYBP*, a calcyclin-binding protein, interacts with different partner proteins and, through these interactions, is involved in various cellular processes, such as proliferation, differentiation, cytoskeletal reorganization, and protein ubiquitination (Lian et al., 2019). For example, the Siah1-interacting protein (SIP)/CACYBP complex participates in the ubiquitination and proteasomal degradation of β-catenin, an oncogenic protein that regulates gene transcription (Jiang et al., 2019). Previous studies have reported that *CACYBP* contributes to the development of a wide range of human malignancies, including gastric, pancreatic, colon, breast, brain, and renal cancer (Topolska-Wos et al., 2016). *CYCS*, also known as cytochrome c, is a small soluble heme protein that shuttles electrons during the oxidative phosphorylation process and engages in the intrinsic apoptotic pathway under proapoptotic conditions (Ow et al., 2008). *PPM1M*, protein phosphatase 1M, is a manganese/magnesium ion-dependent serine/threonine phosphatase (Kamada et al., 2020). It was reported that *PPM1M* inhibits IL-1-induced activation of NF-κB by dephosphorylating IKK β (Henmi et al., 2009).

We also determined the abundance of infiltrated immune cell types in different risk groups using CIBERSORT and ImmuCellAI. The low-risk group with a better prognosis was enriched with $CD4^+$ T cells and Tfhs, which was consistent

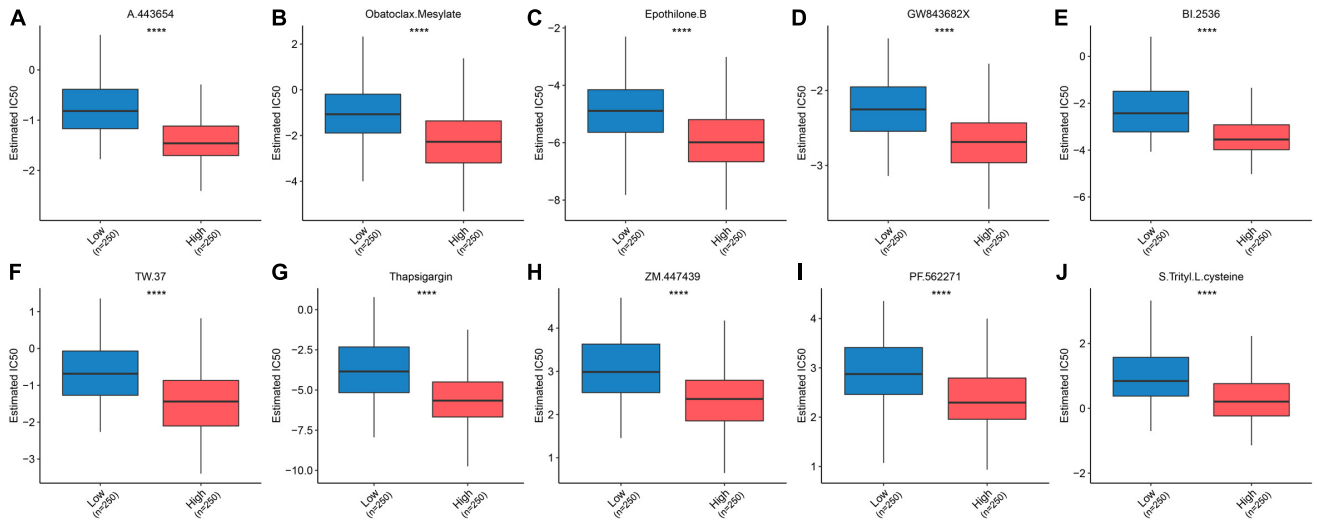


FIGURE 8 | Box plots based on the estimated IC50 of the 10 most significant compounds: **(A)** A.443654, **(B)** obatoclax mesylate, **(C)** epothilone B, **(D)** GW843682X, **(E)** BI.2536, **(F)** TW.37, **(G)** thapsigargin, **(H)** ZM.447439, **(I)** PF.562271, and **(J)** *S*-Trityl-L-cysteine. ****$p < 0.0001$.

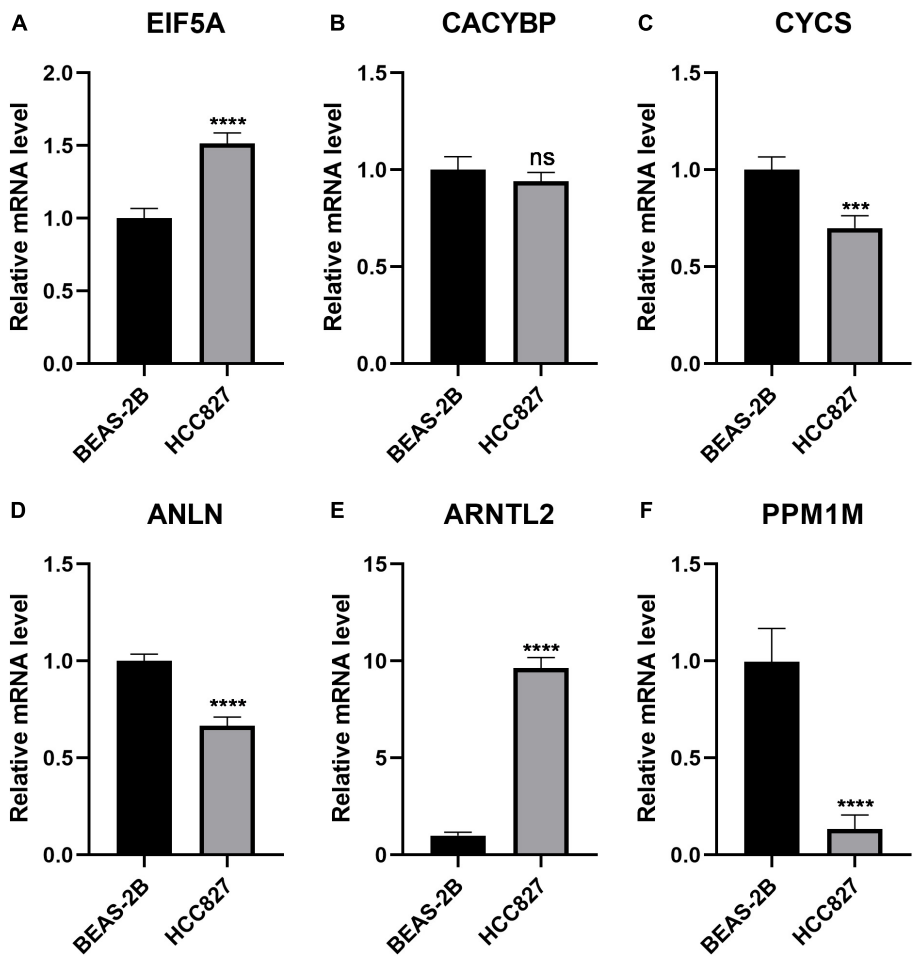


FIGURE 9 | Exploration of the expression of the six genes in normal and LUAD cell lines. Quantification and comparison of mRNA expression in the FRP signature as determined by qPCR: **(A)** EIF5A, **(B)** CACYBP, **(C)** CYCS, **(D)** ANLN, **(E)** ARNTL2, and **(F)** PPM1M. ***$p < 0.001$, ****$p < 0.0001$; ns, non-significant.

with the infiltrating cell types associated with ferroptosis subtype B, and the abundance of nTregs, Tems, and monocytes was relatively low. Follicular helper T (Tfh) cells, emerging as a fifth helper T subset, interact with B cells, driving their differentiation into long-lived antibody-secreting plasma cells or memory B cells (King et al., 2008). In breast cancers, Tfh cells recruit immune cells to the TIME and promote tertiary lymphoid structure formation, in which effective antitumor immune responses can be generated and maintained (Gu-Trantien et al., 2013). Certain TIMEs were affected by several factors including exercise, age, diet, adiposity, the microbiome, sex, and tumor-dependent effects. Comprehensive consideration of the above factors during research will be more conducive to dissect the formation mechanism of TIME and lay the foundation for developing intervention measures (Binnewies et al., 2018). Our results suggested that patients in the low-risk group may have a better prognosis partially because of the infiltration of Tfh cells, which facilitate the formation of an antitumor TIME. Furthermore, we found that potential immunotherapeutic biomarkers, including TMB, the PD-L1 protein, and neoantigens, were elevated in the high-risk group, suggesting that patients in the high-risk group may benefit the most from immunotherapy. Our results revealed distinct immune and mutational profiles between different risk groups. Thus, we recommend to include ferroptosis-related genes or ferroptosis signature in future immunotherapy cohort for biomarker study as compared with previous ferroptosis associated prognostic model (Wang et al., 2021), which was directly constructed utilizing LASSO Cox regression model in TCGA-LUAD dataset and validated in one GEO dataset. By contrast, we firstly used ferroptosis regulators to identify three ferroptosis subtypes with prognostic significance and constructed a FRP signature by repeated LASSO Cox using 308 subtype-distinguishing genes. We also validated the signature in five GEO datasets and evaluated its implications of the immunotherapeutic response and chemotherapy, which was of great importance for further clinical application.

Inevitably, several limitations should be acknowledged in our study: (1) A total of 6 cohorts (a TCGA-LUAD cohort and 5 GEO cohorts) used for the construction and further validation of the FRP signature were downloaded from public databases retrospectively. A study with a prospective clinical trial cohort is needed to verify the clinical application value of our FRP signature. (2) Ferroptosis regulators obtained from the FerrDb website were manually extracted from previous literature reports, and therefore, some unselected crucial ferroptosis-associated genes may be missing in our study. In addition, ferroptosis regulators have been reported to regulate ferroptosis in human cells but not specifically in LUAD. (3) The FRP signature in our study was constructed based on DEGs identified in three ferroptosis subtypes, and the six genes in the FRP signature have not yet been identified to be involved in ferroptosis. Further studies are needed to determine the association and mechanisms between the six FRP genes and ferroptosis in LUAD. (4) Algorithmic calculation of immune cell abundance may not be as accurate as immunofluorescence, and further verification is required.

CONCLUSION

A robust ferroptosis-related prognostic signature was developed and validated with five independent datasets. Different tumor immune microenvironments were observed between the low-risk and high-risk groups. This signature may serve as a potential prognostic biomarker for LUAD in the future.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: the datasets collected in the current study are available in the TCGA (https://portal.gdc.cancer.gov/) and GEO repository (https://www.ncbi.nlm.nih.gov/geo/).

AUTHOR CONTRIBUTIONS

YG and JH directed and designed the study. ZY, JW, RL, and HT extracted the data. SS and YY conducted the data analysis and experiments. SS wrote the manuscript. FT and QX reviewed and edited the manuscript. All authors read and approved the manuscript.

FUNDING

This study was supported by the National Key R&D Program of China (2018YFC1312100 and 2017YFC1311000), CAMS Initiative for Innovative Medicine (2017-I2M-1-005, 2017-I2M-2-003, and 2019-I2M-2-002), Beijing Municipal Science and Technology Commission (Z191100006619116), Innovation Team Development Project of Ministry of Education (IRT_17R10), and Peking Union Medical College Youth Fund (3332018144).

ACKNOWLEDGMENTS

The authors thank all staff in the Department of Thoracic Surgery for their support during the study. The results shown here are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.743724/full#supplementary-material

Supplementary Table 1 | Quantitative RT-PCR primers used in this study.

Supplementary Table 2 | The list of 56 genes for model construction.

Supplementary Table 3 | The significance of differences in infiltrated immune cell between risk groups.

Supplementary Table 4 | List of significance for all 138 compounds.

REFERENCES

Bild, A. H., Yao, G., Chang, J. T., Wang, Q., Potti, A., Chasse, D., et al. (2006). Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 439, 353–357. doi: 10.1038/nature04296

Binnewies, M., Roberts, E. W., Kersten, K., Chan, V., Fearon, D. F., Merad, M., et al. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* 24, 541–550.

Brady, J. J., Chuang, C. H., Greenside, P. G., Rogers, Z. N., Murray, C. W., Caswell, D. R., et al. (2016). An Arntl2-driven secretome enables lung adenocarcinoma metastatic self-sufficiency. *Cancer Cell* 29, 697–710. doi: 10.1016/j.ccell.2016.03.003

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. *Cell Rep.* 18, 248–262. doi: 10.1016/j.celrep.2016.12.019

Dai, E., Han, L., Liu, J., Xie, Y., Kroemer, G., Klionsky, D. J., et al. (2020a). Autophagy-dependent ferroptosis drives tumor-associated macrophage polarization via release and uptake of oncogenic KRAS protein. *Autophagy* 16, 2069–2083. doi: 10.1080/15548627.2020.1714209

Dai, E., Han, L., Liu, J., Xie, Y., Zeh, H. J., Kang, R., et al. (2020b). Ferroptotic damage promotes pancreatic tumorigenesis through a TMEM173/STING-dependent DNA sensor pathway. *Nat. Commun.* 11:6339.

Davis, S., and Meltzer, P. S. (2007). GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 23, 1846–1847. doi: 10.1093/bioinformatics/btm254

Dixon, S. J., Lemberg, K. M., Lamprecht, M. R., Skouta, R., Zaitsev, E. M., Gleason, C. E., et al. (2012). Ferroptosis: an iron-dependent form of nonapoptotic cell death. *Cell* 149, 1060–1072. doi: 10.1016/j.cell.2012.03.042

Ganesh, S., Shui, X., Craig, K. P., Park, J., Wang, W., Brown, B. D., et al. (2018). RNAi-mediated beta-catenin inhibition promotes T cell infiltration and antitumor activity in combination with immune checkpoint blockade. *Mol. Ther.* 26, 2567–2579. doi: 10.1016/j.ymthe.2018.09.005

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* 6:l1.

Gao, M., Yi, J., Zhu, J., Minikes, A. M., Monian, P., Thompson, C. B., et al. (2019). Role of mitochondria in ferroptosis. *Mol. Cell* 73, 354–363.e353.

Geeleher, P., Cox, N., and Huang, R. S. (2014). pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS One* 9:e107468. doi: 10.1371/journal.pone.0107468

Gu-Trantien, C., Loi, S., Garaud, S., Equeter, C., Libin, M., De Wind, A., et al. (2013). CD4(+) follicular helper T cell infiltration predicts breast cancer survival. *J. Clin. Invest.* 123, 2873–2892. doi: 10.1172/jci67428

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 14:7. doi: 10.1007/1-4020-7788-2_2

He, X., and Xu, C. (2020). Immune checkpoint signaling and cancer immunotherapy. *Cell Res.* 30, 660–669. doi: 10.1038/s41422-020-0343-4

Henmi, T., Amano, K., Nagaura, Y., Matsumoto, K., Echigo, S., Tamura, S., et al. (2009). A mechanism for the suppression of interleukin-1-induced nuclear factor kappaB activation by protein phosphatase 2Ceta-2. *Biochem. J.* 423, 71–78. doi: 10.1042/bj20090208

Jennis, M., Kung, C. P., Basu, S., Budina-Kolomets, A., Leu, J. I., Khaku, S., et al. (2016). An African-specific polymorphism in the TP53 gene impairs p53 tumor suppressor function in a mouse model. *Genes. Dev.* 30, 918–930. doi: 10.1101/gad.275891.115

Jiang, L., Kon, N., Li, T., Wang, S. J., Su, T., Hibshoosh, H., et al. (2015). Ferroptosis as a p53-mediated activity during tumour suppression. *Nature* 520, 57–62. doi: 10.1038/nature14344

Jiang, T. X., Zou, J. B., Zhu, Q. Q., Liu, C. H., Wang, G. F., Du, T. T., et al. (2019). SIP/CacyBP promotes autophagy by regulating levels of BRUCE/Apollon, which stimulates LC3-I degradation. *Proc. Natl. Acad. Sci. U.S.A.* 116, 13404–13413. doi: 10.1073/pnas.1901039116

Jiang, X., Stockwell, B. R., and Conrad, M. (2021). Ferroptosis: mechanisms, biology and role in disease. *Nat. Rev. Mol. Cell Biol.* 22, 266–282. doi: 10.1038/s41580-020-00324-8

Kamada, R., Kudoh, F., Ito, S., Tani, I., Janairo, J. I. B., Omichinski, J. G., et al. (2020). Metal-dependent Ser/Thr protein phosphatase PPM family: Evolution, structures, diseases and inhibitors. *Pharmacol .Ther.* 215:107622. doi: 10.1016/j.pharmthera.2020.107622

King, C., Tangye, S. G., and Mackay, C. R. (2008). T follicular helper (TFH) cells in normal and dysregulated immune responses. *Annu Rev Immunol* 26, 741–766. doi: 10.1146/annurev.immunol.26.021607.090344

Kursa, M. B., and Rudnicki, W. R. (2010). Feature selection with the Boruta package. *J. Stat. Softw.* 36, 1–13.

Leone, R. D., Zhao, L., Englert, J. M., Sun, I. M., Oh, M. H., Sun, I. H., et al. (2019). Glutamine blockade induces divergent metabolic programs to overcome tumor immune evasion. *Science* 366, 1013–1021. doi: 10.1126/science.aav2588

Lian, Y. F., Huang, Y. L., Zhang, Y. J., Chen, D. M., Wang, J. L., Wei, H., et al. (2019). CACYBP enhances cytoplasmic retention of P27(Kip1) to promote hepatocellular carcinoma progression in the absence of RNF41 mediated degradation. *Theranostics* 9, 8392–8408. doi: 10.7150/thno.36838

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27, 1739–1740. doi: 10.1093/bioinformatics/btr260

Liu, T., Jiang, L., Tavana, O., and Gu, W. (2019). The deubiquitylase OTUB1 mediates ferroptosis via stabilization of SLC7A11. *Cancer Res.* 79, 1913–1924. doi: 10.1158/0008-5472.can-18-3037

Liu, Z., Zhao, Q., Zuo, Z.-X., Yuan, S.-Q., Yu, K., Zhang, Q., et al. (2020). Systematic analysis of the aberrances and functional implications of ferroptosis in cancer. *Iscience* 23:101302. doi: 10.1016/j.isci.2020.101302

Long, X., Zhou, W., Wang, Y., and Liu, S. (2018). Prognostic significance of ANLN in lung adenocarcinoma. *Oncol. Lett.* 16, 1835–1840.

Mathews, M. B., and Hershey, J. W. (2015). The translation factor eIF5A and human cancer. *Biochim. Biophys. Acta* 1849, 836–844. doi: 10.1016/j.bbagrm.2015.05.002

Miao, Y. R., Zhang, Q., Lei, Q., Luo, M., Xie, G. Y., Wang, H., et al. (2020). ImmuCellAI: a unique method for comprehensive T-cell subsets abundance prediction and its application in cancer immunotherapy. *Adv. Sci.* 7:1902880. doi: 10.1002/advs.201902880

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12, 453–457. doi: 10.1038/nmeth.3337

Okayama, H., Kohno, T., Ishii, Y., Shimada, Y., Shiraishi, K., Iwakawa, R., et al. (2012). Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res.* 72, 100–111. doi: 10.1158/0008-5472.can-11-1403

Ow, Y. P., Green, D. R., Hao, Z., and Mak, T. W. (2008). Cytochrome c: functions beyond respiration. *Nat. Rev. Mol. Cell Biol.* 9, 532–542. doi: 10.1038/nrm2434

Ribas, A. (2015). Releasing the brakes on cancer immunotherapy. *N. Engl. J. Med.* 373, 1490–1492. doi: 10.1056/nejmp1510079

Ribas, A., and Wolchok, J. D. (2018). Cancer immunotherapy using checkpoint blockade. *Science* 359, 1350–1355.

Rousseaux, S., Debernardi, A., Jacquiau, B., Vitte, A.-L., Vesin, A., Nagy-Mignotte, H., et al. (2013). Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. *Sci. Transl. Med.* 5:186ra166.

Sato, M., Larsen, J. E., Lee, W., Sun, H., Shames, D. S., Dalvi, M. P., et al. (2013). Human lung epithelial cells progressed to malignancy through specific oncogenic manipulations. *Mol. Cancer Res.* 11, 638–650. doi: 10.1158/1541-7786.mcr-12-0634-t

Schaer, D. A., Beckmann, R. P., Dempsey, J. A., Huber, L., Forest, A., Amaladas, N., et al. (2018). The CDK4/6 inhibitor abemaciclib induces a T Cell inflamed tumor microenvironment and enhances the efficacy of PD-L1 checkpoint blockade. *Cell Rep.* 22, 2978–2994. doi: 10.1016/j.celrep.2018.02.053

Smyth, G. K. (2005). "Limma: linear models for microarray data," in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, eds R. Gentleman, V. J. Carey, W. Huber, R. A. Irizarry, and S. Dudoit (New York, NY: Springer), 397–420. doi: 10.1007/0-387-29362-0_23

Stockwell, B. R., Friedmann Angeli, J. P., Bayir, H., Bush, A. I., Conrad, M., Dixon, S. J., et al. (2017). Ferroptosis: a regulated cell death nexus linking metabolism. *Redox Biol. Dis. Cell* 171, 273–285. doi: 10.1016/j.cell.2017.09.021

Strnadel, J., Choi, S., Fujimura, K., Wang, H., Zhang, W., Wyse, M., et al. (2017). eIF5A-PEAK1 signaling regulates YAP1/TAZ protein expression and pancreatic

cancer cell growth. *Cancer Res.* 77, 1997–2007. doi: 10.1158/0008-5472.can-16-2594

Suzuki, C., Daigo, Y., Ishikawa, N., Kato, T., Hayama, S., Ito, T., et al. (2005). ANLN plays a critical role in human lung carcinogenesis through the activation of RHOA and by involvement in the phosphoinositide 3-kinase/AKT pathway. *Cancer Res.* 65, 11314–11325. doi: 10.1158/0008-5472.can-05-1507

Tang, D., and Kroemer, G. (2020). Ferroptosis. *Curr. Biol.* 30, R1292–R1297.

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

Tomida, S., Takeuchi, T., Shimada, Y., Arima, C., Matsuo, K., Mitsudomi, T., et al. (2009). Relapse-related molecular signature in lung adenocarcinomas identifies patients with dismal prognosis. *J. Clin. Oncol.* 27, 2793–2799. doi: 10.1200/jco.2008.19.7053

Topolska-Wos, A. M., Chazin, W. J., and Filipek, A. (2016). CacyBP/SIP–Structure and variety of functions. *Biochim. Biophys. Acta* 1860, 79–85. doi: 10.1016/j.bbagen.2015.10.012

Wang, S. J., Li, D., Ou, Y., Jiang, L., Chen, Y., Zhao, Y., et al. (2016). Acetylation Is Crucial for p53-mediated ferroptosis and tumor suppression. *Cell Rep.* 17, 366–373.

Wang, W., Green, M., Choi, J. E., Gijon, M., Kennedy, P. D., Johnson, J. K., et al. (2019). CD8(+) T cells regulate tumour ferroptosis during cancer immunotherapy. *Nature* 569, 270–274. doi: 10.1038/s41586-019-1170-y

Wang, Y., Chen, W., Zhu, M., and Xian, L. (2021). Ferroptosis-related gene signature and patterns of immune infiltration predict the overall survival in patients with lung adenocarcinoma. *Front. Mol. Biosci.* 8:692530.

Wen, Q., Liu, J., Kang, R., Zhou, B., and Tang, D. (2019). The release and activity of HMGB1 in ferroptosis. *Biochem. Biophys. Res. Commun.* 510, 278–283. doi: 10.1016/j.bbrc.2019.01.090

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, J., Zheng, H., Yuan, S., Zhou, B., Zhao, W., Pan, Y., et al. (2019). Overexpression of ANLN in lung adenocarcinoma is associated with metastasis. *Thorac. Cancer* 10, 1702–1709. doi: 10.1111/1759-7714.13135

Zhang, Y., Shi, J., Liu, X., Feng, L., Gong, Z., Koppula, P., et al. (2018). BAP1 links metabolic regulation of ferroptosis to tumour suppression. *Nat. Cell Biol.* 20, 1181–1192. doi: 10.1038/s41556-018-0178-0

Zhang, Y., Zhuang, L., and Gan, B. (2019). BAP1 suppresses tumor development by inducing ferroptosis upon SLC7A11 repression. *Mol. Cell Oncol.* 6:1536845. doi: 10.1080/23723556.2018.1536845

Zhou, N., and Bao, J. (2020). FerrDb: a manually curated resource for regulators and markers of ferroptosis and ferroptosis-disease associations. *Database* 2020:baaa021.

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: 03 November 2021
doi: 10.3389/fcell.2021.745120

Combinatorial Analysis of AT-Rich Interaction Domain 1A and CD47 in Gastric Cancer Patients Reveals Markers of Prognosis

Qianfu Zhao[1†], Qu Cai[1†], Shanhe Yu[2†], Jun Ji[3], Zhenggang Zhu[3], Chao Yan[3] and Jun Zhang[1,4*]*

[1] Department of Oncology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [2] State Key Laboratory of Medical Genomics, National Research Center for Translational Medicine at Shanghai, Shanghai Institute of Hematology, Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China, [3] Shanghai Key Laboratory of Gastric Neoplasms, Department of Surgery, Ruijin Hospital, Shanghai Institute of Digestive Surgery, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [4] State Key Laboratory of Oncogenes and Related Genes, Shanghai Jiao Tong University, Shanghai, China

OPEN ACCESS

Edited by:
Geng Chen, GeneCast Biotechnology Co., Ltd., China

Reviewed by:
KuanHui Chen, Delta State University, United States
Yongyong Yang, Northwestern University, United States

***Correspondence:**
Chao Yan
yc11297@rjh.com.cn
Jun Zhang
junzhang10977@sjtu.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: *21 July 2021*
Accepted: *27 September 2021*
Published: *03 November 2021*

Citation:
Zhao Q, Cai Q, Yu S, Ji J, Zhu Z, Yan C and Zhang J (2021) Combinatorial Analysis of AT-Rich Interaction Domain 1A and CD47 in Gastric Cancer Patients Reveals Markers of Prognosis. Front. Cell Dev. Biol. 9:745120. doi: 10.3389/fcell.2021.745120

Background: The AT-rich interaction domain 1A (*ARID1A*) is thought to be a tumor suppressive gene, and most of its mutations result in loss of expression of ARID1A protein. Combined with SIRPα on the surface of macrophages, CD47 on the surface of cancer cells can send an antiphagocytic "Don't eat me" signal to the immune system that helps to avoid immune surveillance. However, the relationship between ARID1A and CD47 expression and their prognostic value in gastric cancer (GC) are still unknown.

Methods: In this study, we evaluated ARID1A and CD47 expression in 154 GC patients' tissues using tissue microarray. Expressions of ARID1A and CD47 in GC cell lines were determined by western blot and quantitative reverse transcriptase–polymerase chain reaction (qRT-PCR) techniques, and cell membranous CD47 expression was quantified by flow cytometry. In addition, chromatin immunoprecipitation (ChIP)–qPCR was used to determine the aspects of regulation of *CD47* by ARID1A. The proportions of tumor-infiltrating immune cells were estimated on The Cancer Genome Atlas (TCGA) data set by using quanTIseq and EPIC algorithms. The infiltration of M1-polarized macrophages, M2-polarized macrophages, and regulatory T cells (Tregs) in GC tissues was determined by multispectral immunofluorescence.

Results: A significant correlation was found between loss of ARID1A and high expression of CD47 at protein level in GC. By integrating 375 bulk RNA sequencing samples from TCGA data set, we found that mutated *ARID1A* correlated with high *CD47* expression. In GC cell lines, knockdown of *ARID1A* significantly increased CD47 expression both at protein and mRNA levels as measured by western blot, qRT-PCR, and flow cytometry. Moreover, ChIP-qPCR revealed that *CD47* was a direct downstream target gene of ARID1A in GC. Utilizing univariate and multivariate survival analyses, we found that patients with $ARID1A^{loss}CD47^{high}$ expression had a worse prognosis. Estimation of infiltrating immune cells on TCGA data set showed

that a higher infiltration proportion of M2 macrophages and Tregs was found in *ARID1A*^{mutated}*CD47*^{high} expression subgroup. Furthermore, application of multispectral immunofluorescence revealed a higher infiltration proportion of M2 macrophages and Tregs in ARID1A^{loss}CD47^{high} GC tissues.

Conclusion: Loss of ARID1A is strongly correlated with high CD47 expression in GC, and combination of ARID1A and CD47 is a promising prognosis factor in GC.

Keywords: ARID1A, CD47, gastric cancer, prognosis, immunotherapy

INTRODUCTION

Gastric cancer (GC) is the third leading cause of cancer-related death worldwide with the sixth highest morbidity rate (Sung et al., 2021). Among cancers, GC remains to be a huge burden on the public health in China by showing the third highest incidence and mortality rate (Zhang et al., 2021). Because of the population growth and changes of the age structure, the incidence rate of GC is anticipated to be increasing in the future. Despite achievements in the development of diagnostic techniques, patients are often diagnosed at an advanced stage, thus resulting in a high mortality rate. Previous studies have shown that the occurrence and development of GC are closely related to the gene mutation and immune state (Nagarajan et al., 2012; Li et al., 2018; Lin et al., 2019).

The SWI/SNF complex was first identified as a protein complex important for cellular responses to mating-type switching (SWI) or sucrose fermentation (SNF) in the yeast (Lu and Allis, 2017). This complex uses the energy released by ATP hydrolysis to drive nucleosome movement and regulate structuring of chromatin (Roberts and Orkin, 2004). As a core component of SWI/SNF chromatin remodeling complexes, ARID1A participates in various biological processes, including gene transcription, DNA replication, and DNA damage repair (Shen et al., 2015). *ARID1A* is identified as a tumor suppressor gene and is mutated at a high frequency in diverse cancers (Wu and Roberts, 2013; Lo et al., 2021), such as ovarian clear cell carcinoma (∼57%), uterine corpus endometrial carcinoma (∼30%), and stomach adenocarcinoma (∼31%) (Lo et al., 2021). Recently, proof was provided for the context-dependent tumor-suppressive and oncogenic roles of *ARID1A* in the liver cancer (Sun et al., 2017). These findings prompt one to wonder if ARID1A plays a key role in the tumorigenesis and cancer development.

In recent years, immune checkpoint therapy has achieved a great success in cancer treatment and has shifted the paradigm in this field of research. Programmed death 1/programmed death 1 ligand (PD-1/PD-L1) and cytotoxic T lymphocyte–associated protein 4 (CTLA-4) are the most common target molecules for immune checkpoint therapy, and their targeting has shown clinical success in many solid tumors. However, the objective response rate of current immune checkpoint therapy is approximately only 20%, which is low (Topalian et al., 2012). Therefore, it is vital to find new immune checkpoints to develop new immune therapy drugs. CD47, also known as integrin-associated protein, is a cell membrane protein that belongs to the immunoglobulin superfamily (Vonderheide, 2015). CD47 is approximately 50 kDa, with an extracellular N-terminal IgV domain, five transmembrane domains, and a short intracytoplasmic C-terminal domain. CD47 is overexpressed in almost all tumor cells and closely related to clinical prognosis. Besides, in combination with SIRPα on the surface of macrophages, CD47 can send an antiphagocytic "Don't eat me" signal to the immune system that helps cancer cells avoid immune surveillance (Morrissey et al., 2020). Thus, CD47 is expected to become an emerging targeted immune checkpoint.

Results of the clinical and non-clinical research accumulated over the years indicate that ARID1A deficiency is closely related to the immune checkpoint therapy, such as anti–PD-L1 therapy (Shen et al., 2018). However, the relationship between ARID1A and CD47 in GC has rarely been the focus of a study. In the present study, we mainly focused on examining the expression levels of ARID1A and CD47 and their expression relationship and elucidating their prognostic value in GC patients.

MATERIALS AND METHODS

Patients and Samples

Tissue microarray used in this research included a total of 159 GC samples that were purchased from Shanghai Outdo Biotech Co., Ltd. The samples were collected between June 2011 and May 2013, and the follow-up ended on June 2017. After ruling out samples with missing data (3 samples) and distant metastasis (2 samples), 154 samples were included in the present study. Patients received neither chemotherapy nor radiation therapy before surgery. All the samples were validated as gastric adenocarcinoma by the researchers. Clinicopathologic parameters, including sex, age, tumor location, tumor histological grade, tumor size, T stage, N stage, tumor-node-metastasis (TNM) stage, and human epidermal growth factor receptor 2 (HER2) status, were collected. The tumor stage was identified using the eighth version of the TNM staging system of the American Joint Committee on Cancer. The overall survival (OS) time was defined from postoperative to death or to the end of the follow-up. Written informed consents were provided by Shanghai Outdo Biotech Co., Ltd., and the research was conducted in accordance with the recognized ethical guidelines. The studies that involved human participants were reviewed and approved by the Ethics Committee of Shanghai Jiao Tong University School of Medicine Affiliated Ruijin

Hospital. The patients provided written informed consent to participate in this study.

Immunochemistry Staining and Evaluation

Tissue microarray was established, and immunochemistry was conducted following the EnVision two-step procedure of Dako REAL^{TM} Envision^{TM} Detection System (Dako, Agilent Technologies, CA, United States) (Ji et al., 2019). In brief, the sections were deparaffinized and rehydrated before being heated in a microwave oven for antigen retrieval. After having been washed with phosphate-buffered saline (PBS) three times, the sections were incubated with 3% hydrogen peroxide to block the endogenous peroxidase activity. Afterward, the sections were incubated with 1 × Animal-Free Blocking Solution (Cell Signaling Technology, United States) for 1 h at room temperature to reduce non-specific staining. After blocking, the sections were incubated with the primary antibodies overnight at 4°C. On the next day, they were incubated with the horseradish peroxidase (HRP)–labeled second antibodies at 37°C for 30 min and then were incubated with diaminobenzidine solution for visualization.

Anti-ARID1A mouse monoclonal antibody (sc-32761; 1:50 dilution; Santa Cruz Biotechnology, United States) was used to detect the expression level of ARID1A protein in GC tissues. ARID1A was located in the cellular nuclear area and was dyed yellow to varying degrees in cancer cells. Using CaseViewer (magnification, × 40; 3DHISTECH Ltd.), the expression status of ARID1A protein was determined by two independent investigators. Loss of ARID1A was identified once the nuclear staining was vacant in more than 10% of cancer cells, and the rest of the cases were defined as "preserved" (Kim et al., 2016; **Figures 1A,B**).

Anti-CD47 rabbit polyclonal antibody (20305-1-AP; 1:50 dilution; Proteintech Group, United States) was applied to identify CD47 expression level in tissues. The expression of CD47 in GC tissues was evaluated by histochemistry score (*H* score). *H* score = Σ(pi × *i*) = (percentage of weak intensity area × 1) + (percentage of moderate intensity area × 2) + (percentage of strong intensity area × 3), where *i* is the staining intensity, including negative without staining (*i* = 0), weak positive (*i* = 1), medium positive (*i* = 2), and strong positive (*i* = 3), and pi is proportion of the positive cells for the corresponding staining intensity (Paschalis et al., 2019). The median of *H* score was defined as the threshold value to distinguish low and high CD47 expression group. The cutoff value is 179.86 (**Figures 1C,D**).

Cell Lines and Cell Culture

GC cell lines (BGC-823, MKN-28, SNU-1, SUN-5, SNU-16, MKN-45, KATO-III, BGC-803, AGS, SGC-7901, and NCI-N87) and GSE-1 cell lines were preserved in the Ruijin Hospital (Shanghai, China) (Wang et al., 2019). Cells were cultured in DMEM (Dulbecco modified eagle medium) or RPMI-1640 medium supplemented with 10% fetal calf serum at 37°C and with 5% CO_2.

Lentivirus Transduction

The ARID1A lentiviral short hairpin RNAs (shRNAs) were supplied by GenePharma (Shanghai, China). The target sequences are listed in **Supplementary Table 1**. The shRNAs were cloned into PGMLV-SB3 (PGMLV-hU6-MCS-CMV-Puro-WPRE) vector, and an empty vector was used as the control vector. For lentivirus transduction, cells were transduced following the manufacturer's instruction and were selected by puromycin (cat. Ant-pr-5; Invivogen) with the concentration of 2 μg/mL. The expression of ARID1A in BGC-823 and MKN-28 was confirmed by western blot and quantitative reverse transcriptase–polymerase chain reaction (qRT-PCR).

Western Blot Analysis

Proteins were fractionated by 10% sodium dodecyl sulfate–polyacrylamide gel electrophoresis gels and were transferred to polyvinylidene fluoride (PVDF) membranes. The primary antibodies included ARID1A (12354S; 1:1,000 dilution; Cell Signaling Technology, United States), CD47 (20305-1-AP; 1:1,000 dilution; Proteintech Group, United States), and GAPDH (51332S; 1:1,000, Cell Signaling Technology, United States). PVDF membranes were blocked by 5% skim milk for 2 h at room temperature and then were incubated with primary antibodies overnight at 4°C. The membranes were then incubated with HRP-conjugated secondary antibodies (7074, 7076; 1:2,500, Cell Signaling Technology, United States). ECL western blotting substrate (cat. P10300, NCM Biotech, China) and infrared imaging system (LI-COR Biosciences) were used to visualize the protein bands.

Real-Time Quantitative Reverse Transcriptase–Polymerase Chain Reaction

RNA was extracted by Trizol reagent method. Reverse transcription in 20-μL system was performed following the protocol of Fasting gDNA Dispelling RT SuperMix kit (KR118-03; TIANGEN, China). Primers for qRT-PCR were *ARID1A*: 5′-GCATCCTTCCATGAACCAAT-3′ (forward), 5′-CCCATGCCTGTGTGTATCTG-3′(reverse);*GAPDH*:5′-GGAC CTGACCTGCCGTCTAG-3′ (forward), 5′-GTAGCCCAGGA TGCCCTTGA-3′(reverse);*CD47*:5′-AGAAGGTGAAACGATC ATCGAGC-3′ (forward), 5′-CTCATCCATACCACCGGATCT-3′ (reverse). Relative mRNA expression was normalized to *GAPDH* and calculated by $2^{-\triangle CT}$ method. Each experiment was performed in triplicate.

Flow Cytometry

Surface expression of CD47 in GC cells was measured by flow cytometry. In brief, equal amounts of cells were harvested, were Fc-blocked, and were incubated with anti-CD47 antibody (PE) (cat. 12283-MM07-P; SinoBiological) or isotype control for 30 min at room temperature. Afterward, the cells were washed three times and were resuspended in 1% fetal bovine serum/PBS. The fluorescence intensity of CD47 in GC cells was collected on a BD Calibur flow cytometer and was analyzed by FlowJo V10 software.

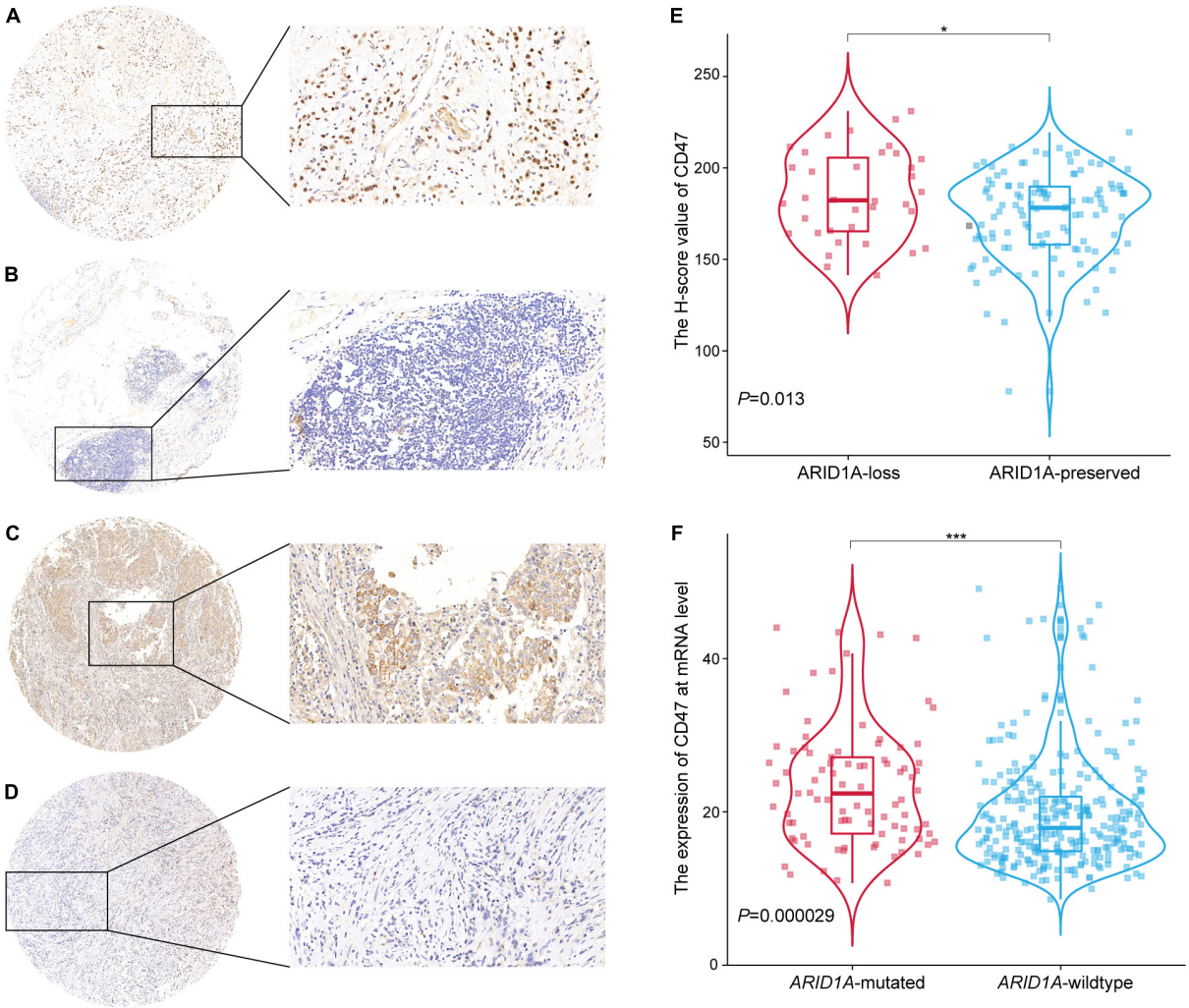


FIGURE 1 | Expression of ARID1A and CD47 and their relationship in GC tissues. **(A,B)** Representative IHC images of ARID1A staining, ARID1A-preserved expression **(A)**, ARID1A-loss expression **(B)**. **(C,D)** Representative IHC images of CD47 staining, CD47 high expression **(C)**, CD47 low expression **(D)**. **(E)** Relationship between loss/preserved expression of ARID1A and CD47 expression (n = 154). **(F)** Relationship between mutated/wild-type expression of *ARD1A* and *CD47* expression (data from TCGA, n = 375). $^{*}p < 0.05$, $^{***}p < 0.001$.

Chromatin Immunoprecipitation Assay

Chromatin immunoprecipitation (ChIP) assays were performed as previously described (Yu et al., 2018). Briefly, BGC-823 and MKN28 cells were harvested and crosslinked in 1% formaldehyde at room temperature, and the fixation reactions were quenched by adding glycine to a final concentration of 0.125 M. After sonication, we separately incubated the soluble chromatins with the antibodies anti-ARID1A (12354S, Cell Signaling Technology) or control IgG (2729, Cell Signaling Technology). Afterward, chromatin immunocomplexes were precipitated with protein A (Millipore, 16-661). The immunoprecipitated complex was washed, and DNA was extracted and purified by QIAquick PCR Purification Kit (Qiagen). ChIP DNA was analyzed by qPCR using specific primers, and the data were normalized by input DNA. The primers for ChIP-qPCR are the following: human *CD47* (5′-AAAGAAGGGGATCCCTAGCA-3′, 5′-CCATCTCCAAATGCACACAC-3′).

Public Database Analysis

Public data from The Cancer Genome Atlas (TCGA), including that of 375 GC patients was included in our analysis. Other cancer data from TCGA including colon adenocarcinoma (COAD), lung adenocarcinoma (LUAD), breast invasive carcinoma (BRCA) and kidney renal clear cell carcinoma (KIRC) was also included in this analysis. TCGA data were downloaded from the Genomic Data Commons data portal,[1] and *ARID1A* mutation information was acquired from the cBioPortal.[2] The cutoff value of *CD47* in TCGA data sets was determined as

[1]https://portal.gdc.cancer.gov/

[2]http://www.cbioportal.org/

median. Tumor-infiltrating immune cells were estimated on the TCGA data by quanTIseq[3] (Finotello et al., 2019) and EPIC[4] (Racle et al., 2017).

Multispectral Immunofluorescence

Multispectral immunofluorescence staining was performed following the manufacturer's instructions of Opal 4-Color Manual IHC Kit (NEL794001KT; PerkinElmer, Waltham, MA, United States). The Opal kit uses tyramine signal amplification conjugated fluorophores to detect targets within an immunofluorescence assay. Initially, the slides were baked in the oven at 65°C for 1 h, dewaxed with xylene, and rehydrated through a graded series of ethanol solutions. After rehydration, slides were fixed in 10% neutral buffered formalin for 20 min. The slides were then placed in microwave for 45 s at 100% power followed by additional 15 min at 20% power in AR buffer. The slides were blocked with blocking buffer (PerkinElmer) and incubated in a humidified chamber for 10 min at room temperature. After removal of the blocking buffer, the slides were incubated with primary antibodies overnight at 4°C. Primary antibodies including CD11c (Abcam, ab11029, 1:75), CD163 (Abcam, ab182422, 1:800), and Foxp3 (Abclonal, A12051, 1:50). The slides were sequentially incubated with Opal Polymer HRP MS + Rb (PerkinElmer) and Opal Fluorophore Working Solution (pal 520, Opal 570, Opal 670; 1:50 dilution) for 10 min at room temperature, respectively. Finally, the slides were incubated with DAPI solution (PerkinElmer) for 5 min at room temperature and were mounted with ProLong Diamond Antifade Mountant (Invitrogen). Imaging was performed using Vectra Quantitative Pathology Imaging Systems (PerkinElmer), and image analysis was performed using the InForm Advanced Image Analysis software (inForm 3.0; PerkinElmer).

Statistical Analysis

The data were analyzed using GraphPad Prism (version 8.0.1), statistical product and service solution (SPSS) (version 22), R (version 4.1.0), and Rstudio (version 4.0.4) software. A χ^2-test was used to analyze the relationship between clinicopathologic parameters and ARID1A status and CD47 expression level. Survival analysis was performed using the Kaplan–Meier method, log-rank test, and Cox regression analysis under various conditions. Analysis of variance and *post hoc* test were used to calculate the difference in the means of multiple groups. Non-parametric test was used to analyze the data, which did not follow normal distribution. $p < 0.05$ was considered statistically significant.

RESULTS

Clinicopathological Characteristics of 154 Gastric Cancer Patients

The clinicopathological characteristics of 154 GC patients enrolled in this study are shown in **Table 1**. In the registered patients, 66.2% of patients were males, and 33.8% were females. In line with a previous study, the incidence rate ratio of male to female was approximately 2:1 in China (Sung et al., 2021). Most patients were 60 years or older (119 cases, 77.3%). For the histological grading, the samples of 66.2% of the patients were poorly differentiated. Regarding T stage, we found that most of the patients were at T4 stage, indicative of the fact that most of the patients were diagnosed at an advanced stage. For the classification according to HER2 status, 116 of patients (75.3%) were found to be negative/weak positive (1 +), which assigned them as HER2 negative, and 9.1% of them were strong positive (3 +), which assigned them as HER2 positive.

TABLE 1 | Clinicopathological characteristics of 154 GC patients.

Factors	No. of patients	Percent
Age (years)		
<60	35	22.7
≥60	119	77.3
Gender		
Male	102	66.2
Female	52	33.8
Location		
Proximal	14	9.1
Middle	70	45.5
Distal	70	45.5
Histological grade		
Moderate	52	33.8
Poor	102	66.2
Tumor size (cm)		
<5	81	52.6
≥5	73	47.4
N stage		
N0	44	28.6
N1	33	21.4
N2	34	22.1
N3	43	27.9
T stage		
T1	4	2.6
T2	15	9.7
T3	18	11.7
T4	117	80
TNM stage		
I	15	9.7
II	36	23.4
III	103	66.9
HER2 status		
0/1+	116	75.3
2+	24	15.6
3+	14	9.1

A Strong Correlation Exists Between a High Expression Level of CD47 and Loss/Mutation of AT-Rich Interaction Domain 1A in Gastric Cancer

Immunohistochemical (IHC) staining was carried out on GC tissue microarrays to evaluate ARID1A (**Figures 1A,B**) and CD47 expression (**Figures 1C,D**). According to the evaluation criteria,

[3]http://icbi.at/quantiseq

[4]http://epic.gfellerlab.org

loss of ARID1A expression was identified in 36 cases of GC tissues (23.4%). The remaining 118 cases (76.6%) exhibited preserved expression of ARID1A. Regarding CD47 expression, 68 of GC tissues (44.2%) were identified as low expression tissues, and 86 of them (55.8%) were identified as high-expression ones. Next, we investigated the correlation between the expression of ARID1A and CD47 in GC tissues. More specifically, in GC tissues, the *H* score of CD47 was higher in the ARID1A-loss group than in the ARID1A-preserved group (**Figure 1E**). Previous studies have reported that most *ARID1A* mutations are inactivated non-sense mutations or frameshift mutations, which result in the loss of ARID1A protein expression (Wang et al., 2004; Wiegand et al., 2010; Wu et al., 2014). Thus, we examined the *ARID1A*-mutated expression as a surrogate of ARID1A-loss expression in TCGA data set. An analogous trend was also identified in CD47 expression at transcriptional level through analysis of the data from the TCGA. More specifically, the expression of CD47 in *ARID1A*-mutated group was found to be higher than that of the *ARID1A* wild-type group (**Figure 1F**). Furthermore, we found that the mRNA expression level of *CD47* was higher in *ARID1A*-mutated samples than *ARID1A* wild-type samples in COAD, LUAD, BRCA, and KIRC, respectively (**Supplementary Figure 1**). Putting side-by-side the information obtained so far, we concluded that loss of ARID1A was significantly correlated with high CD47 expression in GC.

AT-Rich Interaction Domain 1A Status and CD47 Expression Associate With Clinicopathological Features of Gastric Cancer

The relationships between ARID1A status, CD47 expression, and clinicopathologic features of GC patients are shown in **Table 2**. As shown in **Table 2**, ARID1A status and CD47 expression level were significantly associated with T stage. Notably, in early T stage (T1 and T2) cases, the expression of ARID1A was always preserved, and in the majority of instances, the expression of

TABLE 2 | The correlation between ARID1A status and CD47 expression and clinicopathological characteristics.

		ARID1A expression			CD47 expression		
	Total (*n* = 154)	**Loss (%) (*n* = 36)**	**Preserved (%) (*n* = 118)**	***P*-value**	**Low (%) (*n* = 68)**	**High (%) (*n* = 86)**	***P*-value**
Age (years)				0.591			0.17
<60	35	7(20.0)	28(80.0)		19(54.3)	16(45.7)	
≥60	119	29(24.4)	90(75.6)		49(41.2)	70(58.8)	
Gender				0.094			0.484
Male	102	28(27.5)	74(72.5)		43(42.2)	59(57.8)	
Female	52	8(15.4)	44(84.6)		25(48.1)	27(51.9)	
Location				0.327			0.351
Proximal	14	5(35.7)	9(64.3)		8(57.1)	6(42.9)	
Middle	70	13(18.6)	57(81.4)		33(47.1)	37(52.9)	
Distal	70	18(25.7)	52(74.3)		27(38.6)	43(61.4)	
Histological grade				0.642			0.989
Moderate	52	11(21.2)	41(78.8)		23(44.2)	29(55.8)	
Poor	102	25(24.5)	77(75.5)		45(44.1)	57(55.9)	
Tumor size (cm)				**0.002**			0.089
<5	81	11(13.6)	70(86.4)		41(50.6)	40(49.4)	
≥5	73	25(34.2)	48(65.8)		27(37.0)	46(63.0)	
N stage				0.317			0.091
N0	44	6(13.6)	38(86.4)		22(50.0)	22(50.0)	
N1	33	8(24.2)	25(75.8)		11(33.3)	22(66.7)	
N2	34	10(29.4)	24(70.6)		11(32.4)	23(67.6)	
N3	43	12(27.9)	31(72.1)		24(55.8)	19(44.2)	
T stage				**0.011**			**0.024**
T1	4	0(0.0)	4(100.0)		4(100.0)	0(0.0)	
T2	15	0(0.0)	15(100.0)		9(60.0)	6(40.0)	
T3	18	4(22.2)	14(77.8)		9(50.0)	9(50.0)	
T4	117	32(27.4)	85(72.6)		46(39.3)	71(60.7)	
TNM stage				**0.008**	46(39.4)		0.054
I	15	0(0.0)	15(100.0)		11(73.3)	4(26.7)	
II	36	7(19.4)	29(80.6)		14(38.9)	22(61.1)	
III	103	29(28.2)	74(71.8)		43(41.7)	60(58.2)	

P-value < 0.05 is displayed in bold.

CD47 was low. These findings in early T stage samples might be explained by the tumor suppressor effect of ARID1A and the lack of CD47-mediated antiphagocytosis effect (Wu and Roberts, 2013; Morrissey et al., 2020). As for TNM stage, the proportion of ARID1A-loss cases increased with the progression of the TNM stage, and the proportion of high CD47 cases appeared to show a similar trend. These might indicate that ARID1A is often lost at the advanced stages of GC. Moreover, we also found that ARID1A status was correlated with tumor size. No significance difference in the expression of ARID1A was found between groups with age, gender, tumor location, histological grade, and N stage. Besides, no significant difference in the expression of CD47 was found between groups with age, gender, tumor location, histological grade, tumor size, N stage, and TNM stage.

AT-Rich Interaction Domain 1A Regulates CD47 Expression Levels in Gastric Cancer Cell Lines

To validate the relationship between ARID1A and CD47, we examined the expression of ARID1A in GC cell lines using western blot and qRT-PCR (**Figures 2A,B**). We found that the expression of ARID1A was depleted in SNU-16, SNU-5, and SNU-1, and these cell lines were poorly differentiated. Cell lines MKN-28 and AGS were more differentiated, and

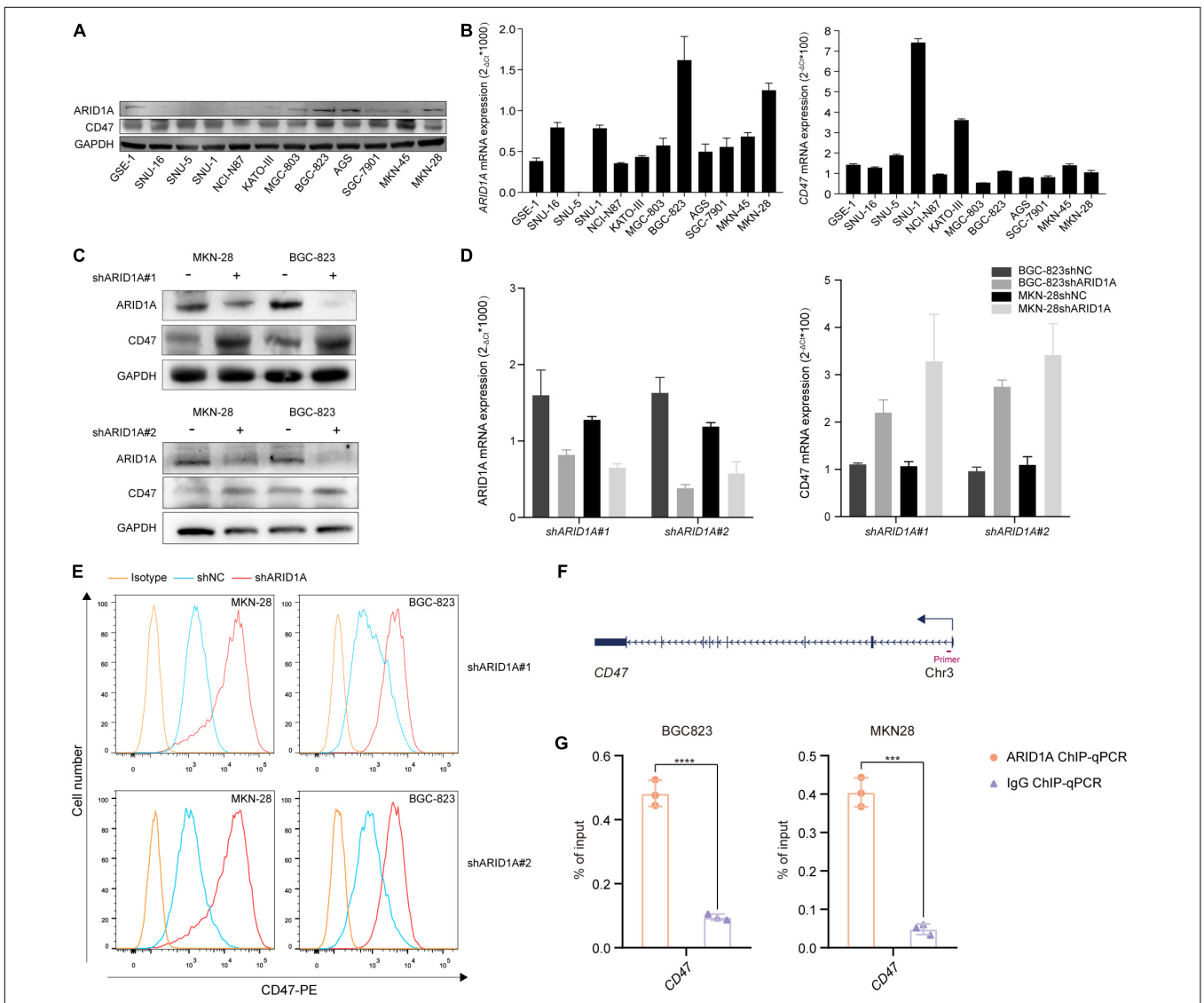


FIGURE 2 | Loss of ARID1A increases CD47 expression. **(A,B)** Protein and mRNA expression levels of ARID1A and CD47 among various GC cell lines. **(C–E)** Knockdown of ARID1A significantly increased total and membranous CD47 protein levels, and *CD47* mRNA level measured by western blot, flow cytometry and qRT-PCR in GC cell lines. **(F)** Schematic illustration of the relative positions of qPCR probes to putative *CD47* promoter region for ChIP-qPCR experiments. **(G)** ARID1A binding to the promoter region of CD47 was determined by ChIP-qPCR in BGC-823 and MKN28 cells, respectively. ***$p < 0.001$, ****$p < 0.0001$.

the expression of ARID1A was relatively higher in them. To study the loss of ARID1A in GC cell lines, we chose BGC-823 and MKN28 cells for their relative abundant expression and found that CD47 expression level was increased at the protein and mRNA level after *ARID1A* knockdown (**Figures 2C,D**). Besides, flow cytometry analysis showed that the expression of membranous CD47 was also significantly increased after *ARID1A* knockdown (**Figure 2E**). Analogous findings were observed using another shRNA. Notably, the extent of the correlation between expressions of ARID1A and CD47 in GC cell lines was the same as that observed in GC tissues, which is indicative of a strong association between loss of ARID1A and a high expression of CD47. To investigate how ARID1A regulates CD47 expression in GC, ChIP-qPCR assay was performed in BGC-823 and MKN28 cell lines, and the results revealed that *CD47* was a direct downstream target gene of ARID1A in GC (**Figures 2F,G**).

ARID1A^{loss}CD47^{high} Expression Indicates Poor Prognosis in Gastric Cancer Patients

To analyze the prognostic value of ARID1A and CD47 in GC, Kaplan–Meier survival analysis and log-rank test were conducted. The endpoint of survival analysis was OS time. As shown in **Figure 3A**, the low expression of CD47 was correlated with a better outcome for GC patients. Meanwhile, loss of expression of ARID1A was associated with a poorer outcome in such patients (**Figure 3B**). The mutation of tumor suppressor gene and a change in the immune microenvironment contribute, to a large extent, to tumor progression. Thus, we combined ARID1A status with CD47 expression to investigate the impact on OS of GC patients. Among all cases, cases in the ARID1A^{preserved}CD47^{low} expression subgroup had the longest OS time (median OS = 53.9 months) than any other subgroups, whereas the outcome of cases in ARID1A^{loss}CD47^{high} expression subgroup was the worst (median OS = 20.4 moths) (**Figure 3C**). It has been found that HER2 is overexpressed in approximately 20% of GC patients, and anti-HER2 targeted therapy significantly improves the prognosis of HER2-positive GC patients (Oh and Bang, 2020). To evaluate the therapeutic value of ARID1A and CD47, HER2 positivity–stratified analysis was conducted. In HER2-negative cases, loss of ARID1A expression or high CD47 expression indicated a poor outcome (**Figures 3F,G**). However, analysis did not reveal a statistically significant difference in the outcome of HER2-positive cases (**Figures 3D,E**). Thus, these results indicate that GC patients with ARID1A^{preserved}CD47^{low} might have a better clinical outcome.

Cox regression analysis was performed to find out independent prognostic factors for the OS of GC. Factors including gender, age, tumor size, histological grade, tumor location, TNM stage, ARID1A status, and CD47 expression were taken into multivariate survival analysis. The results showed that CD47 [multivariate analysis: hazard ratio (HR) = 1.705; 95% confidence interval (CI) = 1.118–2.601; $p < 0.01$] and ARID1A (multivariate analysis: HR = 0.604; 95% CI = 0.374–0.976; $p < 0.05$) were independent prognostic factors for the OS of GC patients (**Figure 4**). In addition, histological grade was also identified as an independent prognostic factor for GC patients.

Lower Infiltration of M2 Macrophage and Regulatory T Cells Was Found in ARID1A^{preserved}CD47^{low} Gastric Cancer Patients

To understand the survival difference in different subgroups stratified by CD47 expression and ARID1A status, we further investigated the tumor-infiltrating immune cell infiltration difference in different subgroups stratified by low/high *CD47* expression and mutated/wild-type *ARID1A* expression by using quanTIseq and EPIC algorithms in the analysis of the data from TCGA. As shown in **Figure 5A**, higher proportion of M1-polarized macrophages were found in the *ARID1A*^{mutated}*CD47*^{high} subgroup than any other subgroups. A previous study showed that high M1-polarized macrophages correlated with a better outcome in GC patients (Zhang et al., 2015). However, a recent study indicated that M1 macrophages have an attenuated prognostic value in *CD47*^{high} GC patients (Shi et al., 2021). Thus, the status of M1 macrophage might lose its prognostic value under high CD47 expression condition in GC. Lower proportions of immunosuppressive regulatory T cells (Tregs) were found in the *ARID1A*^{wildtype}*CD47*^{low} subgroup than any other. Besides, the infiltration proportion of M2-polarized macrophages in *ARID1A*^{mutated}*CD47*^{high} subgroup was significantly higher than that in the *ARID1A*^{wildtype}*CD47*^{low} subgroup. In the *ARID1A*^{wildtype}*CD47*^{low} subgroup, the proportion of CD8^{+} T cells was higher compared with any other subgroups. No difference was found in CD4^{+} T-cell and B-cell infiltration proportions. To further validate the immune cell infiltration implicated by the computational analyses, we carried out multispectral immunofluorescence assay to measure the infiltration of M1 macrophages, M2 macrophages, and Tregs in GC tissues with ARID1A^{loss}CD47^{high} or ARID1A^{preserved}CD47^{low} expression. Consistent with the computational analyses, multispectral immunofluorescence assays showed that higher proportions of M1-polarized macrophages, M2-polarized macrophages, and Tregs were found in the ARID1A^{loss}CD47^{high} subgroup than in the ARID1A^{preserved}CD47^{low} subgroup (**Figures 5B,C**). Putting side-by-side this information, we concluded that better prognosis of GC patients in ARID1A^{preserved}CD47^{low} group might correlate with low Tregs and low M2 macrophages cell infiltration.

DISCUSSION

Because of its durable and robust responses, immune checkpoint therapy has taken oncotherapy into a new era. Although immune checkpoint therapy, including anti–PD-1/PD-L1 and anti–CTLA-4 therapy, has shown great successes in solid tumors (Hodi et al., 2010; Topalian et al., 2012), there are still limitations due to its narrow therapeutic window. CD47 is an emerging target for cancer immune checkpoint therapy by its functioning as a "Don't

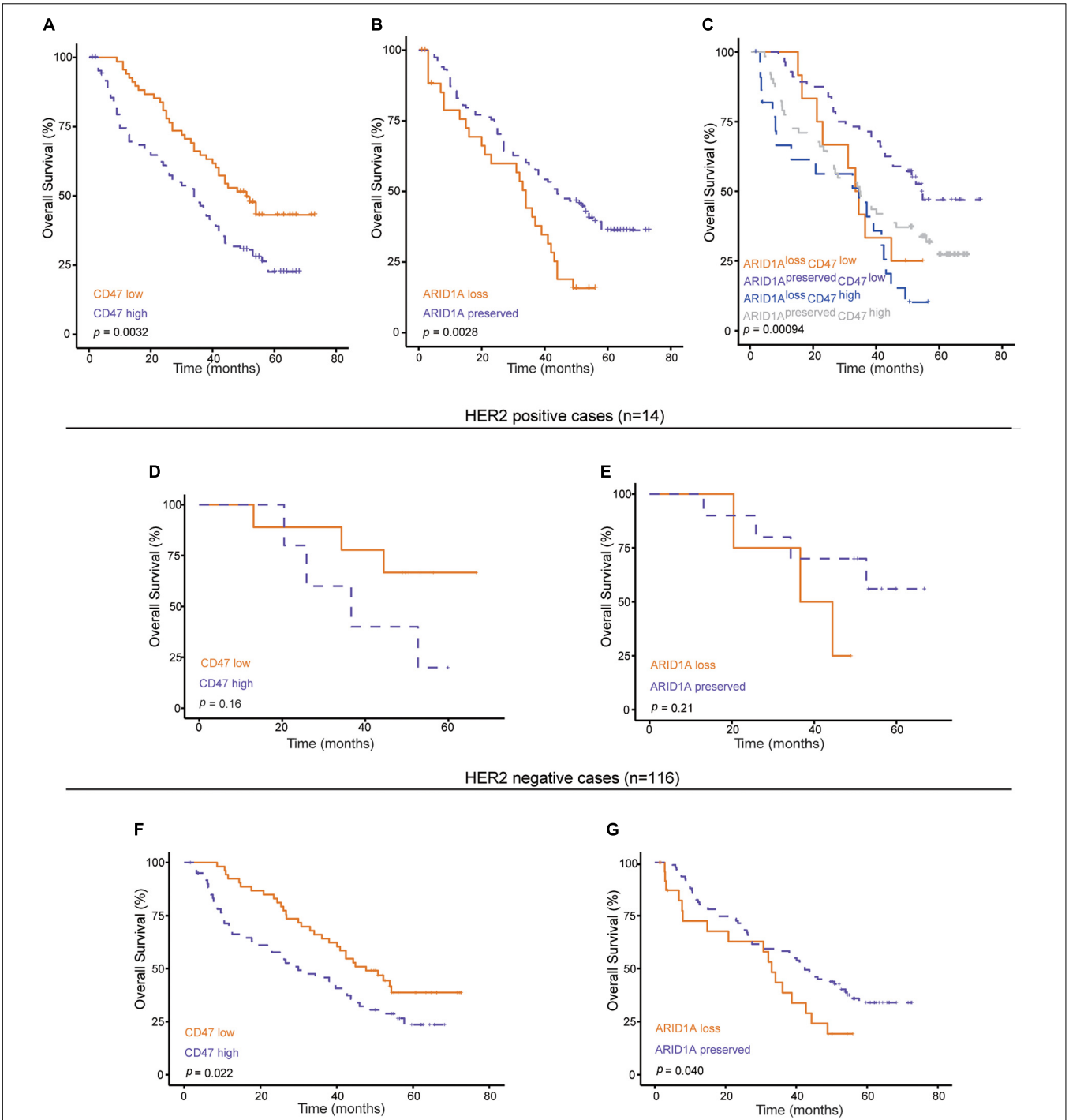


FIGURE 3 | Kaplan–Meier curves under various conditions. **(A,B)** Survival curves according to ARID1A status or CD47 expression (n = 154). **(C)** Survival curves according to ARID1A status combined with CD47 expression level (n = 154). **(D–G)** Survival curves according to ARID1A status or CD47 expression with HER2 positive or negative.

eat me" signal to avoid the phagocytosis by macrophages (Logtenberg et al., 2020). Also, one study showed that $CD8^+$ T cells and dendritic cells could exert antitumor effects through the blocking effect of CD47 (Liu et al., 2015). In line with these findings, we found that high expression of CD47 indicated a poorer outcome and was an independent prognostic factor for GC.

Most *ARID1A* mutations have been shown to be inactivated non-sense mutations or frameshift mutations, which result in the loss of ARID1A protein expression (Wiegand et al., 2010;

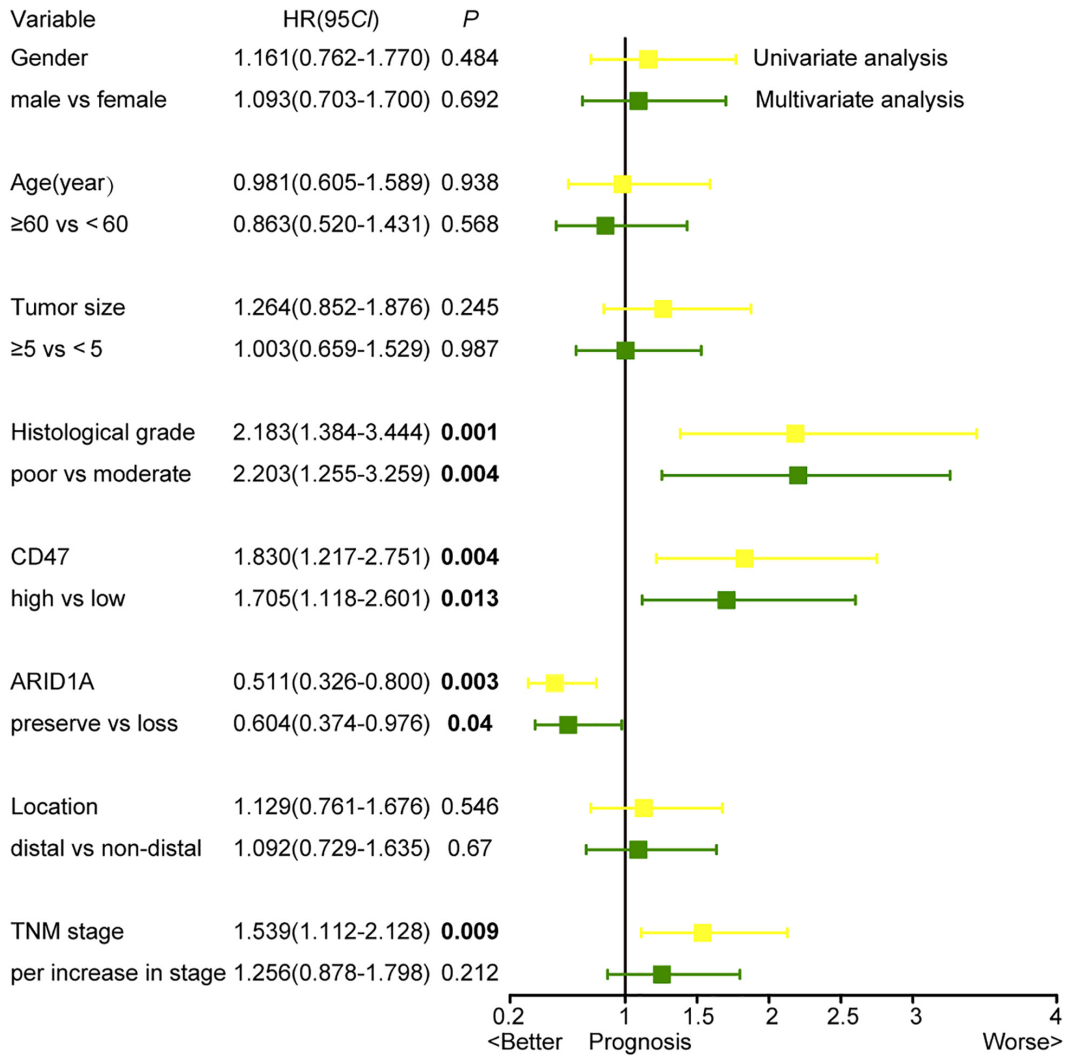


FIGURE 4 | Univariate and multivariate analyses based on clinicopathological characteristics.

Wu and Roberts, 2013; Wu et al., 2014). Thus, we investigated the loss of ARID1A expression in GC tissues. Zhai et al. (2016) found that ARID1A loss impaired the formation of ovarian cancer. Mathur et al. (2017) found that ARID1A loss impaired enhancer-mediated gene regulation and drives colon cancer in mice. In the present article, we found that loss of ARID1A expression was associated with a worse outcome for GC patients and was an independent prognostic factor. In line with our results, Kim et al. (2019) demonstrated that loss of ARID1A expression indicated an inferior outcome in GC patients regardless of OS or recurrence-free survival.

A previous study found that loss of ARID1A could promote tumors to become susceptible to an immune checkpoint inhibitor and could be used as a prognostic biomarker for immune checkpoint therapy (Shen et al., 2018). In terms of mechanism, loss of ARID1A could affect the efficacy of immunotherapy through damaging mismatch repair, promoting tumor mutation, up-regulating the expression of PD-L1 protein, and regulating the immune microenvironment (Shen et al., 2018). Kim et al. (2019) found that loss of ARID1A up-regulated the expression of PD-L1 through activating AKT signaling, and PI3K inhibition could down-regulate PD-L1 expression. In addition, loss of ARID1A could inhibit repayment of DNA single-strand breaks, which up-regulates the expression of PD-L1 (Shen et al., 2015; Sato et al., 2017). Thus, we analyzed the relationship between the promising immune checkpoint CD47 and ARID1A. The results showed that high expression of CD47 was significantly correlated with ARID1A loss/mutation. In GC cell lines, we also found a significant increase of CD47 at protein and mRNA levels after knockdown of ARID1A.

Our analysis on TCGA data showed that the infiltration proportion of M1-polarized macrophage was higher in $ARID1A^{mutated}CD47^{high}$ subgroup than in any other subgroups. Previous research indicated that M1-polarized macrophage played an antitumor role and was correlated with better prognosis in GC (Zhang et al., 2015). Nevertheless, we found

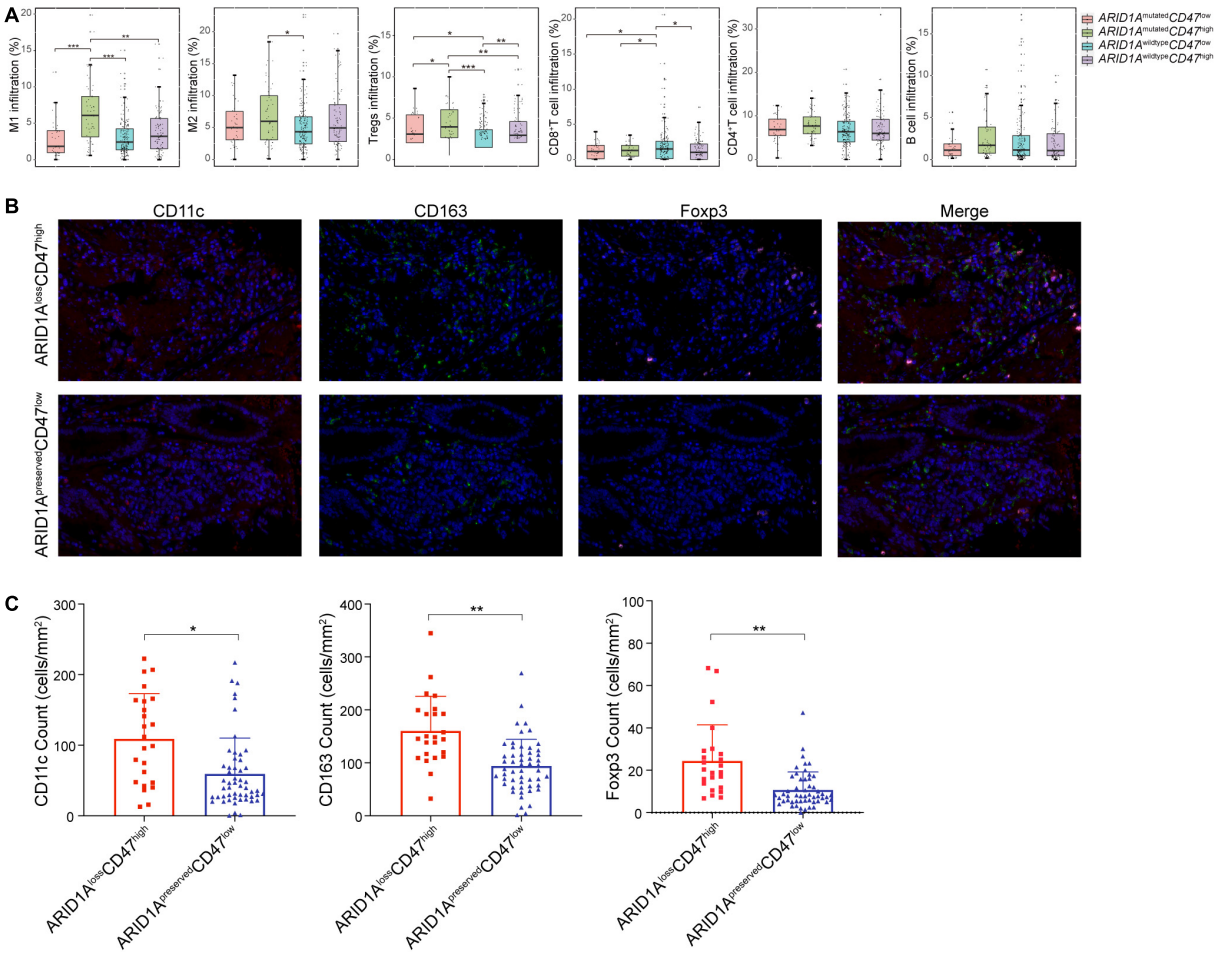


FIGURE 5 | Characterization of tumor-infiltrating immune cells based on ARID1A and CD47 expression. **(A)** Estimation of tumor-infiltrating immune cells in different subgroups stratified by *ARID1A* status and *CD47* expression. **(B)** Multiplex immunofluorescent images illustrate GC tissues with ARID1A^{loss}CD47^{high} or ARID1A^{preserved}CD47^{low} expression that express CD11c, CD163, and Foxp3. **(C)** Quantitative analysis of the densities of immune markers between ARID1A^{loss}CD47^{high} and ARID1A^{preserved}CD47^{low} GC tissues. Error bars represent mean with SD. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

that M1 infiltration failed to predict prognosis, which might be due to the fact that CD47 attenuates the anticancer impact of M1-polarized macrophage (Shi et al., 2021). In addition, one should bear in mind that the prognostic value of immune cells is influenced by many factors, such as cell subpopulation, spatial distribution, cell type, and functional status (Bruni et al., 2020). Also, our finding might be indicative of the loss of function of M1-polarized macrophages in CD47 high expression condition. We also found a higher infiltration proportion of Tregs in *ARID1A*^{mutated}*CD47*^{high} expression subgroup than in any other groups, and a higher proportion of CD8^{+} T cells were infiltrated in *ARID1A*^{wildtype}*CD47*^{low} expression subgroup than in any other one. Multispectral immunofluorescence assay has validated the infiltration proportion of M1 macrophages, M2 macrophages, and Tregs in GC tissues with ARID1A^{loss}CD47^{high} or ARID1A^{preserved}CD47^{low} expression. A previous study reported that CD8^{+} T cells were associated with a good prognosis, and Tregs predicted a poor prognosis (Bruni et al., 2020).

Evidently, it is of value to investigate ARID1A and CD47 expression in GC. Our study is the first to explore the relationship between loss of ARID1A expression and high CD47 expression in GC. Nevertheless, our study bears some limitations, including a not very large sample size, which necessitate conducting further investigations along this line of research.

CONCLUSION

In conclusion, we showed for the first time that loss or mutation of ARID1A was significantly correlated with high CD47 expression in GC. Knockdown of ARID1A significantly increased the expression of CD47 in GC cell lines. With respect to the underlying mechanism in play, we found that ARID1A could bind to the promoter region of *CD47* to regulate its expression. Moreover, we found that ARID1A^{loss}CD47^{high} expression was indicative of a poor prognosis in GC patients, and

combination of ARID1A and CD47 was identified to be a promising prognosis factor in GC.

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 author/s.

ETHICS STATEMENT

Written informed consents were provided by this company and the research was conducted in accordance with recognized ethical guidelines. The studies involving human participants were reviewed and approved by the Ethics Committee of Shanghai Jiao Tong University School of Medicine Affiliated Ruijin Hospital. The patients provided written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CY and JZ: conception and design. QZ, QC, SY, JJ, and ZZ: development of methodology, acquisition of data, and analysis and interpretation of data. QZ and JZ: writing — review of the manuscript. QZ, CY, and JZ: study supervision. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by National Science Foundation of China (81972707) and the Collaborative Innovation Cluster Project of Shanghai Municipal Health Commission (2020CXJQ03).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.745120/full#supplementary-material

REFERENCES

Bruni, D., Angell, H. K., and Galon, J. (2020). The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy. *Nat. Rev. Cancer* 20, 662–680. doi: 10.1038/s41568-020-0285-7

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.

Hodi, F. S., O'Day, S. J., McDermott, D. F., Weber, R. W., Sosman, J. A., Haanen, J. B., et al. (2010). Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. *N. Engl. J. Med.* 363, 711–723.

Ji, J., Zhou, C., Wu, J., Cai, Q., Shi, M., and Zhang, H. (2019). Expression pattern of CDK12 protein in gastric cancer and its positive correlation with CD8(+) cell density and CCL12 expression. *Int. J. Med. Sci.* 16, 1142–1148. doi: 10.7150/ijms.34541

Kim, Y. B., Ahn, J. M., Bae, W. J., Sung, C. O., Lee, D., and Kim, Y. B. (2019). Functional loss of ARID1A is tightly associated with high PD-L1 expression in gastric cancer. *Int. J. Cancer* 145, 916–926. doi: 10.1002/ijc.32140

Kim, Y. B., Ham, I. H., Hur, H., Lee, D., and Kim, Y. B. (2016). Various ARID1A expression patterns and their clinical significance in gastric cancers. *Hum. Pathol.* 49, 61–70. doi: 10.1016/j.humpath.2015.10.008

Li, X., Pasche, B., Zhang, W., and Chen, K. (2018). Association of MUC16 Mutation With Tumor Mutation Load and Outcomes in Patients With Gastric Cancer. *JAMA Oncol.* 4, 1691–1698. doi: 10.1001/jamaoncol.2018.2805

Lin, C., Hongyong, H., Hao, L., Ruochen, L., and Yifan, C. (2019). Tumour-associated macrophages-derived CXCL8 determines immune evasion through autonomous PD-L1 expression in gastric cancer. *Gut* 68, 1764–1773. doi: 10.1136/gutjnl-2018-316324

Liu, X., Pu, Y., Cron, K., Deng, L., Kline, J., Frazier, W. A., et al. (2015). CD47 blockade triggers T cell-mediated destruction of immunogenic tumors. *Nat. Med.* 21, 1209–1215. doi: 10.1038/nm.3931

Lo, Y. H., Kolahi, K. S., Du, Y., Krokhotin, A., Nair, A., Sobba, W. D., et al. (2021). A CRISPR/Cas9-Engineered ARID1A-Deficient Human Gastric Cancer Organoid Model Reveals Essential and Nonessential Modes of Oncogenic Transformation. *Cancer Discov.* 11, 1562–1581. doi: 10.1158/2159-8290.cd-20-1109

Logtenberg, M. E. W., Scheeren, F. A., and Schumacher, T. N. (2020). The CD47-SIRPalpha Immune Checkpoint. *Immunity* 52, 742–752. doi: 10.1016/j.immuni.2020.04.011

Lu, C., and Allis, C. D. (2017). SWI/SNF complex in cancer. *Nat. Genet.* 49, 178–179.

Mathur, R., Alver, B. H., San Roman, A. K., Wilson, B. G., Wang, X., Agoston, A. T., et al. (2017). ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice. *Nat. Genet.* 49, 296–302. doi: 10.1038/ng.3744

Morrissey, M. A., Kern, N., and Vale, R. D. (2020). CD47 Ligation Repositions the Inhibitory Receptor SIRPA to Suppress Integrin Activation and Phagocytosis. *Immunity* 53, 290–302e6.

Nagarajan, N., Bertrand, D., Hillmer, A. M., Zang, Z. J., Yao, F., Jacques, P., et al. (2012). Whole-genome reconstruction and mutational signatures in gastric cancer. *Genome Biol.* 13:R115.

Oh, D. Y., and Bang, Y. J. (2020). HER2-targeted therapies - a role beyond breast cancer. *Nat. Rev. Clin. Oncol.* 17, 33–48. doi: 10.1038/s41571-019-0268-3

Paschalis, A., Sheehan, B., Riisnaes, R., Rodrigues, D. N., Gurel, B., Bertan, C., et al. (2019). Prostate-specific Membrane Antigen Heterogeneity and DNA Repair Defects in Prostate Cancer. *Eur. Urol.* 76, 469–478.

Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E., Gfeller, D., Racle, J., et al. (2017). Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife* 2017:6.

Roberts, C. W., and Orkin, S. H. (2004). The SWI/SNF complex–chromatin and cancer. *Nat. Rev. Cancer* 4, 133–142.

Sato, H., Niimi, A., Yasuhara, T., and Permata, M. (2017). DNA double-strand break repair pathway regulates PD-L1 expression in cancer cells. *Nat. Commun.* 8:1751.

Shen, J., Ju, Z., Zhao, W., Wang, L., Peng, Y., and Ge, Z. (2018). ARID1A deficiency promotes mutability and potentiates therapeutic antitumor immunity unleashed by immune checkpoint blockade. *Nat. Med.* 24, 556–562. doi: 10.1038/s41591-018-0012-z

Shen, J., Peng, Y., Wei, L., Zhang, W., Yang, L., Lan, L., et al. (2015). ARID1A Deficiency Impairs the DNA Damage Checkpoint and Sensitizes Cells to PARP Inhibitors. *Cancer Discov.* 5, 752–767. doi: 10.1158/2159-8290.cd-14-0849

Shi, M., Gu, Y., Jin, K., Fang, H., Chen, Y., Cao, Y., et al. (2021). CD47 expression in gastric cancer clinical correlates and association with macrophage infiltration. *Can. Immunol. Immunother.* 70, 1831–1840. doi: 10.1007/s00262-020-02806-2

Sun, X., Wang, S. C., Luo, X., Jia, Y., Li, L., Gopal, P., et al. (2017). Arid1a Has Context-Dependent Oncogenic and Tumor Suppressor Functions in Liver Cancer. *Cancer Cell* 32, 574–589e6.

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., and Jemal, A. (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. doi: 10.3322/caac.21660

Topalian, S. L., Hodi, F. S., Brahmer, J. R., Gettinger, S. N., Smith, D. C., McDermott, D. F., et al. (2012). Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N. Engl. J. Med.* 366, 2443–2454.

Vonderheide, R. H. (2015). CD47 blockade as another immune checkpoint therapy for cancer. *Nat. Med.* 21, 1122–1123. doi: 10.1038/nm.3965

Wang, C., Shi, M., Ji, J., Cai, Q., Jiang, J., Zhang, H., et al. (2019). A self-enforcing HOXA11/Stat3 feedback loop promotes stemness properties and peritoneal metastasis in gastric cancer cells. *Theranostics* 9, 7628–7647. doi: 10.7150/thno.36277

Wang, X., Nagl, N. G. Jr., Flowers, S., Zweitzig, D., Dallas, P. B., Moran, E., et al. (2004). Expression of p270 (ARID1A), a component of human SWI/SNF complexes, in human tumors. *Int. J. Cancer* 112:636. doi: 10.1002/ijc.20450

Wiegand, K. C., Shah, S. P., Al-Agha, O. M., Zhao, Y., Tse, K., Zeng, T., et al. (2010). ARID1A mutations in endometriosis-associated ovarian carcinomas. *N. Engl. J. Med.* 363, 1532–1543.

Wu, J. N., and Roberts, C. W. (2013). ARID1A mutations in cancer: another epigenetic tumor suppressor? *Cancer Discov.* 3, 35–43. doi: 10.1158/2159-8290.cd-12-0361

Wu, R. C., Wang, T. L., and Shih Ie, M. (2014). The emerging roles of ARID1A in tumor suppression. *Cancer Biol. Ther.* 15, 655–664. doi: 10.4161/cbt.28411

Yu, S. H., Zhu, K.-Y., Chen, J., X-Z, Liu, P-F, Xu, Zhang, W., et al. (2018). JMJD3 facilitates C/EBPβ-centered transcriptional program to exert oncorepressor activity in AML. *Nat. Commun.* 9:3369.

Zhai, Y., Kuick, R., Tipton, C., Wu, R., Sessine, M., Wang, Z., et al. (2016). Arid1a inactivation in an Apc- and Pten-defective mouse ovarian cancer model enhances epithelial differentiation and prolongs survival. *J. Pathol.* 238, 21–30. doi: 10.1002/path.4599

Zhang, H., Wang, X., Shen, Z., Xu, J., Qin, J., Sun, Y., et al. (2015). Infiltration of diametrically polarized macrophages predicts overall survival of patients with gastric cancer after surgical resection. *Gastric. Can.* 18, 740–750. doi: 10.1007/s10120-014-0422-7

Zhang, S., Suna, K., Zheng, R., Zenga, H., Wang, S., Chen, R., et al. (2021). Cancer incidence and mortality in China, 2015. *J. Natl. Can. Cent.* 1, 2–11.

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: 05 November 2021
doi: 10.3389/fcell.2021.758339

Papillary Thyroid Carcinoma Landscape and Its Immunological Link With Hashimoto Thyroiditis at Single-Cell Resolution

Jun Pan[1†], Fang Ye[2*†], Chengxuan Yu[2†], Qinsheng Zhu[1†], Jiaqi Li[2], Yaohui Zhang[1], Hedi Tian[1], Yunjin Yao[1], Minjie Zhu[3], Yibin Shen[1], Feng Zhu[1], Yingying Wang[4], Xinhui Zhou[5], Guoji Guo[2,6,7,8,9*] and Yijun Wu[1*]*

[1] Department of Thyroid Surgery, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China, [2] Center for Stem Cell and Regenerative Medicine, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China, [3] Surgical Department, Hangzhou Third Hospital, Hangzhou, China, [4] Kidney Disease Center, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China, [5] Department of Gynecology, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China, [6] Liangzhu Laboratory, Zhejiang University Medical Center, Hangzhou, China, [7] Zhejiang Provincial Key Laboratory for Tissue Engineering and Regenerative Medicine, Dr. Li Dak Sum & Yip Yio Chin Center for Stem Cell and Regenerative Medicine, Hangzhou, China, [8] Institute of Hematology, Zhejiang University, Hangzhou, China, [9] Stem Cell Institute, Zhejiang University, Hangzhou, China

The tumor microenvironment heterogeneity of papillary thyroid cancer (PTC) is poorly characterized. The relationship between PTC and Hashimoto thyroiditis (HT) is also in doubt. Here, we used single-cell RNA sequencing to map the transcriptome landscape of PTC from eight PTC patients, of which three were concurrent with HT. Predicted copy number variation in epithelial cells and mesenchymal cells revealed the distinct molecular signatures of carcinoma cells. Carcinoma cells demonstrated intertumoral heterogeneity based on *BRAF* V600E mutation or lymph node metastasis, and some altered genes were identified to be correlated with disease-free survival in The Cancer Genome Atlas datasets. In addition, transcription factor regulons of follicular epithelial cells unveil the different transcription activation state in PTC patients with or without concurrent HT. The immune cells in tumors exhibited distinct transcriptional states, and the presence of tumor-infiltrating B lymphocytes was predominantly linked to concurrent HT origin. Trajectory analysis of B cells and plasma cells suggested their migration potential from HT adjacent tissues to tumor tissues. Furthermore, we revealed diverse ligand–receptor pairs between non-immune cells, infiltrating myeloid cells, and lymphocytes. Our results provided a single-cell landscape of human PTC. These data would deepen the understanding of PTC, as well as the immunological link between PTC and HT.

Keywords: papillary thyroid cancer, Hashimoto's thyroiditis, single-cell RNA sequencing, tumor B-cell infiltration, immunological link

Abbreviations: CNV, copy number variations; DCs, dendritic cells; DEGs, differentially expressed genes; DFS, disease-free survival; EMT, epithelial–mesenchymal transition; GSVA, gene set variation analysis; HT, Hashimoto thyroiditis; LNM, lymph node metastasis; NK, natural killer cells; NKT, natural killer T cells; PTC, papillary thyroid cancer; scRNA-seq, single-cell RNA sequencing; RFS, recurrence-free Survival; TF, transcription factor; TME, tumor microenvironment; Treg, regulatory T cells; UMAP, Uniform Manifold Approximation and Projection.

OPEN ACCESS

Edited by:
Geng Chen,
GeneCast Biotechnology Co., Ltd., China

Reviewed by:
Nelson George,
Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), United States
Fei Shen,
South China University of Technology, China

***Correspondence:**
Jun Pan
panjun1928@aliyun.com
Fang Ye
ye-fang@zju.edu.cn
Guoji Guo
ggj@zju.edu.cn
Yijun Wu
wuwu5925@zju.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: 13 August 2021
Accepted: 11 October 2021
Published: 05 November 2021

Citation:
Pan J, Ye F, Yu C, Zhu Q, Li J, Zhang Y, Tian H, Yao Y, Zhu M, Shen Y, Zhu F, Wang Y, Zhou X, Guo G and Wu Y (2021) Papillary Thyroid Carcinoma Landscape and Its Immunological Link With Hashimoto Thyroiditis at Single-Cell Resolution. Front. Cell Dev. Biol. 9:758339. doi: 10.3389/fcell.2021.758339

INTRODUCTION

Thyroid cancer is the most common endocrine malignancy worldwide, with significant increases in incidence around the world over the past three decades (Pellegriti et al., 2013). The global incidence rate in women is 10.2 per 100,000, and that in men is 3.1 per 100,000. This disease represents 5.1% of all estimated female cancer burden (Bray et al., 2018). Papillary thyroid cancer (PTC) is the most prevalent subtype of thyroid cancer, accounting for 70 to 85.9% of the incidence of all thyroid cancers (Kwak et al., 2010). Although PTC is generally indolent and shows a favorable prognosis, some metastasized lesions are not treatable with radioactive iodine or surgery therapy.

With the advances in genomic and transcriptome sequencing technologies, some potential marker genes for PTC diagnosis and progression have been identified. *BRAF/RAS* gene alteration matched treatment represents the gene-targeted therapy. Besides, gene expression–based molecular subtyping of PTC could be potentially used as a prognostic tool with important clinical relevance. *BRAF* V600E mutation has the strong predictive power for the diagnosis and aggressive pathological results of PTC (Kebebew et al., 2007; Xing et al., 2014; Wang et al., 2016). However, conventional genomic and gene expression profiling is applied to characterize a bulk tumor and ignore the intertumoral heterogeneity. PTC is a highly heterogeneous tumor composed of diverse populations of lymphocytes and myeloid cells that affect tumor initiation, progression, and treatment resistance. Some studies have reported a correlation between positive or negative clinical outcomes in thyroid cancer patients and tumor-infiltrating immune cells (Gooden et al., 2011; Gentles et al., 2015). Myeloid-derived suppressor cells are elevated in cancer patients, where they show a strong immunosuppressive potential, and are associated with a poor prognosis (Gabrilovich and Nagaraj, 2009). The frequency of tumor-associated macrophages (TAMs) changes in the distinct subtypes of thyroid cancer (Jung et al., 2015). Natural killer (NK) cells also play a central role in thyroid cancer immunosurveillance (Lee and Sunwoo, 2019). The density of lymphocytes has been reported to be correlated with improved overall survival and lower recurrences of PTC (Kuo et al., 2017). Thus, it is important to characterize tumor cells as well as immune cells to clarify their property in PTC.

Papillary thyroid cancer displayed a slow growth and lymphatic spread with rarely distant metastasis, which might be due to the frequent presence of lymphatic infiltration into tumor sites. It is reported that 18.9 to 23.2% of the PTC is concurrent with Hashimoto thyroiditis (HT), and PTC patients with synchronous appearance of HT had better prognosis compared with those without HT (Ahn et al., 2011; Konturek et al., 2013; Lee et al., 2013; Zhang et al., 2014; Resende de Paiva et al., 2017). In PTC patients with or without concurrent HT, whether tumor-infiltration lymphocytes are attracted by an antitumor immune response or influenced by a preexisting autoimmune process remains unknown. At least, an immunological link between PTC and HT could not be excluded, but the role of HT in shaping the PTC immune milieu is still unclear.

Since the first single-cell RNA sequencing (scRNA-seq) technique was published in 2009 (Tang et al., 2009), scRNA-seq has gradually become a cutting-edge method to resolve intratumor heterogeneity. Single-cell transcriptomes provided new insights into cell–cell interactions in multiple types of cancer (Zhang et al., 2019; Davis et al., 2020; Wieland et al., 2020). Some molecular studies reported the advances in follicular thyroid carcinoma and medullary thyroid carcinoma (Ceolin et al., 2012; Mohammadi and Hedayati, 2017; Borowczyk et al., 2019; Oczko-Wojciechowska et al., 2020). A recent scRNA-seq study made a comparison of tumor microenvironment (TME) in PTC between genders (Peng et al., 2021). Another most recent study delineated dedifferentiation process of anaplastic thyroid cancer and PTC at a single-cell level (Luo et al., 2021). In this study, took the advantage of Microwell-seq, a low-cost and high-throughput scRNA-seq platform that has been reported in our previous studies (Han et al., 2018, 2020), we constructed the single-cell transcriptome landscape of human papillary thyroid carcinoma. The systemic single-cell transcriptome data provided novel insights to understand the TME of PTC. We explored cell composition, functional states, cellular interactions in PTC tumors, and adjacent thyroid tissues (with HT). Specifically, previously uncharacterized HT, which PTC is often concurrent with, was included in this study to investigate the dynamic relationship of immunocytes. Taken together, the precise characterization of PTC and its microenvironment, in combination with the immunological crosstalk with HT, facilitates in-depth understanding of the PTC molecular characteristics. These results would also help in the identification of potential molecular targets for PTC diagnosis and treatment.

MATERIALS AND METHODS

Sample Collection

This study was approved by the Clinical Research Ethics Committee of the First Affiliated Hospital, School of Medicine, Zhejiang University (IIT consent: no. 700,2020), and carried out in accordance with the principles of the Declaration of Helsinki. Eight patients, including three concurrent with HT, diagnosed with PTC were recruited in our study, and all the patients signed the informed written consent for each subject and agreed to donate the specimens. A total of 10 fresh tissue samples (five samples from PTC patients without concurrent HT, three samples from PTC patients with concurrent HT, and two paired adjacent tissues of two PTC patients with concurrent HT) were collected from Chinese PTC patients undergoing thyroidectomy at the Department of Thyroid Surgery of the hospital. Paired adjacent tissues were collected by curettage at the same time as tumor tissue collection. The patients did not receive any other forms of therapy. Diagnosis of PTC and HT cases was histologically confirmed by two independent pathologists, and all of the tumor tissues were assessed by hematoxylin-eosin staining.

Single-Cell Suspension Preparation

Fresh and sterile tumor tissue fragments were initially divided into segments, after two washings with 1 × phosphate-buffered saline, the tumor pieces were dissociated into single-cell

suspensions through Human Tumor Dissociation Kit (Miltenyi Biotec GmbH, Bergisch Gladbach, Germany) according to the kit manuals. Digested tumor pieces were teased through a 40-μm sieve. Then, the dissociated single cells were centrifuged, and cell pellets were resuspended in PRIM1640 (Thermo Fisher Scientific) plus 0.04% bovine serum albumin (Sigma–Aldrich). Viability was confirmed to be >90% in all samples *via* trypan blue (Thermo Fisher Scientific) staining, and the cell suspensions were kept on ice for the scRNA-seq.

Single-Cell RNA Library Preparation and Sequencing

Standard Microwell-seq protocol was performed to treat single-cell suspensions from different samples. In brief, single-cell suspensions and barcode beads were loaded on agarose Microwell array. Beads and cells were trapped in separated Microwells. Transcripts from lysed cell were captured by barcode oligodT bead. Beads were collected in a 1.5-mL tube to do template switch, reverse transcription, exonuclease I treatment, and cDNA amplification. Purified cDNA libraries were tagmented using a customized transposase to enrich 3' ends of transcripts (TruePrep DNA Library Prep Kit V2 for Illumina, Vazyme, cat. no. TD513), Libraries were sequenced on Illumina Hiseq Xten (PE150 mode) by Novogene Co., Ltd., Beijing, China.

Single-Cell RNA Sequencing Data Processing

Drop-seq core computational tool (version 1.12) was used to preprocess the Microwell-seq raw data. As described in Drop-seq computational cookbook[1]. Online R packages of data preprocessing and detailed parameters are available at Github[2]. Filtered reads were used to extract cellular barcode and unique molecular identifier (UMI). We discarded the paired reads if the quality of any base in the barcode was below 10. STAR (version 2.5.2a) with default parameters was used for mapping. Reads were aligned to the *Homo sapiens* GRCh38 reference genome. All multialigned reads were removed, and GTF annotation files from GENCODE were used to tag aligned reads. For UMI count, molecular barcodes with one edit distance were merged to one within a gene. We excluded cells in which there were fewer than 500 UMIs. All the R packages were loaded in R (version 3.6.3), and plots were mapped using R package ggplot2 (version 3.3.5).

Cell Type Clustering and Annotation

After obtaining the digital gene expression data matrix, we used R package Seurat (version 3) for dimension reduction. Preprocessed cells and genes expressed in more than three cells were selected for further clustering and differential gene expression analysis[3]. Filtered data were ln (CPM/100 + 1) transformed. We selected the top 2,000 highly variable genes and top 20 principal components for clustering. Default Wilcoxon rank sum test was used by running FindAllMarkers function in Seurat to find differentially expressed markers in each cluster (min.pct = 0.25, logfc.threshold = 0.25). Uniform Manifold Approximation and Projection (UMAP) was adopted for cluster visualization. We compared known markers with human cell landscape database[4] to achieve cell type annotation.

The Cancer Genome Atlas THCA Dataset Validation

Data Processing and Single-Cell Subsets Correlation Calculation

Transcriptome data from The Cancer Genome Atlas (TCGA) THCA datasets were downloaded from UCSC XENA[5]. Genes with average logFC > 1 were used as marker genes of each cell type. We used Spearman correlation analysis to estimate correlation between immune cell types.

Expression Analyzing of Selected Genes

A total of 568 cases with gene expression data (HTSeq-counts and HTSeq-FPKM) in THCA projects were collected from TCGA. Among them, those with clinical information were included. Then, counts and FPKM data were transformed into TPM for the following analyses. The TPM data for 500 patients were used for further analyses. All statistical analysis and plots in this validation part were produced using R (v4.0.3). Wilcoxon rank sum test and signed rank test were used to analyze the expression of selected genes in PTC samples and PTC combined with HT samples.

Survival Analysis of Selected Genes

To determine the best cutoff value of selected genes to predict disease recurrence in PTC patients, the X-tile software was used. Then, the 500 samples were divided into two groups, high-expression group and low-expression group according to the best cutoff value. Next, Kaplan–Meier method was applied to conduct the survival analysis and plot the survival curves of selected genes. In all tests, $p < 0.05$ was considered statistically significant.

Tumor-Infiltrating Immune Cell Profile

CIBERSORT computational method was applied for estimating the tumor-infiltrating immune cells abundance profile in all 500 samples. The profile of 21 types of immune cells was displayed by boxplot.

RNA Velocity Analysis

Velocyte (version 0.17) was used to calculate RNA velocity of B cell in samples from PTC patients with HT and adjacent tissue samples from PTC patients with HT. The rates of transcriptional changes of each cell were estimated using the ration of spliced and unspliced reads[6] with default parameters. The plot was visualized with UMAP embedding. The differentiation start and end points were estimated using a Markov process with default parameters.

[1]http://mccarrolllab.com/wp-content/uploads/2016/03/Drop-seqAlignmentCookbookv1.2Jan2016.pdf

[2]https://github.com/ggjlab/mca_data_analysis/tree/master/preprocessing

[3]https://satijalab.org/seurat/

[4]https://db.cngb.org/HCL/index.html

[5]https://xena.ucsc.edu/

[6]http://velocyto.org/velocyto.py/tutorial/index.html

Single-Cell Pseudotime Trajectory Analysis

We selected differentially expressed genes (DEGs) of immune cells (B cells, plasma cells). We use monocle2 R package (version 2.4.0) to treat genes expressed in at least three cells in single-cell data. Default settings of DEGs were adopted to construct pseudotime trajectory and heatmap.

Receptor–Ligand Pairing Analysis

We used CellPhoneDB (version 2.1.3) for the analysis of potential receptor–ligand pairings. We aggregated the gene expression levels of all clusters in samples from PTC patients with HT and adjacent tissue from PTC patients with HT. Receptors and ligands expressed in more than 10% of the cells in each cluster were considered. The cutoff was set with the mean expression greater than 0.05 and p values smaller than 0.05. We used the sum of the number of receptor–ligand pairs in each cell–cell pairing to indicate the strength of the cell–cell interactions. The interaction network was visualized using Cytoscape (version 3.7.0) and ggplot2 (version 3.3.5).

Metric Learning

Metric learning is a machine learning model for the purpose of constructing distance metrics from supervised data, and the constructed distance metrics could be used for further clustering, classification, and information retrieval. Here, we used PyTorch Metric Learning tool (version 0.9.90)[7] for designing a binary classifier and applied it to predict the source of follicular epithelial cells (PTC patients with concurrent HT or PTC patients without concurrent HT; by sampling triplets with default parameters). Training sets and validation sets were randomly selected and generated from our scRNA-seq data. To extract and explore genes that have important contributions to classification, we used Captum (version 0.4.0)[8] for model interpretation and understanding. Finally, an R package pheatmap was applied to visualize the gene importance in prediction of the source of follicular epithelial cells.

Immunohistochemistry

Immunohistochemistry (IHC) staining was performed on 4-μm-thick, formalin-fixed, paraffin-embedded sections using an Opal multiplex IHC system (NEL811001KT, PerkinElmer) according to the manufacturer's instructions. For IHC, tissue sections were subjected to antigen retrieval in an induction cooker for 25 min in EDTA buffer (pH 9.0). Followed by treatment with goat serum at 37°C for 40 min, tissue sections were incubated with the following antibodies: TG (ab151539), FOS (ab184938), JUN (ab178858), TFF3 (ab109104), CD55 (ab133684), CCDC80 (ab224050), CD79A/B (ab79414), and CD3D (ab109531) at 4°C overnight. All antibodies were from Abcam (Cambridge, MA, United States). Images were recorded with Metamorph software v7.5.6.0 (Molecular Device) on an Olympus IX81 inverted microscope. The images were evaluated by two independent pathologists who were blinded to the patients' clinical information.

[7]https://github.com/KevinMusgrave/pytorch-metric-learning

[8]https://captum.ai/

RESULTS

Single-Cell Profiling of Papillary Thyroid Cancer

To construct the single-cell landscape of PTC, we applied Microwell-seq to study carcinoma and non-carcinoma microenvironment cells of tumor samples from eight PTC patients diagnosed histologically, with three patients concurrent with HT. Clinical information of all patients is shown in **Table 1**. Diagnosis of all of the PTC patients [including lymph node metastasis (LNM), HT, and *BRAF* V600E mutation] was histologically confirmed by two independent pathologists. Tumor samples from PTC patients and adjacent tissue samples from PTC patients with concurrent HT were digested into single cells for standard workflow of Microwell-seq (**Figure 1A**). After removing low-quality cells, we generated 14,252 cells in eight tumor samples from PTC patients and 9,178 cells in two adjacent tissue samples (HT tissues) from PTC patients with concurrent HT for downstream clustering. The aggregated single-cell transcriptome landscape of eight tumor samples (five samples from PTC patients without concurrent HT, three samples from PTC patients with concurrent HT) was visualized by UMAP plot (**Figure 1B**). We annotated 24 major clusters based on canonical marker genes, including T cells, B cells, plasma cells, macrophages, dendritic cells (DCs), neutrophils, mast cells, endothelial cells, smooth muscle cells, mesenchymal cells, and follicular epithelial cells. Heatmap showed the differentially expressed marker genes of each cluster (**Figure 1C**). Sample proportion in each cell type is shown (**Figures 1D,F**). We kept the original heterogeneity of patients without using any batch remove algorithm. Notably, plasma cells specifically expressing immunoglobulin and XBP1 and B cells specifically expressing CD79A and MS4A1 were restricted in samples from PTC patients with concurrent HT (**Figures 1E,G**). Besides B cells and plasma cells, PTC patients with concurrent HT and PTC patients without concurrent HT shared the most common immune cells. We identified six follicular epithelial cell subsets marked by TG, EPCAM, and malignant genes such as FOS and JUN. Those follicular epithelial cell subsets could also be distinguished by specific marker genes in each cluster. Most follicular epithelial cells_ZCCHC12 high, follicular epithelial cells_C2orf40 high, follicular epithelial cells_TFF3 high, and follicular epithelial cells_PRSS2 high were contributed by samples from PTC patients without concurrent HT (**Figure 1E**).

Intrinsic Parenchymal Cell Heterogeneity of Papillary Thyroid Cancer

We merged parenchymal cell datasets (epithelial cells, stromal cells, mesenchymal cells, muscle cells, and endothelial cells) from eight patients (**Figure 2A**). Most clusters from PTC patients with or without concurrent HT were labeled by their original annotations. It was speculated that tumor

TABLE 1 | Clinical characteristics of patients.

Patient	Gender	Tumor size(cm)	With LNM	With Hashimoto thyroiditis	*BRAF* V600E (mutation:+, wild type: –)	TSH (μIU/mL)[a]	T4 (nmol/L)[b]	T3 (nmol/L)[c]	FT (pmol/L)[d]	FT3 (pmol/L)[e]	TPOAB (IU/mL)[f]	TGAB (IU/mL)[g]
HT_PTC_1	Female	1.15 × 0.49	–	+	+	1.864	112.59	1.42	15.7	4.71	0.46	65.98
PTC_2	Female	1.4 × 1.0	+	–	+	0.553	98.25	1.2	16.78	3.56	0.81	2.08
PTC_3	Female	1.4 × 0.9	+	–	+	3.111	78.99	1.44	12.01	4.24	0.34	2.22
PTC_4	Male	0.7 × 0.4	–	–	–	2.077	88.42	1.68	12.96	4.83	<0.50	1.23
PTC_5	Female	1.1 × 0.8	+	–	+	1.581	81.5	1.53	12.73	4.4	<0.50	0.71
HT_PTC_6	Female	1.5 × 0.8	+	+	+	1.89	131.3	2.28	15.29	6.76	98.8	74.9
PTC_7	Male	1.4 × 1.0 × 0.8	+	–	+	0.794	102.7	1.71	12.22	5.41	<0.50	NA
HT_PTC_8	Female	1.8 × 1.1	+	+	–	0.763	100.6	1.46	12.46	4.49	100.57	490.71

Abbreviations: [a]Thyroid-stimulating hormone. [b]Thyroxine. [c]Triiodothyronine. [d]Free thyroxine. [e]Free triiodothyronine. [f]Thyroid peroxidase antibody. [g]Thyroglobulin antibody.

cells and non-tumor cells were the origin of intratumoral heterogeneity due to the copy number variations (CNVs). Thus, we adopted CopyKAT (Gao et al., 2021) to evaluate the malignancy of follicular epithelial cells and mesenchymal cells based on genomic copy number estimation. Follicular epithelial cells_ ZCCHC12 high, mesenchymal cells_CXCL14 high, and follicular epithelial cells_PRSS2 high harbored highly CNV levels and contained most malignant cells (**Figures 2A,C**). Most predicted tumor cells showed gene amplification in chromosomes 14 and 15 (**Supplementary Figure 1A**). We then examined the expression patterns between predicted normal cells and predicted tumor cells (**Figure 2B**). Malignant mesenchymal cells exhibited high expression levels of CD55, and malignant epithelial cells exhibited high expression levels of TFF3. Gene set variation analysis of pathway demonstrated epithelial–mesenchymal transition (EMT), G2M checkpoint, E2F targets, MYC targets, and P53 pathway were specifically enriched in predicted tumor group with high CNV levels, whereas inflammatory response pathway was down-regulated (**Figure 2E**). These results unveiled an activated EMT state in tumor cell microenvironment of PTC, while immunogenicity of those tumor cells was inhibited.

As an important diagnostic and prognostic molecular marker, we detected the *BRAF* V600E mutation of all patients (Xing, 2010). The LNM was also regarded as an important prognostic factor. Follicular epithelial cells were grouped based on *BRAF* V600E mutation and LNM, respectively, (**Figure 2D**). DEGs in *BRAF* V600E mutation group and LNM group may associate with a poor disease-free survival (DFS) in PTC. Among single-cell DEGs in *BRAF* V600E mutation group, TACSTD2 and CLDN3 showed high expression levels in samples from TCGA THCA cohort (**Figure 2F**). TACSTD2 and CLDN3 were also significantly related to poor DFS in PTC patients. The expression levels of TACSTD2 was correlated with TNM *via* controlling up-regulation of MMP2 (Guan et al., 2017). The overexpression of CLDN3 was also considered to be correlated with thyroid cancer and other malignancies (Hewitt et al., 2006; Hess et al., 2011). In LNM-positive group, epithelial cells showed unique expression of CTSC and B2M that were related to poor DFS in PTC patients (**Figure 2G**; Wu et al., 2018).

Reclustering of Follicular Epithelial Cells

To further profile follicular epithelial cells from PTC patients, we performed unsupervised reclustering of all epithelial cells and generated seven clusters that could be annotated based on their original definition (**Figure 3A** and **Supplementary Figure 2A**). Clusters 3 and 6 were from PTC patients with concurrent HT. Other clusters were contributed by PTC patients without concurrent HT (**Figure 3B**). Clusters in samples from PTC patients without concurrent HT exhibited higher expression levels of TG (**Figure 3C**). CCDC80 was restrictedly expressed in follicular epithelial cells in samples from PTC patients with concurrent HT (**Figure 3A**). It has been reported that in human thyroid carcinoma, CCDC80 played the role as a putative tumor suppressor gene (Ferraro et al., 2013). CCDC80 could positively regulate E-cadherin expression by counteracting the negative regulation of zinc finger protein. Thus, the expression of CCDC80 prevents cancer progression such as EMT. CD55 was found to uniquely expressed in cluster 5 (**Supplementary Figure 2D**), which comprised follicular epithelial cells_JUN high, follicular epithelial cells_NEAT1 high, and follicular epithelial cells_TG high in samples from PTC patients without concurrent HT. CD55 was known as receptor of virus and bacterial and was detected in PTC (Lucas et al., 1996). Previous study has unveiled the role of CD55 to promote the adhesion of medullary thyroid carcinoma cells for rapid spread (Mustafa et al., 2004). Furthermore, the expression of TPO was restricted in follicular epithelial cells_TFF3 high in samples from PTC patients without concurrent HT (**Supplementary Figure 2D**). TFF3 was restrictedly expressed in follicular epithelial cells in samples from PTC patients without concurrent HT (**Figure 3A**). TFF3 has been studied as a promoter of EMT and tumor invasion in PTC (Lin et al., 2018). Enriched expression of TFF3 suggested a stronger invasion ability of cluster 4, which may contribute to LNM in PTC patients without concurrent HT.

Next, we distinguished DEGs of follicular epithelial cells between two groups of patients (**Figure 3D**). Follicular epithelial cells in samples from PTC patients without concurrent HT were observed to highly express FOS, JUN, TG, TPO, DUSP1, IER2, EGR2, and BTG2 (**Figure 3D** and **Supplementary Figures 2B–D**). Associated with low expression levels of ZFP36, those coregulated genes were believed to form a positive network

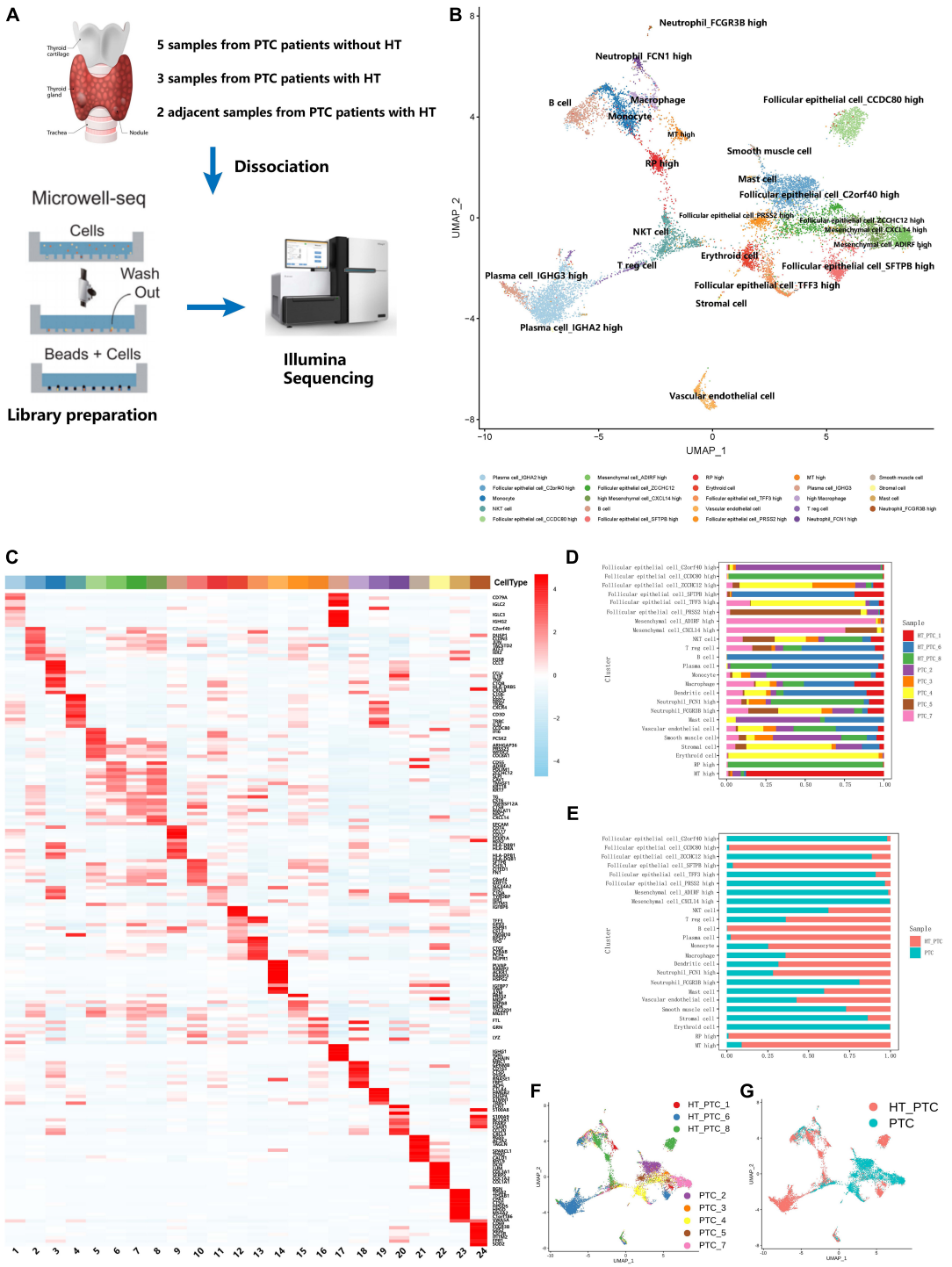


FIGURE 1 | Single-cell transcriptome landscape of papillary thyroid cancer. **(A)** Workflow of sample dissociation, library preparation, and sequencing. **(B)** UMAP plot of subclusters in PTC cell landscape. **(C)** Heatmap of specific marker genes of paired clusters in **(B)**. **(D)** The fractions of samples in each cell subcluster. **(E)** The fractions of sample types in each cell subcluster. **(F)** UMAP plot of cells colored by PTC patients. PTC, samples from PTC patients without concurrent HT; HT_PTC, samples from PTC patients with concurrent HT. **(G)** UMAP plot of single cells colored by PTC types.

in tumor cell proliferation (Canzoneri et al., 2020). FOS and JUN belong to the AP-1 transcription factor (TF) family; their ability was to form heterodimer and homodimer gene regulation network that balanced the progression of malignant follicular epithelial cells (Hess et al., 2004). We further calculated the TF expression patterns and enriched critical regulators of follicular

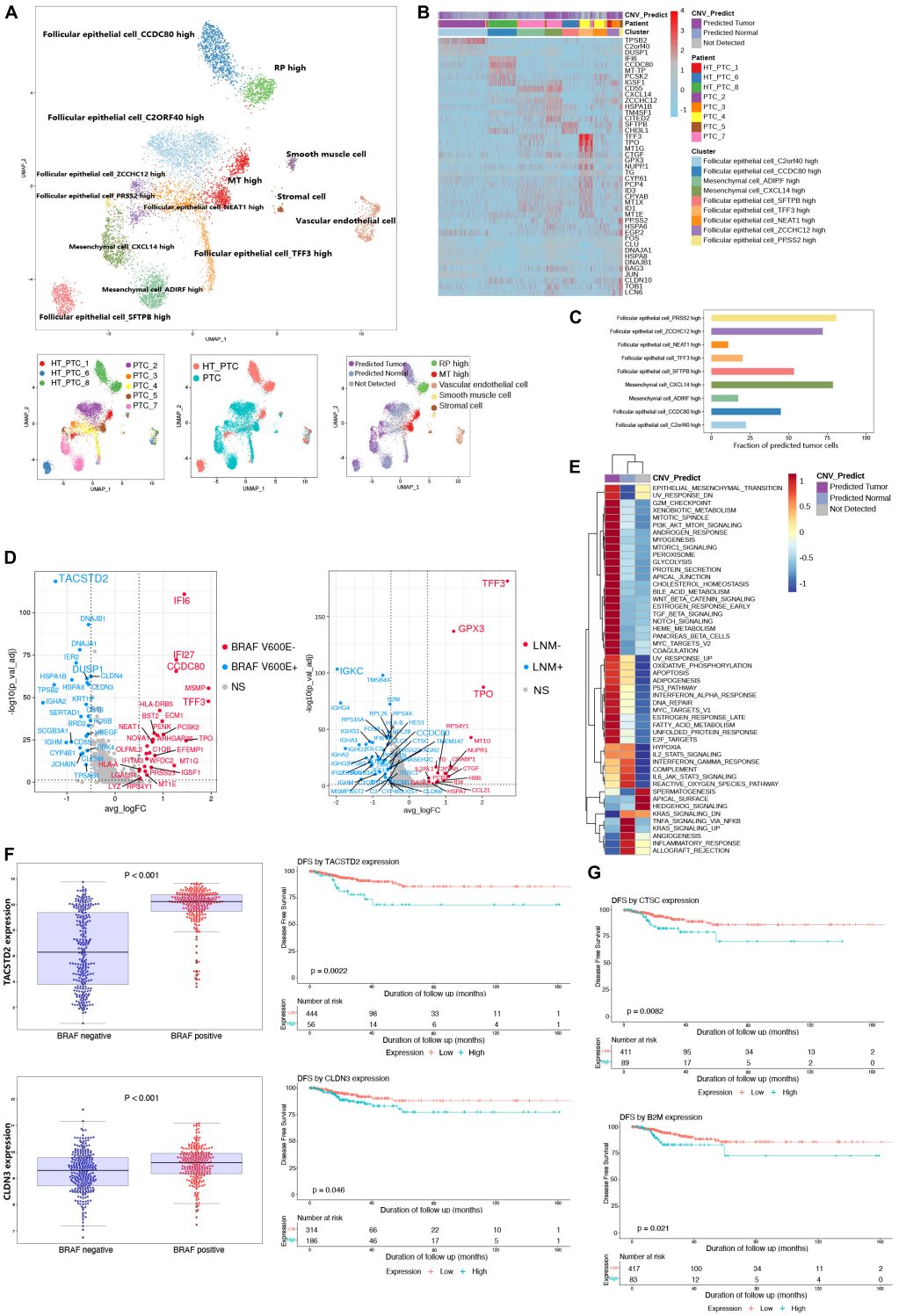


FIGURE 2 | Reclustering of parenchymal cells and identification of malignant cells. **(A)** UMAP plot of parenchymal cell subclusters colored by cell types, patients, PTC types, and CNV prediction. **(B)** Heatmap shows the DEGs of clusters, patients, and CNV groups. PTC, samples from PTC patients without concurrent HT; HT_PTC, samples from PTC patients with concurrent HT. **(C)** Fractions of predicted tumor cells in each cell subcluster. **(D)** Volcano plots show DEGs of follicular epithelial cells based on *BRAF* V600E mutation and LNM. *BRAF* V600E^{+}, highly expressed genes in patients with *BRAF* V600E mutation; *BRAF* V600E^{-}, highly expressed genes in patients without *BRAF* V600E mutation; LNM^{+}, highly expressed genes in patients with LNM; LNM-, highly expressed genes in patients without LNM; NS, genes with no significant expression patterns. **(E)** GSVA pathway enrichment of different group based on CNV level. **(F)** Expression level of TACSTD2 and CLDN3 between different mutation groups and DFS analysis in the TCGA-THCA cohort. **(G)** DFS analysis of CTSC and B2M in the TCGA-THCA cohort.

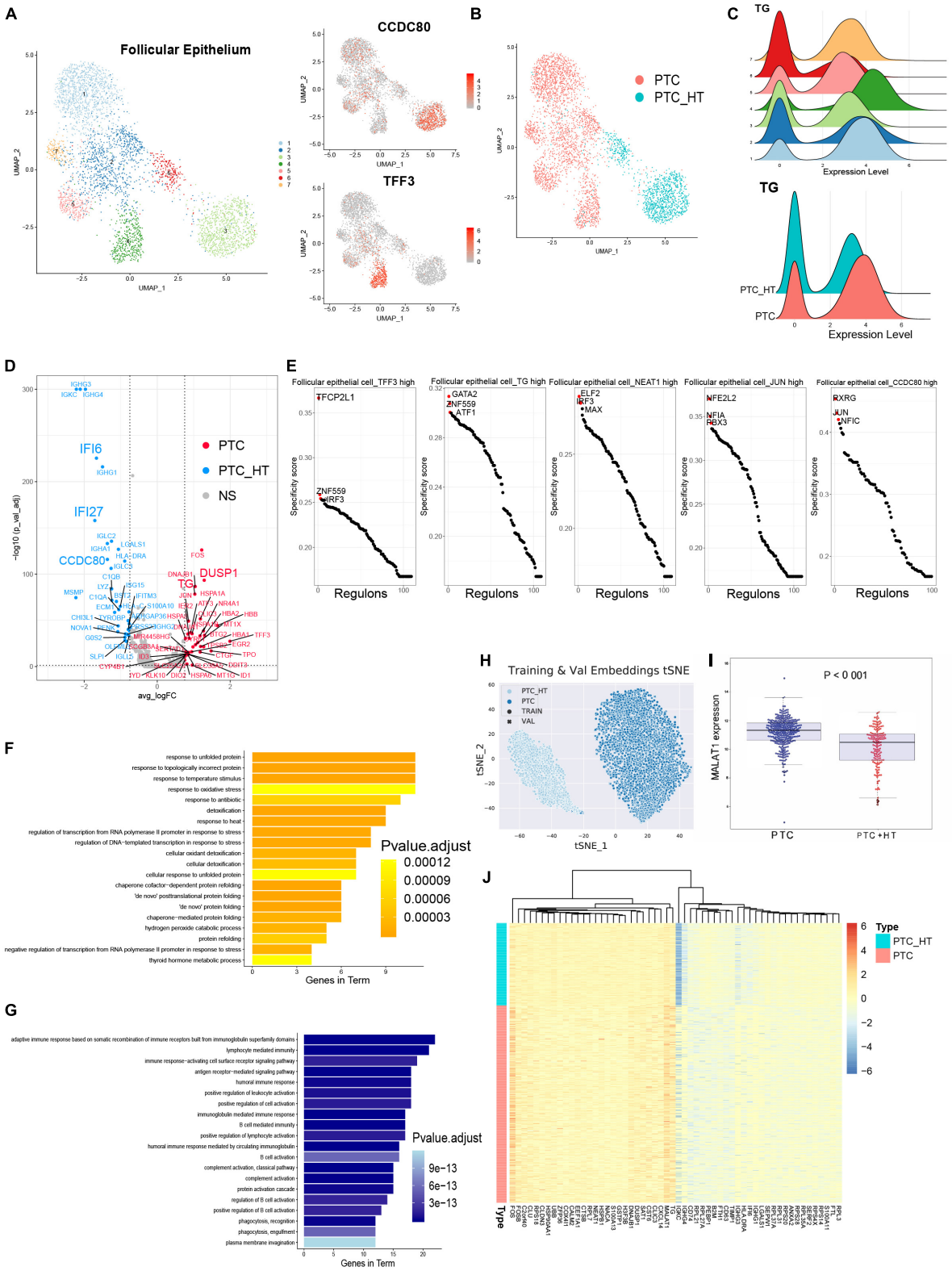


FIGURE 3 | Identifying transcriptome signatures of follicular epithelial cells. **(A)** UMAP plot of follicular epithelial cell subclusters and FeaturePlot of marker genes in two types of PTC. **(B)** UMAP plot of follicular epithelial cells colored by PTC types. **(C)** Density plots of TG expression levels and cell number distribution in follicular epithelial cell subclusters. **(D)** Volcano plot shows DEGs of epithelial cells from two types of PTC. PTC, highly expressed genes in patients without concurrent HT;

(Continued)

FIGURE 3 | (Continued)
PTC_HT, highly expressed genes in patients with concurrent HT; NS, genes with no significant expression patterns. **(E)** Rank of the cell type specific regulons in follicular epithelial cell subclusters. **(F)** Gene ontology enrichment of the top 20 genes in follicular epithelial cells from PTC patients without concurrent HT. **(G)** Gene ontology enrichment of the top 20 genes in follicular epithelial cells from PTC patients with concurrent HT. **(H)** t-SNE plot of follicular epithelial cells clustering patterns using metric learning model. TRAIN, training set; VAL, validation set. **(I)** MALAT1 expression level in the TCGA-THCA cohort based on PTC types. PTC, PTC patients without concurrent HT; PTC + HT, PTC patients with concurrent HT. **(J)** Heatmap of genes contributed to the classification in metric learning model.

epithelial cells using a reported method based on Microwell-seq data (Suo et al., 2018; **Figure 3E**). Binary treated heatmaps presented more complicated TF regulon in samples from PTC patients with concurrent HT (**Supplementary Figures 3A,B**). In samples from PTC patients without concurrent HT, we identified TFCP2L1 and IRF3 as the specific regulons in follicular epithelial cells_TFF3 high. Those regulons were known to play essential roles in tumor cell progression (Kotarba et al., 2018; Arwert et al., 2020). In samples from PTC patients with concurrent HT, RXRG was identified as the most specific regulon in follicular epithelial cells_CCDC80 high. RXRG was found to have positive correlation with better cancer prognostic (Joseph et al., 2019). Gene ontology enrichment of follicular epithelial cells from PTC patients without concurrent HT included oxidative stress response, protein folding, thyroid hormone metabolic process, and detoxification (**Figure 3F**). By contrast, samples from PTC patients with concurrent HT showed significantly high expression levels of immunoglobulin genes and interferon-stimulated genes such as IFI6 and IFI27 (Cheriyath et al., 2011). The expression of IFI27 was involved in cancer epithelial cell proliferation (Suomela et al., 2004). In the opposite way, the function of CCDC80 and lymphocyte infiltration could counteract malignancy progression of follicular epithelial cells in samples from PTC patients with concurrent HT. Interestingly, gene function analysis of specifically expressed genes in samples from PTC patients with HT contained humoral immune response, lymphocyte-mediated immunity, and adaptive immune response based on B cell activation (**Figure 3G**). Those immune responses suggested a potential cell network between follicular epithelial cells, infiltrating B cells, and myeloid cells in HT microenvironment that could affect tumor progression.

We next adopted a machine learning model named metric learning to verify the distinct molecular characteristics in follicular epithelial cells (de Vazelhes et al., 2020). We aimed to identify gene modules that contributed to the epithelial cell characteristics in samples from PTC patients with or without concurrent HT. In this case, a binary classifier was designed to predict the source of follicular epithelial cells in scRNA-seq data. We first generated training datasets and validation datasets by randomly picking expression matrix from follicular epithelial cells annotated in scRNA-seq data. The accuracy of predicting increased with the increase of the iterations (**Supplementary Figure 2E**). The final predicting classification was visualized using *t*-distributed stochastic neighbor embedding plot (PTC_HT: samples from PTC patients with concurrent HT, PTC: samples from PTC patients without concurrent HT; circles represent training datasets, cross represent validation datasets; **Figure 3H**). Follicular epithelial cells had barely no crossover between two classification labels. Training datasets and validation datasets showed a fairly fusion state (**Supplementary Figure 2F**). Heatmap demonstrated genes that contributed to the prediction (**Figure 3J**). Notably, we found that TG, FOS, and MALAT1 contributed greatly to the classification of samples from PTC patients without concurrent HT. The expression pattern of MALAT1 was validated in the TCGA-THCA cohort (**Figure 3I**). As a long non-coding RNA, recent studies have revealed the role of MALAT1 in cancer metastasis and PTC malignant process *via* several mechanisms such as angiogenesis and gene expression regulation (Huang et al., 2016). The metric learning results suggested MALAT1 as a potential biomarker to evaluate the malignancy of PTC in diagnosis.

Profiling of Tumor-Infiltrating Immune Cells in Papillary Thyroid Cancer

Reclustering of all tumor-infiltrating immune cells in tumor samples from PTC patients generated 13 tumor-associated immune cell subsets (**Figure 4A**). Common tumor-infiltrating lymphocytes consisted of NKT cells, regulatory T (Treg) cells, and follicular B cells, whereas most plasma cells and B cells were contributed by tumor samples from PTC patients with concurrent HT. The myeloid clusters included two DC cell subsets, two neutrophils subsets, macrophages, mast cells, and monocytes. A certain degree of immune cell composition variation could be observed between PTC types rather than patients, but a high proportion of lymphocytes could be observed in most patients (**Figure 4C**). Single cells were assigned into annotated clusters to calculate the Spearman correlation between lymphocytes and myeloid cells, respectively, (**Figure 4B**).

Significant enrichment of TF regulon activity could be observed in myeloid cells from PTC patients without concurrent HT. SPI1, ETS2, and CCAAT enhancer binding protein heterodimers (CEBPB, CEBPA, CEBPD, and CEBPG) were enriched in macrophages, neutrophils, DCs, and endothelial cells. Those activation patterns of myeloid cell were believed to participate in the regulation of tumor cell progression (Pitarresi et al., 2016; **Figure 4E** and **Supplementary Figure 3A**). In tumor-infiltrating B cells from PTC patients with HT, we identified POU2F1 and BATH2 as key regulons to maintain B cell identity (Miura et al., 2018; Pankratova et al., 2018; **Figure 4F**).

Papillary thyroid cancer infiltrating T cells consisted of $CD3^+$ cytotoxic T cell, $CD8^+$ cytotoxic cell, and $CTLA4^+$ and $CD4^+$ Treg cells (**Figure 4G**). Plasma cells were characterized by the prominent expression levels of JCHAIN and XBP1. Cycling B cells expressed dogmatic cell cycle genes such as STMN1 and UBE2C (**Figure 4H**). DC subsets expressed low levels of migration marker C-C motif chemokine receptor 7, but interleukin $1B^+$ ($IL1B^+$) DCs expressed high levels of

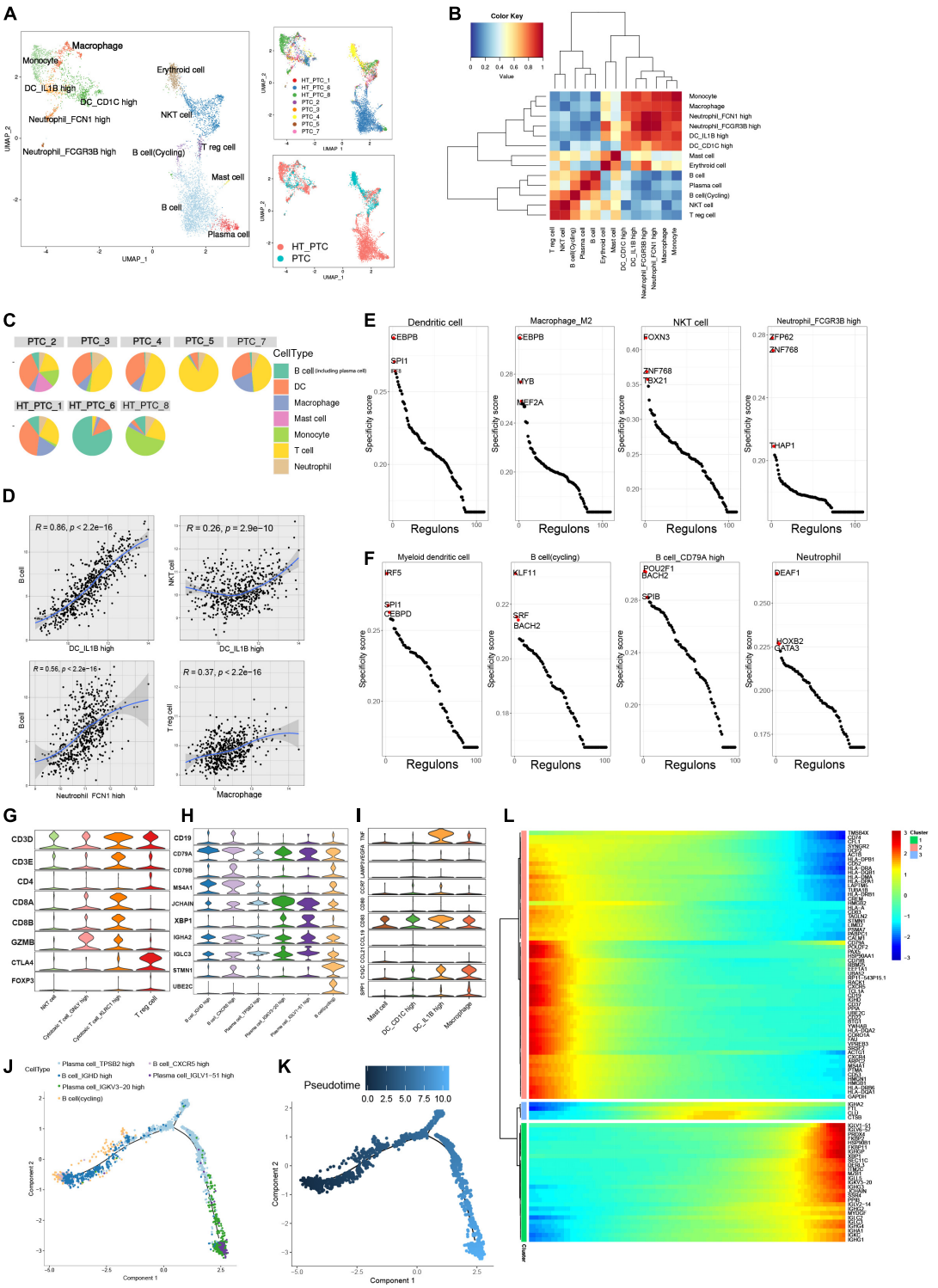


FIGURE 4 | Reclustering of tumor-infiltrating immune cells. **(A)** UMAP plots of immune cell subclusters in PTC colored by cell types, patient, and PTC types. **(B)** Cell–cell Spearman correlation network of immune cell subclusters. **(C)** Pie charts of cell-type fractions for each patient's tumor-infiltrating immune cells. **(D)** Correlation between myeloid cells and lymphocytes subsets in the TCGA-THCA cohort. Coefficient was calculated with Spearman correlation analysis. **(E)** Rank of the cell type–specific regulons in immune cells from PTC patients without concurrent HT. **(F)** Rank of the cell type specific regulon in immune cells from PTC patients with concurrent HT. **(G)** Violin plot of marker genes in T-cell subclusters. **(H)** Violin plot of marker genes in B-cell subclusters. **(I)** Violin plot of marker genes in myeloid cell subclusters. **(J)** Pseudotime trajectory of B-cell subclusters using monocle2. **(K)** Color-coded pseudotime of B-cell subclusters; the start point is dark blue. **(L)** Heatmap of the top DEGs expression levels in three divided clusters from pseudotime trajectory.

proinflammatory mediator genes such as tumor necrosis factor (TNF) and IL1B (Brown et al., 2019; **Figure 4I**). We observed positively strong gene expression signature correlations between $IL1B^+$ DCs and follicular B cells in the TCGA-THCA cohort (**Figure 4D**), indicating that $IL1B^+$ DCs could recruit B cells into TME in PTC patients. Thus, we performed trajectory analysis to investigate the B cell–infiltrating process (**Figure 4J**). Cycling B cells, B cells_IGHD high, and B cells_CXCR5 high were located in the start point of the pseudotime trajectory, whereas three subsets of plasma cells were enriched at the two branches (**Figure 4K**). Along with the immune responses caused by HT, infiltrating B cells proliferated and differentiated into plasma cells to generate thyroid-restricted antibodies. In this differentiation lineage, pseudotemporal expression heatmap showed scaled expression of dynamic genes (**Figure 4L**). Cluster 1 enriched the cycling B cell marker. Function analysis of enriched genes suggested B cell proliferation, differentiation, and leukocyte chemotaxis. Cluster 3 contains up-regulated genes related to the secreting of autoantibodies. Gene function enrichment analysis indicated immunoglobulin receptor binding, antigen binding, and Fc-gamma receptor signaling pathway.

Distinct Molecular Signatures of Papillary Thyroid Cancer With or Without Concurrent Hashimoto's Thyroiditis

As considerable cell heterogeneity was observed between samples from PTC patients with or without concurrent HT, we reclustered cells based on patient type. The clusters in samples from PTC patients without concurrent HT mainly comprised four malignant follicular epithelial cell subsets, NKT cells, Treg cells, mesenchymal cells, DCs, macrophages, neutrophils, and mast cells (**Figures 5A,B** and **Supplementary Figures 4A,C**). Clusters in samples from PTC patients with concurrent HT comprised two plasma cell subsets, B cells, monocytes, NKT cells, myeloid DCs, cytotoxic T cells, Treg cells, follicular epithelial cells, and a small number of mast cells (**Figures 5A,B** and **Supplementary Figure 4B**). As a canonical B cell marker gene, CD79 was validated by IHC staining in samples from PTC patients with concurrent HT (**Figure 5C**). JCHAIN and XBP1 were enriched in plasma cells_IGLL5 high, plasma cells_IGHA2 high, B cells_CD79A high, and cycling B cells (**Supplementary Figure 4D**).

We next examined the cell compositions in two types of patients. Different patients barely showed bias in each cell type (**Figure 5D**). Cell fractions in samples from PTC patients without concurrent HT showed dominated status of follicular epithelial cells, which have been characterized by high expression of TG, JUN, FOS, and EPCAM (**Supplementary Figures 5A–C**). Mesenchymal cells and stromal cells were defined by FN1 and DCN, respectively. Neutrophils expressed high levels of S100A8 and S100A9. Mast cells specifically expressed TPSB2 and TPSAB1 (**Figure 5F**). Samples from PTC patients with HT included four plasma cell and B cell clusters (high expression of immunoglobulin, JCHAIN, XBP1, CD79A, and CD79B). In addition, cycling B cells were defined by UBE2C and STMN1 (**Figure 5G**). Monocyte was defined as high expression of CD14 molecule (CD14). DCs were identified using IRF8, CD1C, and CD86. Vascular endothelial cells showed high expression levels of ENG and PLVAP. Staining intensities of CD55 on samples from PTC patients without concurrent HT also showed relatively higher expression levels (**Supplementary Figure 5D**). To further investigate the patterns of infiltrating B cells in PTC patients with concurrent HT in the clinical datasets, PTC samples in the TCGA-THCA cohort were divided into two groups. We noticed B cell–related markers such as CD79A, CD79B, and MS4A1 were significantly enriched in PTC patients with concurrent HT (**Figure 5E**). Strikingly, tumor-infiltrating immune cell abundance profile of patients in the TCGA-THCA cohort also indicated significant enrichment of plasma cells and naive B cells in PTC patients with concurrent HT (**Figure 5H**).

Tumor-Infiltrating B Lymphocytes With Migratory Capacity Unveil the Immunological Link Between Papillary Thyroid Cancer and Hashimoto's Thyroiditis

It is still not clear whether the infiltrating immune cell phenotypes in samples from PTC patients with concurrent HT were affected by adjacent HT microenvironment. We merge the single cells in tumor samples and adjacent tissue samples from PTC patients with concurrent HT (**Figure 6A** and **Supplementary Figure 6A**). Adjacent tissues from PTC patients with concurrent HT occupied the majority of cycling B cells, Treg cells, cytotoxic T cells, and $MS4A1^+$ B cells (**Supplementary Figures 6B,C**). In contrast, tumor samples from PTC patients with concurrent HT occupied the majority of TAMs and $IGHA2^+$ plasma cells. To delineate the path of B cell–to–plasma cell differentiation, we reclustered B cell subsets in tumor tissues and paired adjacent tissues from PTC patients with concurrent HT (**Supplementary Figures 6D,E**). RNA velocity analysis predicted the migration potential of B cells as well as cycling B cells and revealed their activity to infiltrate tumor tissues (**Figure 6B** and **Supplementary Figure 6C**). Considering the crucial role of myeloid cells in lymphocyte recruitment (Ehlers and Schott, 2014), we investigated the cell–cell ligand–receptor networks in tumor samples and paired adjacent tissues from PTC patients with concurrent HT, respectively. Adjacent tissue showed much fewer ligand–receptor pairs compared to tumor samples (**Figures 6C,D**). TAMs, neutrophils, and DCs enriched most ligand–receptor pairs with lymphocytes. These findings highlight myeloid cells as a potential signal transition hub to regulate the B cell recruitment from adjacent tissues to tumor tissues.

We adopted CellphoneDB to detect common ligand–receptor interactions in thyroid cancer that have been alternately verified in other cancers recently (Chen et al., 2020; Lee et al., 2020). Dot plots showed selected ligand–receptor interactions starting from follicular epithelial cells to immune cells in adjacent tissue and tumor samples from PTC patients with concurrent HT, respectively, (**Figures 6E,F**). These cell–cell interaction networks indicated more positive interactions between follicular epithelial cells and immune cells in tumor samples. Protumor interactions

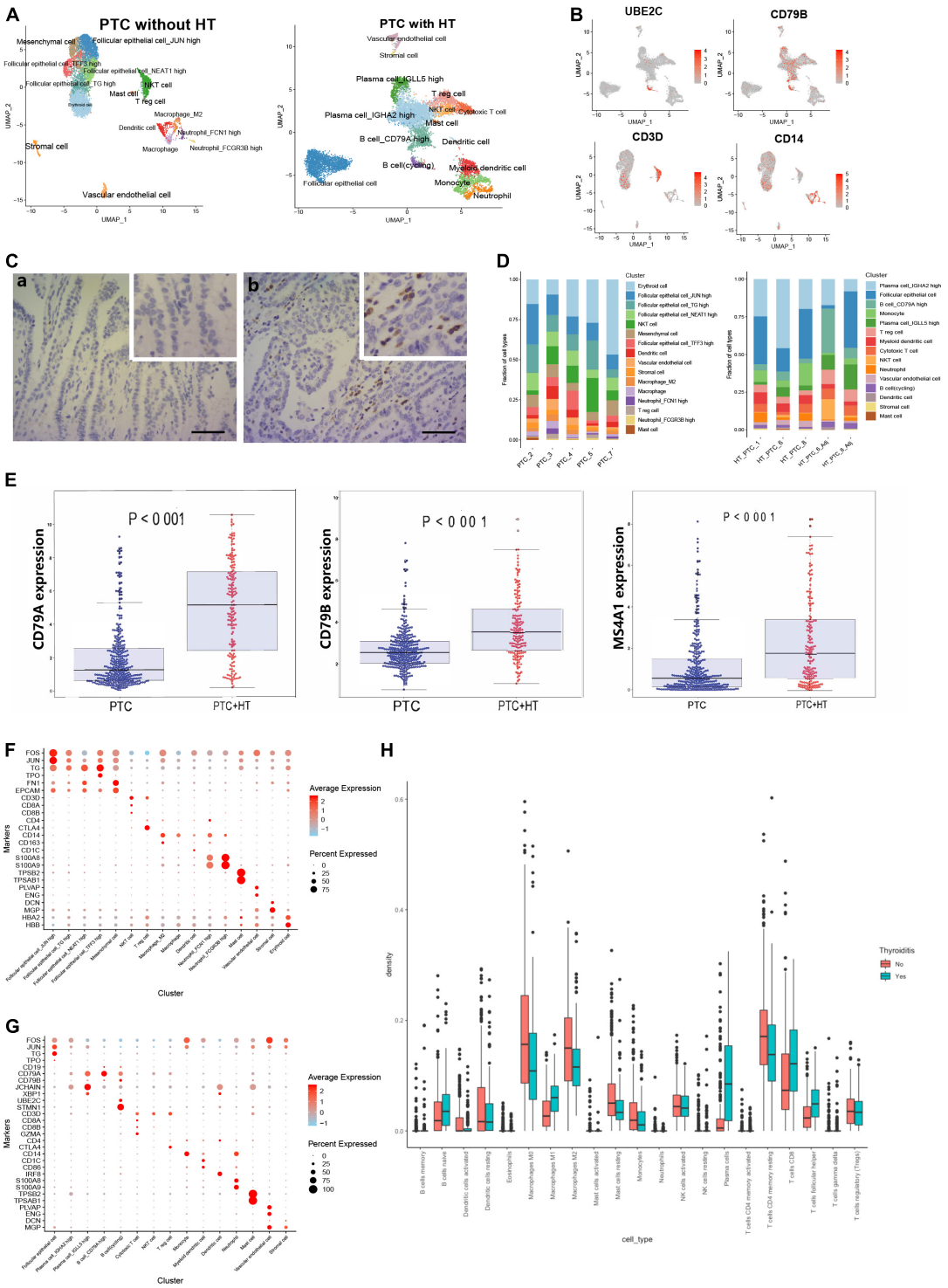


FIGURE 5 | Reclustering of single cells in samples from two types of PTC. **(A)** UMAP plot of single cells in samples from two types of PTC. **(B)** FeaturePlot of selected marker genes in two types of PTC. **(C)** Immunostaining of CD79A/B in two types of samples from PTC patients. Positive signals were stained in brown. Scale bar = 50 μm. (a: sample from PTC patient without concurrent HT, b: sample from PTC patient with concurrent HT). **(D)** The fractions of cell types in samples from two types of PTC. **(E)** The expression levels of B cell marker genes in the TCGA-THCA cohort based on PTC types. PTC, PTC patients without concurrent HT; PTC + HT, PTC patients with concurrent HT. **(F)** Dot plot of marker genes in samples from PTC patients without concurrent HT. Average expression is color-coded. Columns represent the selected marker genes. **(G)** Dot plot of marker genes in samples from PTC patients with concurrent HT. Average expression is color-coded. Columns represent the selected marker genes. **(H)** Density of canonical immune cells in the TCGA-THCA cohort. Thyroiditis Yes, PTC patients with concurrent HT; Thyroiditis No, PTC patients without concurrent HT.

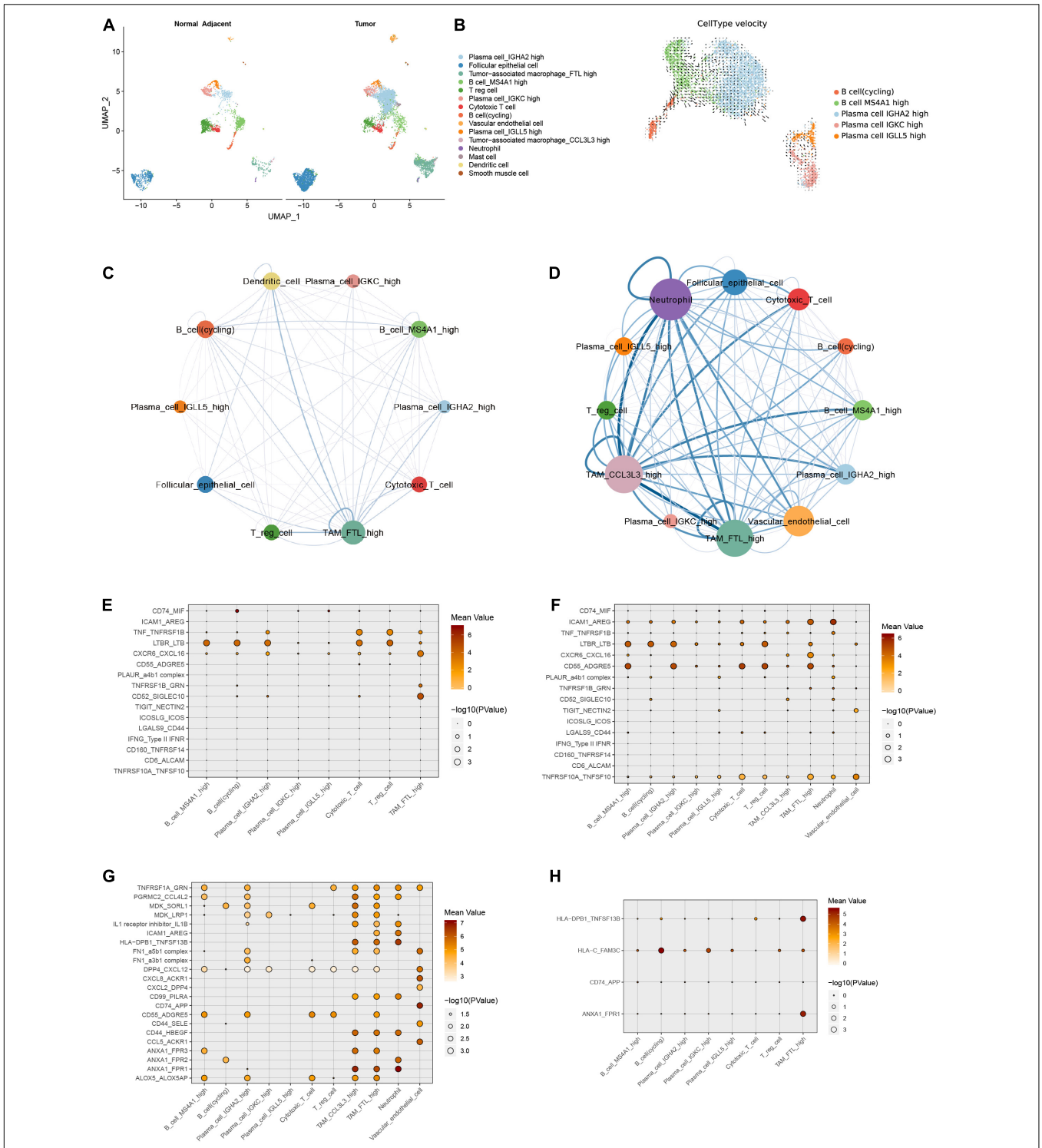


FIGURE 6 | Comparison of adjacent tissues and tumor tissues from PTC patients with concurrent HT. **(A)** UMAP plot of cell types in adjacent tissues and tumor tissues from PTC patients with concurrent HT. **(B)** UMAP plot shows RNA velocities of B cell subsets in adjacent tissues and tumor tissues. **(C)** Cell–cell interaction network of ligand–receptor pairs in adjacent tissues of PTC patients with concurrent HT. The size of the circle represents the total ligand–receptor pairs of each cell type. The line weight represents the ligand–receptor pairs between two linked cell types. **(D)** Cell–cell interaction network of ligand–receptor pairs in tumor tissues of PTC patients with concurrent HT. **(E)** Selected ligand–receptor pairs of follicular epithelial cells from adjacent tissue samples in PTC patients with concurrent HT. *x* axis, receptor cells; *y* axis, ligand–receptor pairs (expression level is color-coded). **(F)** Selected ligand–receptor pairs of follicular epithelial cells from tumor samples in PTC patients with concurrent HT. **(G)** Calculated ligand–receptor pairs of follicular epithelial cells from adjacent tissue samples in PTC patients with concurrent HT. **(H)** Calculated ligand–receptor pairs of follicular epithelial cells from tumor samples in PTC patients with concurrent HT.

such as TNF_TNFRSF1B, LTBR-LTB, and CXCR6_CXCL16 were found to have higher expression levels in follicular epithelial cells. LTBR interacted with lymphocyte LTB and promoted tumor cell metastasis (Das et al., 2019). Transmembranous CXCL16 and soluble CXCL16 played opposite roles in tumor cell migration (Gooden et al., 2014). Another pair TNF_TNFRSF1B could activate the nuclear factor κB pathway and resisted TNF-induced tumor cell death (Rothe et al., 1995). Those previously reported mechanisms in other cancers seem to have the synchronization regulation pattern in PTC with concurrent HT. Follicular epithelial cells in tumor samples from PTC patients with concurrent HT exclusively contained the antitumor interacting pair TNFRSF10A_TNFSF10. TNFSF10 has been reported to induce tumor cell apoptosis (Qu et al., 2019). These potential antitumor interaction networks were constructed by plasma cells, cytotoxic T cells, TAMs, and vascular endothelial cells.

We then calculated other notable interaction pairs of epithelial cells in tumor tissues and adjacent tissues from PTC patients with concurrent HT (**Figures 6G,H**). Interacting pairs HLA-DPB1_TNFSF13B, HLA-C_FAM3C, and ANXA1_FPR1 were observed between follicular epithelial cells and macrophages. TNFSF13B has been recognized as a potent B cell activator cytokine (Nocturne and Mariette, 2015). FAM3C and FPR1 were reported to be related to malignant cell migration (Cao and Zhang, 2018; Yin et al., 2018). In adjacent tissues, the myeloid cell–mediated lymphocyte activation may form an immune response network to resist PTC progress. In follicular epithelial cells from tumor samples, cytokine interactions were observed in ACKR1-positive endothelial cells. Endothelial cells that highly expressed ACKR1 and PLVAP could enhance the transmigration of lymphocytes (Ramachandran et al., 2019), and GRN was found to promote angiogenesis and tumor cell proliferation (Yang et al., 2015). Ligand–receptor pairs of TAMs_FTL^{+} in samples from PTC patients with concurrent HT contained quite a number of cytokines (**Supplementary Figures 7A,B**). Tumor-associated macrophage_CCL3L3 and neutrophils in samples from PTC patients with concurrent HT showed similar patterns (**Supplementary Figures 7C,D**). In fact, a fair proportion of interaction pairs including cytokines and major histocompatibility complex could somehow facilitate the progression of PTC (Miyake et al., 2016). Conversely, lymphocyte infiltration patterns induced by endothelial cells could inhibit the growth and migration of PTC *via* boosting the effect of T cells and improve clinical prognostic (Wouters and Nelson, 2018).

DISCUSSION

Tumor heterogeneous and immune microenvironments have been considered to significantly impact the development of cancer. Different cell subsets in tumors could regulate this process with varying degrees. Conventional bulk RNA-seq of tumors vaguely described the gene expression modules at tissue level. In this work, we generated the single-cell transcriptome landscape of human PTC. The PTC landscape covered 24 individual cell clusters from eight PTC patients, providing a rich resource for multidimensional characterization of parenchymal cells and immune cells in PTC. Inferred CNVs in parenchymal cells identified predicted tumor cells in follicular epithelial cells and mesenchymal cells with strong EMT properties, which may lead to tumor progression and metastasis. The *BRAF* V600E mutation confers an aggressive phenotype in PTC, and a previous study has used RNA-seq to identify DEGs between *BRAF* V600E mutation and *BRAF* wild-type tumors (Smallridge et al., 2014). Nonetheless, single-cell profiling revealed individual tumor cell characteristics that may be obscured in bulk analysis. Our single-cell analysis revealed that *BRAF* V600E mutation tumor epithelial cells highly expressed TACSTD2 and CLDN3, which were then validated in TCGA data. It was intriguing that analysis of TCGA data indicates a strong correlation between TACSTD2, CLDN3, and DFS in PTC, implying that *BRAF* V600E mutation might be related to TACSTD2 through ERK and JNK pathways and the upregulation of MMP2. LNM in PTC has also been shown to be associated with locoregional recurrence and poor prognosis. Single-cell analysis identified that CTSC and B2M were up-regulated in LNM-positive epithelial cells with strong correlation to poor DFS, suggesting that cancer-related signaling pathways such as TNF-α/p38 MAPK signaling pathway could promote LNM in PTC (Zhang et al., 2020). Reclustering of follicular epithelial cells defined epithelial cells high expressed TFF3 in samples from PTC patients without concurrent HT and epithelial cells high expressed CCDC80 in samples from PTC patients with concurrent HT. Machine learning model predicted MALAT1 as a potential biomarker in PTC patients without concurrent HT.

The dynamic responses of immune system in PTC remain to be completely deciphered. We revealed the complexity of immune cells in PTC and identified multiple PTC-specific myeloid cells, T cells, and B cell subsets. In the TCGA-THCA cohort, single-cell gene expression signatures of proinflammatory IL1B^{+} DC cells were highly positively correlated with B cells. The proinflammatory DCs continually recruited lymphocytes into TME and suppressed T cell exhaustion. High proportion of tumor-infiltrating lymphocytes could explain the indolent properties of PTC, which leads to favorable prognosis. Intriguingly, we found B lymphocyte and plasma cell enrichment pattern in PTC tumor tissues that were concurrent with HT, but not in the PTC alone without concurrent HT. B lymphocyte infiltration in TME has been identified in multiple types of cancers such as lung cancer and was recognized as the main effector cells of humoral immunity, which suppress tumor progression (Wang et al., 2019). A great number of ligand–receptor interactions were observed between B cells, T cells, DCs, and TAMs. Precursor–progeny relationship supported by RNA velocity unveiled the migration potential of infiltrating B cells from adjacent tissues to tumor tissues. PTC is always accompanied by the synchronous appearance of HT. Compared with adjacent tissues, more complicated cell–cell interaction networks of ligands and receptors between endothelial cells, epithelial cells, lymphocytes, and myeloid cells in samples from PTC patients with concurrent HT suggested a controlled homeostasis microenvironment of tumor progress regulation. The analyses of PTC and adjacent HT tissues implied that the cellular TME was reshaped by the B lymphocytes derived from adjacent HT tissues. Based on our observations, we hypothesized

that B lymphocyte–related immune response was a possible reason for the better prognosis of PTC patients with concurrent HT. But detailed mechanisms need further investigation and verification. Also, limited number of cases in our study could not cover all the clinical features of PTC, which remains to be further explored.

In summary, our work is a comprehensive systematic single-cell transcriptome survey of human primary PTC. We revealed detailed molecular characteristics of PTC cells as well as their clustering, dynamic, and developmental trajectory. Our analysis uncovered B cells infiltrating in tumor tissues as a distinctive feature for PTC patients with concurrent HT. Our findings are potentially valuable in not only serving as a resource for deeper understanding of PTC in general, but also elucidating the immunological correlation between PTC and HT.

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://www.ncbi.nlm.nih.gov/geo/, GSE163203.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Clinical Research Ethics Committee of The first affiliated hospital, school of medicine, Zhejiang university (IIT consent: NO.700,2020). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YjW, GG, JP, and FY conceived and designed the experiments. JP, QZ, HT, MZ, and YZ collected the patients' samples. JP, FY, YS, FZ, and XZ performed the experiments. FY, CY, YY, JL, and JP analyzed the data. FY, JP, and QZ wrote the manuscript. YnW, JP, and YjW provide funding and supervision. All authors read and approved the final manuscript.

FUNDING

This study was supported by grant from Zhejiang Provincial Natural Science Foundation of China (LQ18H180003 and LQ18H050003), the National Natural Science Foundation of China (81800658), and Zhejiang Medical Science and Technology Projects (2019330597 and 2019330585).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.758339/full#supplementary-material

REFERENCES

Ahn, D., Heo, S. J., Park, J. H., Kim, J. H., Sohn, J. H., Park, J. Y., et al. (2011). Clinical relationship between Hashimoto's thyroiditis and papillary thyroid cancer. *Acta Oncol.* 50, 1228–1234.

Arwert, E. N., Milford, E. L., Rullan, A., Derzsi, S., Hooper, S., Kato, T., et al. (2020). STING and IRF3 in stromal fibroblasts enable sensing of genomic stress in cancer cells to undermine oncolytic viral therapy (vol 56, pg 410, 2020). *Nat. Cell Biol.* 22, 908–908. doi: 10.1038/s41556-020-0527-7

Borowczyk, M., Szczepanek-Parulska, E., Debicki, S., Budny, B., Verburg, F. A., Filipowicz, D., et al. (2019). Differences in mutational profile between follicular thyroid carcinoma and follicular thyroid adenoma identified using next generation sequencing. *Int. J. Mol. Sci.* 26:20.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R., Torre, L., 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, 394–424.

Brown, C. C., Gudjonson, H., Pritykin, Y., Deep, D., Lavallee, V. P., Mendoza, A., et al. (2019). Transcriptional basis of mouse and human dendritic cell heterogeneity. *Cell* 179, 846–863.e24.

Canzoneri, R., Naipauer, J., Stedile, M., Rodriguez Pena, A., Lacunza, E., Gandini, N. A., et al. (2020). Identification of an AP1-ZFP36 regulatory network associated with breast cancer prognosis. *J. Mammary Gland Biol. Neoplasia* 25, 163–172. doi: 10.1007/s10911-020-09448-1

Cao, G. M., and Zhang, Z. Y. (2018). FPR1 mediates the tumorigenicity of human Dove ORIGINAL RESEARCH cervical cancer cells. *Cancer Manag. Res.* 10, 5855–5865. doi: 10.2147/CMAR.S182795

Ceolin, L., Siqueira, D. R., Romitti, M., Ferreira, C. V., and Maia, A. L. (2012). Molecular basis of medullary thyroid carcinoma: the role of RET polymorphisms. *Int. J. Mol. Sci.* 13, 221–239. doi: 10.3390/ijms13010221

Chen, Z. H., Zhou, L. J., Liu, L. L., Hou, Y. X., Xiong, M., Yang, Y., et al. (2020). Single-cell RNA sequencing highlights the role of inflammatory cancer-associated fibroblasts in bladder urothelial carcinoma. *Nat. Commun.* 8:11. doi: 10.1038/s41467-020-18916-5

Cheriyath, V., Leaman, D. W., and Borden, E. C. (2011). Emerging roles of FAM14 family members (G1P3/ISG 6-16 and ISG12/IFI27) in innate immunity and cancer. *J. Interferon Cytokine Res.* 31, 173–181. doi: 10.1089/jir.2010.0105

Das, R., Coupar, J., Clavijo, P. E., Saleh, A., Cheng, T. F., Yang, X., et al. (2019). Lymphotoxin-beta receptor-NIK signaling induces alternative RELB/NF-kappaB2 activation to promote metastatic gene expression and cell migration in head and neck cancer. *Mol. Carcinog.* 58, 411–425. doi: 10.1002/mc.22938

Davis, R., Blake, K., Ma, D., Gabra, M., Hernandez, G., Phung, A., et al. (2020). Transcriptional diversity and bioenergetic shift in human breast cancer metastasis revealed by single-cell RNA sequencing. *Nat. Cell Biol.* 22, 310–320. doi: 10.1038/s41556-020-0477-0

de Vazelhes, W., Carey, C. J., Tang, Y., Vauquier, N., and Bellet, A. (2020). metric-learn: metric learning algorithms in python. *J. Mach. Learn. Res.* 21, 1–6.

Ehlers, M., and Schott, M. (2014). Hashimoto's thyroiditis and papillary thyroid cancer: are they immunologically linked? *Trends Endocrin. Met.* 25, 656–664. doi: 10.1016/j.tem.2014.09.001

Ferraro, A., Schepis, F., Leone, V., Federico, A., Borbone, E., Pallante, P., et al. (2013). Tumor suppressor role of the CL2/DRO1/CCDC80 gene in thyroid carcinogenesis. *J. Clin. Endocrinol. Metab.* 98, 2834–2843.

Gabrilovich, D. I., and Nagaraj, S. (2009). Myeloid-derived suppressor cells as regulators of the immune system. *Nat. Rev. Immunol.* 9, 162–174. doi: 10.1038/nri2506

Gao, R. L., Bai, S. S., Henderson, Y. C., Lin, Y. Y., Schalck, A., Yan, Y., et al. (2021). Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat. Biotechnol.* 39, 599–608. doi: 10.1038/s41587-020-00795-2

Gentles, A. J., Newman, A. M., Liu, C. L., Bratman, S. V., Feng, W. G., Kim, D., et al. (2015). The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat. Med.* 21, 938–945. doi: 10.1038/nm.3909

Gooden, M. J., Wiersma, V. R., Boerma, A., Leffers, N., Boezen, H. M., ten Hoor, K. A., et al. (2014). Elevated serum CXCL16 is an independent predictor of poor survival in ovarian cancer and may reflect pro-metastatic ADAM protease activity. *Br. J. Cancer* 110, 1535–1544. doi: 10.1038/bjc.2014.55

Gooden, M. J. M., de Bock, G. H., Leffers, N., Daemen, T., and Nijman, H. W. (2011). The prognostic influence of tumour-infiltrating lymphocytes in cancer: a systematic review with meta-analysis. *Br. J. Cancer* 28, 93–103. doi: 10.1038/bjc.2011.189

Guan, H. Y., Guo, Z. J., Liang, W. W., Li, H., Wei, G. H., Xu, L. J., et al. (2017). Trop2 enhances invasion of thyroid cancer by inducing MMP2 through ERK and JNK pathways. *BMC Cancer* 17:486. doi: 10.1186/s12885-017-3475-2

Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., et al. (2018). Mapping the mouse cell atlas by microwell-seq. *Cell* 172, 1091–1107.e17.

Han, X., Zhou, Z., Fei, L., Sun, H., Wang, R., Chen, Y., et al. (2020). Construction of a human cell landscape at single-cell level. *Nature* 581, 303–309.

Hess, J., Angel, P., and Schorpp-Kistner, M. (2004). AP-1 subunits: quarrel and harmony among siblings. *J. Cell Sci.* 117, 5965–5973. doi: 10.1242/jcs.01589

Hess, J., Thomas, G., Braselmann, H., Bauer, V., Bogdanova, T., Wienberg, J., et al. (2011). Gain of chromosome band 7q11 in papillary thyroid carcinomas of young patients is associated with exposure to low-dose irradiation. *Proc. Natl. Acad. Sci. U. S. A.* 108, 9595–9600. doi: 10.1073/pnas.1017137108

Hewitt, K. J., Agarwal, R., and Morin, P. J. (2006). The claudin gene family: expression in normal and neoplastic tissues. *BMC Cancer* 6:186. doi: 10.1186/1471-2407-6-186

Huang, J. K., Ma, L., Song, W. H., Lu, B. Y., Huang, Y. B., Dong, H. M., et al. (2016). MALAT1 promotes the proliferation and invasion of thyroid cancer cells *via* regulating the expression of IQGAP1. *Biomed. Pharmacother.* 83, 1–7. doi: 10.1016/j.biopha.2016.05.039

Joseph, C., Al-Izzi, S., Alsaleem, M., Kurozumi, S., Toss, M. S., Arshad, M., et al. (2019). Retinoid X receptor gamma (RXRG) is an independent prognostic biomarker in ER-positive invasive breast cancer. *Br. J. Cancer* 121, 776–785. doi: 10.1038/s41416-019-0589-0

Jung, K. Y., Cho, S. W., Kim, Y. A., Kim, D., Oh, B. C., Park, D. J., et al. (2015). Cancers with higher density of tumor-associated macrophages were associated with poor survival rates. *J. Pathol. Transl. Med.* 49, 318–324. doi: 10.4132/jptm.2015.06.01

Kebebew, E., Weng, J., Bauer, J., Ranvier, G., Clark, O. H., Duh, Q. Y., et al. (2007). The prevalence and prognostic value of BRAF mutation in thyroid cancer. *Ann. Surg.* 246, 466–471. doi: 10.1097/sla.0b013e318148563d

Konturek, A., Barczynski, M., Wierzchowski, W., Stopa, M., and Nowak, W. (2013). Coexistence of papillary thyroid cancer with Hashimoto thyroiditis. *Langenbecks Arch. Surg.* 398, 389–394.

Kotarba, G., Krzywinska, E., Grabowska, A. I., Taracha, A., and Wilanowski, T. (2018). TFCP2/TFCP2L1/UBP1 transcription factors in cancer. *Cancer Lett.* 420, 72–79.

Kuo, C.-Y., Liu, T.-P., Yang, P.-S., and Cheng, S.-P. (2017). Characteristics of lymphocyte-infiltrating papillary thyroid cancer. *J. Cancer Res. Pract.* 4, 95–99.

Kwak, J. Y., Kim, E.-K., Kim, J.-K., Han, J.-H., Hong, S. W., Park, T. S., et al. (2010). Dual priming oligonucleotide-based multiplex PCR analysis for detection of BRAFV600E mutation in FNAB samples of thyroid nodules in BRAFV600E mutation-prevalent area. *Head Neck* 32, 490–498. doi: 10.1002/hed.21210

Lee, E. K., and Sunwoo, J. B. (2019). Natural killer cells and thyroid diseases. *Endocrinol. Metab.* 34, 132–137. doi: 10.3803/enm.2019.34.2.132

Lee, H. K., Ang, K. S., Chevrier, M., Goh, M., Ling, J., Koh, V., et al. (2020). Single cell RNA-seq reveals immunosuppressive gastric stem-like cancer cells as a poor prognostic factor. *bioRxiv*[Preprint] doi: 10.1101/2020.10.23.351726

Lee, J. H., Kim, Y., Choi, J. W., and Kim, Y. S. (2013). The association between papillary thyroid carcinoma and histologically proven Hashimoto's thyroiditis: a meta-analysis. *Eur. J. Endocrinol.* 168, 343–349. doi: 10.1530/EJE-12-0903

Lin, X., Zhang, H., Dai, J., Zhang, W., Zhang, J., Xue, G., et al. (2018). TFF3 contributes to epithelial-mesenchymal transition (EMT) in papillary thyroid carcinoma cells *via* the MAPK/ERK signaling pathway. *J. Cancer* 9, 4430–4439. doi: 10.7150/jca.24361

Lucas, S. D., KarlssonParra, A., Nilsson, B., Grimelius, L., Akerstrom, G., Rastad, J., et al. (1996). Tumor-specific deposition of immunoglobulin G and complement in papillary thyroid carcinoma. *Hum. Pathol.* 27, 1329–1335. doi: 10.1016/s0046-8177(96)90346-9

Luo, H., Xia, X., Kim, G. D., Liu, Y., Xue, Z., Zhang, L., et al. (2021). Characterizing dedifferentiation of thyroid cancer by integrated analysis. *Sci. Adv.* 7:eabf3657. doi: 10.1126/sciadv.abf3657

Miura, Y., Morooka, M., Sax, N., Roychoudhuri, R., Itoh-Nakadai, A., Brydun, A., et al. (2018). Bach2 promotes B Cell receptor-induced proliferation of B lymphocytes and represses cyclin-dependent kinase inhibitors. *J. Immunol.* 200, 2882–2893. doi: 10.4049/jimmunol.1601863

Miyake, M., Hori, S., Morizawa, Y., Tatsumi, Y., Nakai, Y., Anai, S., et al. (2016). CXCL1-mediated interaction of cancer cells with tumor-associated macrophages and cancer-associated fibroblasts promotes tumor progression in human bladder cancer. *Neoplasia* 18, 636–646. doi: 10.1016/j.neo.2016.08.002

Mohammadi, M., and Hedayati, M. (2017). A brief review on the molecular basis of medullary thyroid carcinoma. *Cell J.* 18, 485–492.

Mustafa, T., Klonisch, T., Hombach-Klonisch, S., Kehlen, A., Schmutzler, C., Koehrle, J., et al. (2004). Expression of CD97 and CD55 in human medullary thyroid carcinomas. *Int. J. Oncol.* 24, 285–294.

Nocturne, G., and Mariette, X. (2015). Sjogren Syndrome-associated lymphomas: an update on pathogenesis and management. *Br. J. Haematol.* 168, 317–327. doi: 10.1111/bjh.13192

Oczko-Wojciechowska, M., Czarniecka, A., Gawlik, T., Jarzab, B., and Krajewska, J. (2020). Current status of the prognostic molecular markers in medullary thyroid carcinoma. *Endocr. Connect.* 9, R251–R263. doi: 10.1530/EC-20-0374

Pankratova, E. V., Stepchenko, A. G., Krylova, I. D., Portseva, T. N., and Georgieva, S. G. (2018). The regulatory interplay between Oct-1 isoforms contributes to hematopoiesis and the isoforms imbalance correlates with a malignant transformation of B cells. *Oncotarget* 9, 29892–29905. doi: 10.18632/oncotarget.25648

Pellegriti, G., Frasca, F., Regalbuto, C., Squatrito, S., and Vigneri, R. (2013). Worldwide increasing incidence of thyroid cancer: update on epidemiology and risk factors. *J. Cancer Epidemiol.* 2013, 965212–965212.

Peng, M., Wei, G., Zhang, Y., Li, H., Lai, Y., Guo, Y., et al. (2021). Single-cell transcriptomic landscape reveals the differences in cell differentiation and immune microenvironment of papillary thyroid carcinoma between genders. *Cell Biosci.* 11:39. doi: 10.1186/s13578-021-00549-w

Pitarresi, J. R., Liu, X., Sharma, S. M., Cuitino, M. C., Kladney, R. D., Mace, T. A., et al. (2016). Stromal ETS2 regulates chemokine production and immune cell recruitment during acinar-to-ductal metaplasia. *Neoplasia* 18, 541–552. doi: 10.1016/j.neo.2016.07.006

Qu, Y., Liao, Z., Wang, X., Zhang, J., and Liu, C. (2019). EFLDO sensitizes liver cancer cells to TNFSF10induced apoptosis in a p53dependent manner. *Mol. Med. Rep.* 19, 3799–3806. doi: 10.3892/mmr.2019.10046

Ramachandran, P., Dobie, R., Wilson-Kanamori, J. R., Dora, E. F., Henderson, B. E. P., Luu, N. T., et al. (2019). Resolving the fibrotic niche of human liver cirrhosis at single-cell level. *Nature* 575, 512–518. doi: 10.1038/s41586-019-1631-3

Resende de Paiva, C., Grønhøj, C., Feldt-Rasmussen, U., and von Buchwald, C. (2017). Association between Hashimoto's thyroiditis and thyroid cancer in 64,628 patients. *Front. Oncol.* 7:53. doi: 10.3389/fonc.2017.00053

Rothe, M., Sarma, V., Dixit, V. M., and Goeddel, D. V. (1995). TRAF2-mediated activation of NF-kappa B by TNF receptor 2 and CD40. *Science* 269, 1424–1427. doi: 10.1126/science.7544915

Smallridge, R. C., Chindris, A. M., Asmann, Y. W., Casler, J. D., Serie, D. J., Reddi, H. V., et al. (2014). RNA sequencing identifies multiple fusion transcripts, differentially expressed genes, and reduced expression of immune function genes in BRAF (V600E) mutant vs BRAF wild-type papillary thyroid carcinoma. *J. Clin. Endocr. Metab.* 99, E338–E347. doi: 10.1210/jc.2013-2792

Suo, S. B., Zhu, Q., Saadatpour, A., Fei, L. J., Guo, G. J., and Yuan, G. C. (2018). Revealing the critical regulators of cell identity in the mouse cell atlas. *Cell Rep.* 25, 1436–1445.e3.

Suomela, S., Cao, L., Bowcock, A., and Saarialho-Kere, U. (2004). Interferon alpha-inducible protein 27 (IFI27) is upregulated in psoriatic skin and certain epithelial cancers. *J. Invest. Dermatol.* 122, 717–721. doi: 10.1111/j.0022-202X.2004.22322.x

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6, 377–382.

Wang, S. S., Liu, W., Ly, D., Xu, H., Qu, L., and Zhang, L. (2019). Tumor-infiltrating B cells: their role and application in anti-tumor immunity in lung cancer. *Cell. Mol. Immunol.* 16, 6–18. doi: 10.1038/s41423-018-0027-x

Wang, W. B., Su, X. Y., He, K. F., Wang, Y. L., Wang, H. Y., Wang, H. H., et al. (2016). Comparison of the clinicopathologic features and prognosis of bilateral versus unilateral multifocal papillary thyroid cancer: an updated study with

more than 2000 consecutive patients. *Cancer Am. Cancer Soc.* 122, 198–206. doi: 10.1002/cncr.29689

Wieland, A., Patel, M., Cardenas, M., Eberhardt, C., Hudson, W., Obeng, R., et al. (2020). Defining HPV-specific B cell responses in patients with head and neck cancer. *Nature* 597, 274–278. doi: 10.1038/s41586-020-2931-3

Wouters, M. C. A., and Nelson, B. H. (2018). Prognostic significance of tumor-infiltrating B cells and plasma cells in human cancer. *Clin. Cancer Res.* 24, 6125–6135. doi: 10.1158/1078-0432.ccr-18-1481

Wu, C. C., Lin, J. D., Chen, J. T., Chang, C. M., Weng, H. F., Hsueh, C., et al. (2018). Integrated analysis of fine-needle-aspiration cystic fluid proteome, cancer cell secretome, and public transcriptome datasets for papillary thyroid cancer biomarker discovery. *Oncotarget* 9, 12079–12100. doi: 10.18632/oncotarget.23951

Xing, M. Z. (2010). Prognostic utility of BRAF mutation in papillary thyroid cancer. *Mol. Cell Endocrinol.* 321, 86–93. doi: 10.1016/j.mce.2009.10.012

Xing, M. Z., Liu, R. Y., Liu, X. L., Murugan, A. K., Zhu, G. W., Zeiger, M. A., et al. (2014). BRAF V600E and TERT Promoter mutations cooperatively identify the most aggressive papillary thyroid cancer with highest recurrence. *J. Clin. Oncol.* 32, 2718–2726. doi: 10.1200/JCO.2014.55.5094

Yang, D., Wang, L. L., Dong, T. T., Shen, Y. H., Guo, X. S., Liu, C. Y., et al. (2015). Progranulin promotes colorectal cancer proliferation and angiogenesis through TNFR2/Akt and ERK signaling pathways. *Am. J. Cancer Res.* 5, 3085–3097.

Yin, S., Chen, F. F., Ye, P., and Yang, G. F. (2018). Overexpression of FAM3C protein as a novel biomarker for epithelial-mesenchymal transition and poor outcome in gastric cancer. *Int. J. Clin. Exp. Pathol.* 11, 4247–4256.

Zhang, G. P., Yue, X., and Li, S. Q. (2020). Cathepsin C interacts with TNF-alpha/p38 MAPK signaling pathway to promote proliferation and metastasis in hepatocellular carcinoma. *Cancer Res Treat.* 52, 10–23. doi: 10.4143/crt.2019.145

Zhang, Q., He, Y., Luo, N., Patel, S. J., Han, Y., Gao, R., et al. (2019). Landscape and dynamics of single immune cells in hepatocellular carcinoma. *Cell* 179, 829–845.e20.

Zhang, Y., Dai, J., Wu, T., Yang, N., and Yin, Z. (2014). The study of the coexistence of Hashimoto's thyroiditis with papillary thyroid carcinoma. *J. Cancer Res. Clin. Oncol.* 140, 1021–1026.

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: 12 November 2021
doi: 10.3389/fcell.2021.762260

Deconvolution of Bulk Gene Expression Profiles with Single-Cell Transcriptomics to Develop a Cell Type Composition-Based Prognostic Model for Acute Myeloid Leukemia

*Chengguqiu Dai, Mengya Chen, Chaolong Wang and Xingjie Hao**

Department of Epidemiology and Biostatistics, Key Laboratory for Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Acute myeloid leukemia (AML) is one of the malignant hematologic cancers with rapid progress and poor prognosis. Most AML prognostic stratifications focused on genetic abnormalities. However, none of them was established based on the cell type compositions (CTCs) of peripheral blood or bone marrow aspirates from patients at diagnosis. Here we sought to develop a novel prognostic model for AML in adults based on the CTCs. First, we applied the CIBERSORT algorithm to estimate the CTCs for patients from two public datasets (GSE6891 and TCGA-LAML) using a custom gene expression signature reference constructed by an AML single-cell RNA sequencing dataset (GSE116256). Then, a CTC-based prognostic model was established using least absolute shrinkage and selection operator Cox regression, termed CTC score. The constructed prognostic model CTC score comprised 3 cell types, GMP-like, HSC-like, and T. Compared with the low-CTC-score group, the high-CTC-score group showed a 1.57-fold [95% confidence interval (CI), 1.23 to 2.00; $p = 0.0002$] and a 2.32-fold (95% CI, 1.53 to 3.51; $p < 0.0001$) higher overall mortality risk in the training set (GSE6891) and validation set (TCGA-LAML), respectively. When adjusting for age at diagnosis, cytogenetic risk, and karyotype, the CTC score remained statistically significant in both the training set [hazard ratio (HR) = 2.25; 95% CI, 1.20 to 4.24; $p = 0.0119$] and the validation set (HR = 7.97; 95% CI, 2.95 to 21.56; $p < 0.0001$]. We further compared the performance of the CTC score with two gene expression-based prognostic scores: the 17-gene leukemic stem cell score (LSC17 score) and the AML prognostic score (APS). It turned out that the CTC score achieved comparable performance at 1-, 2-, 3-, and 5-years timepoints and provided independent and additional prognostic information different from the LSC17 score and APS. In conclusion, the CTC score could serve as a powerful prognostic marker for AML and has great potential to assist clinicians to formulate individualized treatment plans.

Keywords: cell type composition, gene expression profiles, transcriptome deconvolution, prognostic model, acute myeloid leukemia

OPEN ACCESS

Edited by:
Lu Xie,
Shanghai Center For Bioinformation Technology, China

Reviewed by:
Sheng Yang,
Nanjing Medical University, China
Shiquan Sun,
Xi'an Jiaotong University, China

****Correspondence:***
Xingjie Hao
xingjie@hust.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: *21 August 2021*
Accepted: *18 October 2021*
Published: *12 November 2021*

Citation:
Dai C, Chen M, Wang C and Hao X (2021) Deconvolution of Bulk Gene Expression Profiles with Single-Cell Transcriptomics to Develop a Cell Type Composition-Based Prognostic Model for Acute Myeloid Leukemia. Front. Cell Dev. Biol. 9:762260. doi: 10.3389/fcell.2021.762260

INTRODUCTION

Acute myeloid leukemia (AML) is characterized by malignant clonal hematopoiesis, which is caused by the accumulation of somatic mutations in hematopoietic stem cells (HSCs) or downstream progenitors (Yamashita et al., 2020). Among diverse leukemia subtypes, AML accounts for most leukemia patients and leukemia-related deaths, and the incidence has been continuously increasing in recent years (Ghazawi et al., 2019; Roman et al., 2016; Shallis et al., 2019). The average 5-years overall survival (OS) probability is approximately 24% by 2016 in the United States, the fifth worst by cancer types, and 17% between 2000 and 2007 in Europe (De Angelis et al., 2015; Shallis et al., 2019). Therefore, accurately stratifying the prognosis is of great significance to formulate individualized treatment plans for AML patients.

As high-throughput sequencing technology becomes affordable, the comprehensive landscape of AML driver mutations has been gradually revealed (Cancer Genome Atlas Research et al., 2013; Papaemmanuil et al., 2016). Identifying the genetic abnormalities, including cytogenetic alterations and molecular variants, greatly contributes to the prognostic assessments for AML patients at diagnosis (Grimwade et al., 2010; Marcucci et al., 2011). Nevertheless, existing prognostic stratifications, such as the 2017 European LeukemiaNet (ELN) risk stratification (Dohner et al., 2017), still require further improvement due to the diversity and heterogeneity of the AML-related genetic abnormalities within and across patients. Some studies attempted to seek novel prognostic markers using gene expression profiles (GEPs), such as the 17-gene leukemic stem cell score (LSC17 score) (Ng et al., 2016) and the AML prognostic score (APS) (Docking et al., 2021). Some of these expression-based prognostic markers showed great performances in evaluating prognosis for AML patients. However, it is difficult to interpret how the genes used to compute the prognostic score affect the prognosis.

It has been suggested that the cell type compositions (CTCs) in the tumor microenvironment are associated with tumor growth, progression, invasion, and metastasis (Hanahan and Weinberg., 2011). Recently, with the application of single-cell sequencing technology in AML, 21 cell types in the bone marrow samples of AML patients were identified, of which six were malignant (van Galen et al., 2019). In addition, it suggested that the CTCs of AML were associated with specific genetic mutation types and different prognoses (van Galen et al., 2019). Therefore, it seemed feasible to construct a novel AML prognostic score based on the CTCs, and how the CTC-based prognostic score would perform remained to be further studied. Experimental methods to acquire the CTCs of samples, including flow cytometry (FCM) (Adan et al., 2017) and single-cell RNA sequencing (scRNA-seq) (Potter., 2018), are costly and infeasible with a large sample size at present. Luckily, increasing computational methods have been developed to infer the CTCs through bulk GEPs (Avila Cobos et al., 2018)—for example, CIBERSORT uses the support vector regression algorithm to deconvolute the bulk GEPs into CTCs based on a reference matrix that comprises the gene expression signatures (GES) of cell types of interest (Newman et al., 2015).

In this study, we aimed to develop a novel prognostic model for *de novo* AML in adults based on the CTCs of patients at diagnosis. Firstly, we constructed a cell type-specific GES reference matrix by conducting a differential expression analysis using AML scRNA-seq profiles. Then, we deconvoluted the bulk GEPs of two AML datasets to CTCs based on the custom reference matrix. Finally, we constructed and evaluated an AML prognostic model, termed CTC score, based on the estimated CTCs. The CTC score showed a comparable performance to previous gene expression-based prognostic models and could act as an independent prognostic factor for AML. In addition, we demonstrated that the CTC score provided additional prognostic information different from LSC17 and APS.

MATERIALS AND METHODS

Data Sources

We downloaded a scRNA-seq dataset, GSE116256 (van Galen et al., 2019), and two bulk gene expression datasets, GSE6891 (Verhaak et al., 2009) and TCGA-LAML (Cancer Genome Atlas Research et al., 2013), for AML from the Gene Expression Omnibus (GEO) data repository (RRID:SCR_005012) and Genomic Data Commons data portal (RRID:SCR_014514), respectively. The scRNA-seq dataset contains single-cell GEPs and cell annotations of 30,712 cells and 27,899 genes from the bone marrow aspirates of 16 AML patients. The cell annotations comprise information such as the number of unique molecular identifiers (UMIs), the number of expressed genes, and the inferred cell type for each cell. A total of 21 cell types were defined, including HSC, HSC-like, progenitor (Prog), Prog-like, granulocyte–monocyte–progenitor (GMP), GMP-like, promonocyte (ProMono), ProMono-like, monocyte (Mono), Mono-like, conventional dendritic cell (cDC), cDC-like, plasmacytoid dendritic cell (pDC), early erythroid progenitor (earlyEry), late erythroid progenitor (lateEry), progenitor B cell (proB), mature B cell (B), plasma cell (plasma), naïve T cell (T), cytotoxic T lymphocyte (CTL), and natural killer cell (NK). Details of the scRNA-seq dataset can be learned from elsewhere (van Galen et al., 2019). For the bulk gene expression dataset GSE6891, 537 GEPs of AML patients profiled by microarray were obtained. For TCGA-LAML, 151 GEPs with fragments per kilobase million normalization were downloaded. The corresponding clinical characteristics and survival information for each sample were downloaded from the cBioPortal database (RRID:SCR_014555).

Study Design

The workflow of this study is illustrated in **Figure 1**. We first constructed the GES reference matrix of the 21 cell types required in CIBERSORT (Newman et al., 2015) (RRID:SCR_016955) using AML scRNA-seq profiles. The CTCs of patients in the bulk gene expression datasets of GSE6891 and TCGA-LAML were subsequently estimated. A CTC-based prognostic model was established, with GSE6891 as the training set, and was validated in TCGA-LAML subsequently.

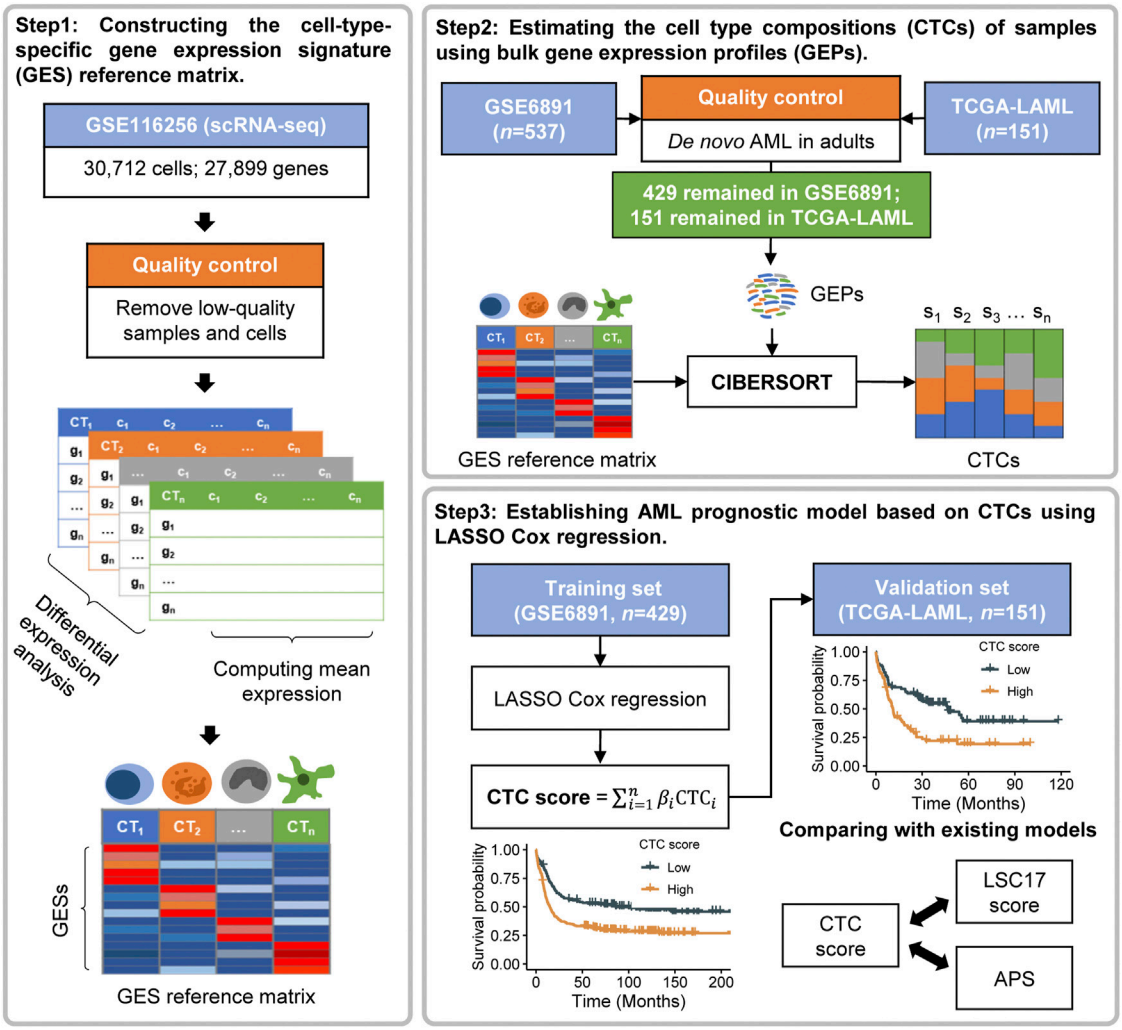


FIGURE 1 | Overview of the workflow.

Data Preprocessing and Quality Control

For the scRNA-seq dataset GSE116256, we excluded cells derived from samples of AML314, AML371, AML722B, and AML997 due to the unconfident cell type annotations (van Galen et al., 2019). Next, we computed the ratio of UMI counts to the number of expressed genes for each cell, termed UTG ratio below. In each cell type, cells with outlier values of UTG ratio were prone to be low in quality. The threshold to filter such cells was determined to be the median UTG ratio plus–minus three times the median absolute deviation (Leys et al., 2013). A total of 27,023 cells remained (**Supplementary Figure S1**).

The bulk GEPs of GSE6891 were generated by Affymetrix Human Genome U133 Plus 2.0 Array (Verhaak et al., 2009). The raw CEL files were processed using affy (version 1.66.0) and normalized by the Gene Chip Robust Multi-array Average (Wu et al., 2004) algorithm using gcrma (version 2.60.0) Bioconductor R package. The probe set IDs were transformed to the corresponding gene symbols according to the chip definition file (GEO accession: GPL570). The probe sets that did not match any gene symbols or matched multiple gene symbols were filtered out. To retain enough genes for subsequent analysis, we computed the mean expression of probe sets that matched the same gene and chose the probe set with the highest average gene expression to represent that gene (Ng et al., 2016). Among the cases in GSE6891, we only retained *de novo* AML cases whose age at diagnosis were greater or equal to 18 with completed survival information.

For TCGA-LAML, the ensemble gene IDs of the downloaded GEPs were transformed to gene symbols according to the comprehensive gene annotation files of GENCODE release 38 (GRCh38.p13; RRID:SCR_014966) in gene transfer format. We filtered out the ensemble gene IDs matching the same gene symbol due to the difficulty in determining which ensemble gene ID to represent that gene. Among the cases in TCGA-LAML, we took the same filtering criteria as implemented for the GSE6891 dataset.

Constructing the Cell Type-Specific Gene Expression Signatures Reference Matrix

We constructed the cell type-specific GES reference matrix based on the AML scRNA-seq GEPs using Seurat (Stuart et al., 2019) (version 3.2.0) R package. First, the single-cell GEPs of AML patients were integrated and imported into a Seurat object. All cells were labeled as the cell type in the annotation file. Then, we normalized the UMI counts to counts per million (CPM) and performed natural-log transformation [log (CPM+1)]. Subsequently, we conducted the differential expression analysis using FindAllMarkers function to acquire highly expressed genes of each cell type by comparing the cells of 1 cell type against all others in turn. The tests of comparisons between groups used the "bimod" method, a likelihood-ratio test for single-cell gene expression (McDavid et al., 2013). The "min.pct" parameter was set to 0. Other parameters were set as default. The acquired highly expressed genes of each cell type with the adjusted *p*-values lower than 0.05 and the average natural-log fold-change (logFC) above 1 were retained (**Supplementary Figure S2** and **Supplementary Table S1**). Notably, the highly expressed genes selected to build the GES reference matrix are the dominant influence factor for CTC estimations, thereby affecting subsequent modeling. Therefore, we extracted the top 25, 50, 100, and 150 most significantly highly expressed genes for each cell type and computed the mean expression by cell type to build 4 cell type-specific GES reference matrices (GES25, GES50, GES100, and GES150; **Supplementary Figure S3**, **Supplementary Table S2**, **Supplementary Table S3**, **Supplementary Table S4**, and **Supplementary Table S5**) (Donovan et al., 2020).

Simulations to Examine the Accuracy of CIBEROSRT and Gene Expression Signatures Matrices

We performed a simulation analysis to examine the accuracy of CIBERSORT using the custom cell type-specific GES reference matrices. Specifically, we first generated 100 artificial samples using scRNA-seq profiles. For each sample, we selected a random number of cells for each cell type from at least 50 to the maximum number of cells for that cell type through the cell barcodes (Donovan et al., 2020). The normalized GEPs of these cells were summed to create the artificial sample with known cell type compositions. Subsequently, we ran CIBERSORT on these artificial samples using different GES matrices. Additionally, two other deconvolution methods, MuSiC (Wang et al., 2019) and MOMF (Sun et al., 2019), were also used for comparisons. The Pearson correlation coefficients of the real proportions and the estimated proportions were computed by each cell type as the metric of accuracy.

Estimating Cell Type Compositions Using CIBERSORT

The simulation results showed that the performances of CIBERSORT, MuSiC, and MOMF were similar (**Supplementary Figure S4**). However, we noticed that MuSiC and MOMF took a much longer running time and much more memory consumptions (data not shown). Accordingly, we chose CIBEROSRT to estimate the relative proportions of 21 AML cell types for the bulk gene expression datasets GSE6891 and TCGA-LAML, setting 100 permutations and disabling the quantile normalization option.

Constructing an Acute Myeloid Leukemia Prognostic Model Based on Cell Type Compositions

After estimating the CTCs (**Supplementary Table S6**, **Supplementary Table S7**, **Supplementary Table S8**, and **Supplementary Table S9**), we found that the estimated proportions of some cell types were almost 0 for most of the samples, probably due to estimation error. To reduce the influence on subsequent modeling, we converted the cell types whose mean proportions were lower than 0.05 or median proportions were equal to 0 to dichotomous variables, with 0 as the cutoff value. Cell types converted to dichotomous or that remained continuous in both datasets and whose Pearson correlation coefficient was $r > 0.8$ in the simulations were used to train and validate the prognostic model.

The bulk gene expression dataset GSE6891 was set as the training set, and TCGA-LAML was set as the validation set to establish and validate a novel prognostic model for AML based on CTCs. With OS as the survival outcome, we performed the least absolute shrinkage and selection operator (LASSO) Cox regression (Simon et al., 2011) and 10-fold cross-validation using glmnet (version 4.1–1) R package. To obtain a robust model, we repeated this process 100 times using different random seeds, and cell types with non-zero coefficients in at least 95 fittings were retained. The coefficients of 100 fitting processes for the retained cell types were averaged as the final coefficient (Elsayed et al., 2020). The linear combination of the selected cell types in the LASSO Cox regression model weighted by the coefficients served as the prognostic marker for AML, called CTC score. For better interpretation and visualization, we partitioned all patients into low- and high-CTC-score groups by median.

The established CTC score was validated in TCGA-LAML. We computed the CTC scores for patients in TCGA-LAML based on the linear equation above (**Supplementary Table S10**). We likewise partitioned the patients in the validation set into low- and high-CTC-score groups based on the median. Kaplan–Meier curves were used to display the different prognoses between low- and high-CTC-score groups.

We considered displaying the CTC score established on the CTCs estimated with GES100 as the reference matrix to be the main results. Other prognostic models based on the CTCs estimated using GES25, GES50, and GES150 were considered as sensitivity analysis and could be accessible in **Supplementary Figure S5**. The Harrell's concordance index (C-index) was used to compare the performance of these models (Harrell et al., 1996).

Verifying the Prognostic Independence of the Cell Type Composition Score

We found that GMP-like has a great weight when computing the CTC score (see results part). It has been reported that GMP-like is

highly associated with two abnormal karyotypes (i.e., *PML-RARA* and *RUNX1-RUNX1T1*), both of which indicate a favorable prognosis (Appelbaum et al., 2006; Wang and Chen., 2008; van Galen et al., 2019). Thus, it is crucial to verify whether the prognostic significance of CTC score was dominantly captured by existing prognostic factors such as karyotypes and cytogenetic risk classifications. To verify this point, we first implemented univariable Cox regressions for clinical characteristics. The clinical characteristics significant in both training and validation dataset and CTC score were introduced to multivariable Cox regressions using survival (version 3.2–7) R package.

Comparing the Cell Type Composition Score with the LSC17 Score and Acute Myeloid Leukemia Prognostic Score

We further evaluated the performance of CTC score by comparing it with the LSC17 score and APS. The LSC17 score was constructed by the expression of 17 genes highly expressed in LSCs, while the APS was constructed by the expression of 16 genes acquired by LASSO Cox regression (Ng et al., 2016; Docking et al., 2021). The LSC17 score and APS for patients in the validation set TCGA-LAML were computed in compliance with the data processing flow and calculation equation according to the original articles (**Supplementary Table S10**) (Ng et al., 2016; Docking et al., 2021). Considering the comparability, all three prognostic scores were not converted to dichotomous variables. We implemented the time-dependent receiver operating characteristic (ROC) curve analysis to evaluate and compare the predictive accuracy using area under the ROC curve (AUC) as the indicator. The predictive sensitivities and specificities of CTC score, LSC17 score, and APS at 1-, 2-, 3-, and 5-years timepoints were computed and compared using timeROC (Blanche et al., 2013) (version 0.4) R package.

Statistical Analysis

For the clinical characteristics of patients in the bulk gene expression datasets GSE6891 and TCGA-LAML, continuous variables were described by medians and ranges, and categorical variables were described by frequencies and proportions. We used the Wilcoxon test or Kruskal–Wallis test for group comparisons of continuous variables and the chi-square test or Fisher's exact test for that of categorical variables. All statistical tests were two-tailed, and *p*-values lower than 0.05 were considered statistically significant. All the analyses were performed in R-4.0.2.

RESULTS

Clinical Characteristics and Cell Type Compositions for Two Bulk Acute Myeloid Leukemia Datasets

For the bulk gene expression dataset GSE6891, 11 patients whose age at diagnosis was lower than 18, 17 patients of myelodysplastic syndrome, and four patients with missing survival information were filtered out. Eventually, 429 patients were eligible, whereas all patients in TCGA-LAML passed the filtering criteria. The descriptive characteristics of patients in these two datasets are shown in **Table 1**. Patients in GSE6891 were younger than those in TCGA-LAML ($p < 0.0001$). FAB classification ($p = 0.0010$) and cytogenetic risk ($p = 0.0313$) were also different between GSE6891 and TCGA-LAML. Patients in GSE6891 comprises more FAB-M5 subtype (23.3% in GSE6891 *vs* 9.9% in TCGA-LAML) and less poor cytogenetic risk strata (19.3% in GSE6891 *vs* 23.8% in TCGA-LAML). The CTCs for patients in the GSE6891 and TCGA-LAML datasets estimated with GES100 as the reference matrix are shown in **Supplementary Figure S6**.

TABLE 1 | Characteristics of acute myeloid leukemia patients in the training set and the validation set.

Characteristic	GSE6891 (Training set; *n* = 429)	TCGA-LAML (Validation set; *n* = 151)	*p*-value
Age at diagnosis, years			<0.0001
Median (range)	44 (18–60)	56 (21–88)	
≤55	361 (84.1)	74 (49.0)	
>55	68 (15.9)	77 (51.0)	
Sex			0.3233
Female	218 (50.8)	69 (45.7)	
Male	211 (49.2)	82 (54.3)	
FAB classification			0.0010
M0	16 (3.7)	15 (9.9)	
M1	94 (21.9)	36 (23.8)	
M2	100 (23.3)	37 (24.5)	
M3	24 (5.6)	15 (9.9)	
M4	81 (18.9)	29 (19.2)	
M5	100 (23.3)	15 (9.9)	
M6	6 (1.4)	2 (1.3)	
M7	0 (0)	1 (0.7)	
NA	8 (1.9)	1 (0.7)	
Cytogenetic risk			0.0313
Good	91 (21.2)	31 (20.5)	
Intermediate	245 (57.1)	81 (53.6)	
Poor	83 (19.3)	36 (23.8)	
NA	10 (2.3)	3 (2.0)	
Karyotype			0.0964
Others	351 (81.8)	127 (84.1)	
PML-RARA	21 (4.9)	14 (9.3)	
RUNX1-RUNX1T1	32 (7.5)	7 (4.6)	
NA	25 (5.8)	3 (2.0)	

Patients with missing values were removed before performing the statistical tests. Chi-square tests were implemented to compare the constituent ratios of characteristics between the training set GSE6891 and the validation set TCGA-LAML, except for FAB classification, for which Fisher's exact test was conducted.
FAB, French–American–British; NA, not available; CTC, cell type composition.

Cell Type Composition-Based Prognostic Score for Acute Myeloid Leukemia

The median follow-up time of patients in the bulk gene expression datasets GSE6891 and TCGA-LAML was 20.11 months [interquartile range (IQR), 7.89–92.78 months] and 19 months (IQR, 6.45–42.1 months), respectively. We fitted a LASSO Cox regression model and defined the CTC score computed by the following equation:

TABLE 2 | Univariable Cox regression with overall survival as the outcome.

Characteristic	GSE6891 (Training set, *n* = 429)		TCGA-LAML (Validation set, *n* = 151)	
	HR (95% CI)	*p*-value	HR (95% CI)	*p*-value
Age at diagnosis, years				
≤55	Reference		Reference	
>55	1.83 (1.36–2.47)	<0.0001	2.71 (1.79–4.11)	<0.0001
Sex				
Female	Reference		Reference	
Male	0.94 (0.74–1.19)	0.6002	1.01 (0.68–1.51)	0.9465
FAB classification				
M0	2.14 (0.96–4.79)	0.0632	3.76 (1.18–12.04)	0.0256
M1	1.43 (0.75–2.72)	0.2770	3.73 (1.29–10.81)	0.0152
M2	1.40 (0.74–2.66)	0.3046	3.33 (1.15–9.64)	0.0262
M3	Reference		Reference	
M4	1.28 (0.67–2.47)	0.4574	3.93 (1.34–11.53)	0.0126
M5	1.66 (0.88–3.14)	0.1186	4.57 (1.42–14.67)	0.0106
M6	0.89 (0.25–3.18)	0.8532	9.69 (1.74–53.99)	0.0095
M7	NA	NA	7.83 (0.86–71.13)	0.0675
Cytogenetic risk				
Good	Reference		Reference	
Intermediate	1.99 (1.39–2.84)	0.0002	3.11 (1.58–6.10)	0.0010
Poor	3.40 (2.27–5.10)	<0.0001	4.36 (2.10–9.03)	<0.0001
Karyotype				
Others	Reference		Reference	
PAML-RARA	0.39 (0.18–0.82)	0.0136	0.28 (0.10–0.78)	0.0143
RUNX1-RUNX1T1	0.39 (0.22–0.70)	0.0017	0.45 (0.14–1.44)	0.1800
CTC score				
Low	Reference		Reference	
High	1.57 (1.23–2.00)	0.0002	2.31 (1.53–3.51)	<0.0001

Patients with missing values were removed before performing the statistical tests.
HR, hazard ratio; CI, confidence interval; FAB, French–American–British; NA, not available; CTC, cell type composition.

CTC score = (−1.7016 × GMP-like) + (0.2015 × HSC-like) + (−0.293 × T), where HSC-like and T were dichotomous. The negative coefficient of GMP-like indicated that lower relative proportions of GMP-like at diagnosis would predict worse survival outcomes. The estimated HSC-like greater than 0 and T equal to 0 would predict worse prognoses.

Comparing with the low-CTC-score group, the high-CTC-score group showed a 1.57-fold (95% CI, 1.23 to 2.00; $p = 0.0002$) higher overall mortality risk in the training set GSE6891 and 2.32-fold (95% CI, 1.53 to 3.51; $p < 0.0001$) in the validation set TCGA-LAML (**Table 2**). The 5-years OS rate for GSE6891 was 47.7% (95% CI, 41.4–54.9%) in the low-CTC-score group and 31.1% (95% CI, 25.5–37.9%) in the high-CTC-score group. For TCGA-LAML, the 5-years OS rate was 41.2% (95% CI, 29.7–57.1%) and 17.7% (95% CI, 10.2–30.7%) in the low-CTC-score group and high-CTC-score group, respectively (**Figure 2**).

The individual-level results of CTCs estimated using GES25, GES50, GSE100, and GES150 could be obtained in **Supplementary Table S6**, **Supplementary Table S7**, **Supplementary Table S8**, and **Supplementary Table S9**. As displayed in **Supplementary Figure S5**, the CTC-based scores established by reference matrices with different GES matrices were robustly associated with the OS of AML in the validation set, with C-index ranging from 0.64 (95% CI, 0.58–0.70) to 0.67 (95% CI, 0.61–0.73).

Cell Type Composition Score Is an Independent Factor in Predicting Acute Myeloid Leukemia Prognosis

We performed univariable and multivariable Cox regressions in both the training and validation sets to test whether the CTC score is an independent factor associated with the OS for AML in adults. Among the clinical characteristics, age at diagnosis, cytogenetics risk, and karyotype were significantly associated with OS in both datasets (**Table 2**). The multivariable Cox regression results showed that CTC score remained statistically significant in GSE6891 (HR = 2.25; 95% CI, 1.20 to 4.24; $p = 0.0119$) and TCGA-LAML (HR = 7.97; 95% CI, 2.95 to 21.56; $p < 0.0001$) when adjusting for age at diagnosis, cytogenetic risk, and karyotype (**Figure 3**). These results suggested that CTC score can predict the prognosis of AML independent of age at diagnosis, cytogenetic risk, and karyotype.

Cell Type Composition Score Provides Additional Prognostic Information Different from LSC17 and Acute Myeloid Leukemia Prognostic Score

In TCGA-LAML, we evaluated the predictive accuracy of 1-, 2-, 3-, and 5-years OS using ROC curves. The corresponding AUCs

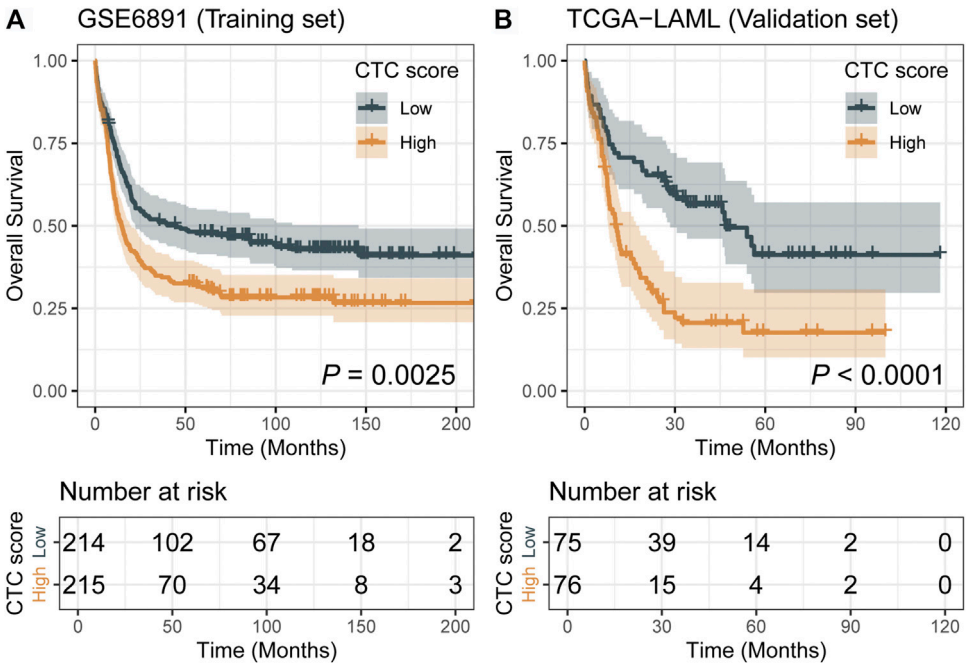


FIGURE 2 | Kaplan–Meier curves of low-cell type composition (CTC)-score and high-CTC-score groups for the training set GSE6891 **(A)** and the validation set TCGA-LAML **(B)**.

and 95% CIs for CTC score, LSC17 score, and APS were computed as shown in **Figure 4**. The differences in AUCs of CTC score *versus* LSC17 score and CTC score *versus* APS at four time points were not statistically significant (**Supplementary Table S11**), suggesting that CTC score can achieve a similar predictive accuracy compared with LSC17 score and APS. Additionally, we simultaneously included CTC score, LSC17 score, and APS into the multivariable Cox regression (**Figure 5**). CTC score (HR = 3.65; 95% CI, 1.37 to 9.7; p = 0.0095) and APS (HR = 1.84; 95% CI, 1.06 to 3.18; p = 0.0297) remained statistically significant, suggesting that both CTC score and APS could capture additional prognostic information compared with LSC17 score. Furthermore, the additional prognostic information captured by the CTC score was different from that captured by APS.

DISCUSSION

In the present study, we have constructed an AML prognostic score based on the assumption that the CTCs of AML patients at diagnosis can reflect the genetic abnormalities and are thus correlated with their prognosis (van Galen et al., 2019). To estimate CTCs, we first constructed a cell type-specific GES reference matrix GES100 through a differential expression analysis of the AML scRNA-seq dataset. Then, we applied the CIBERSORT algorithm to deconvolute the bulk GEPs of AML samples to CTCs by the custom GES reference matrix. Subsequently, an AML prognostic score based on the CTCs (i.e., CTC score) comprising 3 cell types, GMP-like, HSC-like, and T, was established for *de novo* AML in adults. CTC score was significantly associated with the OS in both the training set and the validation set.

Previous studies applying CIBERSORT to estimate the immune microenvironment for AML all used LM22, which contains the GESs of 22 immunocytes provided by the author as the reference matrix (Newman et al., 2015; Xu et al., 2020; Cheng et al., 2021; Jia et al., 2021). However, the estimates of CTCs might be inaccurate in these studies because of the resemblance between normal immunocytes and malignant leukemic blasts, especially for the myeloid lineages—for example, both Xu et al. (2020) and Cheng et al. (2021) identified that higher relative proportions of M2 macrophage were associated with a poorer prognosis for AML. Additionally, Xu et al. (2020) suggested the marker gene of M2 macrophage CD206, also presenting in immature dendritic cells (DCs) (Wollenberg et al., 2002), as a novel prognostic predictor. However, we found that CD206 was highly expressed in cDC-like (**Supplementary Figure S7**). Thus, the estimated proportions for M2 macrophage might be overestimated due to the similarity between cDC-like and M2 macrophage when using LM22 as the reference. To fix this issue, we constructed custom GES reference matrices containing all 21 cell types of the bone marrow annotated by the single-cell GEPs. In this manner, the estimated relative proportions of using CIBERSORT could reflect the real proportions of each cell type in the sample. When considering both the normal and the malignant cell types in AML samples, the established CTC score showed a powerful prognostic significance.

We noticed that the coefficient of GMP-like in CTC score was greater than the other 2 cell types. It has been revealed before that

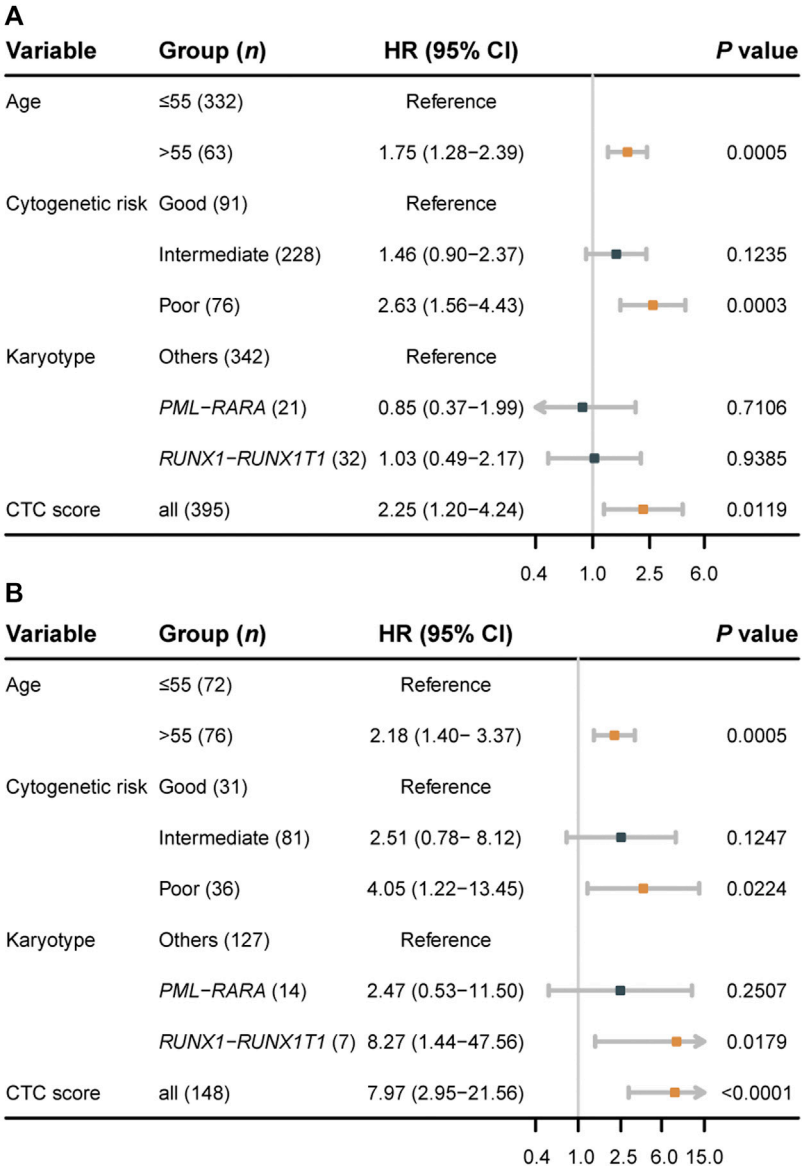


FIGURE 3 | Forest plots of multivariable Cox regression results for the training set GSE6891 **(A)** and the validation set TCGA-LAML **(B)**.

GMP-like was associated with PML-RARA and RUNX1–RUNX1T1 fusion in the TCGA-LAML dataset (van Galen et al., 2019). This finding was repeated in the bulk gene expression dataset GSE6891 (**Supplementary Figure S8**). Researchers found that the PML–RARA fusion leads to a block in the differentiation of myeloid cells at the promyelocytic stage (Grisolano et al., 1997). In recent decades, the PML–RARA fusion-induced AML has become highly curable since the broad application of target chemotherapy drugs, all-trans retinoic acid and arsenic trioxide, into clinical use (Wang and Chen., 2008). The RUNX1–RUNX1T1 fusion-induced AML has also been determined to have a good prognosis (Appelbaum et al., 2006). It is characterized by the expressed myeloperoxidase, a protein expressed mainly in neutrophils, in more than 90% of leukemia blasts (Schlaifer et al., 1993; Aratani., 2018). Both of these two gene fusions are considered to be of good prognosis in cytogenetic risk classification (Slovak et al., 2000). In other words, the CTC score is probably confounded by these two gene fusions for the great weight of GMP-like. Analogously, other covariates imbalanced such in the training and validation sets as the cytogenetic risk might also confound the results. Therefore, it is crucial to figure out whether the CTC score can provide additional and independent prognostic information to AML prognosis in comparison to the existing classifications. In our study, we have justified this by conducting multivariable Cox regression analyses. We introduced age at diagnosis, karyotype, and cytogenetic risk as covariates for both the training and validation datasets, and the CTC score remained statistically significant.

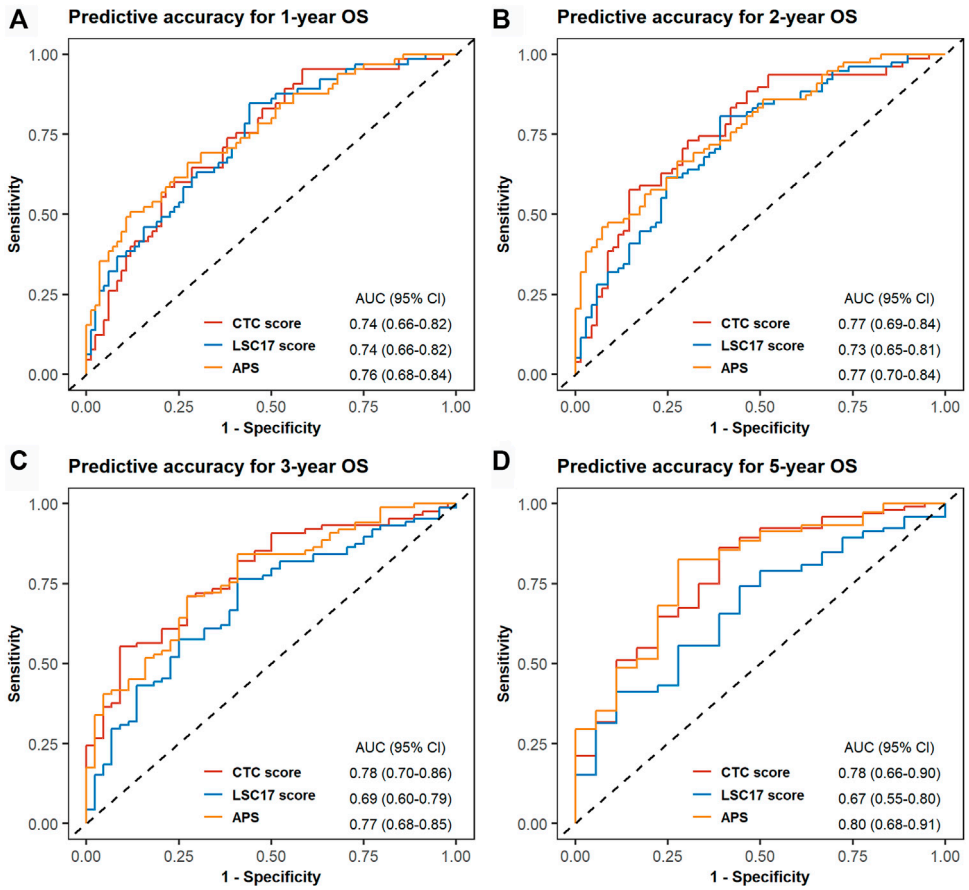


FIGURE 4 | Time-dependent receiver operating characteristic (ROC) curves for cell type composition score, LSC17 score, and acute myeloid leukemia prognostic score of the validation set TCGA-LAML. One-year **(A)**, 2-years **(B)**, 3-years **(C)**, and 5-years **(D)** ROC curves and the corresponding areas under ROC curve with 95% CI are displayed.

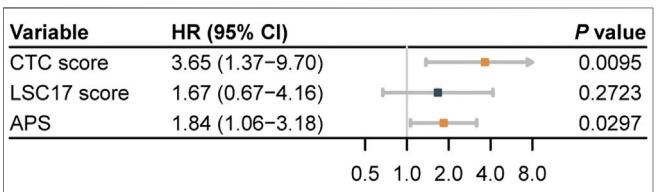


FIGURE 5 | Forest plot for multivariable Cox regression, incorporating three acute myeloid leukemia prognostic scores of the validation set TCGA-LAML.

Except for the LSC17 score and APS, most of the existing studies were based on transcriptomic profiles aiming to construct prognostic scores or find genes associated with the prognosis of AML in adults or pediatric AML were based on transcriptomic profiles (Duployez et al., 2019; Huang et al., 2019; Elsayed et al., 2020; Wang et al., 2020). Some of the genes in these models were inexplicable. Few AML prognostic studies focused on the CTCs of samples from AML patients at diagnosis. In our study, we showed that the AML prognostic model established on the CTCs could independently assess the overall survival of AML patients. The CTC score achieved comparative performance in predicting AML prognosis compared with the gene expression-based prognostic scores. Furthermore, we found that the CTC score could provide additional information different from the LSC17 score and APS. The CTC score clarified that GMP-like was a powerful cell marker predicting the prognosis for AML. Rapid detection of the proportions of GMP-like in the samples from AML patients at diagnosis was expected to aid prognostic classification in the future. Nevertheless, more datasets are required to further verify the effectiveness of the CTC score. Besides this, to incorporate CTC score, APS, and other prognostic factors into a more powerful prognostic model for AML is expected in further studies.

There exist several limitations in the present study. First, the similarity between different cell types inevitably affects the estimation of CIBERSORT. At present, the highly expressed genes of each cell type are typically obtained by comparing 1 cell type against all others. Such a method makes it difficult to distinguish 1 cell type from another similar cell type, especially when the number of one of the cell types is relatively small. To mitigate this influence, we filtered out highly expressed genes with logFC lower than 1 and chose the

most significant for each cell type. Second, the discrepancies of distribution for some cell types (e.g., ProMono-like) between the training set and the validation set, as shown in **Supplementary Figure S6**, might be caused by estimation error, different composition in AML subtypes between datasets, and different transcriptome sequencing approach. This might limit the power to identify the associations of these cell types with AML prognosis. Third, we assumed that samples from bone marrow aspirates and peripheral blood comprised the same cell types. The samples of bulk GEPs datasets GSE6891 and TCGA-LAML were from different tissues, bone marrow aspirates, or peripheral blood, which might cover the prognostic role of some anti-tumor cell types—for example, T cells accounted for a great part in the single-cell dataset (**Supplementary Figure S1**), whereas the estimated proportions of bulk datasets were less (**Supplementary Figure S6**).

In conclusion, our study established a novel AML prognostic score using CTCs for *de novo* AML in adults. CTC score has great potential to assist clinicians to formulate individualized treatment plans, thereby improving the prognosis for AML patients.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgiacc=GSE116256, Gene Expression Ominibus, accession number:GSE116256; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgiacc=GSE6891, Gene Expression Ominibus, accession number: GSE6891; https://portal.gdc.cancer.gov/, Genomic Data Commons Data Portal, TCGA-LAML; https://www.cbioportal.org/study/clinicalDataid=laml_tcga_pub, cBioPortal, Acute Myeloid Leukemia (TCGA, NEJM 2013).

AUTHOR CONTRIBUTIONS

Conception and design: CD, XH. Development of methodology: CD, XH. Acquisition of data. Financial support: CW, XH. Collection and assembly of data: CD, XH. Data analysis and interpretation: CD, MC, XH. Manuscript writing: All authors. Final approval of manuscript: All authors. Accountable for aspects of the work: All authors.

FUNDING

This study was funded by the National Natural Science Foundation of China (Award number: 82003561 and 81973148).

ACKNOWLEDGMENTS

We thank RV (The JAX Cancer Center, Roux Center for Genomics and Computational Biology, Farmington, Connecticut, United States) for providing us with the survival information of AML patients in the GSE6891 dataset.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.762260/full#supplementary-material

REFERENCES

Adan, A., Alizada, G., Kiraz, Y., Baran, Y., and Nalbant, A. (2017). Flow Cytometry: Basic Principles and Applications. *Crit. Rev. Biotechnol.* 37, 163–176. doi:10.3109/07388551.2015.1128876

Appelbaum, F. R., Kopecky, K. J., Tallman, M. S., Slovak, M. L., Gundacker, H. M., Kim, H. T., et al. (2006). The Clinical Spectrum of Adult Acute Myeloid Leukaemia Associated with Core Binding Factor Translocations. *Br. J. Haematol.* 135, 165–173. doi:10.1111/j.1365-2141.2006.06276.x

Aratani, Y. (2018). Myeloperoxidase: Its Role for Host Defense, Inflammation, and Neutrophil Function. *Arch. Biochem. Biophys.* 640, 47–52. doi:10.1016/j.abb.2018.01.004

Avila Cobos, F., Vandesompele, J., Mestdagh, P., and De Preter, K. (2018). Computational Deconvolution of Transcriptomics Data from Mixed Cell Populations. *Bioinformatics* 34, 1969–1979. doi:10.1093/bioinformatics/bty019

Blanche, P., Dartigues, J.-F., and Jacqmin-Gadda, H. (2013). Estimating and Comparing Time-dependent Areas under Receiver Operating Characteristic Curves for Censored Event Times with Competing Risks. *Statist. Med.* 32, 5381–5397. doi:10.1002/sim.5958

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

De Angelis, R., Minicozzi, P., Sant, M., Dal Maso, L., Brewster, D. H., Osca-Gelis, G., et al. (2015). Survival Variations by Country and Age for Lymphoid and Myeloid Malignancies in Europe 2000-2007: Results of EUROCARE-5 Population-Based Study. *Eur. J. Cancer* 51, 2254–2268. doi:10.1016/j.ejca.2015.08.003

Docking, T. R., Parker, J. D. K., Jädersten, M., Duns, G., Chang, L., Jiang, J., et al. (2021). A Clinical Transcriptome Approach to Patient Stratification and Therapy Selection in Acute Myeloid Leukemia. *Nat. Commun.* 12, 2474. doi:10.1038/s41467-021-22625-y

Döhner, H., Estey, E., Grimwade, D., Amadori, S., Appelbaum, F. R., Büchner, T., et al. (2017). Diagnosis and Management of AML in Adults: 2017 ELN Recommendations from an International Expert Panel. *Blood* 129, 424–447. doi:10.1182/blood-2016-08-733196

Donovan, M. K. R., D'Antonio-Chronowska, A., D'Antonio, M., and Frazer, K. A. (2020). Cellular Deconvolution of GTEx Tissues powers Discovery of Disease and Cell-type Associated Regulatory Variants. *Nat. Commun.* 11, 955. doi:10.1038/s41467-020-14561-0

Duployez, N., Marceau-Renaut, A., Villenet, C., Petit, A., Rousseau, A., Ng, S. W. K., et al. (2019). The Stem Cell-Associated Gene Expression Signature Allows Risk Stratification in Pediatric Acute Myeloid Leukemia. *Leukemia* 33, 348–357. doi:10.1038/s41375-018-0227-5

Elsayed, A. H., Rafiee, R., Cao, X., Raimondi, S., Downing, J. R., Ribeiro, R., et al. (2020). A Six-Gene Leukemic Stem Cell Score Identifies High Risk Pediatric Acute Myeloid Leukemia. *Leukemia* 34, 735–745. doi:10.1038/s41375-019-0604-8

Ghazawi, F. M., Le, M., Cyr, J., Netchiporouk, E., Rahme, E., Alakel, A., et al. (2019). Analysis of Acute Myeloid Leukemia Incidence and Geographic Distribution in Canada from 1992 to 2010 Reveals Disease Clusters in Sarnia and Other Industrial US Border Cities in Ontario. *Cancer* 125, 1886–1897. doi:10.1002/cncr.32034

Grimwade, D., Hills, R. K., Moorman, A. V., Walker, H., Chatters, S., Goldstone, A. H., et al. (2010). Refinement of Cytogenetic Classification in Acute Myeloid Leukemia: Determination of Prognostic Significance of Rare Recurring Chromosomal Abnormalities Among 5876 Younger Adult Patients Treated in the United Kingdom Medical Research Council Trials. *Blood* 116, 354–365. doi:10.1182/blood-2009-11-254441

Grisolano, J. L., Wesselschmidt, R. L., Pelicci, P. G., and Ley, T. J. (1997). Altered Myeloid Development and Acute Leukemia in Transgenic Mice Expressing PML-Rarα under Control of Cathepsin G Regulatory Sequences. *Blood* 89, 376–387. doi:10.1182/blood.v89.2.376

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of Cancer: the Next Generation. *Cell* 144, 646–674. doi:10.1016/j.cell.2011.02.013

Harrell, F. E., Jr., Lee, K. L., and Mark, D. B. (1996). Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors. *Statist. Med.* 15, 361–387. doi:10.1002/(sici)1097-0258(19960229)15:4<361:aid-sim168>3.0.co;2-4

Huang, S., Zhang, B., Fan, W., Zhao, Q., Yang, L., Xin, W., et al. (2019). Identification of Prognostic Genes in the Acute Myeloid Leukemia Microenvironment. *Aging* 11, 10557–10580. doi:10.18632/aging.102477

Jia, M., Zhang, H., Wang, L., Zhao, L., Fan, S., and Xi, Y. (2021). Identification of Mast Cells as a Candidate Significant Target of Immunotherapy for Acute Myeloid Leukemia. *Hematology* 26, 284–294. doi:10.1080/16078454.2021.1889158

Cancer Genome Atlas Research NLey, T. J., Miller, C., Ding, L., Raphael, B. J., Mungall, A. J., et al. (2013). Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *N. Engl. J. Med.* 368, 2059–2074. doi:10.1056/NEJMoa1301689

Leys, C., Ley, C., Klein, O., Bernard, P., and Licata, L. (2013). Detecting Outliers: Do Not Use Standard Deviation Around the Mean, Use Absolute Deviation Around the Median. *J. Exp. Soc. Psychol.* 49, 764–766. doi:10.1016/j.jesp.2013.03.013

Marcucci, G., Haferlach, T., and Döhner, H. (2011). Molecular Genetics of Adult Acute Myeloid Leukemia: Prognostic and Therapeutic Implications. *Jco* 29, 475–486. doi:10.1200/jco.2010.30.2554

McDavid, A., Finak, G., Chattopadyay, P. K., Dominguez, M., Lamoreaux, L., Ma, S. S., et al. (2013). Data Exploration, Quality Control and Testing in Single-Cell qPCR-Based Gene Expression Experiments. *Bioinformatics* 29, 461–467. doi:10.1093/bioinformatics/bts714

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust Enumeration of Cell Subsets from Tissue Expression Profiles. *Nat. Methods* 12, 453–457. doi:10.1038/nmeth.3337

Ng, S. W. K., Mitchell, A., Kennedy, J. A., Chen, W. C., McLeod, J., Ibrahimova, N., et al. (2016). A 17-gene Stemness Score for Rapid Determination of Risk in Acute Leukaemia. *Nature* 540, 433–437. doi:10.1038/nature20598

Papaemmanuil, E., Gerstung, M., Bullinger, L., Gaidzik, V. I., Paschka, P., Roberts, N. D., et al. (2016). Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N. Engl. J. Med.* 374, 2209–2221. doi:10.1056/NEJMoa1516192

Potter, S. S. (2018). Single-cell RNA Sequencing for the Study of Development, Physiology and Disease. *Nat. Rev. Nephrol.* 14, 479–492. doi:10.1038/s41581-018-0021-7

Roman, E., Smith, A., Appleton, S., Crouch, S., Kelly, R., Kinsey, S., et al. (2016). Myeloid Malignancies in the Real-World: Occurrence, Progression and Survival in the UK's Population-Based Haematological Malignancy Research Network 2004-15. *Cancer Epidemiol.* 42, 186–198. doi:10.1016/j.canep.2016.03.011

Schlaifer, D., Cooper, M., Attal, M., Sartor, A., Trepel, J., Laurent, G., et al. (1993). Myeloperoxidase: an Enzyme Involved in Intrinsic Vincristine Resistance in Human Myeloblastic Leukemia. *Blood* 81, 482–489. doi:10.1182/blood.V81.2.482.48210.1182/blood.v81.2.482.bloodjournal812482

Shallis, R. M., Wang, R., Davidoff, A., Ma, X., and Zeidan, A. M. (2019). Epidemiology of Acute Myeloid Leukemia: Recent Progress and Enduring Challenges. *Blood Rev.* 36, 70–87. doi:10.1016/j.blre.2019.04.005

Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2011). Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J. Stat. Soft.* 39, 1–13. doi:10.18637/jss.v039.i05

Slovak, M. L., Kopecky, K. J., Cassileth, P. A., Harrington, D. H., Theil, K. S., Mohamed, A., et al. (2000). Karyotypic Analysis Predicts Outcome of Preremission and Postremission Therapy in Adult Acute Myeloid Leukemia: a Southwest Oncology Group/Eastern Cooperative Oncology Group Study. *Blood* 96, 4075–4083. doi:10.1182/blood.v96.13.4075.h8004075_4075_4083

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., et al. (2019). Comprehensive Integration of Single-Cell Data. *Cell* 177, 1888–1902. e1821. doi:10.1016/j.cell.2019.05.031

Sun, X., Sun, S., and Yang, S. (2019). An Efficient and Flexible Method for Deconvoluting Bulk RNA-Seq Data with Single-Cell RNA-Seq Data. *Cells* 8, 1161. doi:10.3390/cells8101161

van Galen, P., Hovestadt, V., Wadsworth Ii, M. H., Hughes, T. K., Griffin, G. K., Battaglia, S., et al. (2019). Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell* 176, 1265–1281. e1224. doi:10.1016/j.cell.2019.01.031

Verhaak, R. G. W., Wouters, B. J., Erpelinck, C. A. J., Abbas, S., Beverloo, H. B., Lugthart, S., et al. (2009). Prediction of Molecular Subtypes in Acute Myeloid Leukemia Based on Gene Expression Profiling. *Haematologica* 94, 131–134. doi:10.3324/haematol.13299

Wang, X., Park, J., Susztak, K., Zhang, N. R., and Li, M. (2019). Bulk Tissue Cell Type Deconvolution with Multi-Subject Single-Cell Expression Reference. *Nat. Commun.* 10, 380. doi:10.1038/s41467-018-08023-x

Wang, Y., Hu, F., Li, J.-y., Nie, R.-c., Chen, S.-l., Cai, Y.-y., et al. (2020). Systematic Construction and Validation of a Metabolic Risk Model for Prognostic Prediction in Acute Myelogenous Leukemia. *Front. Oncol.* 10, 540. doi:10.3389/fonc.2020.00540

Wang, Z.-Y., and Chen, Z. (2008). Acute Promyelocytic Leukemia: from Highly Fatal to Highly Curable. *Blood* 111, 2505–2515. doi:10.1182/blood-2007-07-102798

Wollenberg, A., Oppel, T., Schottdorf, E.-M., Günther, S., Moderer, M., and Mommaas, M. (2002). Expression and Function of the Mannose Receptor CD206 on Epidermal Dendritic Cells in Inflammatory Skin Diseases. *J. Invest. Dermatol.* 118, 327–334. doi:10.1046/j.0022-202x.2001.01665.x

Wu, Z., Irizarry, R. A., Gentleman, R., Martinez-Murillo, F., and Spencer, F. (2004). A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. *J. Am. Stat. Assoc.* 99, 909–917. doi:10.1198/016214504000000683

Xu, Z.-J., Gu, Y., Wang, C.-Z., Jin, Y., Wen, X.-M., Ma, J.-C., et al. (2020). The M2 Macrophage Marker CD206: a Novel Prognostic Indicator for Acute Myeloid Leukemia. *Oncoimmunology* 9, 1683347. doi:10.1080/2162402X.2019.1683347

Yamashita, M., Dellorusso, P. V., Olson, O. C., and Passegué, E. (2020). Dysregulated Haematopoietic Stem Cell Behaviour in Myeloid Leukaemogenesis. *Nat. Rev. Cancer* 20, 365–382. doi:10.1038/s41568-020-0260-3

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: 16 November 2021
doi: 10.3389/fcell.2021.780461

Comprehensive Analysis Uncovers Prognostic and Immunogenic Characteristics of Cellular Senescence for Lung Adenocarcinoma

Weihao Lin[1†], *Xin Wang*[1†], *Zhen Wang*[1], *Fei Shao*[1], *Yannan Yang*[1], *Zheng Cao*[2], *Xiaoli Feng*[2], *Yibo Gao*[1,3*] and *Jie He*[1,3*]

[1]Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [2]Department of Pathology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [3]State Key Laboratory of Molecular Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

OPEN ACCESS

Edited by:
Geng Chen,
GeneCast Biotechnology Co., Ltd., China

Reviewed by:
Beatriz Martín-Antonio,
University Hospital Fundacion Jimenez Diaz, Spain
Nicola Alessio,
Universita della Campania Luigi Vanvitelli, Italy

***Correspondence:**
Yibo Gao
gaoyibo@cicams.ac.cn
Jie He
hejie@cicams.ac.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: 21 September 2021
Accepted: 27 October 2021
Published: 16 November 2021

Citation:
Lin W, Wang X, Wang Z, Shao F, Yang Y, Cao Z, Feng X, Gao Y and He J (2021) Comprehensive Analysis Uncovers Prognostic and Immunogenic Characteristics of Cellular Senescence for Lung Adenocarcinoma. Front. Cell Dev. Biol. 9:780461. doi: 10.3389/fcell.2021.780461

Cellular senescence plays a crucial role in tumorigenesis, development and immune modulation in cancers. However, to date, a robust and reliable cellular senescence-related signature and its value in clinical outcomes and immunotherapy response remain unexplored in lung adenocarcinoma (LUAD) patients. Through exploring the expression profiles of 278 cellular senescence-related genes in 936 LUAD patients, a cellular senescence-related signature (SRS) was constructed and validated as an independent prognostic predictor for LUAD patients. Notably, patients with high SRS scores exhibited upregulation of senescence-associated secretory phenotype (SASP) and an immunosuppressive phenotype. Further analysis showed that SRS combined with immune checkpoint expression or TMB served as a good predictor for patients' clinical outcomes, and patients with low SRS scores might benefit from immunotherapy. Collectively, our findings demonstrated that SRS involved in the regulation of the tumor immune microenvironment through SASP was a robust biomarker for the immunotherapeutic response and prognosis in LUAD.

Keywords: lung adenocarcinoma, cellular senescence, senescence-associated secretory phenotype, tumor microenvironment, prognosis, immunotherapy

INTRODUCTION

Lung cancer has the highest incidence and mortality of cancer worldwide (Sung, et al., 2021). The 5-years survival rate is less than 20% (Miller, et al., 2019). Lung adenocarcinoma (LUAD) is the main histological subtype of non-small-cell lung cancer (NSCLC), accounting for approximately 60% of NSCLC cases (Behera, et al., 2016). Although understanding LAUD genomics and breakthroughs of targeted therapies and immunotherapies have substantially expanded treatment modalities, challenges associated with LAUD remain elusive. Therefore, better prognostic tools and biomarkers accurately predicting the characteristics of tumors are urgently needed to stratify patients and personalize treatment strategies for LUAD.

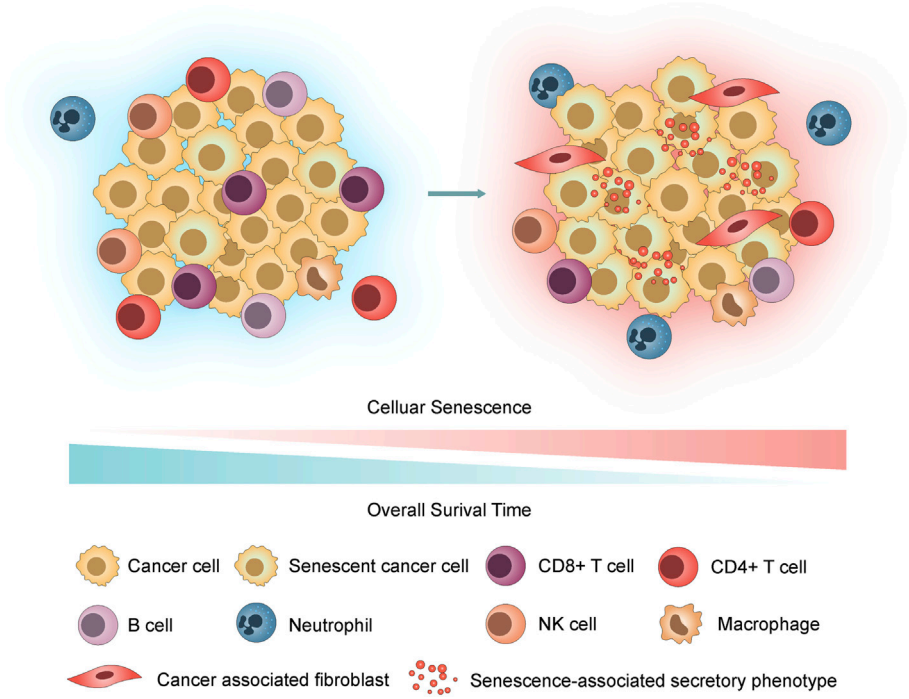


FIGURE 1 | Hypothesis graph. With the accumulation of senescent cells, the inflammatory SASP remodels the TIME by recruiting immunosuppressive cells, thus promoting tumor cell invasion through immunosurveillance, proliferation, and metastasis and contributing to poor prognosis in LUAD.

Cellular senescence is one of the key processes of ageing (Campisi 2013) and serves as a link between ageing and cancer (Partridge, et al., 2018). However, the linkage of senescence and cancer, especially in lung cancer, is complex and poorly understood at present. Previous studies have highlighted that the existence of senescence plays a double-edged sword in the process of tumorigenesis and development. On the one hand, in the context of senescent cells entering permanent cell cycle arrest, senescence ensures tissue homeostasis and prevents tumorigenesis (Krizhanovsky, et al., 2008; Perez-Mancera, et al., 2014). Senescence acts as a barrier from tumor development in early tumorigenesis when it is followed by immune clearance and tissue remodelling (Xue, et al., 2007). On the other hand, cellular senescence can present a detrimental outcome when senescent cells are not cleared by the immune system and accumulate. This accumulation promotes the senescence-associated secretory phenotype (SASP), which releases cytokines, growth factors, extracellular matrix (ECM) components and ECM-degrading enzymes (Lasry and Ben-Neriah 2015; Lopes-Paciencia, et al., 2019), leading to both the ageing process and tumor development (Coppe, et al., 2010; Cuollo, et al., 2020). Therefore, an improved understanding of the impact of senescence on tumor immunity associated with invasion and development is required to frame novel treatment paradigms for tumors.

According to recent studies, tumor cells can undergo senescence as an evolutionary process, including both tumor-intrinsic characteristics (dramatic gene expression changes along with chromatin remodelling and engagement of a persistent DDR) and extrinsic immune pressure (a temporal cascade in the development of SASP) (Lasry and Ben-Neriah 2015; Berben, et al., 2021; Eggert, et al., 2016; Hernandez-Segura, et al., 2018; Kumari and Jat 2021). Notably, the deleterious effects of SASP overshadow its beneficial properties (Cuollo, et al., 2020). We hypothesized that accompanied by the accumulation of senescent cells, inflammatory SASP remodels the tumor immune microenvironment (TIME) by recruitment of immunosuppressive protumorigenic cells, such as cancer-associated fibroblasts (CAFs), macrophages and neutrophils, and a decrease in cytotoxic lymphocytes (T and NK cells) and promotes tumor cell evasion from immunosurveillance, growth, and metastasis, contributing to poor prognosis in LUAD (**Figure 1**).

To systematically assess the correlations between cellular senescence and prognosis in LUAD, we established a novel risk model based on cellular senescence-related genes and explored their potential importance as predictive biomarkers for prognosis and immunotherapy response. Subsequently, the relationships among risk subgroups, immune checkpoints, and immune cell infiltration were thoroughly analysed based on cell senescence-related signature. Further exploration of the mechanisms suggested that tumor cellular senescence affected the TIME through SASP. This study provided new insights into the regulatory mechanisms of cellular senescence associated with the TIME and strategies for LUAD immunotherapy.

MATERIALS AND METHODS

Data Acquisition and Processing

Clinical information and transcriptional profiles of patients with LUAD were obtained from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov) and the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo). After filtering, a total of 500 patients with both mRNA expression and corresponding clinical data in the TCGA cohort were included in the training cohort. Fragments per kilobase million (FPKM) data of the TCGA cohort were then transformed into transcripts per million (TPM) data for further analysis. Three additional independent datasets, GSE30219 (Rousseaux, et al., 2013) (n = 83), GSE31210 (Okayama, et al., 2012) (n = 226) and GSE50081 (Der, et al., 2014) (n = 127), were enrolled as the validation cohorts. For microarray data processing, the mean expression values were used when genes matched with multiple probes. Moreover, IMvigor210 (Mariathasan, et al., 2018), an immunotherapy cohort with 348 metastatic urothelial cancer patients treated with anti-PD-L1 agent, was downloaded from http://research-pub.gene.com/IMvigor210CoreBiologies/, and data processing methods were also provided in the IMvigor210CoreBiologies package. Detailed clinical information of the five patient datasets is shown in **Supplementary Table S1**. The flow diagram of this study is depicted in **Supplementary Figure S1**.

Development and Validation of the Cellular Senescence-Related Signature

The list of genes was obtained from CellAge (Avelar, et al., 2020) (https://genomics.senescence.info/cells/), which contains manually curated data of human genes associated with cellular senescence. A total of 278 genes (**Supplementary Table S2**) were included in this study. We first screened cellular senescence-related differentially expressed genes (DEGs) between normal samples (n = 59) and tumor samples (n = 513) based on the thresholds of an adjusted $p < 0.01$ and $|\log2(\text{fold change})| > 1$. Univariate Cox proportional hazard regression analysis was performed to identify cellular senescence-related prognostic genes ($p < 0.001$). Next, the DEGs and prognostic genes were investigated using the R package "veen" to acquire prognostic cellular senescence-related DEGs, and correlations were visualized by the R package "circlize" (Gu, et al., 2014). Least absolute shrinkage and selection operator (LASSO) Cox regression (Tibshirani 1997) was conducted with a random seed using the R package "glmnet" (Friedman, et al., 2010) to construct the risk score model (cellular senescence-related signature, SRS) for best predicting survival in the training cohort and was repeated 1,000 times. The optimal values of the penalty parameter lambda were determined through 10-fold cross-validations. Based on the median risk score calculated by SRS, patients in the training and validation cohorts were divided into high- and low-risk groups, and the performance of SRS was subsequently evaluated.

Signature Genes Analyses

Expression of the five signature genes was analysed in The Gene Expression Profiling Interactive Analysis (GEPIA2 (Tang, et al., 2019), http://gepia2.cancer-pku.cn/) database, and these analyses were based on tumor and normal samples from the TCGA and GTEx databases. Pancancer expression analysis of the five genes was also performed using the Oncomine (https://www.oncomine.org/) database. UALCAN (Chandrashekar, et al., 2017) (http://ualcan.path.uab.edu), another powerful interactive online tool, was used to reveal the promoter methylation levels of signature genes.

Pathway and Functional Enrichment Analysis

Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa, et al., 2016) and Gene Ontology (GO) (The Gene Ontology 2019) enrichment analyses were applied using the R package clusterProfiler (Yu, et al., 2012). The DEGs between the high- and low-risk groups were subjected to pathway and functional enrichment analysis. Gene set enrichment analysis (GSEA) (Subramanian, et al., 2005) was also performed in the javaGSEA desktop application (GSEA 4.1.0) to identify the underlying pathways or processes in patients with high or low scores. Significantly enriched gene sets were defined as gene sets with a normalized enrichment score (NES) > 1.5 and $p < 0.05$.

Correlation Between Immune Cell Infiltration and SRS

We integrated several computational tools to estimate immune cell infiltration in TCGA RNA-seq cohorts. Immune infiltration estimations for characterizing the cell composition of complex tissues from the gene expression profiles were performed using TIMER (Li, et al., 2016), EPIC(Racle, et al., 2017), xCELL (Aran, et al., 2017), CIBERSORT(Newman, et al., 2015) and quanTIseq (Finotello, et al., 2019) algorithms in TIMER2.0 (Li, et al., 2016; Li, et al., 2017; Li, et al., 2020; Sturm, et al., 2019) (http://timer.comp-genomics.org/). Using the gsva (Hanzelmann, et al., 2013) algorithm, gene signatures of tumor-infiltrating lymphocytes downloaded in TISIDB(Ru, et al., 2019) (http://cis.hku.hk/TISIDB/index.php) were also employed to estimate the immune cell infiltration level of each sample. Pearson correlation analysis was conducted to clarify the correlation between SRS and immune cell infiltration.

Assessment of SRS and Response to Immune Checkpoint Inhibitors

The immunophenoscore (IPS), which has been demonstrated to predict patients' response to immune checkpoint inhibitor (ICI) treatment, was downloaded from The Cancer Immunome Atlas (TCIA (Charoentong, et al., 2017), https://tcia.at). A higher IPS score indicates a better immunotherapy response. Tumor Immune Dysfunction and Exclusion (TIDE (Fu, et al., 2020; Jiang, et al., 2018), http://tide.dfci.harvard.edu/), which was

developed to assess immune evasion mechanisms, is another robust biomarker used to predict immunotherapy response. A higher TIDE score means that tumor cells are more prone to escape from immunosurveillance, suggesting a lower response rate to immunotherapy. TIDE scores were obtained after unloading the input data as described in the instructions. The TMB for each patient in the TCGA cohort was calculated as the number of nonsynonymous mutations per mega-base. PD-L1 expression on tumor-infiltrating immune cells (ICs) of patients in the IMvigor210 cohort was assessed by immunohistochemistry. IC0 and IC1 exhibit low PD-L1 expression, while IC2 indicates high PD-L1 expression in our study.

Clinical Specimens

We retrospectively collected 74 paraffin-embedded LUAD specimens and 74 adjacent normal tissues from the biobank of National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital in Chinese Academy of Medical Sciences and Peking Union Medical College (Beijing, China) and constructed tissue microarray (TMA). All the biospecimens were obtained from LAUD patients who underwent radical resection and had received no prior chemotherapy or radiotherapy. Informed consent was obtained from all patients. This study was approved by the Ethics and Research Committees of the National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences, and Peking Union Medical College.

Immunohistochemistry

TMA slides were incubated at 4°C overnight with primary antibodies against FOXM1 (Proteintech, 13147-1-AP, 1: 200), HJURP (Proteintech, 15283-1-AP, 1: 200), and PTTG1 (ABclonal, A8307, 1: 500) after deparaffinization, rehydration and antigen retrieval. Next, the slides were incubated with anti-rabbit secondary antibody and followed by DAB staining and hematoxylin counterstaining. Two pathologists who were blind to the information of patients independently evaluated the IHC results. The TMA sections were scored according to the percentage of positive cells and staining intensity. Staining intensity was scored as 0 (negative), 1 (weak), 2 (moderate) or 3 (strong) and the expression proportion of positive cells was scored as 1 (0–25%), 2 (26–50%), 3 (51–75%) or 4 (76–100%). The proportion and intensity scores were then integrated to obtain a final score.

Statistical Analysis

Data analysis and graph generation were all performed in R version 3.5.1 (https://www.r-project.org), SPSS Statistics V25.0 and GraphPad Prism 8.0. For comparisons of two groups, unpaired Student's t-test was applied to analyse the statistical significance of normally distributed variables, and the Wilcoxon rank-sum test was adopted to estimate nonnormally distributed variables. Categorical variables were compared using the χ^2 test. The Kaplan-Meier survival curve for overall survival (OS) analysis was plotted with the R package "survminer". Receiver operating characteristic (ROC) curves for 1-, 3-, and 5-years survival were delineated to evaluate the predictive efficacy of the SRS score, which was generated using timeROC(Blanche, et al., 2013). Univariate and multivariate Cox regression analyses were utilized to evaluate the association between OS and clinicopathological characteristics as well as SRS scores. All statistical analyses were two-tailed, and $p < 0.05$ was considered statistically significant.

RESULTS

Identification of Differentially Expressed Senescence-Related Genes in LUAD

To comprehensively characterize the expression pattern of cellular senescence-related genes, the 278 genes downloaded in CellAge (Avelar, et al., 2020) were compared in tumor tissues versus normal tissues in the TCGA-LUAD cohort, and we identified 69 differentially expressed genes (DEGs). Among them, 42 genes were upregulated, whereas 27 were downregulated (**Figure 2A**; **Supplementary Table S3**). GO and KEGG analyses were performed to clarify the biological process of the DEGs. As expected, the DEGs were remarkably enriched in cell cycle- and cellular senescence-related pathways, as they were obtained from a website of genes related to cellular senescence (**Supplementary Figure S2**; **Supplementary Table S4**).

Next, univariate Cox proportional hazard regression analysis was initially performed to identify cellular senescence-related genes associated with overall survival (OS). A total of 18 genes were significantly associated with OS ($p < 0.001$, **Supplementary Table S5**), ten of which overlapped with the DEGs (**Figure 2B**). All ten genes (AURKA, CDK1, CENPA, CHEK1, FOXM1, HJURP, MAD2L1, PKM, PTTG1 and TACC3) were upregulated in LUAD and considered risk factors ($p < 0.001$, HR > 1) (**Figure 2C**). Besides, 10 prognostic DEGs were positively correlated with each other (**Figure 2D**).

Development of a Cellular Senescence-Related Signature in LUAD

To construct a cellular senescence-related signature (SRS) for survival prediction, the 10 genes mentioned above were analysed by LASSO-Cox regression analysis. A 5-gene signature was constructed according to the optimum λ value (**Figures 2E,F**). We then established a risk score formula based on the expression of the five genes for patients with LUAD: risk score = (0.0089 × expression value of FOXM1) + (0.1233 × expression value of HJURP) + (0.3092 × expression value of PKM) + (0.0851 × expression value of PTTG1) + (0.0003 × expression value of TACC3). The risk score of every patient was then calculated using this formula, and patients in the TCGA cohort were stratified into low- and high-risk groups according to the median value of the risk score.

The distribution of the SRS score, the survival status, and a heatmap exhibiting the expression profiles of the selected genes in the high- and low-risk groups are presented in **Figures 3A–C**. Kaplan-Meier survival analysis demonstrated that patients in the high-risk group had a significantly shorter OS time compared with those in the low-risk group (**Figure 3D**, HR = 2.048, 95% CI 1.529–2.743, log-rank $p < 0.0001$). The 5-years survival rate of the

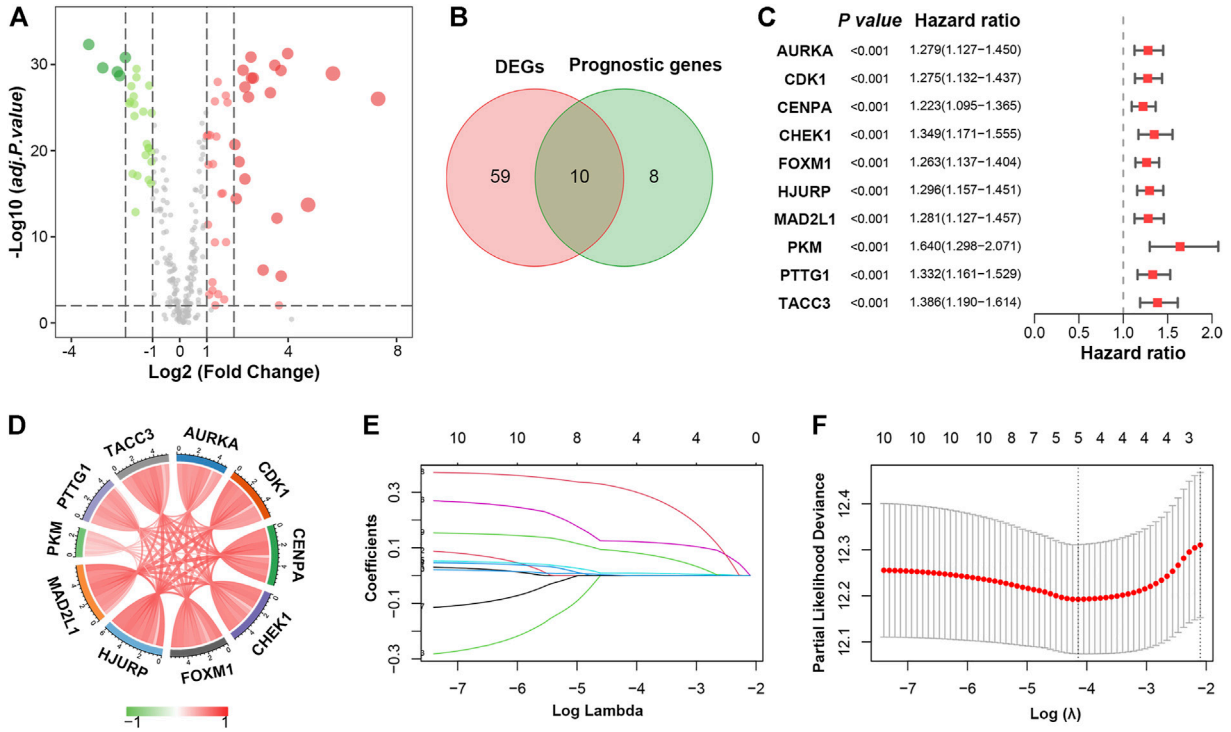


FIGURE 2 | Identification of prognostic cellular senescence-related DEGs. **(A)**. Volcano plot of cellular senescence-related genes in TCGA database; red and green nodes indicate upregulated and downregulated genes, respectively. **(B)**. Venn diagram showing 10 overlapping genes in DEGs and prognostic genes. **(C)**. Forrest plot of the univariate Cox analysis of 10 overlapping genes. **(D)**. Correlation network of 10 candidate genes. **(E)**. LASSO coefficient profiles of 10 candidate genes. **(F)**. Cross-validation for tuning parameter selection in the LASSO regression.

high-risk group was 30.2%, which was significantly lower than that of the low-risk group (49.7%). Time-dependent receiver operating characteristic (ROC) analysis was performed, and the areas under the curve (AUCs) for 2-, 3-, and 5-years OS were 0.675, 0.660, and 0.607, respectively (**Figure 3E**). In addition, our formula also worked well when applied to patients with different clinical stages. As shown in **Figures 3F,G** significant difference in OS time was observed in both early-stage (HR = 1.955, 95% CI 1.357–2.816, log-rank p = 0.0002) and advanced-stage LUAD (HR = 1.725, 95% CI 1.034–2.879, log-rank p = 0.0478).

To further verify whether the SRS-based risk score was an independent prognostic factor for LUAD, univariate and multivariate Cox regression analyses of clinicopathological factors in the TCGA cohort were performed. The T stage, N stage, TNM stage and risk score were correlated with OS in univariable analysis. After multivariable adjustment, the risk score remained a significantly independent prognostic factor (HR = 2.746, 95% CI: 1.738–4.339, $p < 0.001$) for patients with LUAD (**Figure 4A**). We also analysed the correlation between SRS and patients' clinicopathological parameters, including age, sex, T stage, N stage and TNM stage, in the TCGA cohort. Significantly higher percentages of patients with lymphatic metastasis and late-stage LUAD were identified in the high-risk group (**Supplementary Figure S3**), indicating that a higher SRS score was related to the malignant progression of LUAD.

Validation of SRS in Three Independent Cohorts

To validate the predictive function of SRS on OS benefit, three independent data sets from the GEO database were enrolled. As illustrated in **Figures 4B–D**, patients with high-risk scores exhibited significantly worse OS in all three cohorts, including GSE30219 (HR 2.163, 95% CI 1.188–3.938, p = 0.0118), GSE31210 (HR 6.699, 95% CI 3.450–13.01, $p < 0.0001$) and GSE50081 (HR 2.842, 95% CI 1.628–4.962, p = 0.0002). The area under the ROC curve (AUC) values in the GSE30219 cohort were 0.687, 0.722, and 0.732 for 2, 3 and 5 years, respectively (**Figure 4E**). In the GSE31210 cohort, all AUC values were greater than 0.7 (**Figure 4F**). For the GSE50081 cohort, the AUCs of SRS at 2, 3, and 5 years were 0.690, 0.681, and 0.717, respectively (**Figure 4G**). Moreover, we also observed that high expression of the five genes in four different cohorts was consistently indicative of poor prognosis for LUAD patients (**Supplementary Figure S4**). These results confirmed that SRS could serve as a good predictive factor to classify patients with different OS.

Biological Processes Analysis of SRS

Multicohort evaluation confirmed a robust prognostic value of SRS, which prompted us to further explore the possible mechanism underlying the predictive role of the signature. As shown in **Supplementary Figures S5A,B**, all five genes were abnormally

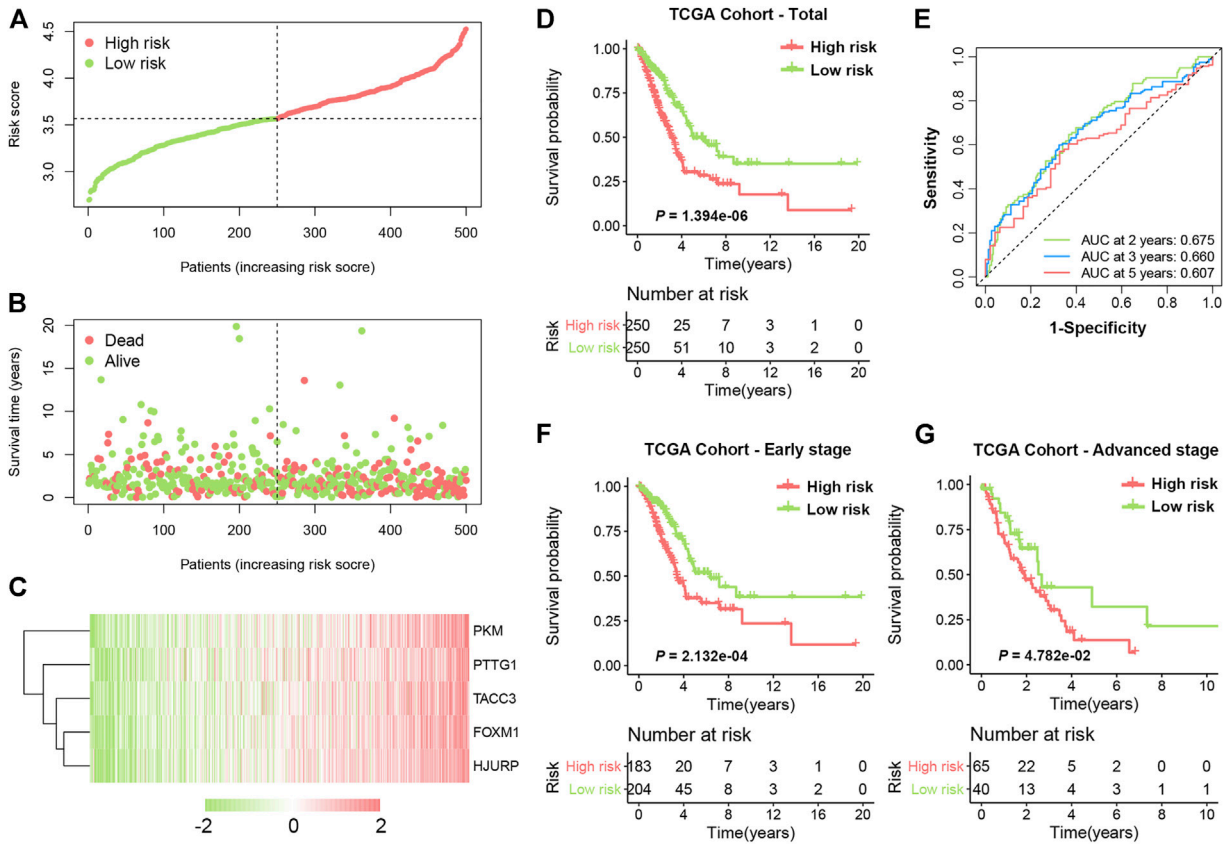


FIGURE 3 | Development of SRS in TCGA cohort. **(A)**. The risk score distribution of LUAD patients in the training cohort. **(B)**. Survival status scatter plots for LUAD patients in the training cohort. **(C)**. Heatmap of the expression profiles of five signature genes in the high- and low-risk groups. **(D)**. Kaplan-Meier curves of OS in total LUAD patients of TCGA cohort based on risk score. **(E)**. Time-dependent ROC curve analysis of the prognostic model (2, 3, and 5 years). **(F)** and **(G)**. Kaplan-Meier curves of OS in patients with early-stage and advanced-stage LUAD based on risk score.

upregulated in LUAD and many other types of cancer, including colorectal cancer, liver cancer, and brain cancer. We then analysed the relationship between methylation and the expression of the five genes. Significantly lower methylation levels of PTTG1 and TACC3 promoters were found in tumor tissues compared with normal tissues, which may account for the abnormal expression of the signature genes in LUAD (**Supplementary Figure S5C**).

Regarding the downstream effects, we first extracted the DEGs between subgroups categorized by the risk signature by applying the criteria FDR <0.05 and $|\log2FC| \geq 1$. In total, 1,164 genes were differentially expressed between the two groups (**Supplementary Figure S6A**; **Supplementary Table S6**). Based on these SRS-related DEGs, GO analysis and KEGG analysis were performed. As expected, the results indicated that DEGs were involved in cellular senescence and cell cycle-related biological processes, such as nuclear division and chromosome segregation (**Supplementary Figures S6B,C**; **Supplementary Table S7**). In addition, GSEA revealed prominent enrichment in hallmark gene sets, such as MTORC1 signalling, glycolysis, and the unfolded protein response, in the high-risk group compared to the low-risk group in TCGA cohort. Similar trends were also observed in the three validation cohorts (**Supplementary Figure S6D**). These results suggested a more malignant phenotype in patients with high-risk scores, which may lead to a poor prognosis in LUAD patients.

SRS Is Associated With Alterations in SASP and Immune Cell Infiltration

Cellular senescence occurs when cells are confronted by excessive extracellular or intracellular stress *in vivo* or *in vitro* (Ben-Porath and Weinberg 2004; Ben-Porath and Weinberg 2005). As displayed in **Figure 5A**, cellular senescence-associated pathways, including oncogene-induced, DNA damage telomere stress-induced, and oxidative stress-induced senescence, were significantly enriched in patients with high SRS scores. Intriguingly, we noticed that the senescence-associated secretory phenotype (SASP) pathway was also prominently enriched. SASP indicates an enormous number of secretory proteins secreted by senescent cells, which may induce changes in the tumor microenvironment, thus promoting tumor recurrence and progression (Acosta, et al., 2008; Coppe, et al., 2008; Green 2008; Kuilman, et al., 2008). Our results revealed overexpression of different types of SASP in the high-risk group (**Figure 5B**). Interleukins (IL-1A, IL-1B, IL-6, and IL-15), chemokines (CCL3, CCL8, CCL11, CCL20, CCL26, CXCL1, CXCL5, CXCL8, and CXCL11), growth factors and regulators

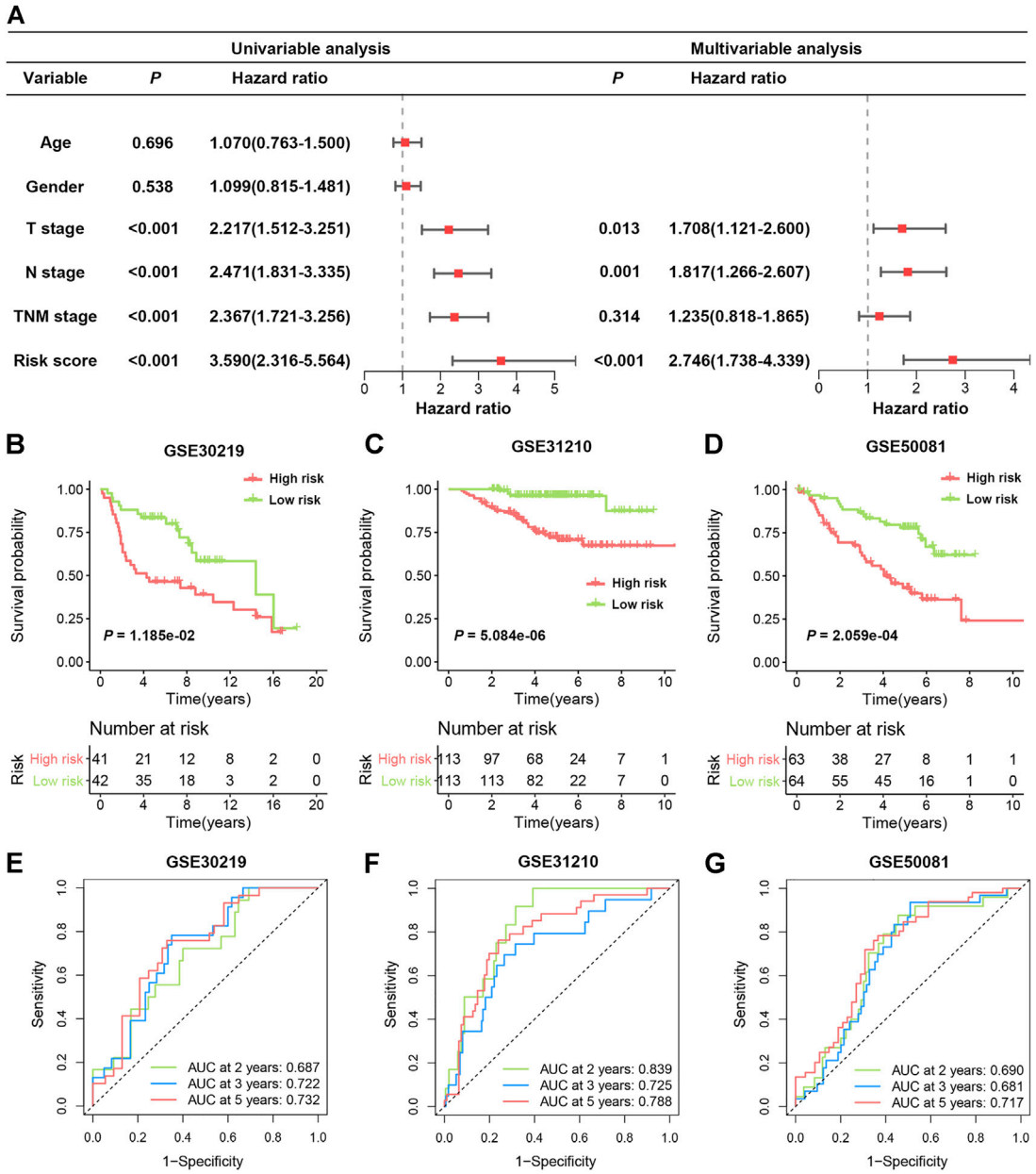


FIGURE 4 | Validation of SRS in independent cohorts. **(A)**. Univariable and multivariable Cox regression analysis of SRS and overall survival in the TCGA cohort. **(B–D)**: Kaplan-Meier curves of OS in high- and low-risk groups in the GSE30219, GSE31210, and GSE50081 datasets. **(E–G)**: ROC curve analyses in the GSE30219, GSE31210, and GSE50081 datasets.

(AREG, EREG, IGFBP3, PIGF, and VEGFA), proteases and regulators (CTSB, MMP1, MMP3, MMP10, MMP12, MMP14, PLAU, SERPINE1, and TIMP2), and soluble or shed receptors or ligands (PLAUR, TNFRSF1A, and TNFRSF11B) were significantly upregulated, further confirming higher levels of the SASP in high-risk patients.

Notably, some upregulated SASPs, including IL-6, CXCL8, and VEGF, possess immunosuppressive properties (Kato, et al., 2018; Lamano, et al., 2019; Liu, et al., 2013; Sharma, et al., 2017). For example, IL-6, secreted by CAFs, regulates immunosuppressive TIL populations in the TIME (Kato, et al., 2018). Thus, we hypothesized that patients with high SRS scores who had higher SASP levels might exhibit an immunosuppressive phenotype via SASP. To characterize the SRS-related immune landscape, RNA-seq-derived infiltrating immune cell populations were estimated by TIMER, EPIC, xCell, CIBERSORT-ABS, and quanTIseq algorithms in TIMER2.0 and TISIDB. We found that patient risk groups stratified by SRS showed distinct immune infiltrate patterns. Correlation analysis showed that the infiltration levels of B cells, $CD4^+$ T cells, and $CD8^+$ T cells were negatively correlated with the SRS score, whereas a higher SRS score indicated increased abundance of neutrophils, cancer-associated fibroblasts, Tregs, and

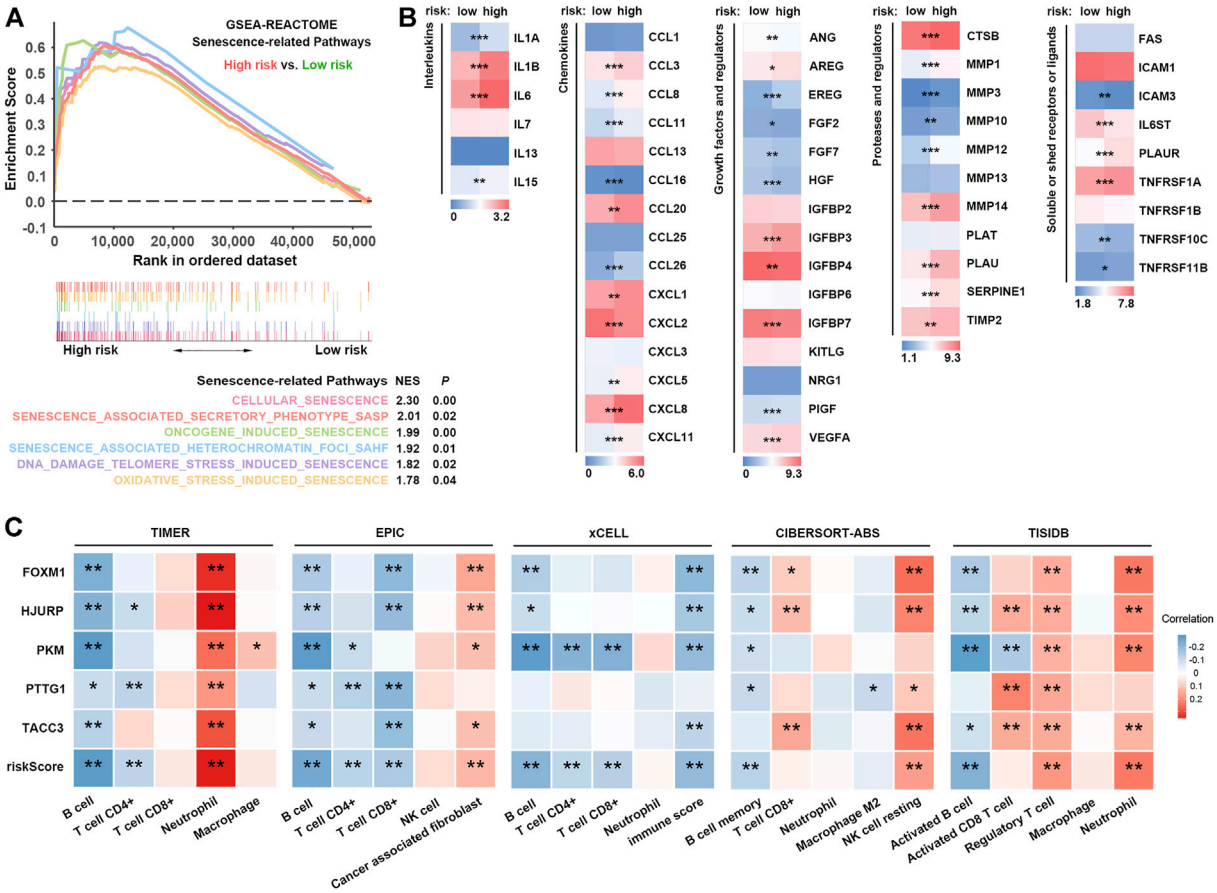


FIGURE 5 | SRS-related SASP and immune cell infiltration. **(A)**. Comparison of signatures related to multiple pathways of cellular senescence between high- and low-risk groups. **(B)**. Expression of different types of SASP factors between high- and low-risk groups. **(C)**. Correlation analysis between risk scores and different immune cells estimated by TIMER, EPIC, xCELL, CIBERSORT-ABS and TISIDB. *, **, and *** represent $P < 0.05$, $P < 0.01$, and $P < 0.001$, respectively.

resting NK cells (**Figure 5C**; **Supplementary Figure S7A**). GSEA revealed a significant enrichment of signatures associated with upregulation of TGF-β signalling, whereas no significant difference in IFN-γ signalling was observed in the high SRS score versus the low SRS score group (**Supplementary Figure S7B**). Taken together, our results implied that a high level of cellular senescence may remodel a suppressive TIME via SASP.

Impact of SRS and Immune Checkpoints on Clinical Outcome

Previous studies have emphasized the importance of immune checkpoint genes in modulating immune infiltration (Juneja, et al., 2017; Kumagai, et al., 2020), and our results also revealed significant relevance between cellular senescence and tumor immunity. Thus, to further investigate the complex crosstalk that occurs among immune infiltration, immune checkpoint genes and SRS, we first compared the expression pattern of immune checkpoint genes between patient groups divided based on the SRS. As shown in **Figures 6A–D**, patients with high SRS scores tended to express high levels immune checkpoint genes (PD-L1, PD1 and CTLA4) compared with the low SRS group, which was further confirmed in 3 validation cohorts. Other immune checkpoints, such as LAG3 and TIM3, which are also considered exhausted T cell markers, exhibited a trend of overexpression in the high SRS score group in the multicohort, suggesting that SRS has the potential to identify immune dysfunction in LUAD patients (**Supplementary Figure S8A**).

Next, we considered SRS in combination with immune checkpoint expression to assess whether SRS influences OS in patients with similar immune checkpoint expression. Survival analysis of the four groups stratified by SRS and immune checkpoint gene expression was conducted. As depicted in **Figures 6E–G**, patients with low PD-L1 and low risk had prolonged OS compared to those with low PD-L1 and high risk ($p = 0.0064$). Among patients with high PD-L1 expression, a lower risk score signified a remarkably better survival ($p = 0.0005$). Similar survival patterns were also observed among the four patient groups stratified by SRS and PD1 or CTLA4 expression in TCGA cohort. Besides, among various immune checkpoint genes, multivariate Cox regression modelling showed that SRS score remained an independently predictor for overall survival (HR = 2.083, 95% CI: 1.533–2.830, $p < 0.001$). We then repeated the same analysis in three validation cohorts. Consistent with TCGA dataset, patients

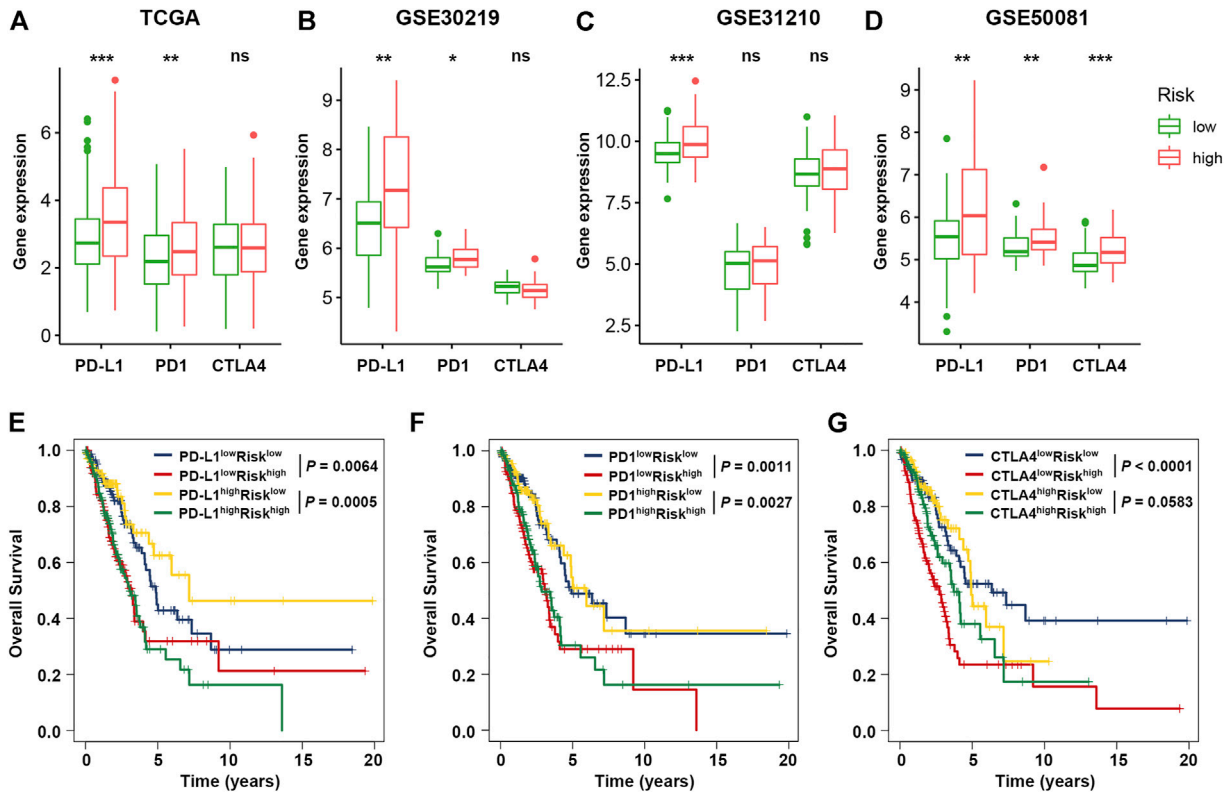


FIGURE 6 | Impact of immune checkpoint genes and SRS on clinical outcome. **(A–D)**. Comparison of the expression levels of immune checkpoint genes (PD-L1, PD1, and CTLA4) between the high- and low-risk groups in the training and validation cohorts. **(E–G)**. Kaplan-Meier survival curves of OS among four patient groups divided by the SRS and PD-L1, PD-1, or CTLA-4. *, **, and *** represent $P < 0.05$, $P < 0.01$, and $P < 0.001$, respectively.

with low SRS scores had significantly better survival than those with high SRS scores despite the fact that similar expression levels of immune checkpoint genes were observed in cohorts GSE50081 (**Supplementary Figure S8B**) and GSE31210 (**Supplementary Figure S8C**). However, no significant result was observed in cohort GSE30219 (**Supplementary Figure S8D**).

In addition to immune checkpoints, TMB is also considered an independent prognostic predictor in various cancer types. We first calculated the TMB of each group and found that patients with higher SRS scores had noticeably increased TMB relative to the low-risk group (**Supplementary Figure S9A**). Subsequently, the survival distribution of patient groups classified by SRS and TMB level was also compared. As shown in **Supplementary Figure S9B**, patients with high SRS scores suffered unfavourable OS irrespective of patients' TMB level.

These results imply that SRS combined with immune checkpoint expression or TMB might serve as a promising predictor of patients' clinical outcomes.

Predictive Potential of SRS in Immunotherapy Response

Growing evidence has shown that immune checkpoint inhibitors (ICIs) have improved the survival of NSCLC, but responses vary. Thus, accurate predictive biomarkers are urgently needed (Mok, et al., 2019; Reck, et al., 2016). Given the association between SRS and immune infiltration, we further explored the predictive potential of SRS of ICIs by analysing the correlation of SRS and recognized immunotherapy predictors, including TIDE (Fu, et al., 2020; Jiang, et al., 2018) and IPS(Charoentong, et al., 2017). We discovered that patients in the high-risk group tended to achieve higher TIDE scores in TCGA cohort, and this result was further confirmed in three validation cohorts (**Figure 7A**). In addition, IPS was significantly increased in the low SRS score group ($p < 0.001$), and patients' response to anti-CTLA4 treatment was relatively higher in the low-risk group ($p < 0.001$, **Figure 7B**). These results indicate that patients with low SRS scores may benefit from ICIs.

Considering the immunotherapy response predictive potential of SRS, we next performed Kaplan-Meier survival analysis to investigate the predictive role of immunotherapeutic overall survival using the immunotherapy cohort IMvigor210. As expected, a beneficial trend of low SRS scores in immunotherapeutic OS was observed in the IMvigor210 cohort (HR = 1.368 95% CI 1.036–1.808 $p = 0.0197$, **Figure 7C**), and the low-risk group also exhibited significantly better OS than the high-risk group among the PD-L1 high ($p = 0.0445$, **Figure 7D**) or TMB-low population ($p = 0.0366$, **Supplementary Figure S9C**). Collectively, SRS in combination with TMB or PD-L1 is a promising candidate for predicting the

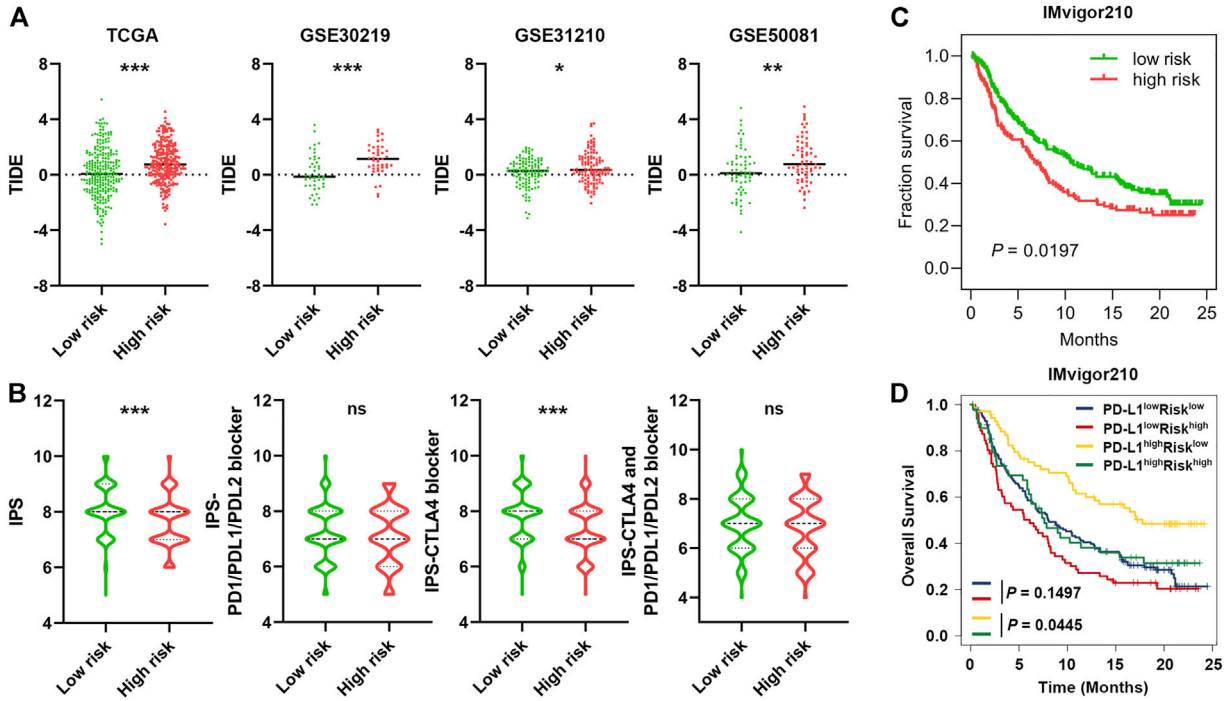


FIGURE 7 | Predictive potential of SRS in immunotherapeutic benefits. **(A)**. The distribution of TIDE scores between patients with a higher SRS score and those with a lower SRS score in four different cohorts as indicated. **(B)**. The distribution of IPS in the high-risk and low-risk groups in the TCGA dataset. **(C)**. Kaplan-Meier curves for high and low SRS score patient groups in the IMvigor210 cohort. **(D)**. Kaplan-Meier curves for four patient groups stratified by SRS and PD-L1 expression. *, **, and *** represent $P < 0.05$, $P < 0.01$, and $P < 0.001$, respectively.

response to ICIs among patients with LUAD and improving therapeutic strategies.

Validation of Signature Gene Expressions in LUAD Tissues

To further explore the protein expression of genes that constituted SRS, three genes that exhibited significantly higher expression level in lung cancer in GEPIA2 (**Supplementary Figure S5A**) were quantified by IHC and compared between in tumor tissues ($n = 74$) and adjacent normal tissues ($n = 74$). As expected, IHC staining revealed that protein expressions of FOXM1, HJURP, and PTTG1 were all significantly elevated in tumors compared with adjacent normal tissues (**Figures 8A–F**), indicating that SRS genes may play an important in lung cancer progression.

DISCUSSION

Senescence is a complex biological process with both cell autonomous and paracrine effects that has a significant impact on the microenvironment (Lasry and Ben-Neriah 2015; Hernandez-Segura, et al., 2018). Increasing evidence indicates that senescent cells can be eliminated through a SASP-provoked immune response, which involves both innate and adaptive immunity (Schneider, et al., 2021). Conceivably, the SASP has several positive functions in the short term. However, in the long term, these functions can become detrimental in the immunosuppressive context of cancer to promote tumor development (Lopes-Paciencia, et al., 2019; Coppe, et al., 2010; Basisty, et al., 2020; Birch and Gil 2020). However, how senescent cells interact with tumor immune infiltration and their value in evaluating the immune infiltrate of tumors and clinical outcomes have not been reported, particularly in lung cancer. Thus, modelling lung cancer will be important to decipher whether senescence molecular determinants reshape tumor microenvironments and whether this modification has implications for the prognosis and immunotherapy response of LUAD patients. Importantly, uncovering how cellular senescence influences the TIME can provide a window for discoveries of how we can effectively improve the immunosuppressive milieu by senolytic therapies (van Deursen 2019).

In this study, we analysed the expression patterns, prognostic values, and effects on the TIME of cellular senescence-related genes in LUAD. Using the LASSO method, we constructed a novel survival prediction model (SRS) based on the expression of five senescence features in the TCGA dataset. Furthermore, the SRS was well validated in three different public GEO datasets. We also explored the features of the immune microenvironment, including immune cell distribution and inflammatory activities, in patients with high and low SRS scores. Markedly, we distinguished different SASP affecting TIME remodelling as potential mechanisms underlying immune escape and tumor progression. Additionally, we found that the SRS score was an

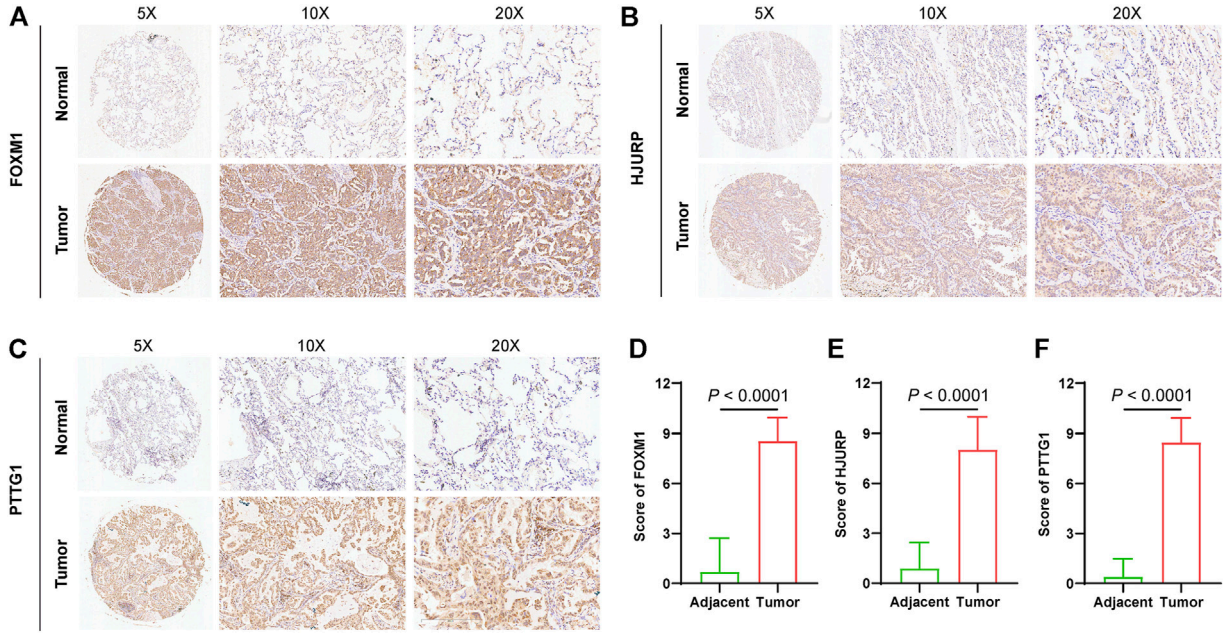


FIGURE 8 | Validation of signature gene expressions in LUAD tissues. Quantification and comparison of protein expression of FOXM1 **(A** and **D)**, HJURP **(B** and **E)**, and PTTG1 **(C** and **F)** by IHC.

independent prognostic factor for LUAD patients and was coupled with specific immune checkpoint factors or TMB as predictive biomarkers of ICI response.

This study represents one of the first reports to examine the differential expression analysis of cellular senescence-relevant genes and then identify their prognostic values using TCGA and GEO databases. Markedly, we identified five significantly upregulated genes, including FOXM1, HJURP, PKM, PTTG1, and TACC3, which were also included in the cellular senescence-related signature reported in this study. Interestingly, these five signature genes are reported as negative regulators of cellular senescence in many human cancers and play important roles in tumor development (Ao, et al., 2017; Caporali, et al., 2012; Chen, et al., 2010; Francica, et al., 2016; Kato, et al., 2007; Schmidt, et al., 2010; Tao, et al., 2014). Forkhead box protein M1 (FOXM1) is significantly associated with immunotherapy resistance in lung cancer and patients (Galbo, et al., 2021; Wang, et al., 2014). Holliday junction recognition protein (HJURP), a histone H3 chaperone, affects cell cycle progression, DNA repair and chromosome segregation during mitosis. HJURP is overexpressed in cancers (Wang, et al., 2020) and associated with poor prognosis in NSCLC (Wei, et al., 2019). Pyruvate kinase M (PKM) is a glycolytic enzyme required for tumor proliferation and progression. PKM2 acts as the key factor mediating Th17 cell differentiation (Damasceno, et al., 2020), and silencing PKM2 mRNA could decrease PD-L1 expression and cancer evasion of immune surveillance (Guo, et al., 2019). As an oncogene during spindle formation or chromosome segregation (Bernal, et al., 2002; Li, et al., 2013), pituitary tumor-transforming gene-1 (PTTG1) is an independent poor prognostic factor in NSCLC patients (Wang, et al., 2016) and can elicit an immunogenic response in NSCLC patients (Chiriva-Internati, et al., 2014). Transforming acidic coiled-coil protein 3 (TACC3) is involved in chromosomal alignment, separation, and cytokinesis, which is correlated with p53-mediated apoptosis (Schneider, et al., 2008; Zhang, et al., 2018). Additionally, TACC3 exerts as a prognostic biomarker for prostate cancer (Qie, et al., 2020), osteosarcoma (Matsuda, et al., 2018) and NSCLC (Jiang, et al., 2016), and high TACC3 expression is associated with increased immune cell infiltration and T cell exhaustion (Fan, et al., 2021). These published experimental efforts provide further evidence supporting that SRS has the potential to mirror LUAD prognosis based on immune landscape alterations.

As the role of cellular senescence is largely underexplored in cancer, it is important to gain more extensive insight into the linkage of cancer, senescence and the immune environment. However, to date, the effect of senescence on the tumor immune infiltrate and whether this would impact the response to ICIs have only been poorly studied. By performing a detailed characterization of the tumor immune infiltrate in patients with LUAD, we observed that cellular senescence-related genes could have substantial effects on the composition and distribution of the tumor immune infiltrate. In this study, we found that the SRS score was inversely associated with the infiltration levels of B cells, $CD4^+$ T cells and $CD8^+$ T cells, whereas neutrophils, CAFs, Tregs and resting NK cells were positively correlated with the SRS score in LUAD. This result suggested that patients with higher SRS score might have an immunosuppressive tumor microenvironment, which prevented immune clearance of tumor cells. GSEA also showed that upregulation of the TGF-β-associated pathway, which has been widely reported as an important factor to restrain antitumor immunity (Mariathasan, et al., 2018; Sheng, et al.,

2021), was prominently enriched in the high-risk group. Second, to further explore the mechanisms of immune remodelling by the increasing burden of senescent cells in tumors, we uncovered that alterations of SASP could impact TIME establishment, which ultimately contributes to immune escape and provokes tumor development. The high SRS score group exhibited increases in proinflammatory cytokines, including IL-1b, IL-6, and IL-8; growth factors, such as EGF, VEGF and IGFBP; receptors, such as ICAMs; and proteases, such as MMPs. These factors may modulate immune cell recruitment and have a tumor-promoting effect (Coppe, et al., 2010; Cuollo, et al., 2020; Basisty, et al., 2020; Lau and David 2019). Third, as mentioned above, cellular senescence-related immune remodelling may explain the diminished efficacy of checkpoint blockade. Intriguingly, we noticed that the exhausted T cell markers PD-(L)1, CTLA-4, LAG-3, and TIM3 were aberrantly increased in LUAD specimens with high SRS scores, indicating that T cells become progressively hypofunctional and hyporesponsive with senescence upregulation. This finding may explain the lower response to immunotherapy in older individuals.

Therefore, our findings provide obvious clinical significance. On the one hand, significantly prolonged survival was observed for patients with low SRS scores, suggesting that high-risk patients should receive more frequent clinical surveillance and corresponding measures to prevent disease recurrence and progression. On the other hand, given that only a proportion of patients can derive durable benefits from ICIs, we need more accurate biomarkers with clinical utility. The developed cellular senescence-related signature can be applied not only as a prognostic tool but also as guidance for individualized immunotherapy. Besides, Small molecules targeting FOXM1 (Gormally, et al., 2014; Li, et al., 2020), PKM2 (Ning, et al., 2017) and TACC3(Akbulut, et al., 2020; Polson, et al., 2018) have been developed, and demonstrated promising anticancer capacity *in vitro* and *in vivo* experiments. These findings highlight the potential for using these compounds for future clinical application. Furthermore, we propose that controlling cellular senescence-associated inflammation by targeting specific inflammatory mediators may have a beneficial therapeutic effect in the treatment of cancer. A new group of drugs, named senolytic drugs, including quercetin, navitoclax, and fisetin, have received increased attention, and preclinical clinical data of their potential role in combination with immunotherapy are emerging. Thus, this group of drugs may have vast implications (van Deursen 2019; Campisi, et al., 2019; Kolb, et al., 2021; Prasanna, et al., 2021).

Although our study reports the benefits of immunotherapy and prognosis in LUAD, this study has several limitations. First, the five-gene signature was developed and validated in a public dataset; thus, external validation in multicentre cohorts is needed. Second, it is necessary to perform prospective clinical trials to verify the applicability of our research results in LUAD patients receiving immunotherapy. Third, the regulatory mechanisms by which cellular senescence-related genes reshape the TIME warrant further *in vivo* and *in vitro* investigations. Moreover, further studies are also needed to illustrate how the aged TIME contributes to lung cancer development. Finally, the preliminary interpretation of mechanisms underlying the association between cellular senescence-related genes and worse response to ICIs must be further elucidated using basic experiments.

In conclusion, our study identified and validated a cellular senescence-related signature that is based on five cellular senescence-related genes as an indicator of immune cell infiltration in the TIME and had independent prognostic significance for patients with LUAD. Importantly, the SRS was significantly associated with the immune cell infiltration levels of LUAD patients and involved in the regulation of the LUAD immune microenvironment by SASP. Finally, we characterized the complex interplay between the SRS score and immune checkpoint genes in patient outcomes and suggested the potential of the SRS score coupled with specific immune checkpoint factors as predictive biomarkers of ICI response to enable a more precise selection of patients who will benefit from checkpoint inhibitor immunotherapy. Therefore, identifying cellular senescence-related genes affecting tumor immune responses and further studying their regulatory mechanisms might assist risk stratification and provide promising targets for improving the response of LUAD to immunotherapy.

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 National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences, and Peking Union Medical College. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YG and JH designed the study. ZW, FS, YY, ZC, and XF collected the data. WL and XW performed the data analysis and interpretated the data. XW and WL drafted the manuscript. YG revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by National Key R&D Program of China (2020AAA0109500), National Natural Science Foundation of China (82122053), the Beijing Municipal Science & Technology Commission (Z191100006619115), R&D Program of Beijing Municipal Education commission (KJZD20191002302), CAMS Initiative for Innovative Medicine (2021-1-I2M-012), Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (2021-PT310-001), and Aiyou Foundation (KY201701).

ACKNOWLEDGMENTS

The authors would like to thank the staff members of the TCGA Research Network, the GEO data as well as all the authors for making their valuable research data public.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.780461/full#supplementary-material

REFERENCES

Acosta, J. C., O'Loghlen, A., Banito, A., Guijarro, M. V., Augert, A., Raguz, S., et al. (2008). Chemokine Signaling via the CXCR2 Receptor Reinforces Senescence. *Cell* 133, 1006–1018. doi:10.1016/j.cell.2008.03.038

Akbulut, O., Lengerli, D., Saatci, O., Duman, E., Seker, U. O. S., Isik, A., et al. (2020). A Highly Potent Tacc3 Inhibitor as a Novel Anticancer Drug Candidate. *Mol. Cancer Ther.* 19, 1243–1254. doi:10.1158/1535-7163.mct-19-0957

Ao, R., Guan, L., Wang, Y., and Wang, J.-N. (2017). Effects of PKM2 Gene Silencing on the Proliferation and Apoptosis of Colorectal Cancer LS-147T and SW620 Cells. *Cell Physiol. Biochem.* 42, 1769–1778. doi:10.1159/000479456

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

Avelar, R. A., Ortega, J. G., Tacutu, R., Tyler, E. J., Bennett, D., Binetti, P., et al. (2020). A Multidimensional Systems Biology Analysis of Cellular Senescence in Aging and Disease. *Genome Biol.* 21, 91. doi:10.1186/s13059-020-01990-9

Basisty, N., Kale, A., Jeon, O. H., Kuehnemann, C., Payne, T., Rao, C., et al. (2020). A Proteomic Atlas of Senescence-Associated Secretomes for Aging Biomarker Development. *Plos Biol.* 18, e3000599. doi:10.1371/journal.pbio.3000599

Behera, M., Owonikoko, T. K., Gal, A. A., Steuer, C. E., Kim, S., Pillai, R. N., et al. (2016). Lung Adenocarcinoma Staging Using the 2011 Iaslc/ats/ers Classification: A Pooled Analysis of Adenocarcinoma *In Situ* and Minimally Invasive Adenocarcinoma. *Clin. Lung Cancer* 17, e57–e64. doi:10.1016/j.cllc.2016.03.009

Ben-Porath, I., and Weinberg, R. A. (2005). The Signals and Pathways Activating Cellular Senescence. *Int. J. Biochem. Cel. Biol.* 37, 961–976. doi:10.1016/j.biocel.2004.10.013

Ben-Porath, I., and Weinberg, R. A. (2004). When Cells Get Stressed: An Integrative View of Cellular Senescence. *J. Clin. Invest.* 113, 8–13. doi:10.1172/jci200420663

Berben, L., Floris, G., and Wildiers, H., 2021. Cancer and Aging: Two Tightly Interconnected Biological Processes Cancers (Basel) 13. doi:10.3390/cancers13061400

Bernal, J. A., Luna, R., Espina, Á., Lázaro, I., Ramos-Morales, F., Romero, F., et al. (2002). Human Securin Interacts with P53 and Modulates P53-Mediated Transcriptional Activity and Apoptosis. *Nat. Genet.* 32, 306–311. doi:10.1038/ng997

Birch, J., and Gil, J. (2020). Senescence and the SASP: many Therapeutic Avenues. *Genes Dev.* 34, 1565–1576. doi:10.1101/gad.343129.120

Blanche, P., Dartigues, J.-F., and Jacqmin-Gadda, H. (2013). Estimating and Comparing Time-dependent Areas under Receiver Operating Characteristic Curves for Censored Event Times with Competing Risks. *Statist. Med.* 32, 5381–5397. doi:10.1002/sim.5958

Campisi, J. (2013). Aging, Cellular Senescence, and Cancer. *Annu. Rev. Physiol.* 75, 685–705. doi:10.1146/annurev-physiol-030212-183653

Campisi, J., Kapahi, P., Lithgow, G. J., Melov, S., Newman, J. C., and Verdin, E. (2019). From Discoveries in Ageing Research to Therapeutics for Healthy Ageing. *Nature* 571, 183–192. doi:10.1038/s41586-019-1365-2

Caporali, S., Alvino, E., Levati, L., Esposito, A. I., Ciomei, M., Brasca, M. G., et al. (2012). Down-regulation of the Pttg1 Proto-Oncogene Contributes to the Melanoma Suppressive Effects of the Cyclin-dependent Kinase Inhibitor Pha-848125. *Biochem. Pharmacol.* 84, 598–611. doi:10.1016/j.bcp.2012.06.004

Chandrashekar, D. S., Bashel, B., Balasubramanya, S. A. H., Creighton, C. J., Ponce-Rodriguez, I., Chakravarthi, B. V. S. K., et al. (2017). Ualcan: A portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia* 19, 649–658. doi:10.1016/j.neo.2017.05.002

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, W.-S., Yu, Y.-C., Lee, Y.-J., Chen, J.-H., Hsu, H.-Y., and Chiu, S.-J. (2010). Depletion of Securin Induces Senescence after Irradiation and Enhances Radiosensitivity in Human Cancer Cells Regardless of Functional P53 Expression. *Int. J. Radiat. Oncology*Biology*Physics* 77, 566–574. doi:10.1016/j.ijrobp.2009.12.013

Chiriva-Internati, M., Mirandola, L., and Figueroa, J. A., 2014. Selective Expression and Immunogenicity of the Cancer/testis Antigens Sp17, Akap4 and Pttg1 in Non-small Cell Lung Cancer: New Candidates for Active Immunotherapy Chest.

Coppé, J.-P., Desprez, P.-Y., Krtolica, A., and Campisi, J. (2010). The Senescence-Associated Secretory Phenotype: The Dark Side of Tumor Suppression. *Annu. Rev. Pathol. Mech. Dis.* 5, 99–118. doi:10.1146/annurev-pathol-121808-102144

Coppé, J. P., Patil, C. K., Rodier, F., Sun, Y., Muñoz, D. P., Goldstein, J., et al. (2008). Senescence-associated Secretory Phenotypes Reveal Cell-Nonautonomous Functions of Oncogenic RAS and the P53 Tumor Suppressor. *Plos Biol.* 6, 2853–2868. doi:10.1371/journal.pbio.0060301

Cuollo, L., Antonangeli, F., Santoni, A., and Soriani, A. (2020). The Senescence-Associated Secretory Phenotype (SASP) in the Challenging Future of Cancer Therapy and Age-Related Diseases. *Biology (Basel)* 9. doi:10.3390/biology9120485

Damasceno, L. E. A., Prado, D. S., Veras, F. P., Fonseca, M. M., Toller-Kawahisa, J. E., Rosa, M. H., et al. (2020). Pkm2 Promotes Th17 Cell Differentiation and Autoimmune Inflammation by fine-tuning Stat3 Activation. *J. Exp. Med.* 217. doi:10.1084/jem.20190613

Der, S. D., Sykes, J., Pintilie, M., Zhu, C.-Q., Strumpf, D., Liu, N., et al. (2014). Validation of a Histology-independent Prognostic Gene Signature for Early-Stage, Non-small-cell Lung Cancer Including Stage Ia Patients. *J. Thorac. Oncol.* 9, 59–64. doi:10.1097/jto.0000000000000042

Eggert, T., Wolter, K., Ji, J., Ma, C., Yevsa, T., Klotz, S., et al. (2016). Distinct Functions of Senescence-Associated Immune Responses in Liver Tumor Surveillance and Tumor Progression. *Cancer Cell* 30, 533–547. doi:10.1016/j.ccell.2016.09.003

Fan, X., Liu, B., Wang, Z., and He, D. (2021). TACC3 Is a Prognostic Biomarker for Kidney Renal clear Cell Carcinoma and Correlates with Immune Cell Infiltration and T Cell Exhaustion. *Aging* 13, 8541–8562. doi:10.18632/aging.202668

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

Francica, P., Nisa, L., Aebersold, D. M., Langer, R., Bladt, F., Blaukat, A., et al. (2016). Depletion of Foxm1 via Met Targeting Underlies Establishment of a DNA Damage-Induced Senescence Program in Gastric Cancer. *Clin. Cancer Res.* 22, 5322–5336. doi:10.1158/1078-0432.ccr-15-2987

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* 33, 1–22. doi:10.18637/jss.v033.i01

Fu, J., Li, K., Zhang, W., Wan, C., Zhang, J., Jiang, P., et al. (2020). Large-scale Public Data Reuse to Model Immunotherapy Response and Resistance. *Genome Med.* 12, 21. doi:10.1186/s13073-020-0721-z

Galbo, P. M., Jr., Zang, X., and Zheng, D. (2021). Molecular Features of Cancer-Associated Fibroblast Subtypes and Their Implication on Cancer Pathogenesis, Prognosis, and Immunotherapy Resistance. *Clin. Cancer Res.* 27, 2636–2647. doi:10.1158/1078-0432.ccr-20-4226

Gormally, M. V., Dexheimer, T. S., Marsico, G., Sanders, D. A., Lowe, C., Matak-Vinković, D., et al. (2014). Suppression of the Foxm1 Transcriptional Programme via Novel Small Molecule Inhibition. *Nat. Commun.* 5, 5165. doi:10.1038/ncomms6165

Green, M. R. (2008). Senescence: Not Just for Tumor Suppression. *Cell* 134, 562–564. doi:10.1016/j.cell.2008.08.003

Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). Circlize Implements and Enhances Circular Visualization in R. *Bioinformatics* 30, 2811–2812. doi:10.1093/bioinformatics/btu393

Guo, C.-Y., Zhu, Q., Tou, F.-F., Wen, X.-M., Kuang, Y.-K., and Hu, H. (2019). The Prognostic Value of Pkm2 and its Correlation with Tumour Cell Pd-L1 in Lung Adenocarcinoma. *BMC Cancer* 19, 289. doi:10.1186/s12885-019-5519-2

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

Hernandez-Segura, A., Nehme, J., and Demaria, M. (2018). Hallmarks of Cellular Senescence. *Trends Cel. Biol.* 28, 436–453. doi:10.1016/j.tcb.2018.02.001

Jiang, F., Kuang, B., Que, Y., Lin, Z., Yuan, L., Xiao, W., et al. (2016). The Clinical Significance of Transforming Acidic Coiled-Coil Protein 3 Expression in Non-small Cell Lung Cancer. *Oncol. Rep.* 35, 436–446. doi:10.3892/or.2015.4373

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

Juneja, V. R., McGuire, K. A., Manguso, R. T., LaFleur, M. W., Collins, N., Haining, W. N., et al. (2017). Pd-l1 on Tumor Cells Is Sufficient for Immune Evasion in Immunogenic Tumors and Inhibits Cd8 T Cell Cytotoxicity. *J. Exp. Med.* 214, 895–904. doi:10.1084/jem.20160801

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). Kegg as a Reference Resource for Gene and Protein Annotation. *Nucleic Acids Res.* 44, D457–D462. doi:10.1093/nar/gkv1070

Kato, T., Noma, K., Ohara, T., Kashima, H., Katsura, Y., Sato, H., et al. (2018). Cancer-Associated Fibroblasts Affect Intratumoral CD8+ and FoxP3+ T Cells via IL6 in the Tumor Microenvironment. *Clin. Cancer Res.* 24, 4820–4833. doi:10.1158/1078-0432.ccr-18-0205

Kato, T., Sato, N., Hayama, S., Yamabuki, T., Ito, T., Miyamoto, M., et al. (2007). Activation of Holliday Junction-Recognizing Protein Involved in the Chromosomal Stability and Immortality of Cancer Cells. *Cancer Res.* 67, 8544–8553. doi:10.1158/0008-5472.can-07-1307

Kolb, R., De, U., Khan, S., Luo, Y., Kim, M.-C., Yu, H., et al. (2021). Proteolysis-targeting Chimera against Bcl-Xl Destroys Tumor-Infiltrating Regulatory T Cells. *Nat. Commun.* 12, 1281. doi:10.1038/s41467-021-21573-x

Krizhanovsky, V., Xue, W., Zender, L., Yon, M., Hernando, E., and Lowe, S. W. (2008). Implications of Cellular Senescence in Tissue Damage Response, Tumor Suppression, and Stem Cell Biology. *Cold Spring Harbor Symposia Quantitative Biol.* 73, 513–522. doi:10.1101/sqb.2008.73.048

Kuilman, T., Michaloglou, C., Vredeveld, L. C. W., Douma, S., van Doorn, R., Desmet, C. J., et al. (2008). Oncogene-Induced Senescence Relayed by an Interleukin-dependent Inflammatory Network. *Cell* 133, 1019–1031. doi:10.1016/j.cell.2008.03.039

Kumagai, S., Togashi, Y., Kamada, T., Sugiyama, E., Nishinakamura, H., Takeuchi, Y., et al. (2020). The Pd-1 Expression Balance between Effector and Regulatory T Cells Predicts the Clinical Efficacy of Pd-1 Blockade Therapies. *Nat. Immunol.* 21, 1346–1358. doi:10.1038/s41590-020-0769-3

Kumari, R., and Jat, P. (2021). Mechanisms of Cellular Senescence: Cell Cycle Arrest and Senescence Associated Secretory Phenotype. *Front. Cel. Dev. Biol.* 9, 645593. doi:10.3389/fcell.2021.645593

Lamano, J. B., Lamano, J. B., Li, Y. D., DiDomenico, J. D., Choy, W., Veliceasa, D., et al. (2019). Glioblastoma-derived Il6 Induces Immunosuppressive Peripheral Myeloid Cell Pd-L1 and Promotes Tumor Growth. *Clin. Cancer Res.* 25, 3643–3657. doi:10.1158/1078-0432.ccr-18-2402

Lasry, A., and Ben-Neriah, Y. (2015). Senescence-associated Inflammatory Responses: Aging and Cancer Perspectives. *Trends Immunol.* 36, 217–228. doi:10.1016/j.it.2015.02.009

Lau, L., and David, G. (2019). Pro- and Anti-tumorigenic Functions of the Senescence-Associated Secretory Phenotype. *Expert Opin. Ther. Targets* 23, 1041–1051. doi:10.1080/14728222.2019.1565658

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, H., Yin, C., Zhang, B., Sun, Y., Shi, L., Liu, N., et al. (2013). Pttg1 Promotes Migration and Invasion of Human Non-small Cell Lung Cancer Cells and Is Modulated by Mir-186. *Carcinogenesis* 34, 2145–2155. doi:10.1093/carcin/bgt158

Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. S., et al. (2017). TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* 77, e108–e110. doi:10.1158/0008-5472.can-17-0307

Li, T., Fu, J., Zeng, Z., Cohen, D., Li, J., Chen, Q., et al. (2020). Timer2.0 for Analysis of Tumor-Infiltrating Immune Cells. *Nucleic Acids Res.* 48, W509–W514. doi:10.1093/nar/gkaa407

Li, Y., Lu, L., Tu, J., Zhang, J., Xiong, T., Fan, W., et al. (2020). Reciprocal Regulation between Forkhead Box M1/NF-κB and Methionine Adenosyltransferase 1A Drives Liver Cancer. *Hepatology* 72, 1682–1700. doi:10.1002/hep.31196

Liu, C., Peng, W., Xu, C., Lou, Y., Zhang, M., Wargo, J. A., et al. (2013). Braf Inhibition Increases Tumor Infiltration by T Cells and Enhances the Antitumor Activity of Adoptive Immunotherapy in Mice. *Clin. Cancer Res.* 19, 393–403. doi:10.1158/1078-0432.ccr-12-1626

Lopes-Paciencia, S., Saint-Germain, E., Rowell, M.-C., Ruiz, A. F., Kalegari, P., and Ferbeyre, G. (2019). The Senescence-Associated Secretory Phenotype and its Regulation. *Cytokine* 117, 15–22. doi:10.1016/j.cyto.2019.01.013

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, 544–548. doi:10.1038/nature25501

Matsuda, K., Miyoshi, H., Hiraoka, K., Hamada, T., Nakashima, K., Shiba, N., et al. (2018). Elevated Expression of Transforming Acidic Coiled-Coil Containing Protein 3 (Tacc3) Is Associated with a Poor Prognosis in Osteosarcoma. *Clin. Orthop. Relat. Res.* 476, 1848–1855. doi:10.1097/corr.0000000000000379

Miller, K. D., Nogueira, L., Mariotto, A. B., Rowland, J. H., Yabroff, K. R., Alfano, C. M., et al. (2019). Cancer Treatment and Survivorship Statistics, 2019. *CA A. Cancer J. Clin.* 69, 363–385. doi:10.3322/caac.21565

Mok, T. S. K., Wu, Y. L., Kudaba, I., Kowalski, D. M., Cho, B. C., Turna, H. Z., et al. (2019). Pembrolizumab versus Chemotherapy for Previously Untreated, Pd-L1-Expressing, Locally Advanced or Metastatic Non-small-cell Lung Cancer (Keynote-042): A Randomised, Open-Label, Controlled, Phase 3 Trial. *Lancet* 393, 1819–1830. doi:10.1016/S0140-6736(18)32409-7

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust Enumeration of Cell Subsets from Tissue Expression Profiles. *Nat. Methods* 12, 453–457. doi:10.1038/nmeth.3337

Ning, X., Qi, H., Li, R., Li, Y., Jin, Y., McNutt, M. A., et al. (2017). Discovery of Novel Naphthoquinone Derivatives as Inhibitors of the Tumor Cell Specific M2 Isoform of Pyruvate Kinase. *Eur. J. Med. Chem.* 138, 343–352. doi:10.1016/j.ejmech.2017.06.064

Okayama, H., Kohno, T., Ishii, Y., Shimada, Y., Shiraishi, K., Iwakawa, R., et al. (2012). Identification of Genes Upregulated in ALK-Positive and EGFR/KRAS/ALK-Negative Lung Adenocarcinomas. *Cancer Res.* 72, 100–111. doi:10.1158/0008-5472.can-11-1403

Partridge, L., Deelen, J., and Slagboom, P. E. (2018). Facing up to the Global Challenges of Ageing. *Nature* 561, 45–56. doi:10.1038/s41586-018-0457-8

Pérez-Mancera, P. A., Young, A. R. J., and Narita, M. (2014). Inside and Out: The Activities of Senescence in Cancer. *Nat. Rev. Cancer* 14, 547–558. doi:10.1038/nrc3773

Polson, E. S., Kuchler, V. B., Abbosh, C., Ross, E. M., Mathew, R. K., Beard, H. A., et al. (2018). Khs101 Disrupts Energy Metabolism in Human Glioblastoma Cells and Reduces Tumor Growth in Mice. *Sci. Transl. Med.* 10. doi:10.1126/scitranslmed.aar2718

Prasanna, P. G., Citrin, D. E., Hildesheim, J., Ahmed, M. M., Venkatachalam, S., Riscuta, G., et al. (2021). Therapy-Induced Senescence: Opportunities to Improve Anti-cancer Therapy. *J. Natl. Cancer Inst.* 113, 1285–1298. doi:10.1093/jnci/djab064

Qie, Y., Wang, L., Du, E., Chen, S., Lu, C., Ding, N., et al. (2020). Tacc3 Promotes Prostate Cancer Cell Proliferation and Restrains Primary Cilium Formation. *Exp. Cel. Res.* 390, 111952. doi:10.1016/j.yexcr.2020.111952

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. doi:10.7554/eLife.26476

Reck, M., Rodríguez-Abreu, D., Robinson, A. G., Hui, R., Csőszi, T., Fülöp, A., et al. (2016). Pembrolizumab versus Chemotherapy for Pd-L1-Positive Non-small-cell Lung Cancer. *N. Engl. J. Med.* 375, 1823–1833. doi:10.1056/nejmoa1606774

Rousseaux, S., Debernardi, A., Jacquiau, B., Vitte, A. L., Vesin, A., Nagy-Mignotte, H., et al. (2013). Ectopic Activation of Germline and Placental Genes Identifies Aggressive Metastasis-Prone Lung Cancers. *Sci. Transl. Med.* 5. 186ra66. doi:10.1126/scitranslmed.3005723

Ru, B., Wong, C. N., Tong, Y., Zhong, J. Y., Zhong, S. S. W., Wu, W. C., et al. (2019). Tisidb: An Integrated Repository portal for Tumor-Immune System Interactions. *Bioinformatics* 35, 4200–4202. doi:10.1093/bioinformatics/btz210

Schmidt, S., Schneider, L., Essmann, F., Cirstea, I. C., Kuck, F., Kletke, A., et al. (2010). The Centrosomal Protein Tacc3 Controls Paclitaxel Sensitivity by Modulating a Premature Senescence Program. *Oncogene* 29, 6184–6192. doi:10.1038/onc.2010.354

Schneider, J. L., Rowe, J. H., Garcia-de-Alba, C., Kim, C. F., Sharpe, A. H., and Haigis, M. C. (2021). The Aging Lung: Physiology, Disease, and Immunity. *Cell* 184, 1990–2019. doi:10.1016/j.cell.2021.03.005

Schneider, L., Essmann, F., Kletke, A., Rio, P., Hanenberg, H., Schulze-Osthoff, K., et al. (2008). Tacc3 Depletion Sensitizes to Paclitaxel-Induced Cell Death and Overrides P21waf-Mediated Cell Cycle Arrest. *Oncogene* 27, 116–125. doi:10.1038/sj.onc.1210628

Sharma, P., Hu-Lieskovan, S., Wargo, J. A., and Ribas, A. (2017). Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy. *Cell* 168, 707–723. doi:10.1016/j.cell.2017.01.017

Sheng, W., Liu, Y., and Chakraborty, D. (2021). *Simultaneous Inhibition of Lsd1 and Tgf-B Enables Eradication of Poorly Immunogenic Tumors with Anti-pd-1 Treatment Cancer Discov*.

Sturm, G., Finotello, F., Petitprez, F., Zhang, J. D., Baumbach, J., Fridman, W. H., et al. 2019, Comprehensive Evaluation of Transcriptome-Based Cell-type Quantification Methods for Immuno-Oncology 35: i436-i445.doi:10.1093/bioinformatics/btz363

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. *Proc. Natl. Acad. Sci.* 102, 15545–15550. doi:10.1073/pnas.0506580102

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, 209–249. doi:10.3322/caac.21660

Tang, Z., Kang, B., Li, C., Chen, T., and Zhang, Z. (2019). Gepia2: An Enhanced Web Server for Large-Scale Expression Profiling and Interactive Analysis. *Nucleic Acids Res.* 47, W556–W560. doi:10.1093/nar/gkz430

Tao, J., Xu, X.-S., Song, Y.-Z., Qu, K., Wu, Q.-F., Wang, R.-T., et al. (2014). Down-regulation of FoxM1 Inhibits Viability and Invasion of Gallbladder Carcinoma Cells, Partially Dependent on Inducement of Cellular Senescence. *Wjg* 20, 9497–9505. doi:10.3748/wjg.v20.i28.9497

The Gene Ontology, C. (2019). The Gene Ontology Resource: 20 Years and Still Going strong. *Nucleic Acids Res.* 47, D330–D338. doi:10.1093/nar/gky1055

Tibshirani, R. (1997). The Lasso Method for Variable Selection in the Cox Model. *Statist. Med.* 16, 385–395. doi:10.1002/(sici)1097-0258(19970228)16:4<385:aid-sim380>3.0.co;2-3

van Deursen, J. M. (2019). Senolytic Therapies for Healthy Longevity. *Science* 364, 636–637. doi:10.1126/science.aaw1299

Wang, C.-j., Li, X., Shi, P., Ding, H.-y., Liu, Y.-p., Li, T., et al. (2020). Holliday junction Recognition Protein Promotes Pancreatic Cancer Growth and Metastasis via Modulation of the Mdm2/p53 Signaling. *Cell Death Dis.* 11, 386. doi:10.1038/s41419-020-2595-9

Wang, F., Liu, Y., and Chen, Y. (2016). Pituitary Tumor Transforming Gene-1 in Non-small Cell Lung Cancer: Clinicopathological and Immunohistochemical Analysis. *Biomed. Pharmacother.* 84, 1595–1600. doi:10.1016/j.biopha.2016.10.047

Wang, I.-C., Ustiyan, V., Zhang, Y., Cai, Y., Kalin, T. V., and Kalinichenko, V. V. (2014). Foxm1 Transcription Factor Is Required for the Initiation of Lung Tumorigenesis by Oncogenic KrasG12D. *Oncogene* 33, 5391–5396. doi:10.1038/onc.2013.475

Wei, Y., Ouyang, G. L., Yao, W. X., Zhu, Y. J., Li, X., Huang, L. X., et al. (2019). Knockdown of HJURP Inhibits Non-small Cell Lung Cancer Cell Proliferation, Migration, and Invasion by Repressing Wnt/β-Catenin Signaling. *Eur. Rev. Med. Pharmacol. Sci.* 23, 3847–3856. doi:10.26355/eurrev_201905_17812

Xue, W., Zender, L., Miething, C., Dickins, R. A., Hernando, E., Krizhanovsky, V., et al. (2007). Senescence and Tumour Clearance Is Triggered by P53 Restoration in Murine Liver Carcinomas. *Nature* 445, 656–660. doi:10.1038/nature05529

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16, 284–287. doi:10.1089/omi.2011.0118

Zhang, Y., Tan, L., Yang, Q., Li, C., and Liou, Y.-C. (2018). The Microtubule-Associated Protein Hurp Recruits the Centrosomal Protein Tacc3 to Regulate K-Fiber Formation and Support Chromosome Congression. *J. Biol. Chem.* 293, 15733–15747. doi:10.1074/jbc.ra118.003676

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: 30 November 2021
doi: 10.3389/fcell.2021.781285

Identification of Pan-Cancer Biomarkers Based on the Gene Expression Profiles of Cancer Cell Lines

ShiJian Ding[1†], Hao Li[2†], Yu-Hang Zhang[3†], XianChao Zhou[4], KaiYan Feng[5], ZhanDong Li[2], Lei Chen[6], Tao Huang[7,8*] and Yu-Dong Cai[1*]*

[1]School of Life Sciences, Shanghai University, Shanghai, China, [2]College of Food Engineering, Jilin Engineering Normal University, Changchun, China, [3]Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States, [4]Center for Single-Cell Omics, School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [5]Department of Computer Science, Guangdong AIB Polytechnic College, Guangzhou, China, [6]College of Information Engineering, Shanghai Maritime University, Shanghai, China, [7]CAS Key Laboratory of Computational Biology, Bio-Med Big Data Center, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China, [8]CAS Key Laboratory of Tissue Microenvironment and Tumor, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China

OPEN ACCESS

Edited by:
Lu Xie,
Shanghai Center For Bioinformation Technology, China

Reviewed by:
Bo Zhou,
Shanghai University of Medicine and Health Sciences, China
Qi Dai,
Zhejiang Sci-Tech University, China

***Correspondence:**
Lei Chen
chen_lei1@163.com
Tao Huang
tohuangtao@126.com
Yu-Dong Cai
cai_yud@126.com

[†]These authors 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: 22 September 2021
Accepted: 16 November 2021
Published: 30 November 2021

Citation:
Ding S, Li H, Zhang Y-H, Zhou X, Feng K, Li Z, Chen L, Huang T and Cai Y-D (2021) Identification of Pan-Cancer Biomarkers Based on the Gene Expression Profiles of Cancer Cell Lines. Front. Cell Dev. Biol. 9:781285. doi: 10.3389/fcell.2021.781285

There are many types of cancers. Although they share some hallmarks, such as proliferation and metastasis, they are still very different from many perspectives. They grow on different organ or tissues. Does each cancer have a unique gene expression pattern that makes it different from other cancer types? After the Cancer Genome Atlas (TCGA) project, there are more and more pan-cancer studies. Researchers want to get robust gene expression signature from pan-cancer patients. But there is large variance in cancer patients due to heterogeneity. To get robust results, the sample size will be too large to recruit. In this study, we tried another approach to get robust pan-cancer biomarkers by using the cell line data to reduce the variance. We applied several advanced computational methods to analyze the Cancer Cell Line Encyclopedia (CCLE) gene expression profiles which included 988 cell lines from 20 cancer types. Two feature selection methods, including Boruta, and max-relevance and min-redundancy methods, were applied to the cell line gene expression data one by one, generating a feature list. Such list was fed into incremental feature selection method, incorporating one classification algorithm, to extract biomarkers, construct optimal classifiers and decision rules. The optimal classifiers provided good performance, which can be useful tools to identify cell lines from different cancer types, whereas the biomarkers (e.g. NCKAP1, TNFRSF12A, LAMB2, FKBP9, PFN2, TOM1L1) and rules identified in this work may provide a meaningful and precise reference for differentiating multiple types of cancer and contribute to the personalized treatment of tumors.

Keywords: pan-cancer study, feature selection, classification algorithm, decision rule, biomarker

INTRODUCTION

"Cancer" is the term used to describe a series of diseases that is characterized by the spontaneous expansion and spread of somatic cell clones. It is becoming a serious public health problem worldwide. In 2020 alone, over 19.29 million new cases of cancer were diagnosed, and more than 9.58 million people died from cancer (World Health Organization, 2019). The hallmarks of cancer have been extensively described as six biological capabilities, namely, enhanced proliferative signaling, growth suppressor escape, cell death resistance, replicative immortality, angiogenesis induction, and invasion and metastasis activation (Hanahan and Weinberg, 2011). In other words, the pro-oncogenic function is the abnormal expression of various genes based on these six biological capabilities. Therefore, cancer genomic data, particularly gene expression signatures, can provide insight into the occurrence and development of cancers and, importantly, can be used to develop targeted therapies for cancers (Garman et al., 2007).

Although many cancers share the hallmarks of cancer, they are still very different. They grow on different organs and tissue. Pan-cancer studies provide an opportunity to understand the commonalities, heterogeneity, and emergent themes of multiple tumors (Andor et al., 2016). Increased numbers of tumor sample datasets provide scientists with a clear picture of tumors, rare driver events in heterogeneous tumor samples, and new molecular carcinogenic mechanisms that may be readily detected (Weinstein et al., 2013). For example, a study on the genomic predictors of the drug sensitivity of 947 human cancer cell lines based on a cancer cell line encyclopedia revealed known and novel response candidate biomarkers, which may contribute to cancer biology and therapeutic development (Barretina et al., 2012). Another study on long noncoding RNA (lncRNA) in 5185 TCGA tumors demonstrated that although tumor-specific dysregulated lncRNAs are commonly observed in a variety of tumors, genes and pathways could be synergistically regulated in different cancers by the same group of lncRNAs; this information may provide useful ideas for the development of broad-spectrum antineoplastic drugs (Chiu et al., 2018).

The sample size of TCGA based pan-cancer studies is already very large as a multi-omics data source. But it is still not enough to get robust pan-cancer biomarkers if we consider the large variances among cancer patients across cancer types and even within the same cancer. Tumor heterogeneity can be broadly categorized into intratumor heterogeneity and inter tumor heterogeneity (Burrell and Swanton, 2014). Inter tumor heterogeneity refers to the heterogeneity between patients with the same histological tumor type and has been considered to be caused by patient-specific factors, including germline mutations, individualized somatic mutations, and environmental factors. Intratumor heterogeneity can be divided into spatial heterogeneity (different regions of the tumor have different genetic aberrations) and temporal heterogeneity (during disease progression) (Dagogo-Jack and Shaw, 2018). Studies across multiple cancers have suggested that intratumor heterogeneity promotes tumor growth, metastasis, and drug resistance in human cancers (Hyo-eun et al., 2015; Russo et al., 2016). Therefore, treatment strategies with increased effectiveness and durability still need to be developed on the basis of a comprehensive understanding of tumor dynamics.

To get robust pan-biomarkers, there are two approaches: increase the sample size or reduce the variance. TCGA and the following works tried the first approach of increasing sample size. In this study, we would like to try the second approach of reducing the variance by analyzing the cancer cell line data from Cancer Cell Line Encyclopedia (CCLE) (Ghandi et al., 2019). The important genes were extracted by using the Boruta method (Kursa and Rudnicki, 2010). These genes were further analyzed with the max-relevance and min-redundancy (mRMR) method to evaluate their importance and sort them in a feature list. This list was fed into the incremental feature selection (IFS) method (Liu and Setiono, 1998) that combined support vector machine (SVM) (Cortes and Vapnik, 1995) or decision tree (DT) (Safavian and Landgrebe, 1991) to identify important genes and decision rules and build powerful classifiers. Further analysis was performed through a literature review of the top-ranked genes and portion decision rules to confirm the validity and reliability of the results. This study gives new insight into pan-cancer studies and may provide novel targets of tumor-specific therapies.

MATERIALS AND METHODS

Datasets

Xiao et al. (Xiao et al., 2019) downloaded the raw RNA-Seq data from Cancer Cell Line Encyclopedia (CCLE) (Ghandi et al., 2019) and quantified the gene expression levels as Transcripts Per kilobase Million (TPM) using RSEM (Li and Dewey, 2011). We used the processed gene expression data by Xiao et al. (Xiao et al., 2019). The cancer types with sample sizes of less than 10 was removed. Finally, there were 988 cancer cell lines from 20 cancer types. The sample size of each tumor is listed in **Table 1**. For each sample, 57,820 gene features were included. We investigated the expression patterns of genes in different tumor types and whether these tumor types could be distinguished on the basis of expression profiles.

Boruta Feature Selection

The CCLE data involved a large number of genes (features). Obviously, not all genes are associated with the investigated tumor types. Therefore, filtering the important genes is necessary. Here, we applied the Boruta (Kursa and Rudnicki, 2010) method to select a set of relevant features with multiple tumor labels.

The Boruta method is a wrapping algorithm that is based on random forest (RF) and involves the following steps: 1) the new shuffled data are generated by copying the original dataset and shuffling original features; 2) a RF classifier that can output the importance score of each feature is trained by using the new feature matrix as the input; and 3) the features in the original features that are sincerely relevant to the labels are retained, and the shuffled data are removed. Boruta finally selects the relevant features after several iterations of the above three steps.

TABLE 1 | Distribution of samples and decision rules in different cancer cell lines.

Cancer cell line types	Number of cell lines	Number of decision rules	Number of criteria	Number of involved genes
Autonomic ganglia	16	4	42	17
Bone	20	4	46	19
Breast	51	23	325	63
Central nervous system	65	18	219	57
Endometrium	28	16	191	52
Fibroblast	37	3	28	15
Haematopoietic and lymphoid tissue	173	8	96	34
Kidney	32	7	88	38
Large intestine	56	9	123	47
Liver	25	9	115	42
Lung	188	51	740	80
Oesophagus	27	8	145	47
Ovary	47	25	306	67
Pancreas	41	14	208	56
Skin	49	7	83	34
Soft tissue	28	9	126	41
Stomach	37	26	361	72
Thyroid	12	7	79	38
Upper aerodigestive tract	31	11	162	47
Urinary tract	25	16	216	59

The Boruta program that we used in this research was downloaded from https://github.com/scikit-learn-contrib/boruta_py and was set to default parameters for execution.

mRMR

After feature filtering by using the Boruta method, the mRMR (Peng et al., 2005) feature selection method was used to evaluate the importance of the remaining features. This approach has been widely used to analyze complicated systems.

The mRMR method evaluates the importance of target features by using max-relevance and min-redundancy. Features with great relevance to the category labels and low redundancy with other features are considered to be influential. It uses mutual information (MI) to measure relevance and redundancy. The score of the MI between two variables X and Y is calculated as

$$I(X,Y) = \int\int p(x,y) log \frac{p(x,y)}{p(x)p(y)} dxdy, \quad (1)$$

where $p(x,y)$ represents the joint probability distribution function of X and Y, and $p(x)$ and $p(y)$ represent the marginal probability distribution function of X and Y, respectively. The mRMR constructs order feature lists on the basis of the importance of features. Specifically, the program loops several times, and each loop selects a feature that has the greatest correlation with the target variable and the least correlation with the selected features. Finally, a list of sorting features is obtained in accordance with the selected orders.

The mRMR program used in this research was obtained from http://penglab.janelia.org/proj/mRMR/and executed with default parameters.

Incremental Feature Selection

Even though the mRMR method produces a ranked list of features on the basis of the importance of features, we still cannot determine the influential features. The IFS (Liu and Setiono, 1998) method can determine the optimal number of key features in combination with one classification algorithm. First, IFS generates a series of feature subsets from the above feature list in accordance with the step size. For example, if the step size is 10, the first feature subset will be the top 10 features, and the second subset will be the top 20 features. Next, one classifier is built based on the training set, where the samples are represented by features from each feature subset. The classifiers are evaluated by 10-fold cross-validation (Kohavi, 1995) to obtain evaluation metrics. Finally, the optimal feature subset and the best classifier are determined, and features in this subset are called optimum features.

Decision Tree

DT (Safavian and Landgrebe, 1991) is a model that presents decision rules and classification results in a tree-like structure and is widely used in the biological and biomedical fields. DT is a supervised learning approach that builds a model based on the IF–THEN format. It achieves superior model performance through low computational complexity. The common decision trees are Iterative Dichotomizer 3, C4.5, and Classification and Regression Tree. They use different partition strategies when building a prediction model. In this study, we used the Scikit-learn (Pedregosa et al., 2011) module in Python to construct a DT classifier.

Support Vector Machine

SVM (Cortes and Vapnik, 1995; Zhou J.-P. et al., 2020; Wang et al., 2021) is a supervised learning algorithm in statistical

learning methods that is commonly used in classification and regression problems. SVM maps the data from a low-dimensional space to a high-dimensional space by using a kernel function. Then, a hyperplane with the maximum interval existing in the high-dimensional space makes two classes of samples linearly separable.

In this study, 20 tumor types needed to be classified. This task was a multiclass classification problem. Therefore, we applied the one-versus-rest strategy to train a multiclass SVM, which was split into numerous binary SVMs. For each binary SVM, samples of one class were regarded as positive examples, and samples of all other classes were used as negative examples. We directly used the tool "SMO" in Weka software (Gewehr et al., 2007) in this study. The sequential minimum optimization algorithm (Platt, 1998; Keerthi et al., 2001) was utilized to optimize the training procedure. The kernel function was set as a polynomial function.

Synthetic Minority Oversampling Technique

As can be seen from **Table 1**, the sample sizes for all tumor types were quite different. For example, types "lung" and "hematopoietic and lymphoid tissue" contained 188 and 173 samples, respectively, whereas types "autonomic ganglion" and "bone" had only 16 and 20 samples, respectively. These results indicated that the whole dataset of this study was unbalanced. Accordingly, we adopted the synthetic minority oversampling technique (SMOTE) (Chawla et al., 2002) to balance the dataset when building classifiers. This method uses the k-nearest neighbor algorithm to expand the sample sizes of each minority class. SMOTE first selects random data from one minority class and then finds the k-nearest neighbors in this class. Next, the new sample data are synthesized between the random data and the randomly generated k-nearest neighbor. After SMOTE processing, the sample size of each minority class is equal to that of the majority class. In other words, the sample sizes of the 20 tumor types in this study were equal. In this study, we oversampled data by using the tool "SMOTE" in Weka software (Gewehr et al., 2007).

Performance Measurement

In this study, several multiclass classifiers were used to distinguish samples from 20 tumor types. We adopted 10-fold cross-validation (Kohavi, 1995; Chen et al., 2017; Zhao et al., 2018; Zhou JP. et al., 2020; Jia et al., 2020; Liang et al., 2020; Zhang et al., 2021c; Yang and Chen, 2021; Zhu et al., 2021) to evaluate the performance of each multiclass classifier. We correlation coefficient (MCC) (Matthews, 1975; Gorodkin, 2004; Liu H. et al., 2021; Zhang et al., 2021a; Zhang et al., 2021b; Pan et al., 2021) to measure and evaluate the prediction quality of the results of 10-fold cross-validation. Let X be a matrix representing predicted labels yielded by one classifier and Y be another matrix indicating the actual labels of samples. The calculation formula of MCC is as follows:

$$MCC = \frac{cov(X,Y)}{\sqrt{cov(X,X)cov(Y,Y)}} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{C}\left(x_{ij}-\bar{x}_j\right)\left(y_{ij}-\bar{y}_j\right)}{\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{C}\left(x_{ij}-\bar{x}_j\right)^2\sum_{i=1}^{n}\sum_{j=1}^{C}\left(y_{ij}-\bar{y}_j\right)^2}}, \quad (2)$$

where $cov(X,Y)$ denotes the correlation coefficient of X and Y, and $\bar{x}_j$ and $\bar{y}_j$ are the average values in the j th column of X and Y, respectively. In addition, C denotes the number of tumor types, and n denotes the total number of samples.

In addition to MCC, we also calculated accuracy on each cancer type and overall accuracy. The accuracy on the ith cancer type was computed by

$$Accuracy_i = \frac{n_i}{N_i} \qquad i = 1, 2, \cdots, 20, \quad (3)$$

where N_i represents the number of samples in the ith cancer type and n_i denotes correctly predicted samples in the ith cancer type. The overall accuracy was calculated by

$$\text{Overall accuracy} = \frac{\sum_{i=1}^{20} n_i}{\sum_{i=1}^{20} N_i} \qquad i = 1, 2, \cdots, 20, \quad (4)$$

MCC was set as the key measurement and others were also provided for reference in this study.

RESULTS

In this study, several computational methods were used to analyze the CCLE dataset of 20 tumor types. The analysis process is shown in **Figure 1**.

Results of the Boruta and max-Relevance and Min-Redundancy Methods

All features were first analyzed by using the Boruta method. A total of 54,634 features were removed, and 3,186 features were retained. These retained features are provided in **Supplementary Table S1**. These 3,186 features were further analyzed by using the mRMR method, and a feature ranking list was generated on the basis of their importance. This list can also be found in **Supplementary Table S1**.

Results of the Incremental Feature Selection Method

The feature list produced by the mRMR method was fed into the IFS method. A series of feature subsets were generated by setting the step size to 10. The DT and SVM were used to build classifiers on each feature subset. Then, all classifiers were evaluated by 10-fold cross-validation to obtain evaluation metrics, such as accuracy on each cancer type, overall accuracy and MCC. The above measurements that were acquired by the two classification algorithms for each subset of features are shown in **Supplementary Table S2**. We plotted the IFS curve to visualize the results. For SVM, the MCC was set as the Y-axis,

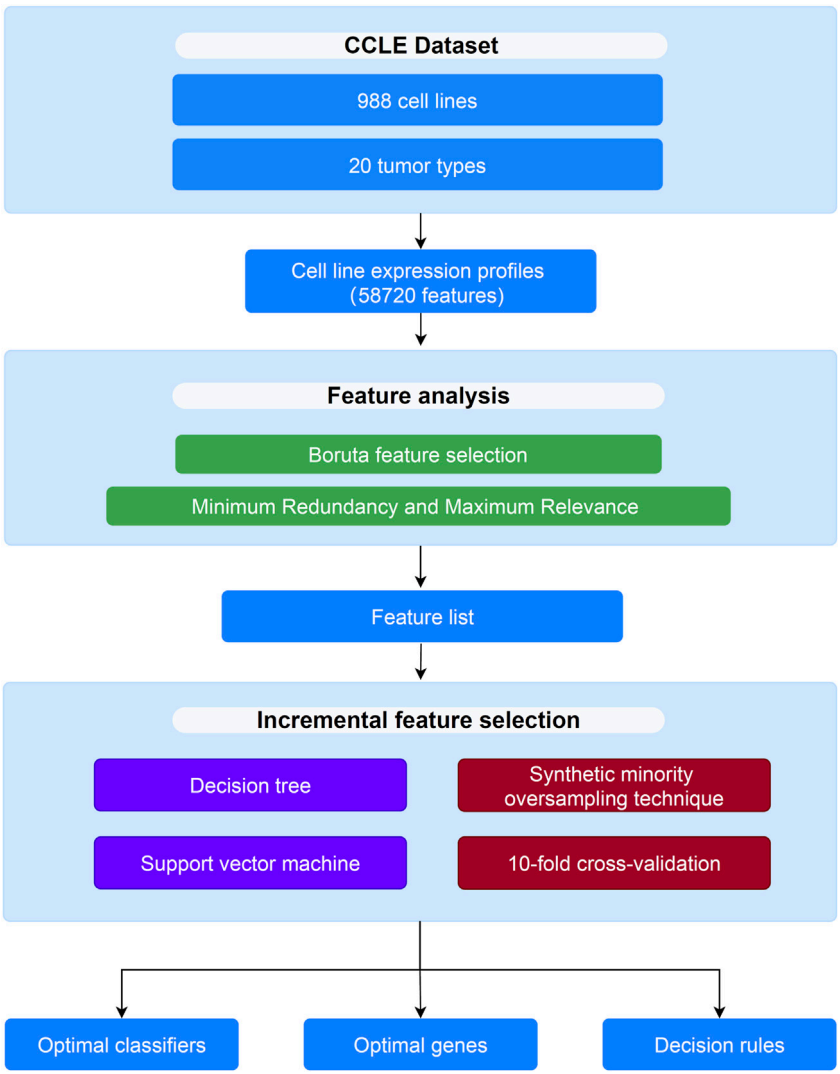


FIGURE 1 | Flow chart to show the entire analysis procedures. The CCLE dataset which includes 988 cell lines and 20 tumor types is analyzed by Boruta and mRMR methods, resulting in a feature list. The list is then fed into the incremental feature selection method to extract optimal genes, build the optimal classifiers and construct decision rules.

and the number of features was set as the X-axis. As shown in **Figure 2**, when the number of features reached 3,130, the highest value of MCC was 0.976. The corresponding overall accuracy was 0.978 (**Table 2**). Accordingly, the best SVM classifier can be built based on these top 3,130 features. Although this classifier provided the highest performance, its efficiency was not very high because an excessively high number of features were used. The IFS results of SVM were carefully checked (**Supplementary Table S2** and **Figure 2**). When the top 400 features were adopted, the MCC reached 0.951, which was only slightly lower than the highest MCC. The overall accuracy was 0.954 (**Table 2**). It was also a little lower than that of the best SVM classifier. It can be concluded that these two SVM classifiers provided almost equal performance. To further confirm this fact, we also investigated the accuracies on 20 cancer types yielded by these two classifiers. A radar graph was plotted, as shown in **Figure 3**. Clearly, the areas inside the curves of two classifiers were almost same, suggesting the equal performance of these two classifiers. However, the number of features was considerably lower. The SVM classifier with these features had drastically higher efficiency than the best SVM classifier. Thus, this classifier could be the proposed classifier for assigning samples to the correct cancer type.

In addition to the above SVM algorithm, we used the DT, which is a white-box classification algorithm. In this process, the step size of IFS with the DT was also set to 10, and only the top 400 features in the mRMR list were considered. The IFS results are also available in **Supplementary Table S2**, and the IFS curve is presented in **Figure 4**. The best DT classifier yielded an MCC value of 0.754, which was based on the top 390 features. The

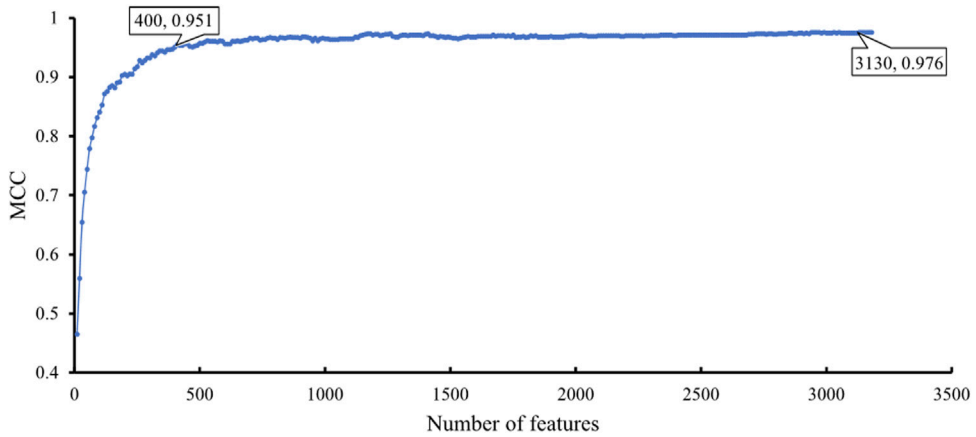


FIGURE 2 | IFS curve with SVM classification algorithm on the different number of features. The SVM provides the highest MCC of 0.976 when the top 3,130 features are adopted. When top 400 features are adopted, SVM provides good performance with MCC of 0.951.

TABLE 2 | Performance of some key classifiers.

Classification algorithm	Number of features	Overall accuracy	MCC
Support vector machine	3,130	0.978	0.976
	400	0.954	0.951
Decision tree	390	0.771	0.754
	100	0.757	0.739

overall accuracy of this classifier was 0.771, as listed in **Table 2**. Likewise, we also wanted to obtain an accepted classifier that used few features and provided high performance. As can be seen from **Figure 4**, the MCC reached 0.739 when the top 100 features were used. The overall accuracy was 0.757 (**Table 2**). They were only a little lower than those of the best DT classifier. Furthermore, the accuracies on 20 cancer types of these two classifiers were also investigated, as illustrated in **Figure 3**. Evidently, these two DT classifiers were almost at the same level. Therefore, these top 100 features were considered to build the DT classifier.

Classification Rules

As mentioned above, the DT classifier with the top 100 features exhibited high performance. Thus, we constructed a DT with these features and all samples. Consequently, we obtained 275 rules, which are presented in **Supplementary Table S3**. The number of decision rules and criteria used for 20 tumor types are shown in **Table 1**. Each cancer type was assigned some decision rules. The cancer type "Lung" was assigned most decision rules, whereas 'Fibroblast' received least rules. The further analysis of these rules can be found in *Analysis of Decision Rules*.

GO and KEGG Enrichment Analysis

As mentioned in *Results of the Incremental Feature Selection Method*, the SVM classifiers with top 400 features gave a little lower performance than the best SVM classifier. However, it had much higher efficiency because much less features were used in this classifier. Thus, these 400 features may be highly related to distinguish different cancer types. Thus, we conducted GO and KEGG enrichment analysis on these features (genes). The results can be found in **Supplementary Table S4**. Some top GO terms and KEGG pathways are illustrated in **Figures 5**, **6**. In *Analysis of Essential Genes*, the discussion on the enrichment analysis results would be given.

DISCUSSION

In this study, we used the Boruta and mRMR methods to analyze features and applied the IFS method combined with SVM and DT to construct classifiers and decision rules. Some essential features (genes) (see **Supplementary Table S1**) were extracted. Furthermore, we obtained several decision rules. In this section, we provide an extensive analysis of these essential genes and decision rules.

Analysis of Essential Genes

Firstly, we performed GO/KEGG enrichment analysis to find whether our 400 selected features were significantly enriched in specific terms. Results were described in *GO and KEGG Enrichment Analysis*, top GO terms and KEGG pathways are illustrated in **Figures 5**, **6**.

We found that the significantly enriched GO terms mainly involve actin organization, cell matrix components and cell polarization. Actin assembly is very important for cell migration, and abnormal regulation of cell migration can drive cancer invasion and metastasis (Yamaguchi and Condeelis, 2007). Different cancers and different differentiation states of cancers often show different patterns of cell migration and the migration of these cancer cells is regulated by various signals. Although it is difficult to use a single strategy to regulate the motility of all cancer cells, inhibiting actin polymerization can inhibit migration of most types of cancer cells (Yamazaki et al.,

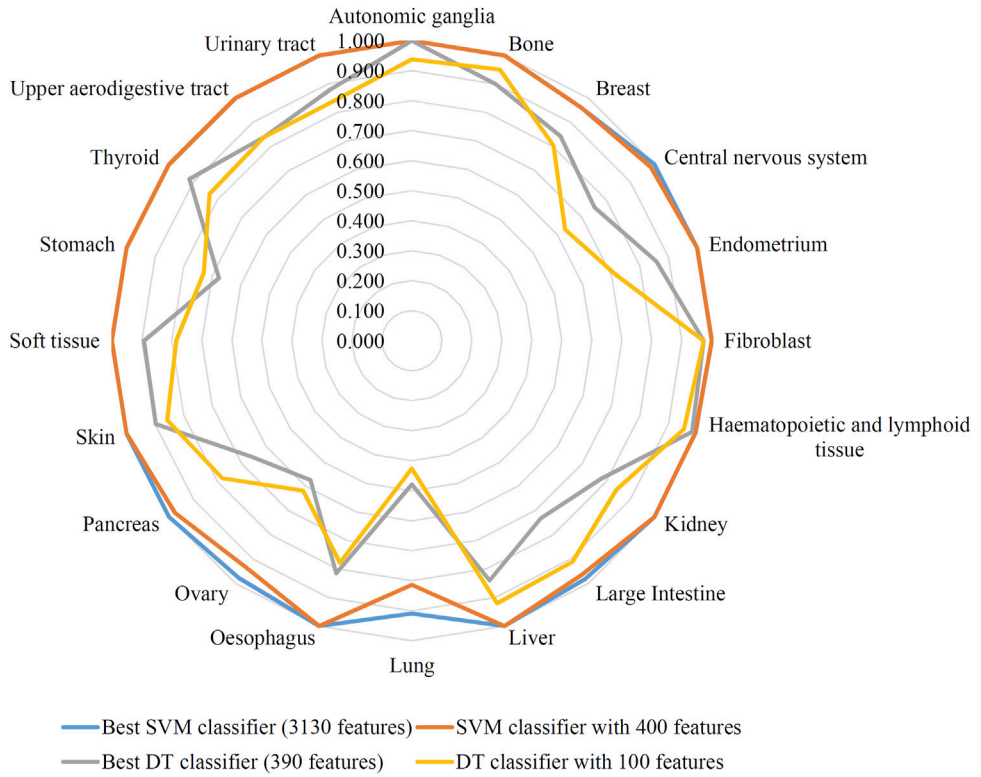


FIGURE 3 | Radar graph to show the performance of two support vector machine (SVM) classifiers and two decision tree (DT) classifiers on 20 cancer types. Two SVM classifiers provide almost equal performance, also for two DT classifiers.

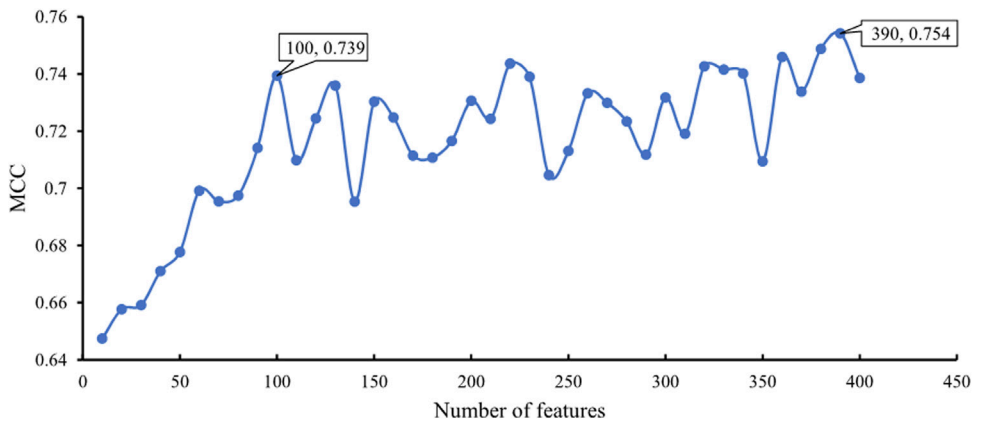


FIGURE 4 | IFS curve with DT classification algorithm on the different number of features. The DT provides the highest MCC of 0.754 when the top 390 features are adopted. DT yields high performance with MCC of 0.739 when only 100 features are used.

2005). The loss of cell polarity has been shown to be related to tumor progression (Wodarz and Näthke, 2007). Generally, aggressive tumors lack polarity, and study have shown that different cancers have different abnormal expression or localization of polar proteins, which may also serve as the basis for our classifier (Ellenbroek et al., 2012). The KEGG results also showed similar results, which are mainly related to migration and actin cytoskeleton. This reflects both the importance of cell migration ability to tumors and the difference in invasion of different tumors.

Secondly, among the 400 selected features (genes), the top-ranked genes were usually highly decisive for distinguishing different cell lines. Therefore, some of them were selected for analysis, which are listed in **Table 3**.

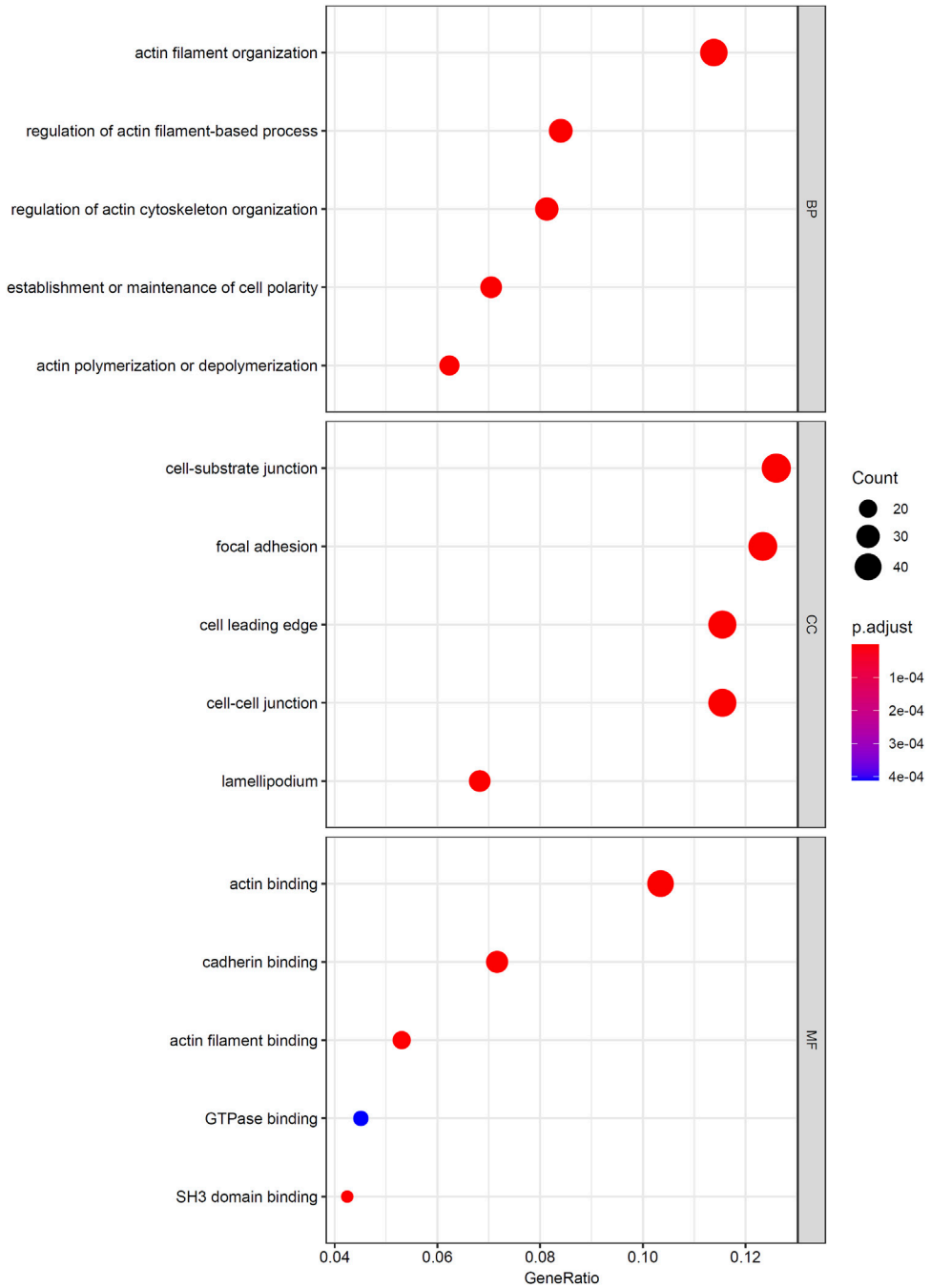


FIGURE 5 | Top enriched GO terms to the top 400 genes in the mRMR feature list.

The highest-ranking feature is NCKAP1 (ENSG00000061676). It encodes the NCK-associated protein 1 as a part of the WAVE (WASF) complex that regulates lamellipodia formation. Past studies have revealed that NCKAP1 is associated with multiple types of human cancer. A previous study showed that the WASF3 gene is a promoter of cell invasion in breast cancer and the Nckap1 can keep WASF3 in an inactive conformation through binding to the WASF homology domain at the N-terminus. The activation of WASF3 depends on the combination with RAC1 which can be prevented by the absence of NCKAP1. Thus, the downregulation of NCKAP1 inhibits the activity of WASF3 and may suppresses metastasis in breast cancer cells. In addition, univariate survival analysis have found that high expression level of NCKAP1 is correlated with short overall survival (Teng et al., 2016). The function of NCKAP1 in liver cancer has recently been clarified. Specifically,

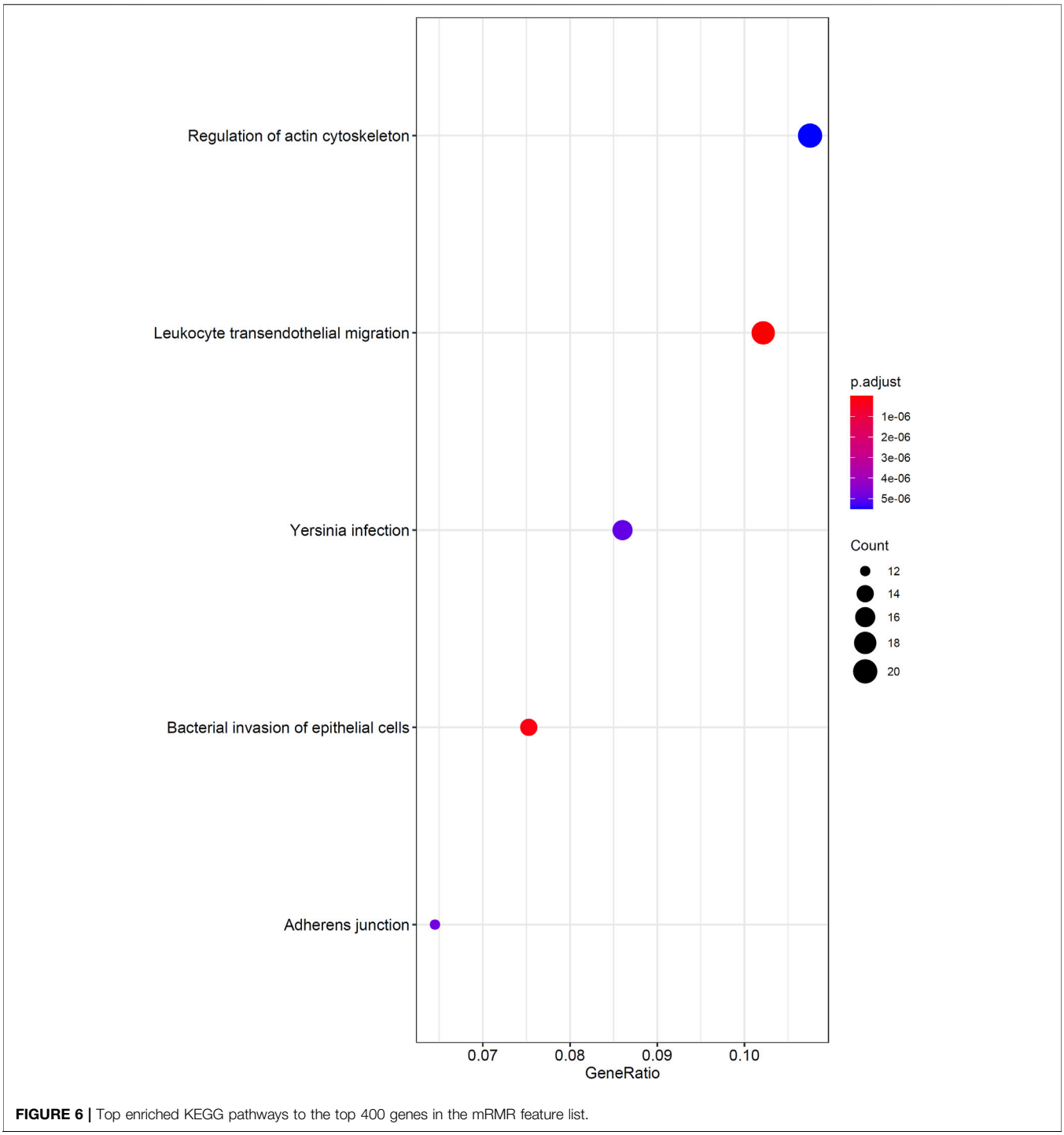


FIGURE 6 | Top enriched KEGG pathways to the top 400 genes in the mRMR feature list.

NCKAP1 can control tumor growth and improve prognosis by enhancing Rb1/p53 activation in hepatocellular carcinoma (HCC) (Zhong et al., 2019). Similarly, a recent study discovered that NCKAP1 is highly expressed in primary non-small-cell lung cancer (NSCLC) and is significantly associated with histologic tumor grade, metastasis, and poor survival rate. It is also related to the HSP90-mediated invasion and metastasis of NSCLC by stimulating MMP9 activation and the epithelial–mesenchymal transition (EMT) (Xiong et al., 2019). In conclusion, NCKAP1 is aberrantly expressed in a variety of cancer types and could be a biomarker and potential therapeutic target.

TNFRSF12A (ENSG00000006327) encodes the receptor of TNFSF12/TWEAK, which is also known as fibroblast growth factor-inducible molecule 14. It can promote endothelial cell proliferation and angiogenesis. Studies have demonstrated that

TABLE 3 | Information of essential genes.

Ensembl ID	Gene symbol	Description
ENSG00000061676	NCKAP1	NCK Associated Protein 1
ENSG00000006327	TNFRSF12A	TNF Receptor Superfamily Member 12A
ENSG00000172037	LAMB2	Laminin Subunit Beta 2
ENSG00000122642	FKBP9	FKBP Prolyl Isomerase 9
ENSG00000070087	PFN2	Profilin 2
ENSG00000141198	TOM1L1	Target Of Myb1 Like 1 Membrane Trafficking Protein

TNFRSF12A is highly expressed in breast cancer, and a high TNFRSF12A level associated with matrix metalloproteinase (MMP)-9 overexpression is related to cancer progression; thus, TNFRSF12A-targeting therapy could improve survival rates in cancer (Yang et al., 2018). Furthermore, through modulating the expression of MMP-9, the overexpression of TNFRSF12A can promote prostate cancer progression and result in poor treatment outcomes (Huang et al., 2011). TNFRSF12A is also highly expressed in human HCC, and *in vivo* experiments have revealed that TNFRS12A knockdown can inhibit cancer cell proliferation and migration (Wang et al., 2017). TNFRSF12A has also been demonstrated to be highly expressed in NSCLC and contribute to NSCLC cell migration and invasion *in vitro* (Whitsett et al., 2012). Other studies have also confirmed that TNFRSF12A is overexpressed in melanomas, gliomas, and esophageal and pancreatic cancers (Han et al., 2005; Tran et al., 2006; Watts et al., 2007; Zhou et al., 2013). Interestingly, in certain tumor types, TNFRSF12A exhibits a low expression level. A study on TCGA data suggested that the downregulation of TNFRSF12A in thyroid cancer could be a potential molecular biomarker for the prediction of poor prognosis (Wu et al., 2020). Therefore, TNFRSF12A has different expression patterns in different cancers and could be a remarkable feature for distinguishing different cancer cell lines. In addition, it could also be a critical therapeutic target, and preclinical studies have shown that the use of inhibitors in cancer with high TNFRSF12A expression has certain effects (Wajant, 2013).

LAMB2 (ENSG00000172037) encodes a subunit of laminins, which are one of the major glycoproteins present in the basement membrane of the extracellular matrix and are related to tumor angiogenesis, invasion, and metastasis. A previous study revealed that the downregulation of LAMB2 caused by HE4 gene interference results in the invasion and metastasis of ovarian cancer cells (Zhuang et al., 2014). Studies on pancreatic cancer have demonstrated that the lack of basement membrane continuity, which is determined by limited laminin expression, is associated with poor postoperative outcomes. In other words, in pancreatic cancer, the downregulation of LAMB2 is correlated with poor prognosis (Van Der Zee et al., 2012).

FKBP9 (ENSG00000122642), which encodes FKBP prolyl isomerase 9, is known to be associated with chaperonin-mediated protein folding and protein metabolism. A recent study has found that FKBP9 could be an independent prognostic marker for predicting the poor prognosis of patients with prostate cancer; that high FKBP9 levels and short biochemical-recurrence-free survival are significantly correlated (p = 0.041); and that FKBP9 may be a cancer promoter that enhances prostate cancer progression (Jiang et al., 2020). Another study found that FKBP9 is a critical factor for promoting the malignant behaviors of glioblastoma cells; high FKBP9 level is related to poor prognosis and could confer malignant cells with the capability to resist endoplasmic reticulum stress inducers (Xu et al., 2020). Other studies have also confirmed that FKBP9 is connected with other cancers, such as colorectal and breast cancers (Bianchini et al., 2006; Chang et al., 2020). Thus, FKBP9 may be an effective feature of many cancer cell lines.

PFN2 (ENSG00000070087) encodes an actin monomer-binding protein. It participates in regulating actin aggregation in response to extracellular signals and cell motility. Recently, PFN2 has emerged as a key regulator of cancer development and progression. PFN2 has been reported to be highly expressed in triple-negative breast cancer (TNBC); it could promote the proliferation, migration, and invasion of TNBC cells and may be partially responsible for the worsened survival associated with high PFN2 levels (Ling et al., 2021). In esophageal squamous cell carcinoma, a high PFN2 level is related to short overall survival. Moreover, PFN2 expression is positively associated with tumor invasion depth and lymph node metastasis (Cui et al., 2016). Another study demonstrated that PFN2 is highly expressed in head and neck squamous cell cancer (HNSCC) tissues and cell lines and that the activation of the PI3K/Akt/β-catenin signaling pathway by PFN2 results in the proliferation and metastasis promotion of HNSCC, whereas PFN2 knockdown produces the opposite effects (Zhou et al., 2019). However, another study suggested a different result: the degree of tumor metastasis is negatively associated with PFN2 expression level likely because of the enhancement in EMT induced by low PFN2 levels considering that enhanced EMT may increase migratory capabilities (Zhang H. et al., 2018). Other studies have also found that PFN2 has different expression patterns and effects in NSCLC, small cell lung cancer, and gastric cancer (Hippo et al., 2002; Yan et al., 2017; Cao et al., 2020). In conclusion, PFN2 plays an important role in a variety of cancers and could be an important biomarker for different cancer cells, as well as an attractive therapeutic target.

TOM1L1 (ENSG00000141198) encodes the target of myb1-like 1 membrane trafficking protein. ERBB2-induced breast cancer cell invasion has been documented to be caused by the TOM1L1-derived membrane delivery of MT1-MMP, and ERBB2 and TOM1L1 are frequently coamplified in the breast (Chevalier

et al., 2016). Other studies have also found that TOM1L1 is related to colorectal cancer and is highly expressed in bladder cancer (Emaduddin et al., 2008; Zhang Y. et al., 2018).

As analyzed above, the selected genes from our results showed strong expression differences in multiple cancer cells. These genes could be good therapeutic targets. By the same token, distinct gene expression patterns could also be remarkably decisive features for different cancer cell lines.

Analysis of Decision Rules

Previously, we constructed 275 decision rules on the basis of the top 100 selected features and all cell lines. Each rule contained several criteria. The numbers of rules and criteria for each cancer type are listed in **Table 1**. In addition, the number of genes involving rules for each cancer type is also listed in this table. In the following, we provide our interpretation and experimental evidence for some rules based on published literature. These evidences indicate the effects of the high/low expression of key genes on tumors which also found to have similar expression patterns in the decision rules of the corresponding tumor cell line (relatively high/low expression level compared to other cell lines).

The 23 rules for identifying breast cancer cell lines included 325 criteria, which involved 62 genes. These genes have considerable experimental support, and here we show some evidence. LDHB (ENSG00000111716) encodes the B subunit of lactate dehydrogenase enzyme, which participates in glycolysis. A study found that LDHB is specifically upregulated in basal-like TNBC, and the loss of LDBH arrests tumor growth *in vivo* (Cui et al., 2015). One study discovered that PAX8 (ENSG00000125618) is the best discriminatory marker between ovarian and breast carcinomas (Nonaka et al., 2008). The same study also reported that PAX8 is negatively expressed in serous carcinoma but is positively expressed in breast carcinomas. This expression pattern is in agreement with our decision rules. RAB34 (ENSG00000109113) regulates the spatial distribution of lysosomes, secretion, and micropinocytosis and is expressed at high levels in breast cancer cell lines. A recent study has found that RAB34 is overexpressed in breast cancer and that the high expression of RAB34 is closely linked to breast cancer cell adhesion, migration, and invasion (Sun et al., 2018).

Among the 275 rules, 51 could identify lung cancer cell lines with 740 criteria involving 80 genes. Here, we provide clear experimental evidence for some genes that are well established in the literature. SOX10 (ENSG00000100146) is a transcription factor that encodes genes involved in the regulation of embryonic development and cell-fate decisions. Studies have demonstrated that SOX10 is usually overexpressed in multiple cancers. It can activate stem/progenitor cells through the Wnt/β-catenin signaling pathway and induces mesenchymal transformation expression (Zhou et al., 2014; Miettinen et al., 2015). However, it appeared in all the lung cancer decision rules with a low expression. A recent experimental study on 1085 NSCLC tumor tissue samples has given direct support for our results. A microarray analysis study revealed that SOX10 is negatively expressed in NSCLC, with only 5 (<1%) cases showing positive results (Kriegsmann et al., 2018). ARHGAP30 (ENSG00000186517) encodes Rho GTPase-activating protein 30, which plays an important role in cell adhesion and cytoskeleton organization regulation. It is downregulated in lung cancer cell lines. Moreover, a low ARHGAP30 level is associated with the activation of Wnt/β-catenin signaling pathways and further leads to lung cancer cell proliferation, migration, and invasion (Mao and Tong, 2018). CTDSPL/RBSP3 (ENSG00000144677) was also downregulated in our rules. It has been reported to be a tumor-suppressor gene in multiple cancers (Kashuba et al., 2009) and to be downregulated in lung cancer (Senchenko et al., 2010). TSPAN4 (ENSG00000214063) was highly expressed in our rules. The transcriptional product of TSPAN4 is a circular RNA that is upregulated in lung adenocarcinoma; circ-TSPAN4 can promote metastasis by increasing the expression of ZEB1 (Ying et al., 2019). In our rules, S100A13 (ENSG00000189171) was required to be relatively highly expressed. As has been seen in another study, S100A13 is overexpressed in NSCLC, especially in the advanced stage. High S100A13 level is strongly associated with tumor angiogenesis and poor prognosis (Miao et al., 2018).

Liver cancer cell lines had nine decision rules containing 115 criteria. These criteria involved 42 genes. The expression patterns of many genes in these rules have been confirmed in several other studies. A1BG-AS1 (ENSG00000268895) is a RNA gene, and its transcriptional product is a lncRNA. A study found that A1BG-AS1 inhibits HCC cell proliferation, migration, and invasion *in vitro*. Clinical association analysis revealed that A1BG-AS1 is downregulated in HCC, and low A1BG-AS1 level is also associated with advanced tumor stage, microvascular invasion, and high tumor grade (Bai et al., 2019). CMTM4 (ENSG00000183723) also showed a low expression level in our rules. As found in other studies, CMTM4 plays a tumor-suppressor role in HCC, wherein it inhibits tumor activities by regulating cell growth and cell cycle (Bei et al., 2017). Thus, consistent with our results, CMTM4 showed negative expression in HCC. AKAP1 (ENSG00000121057) encodes A-kinase anchoring protein 1 and plays an important role in the regulation of mitochondrial function and oxidative metabolism. A previous study identified that AKAP1 is overexpressed in HHC; this expression pattern also provides supporting evidence for our decision rules. AKAP1 may contribute to tumor progression and result in poor overall and disease-free survival rates in patients with HCC (Yu et al., 2018).

The identification of ovarian cancer cell lines had 25 rules with 306 criteria. These criteria involved 67 genes. The validity of our results was supported by other studies that have confirmed some of these genes. As mentioned in the rules for breast cancer cell lines, PAX8 is highly expressed in ovarian cancer and could be a remarkable feature for discriminating between breast cancer and ovarian cancer (Nonaka et al., 2008). In addition, another study found that the knockdown of PAX8 significantly reduces cancer cell proliferation, migration, and invasion (Di Palma et al., 2014). GNAI2 (ENSG00000114353) encodes heterotrimeric G protein, which plays a direct role in regulating the cAMP response element-binding protein. In agreement with our findings, the results of the direct sequencing and qPCR analysis of 589 human ovarian cancer revealed that 85.9% (506) of patients

have decreased GNAI2 messaging (Raymond et al., 2014). SRPX (ENSG00000101955) is reported to be a tumor-suppressor gene and is downregulated in multiple cancer cells and tissues (Tambe et al., 2016). This result is consistent with our decision rules for endometrial, pancreatic, and urinary tract cancers. However, one difference is worth noting: we found that SRPX was overexpressed in most rules for ovarian cancer. The overexpression of SRPX has been affirmed by a recent study based on clinical specimens, wherein the upregulation of SRPX is associated with tumor invasion and migration activity in ovarian cancer (Liu et al., 2019).

At the same time, we noted the exclusive genes for some cancer cell line may be quite important. For example, CD276 (ENSG00000103855) only been shown in the rules of lung cancer cell lines. CD276 (B7-H3) encode a member of the immunoglobulin superfamily and is an important immune checkpoint member of the B7/CD28 families. It is induced by antigen presenting cells and participates in the regulation of T cell-mediated immune response (Picarda et al., 2016). Studies have found that CD276 is associated with Mycoplasma pneumoniae pneumonia. It is up-regulated in patients' plasma and may be involved in the progression of pneumonia by increasing the concentration of TNF-α and the activation of neutrophils (Chen et al., 2013). At the same time, CD276 is also abnormally expressed in a variety of tumors and participates in tumor proliferation, apoptosis, differentiation, invasion and interepithelial transformation. Usually CD276 is up-regulated in tumors and is associated with poor prognosis of patients (Liu S. et al., 2021). In NSCLC, a previous meta-analysis found that the high expression of CD276 was significantly associated with patients' lymph node metastasis and advanced TNM staging (Wu et al., 2016). Other studies have also found that the expression of CD276 is related to the smoking history and pathological types of patients. Usually, the expression of CD276 in patients with lung adenocarcinoma or smoking history is associated with a shorter overall survival (Inamura et al., 2017; Zhang and Hao, 2019). Although CD276 is highly expressed in a variety of tumors, and its molecular mechanism to promote cancer progression is not clear, our results show that it may be more important for lung cancer. At the same time, other study also found that CD276 was up-regulated in tumor cells of lung cancer patients treated with trametinib, which can achieve better therapeutic effect after combined B7-H3 × T cell bispecific antibody treatment, and this also proves that CD276 is a potential therapeutic target for lung cancer.

We provided pieces of evidence for some decisive genes in the decision rules for four classes of cancer cell lines in the preceding discussion. Although these genes also have different expression patterns in other cancer cell lines and a large number of remaining genes have not yet been explained in detail, we can confirm the reliability of our results from the substantial evidence that we presented. Notably, some distinctive and decisive genes in our rules have not previously been investigated by other researchers. These genes may give new insight into tumor growth and progression, as well as novel potential therapeutic targets.

CONCLUSION

This study gave a computational investigation on the cell line gene expression data of cancer cell lines. Several machine learning algorithms were applied on such data. On one hand, we constructed efficient classifiers, which can be latent tools to identify different cancer types. On the other hand, a new set of potential biomarkers (*NCKAP1*, *TNFRSF12A*, *LAMB2*, *FKBP9*, *PFN2*, *TOM1L1*) and expression rules for the identification of different cancers at the transcriptome level were discovered. These biomarkers and rules can be useful materials to uncover mechanism underlying different cancer types, thereby improving our understanding on cancer.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://figshare.com/articles/dataset/scRNA-seq_Datasets/7174922?file=14847260.

AUTHOR CONTRIBUTIONS

LC, TH and Y-DC designed the study. SD, HL and Y-HZ performed the experiments. SD, Y-HZ, XZ, KF and ZL analyzed the results. HL and Y-HZ wrote the manuscript. All authors contributed to the research and reviewed the manuscript.

FUNDING

This work was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (XDB38050200, XDA26040304), National Key R&D Program of China (2018YFC0910403), the Fund of the Key Laboratory of Tissue Microenvironment and Tumor of Chinese Academy of Sciences (202002).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.781285/full#supplementary-material

REFERENCES

Andor, N., Graham, T. A., Jansen, M., Xia, L. C., Aktipis, C. A., Petritsch, C., et al. (2016). Pan-cancer Analysis of the Extent and Consequences of Intratumor Heterogeneity. *Nat. Med.* 22, 105–113. doi:10.1038/nm.3984

Bai, J., Yao, B., Wang, L., Sun, L., Chen, T., Liu, R., et al. (2019). lncRNA A1BG-AS1 Suppresses Proliferation and Invasion of Hepatocellular Carcinoma Cells by Targeting miR-216a-5p. *J. Cel. Biochem.* 120, 10310–10322. doi:10.1002/jcb.28315

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., et al. (2012). The Cancer Cell Line Encyclopedia Enables Predictive Modelling of Anticancer Drug Sensitivity. *Nature* 483, 603–607. doi:10.1038/nature11003

Bei, C., Zhang, Y., Wei, R., Zhu, X., Wang, Z., Zeng, W., et al. (2017). Clinical Significance of CMTM4 Expression in Hepatocellular Carcinoma. *Onco. Targets Ther.* 10, 5439–5443. doi:10.2147/ott.s149786

Bianchini, M., Levy, E., Zucchini, C., Pinski, V., Macagno, C., De Sanctis, P., et al. (2006). Comparative Study of Gene Expression by cDNA Microarray in Human Colorectal Cancer Tissues and normal Mucosa. *Int. J. Oncol.* 29, 83–94. doi:10.3892/ijo.29.1.83

Burrell, R. A., and Swanton, C. (2014). Tumour Heterogeneity and the Evolution of Polyclonal Drug Resistance. *Mol. Oncol.* 8, 1095–1111. doi:10.1016/j.molonc.2014.06.005

Cao, Q., Liu, Y., Wu, Y., Hu, C., Sun, L., Wang, J., et al. (2020). Profilin 2 Promotes Growth, Metastasis, and Angiogenesis of Small Cell Lung Cancer through Cancer-Derived Exosomes. *Aging* 12, 25981–25999. doi:10.18632/aging.202213

Chang, Y.-S., Chang, C.-M., Lin, C.-Y., Chao, D.-S., Huang, H.-Y., and Chang, J.-G. (2020). Pathway Mutations in Breast Cancer Using Whole-Exome Sequencing. *Oncol. Res.* 28, 107–116. doi:10.3727/096504019x15698362825407

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *jair* 16, 321–357. doi:10.1613/jair.953

Chen, Z.-R., Zhang, G.-B., Wang, Y.-Q., Yan, Y.-D., Zhou, W.-F., Zhu, C.-H., et al. (2013). Soluble B7-H3 Elevations in Hospitalized Children with Mycoplasma Pneumoniae Pneumonia. *Diagn. Microbiol. Infect. Dis.* 77, 362–366. doi:10.1016/j.diagmicrobio.2013.09.006

Chen, L., Wang, S., Zhang, Y.-H., Li, J., Xing, Z.-H., Yang, J., et al. (2017). Identify Key Sequence Features to Improve CRISPR sgRNA Efficacy. *IEEE Access* 5, 26582–26590. doi:10.1109/access.2017.2775703

Chevalier, C., Collin, G., Descamps, S., Touaitahuata, H., Simon, V., Reymond, N., et al. (2016). TOM1L1 Drives Membrane Delivery of MT1-MMP to Promote ERBB2-Induced Breast Cancer Cell Invasion. *Nat. Commun.* 7, 10765. doi:10.1038/ncomms10765

Chiu, H. S., Somvanshi, S., Patel, E., Chen, T. W., Singh, V. P., Zorman, B., et al. (2018). Pan-cancer Analysis of lncRNA Regulation Supports Their Targeting of Cancer Genes in Each Tumor Context. *Cell Rep.* 23, 297–312.e12. doi:10.1016/j.celrep.2018.03.064

Cortes, C., and Vapnik, V. (1995). Support-Vector Networks. *Mach Learn.* 20, 273–297. doi:10.1007/bf00994018

Cui, J., Quan, M., Jiang, W., Hu, H., Jiao, F., Li, N., et al. (2015). Suppressed Expression of LDHB Promotes Pancreatic Cancer Progression via Inducing Glycolytic Phenotype. *Med. Oncol.* 32, 143. doi:10.1007/s12032-015-0589-8

Cui, X.-B., Zhang, S.-M., Xu, Y.-X., Dang, H.-W., Liu, C.-X., Wang, L.-H., et al. (2016). PFN2, a Novel Marker of Unfavorable Prognosis, Is a Potential Therapeutic Target Involved in Esophageal Squamous Cell Carcinoma. *J. Transl Med.* 14, 137. doi:10.1186/s12967-016-0884-y

Dagogo-Jack, I., and Shaw, A. T. (2018). Tumour Heterogeneity and Resistance to Cancer Therapies. *Nat. Rev. Clin. Oncol.* 15, 81–94. doi:10.1038/nrclinonc.2017.166

Di Palma, T., Lucci, V., De Cristofaro, T., Filippone, M. G., and Zannini, M. (2014). A Role for PAX8 in the Tumorigenic Phenotype of Ovarian Cancer Cells. *BMC cancer* 14, 292. doi:10.1186/1471-2407-14-292

Ellenbroek, S. I., Iden, S., and Collard, J. G. (2012). Cell Polarity Proteins and Cancer. *Semin. Cancer Biol.* 22, 208–215. doi:10.1016/j.semcancer.2012.02.012

Emaduddin, M., Edelmann, M. J., Kessler, B. M., and Feller, S. M. (2008). Odin (ANKS1A) Is a Src Family Kinase Target in Colorectal Cancer Cells. *Cell Commun Signal* 6, 7. doi:10.1186/1478-811x-6-7

Garman, K. S., Nevins, J. R., and Potti, A. (2007). Genomic Strategies for Personalized Cancer Therapy. *Hum. Mol. Genet.* 16, R226–R232. doi:10.1093/hmg/ddm184

Gewehr, J. E., Szugat, M., and Zimmer, R. (2007). BioWeka Extending the Weka Framework for Bioinformatics. *Bioinformatics* 23, 651–653. doi:10.1093/bioinformatics/btl671

Ghandi, M., Huang, F. W., Jané-Valbuena, J., Kryukov, G. V., Lo, C. C., Mcdonald, E. R., 3rd, et al. (2019). Next-generation Characterization of the Cancer Cell Line Encyclopedia. *Nature* 569, 503–508. doi:10.1038/s41586-019-1186-3

Gorodkin, J. (2004). Comparing Two K-Category Assignments by a K-Category Correlation Coefficient. *Comput. Biol. Chem.* 28, 367–374. doi:10.1016/j.compbiolchem.2004.09.006

Han, H., Nagle, R., and Von Hoff, D. D. (2005). Overexpression of FN14/TWEAK Receptor in Pancreatic Cancer. *Cancer Res.* 65, 554–555.

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of Cancer: the Next Generation. *Cell* 144, 646–674. doi:10.1016/j.cell.2011.02.013

Hippo, Y., Taniguchi, H., Tsutsumi, S., Machida, N., Chong, J. M., Fukayama, M., et al. (2002). Global Gene Expression Analysis of Gastric Cancer by Oligonucleotide Microarrays. *Cancer Res.* 62, 233–240.

Huang, M., Narita, S., Tsuchiya, N., Ma, Z., Numakura, K., Obara, T., et al. (2011). Overexpression of Fn14 Promotes Androgen-independent Prostate Cancer Progression through MMP-9 and Correlates with Poor Treatment Outcome. *Carcinogenesis* 32, 1589–1596. doi:10.1093/carcin/bgr182

Hyo-eun, C. B., Ruddy, D. A., Radhakrishna, V. K., Caushi, J. X., Zhao, R., Hims, M. M., et al. (2015). Studying Clonal Dynamics in Response to Cancer Therapy Using High-Complexity Barcoding. *Nat. Med.* 21, 440–448. doi:10.1038/nm.3841

Inamura, K., Yokouchi, Y., Kobayashi, M., Sakakibara, R., Ninomiya, H., Subat, S., et al. (2017). Tumor B7-H3 (CD276) Expression and Smoking History in Relation to Lung Adenocarcinoma Prognosis. *Lung Cancer* 103, 44–51. doi:10.1016/j.lungcan.2016.11.013

Jia, Y., Zhao, R., and Chen, L. (2020). Similarity-Based Machine Learning Model for Predicting the Metabolic Pathways of Compounds. *IEEE Access* 8, 130687–130696. doi:10.1109/access.2020.3009439

Jiang, F.-N., Dai, L.-J., Yang, S.-B., Wu, Y.-D., Liang, Y.-X., Yin, X.-L., et al. (2020). Increasing of FKBP9 Can Predict Poor Prognosis in Patients with Prostate Cancer. *Pathol. - Res. Pract.* 216, 152732. doi:10.1016/j.prp.2019.152732

Kashuba, V. I., Pavlova, T. V., Grigorieva, E. V., Kutsenko, A., Yenamandra, S. P., Li, J., et al. (2009). High Mutability of the Tumor Suppressor Genes RASSF1 and RBSP3 (CTDSPL) in Cancer. *PloS one* 4, e5231. doi:10.1371/journal.pone.0005231

Keerthi, S. S., Shevade, S. K., Bhattacharyya, C., and Murthy, K. R. K. (2001). Improvements to Platt's SMO Algorithm for SVM Classifier Design. *Neural Comput.* 13, 637–649. doi:10.1162/089976601300014493

Kohavi, R. (1995). "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," in Proceedings of the 14th International Joint Conference on Artificial Intelligence, Montreal, Quebec, Canada (Lawrence Erlbaum Associates), 1137–1145.

Kriegsmann, M., Kriegsmann, K., Harms, A., Longuespée, R., Zgorzelski, C., Leichsenring, J., et al. (2018). Expression of HMB45, MelanA and SOX10 Is Rare in Non-small Cell Lung Cancer. *Diagn. Pathol.* 13, 68. doi:10.1186/s13000-018-0751-7

Kursa, M. B., and Rudnicki, W. R. (2010). Feature Selection with the Boruta Package. *J. Stat. Softw.* 36, 1–13. doi:10.18637/jss.v036.i11

Li, B., and Dewey, C. N. (2011). RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome. *BMC Bioinf.* 12, 323. doi:10.1186/1471-2105-12-323

Liang, H., Chen, L., Zhao, X., and Zhang, X. (2020). Prediction of Drug Side Effects with a Refined Negative Sample Selection Strategy. *Comput. Math. Methods Med.* 2020, 1573543. doi:10.1155/2020/1573543

Ling, Y., Cao, Q., Liu, Y., Zhao, J., Zhao, Y., Li, K., et al. (2021). Profilin 2 (PFN2) Promotes the Proliferation, Migration, Invasion and Epithelial-To-Mesenchymal Transition of Triple Negative Breast Cancer Cells. *Breast Cancer* 28, 368–378. doi:10.1007/s12282-020-01169-x

Liu, H., and Setiono, R. (1998). Incremental Feature Selection. *Appl. Intell.* 9, 217–230. doi:10.1023/a:1008363719778

Liu, C. L., Pan, H. W., Torng, P. L., Fan, M. H., and Mao, T. L. (2019). SRPX and HMCN1 Regulate Cancer-Associated Fibroblasts to Promote the Invasiveness of Ovarian Carcinoma. *Oncol. Rep.* 42, 2706–2715. doi:10.3892/or.2019.7379

Liu, H., Hu, B., Chen, L., and Lu, L. (2021a). Identifying Protein Subcellular Location with Embedding Features Learned from Networks. *Curr. Proteomics* 18, 646–660. doi:10.2174/1570164617999201124142950

Liu, S., Liang, J., Liu, Z., Zhang, C., Wang, Y., Watson, A. H., et al. (2021b). The Role of CD276 in Cancers. *Front. Oncol.* 11, 847. doi:10.3389/fonc.2021.654684

Mao, X., and Tong, J. (2018). ArhgaP30 Suppressed Lung Cancer Cell Proliferation, Migration, and Invasion through Inhibition of the Wnt/β-Catenin Signaling Pathway. *Onco Targets Ther.* 11, 7447–7457. doi:10.2147/ott.s175255

Matthews, B. W. (1975). Comparison of the Predicted and Observed Secondary Structure of T4 Phage Lysozyme. *Biochim. Biophys. Acta (Bba) - Protein Struct.* 405, 442–451. doi:10.1016/0005-2795(75)90109-9

Miao, S., Qiu, T., Zhao, Y., Wang, H., Sun, X., Wang, Y., et al. (2018). Overexpression of S100A13 Protein Is Associated with Tumor Angiogenesis and Poor Survival in Patients with Early-Stage Non-Small Cell Lung Cancer. *Thorac. Cancer* 9, 1136–1144. doi:10.1111/1759-7714.12797

Miettinen, M., Mccue, P. A., Sarlomo-Rikala, M., Biernat, W., Czapiewski, P., Kopczynski, J., et al. (2015). Sox10-A Marker for Not Only Schwannian and Melanocytic Neoplasms but Also Myoepithelial Cell Tumors of Soft Tissue. *Am. J. Surg. Pathol.* 39, 826–835. doi:10.1097/pas.0000000000000398

Nonaka, D., Chiriboga, L., and Soslow, R. A. (2008). Expression of Pax8 as a Useful Marker in Distinguishing Ovarian Carcinomas from Mammary Carcinomas. *Am. J. Surg. Pathol.* 32, 1566–1571. doi:10.1097/pas.0b013e31816d71ad

Pan, X., Li, H., Zeng, T., Li, Z., Chen, L., Huang, T., et al. (2021). Identification of Protein Subcellular Localization with Network and Functional Embeddings. *Front. Genet.* 11, 626500. doi:10.3389/fgene.2020.626500

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-Learn: Machine Learning in Python. *J. Machine Learn. Res.* 12, 2825–2830.

Hanchuan Peng, H., Fuhui Long, F., and Ding, C. (2005). Feature Selection Based on Mutual Information Criteria of max-dependency, max-relevance, and Min-Redundancy. *IEEE Trans. Pattern Anal. Machine Intell.* 27, 1226–1238. doi:10.1109/tpami.2005.159

Picarda, E., Ohaegbulam, K. C., and Zang, X. (2016). Molecular Pathways: Targeting B7-H3 (CD276) for Human Cancer Immunotherapy. *Clin. Cancer Res.* 22, 3425–3431. doi:10.1158/1078-0432.ccr-15-2428

J. Platt (Editor) (1998). *Fast Training of Support Vector Machines Using Sequential Minimal Optimization* (Cambridge, MA: MIT Press).

Raymond, J. R., Appleton, K. M., Pierce, J. Y., and Peterson, Y. K. (2014). Suppression of GNAI2 Message in Ovarian Cancer. *J. Ovarian Res.* 7, 6. doi:10.1186/1757-2215-7-6

Russo, M., Siravegna, G., Blaszkowsky, L. S., Corti, G., Crisafulli, G., Ahronian, L. G., et al. (2016). Tumor Heterogeneity and Lesion-specific Response to Targeted Therapy in Colorectal Cancer. *Cancer Discov.* 6, 147–153. doi:10.1158/2159-8290.cd-15-1283

Safavian, S. R., and Landgrebe, D. (1991). A Survey of Decision Tree Classifier Methodology. *IEEE Trans. Syst. Man. Cybern.* 21, 660–674. doi:10.1109/21.97458

Senchenko, V. N., Anedchenko, E. A., Kondratieva, T. T., Krasnov, G. S., Dmitriev, A. A., Zabarovska, V. I., et al. (2010). Simultaneous Down-Regulation of Tumor Suppressor Genes RBSP3/CTDSPL, NPRL2/G21 and RASSF1A in Primary Non-small Cell Lung Cancer. *BMC cancer* 10, 75. doi:10.1186/1471-2407-10-75

Sun, L., Xu, X., Chen, Y., Zhou, Y., Tan, R., Qiu, H., et al. (2018). Rab34 Regulates Adhesion, Migration, and Invasion of Breast Cancer Cells. *Oncogene* 37, 3698–3714. doi:10.1038/s41388-018-0202-7

Tambe, Y., Hasebe, M., Kim, C. J., Yamamoto, A., and Inoue, H. (2016). The Drs Tumor Suppressor Regulates Glucose Metabolism via Lactate Dehydrogenase-B. *Mol. Carcinog.* 55, 52–63. doi:10.1002/mc.22258

Teng, Y., Qin, H., Bahassan, A., Bendzunas, N. G., Kennedy, E. J., and Cowell, J. K. (2016). The WASF3-NCKAP1-CYFIP1 Complex Is Essential for Breast Cancer Metastasis. *Cancer Res.* 76, 5133–5142. doi:10.1158/0008-5472.can-16-0562

Tran, N. L., Mcdonough, W. S., Savitch, B. A., Fortin, S. P., Winkles, J. A., Symons, M., et al. (2006). Increased Fibroblast Growth Factor-Inducible 14 Expression Levels Promote Glioma Cell Invasion via Rac1 and Nuclear Factor-κB and Correlate with Poor Patient Outcome. *Cancer Res.* 66, 9535–9542. doi:10.1158/0008-5472.can-06-0418

Van Der Zee, J. A., Van Eijck, C. H., Hop, W. C., Biermann, K., Dicheva, B. M., Seynhaeve, A. L., et al. (2012). Tumour Basement Membrane Laminin Expression Predicts Outcome Following Curative Resection of Pancreatic Head Cancer. *Br. J. Cancer* 107, 1153–1158. doi:10.1038/bjc.2012.373

Wajant, H. (2013). The TWEAK-Fn14 System as a Potential Drug Target. *Br. J. Pharmacol.* 170, 748–764. doi:10.1111/bph.12337

Wang, T., Ma, S., Qi, X., Tang, X., Cui, D., Wang, Z., et al. (2017). Knockdown of the Differentially Expressed Gene TNFRSF12A Inhibits Hepatocellular Carcinoma Cell Proliferation and Migration *In Vitro*. *Mol. Med. Rep.* 15, 1172–1178. doi:10.3892/mmr.2017.6154

Wang, Y., Xu, Y., Yang, Z., Liu, X., and Dai, Q. (2021). Using Recursive Feature Selection with Random Forest to Improve Protein Structural Class Prediction for Low-Similarity Sequences. *Comput. Math. Methods Med.* 2021, 5529389. doi:10.1155/2021/5529389

Watts, G. S., Tran, N. L., Berens, M. E., Bhattacharyya, A. K., Nelson, M. A., Montgomery, E. A., et al. (2007). Identification of Fn14/TWEAK Receptor as a Potential Therapeutic Target in Esophageal Adenocarcinoma. *Int. J. Cancer* 121, 2132–2139. doi:10.1002/ijc.22898

Weinstein, J. N., Collisson, E. A., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., et al. (2013). The Cancer Genome Atlas Pan-Cancer Analysis Project. *Nat. Genet.* 45, 1113–1120. doi:10.1038/ng.2764

Whitsett, T. G., Cheng, E., Inge, L., Asrani, K., Jameson, N. M., Hostetter, G., et al. (2012). Elevated Expression of Fn14 in Non-small Cell Lung Cancer Correlates with Activated EGFR and Promotes Tumor Cell Migration and Invasion. *Am. J. Pathol.* 181, 111–120. doi:10.1016/j.ajpath.2012.03.026

Wodarz, A., and Näthke, I. (2007). Cell Polarity in Development and Cancer. *Nat. Cel Biol.* 9, 1016–1024. doi:10.1038/ncb433

World Health Organization (2019). International Agency for Research on Cancer. Available at https://gco.iarc.fr/today (Accessed January 4, 2021).

Wu, S., Zhao, X., Wu, S., Du, R., Zhu, Q., Fang, H., et al. (2016). Overexpression of B7-H3 Correlates with Aggressive Clinicopathological Characteristics in Non-small Cell Lung Cancer. *Oncotarget* 7, 81750–81756. doi:10.18632/oncotarget.13177

Wu, Z.-H., Niu, X., Wu, G.-H., and Cheng, Q. (2020). Decreased Expression of TNFRSF12A in Thyroid Gland Cancer Predicts Poor Prognosis. *Medicine* 99, e21882. doi:10.1097/md.0000000000021882

Xiao, Z., Dai, Z., and Locasale, J. W. (2019). Metabolic Landscape of the Tumor Microenvironment at Single Cell Resolution. *Nat. Commun.* 10, 3763. doi:10.1038/s41467-019-11738-0

Xiong, Y., He, L., Shay, C., Lang, L., Loveless, J., Yu, J., et al. (2019). Nck-associated Protein 1 Associates with HSP90 to Drive Metastasis in Human Non-small-cell Lung Cancer. *J. Exp. Clin. Cancer Res.* 38, 122. doi:10.1186/s13046-019-1124-0

Xu, H., Liu, P., Yan, Y., Fang, K., Liang, D., Hou, X., et al. (2020). FKBP9 Promotes the Malignant Behavior of Glioblastoma Cells and Confers Resistance to Endoplasmic Reticulum Stress Inducers. *J. Exp. Clin. Cancer Res.* 39, 44. doi:10.1186/s13046-020-1541-0

Yamaguchi, H., and Condeelis, J. (2007). Regulation of the Actin Cytoskeleton in Cancer Cell Migration and Invasion. *Biochim. Biophys. Acta (Bba) - Mol. Cel Res.* 1773, 642–652. doi:10.1016/j.bbamcr.2006.07.001

Yamazaki, D., Kurisu, S., and Takenawa, T. (2005). Regulation of Cancer Cell Motility through Actin Reorganization. *Cancer Sci.* 96, 379–386. doi:10.1111/j.1349-7006.2005.00062.x

Yan, J., Ma, C., and Gao, Y. (2017). MicroRNA-30a-5p Suppresses Epithelial-Mesenchymal Transition by Targeting Profilin-2 in High Invasive Non-small Cell Lung Cancer Cell Lines. *Oncol. Rep.* 37, 3146–3154. doi:10.3892/or.2017.5566

Yang, Y., and Chen, L. (2021). Identification of Drug–Disease Associations by Using Multiple Drug and Disease Networks. *Curr. Bioinf.* 16. doi:10.2174/1574893616666210825115406

Yang, J., Min, K.-W., Kim, D.-H., Son, B. K., Moon, K. M., Wi, Y. C., et al. (2018). High TNFRSF12A Level Associated with MMP-9 Overexpression Is Linked to Poor Prognosis in Breast Cancer: Gene Set Enrichment Analysis and Validation in Large-Scale Cohorts. *PloS one* 13, e0202113. doi:10.1371/journal.pone.0202113

Ying, X., Zhu, J., and Zhang, Y. (2019). Circular RNA Circ-TSPAN4 Promotes Lung Adenocarcinoma Metastasis by Upregulating ZEB1 via Sponging miR-665. *Mol. Genet. Genomic Med.* 7, e991. doi:10.1002/mgg3.991

Yu, J., Zhang, Y., Zhou, D., Wu, J., and Luo, R. (2018). Higher Expression of A-Kinase Anchoring-Protein 1 Predicts Poor Prognosis in Human Hepatocellular Carcinoma. *Oncol. Lett.* 16, 131–136. doi:10.3892/ol.2018.8685

Zhang, C., and Hao, X. (2019). Prognostic Significance of CD276 in Non-small Cell Lung Cancer. *Open Med.* 14, 805–812. doi:10.1515/med-2019-0076

Zhang, H., Yang, W., Yan, J., Zhou, K., Wan, B., Shi, P., et al. (2018a). Loss of Profilin 2 Contributes to Enhanced Epithelial-Mesenchymal Transition and Metastasis of Colorectal Cancer. *Int. J. Oncol.* 53, 1118–1128. doi:10.3892/ijo.2018.4475

Zhang, Y., Fang, L., Zang, Y., and Xu, Z. (2018b). Identification of Core Genes and Key Pathways via Integrated Analysis of Gene Expression and DNA Methylation Profiles in Bladder Cancer. *Med. Sci. Monit.* 24, 3024–3033. doi:10.12659/msm.909514

Zhang, Y.-H., Li, H., Zeng, T., Chen, L., Li, Z., Huang, T., et al. (2021a). Identifying Transcriptomic Signatures and Rules for SARS-CoV-2 Infection. *Front. Cel Dev. Biol.* 8, 627302. doi:10.3389/fcell.2020.627302

Zhang, Y.-H., Li, Z., Zeng, T., Chen, L., Li, H., Huang, T., et al. (2021b). Detecting the Multiomics Signatures of Factor-specific Inflammatory Effects on Airway Smooth Muscles. *Front. Genet.* 11, 599970. doi:10.3389/fgene.2020.599970

Zhang, Y.-H., Zeng, T., Chen, L., Huang, T., and Cai, Y.-D. (2021c). Determining Protein-Protein Functional Associations by Functional Rules Based on Gene Ontology and KEGG Pathway. *Biochim. Biophys. Acta (Bba) - Proteins Proteomics* 1869, 140621. doi:10.1016/j.bbapap.2021.140621

Zhao, X., Chen, L., and Lu, J. (2018). A Similarity-Based Method for Prediction of Drug Side Effects with Heterogeneous Information. *Math. Biosci.* 306, 136–144. doi:10.1016/j.mbs.2018.09.010

Zhong, X.-P., Kan, A., Ling, Y.-H., Lu, L.-H., Mei, J., Wei, W., et al. (2019). NCKAP1 Improves Patient Outcome and Inhibits Cell Growth by Enhancing Rb1/p53 Activation in Hepatocellular Carcinoma. *Cell Death Dis.* 10, 369. doi:10.1038/s41419-019-1603-4

Zhou, H., Ekmekcioglu, S., Marks, J. W., Mohamedali, K. A., Asrani, K., Phillips, K. K., et al. (2013). The TWEAK Receptor Fn14 Is a Therapeutic Target in Melanoma: Immunotoxins Targeting Fn14 Receptor for Malignant Melanoma Treatment. *J. Invest. Dermatol.* 133, 1052–1062. doi:10.1038/jid.2012.402

Zhou, D., Bai, F., Zhang, X., Hu, M., Zhao, G., Zhao, Z., et al. (2014). SOX10 Is a Novel Oncogene in Hepatocellular Carcinoma through Wnt/β-catenin/TCF4 cascade. *Tumor Biol.* 35, 9935–9940. doi:10.1007/s13277-014-1893-1

Zhou, K., Chen, J., Wu, J., Xu, Y., Wu, Q., Yue, J., et al. (2019). Profilin 2 Promotes Proliferation and Metastasis of Head and Neck Cancer Cells by Regulating PI3K/AKT/β-Catenin Signaling Pathway. *Oncol. Res.* 27, 1079–1088. doi:10.3727/096504019x15579146061957

Zhou, J. P., Chen, L., and Guo, Z. H. (2020a). iATC-NRAKEL: An Efficient Multi-Label Classifier for Recognizing Anatomical Therapeutic Chemical Classes of Drugs. *Bioinformatics* 36, 1391–1396. doi:10.1093/bioinformatics/btz757

Zhou, J.-P., Chen, L., Wang, T., and Liu, M. (2020b). iATC-FRAKEL: a Simple Multi-Label Web Server for Recognizing Anatomical Therapeutic Chemical Classes of Drugs with Their Fingerprints Only. *Bioinformatics* 36, 3568–3569. doi:10.1093/bioinformatics/btaa166

Zhu, Y., Hu, B., Chen, L., and Dai, Q. (2021). iMPTCE-Hnetwork: A Multilabel Classifier for Identifying Metabolic Pathway Types of Chemicals and Enzymes with a Heterogeneous Network. *Comput. Math. Methods Med.* 2021, 6683051. doi:10.1155/2021/6683051

Zhuang, H., Tan, M., Liu, J., Hu, Z., Liu, D., Gao, J., et al. (2014). Human Epididymis Protein 4 in Association with Annexin II Promotes Invasion and Metastasis of Ovarian Cancer Cells. *Mol. Cancer* 13, 243. doi:10.1186/1476-4598-13-243

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: 09 December 2021
doi: 10.3389/fcell.2021.767340

Quantified CIN Score From Cell-free DNA as a Novel Noninvasive Predictor of Survival in Patients With Spinal Metastasis

Su Chen[1†], Minglei Yang[1†], Nanzhe Zhong[1†], Dong Yu[2], Jiao Jian[1], Dongjie Jiang[1], Yasong Xiao[1], Wei Wei[1], Tianzhen Wang[3], Yan Lou[1], Zhenhua Zhou[1], Wei Xu[1], Wan Wan[1], Zhipeng Wu[1], Haifeng Wei[1], Tielong Liu[1], Jian Zhao[1], Xinghai Yang[1] and Jianru Xiao[1]**

[1]Department of Orthopedic Oncology, Changzheng Hospital, Naval Medical University, Shanghai, China, [2]Center of Translational Medicine, Naval Medical University, Shanghai, China, [3]Weizmann Institute of Sciences, Rehovot, Israel

Purpose: Most currently available scores for survival prediction of patients with bone metastasis lack accuracy. In this study, we present a novel quantified CIN (Chromosome Instability) score modeled from cfDNA copy number variation (CNV) for survival prediction.

Experimental Design: Plasma samples collected from 67 patients with bone metastases from 11 different cancer types between November 2015 and May 2016 were sent through low-coverage whole genome sequencing followed by CIN computation to make a correlation analysis between the CIN score and survival prognosis. The results were validated in an independent cohort of 213 patients.

Results: During the median follow-up period of 598 (95% CI 364–832) days until December 25, 2018, 124 (44.3%) of the total 280 patients died. Analysis of the discovery dataset showed that CIN score = 12 was the optimal CIN cutoff. Validation dataset showed that CIN was elevated (score ≥12) in 87 (40.8%) patients, including 5 (5.75%) with head and neck cancer, 11 (12.6%) with liver and gallbladder cancer, 11 (12.6%) with cancer from unidentified sites, 21 (24.1%) with lung cancer, 7 (8.05%) with breast cancer, 4 (4.60%) with thyroid cancer, 6 (6.90%) with colorectal cancer, 4 (4.60%) with kidney cancer, 2 (2.30%) with prostate cancer, and 16 (18.4%) with other types of cancer. Further analysis showed that patients with elevated CIN were associated with worse survival ($p < 0.001$). For patients with low Tokuhashi score (≤8) who had predictive survival of less than 6 months, the CIN score was able to distinguish patients with a median overall survival (OS) of 443 days (95% CI 301–585) from those with a median OS of 258 days (95% CI 184–332).

OPEN ACCESS

Edited by:
Geng Chen,
GeneCast Biotechnology Co., Ltd., China

Reviewed by:
Kosuke Nakano,
Osaka University, Japan
Maria Neofytou,
University of Cambridge, United Kingdom

***Correspondence:**
Jianru Xiao
Jianruxiao83@163.com
Xinghai Yang
cnspineyang@163.com

[†]These authors 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: 30 August 2021
Accepted: 18 October 2021
Published: 09 December 2021

Citation:
Chen S, Yang M, Zhong N, Yu D, Jian J, Jiang D, Xiao Y, Wei W, Wang T, Lou Y, Zhou Z, Xu W, Wan W, Wu Z, Wei H, Liu T, Zhao J, Yang X and Xiao J (2021) Quantified CIN Score From Cell-free DNA as a Novel Noninvasive Predictor of Survival in Patients With Spinal Metastasis. Front. Cell Dev. Biol. 9:767340. doi: 10.3389/fcell.2021.767340

Abbreviations: AFP: alpha fetoprotein; AUC, area under ROC; bp, base pair; CEA, carcinoembryonic antigen; cen, centromere; cfDNA, Cell-free DNA; chr, chromosome; CIN, chromosome instability; CNV, copy number variation; DKK, dickkopf; DLC, deleted in liver cancer; DNA, deoxyribonucleic acid; EDTA, ethylenediaminetetraacetate; KPS, Karnofsky Performance Status; MCL, mantle cell lymphoma; ng, nanogram; NS, not significant; OS, overall survival; p, petit; PTEN, phosphatase and tensin homolog; ROC, operating characteristic curve; TOK, Tokuhashi score; TP53, tumor protein 53; VEGFA, vascular endothelial cell growth factor A.

Conclusion: CNV examination in bone metastatic cancer from cfDNA is superior to the traditional predictive model in that it provides a noninvasive and objective method of monitoring the survival of patients with spine metastasis.

Keywords: tokuhashi score, prognosis, CIN score, CNV, spine metastasis

INTRODUCTION

The skeletal system is the third most common metastatic site of most cancers (Sutcliffe et al., 2013). Postmortem examination showed that about 70% of cancer patients had spinal metastases (Bartels et al., 2008). Intractable pain, neurologic deficits, and paralysis are common symptoms of these patients, fundamentally impacting their survival and quality of life (Tang et al., 2016). Survival time of the patients with bone metastases is a key consideration for the decision-making of subsequent treatments. However, there is a lack of robust clinical tools or biomarkers to accurately predict the prognosis of bone metastatic patients.

Several scoring systems including Tomita, Sioutos, and Van der Linden scores have been developed (Tomita et al., 2001; Sioutos et al., 1995; van der Linden et al., 2005), and the revised Tokuhashi score is the most commonly used in survival prediction of cancer patients (Tokuhashi et al., 1990; Tokuhashi et al., 2005). However, previous studies have demonstrated that the revised Tokuhashi score is not accurate enough in prognostic prediction of patients with spinal metastasis, and therefore the survival of patients with low a Tokuhashi score (≤8) might be underestimated (Tan et al., 2016; Yang et al., 2019). Precise selection of these patients could improve therapeutic planning and ameliorate the clinical outcomes.

The implementation of molecular diagnostics involves scalability, sensitivity, and specificity. In patients with spinal metastasis, whole-genome sequencing analysis of cell-free DNA has emerged to be a promising strategy for reflecting the burden and genetic alterations of tumors (Ellinger et al., 2011). The profiling of genetic alterations in circulating cell-free DNA from patient plasma has potential clinical applications including early disease detection, real-time prediction of treatment response, and prognostication (Tran et al., 2021). In various cancer types, cfDNA unveils abundant information including the metastatic status, microsatellite instability, somatic copy number alterations, and the structural rearrangements (Vanderstichele et al., 2017; Mayrhofer et al., 2018). Copy number variation (CNV) is the result of ongoing changes in the number or structure of chromosomes, which was a hallmark in various cancers (Vanderstichele et al., 2017; Gronroos and López-García, 2018). Initially, maternal neoplasia was incidentally detected by identifying chromosomal and/or subchromosomal abnormalities during noninvasive prenatal testing (Bianchi et al., 2015; Dharajiya et al., 2018). A later study on the large population validated a novel cancer detection pipeline algorithm to enhance the detection of occult maternal malignancies through identifying multiple chromosomal aneuploidies (Ji et al., 2019). In addition to noninvasive prenatal testing, other noninvasive examinations focusing on chromosomal instability analysis suggested high CNV detected in cfDNA could serve as surrogates for predicting and monitoring malignancies (Oellerich et al., 2017).

In this study, low-coverage whole-genome sequencing of cfDNA was conducted to examine blood plasma from patients with spinal metastasis. An analysis pipeline was developed and validated to evaluate the CNV status in cfDNA, in an attempt to determine whether the CIN score, which was defined as a prognostic factor in patients with metastatic cancer like breast cancer (Mo et al., 2020), could also predict the survival of pan-cancer patients with spinal metastasis and especially identify the patients with actual good prognosis but ranked with a low Tokuhashi score.

Patients and Methods

Patients and Study Design

This cohort study was approved by the institutional review board. Written informed consent was obtained from all participating patients. Between November 2015 and May 2016, 67 patients with spinal metastasis were retrospectively included in the discovery group, and between June 2016 and September 2018, 213 patients were included as a prospective cohort for final validation (validation group), shown in **Figure 1**. The inclusion criteria are as follows: patients diagnosed with radiologically and pathologically confirmed spinal metastases. Blood samples were collected in EDTA anticoagulant tubes (BD, USA) before the surgical treatment and centrifuged to obtain plasma within 40 min. Patients with unqualified blood samples were excluded. All the patients were preoperatively assessed according to the revised Tokuhashi score (Tokuhashi et al., 2005). Clinical data were abstracted from the medical records, including the general condition, primary tumor site, visceral metastasis, vertebral metastasis, extraspinal metastasis, and neurologic deficits. The patients' general conditions were evaluated according to the Karnofsky Performance Status (KPS). The neurological status was assessed based on Frankel scale. Overall survival (OS) was defined as the interval between the first diagnosis and death or the date of last follow-up. The study was terminated in December 2018, when more than one third of patients died in the validation group.

cfDNA Extraction and Whole-Genome Low-Coverage Sequencing

Total genomic DNA and cfDNA were isolated from plasma using the Amp Genomic DNA Kit (TIANGEN) and QIAseq cfDNA Extraction kit (Qiagen), respectively. Next-generation sequencing was performed as previously described (Stover et al., 2018). cfDNA (10 ng) was used for preparation of sequencing libraries (NEBnext Ultra II). Barcoded sequencing adaptors (8 bp) were then ligated with DNA fragments and amplified by PCR. Purified sequencing libraries were massively parallel sequenced by Illumina NovaSeq 6000 platform. 4G sequencing of raw data

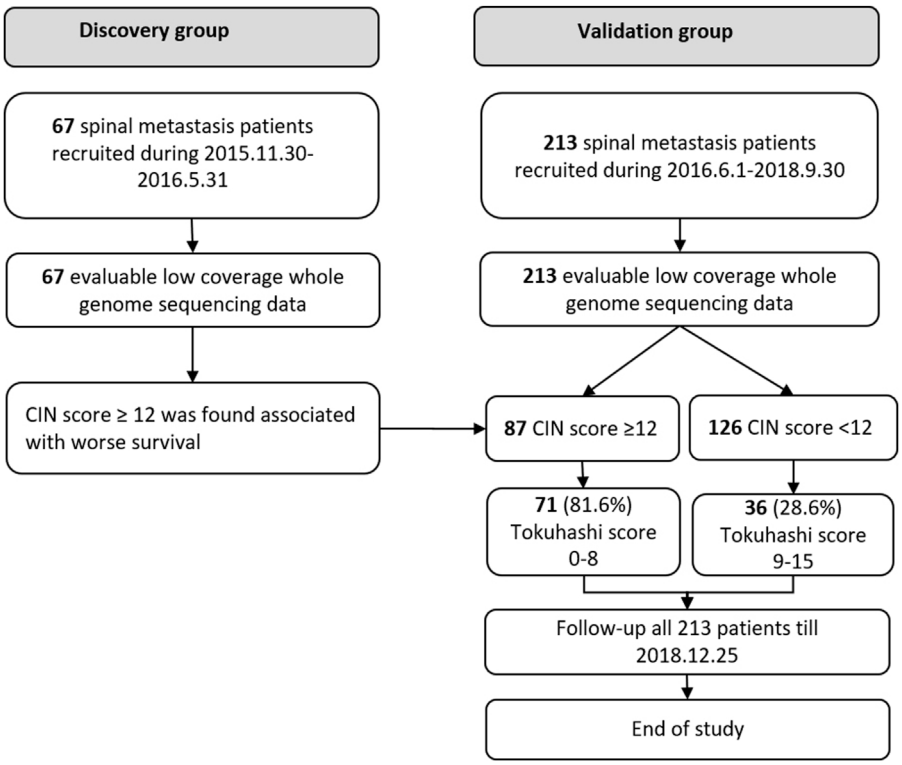


FIGURE 1 | Study design.

per sample was filtered and aligned to the human reference genome.

CIN Score in cfDNA

Plasma cell-free DNA was extracted and analyzed by Illumina NovaSeq 6000 platform. At least 10M paired reads were collected for each sample. The reads were mapped to human reference genome hg19. Genomic coverage was then counted by using software samtools mpileup (Ji et al., 2019). The mean coverage for each 200k bin was then calculated as previously reported (Mo et al., 2020). Then 100 kb, 200 kb, and 400 kb bin size were compared in **Supplementary Figure S3**. Z-scores for each bin were then normalized by Z-score by using the following formula:

$$Coverage_{normalized} = \frac{coverege_{raw} - mean(coverage_{controls,\ raw})}{stdev(coverage_{controls,\ raw})} \quad (1)$$

Circular binary segmentation (CBS) algorithm from R package DNA Copy (Oellerich et al., 2017) was then used to find significant genomic breakpoints and copy number changed genomics segments.

The Overall Chromosomal Instability Scores Are Summarized as

$$CIN\_score = \log2\left(\sum_{s \in segments} V_s \times L_s\right)$$

where V_s is the value of the segment and, L_s is the length of the segments in unit of 200,000 base pairs.

Statistical Analysis

Statistical analysis was performed by R project. Continuous variables are expressed as the mean ± SD. Categorical variables were compared using the Fisher exact test. The Kaplan–Meier plots were generated to estimate survival and compared using the log-rank test, using packHV package. The median follow-up was estimated by reverse Kaplan–Meier method. Univariate and multivariable Cox proportional hazards models were calculated using the survival package. The receiver operating characteristic curve (ROC) was constructed, and the area under ROC (AUC) was calculated using the ROCR package.

Raw sequencing data reported in this paper have been deposited in the Genome Sequence Archive in the BIG Data Center, Chinese Academy of Sciences that are accessible at http://bigd.big.ac.cn/gsa (ref: Database resources of the BIG Data Center in 2018). The BioProject number is PRJCA006730.

RESULTS

Patient Characteristics

In the discovery group, 67 plasma samples were identified for low-coverage whole-genome sequencing. The control data

TABLE 1 | Patient characterization in discovery and validation group.

Factors	Discovery (2015.11-2016.5) N=67	Validation (2016.6 -2018.09) N=213	Fisher test
Age			
≥58	42	108	NS
≤57	25	105	
Gender			
Female	17	89	0.02
Male	50	124	
Primary Site			
Lung	12	49	NS
Unidentified	22	28	<0.001
Kidney	7	31	NS
Liver, Gallbladder	9	23	NS
Thyroid	7	19	NS
Breast	3	15	NS
Prostate	1	6	NS
Urine	2	2	NS
Others	4	40	0.01
Count of vertebral metastasis			
≥3	27	87	NS
2	15	46	NS
1	25	80	NS
Visceral metastases			
Unremovable	10	43	NS
Removable	3	11	NS
No Metastases	54	159	NS
Palsy			
Complete	23	75	NS
Incomplete	36	122	NS
None	8	16	NS

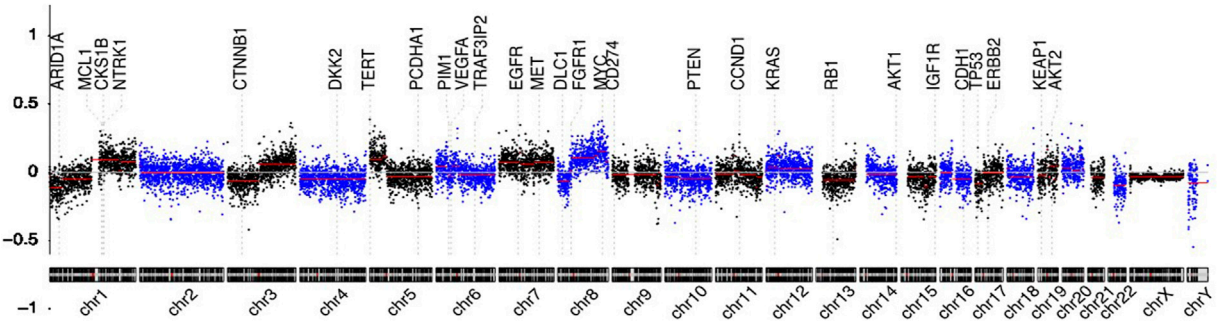


FIGURE 2 | Circulating cell-free genome landscape of bone metastasis tumors. Circulating cell-free chromosomal instability of spinal metastasis cancers. Chromosomes 1, 2, . . . , and Y are plotted from left to right. Each dot indicates the normalized coverage value of a 200k bin. Genes of interest are marked with dashed lines.

acquired from healthy Chinese blood donors were generously provided by Suzhou Hongyuan Biotech Inc. (Suzhou, China). The 213 plasma samples from the validation group were used for final analysis. The mean age of the patients in the discovery and validation groups was 57.81 ± 11.33 (median 58, range 27–80) and 57.09 ± 12.20 (median 58, range 21–85) years, respectively. The median follow-up time in the two groups was 1,002 (95% CI 984–1020) and 315 (95% CI 234–396) days, with 45 (67.2%) patients and 79 (37.1%) patients passing away before the last follow-up. Differences in baseline data between the two groups were compared (**Table 1**).

Assessment of CNVs in cfDNA

Genome-wide overview of CNVs in 280 patients of the two groups are summarized in **Figure 2**. Chromosome arms 1q, 6p, 7, 8q, 20, 10p, 5p15.33, 10cen, 15q, 17q, 19, 11q13.3, and

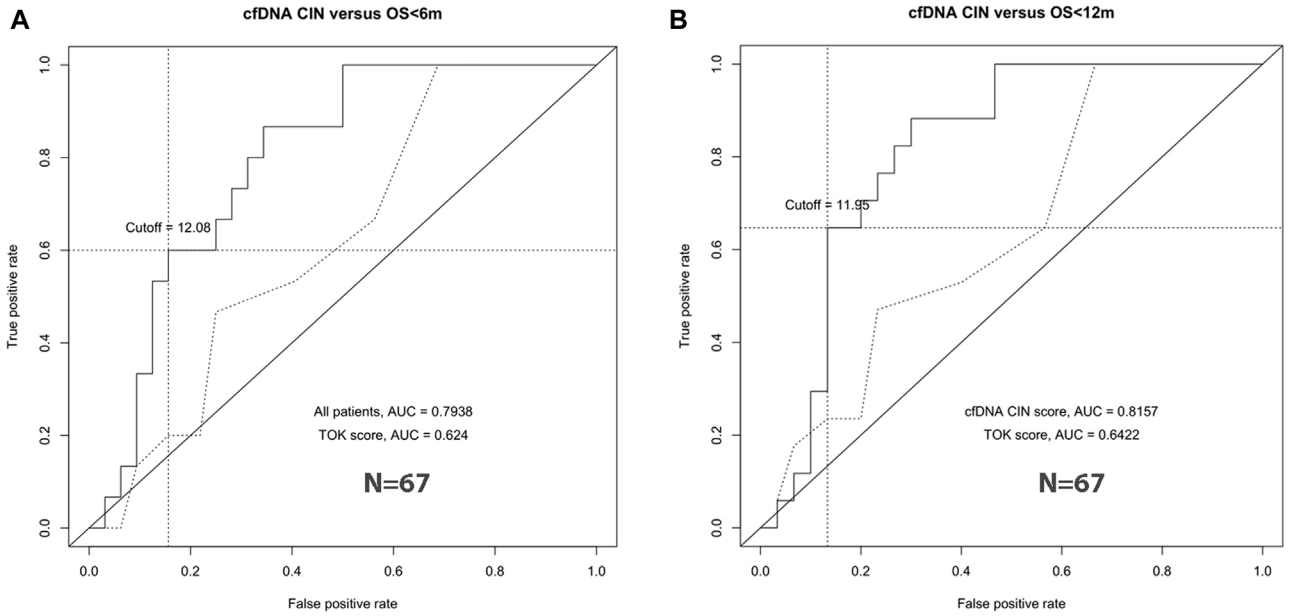


FIGURE 3 | CIN cutoff by ROC curve analyses. The best cutoff to predict 6-months survival **(left, A)** and 12-months survival **(right, B)**.

22q were found with statistically significant copy number gains in 78.26, 47.83, 34.78, 91.30, 52.17, 30.43, 39.13, 21.74, 17.39, 39.13, 8.70, 13.04, and 13.04% samples, respectively, where well-studied oncogenes, MYC(8q), MCL1(1q), and VEGFA(6p), were located. Chromosome arms 1p, chr4, 6q, 10q, 13q, 8p, 11q, 11p, chr16, chr9, 17p, 21q, 14q, chr3, chr18, and 8p were found with statistically significant copy loss in 65.22, 60.87, 47.83, 60.87, 65.22, 52.17, 39.13, 43.48, 43.48, 47.83, 52.17, 30.43, 43.48, 21.74, 26.09, and 34.78% of samples, respectively, where potential tumor suppressor genes DLC1(8p), DKK2(4q), PTEN(10q), and TP53(17p) were located.

Detectable chromosomal changes were much less pronounced in thyroid and kidney cancer than those in lung, liver, breast, and head and neck cancers. For tumors with high CNVs, chromosome 1q gain and 8q gain were common. Chr6p gain and 4q loss were the most frequent in cancers originating from the liver. 3q gain was the most frequent in lung, head and neck, and breast cancers. Chr7 gain was the most in lung cancer and head and neck cancer. Chr20 gain was detectable in liver and breast cancers.

We also ran our data through ichorCNA algorithm (https://github.com/broadinstitute/ichorCNA) and found a linear relationship of tumor fraction between our pipeline and ichorCNA pipeline. ctDNA% also showed a positive correlation with the estimated CIN score in this study (**Supplementary Figure S2**).

Increased CNVs Predict Worse Survival

In the discovery group, survival ROC curves were generated to identify the specific threshold of the CIN score for the classification of long-survival patients and short-survival patients. For 6- and 12-months survival prediction, the AUC value was 0.7938 (95% CI 0.6657–0.9218) and 0.8157 (95% CI 0.6922–0.9392) using CIN score, with the threshold of 12.08 and 11.95 (**Figures 3A,B**). Based on the results in the discovery group, patients with CIN score ≥12.0 were categorized as high-CIN status and those with a lower score as low-CIN status. The median OS of patients with high- or low-CIN status was 213 (95% CI 66.2–360) and 810 (95% CI 369–1251) days, respectively.

In the validation group, univariate analysis showed that patients with cfDNA CIN score <12 (HR 0.314, 95% CI 0.201–0.493, $p < 0.001$), Tokuhashi score, >8 (HR 2.252, 95% CI 1.559–3.253, $p < 0.001$), better general condition (HR 0.604, 95% CI 0.418–0.873, $p = 0.007$), low malignant primary tumor (HR 0.735, 95% CI 0.641–0.843, $p < 0.001$), a better Frankel scale score (HR 0.577, 95% CI 0.398–0.836, $p = 0.004$), inoperable visceral metastasis (HR 0.731, 95% CI 0.564–0.948, $p = 0.018$), age ≥58 (HR 1.634, 95% CI 1.044–2.557, $p = 0.032$), and extra-spinal metastasis (HR 0.563, 95% CI 0.412–0.768, $p < 0.001$) were statistically significant for short survival (**Figure 4**).

Furthermore, cfDNA CIN score <12 (**Table 2**, HR 0.561, 95% CI 0.447–0.703, $p < 0.001$) and Tokuhashi score, ≤8 (**Table 2**, HR 2.258, 95% CI 1.561–3.266, $p < 0.001$) both reached statistical significance by two-parameter Cox regression analysis. Also, all these eight variables were used for multivariate analysis, finding that the primary tumor site (HR 0.728, 95% CI 0.620–0.855, $p < 0.001$), Tokuhashi score, ≤8 (HR 2.886, 95% CI 1.318–6.323, $p = 0.008$), extra-spinal metastasis (HR 0.529, 95% CI 0.384–0.729, $p < 0.001$), and cfDNA CIN score <12 (HR 0.322, 95% CI 0.204–0.507, $p < 0.001$) were predictors of survival for patients with spinal metastasis (**Table 3**).

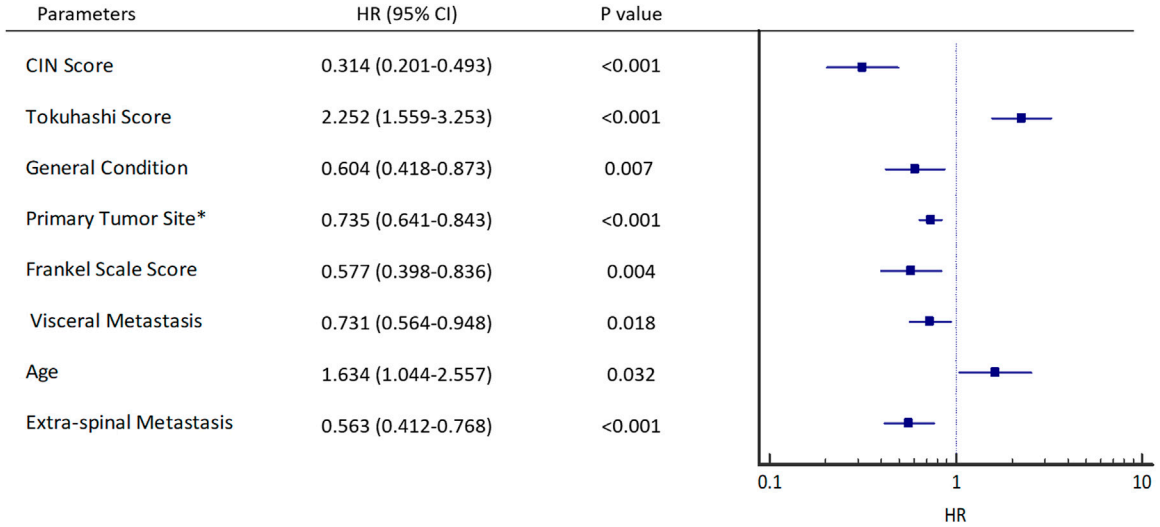


FIGURE 4 | Univariant hazard ratios. *Primary site of cancer according to Tokuhashi score, 5 (Thyroid, Prostate, Breast), 4 (Rectum), 3 (Kidney, Uterus), 2 (Others), 1 (Liver, Gallbladder, Unidentified), 0 (Lung, Osteosarcoma, Stomach). General condition is estimated by KPS score, KPS (Good: 80–100, Moderate: 50–70, Poor: 10–40).

TABLE 2 | Two-parameter Cox-regression analysis.

	P	HR	95% CI
cfDNA CIN_score < 12	<0.001	0.561	0.447–0.703
TOK_Score ≤ 8	<0.001	2.258	1.561–3.266

TABLE 3 | Eight-parameter multivariate survival analyses.

	P	HR	95% CI
cfDNA CIN_score < 12	<0.001	0.322	0.204–0.507
Primary site	<0.001	0.728	0.620–0.855
Tokuhashi score ≤ 8	0.008	2.886	1.318–6.323
Ex- vertebral bone metastases	<0.001	0.529	0.384–0.729

Discrimination of Long-Survival Patients With Poor Tokuhashi Score

Knowing that Tokuhashi score is widely used in prognostic prediction of spinal metastasis patients, we tested its accuracy in the validation group (**Figure 5B**) and found that 161 patients with low Tokuhashi score (≤8) in the validation group had significantly shorter median OS (392 days, 95% CI 313–471), as compared with the median OS of the other 52 patients with Tokuhashi score >8 (median survival unreached, HR 2.252, 95% CI 1.559–3.253, $p < 0.001$). The median OS was 298 days (95% CI 129–467) for patients with high-CIN scores versus 707 days (95% CI 501–913) for patients with low-CIN scores (HR 0.314, 95% CI 0.201–0.493, $p < 0.001$, **Figure 5A**). Distribution of CIN score on each Tokuhashi score is shown in **Figure 5C**. Patients with high Tokuhashi score had lower CIN score ($p < 0.01$).

However, 99 (61.5%) of the 161 patients with low Tokuhashi score had predictive survival of less than 6 months, while their actual survival was more than 6 months. Next, in the low Tokuhashi score group, we tested whether the CNV status could discriminate the good-prognosis patients from the poor-prognosis ones. The median OS was 258 days (95% CI 184–332) for patients with high-CIN scores versus 443 days (95% CI 301–585) for patients with low-CIN scores ($p = 0.002$, **Figure 5E**). With the help of the CIN score, spinal metastasis patients with low Tokuhashi score could be divided into two different subgroups in terms of survival. Similarly, the median OS in the high Tokuhashi score group was also longer in the low-CIN patients than that in the high CIN patients ($p = 0.013$, **Figure 5D**).

The Performance of Combined use of Tokuhashi and cfDNA CIN Scores in Predicting Short- and Long-Survival Patients

We then combined Tokuhashi and CIN scores as a predictor of patient OS. As shown in **Table 4**, CIN scores larger than 12 defined the worst survival group of 87 (40.8%) patients, of whom 50 patients (57.5%) died within 6 months. The 90 (42.3%) patients with Tokuhashi scores ≤8 and CIN scores less than 12 showed a better prognosis, with the 6-months death rate of 22.2%. The other 36 patients with Tokuhashi scores >8 and CIN scores less than 12 showed the best survival, with the 6-months death rate of 5.56%, and 18 (50.0%) of them showed OS longer than 12 months.

DISCUSSION

Spinal metastasis is a common occurrence in multiple advanced cancers (Bartels et al., 2008). Pretreatment survival prediction

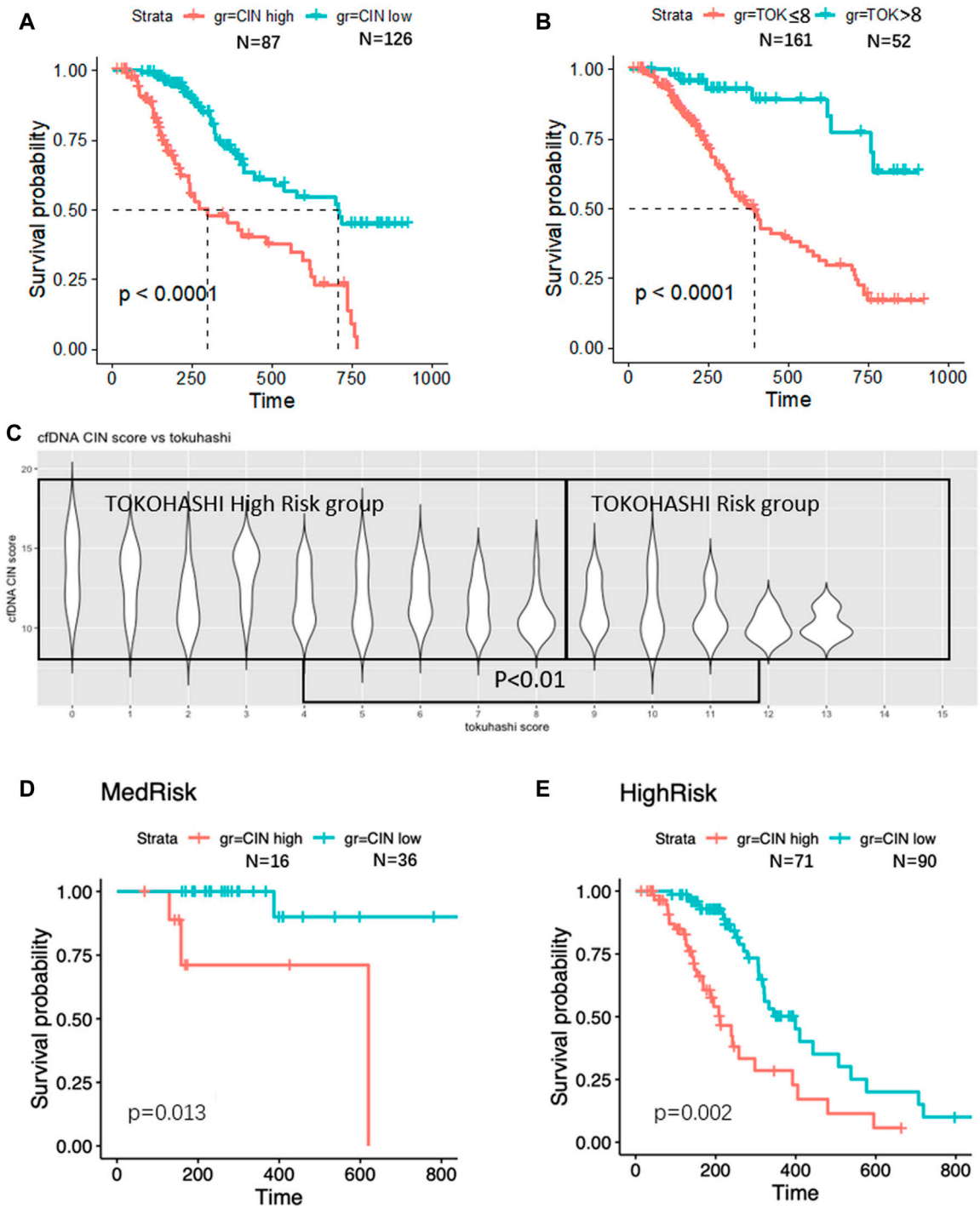


FIGURE 5 | Survival analyses. Significantly worse survival was found in CIN high **(A)** and low TOK score **(B)** patients. CIN is higher in TOK high-risk group **(C)**. In TOK medium-risk group, higher CIN was found associated with poor survival **(D)**. In TOK high-risk group, higher CIN was found associated with poor survival **(E)**.

may help clinicians in the decision-making process for subsequent treatments. Several survival predictive models have been developed over the past decades, but their accuracy is unsatisfactory (Tan et al., 2016). On the other hand, there is a lack of real-time biomarkers with consideration of the molecular characteristics of cancer cells for survival prediction (Bauer et al., 2002). It was found in this study that CNVs from cfDNA could act as a potential biomarker for prognosis prediction of patients

TABLE 4 | The performance of predicting of short and long-survival patients by combining Tokuhashi score and cfDNA CIN scores.

		<6 months	6–12 months	≥12 months
High risk (N=87)	UCAD≥12	50 (57.5%)	18 (20.7%)	19(21.8%)
Median risk (N=90)	TOK≤8, UCAD<12	20 (22.2%)	43 (47.8%)	27 (30.0%)
Low risk (N=36)	TOK>8, UCAD<12	2 (5.56%)	16 (44.4%)	18 (50.0%)

with spinal metastasis and help discriminate long-survival patients from poor-prognosis patients assessed by traditional prediction models in the perspective of molecular characteristics of malignancies.

Copy number variation is a hallmark in various cancers, including metastatic cancers, and deemed a prognostic predictor associated with an increased risk of recurrence or death (Gronroos and López-García, 2018; Tan et al., 2019). In this study, cfDNA CNV was first proved to be a potential biomarker for predicting the survival of patients with spinal metastasis. Vanderstichele et al. (Vanderstichele et al., 2017) reported that evaluating the cfDNA with CNV could help distinguish the patients with an adnexal mass caused by an invasive carcinoma from patients presenting with a benign adnexal mass. In addition, more genetic and epigenetic alterations could also be detected by cfDNA sequencing (Ellinger et al., 2011). Mehra et al. (2018) demonstrated that the cfDNA fragment concentration was an independent prognostic variable for radiological progression-free survival and OS of patients with metastatic prostate cancer. Stover et al. (2018) reported that tumor fraction was associated with worse survival in metastatic triple-negative breast cancer patients, and specific somatic copy number alterations were enriched in these patients. Spinal metastases originating from multiple types of cancer are featured by heterogeneous molecular characters, which impose a significant challenge on clinicians in precise diagnosis, evaluation, and treatment (Jansson and Bauer, 2006; Kawahara et al., 2009). As aforementioned, CNV could present the common properties of metastatic malignancies. Our results demonstrated that CNV in cfDNA was significantly associated with the survival of patients with spinal metastasis, which provides a universal and convenient tool of pretreatment evaluation for different cancers as an integrated entity. This interpretation of CNV could be applicable to the surveillance of progression of malignancies in other scenarios.

Another intriguing finding of our study is that CNV successfully discriminated the good-prognosis patients with low Tokuhashi score from poor-prognosis patients. According to the revised Tokuhashi score, the predicted survival of patients with a total score ≤8 was less than 6 months (Tokuhashi et al., 2005). However, it was found that Tokuhashi score might not be qualified to establish a precise prediction of survival, especially in predicting short-term survival (Ahmed et al., 2018). Some patients ranked with a low Tokuhashi score may possess a relatively longer survival than expected. Ahmed et al. (2018) reported that the revised Tokuhashi score was not accurate in 90-days survival prediction after surgery, with the AUC being 0.67. The data of Tan et al. (2016) suggested that only 41.7% patients had a Tokuhashi-predicted survival that was correlated with actual survival. Judged by Tokuhashi score, only 6.1% patients were predicted to survive more than 6 months, while 44.4% of patients in the cohort attained survival longer than half a year. Moreover, the mean OS of patients with spinal metastases have been observed to be improved and longer than 1 year in the recent literature (Tang et al., 2015; Yang et al., 2019), which is in congruence with our data. As CNV level can reflect the disease progression through the actual activity of the tumor individually, it proved capable of predicting the prognosis of patients with highly malignant spinal metastases in the current study, including lung cancer. Identification of the patients who have long-term survival in the low Tokuhashi score group can contribute to a more positive alteration of treatment, thus improving their actual prognosis. Pelegrini de Almeida et al. (2018) retrospectively analyzed 117 patients and found that the patients with low Tokuhashi score for whom surgical treatment was not recommended traditionally had better quality of life and longer survival after adequate surgery than the results inferred by the Tokuhashi system. It was found in our study that patients with poor Tokuhashi score but low CNV had relatively good prognosis, with the median OS of 433 days. Accurate survival prediction of this part of patients through cfDNA test might optimize the selection of their treatment procedures.

Unlike the traditional predictive models for metastatic malignancies such as the Tokuhashi score, which may be influenced by subjective confounders, CNV is calculated based on the plasma cfDNA, which represents a more objective set of criteria. Almost all previous models focus on clinical characteristics of the patients with spinal metastases, while CNV reflects the tumor status of progression in circulation (Batista et al., 2016; Vanderstichele et al., 2017). Although other blood tests including certain tumor biomarkers (CEA, AFP, etc.), inflammation-based biomarkers (platelet–lymphocyte ratio, neutrophil–lymphocyte ratio, etc.), and hemostasis biomarkers (D-dimer, fibrinogen, etc.) were also reported to be associated with survival of cancer patients, they are either specific only for a single tumor type or lack sensitivity (Ay et al., 2012; Li et al., 2018; Yang et al., 2019). Examining CNV in cfDNA for prognosis prediction provides more

information on the innate and dynamic character of metastatic malignancies, which are often ignored in most traditional models. Furthermore, accurate survival prediction through CNV may help make a more precise treatment decision. It has been deemed as the principle that the choice of treatment should be judged by the prognosis of spinal metastatic patients (Tokuhashi et al., 1990; Tomita et al., 2001). For good-prognosis patients, radical excision with wide or marginal margin is suggested for long-term local control, while for patients with poor prognostic prediction, intralesional excision or palliative treatment can reduce unnecessary iatrogenic injury (Tomita et al., 2001).

The main limitation of this study is that all blood samples were obtained from the cohort in one single institution. The patients recruited in our study are relatively homogeneous, and were all diagnosed with spinal metastases, implying that our conclusion should be interpreted cautiously in patients with metastatic sites involving other systems. Besides, we noticed different features of CIN and different clinical outcomes between tumors from various origins, but the score was worse when estimating each carcinoma (**Supplementary Figure S1**) than that of pan-cancer as a whole (**Supplementary Figure S1**). Thereby, an increased number of patients and further investigation taking into account the differences between malignancies are required to improve the performance of the algorithm. Second, as the cohorts of the discovery group and validation group were not evaluated synchronously, relevant biases may not be avoidable. The constitution of the malignant types is somewhat different between the two groups; for instance, fewer patients in the validation group were diagnosed with cancers of unknown primary sites than those in the discovery group ($p < 0.001$). Third, clinical advances in imaging and histopathological examinations with specific immunohistochemistry in more recent years may also affect tumor diagnosis and increase the sensitivity of primary tumor detection (Collado Martín et al., 2014; Varadhachary and Raber, 2014). Finally, more female patients were included in the validation group than those in the discovery group ($p = 0.02$).

In summary, we proposed a framework for the minimally invasive examination of genetic architecture of spinal metastatic cancers. The interpretation of its result was proved to be correlated with the clinical outcome of patients. A high cfDNA CIN score suggests a dismal survival of patients with spinal metastasis. cfDNA testing for CNV could be applied into clinical practice as a new indicator for the prognosis of spinal metastasis. Accurate selection of patients with either predicted long or short survival might improve the treatment planning and ameliorate the clinical outcome.

KEY POINTS

A CIN score as a new quantified survival indicator is developed for survival prediction of patients with spine metastasis.

IMPORTANCE OF THE STUDY

The skeletal system is the third most common metastatic site of most cancers. About 19.7% of the terminal-stage cancer patients with metastasis diagnosed during 2015–2018 survived more than 2 years, and over half of them were still alive until the last follow-up in our center. However, the widely used revised Tokuhashi score lacks accuracy in prognostic prediction. About 50% of the patients with a Tokuhashi score under 8 survived for more than 1 year but were tortured by both physical suffering and mental desperation. In this study, we present a CIN score as a new quantified indicator for the survival prediction of such patients. The CIN score is a more objective set of criteria that can reflect both molecular characteristics of all cancer subclones and tumor burden instead of traditional symptoms and clinically oriented scores. Therefore, it can ensure the optimal therapeutic efficiency by helping make a more precise clinical decision.

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: National Genomics Data Center under BioProject number PRJCA006730.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Medical Ethics Committee, Shanghai Changzheng Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JX, SC, XY, and NZ developed the concept and designed the experiments. SC, MY, NZ, YX, WWe, DJ, JJ, YL, ZZ, WX, WWa, ZW, HW, TL, JZ, and XY. arranged the clinical documents and communicated with the patients. SC, MY, and NZ. wrote the article. SC, MY, YX, NZ, and TW performed the molecular biology experiments. MY, SC, and NZ performed the bioinformatics and statistical analyses. JX, SC, MY, NZ, YX, WWe, DJ, JJ, ZZ, LY, and WWa provided patient specimens. SC, MY, and NZ contributed equally as first authors. YD performed re-analysis and verification of the data in the revised article. All authors commented on the article at all stages.

FUNDING

The National Key R & D Programs of the Ministry of Science and Technology (No. 2016YFC0902100); Shanghai Science and Technology Innovation Plan (No. 17411950300 and No. 18ZR1439000); the Foundation for Young Scholars of the Second Military Medical University (No. 2017QN16; No.

2017QN17); Shanghai Sailing Program (No. 18YF-1423000); Youth Doctor Assistance Funds "Pyramid Program" of Shanghai Changzheng Hospital (recipient: MY).

ACKNOWLEDGMENTS

We thank Shanghai Your Science Biotechnology Co., Ltd. for its editorial assistance and sample transportation. We also thank Dr. Ziliang Qian from Suzhou Hongyuan Biotech Co. for the UCAD (ultrasensitive chromosomal aberration detector) technical assistance and providing control data from healthy blood donors of Chinese.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.767340/full#supplementary-material

Supplementary Figure S1 | The survival curve of different carcinomas. The prognosis was compared between high CIN status and low CIN status in each carcinoma.

Supplementary Figure S2 | (A) The comparison of tumor fraction estimation performed by our method and ichorCNA algorithm. **(B)** The correlation of ctDNA% and the estimated CIN score.

Supplementary Figure S3 | (A) The correlation of 100 kb bin size and 200 kb bin size. **(B)** The correlation of 400 kb bin size and 200 kb bin size.

REFERENCES

Ahmed, A. K., Goodwin, C. R., Heravi, A., Kim, R., Abu-Bonsrah, N., Sankey, E., et al. (2018). Predicting Survival for Metastatic Spine Disease: a Comparison of Nine Scoring Systems. *Spine J.* 18, 1804–1814. doi:10.1016/j.spinee.2018.03.011

Ay, C., Dunkler, D., Pirker, R., Thaler, J., Quehenberger, P., Wagner, O., et al. (2012). High D-Dimer Levels Are Associated with Poor Prognosis in Cancer Patients. *HAEMATOLOGICA* 97, 1158–1164. doi:10.3324/haematol.2011.054718

Bartels, R. H. M. A., van der Linden, Y. M., and van der Graaf, W. T. A. (2008). Spinal Extradural Metastasis: Review of Current Treatment Options. *CA: A Cancer J. Clinicians* 58, 245–259. doi:10.3322/ca.2007.0016

Batista, N., Tee, J., Sciubba, D., Sahgal, A., Laufer, I., Weber, M., et al. (2016). Emerging and Established Clinical, Histopathological and Molecular Parametric Prognostic Factors for Metastatic Spine Disease Secondary to Lung Cancer: Helping Surgeons Make Decisions. *J. Clin. Neurosci.* 34, 15–22. doi:10.1016/j.jocn.2016.05.023

Bauer, H., Tomita, K., Kawahara, N., Abdel-Wanis, M. E., and Murakami, H. (2002). Surgical Strategy for Spinal Metastases. *Spine (Phila Pa 1976)* 27, 1124–1126. doi:10.1097/00007632-200205150-00027

Bianchi, D. W., Chudova, D., Sehnert, A. J., Bhatt, S., Murray, K., Prosen, T. L., et al. (2015). Noninvasive Prenatal Testing and Incidental Detection of Occult Maternal Malignancies. *JAMA* 314, 162–169. doi:10.1001/jama.2015.7120

Collado Martín, R., García Palomo, A., de la Cruz Merino, L., Borrega García, P., and Barón Duarte, F. J. (2014). Clinical Guideline SEOM: Cancer of Unknown Primary Site. *CLIN. TRANSL ONCOL.* 16, 1091–1097. doi:10.1007/s12094-014-1244-0

Dharajiya, N. G., Grosu, D. S., Farkas, D. H., McCullough, R. M., Almasri, E., Sun, Y., et al. (2018). Incidental Detection of Maternal Neoplasia in Noninvasive Prenatal Testing. *CLIN. CHEM.* 64, 329–335. doi:10.1373/clinchem.2017.277517

Ellinger, J., Müller, S. C., Stadler, T. C., Jung, A., von Ruecker, A., and Bastian, P. J. (2011). The Role of Cell-free Circulating DNA in the Diagnosis and Prognosis of Prostate Cancer. *Urol. Oncol. Semin. Original Invest.* 29, 124–129. doi:10.1016/j.urolonc.2009.05.010

Gronroos, E., and López-García, C. (2018). Tolerance of Chromosomal Instability in Cancer: Mechanisms and Therapeutic Opportunities. *CANCER RES.* 78, 6529–6535. doi:10.1158/0008-5472.can-18-1958

Jansson, K.-A., and Bauer, H. C. F. (2006). Survival, Complications and Outcome in 282 Patients Operated for Neurological Deficit Due to Thoracic or Lumbar Spinal Metastases. *EUR. SPINE J.* 15, 196–202. doi:10.1007/s00586-004-0870-6

Ji, X., Li, J., Huang, Y., Sung, P.-L., Yuan, Y., Liu, Q., et al. (2019). Identifying Occult Maternal Malignancies from 1.93 Million Pregnant Women Undergoing Noninvasive Prenatal Screening Tests. *GENET. MED.* 21, 2293–2302. doi:10.1038/s41436-019-0510-5

Kawahara, N., Tomita, K., Murakami, H., and Demura, S. (2009). Total *en bloc* spondylectomy for spinal tumors: surgical techniques and related basic background. *Orthop. Clin. North America* 40, 47–63. doi:10.1016/j.ocl.2008.09.004

Li, X., An, B., Zhao, Q., Qi, J., Wang, W., Zhang, D., et al. (2018). Combined Fibrinogen and Neutrophil–Lymphocyte Ratio as a Predictive Factor in Resectable Colorectal Adenocarcinoma. *Cmar* 10, 6285–6294. doi:10.2147/cmar.s161094

Mayrhofer, M., De Laere, B., Whitington, T., Van Oyen, P., Ghysel, C., Ampe, J., et al. (2018). Cell-free DNA Profiling of Metastatic Prostate Cancer Reveals Microsatellite Instability, Structural Rearrangements and Clonal Hematopoiesis. *GENOME MED.* 10, 85. doi:10.1186/s13073-018-0595-5

Mehra, N., Dolling, D., Sumanasuriya, S., Christova, R., Pope, L., Carreira, S., et al. (2018). Plasma Cell-free DNA Concentration and Outcomes from Taxane Therapy in Metastatic Castration-Resistant Prostate Cancer from Two Phase III Trials (FIRSTANA and PROSELICA). *Eur. Urol.* 74, 283–291. doi:10.1016/j.eururo.2018.02.013

Mo, H., Wang, X., Ma, F., Qian, Z., Sun, X., Yi, Z., et al. (2020). Genome-wide Chromosomal Instability by Cell-free DNA Sequencing Predicts Survival in Patients with Metastatic Breast Cancer. *The Breast* 53, 111–118. doi:10.1016/j.breast.2020.07.004

Oellerich, M., Schütz, E., Beck, J., Kanzow, P., Plowman, P. N., Weiss, G. J., et al. (2017). Using Circulating Cell-free DNA to Monitor Personalized Cancer Therapy. *Crit. Rev. Clin. Lab. Sci.* 54, 205–218. doi:10.1080/10408363.2017.1299683

Pelegrini de Almeida, L., Vidaletti, T., Martins de Lima Cecchini, A., Sfreddo, E., Martins de Lima Cecchini, F., and Falavigna, A. (2018). Reliability of Tokuhashi Score to Predict Prognosis: Comparison of 117 Patients. *World Neurosurg.* 111, e1–e6. doi:10.1016/j.wneu.2017.11.033

Sioutos, P. J., Arbit, E., Meshulam, C. F., and Galicich, J. H. (1995). Spinal Metastases from Solid Tumors. Analysis of Factors Affecting Survival. *Cancer* 76, 1453–1459. doi:10.1002/1097-0142(19951015)76:8<1453:aid-cncr2820760824>3.0.co;2-t

Stover, D. G., Parsons, H. A., Ha, G., Freeman, S. S., Barry, W. T., Guo, H., et al. (2018). Association of Cell-free DNA Tumor Fraction and Somatic Copy Number Alterations with Survival in Metastatic Triple-Negative Breast Cancer. *Jco* 36, 543–553. doi:10.1200/jco.2017.76.0033

Sutcliffe, P., Connock, M., Shyangdan, D., Court, R., Kandala, N. B., and Clarke, A. (2013). A Systematic Review of Evidence on Malignant Spinal Metastases: Natural History and Technologies for Identifying Patients at High Risk of Vertebral Fracture and Spinal Cord Compression. *Health Technol. Assess.* 17, 1–274. doi:10.3310/hta17420

Tan, J. H., Tan, K. A., Zaw, A. S., Thomas, A. C., Hey, H. W., Soo, R. A., et al. (2016). Evaluation of Scoring Systems and Prognostic Factors in Patients with Spinal Metastases from Lung Cancer. *Spine (Phila Pa 1976)* 41, 638–644. doi:10.1097/BRS.0000000000001279

Tan, Z., Chan, Y. J. A., Chua, Y. J. K., Rutledge, S. D., Pavelka, N., Cimini, D., et al. (2019). Environmental Stresses Induce Karyotypic Instability in Colorectal Cancer Cells. *MBoC* 30, 42–55. doi:10.1091/mbc.e18-10-0626

Tang, Y., Qu, J., Wu, J., Li, S., Zhou, Y., and Xiao, J. (2015). Metastatic Spinal Cord Compression from Non-small-cell Lung Cancer Treated with Surgery and Adjuvant Therapies. *The J. Bone Jt. Surgery-American Volume* 97, 1418–1425. doi:10.2106/jbjs.n.01124

Tang, Y., Qu, J., Wu, J., Liu, H., Chu, T., Xiao, J., et al. (2016). Effect of Surgery on Quality of Life of Patients with Spinal Metastasis from Non-small-cell Lung Cancer. *J. Bone Jt. Surg.* 98, 396–402. doi:10.2106/jbjs.o.00629

Tokuhashi, Y., Matsuzaki, H., Oda, H., Oshima, M., and Ryu, J. (2005). A Revised Scoring System for Preoperative Evaluation of Metastatic Spine Tumor Prognosis. *Spine (Phila Pa 1976)* 30, 2186–2191. doi:10.1097/01.brs.0000180401.06919.a5

Tokuhashi, Y., Matsuzaki, H., Toriyama, S., Kawano, H., and Ohsaka, S. (1990). Scoring System for the Preoperative Evaluation of Metastatic Spine Tumor Prognosis. *Spine (Phila Pa 1976)* 15, 1110–1113. doi:10.1097/00007632-199011010-00005

Tomita, K., Kawahara, N., Kobayashi, T., Yoshida, A., Murakami, H., and Akamaru, T. (2001). Surgical Strategy for Spinal Metastases. *Spine (Phila Pa 1976)* 26, 298–306. doi:10.1097/00007632-200102010-00016

Tran, N. H., Kisiel, J., and Roberts, L. R. (2021). Using Cell-free DNA for HCC Surveillance and Prognosis. *JHEP Rep.* 3, 100304. doi:10.1016/j.jhepr.2021.100304

van der Linden, Y. M., Dijkstra, S. P. D. S., Vonk, E. J. A., Marijnen, C. A. M., and Leer, J. W. H. (2005). Prediction of Survival in Patients with Metastases in the Spinal Column. *Cancer* 103, 320–328. doi:10.1002/cncr.20756

Vanderstichele, A., Busschaert, P., Smeets, D., Landolfo, C., Van Nieuwenhuysen, E., Leunen, K., et al. (2017). Chromosomal Instability in Cell-free DNA as a Highly Specific Biomarker for Detection of Ovarian Cancer in Women with Adnexal Masses. *CLIN. CANCER RES.* 23, 2223–2231. doi:10.1158/1078-0432.ccr-16-1078

Varadhachary, G. R., and Raber, M. N. (2014). Cancer of Unknown Primary Site. *N. Engl. J. Med.* 371, 757–765. doi:10.1056/NEJMra1303917

Yang, M., Xu, W., Liu, T., Yang, X., Wang, P., Wu, S., et al. (2019). Development and Validation of a Novel Survival Prediction Model in Patients with Spinal Metastasis from Non-small Cell Lung Cancer. *Spine (Phila Pa 1976)* 44, 246–257. doi:10.1097/BRS.0000000000002816

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: 21 December 2021
doi: 10.3389/fcell.2021.797984

An Immune Panel Signature Predicts Prognosis of Lung Adenocarcinoma Patients and Correlates With Immune Microenvironment

Yuan Zhou[1,2], *Lu Tang*[3], *Yuqiao Chen*[1,2], *Youyu Zhang*[1,2] *and Wei Zhuang*[1,2]*

[1]*Department of Thoracic Surgery, Xiangya Hospital of Central South University, Changsha, China,* [2]*National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University,Changsha, China,* [3]*Department of Anesthesiology, Xiangya Hospital of Central South University, Changsha, China*

OPEN ACCESS

Edited by:
D. P. Kreil,
Boku University, Austria

Reviewed by:
Xianghui Fu,
Sichuan University, China
Ranjie Xu,
Rutgers, The State University of New Jersey–Busch Campus, United States
Weiyi Fang,
Southern Medical University, China

***Correspondence:**
Wei Zhuang
zhuangwei@csu.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: *19 October 2021*
Accepted: *26 November 2021*
Published: *21 December 2021*

Citation:
Zhou Y, Tang L, Chen Y, Zhang Y and Zhuang W (2021) An Immune Panel Signature Predicts Prognosis of Lung Adenocarcinoma Patients and Correlates With Immune Microenvironment. Front. Cell Dev. Biol. 9:797984. doi: 10.3389/fcell.2021.797984

Background: Lung cancer, especially lung adenocarcinoma (LUAD) with high incidence, seriously endangers human life. The immune microenvironment is one of the malignant foundations of LUAD, but its impact at the molecular level is incompletely understood.

Method: A total of 34 LUAD samples from Xiangya Hospital were collected for immune oncology (IO) profiling. Univariate Cox analysis was performed to profile prognostic immune genes based on our immune panel sequencing data. The least absolute shrinkage and selection operator (LASSO) algorithm was applied to construct a risk signature. The cut-off threshold of risk score was determined using X-tile software. Kaplan–Meier survival curves and receiver operating characteristic (ROC) curves were employed to examine the performance of this risk signature for predicting prognosis. The immune infiltration was estimated using a single-sample gene set enrichment analysis (ssGSEA) algorithm.

Result: Thirty-seven immune genes were profiled to be significantly correlated with the progression-free survival (PFS) in our cohort. Among them, *BST2*, *KRT7*, *LAMP3*, *MPO*, *S100A8*, and *TRIM29* were selected to construct a risk signature. Patients with a higher risk score had a significantly shorter PFS ($p = 0.007$). Time-dependent ROC curves indicated that our risk signature had a robust performance in accurately predicting survival. Specifically, the 6-, 12-, and 18-month area under curve (AUC) was 0.800, 0.932, and 0.912, respectively. Furthermore, the risk signature was positively related to N stage, tumor stage, and tumor malignancy. These results were validated using two external cohorts. Finally, the risk signature was **s**ignificantly and uniquely correlated with abundance of neutrophil.

Conclusion: Our study revealed an immune panel-based signature that could predict the prognosis of LUAD patients and was associated with the infiltration of neutrophils.

Keywords: lung adenocarcinoma, progression-free survival, immune panel sequencing, gene signature, metastasis

INTRODUCTION

Lung cancer is nowadays the leading cause of cancer-related morbidity and mortality worldwide, accounting for nearly 20% of cancer deaths (Bray et al., 2018). Approximately 85% of patients have a group of histological subtypes collectively termed as non-small cell lung cancer (NSCLC), of which lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) are the most common subtypes (Herbst et al., 2018). Since they are usually asymptomatic at an early stage, most lung cancers (61%) are diagnosed at stage III or IV, with only 21% at stage I. More importantly, advanced lung cancer confers an extremely poor prognosis. In brief, the 5-year relative survival rate for stage I patients is 57% and decreases to 29% for patients with stage III lung cancers (Miller et al., 20192019).

Although improvements in surgical techniques and chemoradiotherapy, as well as individualized treatment regimens with tyrosine kinase inhibitors (TKIs) as the mainstay, have led to inspiring clinical advances (Pao and Girard, 2011; Miller et al., 20192019), there are still some populations that exhibit limited responses or acquire resistance. In addition, previous studies have explored more factors to predict the patient's prognosis in a more accurate manner, such as vascular spreading (Gabor et al., 2004) and lymphatic spreading (Popper, 2016). But the prediction of metastasis as well as the patient's prognosis at the clinical level is far from satisfactory. Thus, how to predict patient prognosis (herein specifically refers to the metastasis of lung cancer) is an urgent issue that remains unaddressed for clinicians. Therefore, we need to focus on the biological process and intrinsic malignant basis of LUAD at the molecular level in the hope of establishing good prognostic indicators or targeting regimens.

In 2018, the immune landscape of cancer has been depicted by conducting a comprehensive immunogenomic analysis of more than 10,000 samples across 33 cancer types (Thorsson et al., 2018). Samples could be well clustered based on the immune profiles and exhibit different molecular profiles and distinct survivals. Since then, efforts have been devoted to analyzing tumor–immune cell interaction in specific malignancies at the transcriptional level. However, we noticed that although several studies have applied bioinformatics in the context of lung cancers, most of them focused on publicly available databases such as classical The Cancer Genome Atlas (TCGA) database, making the conclusions limited and ungeneralizable.

It is well known that cancer cells can functionally construct a tumor microenvironment (TME) by regulating the reprogramming of surrounding cells, which play a decisive role in tumor survival and progression. Immune cells are important components of TME and play a critical role in this process. Both innate immune cells (macrophages, neutrophils, dendritic cells, and natural killer cells) and adaptive immune cells (T and B cells) contribute to tumor progression within the TME context. Dialogs between cancer cells and surrounding immune infiltrates ultimately result in a complexed network that promotes tumor growth and metastasis (Hinshaw and Shevde, 2019). Particularly, the TME has largely affected the efficacy of immunotherapies based on immune checkpoint blockade (He et al., 2015; Sharma et al., 2017; Herbst et al., 2018), which highlights the importance and the urgent need of deeply understanding how TME orchestrates the therapeutic and prognostic outcomes. Therefore, we employed our sequencing data to employ the prognostic immune genes and construct a risk signature for prognosis prediction and potential therapy target.

MATERIALS AND METHODS

Data Extraction

We enrolled 34 LUAD samples in our institute from December 2014 to December 2016. The patients who met the following inclusion criteria were included: 1) postoperative pathology confirmed advanced lung cancer, and the pathological stage was T1–4, N1–2, M0; 2) no neoadjuvant therapy was administrated before surgery; and 3) disease progression events such as local recurrence or distant metastasis occurred during postoperative follow-up. Finally, a total of 34 corresponding samples were collected, and these samples were subsequently sequenced by Genecast Biotechnology, Beijing, China. Specifically, RNA immune oncology (IO) profiling was performed to quantify 395 IO associated genes related to tumor markers, basic signaling pathways, tumor-specific antigens, immune responses, infiltrating immune cells, and housekeeping (HK) genes in human solid cancers (**Supplementary Table S1**).

All of the patients (or their family representatives) have signed written inform consent, and this study was approved by the Ethics Committee of Xiangya Hospital, Central South University. The clinical information of included samples is summarized in **Supplementary Table S2**.

RNA-seq data and clinical information were extracted for validation from TCGA (https://portal.gdc.cancer.gov/) and the Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo) databases. A total of 510 and 443 patients with LUAD were extracted from TCGA-LUAD and GSE68465 datasets (Shedden et al., 2008), respectively.

Identification of Prognostic Immune Genes

Univariate Cox analysis was performed on immune genes obtained in our immune panel sequencing data to profile those with significant correlations with progression-free survival (PFS) of patients with LUAD.

Construction and Validation of Risk Signature

The least absolute shrinkage and selection operator (LASSO) algorithm was applied to refine the most representative of the prognostic immune genes and assign them corresponding coefficients (Boyd et al., 2011). The genes at the smallest lambda value were included to build the signature. And the risk score for each individual in the training and validation cohorts was calculated using the following formula: Risk $score = \sum_{i}^{n}(\beta_i \times x_i)$. n was the number of genes. x_i represented the mRNA expression level of each model gene, and β_i meant the coefficient.

The cut-off threshold to divide patients into high-risk and low-risk groups was determined using X-tile software (Camp et al., 2004). Kaplan–Meier survival curves and time-dependent

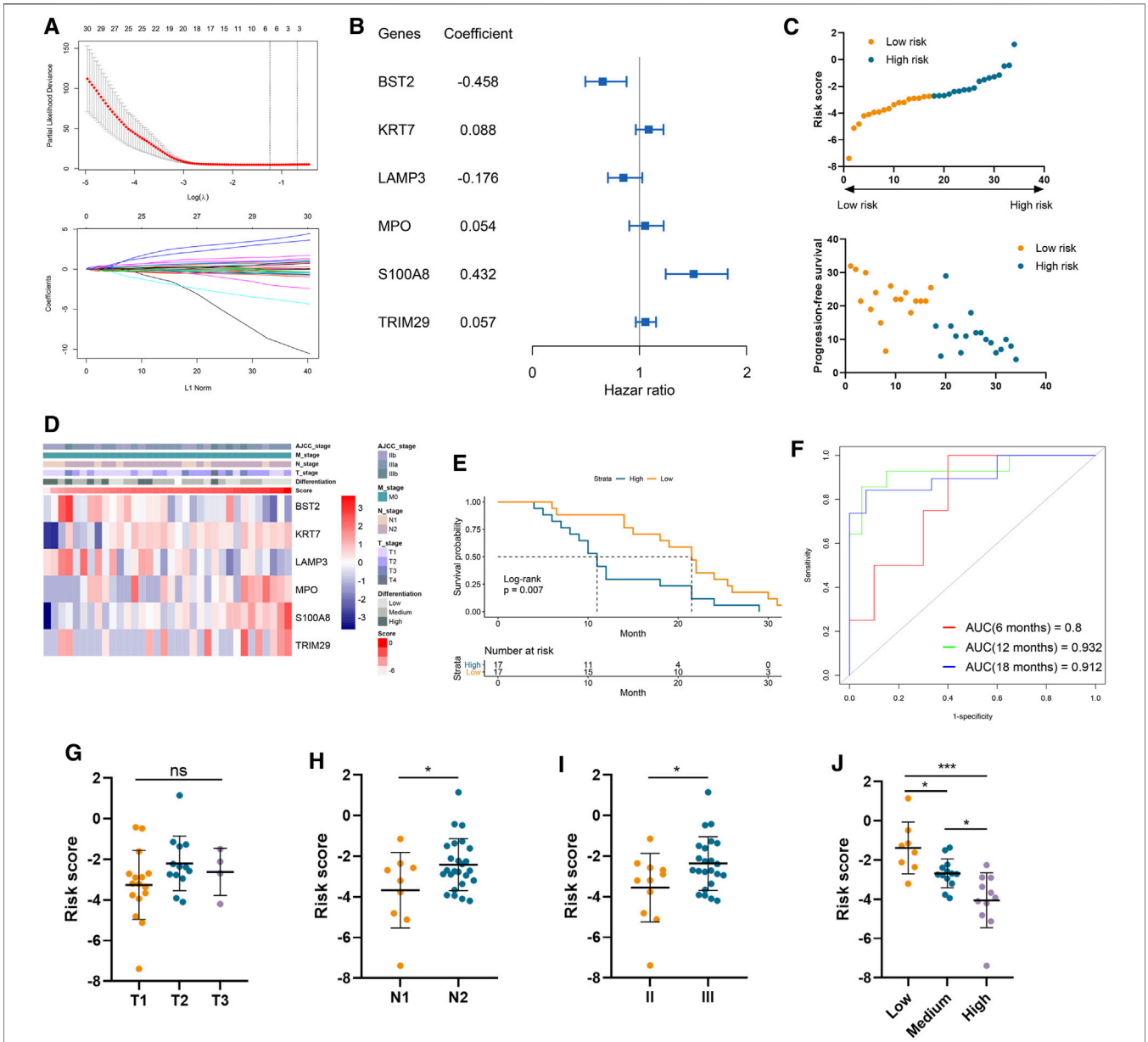


FIGURE 1 | Construction of risk signature to predict the metastasis of lung adenocarcinoma. **(A)** LASSO analysis of prognostic genes with the minimum lambda value. **(B)** The coefficient of six genes to construct the risk signature. **(C)** Risk score of each sample was calculated based on the coefficient and expression of six genes. **(D)** Heatmap of the expression of six genes in samples with increasing risk score. **(E)** Kaplan–Meier analysis of high-risk and low-risk patients. **(F)** Time-dependent ROC analysis of risk score to predict the 6-, 12-, and 18-month survival. **(G–J)** The risk score in tumors with different T stages **(G)**, N stages **(H)**, AJCC stages **(I)**, and differentiation degree **(J)**.

receiver operating characteristic (ROC) curves were employed to examine the performance of this risk signature for predicting prognosis. Furthermore, the relationship between risk score and clinicopathological features, including American Joint Committee on Cancer (AJCC) tumor stage and degree of tumor differentiation, was also assessed.

Immune Infiltration

Gene set variation analysis (GSVA) analysis was conducted using the "GSVA" R package (Hänzelmann et al., 2013). The immune infiltration of multiple cell types was estimated using single-sample gene set enrichment analysis (ssGSEA) algorithm.

Statistical Analysis

Data analysis and visualization were conducted using GraphPad Prism version 8.0.1 and R language version 3.6.3. The cut-off value for risk score was determined using X-tile software. Student's t-test and Wilcoxon test were used to compare the difference between two groups. One-way analysis of variance (ANOVA) test and Kruskal–Wallis test were used to compare the

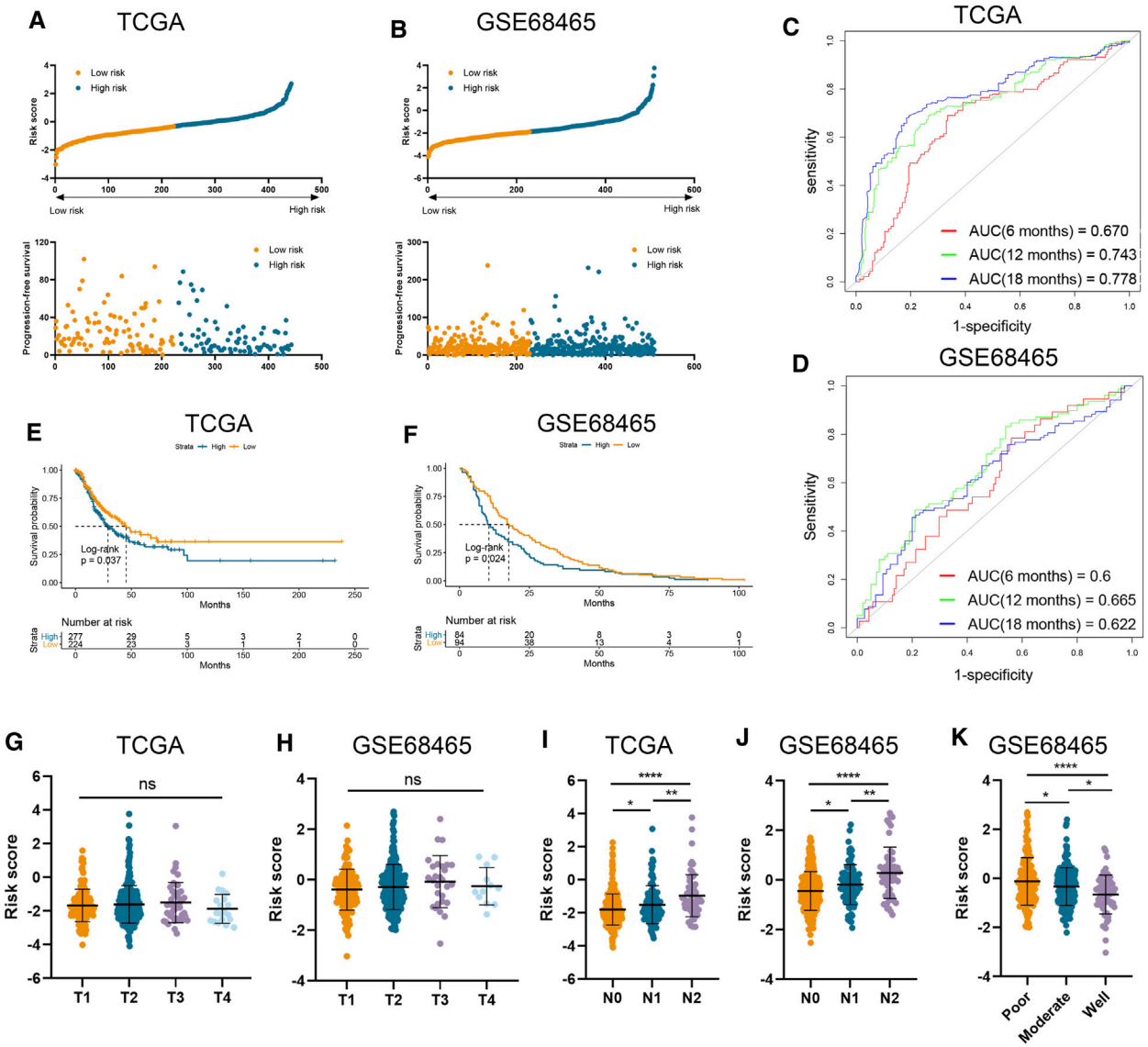


FIGURE 2 | Validation of risk signature in two independent datasets. **(A,B)** Risk score was calculated in the TCGA **(A)** and GSE68465 **(B)** datasets. **(C,D)** Time-dependent ROC analysis of risk score to predict the 6-, 12-, and 18-month survival in TCGA **(C)** and GSE68465 **(D)** datasets. **(E,F)** Kaplan–Meier analysis of progression-free survival of high-risk and low-risk patients in TCGA **(E)** and GSE68465 **(F)** datasets. **(G–J)** The risk score in tumors with different T stages, and N stages in two datasets. **(K)** The risk score in tumors with different differentiation degrees in GSE68465.

difference among more than two groups. Kaplan–Meier analysis was used for survival analysis. A p-value <0.05 was considered statistically significant.

RESULTS

Immune Gene-Based Risk Signature Predicts Prognosis and Correlates With Malignant Clinical Features

A total of 37 genes were profiled to be significantly correlated with the PFS in our cohort involving 34 LUAD samples (**Supplementary Table S3**). And six of them, including bone marrow stromal cell antigen 2 (*BST2*), keratin 7 (*KRT7*), lysosomal associated membrane protein 3 (*LAMP3*), myeloperoxidase (*MPO*), S100 calcium binding protein A8 (*S100A8*), and tripartite motif containing 29 (*TRIM29*), were selected by the LASSO model to construct a risk signature (**Figure 1A**). Among them, *S100A8*, *TRIM29*, *MPO*, and *KRT7* were significant protective factors for LUAD patients, while LAMP3 and BST2 were risk factors in our cohort (**Supplementary Figure S1**). More importantly, these model genes exhibited consistent prognostic values in two external validation cohorts (**Supplementary Figure S2**). Their coefficients are displayed in **Figure 1B** and the distributions of

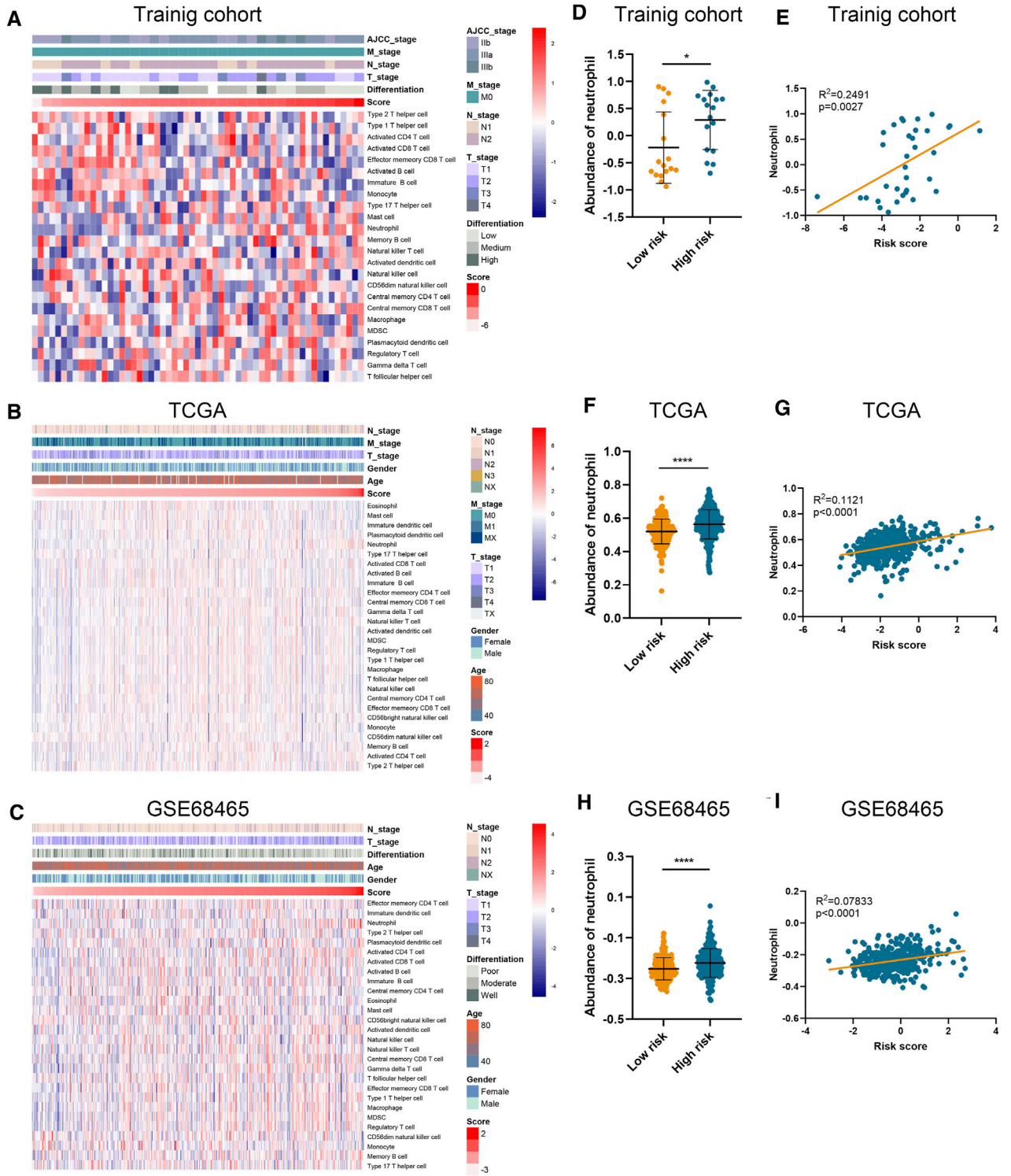


FIGURE 3 | Association between risk score and immune infiltration. **(A–C)** The abundance of different immune cells in the training cohort **(A)**, TCGA **(B)**, and GSE68465 **(C)** datasets. **(D–I)** The abundance of neutrophil in high-risk and low-risk patients and its correlation with risk score in training cohort **(D,E)**, TCGA **(F,G)**, and GSE68465 **(H,I)** datasets.

risk scores and PFS status were illustrated in **Figure 1C**. The expression pattern of *BST2* as well as *LAMP3* was opposite to that of the risk score, and conversely, the same trend was observed for the four remaining model genes and the risk score (**Figure 1D**). Importantly, patient prognosis differed between the two risk groups, as patient with a higher risk score significantly demonstrated shorter PFS (Log-rank $p = 0.007$; **Figure 1E**). Furthermore, time-dependent ROC curves indicated that our risk signature had a robust performance in accurately predicting survival. Specifically, the 6-, 12-, and 18-month area under curve (AUC) was 0.800, 0.932, and 0.912, respectively (**Figure 1F**). Regarding clinical characteristics, risk score did not correlate with tumor T stage but was positively related to N stage, AJCC stage, and tumor malignancy (as negatively related to the degree of tumor differentiation). (**Figures 1G–J**).

Validation of the Risk Signature Using Independent Cohorts With Large Sample Size

The clinical and RNA sequencing data of TCGA-LUAD and GSE68465 cohorts with 510 and 443 LUAD patients were extracted for external validation. Similarly, patients with lower risk scores presented with longer PFS time than those with higher risk scores (**Figures 2A, B**). The performance of our risk signature in predicting prognosis was still excellent under validation, as the AUC was all above 0.6 in two cohorts (**Figures 2C, D**). Besides, Kaplan–Meier analyses also revealed significantly poorer outcomes in the high-risk groups (TCGA-LUAD log-rank $p = 0.037$; GSE68465 log-rank $p = 0.024$; **Figures 2E, F**). There was no significant correlation between risk signature and tumor T stage, which was consistent with the result in our training cohort (**Figures 2G, H**). And LUAD patients with higher AJCC N stage and lower degree of tumor differentiation had higher risk scores in both validation cohorts (**Figures 2I–K**).

Risk Signature Correlates With Infiltrating Neutrophil in Tumor Microenvironment

Using ssGSEA algorithm, we obtained the abundances of 24 infiltrating cell types (**Supplementary Table S4**). Significant and unique correlation was detected between our risk signature and neutrophil (**Figures 3A–C**), as LUAD samples in the high-risk group possessed higher abundance of neutrophil than those in the low-risk group (**Figures 3D, F, H**). Moreover, the risk score was correlated with the abundance of neutrophil in the training cohort ($p = 0.0027$; **Figure 3E**), TCGA-LUAD cohort ($p < 0.0001$; **Figure 3G**), and GSE68465 cohort ($p < 0.0001$; **Figure 3I**). However, the risk score is not relevant to other types of immune cells.

DISCUSSION

Prodigious effort has been devoted to explore the underlying mechanisms of lung cancer at the molecular level, but the current understanding of TME and prognostic factors is still far from satisfactory. In this study, we first screened out the significant prognostic gene sets for LUAD in immune panel profiles and further investigated them. We constructed a novel risk signature using representative immune genes with significant prognostic value for LUAD. Gratifyingly, our six-gene risk signature allowed robust risk stratification as high-risk patients had a significantly worse prognosis ($p = 0.007$). And this model presented a predictive power with high specificity (AUC > 0.85). To our surprise, the trained model showed higher accuracy of 12- and 18-month survival prediction compared with the result from the 6-month survival prediction, and the superiority was kept when the prediction was done using the test sets. This is likely because the immune-related genetic characteristics have increasingly amplified with certain and consistent proportions during cancer progression.

Events for PFS refer to tumor progression (recurrence, enlargement, or metastasis) or death from any cause. And for LUAD, the length of PFS strongly depends on patients' metastatic events. Our risk signature was found to be significantly correlated with tumor N stage, AJCC stage, and tumor malignant property, which are independent predictors of lung cancer metastasis (Lim et al., 2018). This also, to some extent, explains the sound ability of our model in predicting metastasis.

A key finding in our study is that the established risk signature is significantly correlated with the abundance of neutrophil in all the cohorts. Tumor-associated neutrophils predict poor overall survival in many cancer types, with their location in the tumor and specific markers being important deferential determinants (Shaul and Fridlender, 2019). For example, massive expression of inflammation-related genes is transcriptionally activated by epigenetic remodeling in advanced clear cell renal cell carcinoma, which is related to metastasis in a neutrophil-dependent manner (Nishida et al., 2020). Meanwhile, neutrophils in lung mesenchyme are essential for breast cancer lung metastasis (Wculek and Malanchi, 2015; Li et al., 2020). More importantly, using single-cell RNA sequencing (scRNA-seq) to map tumor-infiltrating myeloid cells in non-small-cell lung cancer patients, neutrophils were found to be a key regulator of cancer growth across individuals and species (Zilionis et al., 2019). Combining these findings, we propose that our risk signature correlates with neutrophil population intrinsically and reflects the risk of metastasis in LUAD patients.

As for the functions of proteins encoded by signature genes, KRT7 is found to be tightly linked to cancer metastasis in colorectal cancer (Yu et al., 2017) and breast cancer (Chen et al., 2021). LAMP3 is implied to be involved in metastasis in esophageal squamous cell carcinoma (Huang et al., 2020) and uterine cervical cancer (Kanao et al., 2005) but in different pathways. Neutrophils play a vital role in both chemically mediating inflammatory response through MPO and biologically promoting metastasis during inflammation triggered by the environmental stimuli (Tang et al., 2020). Moreover, S100A8 can induce the activation of MPO, and novel monoclonal antibody against it efficiently prevents lung cancer metastasis (Kinoshita et al., 2019). TRIM29 is a selective regulator of the activation of alveolar macrophages and the production of proinflammatory cytokines in the lungs and is

found to mediate lung squamous cell carcinoma cell metastasis by regulating autophagic degradation of E-cadherin.

There are some limitations in this study. First, although we validated the results using external cohorts, the sample size in our cohort is small, reducing the credibility and generalizability of the conclusion. Second, there is no experiment verification of our proposal, for instance, the relationship between model genes and neutrophil. Third, we cannot demonstrate that the affected survival is due to neutrophil population, which requires further *in vivo* analysis.

CONCLUSION

Using the immune panel sequencing of our samples, we profiled prognostic immune genes and constructed a risk signature, which confers an excellent and robust predictive power. And this signature is associated with tumor stage, malignant property, and, more importantly, the abundance of neutrophil.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://portal.gdc.cancer.gov/ and https://www.ncbi.nlm.nih.gov/geo/. The raw data supporting the conclusions of this article will be available on reasonable request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committe of Xiangya Hospital, Central South University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

WZ conceived and supervised the study. YZo, LT, YC, and YZa collected the clinical specimens. YZo performed the data analysis and drafted the manuscript. All authors have read and approved the final version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.797984/full#supplementary-material.

Supplementary Figure S1 | Kaplan–Meier analysis of six genes in the training cohort.

Supplementary Figure S2 | Kaplan–Meier analysis of six genes in the TCGA and GSE68465 datasets.

Supplementary Table S1 | The list of immune panel genes.

Supplementary Table S3 | The list of 37 prognostic immune genes.

REFERENCES

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011). Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers2011 5/23/2011. *Foundations Trends Machine Learn.* 3, 1–122. doi:10.1561/2200000016

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. Clinicians* 68 (6), 394–424. doi:10.3322/caac.21492

Camp, R. L., Dolled-Filhart, M., and Rimm, D. L. (2004). X-tile. *Clin. Cancer Res.* 10 (21), 7252–7259. doi:10.1158/1078-0432.ccr-04-0713

Chen, F., Chen, Z., Guan, T., Zhou, Y., Ge, L., Zhang, H., et al. (2021). N6-Methyladenosine Regulates mRNA Stability and Translation Efficiency of KRT7 to Promote Breast Cancer Lung Metastasis. *Cancer Res.* 81 (11), 2847–2860. doi:10.1158/0008-5472.can-20-3779

Gabor, S., Renner, H., Popper, H., Anegg, U., Sankin, O., Matzi, V., et al. (2004). Invasion of Blood Vessels as Significant Prognostic Factor in Radically Resected T1-3N0M0 Non-small-cell Lung Cancer. *Eur. J. Cardio-Thoracic Surg.* 25 (3), 439–442. doi:10.1016/j.ejcts.2003.11.033

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *bmc bioinformatics* 14 (1), 7. doi:10.1186/1471-2105-14-7

He, J., Hu, Y., Hu, M., and Li, B. (2015). Development of PD-1/pd-L1 Pathway in Tumor Immune Microenvironment and Treatment for Non-small Cell Lung Cancer. *Sci. Rep.* 5 (1), 13110. doi:10.1038/srep13110

Herbst, R. S., Morgensztern, D., and Boshoff, C. (2018). The Biology and Management of Non-small Cell Lung Cancer. *nature* 553 (7689), 446–454. doi:10.1038/nature25183

Hinshaw, D. C., and Shevde, L. A. (2019). The Tumor Microenvironment Innately Modulates Cancer Progression. *Cancer Res.* 79 (18), 4557–4566. doi:10.1158/0008-5472.can-18-3962

Huang, F., Ma, G., Zhou, X., Zhu, X., Yu, X., Ding, F., et al. (2020). Depletion of LAMP3 Enhances PKA-Mediated VASP Phosphorylation to Suppress Invasion and Metastasis in Esophageal Squamous Cell Carcinoma. *Cancer Lett.* 479, 100–111. doi:10.1016/j.canlet.2020.03.014

Kanao, H., Enomoto, T., Kimura, T., Fujita, M., Nakashima, R., Ueda, Y., et al. (2005). Overexpression ofLAMP3/TSC403/DC-LAMPPromotes Metastasis in Uterine Cervical Cancer. *Cancer Res.* 65 (19), 8640–8645. doi:10.1158/0008-5472.can-04-4112

Kinoshita, R., Sato, H., Yamauchi, A., Takahashi, Y., Inoue, Y., Sumardika, I. W., et al. (2019). Newly Developed anti-S100A8/A9 Monoclonal Antibody Efficiently Prevents Lung Tropic Cancer Metastasis. *Int. J. Cancer* 145 (2), 569–575. doi:10.1002/ijc.31982

Li, P., Lu, M., Shi, J., Gong, Z., Hua, L., Li, Q., et al. (2020). Lung Mesenchymal Cells Elicit Lipid Storage in Neutrophils that Fuel Breast Cancer Lung Metastasis. *Nat. Immunol.* 21 (11), 1444–1455. doi:10.1038/s41590-020-0783-5

Lim, W., Ridge, C. A., Nicholson, A. G., and Mirsadraee, S. (2018). The 8th Lung Cancer TNM Classification and Clinical Staging System: Review of the Changes and Clinical Implications. *Quant. Imaging Med. Surg.* 8 (7), 709–718. doi:10.21037/qims.2018.08.02

Miller, K. D., Nogueira, L., Mariotto, A. B., Rowland, J. H., Yabroff, K. R., Alfano, C. M., et al. (20192019). Cancer Treatment and Survivorship Statistics, 2019. *CA A. Cancer J. Clin.* 69 (5), 363–385. doi:10.3322/caac.21565

Nishida, J., Momoi, Y., Miyakuni, K., Tamura, Y., Takahashi, K., Koinuma, D., et al. (2020). Epigenetic Remodelling Shapes Inflammatory Renal Cancer and Neutrophil-dependent Metastasis. *Nat. Cel Biol* 22 (4), 465–475. doi:10.1038/s41556-020-0491-2

Pao, W., and Girard, N. (2011). New Driver Mutations in Non-small-cell Lung Cancer. *Lancet Oncol.* 12 (2), 175–180. doi:10.1016/s1470-2045(10)70087-5

Popper, H. H. (2016). Progression and Metastasis of Lung Cancer. *Cancer Metastasis Rev.* 35 (1), 75–91. doi:10.1007/s10555-016-9618-0

Sharma, P., Hu-Lieskovan, S., Wargo, J. A., and Ribas, A. (2017). Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy. *cell* 168 (4), 707–723. doi:10.1016/j.cell.2017.01.017

Shaul, M. E., and Fridlender, Z. G. (2019). Tumour-associated Neutrophils in Patients with Cancer. *Nat. Rev. Clin. Oncol.* 16 (10), 601–620. doi:10.1038/s41571-019-0222-4

Shedden, K., Shedden, K., Taylor, J. M., Enkemann, S. A., Tsao, M. S., Yeatman, T. J., et al. (2008). Gene Expression-Based Survival Prediction in Lung Adenocarcinoma: a Multi-Site, Blinded Validation Study. *Nat. Med.* 14 (8), 822–827. doi:10.1038/nm.1790

Tang, L., Wang, Z., Mu, Q., Yu, Z., Jacobson, O., Li, L., et al. (2020). Targeting Neutrophils for Enhanced Cancer Theranostics. *Adv. Mater.* 32 (33), e2002739. doi:10.1002/adma.202002739

Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Ou Yang, T. H., et al. (2018). The Immune Landscape of Cancer. *immunity* 51 (4), 411–412. doi:10.1016/j.immuni.2019.08.004

Wculek, S. K., and Malanchi, I. (2015). Neutrophils Support Lung Colonization of Metastasis-Initiating Breast Cancer Cells. *nature* 528 (7582), 413–417. doi:10.1038/nature16140

Yu, T., Guo, F., Yu, Y., Sun, T., Ma, D., Han, J., et al. (2017). Fusobacterium Nucleatum Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. *cell* 170 (3), 548–563. doi:10.1016/j.cell.2017.07.008

Zilionis, R., Engblom, C., Pfirschke, C., Savova, V., Zemmour, D., Saatcioglu, H. D., et al. (2019). Single-Cell Transcriptomics of Human and Mouse Lung Cancers Reveals Conserved Myeloid Populations across Individuals and Species. *immunity* 50 (5), 1317–e10. doi:10.1016/j.immuni.2019.03.009

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: 05 January 2022
doi: 10.3389/fcell.2021.794144

Single-Cell Transcriptomes Combining with Consecutive Genomics Reveal Clonal Evolution and Gene Regulatory Networks in Relapsed and Refractory Multiple Myeloma

Jiadai Xu[1], *Yue Wang*[1], *Zheng Wei*[1], *Jingli Zhuang*[1], *Jing Li*[1], *Yifeng Sun*[1], *Liang Ren*[1], *Yawen Wang*[1], *Panpan Li*[1], *Shiyang Gu*[1], *Yian Zhang*[1], *Jifeng Jiang*[1], *Chen Chen*[1], *Yu Zhang*[2] *and Peng Liu*[1]*

[1]Department of Hematology, Zhongshan Hospital, Fudan University, Shanghai, China, [2]Department of Cardiovascular Surgery, Renji Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, China

This study attempted to investigate how clonal structure evolves, along with potential regulatory networks, as a result of multiline therapies in relapsed/refractory multiple myeloma (RRMM). Eight whole exome sequencing (WES) and one single cell RNA sequencing (scRNA-seq) were performed in order to assess dynamic genomic changes in temporal consecutive samples of one RRMM patient from the time of diagnosis to death (about 37 months). The 63-year-old female patient who suffered from MM (P1) had disease progression (PD) nine times from July 2017 [newly diagnosed (ND)] to Aug 2020 (death), and the force to drive branching-pattern evolution of malignant PCs was found to be sustained. The mutant-allele tumor heterogeneity (MATH) and tumor mutation burden (TMB) initially exhibited a downward trend, which was then upward throughout the course of the disease. Various somatic single nucleotide variants (SNVs) that had disappeared after the previous treatment were observed to reappear in later stages. Chromosomal instability (CIN) and homologous recombination deficiency (HRD) scores were observed to be increased during periods of all progression, especially in the period of extramedullary plasmacytoma. Finally, in combination with WES and scRNA-seq of P1-PD9 (the nineth PD), the intro-heterogeneity and gene regulatory networks of MM cells were deciphered. As verified by the overall survival of MM patients in the MMRF CoMMpass and GSE24080 datasets, RUNX3 was identified as a potential driver for RRMM.

Keywords: multiple myeloma, clonal evolution, relapsed, refractory, heterogeneity

OPEN ACCESS

Edited by:
Lu Xie,
Shanghai Center for Bioinformation Technology, China

Reviewed by:
Jiantao Shi,
Shanghai Institute of Biochemistry and Cell Biology (CAS), China
Ernesto Diaz-Flores,
University of California, San Francisco, United States

***Correspondence:**
Peng Liu
liu.peng@zs-hospital.sh.cn

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

Received: *13 October 2021*
Accepted: *16 November 2021*
Published: *05 January 2022*

Citation:
Xu J, Wang Y, Wei Z, Zhuang J, Li J, Sun Y, Ren L, Wang Y, Li P, Gu S, Zhang Y, Jiang J, Chen C, Zhang Y and Liu P (2022) Single-Cell Transcriptomes Combining with Consecutive Genomics Reveal Clonal Evolution and Gene Regulatory Networks in Relapsed and Refractory Multiple Myeloma. Front. Cell Dev. Biol. 9:794144. doi: 10.3389/fcell.2021.794144

INTRODUCTION

Multiple myeloma (MM) is a plasma cell malignancy that is characterized by highly intra-clonal heterogeneity. Despite emerging novel therapeutic strategies, response to treatment and final outcomes continue to vary in MM patients, and the disease is considered to be mostly incurable. During MM progression, therapeutic pressure may bestow a selective advantage on sub-clonal expansion through which the fittest subclone will dominate. In turn, the presence of sub-clonal mutations may affect the subsequent efficacy of the targeted therapy, resulting in a vicious cycle. The development of anti-MM therapy requires an understanding of driver genetic alterations, gene regulatory networks as well as an appreciation of the evolutionary processes of malignant plasma cells (PCs).

Recently, large-scale genomic sequencing studies have provided deep insights into the genetic landscape of MM (Robiou du Pont et al., 2017; Walker et al., 2018), which may provide potential targeted candidates for personalized therapy. However, most of these studies obtained only a single, or at most two, samples from each individual patient. Little is known about the temporal clonal evolutionary processes in refractory or multiple relapsed MM. In 2012, three landmark studies provided insight into the intra-clonal heterogeneity early at the diagnosis and different stages of MM after relapse (Egan et al., 2012; Keats et al., 2012; Walker et al., 2012), suggesting a Darwinian model of tumor evolution in MM.

In this study, whole exome sequencing (WES) and single cell RNA-sequencing (scRNA-seq) are simultaneously applied in order to determine dynamic genomic changes among eight temporal consecutive samples of one relapsed and refractory MM (RRMM) patient from the time of diagnosis to death (about 37 months). By searchingdatasets from the Multiple Myeloma Research Foundation (MMRF) CoMMpass (Clinical Outcomes in MM to Personal Assessment of Genetic Profile) study (n = 766) and Gene Expression Omnibus (accession GSE24080, n = 559), RUNX3, a member of runt-related transcription factor family (referred to as RUNXs), is shown to serve as a potential driver for RRMM or secondary plasma cell leukemia (PCL). Accordingly, from genomics to single cell transcriptomes, the manner in which the clonal structure and regulatory networks evolved under the pressure of multiline therapies was investigated.

MATERIALS AND METHODS

Patients and Samples

A 63-year-old female patient with MM (P1) suffered progressive disease (PD) nine times from July 12, 2017 (new diagnosis (ND)) to August 8, 2020 (death) at Zhongshan Hospital, Fudan University. Since she also suffered from coronary atherosclerotic heart disease and undergone twice percutaneous coronary intervention in the period from ND to PD2, she missed benefiting from the best opportunity for receiving autologous hematopoietic stem cells. Eight temporal consecutive samples (ND, PD1, PD2, PD3, PD4, PD6, PD8, and PD9) were collected. The samples from ND to PD6 were CD138 positive PCs from bone marrow, while PD8 sample were tissues from extramedullary plasmacytoma. In the end-stage of MM (PD9), malignant PCs appeared in her peripheral blood (PB), the proportion of which reached 10%. The patient refused to undergo bone marrow aspiration again. Therefore, for PD9, CD138 positive PCs from PB were used for WES, PB mononuclear cells (PBMC) were used for scRNA-seq.

The diagnosis for MM, International Staging System stage (ISS), and Durie-Salmon stage (DS) were determined in accordance with the criteria of International Myeloma Working Group (IMWG), 2018 (Kumar and Rajkumar, 2018). The definition of progressive disease (PD) adhered to the IMWG consensus criteria for response in 2016 (Kumar et al., 2016). Time to progression (TTP) was calculated from the initiation of therapy to progression. Electronic records of this RRMM patient were reviewed. Written informed consent was provided by the patient according to the Declaration of Helsinki. This study was approved by the ethics committee of Fudan University, Zhongshan Hospital (B2017-031R).

Collection of Mononuclear Cells and CD138+/CD138- Cell Sorting

Collecting mononuclear cells from bone marrow (BM) aspirate and peripheral blood (PB), along with sorting CD138+/CD138- plasma cells from BM mononuclear cells (BMMC), were performed as previously described (Xu et al., 2020).

Deoxyribonucleic Acid Extraction, Library Construction and Whole Exome Sequencing Basic Data Analysis

Genomic DNA were extracted from the sorted CD138 + plasma cells and matched PB mononuclear cells (PBMC) by utilizing the QIAamp DNA Mini kit (250) (51,306, QIAGEN). DNA was subsequently quantified using the Qubit 3.0 (Invitrogen) and Nanodrop spectrophotometer (Thermo Fisher Scientific), while integrity was assessed using 1% agarose electrophoresis. Genomic libraries were then captured using the Agilent SureSelect Human All Exon V6 kit (Agilent Technologies, United States). Approximately 2–3 μg genomic DNA was sheared to 150–220 bp small fragments using a sonicator (Covaris, Inc., Woburn, MA). DNA was purified and treated with reagents supplied with the kit according to the given protocol. Adapters from Agilent were ligated onto the polished ends, and the libraries were amplified using polymerase chain reaction. The amplified libraries were then hybridized with the Agilent SureSelect Human All Exon V6 (Müeller-Pillasch et al., 1997) probes. The DNA fragments bound with the probes were washed and eluted with the buffer provided in the kit. These libraries were sequenced on the Illumina sequencing platform (HiSeq X-10, Illumina, Inc., San Diego, CA), after which 150 bp paired-end reads were generated.

The raw data (FASTQ format) were pre-processed with fastp (version: 0.19.5) (Chen et al., 2018). Reads containing less than 70% bases with average quality value below 20 (Q20) were filtered out using the NGSQC toolkit (version 2.3.2). Bases with a quality below Q20 were trimmed from the 3' end. Reads with ambiguous bases or those

shorter than 75 bp were also removed. High quality and clean reads were aligned to the reference human genome (GRCh37/hg19) using the Burrows-Wheeler aligner (BWA) (version 0.7.12) (Li and Durbin, 2009). Duplicate reads were removed using Picard (version 4.1.0.0). The mapped reads were sorted and indexed using the Sequence Alignment Map tool (SAMtools, version 1.4) (Li et al., 2009). The Genome Analysis Toolkit (GATK, version 4.1.0.0) (McKenna et al., 2010) was used for the recalibration of base quality score and realignment of single nucleotide polymorphisms (SNPs) and short insertion/deletions (INDELs). The final BAM files were used as input files for variant calling. Variants were annotated using ANNOVAR (Wang et al., 2010). The somatic mutations including somatic single nucleotide variants (SNVs) and somatic INDELs were screened out by comparing the variants between $CD138^+$ plasma cells and the matched PBMC using MuTect2 (version 1.1.7) (Cibulskis et al., 2013). Identification of copy number alterations (CNA) were called using the Control-FREEC software (Boeva et al., 2012). The whole exome sequencing was conducted by OE Biotech Co., Ltd (Shanghai, China).

Clonal Evolution Analysis

All somatic mutations of the eight consecutive samples at different stages (ND-PD1-PD2-PD3-PD4-PD6-PD8-PD9) were merged. Nonsynonymous somatic mutations and INDELs that changed the protein amino acid sequence were then filtered out. The sites with maf >0.01 in the population database of 1000 g, exac, gnomad, esp6500, as well as sites with a depth of less than 30, were then removed. Pyclone software was used to analyze the clone type of each sample. Based on the results of Pyclone, ClonEvol was used to analyze the evolution of clones between samples. According to the ClonEvol findings, the fishplot R package was used to draw a fish pattern (Roth et al., 2014).

Screening of Known Driver Genes

In order to identify the mutated driver genes, all somatic variations in each sample were then compared with the known driver genes. The sources of the driver genes for comparison were: 1) CGC513: driver gene listed in Cancer Gene Census list; 2) Bert Vogelstein125 (Vogelstein et al., 2013): 125 mut-driver genes in Bert Vogelstein's paper; and 3) SMG127: a significantly mutated gene found via TCGA pan-cancer data. The potential driver genes in these MM samples were screened out.

DNA Ploidy Profiling

Hypodiploidy or whole chromosome deletion is defined by a total of less than 45 whole chromosomes. However, hyper-diploidy is defined by chromosome amplification in the form of trisomies or tetrasomies of odd number chromosomes (48–74 chromosomes). Estimation of ploidy was analyzed using the Sequenza2 software (Favero et al., 2015).

Tumor Mutation Burden, Homologous Recombination Deficiency Scores and Mutant-Allele Tumor Heterogeneity

TMB = Number of nonsynonymous somatic mutations in the area of coding sequence (CDS)/Length of CDS. HRD scores were calculated as previously described using the scarHRD R package (Davies et al., 2017). MATH = 100 × median absolute deviation (MAD)/median (Mayakonda et al., 2018).

Chromium 10× Single-Cell 3' mRNA Sequencing and Data Processing

The main steps of scRNA-seq for PD9 were cell preparation, cDNA synthesis, library construction, and sequencing. This protocol refers specifically to the CG00052_SingleCell3_ReagentKitv2UserGuide_RevD, which is downloadable from the 10× Genomics website. The three PBMC samples of the healthy people were taken from official Chromium 10× (PBMC1 https://cg.10xgenomics.com/samples/cell-exp/3.0.0/pbmc_10k_v3/pbmc_10k_v3_fastqs.tar;PBMC2 https://cg.10xgenomics.com/samples/cell-exp/4.0.0/Parent_NGSC3_DI_PBMC/Parent_NGSC3_DI_PBMC_fastqs.tar; PBMC3 https://cg.10xgenomics.com/samples/cell-exp/4.0.0/SC3_v3_NextGem_DI_PBMC_10K/SC3_v3_NextGem_DI_PBMC_10K_fastqs.tar). The Cell Ranger software pipeline (version 3.1.0) provided by 10 × Genomics was used to demultiplex cellular barcodes, map reads to the genome and transcriptome using the STAR aligner, and down-sample reads as required to generate normalized aggregate data across samples, thereby producing a matrix of gene counts versus cells. The Seurat R package (version 3.1.1) was used to process the unique molecular identifier (UMI) count matrix, remove low quality cells and likely multiplet captures, obtain the normalized count and gene expression, perform graph-based clustering and identify marker genes of each cluster. Cells were visualized using a 2-dimensional t-distributed stochastic neighbor embedding (t-SNE) algorithm. Differentially expressed genes (DEGs) were identified using the FindMarkers function. p value <0.05 and $|\log2foldchange| > 0.58$ were set as the thresholds for significantly differential expression. Cell trajectory was carried out by Monocle. Single-cell regulatory network inference, and clustering (SCENIC) analysis was performed in order to infer the regulon activity score (Hänzelmann et al., 2013; Trapnell et al., 2014; Haghverdi et al., 2016).

RESULTS

Interpretation of Therapy-Induced Evolution of MM Cells Through Analyzing Eight Temporal Consecutive WES Data

Treatment timeline for P1 is shown in **Figure 1A**. WES was conducted with a mean coverage depth of 167× (range: 126-220X) for CD138 + plasma cells and 173× for PBMCs (control), consistent with the recommendations for WES. The vast majority of genomic sequences (99.78–99.97%) were mapped to the hg19 (GRCh37) reference genome. The detailed clinical data of baseline, as well as each progression of P1, are given in **Supplementary Table S1**. The total number of different types of somatic single nucleotide variants (SNVs) and short insertion/deletions (INDELs), summary of quality control and mapping results in the eight consecutive samples are summarized in

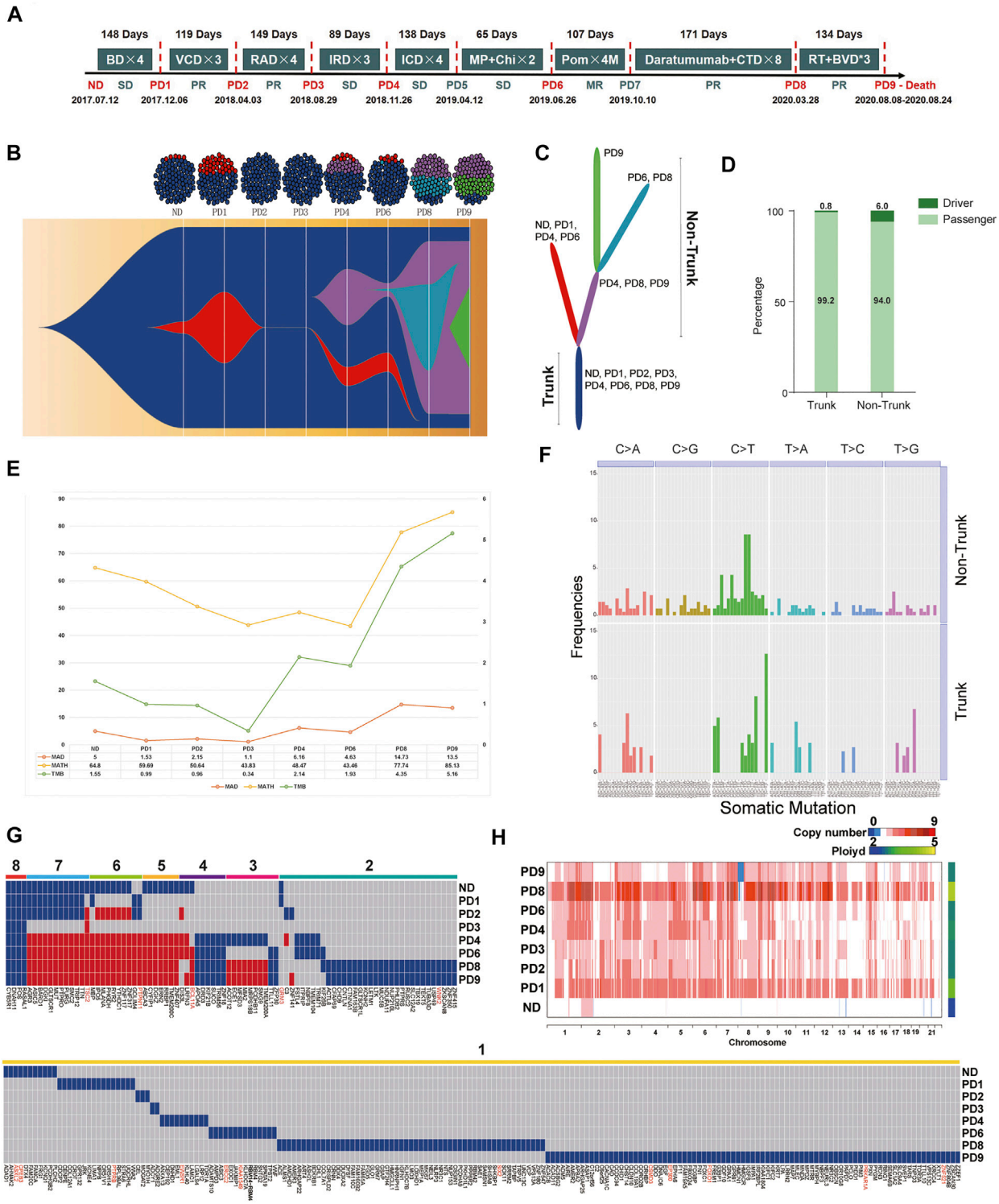


FIGURE 1 | (A) Treatment timeline for P1: After diagnosis with active MM (2017.7.12), P1 received Bortezomib/Dexamethasone (BD) as the primary therapy. After four cycles of BD, P1 suffered the first PD (PD1, day 1–148, TTP = 148 days). The second line therapy were Bortezomib/Cyclophosphamide/Dexamethasone (VCD). After three cycles of VCD, P1 suffered the second PD (PD2, day 148–266, TTP = 119 days). Lenalidomide/Adriamycin/Dexamethasone (RAD) were selected as the third line therapy regimen. The third PD (PD3, day 266–414, TTP = 149 days) occurred after three cycles of RAD. After three cycles of the fourth-line therapy, Ixazomib/Lenalidomide/Dexamethasone (IRD), P1 suffered the fourth PD (PD4, day 414–503, TTP = 89 days). Ixazomib/Cyclophosphamide/Dexamethasone (ICD) were the fifth-line therapy regimen. After four cycles of ICD, P1 had a clinical relapse (PD5, day 503–640, TTP = 138 days). The therapy further changed to Melphalan/Prednisone/

(*Continued*)

FIGURE 1 | Chidamide (MP + Chi). However, P1 had another PD (PD6, day 640–704, TTP = 65 days). The patient received Pomadomide (Pom) as the seventh-line therapy until the seventh PD (PD7, day 704–811, TTP = 107 days). The eighth-line therapy was Daratumumab and Cyclophosphamide, Thalidomide and Dexamethasone (CTD) which sustained 171 days until extramedullary plasmacytoma occurred on May 28, 2020 (PD8, day 811–982, TTP = 171 days). Finally, P1 received radiotherapy and Bendamustine/Bortezomib/Dexamethasone (BVD) for three cycles. In the end-stage of disease, malignant PCs appeared in her peripheral blood (PB), the proportion of which reached 10%. She passed away on Aug 24th, 2020. **(B)** Fish model showing heterogeneity of malignant PCs from P1 was analyzed by Pyclone and ClonEvol. After four times internal balances between red cluster and dark blue cluster (ND-PD3), a new clonal cluster (purple) was acquired under the pressure of multiple-line therapy interventions on PD4. Malignant PCs of PD9 consisted of 1) the always-existing dark-blue trunk sub-group, 2) the purple non-trunk sub-group evolved from PD4, and 3) the newly emerging green non-trunk sub-group. **(C)** The branching-pattern phylogenetic relationships in all malignant PCs from P1. **(D)** These driver genes were scattered along the phylogenetic tree, surprisingly, the number of driver genes in the non-trunk cluster (6.0%) was higher than the trunk cluster (0.8%). **(E)** From the first diagnosis to the nineth relapse, MATH of each sample was 64.80, 59.65, 50.64, 43.83, 48.47, 43.46, 77.74, and 85.13. TMB of each sample was 1.55, 0.99, 0.96, 0.34, 2.14, 1.93, 4,35 and 5.16, respectively. **(F)** Among all mutations, C > T transitions were the predominant change. In addition, T [C > T]T (12.61%) and T [C > T]A (8.11%) transitions dominated the trunk mutations, while G [C > T]G and G [C > T]C transitions (8.57 and 8.57%) dominated the non-trunk mutations. **(G)** The temporal distribution of depth >50, copy number >0 and nonsynonymous SNVs detected by WES in a heat map, with dark blue or red indicating the presence of a mutation and gray indicating the absence of a mutation. The eight color bars above the heat map indicate classification of these SNVs according to the total number of occurrences in the samples. For the gene names, red indicates that the mutation maybe a known driver gene, and black indicates a passenger gene. **(H)** Heterogeneity at the copy number (CN) level was analyzed. The ploidy of ND-PD9 samples were 2.029, 3.532, 2.768, 2.753, 2.873, 2.817, 4.556, and 2.774, respectively.

Supplementary Figure S1, **Supplementary Table S2** and **Supplementary Table S3**, respectively.

In order to investigate the changes in the clonal architecture and decipher the clonal evolution of MM cells in light of different treatments, Pyclone and ClonEvol were used. Heterogeneity (**Figure 1B**) and the branching-pattern phylogenetic relationships in these malignant PCs were then detected (**Figure 1C**). Notably, after conducting internal balances between red cluster and dark blue cluster (ND-PD3) four times, a new clonal cluster (purple) was acquired under the pressure of multiple-line therapy interventions on PD4. Branch lengths of the phylogenetic tree were found to be proportional to the number of nonsynonymous mutations separating the branching points. The trunk (the dark blue cluster), which were the most persistent clones, existed throughout the entire course and formed the mutational profile of the founder population. The non-trunk cluster, namely, the red cluster, purple cluster, light blue cluster and green cluster, were present in some stages.

According to the known driver genes from annotations of three pan-cancer databases, including CGC513, BV125 and SMG127, a total of 18 driver variants were identified (**Supplementary Table S4**). These driver genes were found to be scattered along the phylogenetic tree. Surprisingly, the number of driver genes in the non-trunk cluster (6.0%) was observed to be higher than that of the trunk cluster (0.8%) (**Figure 1D**). This indicated that the force to drive branching-pattern evolution of malignant PCs was sustained. In view of the heterogeneity and complex regulating network in MM progression, trunk and non-trunk genes were enriched into multiple pathways and functions using the Gene Ontology (GO) databases (**Supplementary Figure S2**).

From first diagnosis to the nineth relapse, the MATH of each sample was determined to be 64.80, 59.65, 50.64, 43.83, 48.47, 43.46, 77.74, and 85.13. Meanwhile, the TMB of each sample was 1.55, 0.99, 0.96, 0.34, 2.14, 1.93, 4,35, and 5.16, respectively. The trend of MATH and TMB presented an initial gradual downward trend followed by an upward trend throughout the course of the disease (**Figure 1E**). PD3 showed the lowest MATH and TMB in the eight continues samples. Moreover, only the dark blue cluster (Trunk) was detected in PD3, which was actually an imbalanced status of the bulk malignant PCs. In conjunction with the clinical data presented in **Figure 1A**, this status actually led to the shortest TTP as well as a strong and rapid counterattack of the red and purple clusters (PD4). Accordingly, we attempted to understand this phenomenon from an evolutionary perspective; administration according to a fixed and linear protocol is a selective perturbation that may lead to the emergence of drug-resistant subclones. Dynamic imbalances may not be conducive to stabilizing malignant MM cells. Adaptive therapy based on evolutionary law may provide a potential strategy for RRMM treatment to attain a fixed tumor population and prolong PFS rather than reducing malignant cells entirely.

The trinucleotide mutational spectrum of trunk (up) and non-trunk (down) mutations based on the phylogenetic tree is shown in **Figure 1F**. Among all mutations, C > T transitions were found to be the predominant change. In addition, T [C > T]T (12.61%) and T [C > T]A (8.11%) transitions dominated the trunk mutations, while G [C > T]G and G [C > T]C transitions (8.57 and 8.57%) dominated the non-trunk mutations.

During therapy, most MM patients initially go into remission, however, drug-resistant mutations may later cause disease progression. The mutational spectrum was further analyzed based on the timeline of mutation acquisition (**Figure 1G**). This figure illustrates the temporal distribution of depth >50, copy number >0 and nonsynonymous SNVs detected by WES in a heat map, with dark blue or red indicating the presence of a mutation and gray indicating the absence of a mutation. The eight color bars above the heat map indicate classification of these SNVs according to the total number of occurrences in the samples. In terms of gene names, red indicated that the mutation may be a known driver gene, while black indicated a passenger gene. Here, only four identical somatic passenger mutations (FOS, RASAL1, CYB5R1, and DNAH11) were found to be ubiquitously detectable in all samples. No known-driver gene was shared between the eight specimens. By observing the genomics of P1, some SNVs that had disappeared after the previous treatment were observed to reappear in later stages (red blocks). Mutations were further classified into susceptible SNVs and resistant SNVs for each treatment course. GO and Kyoto

TABLE 1 | HRD-scores of the P1 patient.

Samples ID	LOH*	TAI*	LST*	HRD-sum*
ND	0	3	2	5
PD1	0	39	24	63
PD2	0	36	8	44
PD3	0	32	18	50
PD4	0	37	27	64
PD6	0	31	21	52
PD8	0	39	28	67
PD9	0	34	26	60

Abbreviations: LOH: Loss of Heterozygosity; TAI: Number of Telomeric Allelic Imbalances; LST: Large Scale Transitions; HRD-sum: Heterozygosity scar.

Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were conducted, as shown in **Supplementary Figure S3** and **Supplementary Figure S4**. Based on personal genomics, a potential treatment strategy—recycle therapy may be considered.

The heterogeneity at the copy number (CN) level was then analyzed (**Figure 1H**). The ploidy of ND-PD9 samples were found to be 2.029, 3.532, 2.768, 2.753, 2.873, 2.817, 4.556, and 2.774, respectively. Remarkably, compared to the ND sample, the PD samples were observed to possess higher-ploidy karyotypes, indicating that chromosomal instability (CIN) increased during the period of progression. Among the PD samples, PD8 (extramedullary plasmacytoma) had the highest ploidy after the eighth-line therapy, which combined Daratumumab with Cyclophosphamide, Thalidomide and Dexamethasone (CTD).

HRD scores of each sample are summarized in **Table 1**, for which the HRD-sum scores [Loss of Heterozygosity (LDH) + Number of Telomeric Allelic Imbalances (TAI) + Large Scale Transitions (LST)] of each sample were found to be 5, 63, 44, 50, 64, 52, 67, and 60. In MM, extramedullary progression is always associated with treatment resistance as well as a high mortality rate. The changing trend of HRD scores were observed to be consistent with ploidy.

Combination With Genomics and Single Cell Transcriptomes Deciphered the Heterogeneity and Distinct Evolutional Subpopulations in MM Cells

Next, the single-cell transcriptome data of PD9, from peripheral blood of P1 (about 10% PCs), were analyzed. After conducting quality control, 553 356 446 sequence reads and 39 525 reads per cell for an estimated 14 000 cells were obtained, with 87.3% confidently mapped to the human reference transcriptome GRCh38–3.0.0. On average, 937 genes and 2,369 unique molecular identifiers (UMIs) per cell were detected.

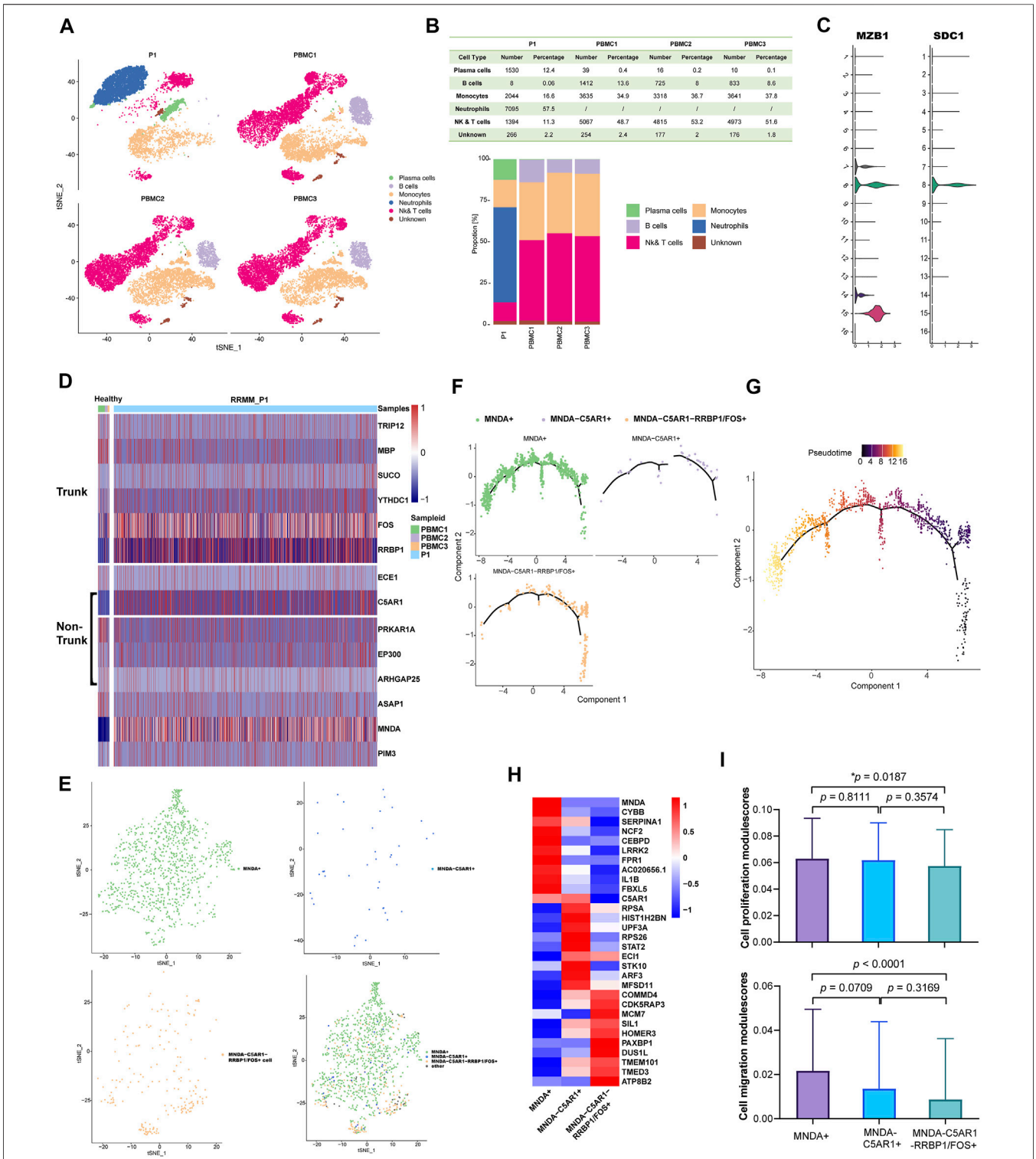
In order to explore the cellular composition of PBMC from P1_PD9, a comparison was made with the data of three healthy PBMC controls from the 10 × platform database described in supplementary methods, while unsupervised clustering was applied to distinguish the cell types. Finally, PCs, B cells, monocytes, neutrophils, nature killer cells (NK) and T cells were categorized and visualized using t-distributed stochastic neighbor embedding (t-SNE) (**Figure 2A**). Stacked bar plots (**Figure 2B**) demonstrated that, compared to normal specimens, PCs were significantly increased in the P1 sample. Identification of PCs subpopulations (cluster 8) was then performed according to gene markers, including SDC1 (CD138) and MZB1 (**Figure 2C**).

According to the above WES results of evolutionary tree from P1 (**Figures 1B,C**), malignant PCs of PD9 consisted of: 1) the always-existing dark-blue trunk sub-group; 2) the purple non-trunk sub-group evolved from PD4; and 3) the newly emerging green non-trunk sub-group. In view of the mutational profiling of the above sub-groups, the expression levels of these mutant genes between PCs from normal control (PBMC1, 2, 3) and P1 were then compared (**Figure 2D**). Genes that were not expressed in the displayed cells were filtered out and not shown on the heatmap. Based on the results of the heatmap, PCs from P1 was further grouped as three sub-clusters: 1) PCs expressed higher level of MNDA (MNDA+, the green non-trunk sub-group); 2) PCs expressed higher level of C5AR1 but no MNDA (C5AR1+ MNDA-, the purple non-trunk sub-group); and 3) PCs expressed higher level of FOS or RRBP1 but no MNDA and C5AR1 (MNDA-C5AR1-RRBP1/FOS+, the dark-blue trunk sub-group) (**Figure 2E**). Minimum spanning tree (MST) of malignant PCs performed by pseudo-time analysis verified that the MNDA + sub-cluster evolved from MNDA-C5AR1-RRBP1/FOS + sub-cluster (**Figures 2F,G**). The different expression profiles of the three sub-clusters are illustrated in the corresponding heatmap (**Figure 2H**), reflecting dynamic gene expression profiles during the malignant evolution of MM cells. The AddModuleScore tool from Seurat was then utilized in order to calculate the module scores for cell proliferation and cell migration expression in the three sub-clusters, respectively. Compared to the trunk cluster, the MNDA + cluster exhibited significantly higher scores in cell proliferation ($p = 0.0187$) and cell migration ($p < 0.0001$) (**Figure 2I**). Together, the corresponding data highlighted the heterogeneity and 180 distinct evolutional subpopulations among MM cells.

Gene Regulatory Networks Revealed by scRNA-seq Identify RUNX3 Gene as a Potential Driver for RRMM

Transcription factors (TFs) and their targeted genes comprise a complex gene network regulation, referred to as a regulon, that could determine cell functional identity. Single-cell regulatory network inference and clustering (SCENIC) analysis was performed to infer the regulon activity score (RAS) for the MNDA+, C5AR1+ MNDA- and MNDA-C5AR1-RRBP1/FOS + sub-clusters, respectively (**Figure 3A**). Five regulon modules were then identified for the three sub-groups according to the Connection Specificity Index (CSI), indicatinga significant correlation between different regulons with minimization of the effects of non-specific crosstalk (**Figure 3B**). Inter-gene expression correlations as well as the specifically involved genes in each module are shown in **Figure 3C**.

In terms of the most malignant sub-cluster, namely the MNDA + sub-cluster, the intersection of TFs marked with red boxes in **Figures 3A,B** were determined by incorporating the



Cell Type	P1 Number	P1 Percentage	PBMC1 Number	PBMC1 Percentage	PBMC2 Number	PBMC2 Percentage	PBMC3 Number	PBMC3 Percentage
Plasma cells	1530	12.4	39	0.4	16	0.2	10	0.1
B cells	8	0.06	1412	13.6	725	8	833	8.6
Monocytes	2044	16.6	3635	34.9	3318	36.7	3641	37.8
Neutrophils	7095	57.5	/	/	/	/	/	/
NK & T cells	1394	11.3	5067	48.7	4815	53.2	4973	51.6
Unknown	266	2.2	254	2.4	177	2	176	1.8

FIGURE 2 | (A) Comparing with the data of three healthy PBMC controls from the 10× platform database described above, cell types were identified in P1_PD9 and were visualized by t-SNE (16 clusters in total). **(B)** Stacked bar plots show that compared with normal specimens, plasma cells significantly increased and B cells significantly decreased in the P1 sample. **(C)** According to gene markers, including SDC1 (CD138) and MZB1, cluster 8 was finally determined to be the malignant PCs. **(D)** According to the WES results of evolutionary tree from P1 in **Figure 1B** and the mutational profiling of trunk and non-trunk, the mRNA expression levels of these mutant genes between PCs from normal control (PBMC1, 2, 3) and malignant PCs from P1 were compared and showed in the heatmap. **(E)** Malignant PCs in P1 was further grouped as three sub-clusters: 1) PCs expressed higher level of MNDA (MNDA+, the green non-trunk sub-group in WES results); 2) PCs expressed higher level of

(*Continued*)

FIGURE 2 | C5AR1 but no MNDA (C5AR1+ MNDA-, the purple non-trunk sub-group in WES results); and 3) PCs expressed higher level of FOS or RRBP1 but no MNDA and C5AR1 (MNDA-C5AR1-RRBP1/FOS+, the dark-blue trunk sub-group in WES results). **(F)** Minimum spanning tree (MST) of malignant PCs performed by pseudo-time analysis also revealed that the MNDA + non-trunk sub-group evolved from MNDA-C5AR1-RRBP1/FOS + trunk sub-group (from right to left). **(G)** The pseudo-time of malignant PCs. **(H)** The different expression gene profiles of the three sub-clusters. **(I)** Cell proliferation and migration modulescores among the three sub-group.

results of regulation activity and specific modules. Enrichment analysis using KEGG (**Figure 3D**) and GO (**Figure 3E**) of the 9 TFs, including JUNB, CEBPD, IRF1, CEBPB, SPL1, IKZF1, CEBPA, BACH1, RUNX3, suggested that the TNF signaling pathway, transcriptional mis-regulation in cancer, regulation of transcription involved in G1/S transition of mitotic cell cycle may serve vital roles.

Next, using the MMRF-CoMMpass datasets, a progressive increase of RUNX3 mRNA expression level with times of PD were found (Baseline vs. PD 2-4: $p = 0.0275$, Baseline vs. PD 5-6: $p = 0.0036$) (**Figure 3F**). Moreover, the overall survival (OS) of the 766 newly diagnosed MM (NDMM) were calculated *via* Kaplan-Meier survival analysis in order to estimate the effect of the above 9 TFs. Accordingly, by univariate survival analysis (Kaplan-Meier analysis, log-rank test), the RUNX3 mRNA expression level was found to be the only poor prognostic factor among the 9 TFs (low vs. high: $p = 0.024$, median vs. high: $p = 0.025$) (**Figure 3G**). Furthermore, multivariate survival analysis by Cox regression, including age, Eastern Cooperative Oncology Group performance status (ECOG) and the Revised International Staging System (R-ISS), identified the RUNX3 gene as an independent prognostic factor for OS (**Table 2**, $p = 0.027$). The survival effect of the RUNX3 gene was then validated in the GSE24080 dataset (Kaplan-Meier analysis, log-rank test, median vs high: $p = 0.013$) (**Figure 3H**). While for the multivariate survival analysis, including age, cytogenetic abnormalities, albumin and beta-2 micro-globulin, RUNX3 showed a trend to be an independent prognositic factor for OS (**Table 3**, $p = 0.070$). The corresponding findings suggest that the RUNX3 gene may be a potential therapeutic target for the treatment of MM.

DISCUSSION

In this study, temporal consecutive genetic analysis provided evidence of heterogeneity in MM and unmasked the evolutionary trajectories of PCs derived by multi-line therapies. Based on somatic mutations, the evolution trajectory was constructed to reflect the development of this RRMM patient, showing that the dominant pattern was branch evolution. Final scRNA-seq data provided new evidence and insights into deciphering the heterogeneity and evolution in RRMM, which are pivotal in dissecting MM-related mechanisms in detail. We hypothesized that MM clonal evolution and progression underlie Darwinian selection, which are mediated by tumor-intrinsic characteristics as well as extrinsic pressure.

In the past decades, the survival of most MM patients has improved significantly due to the development of autologous stem cell transplantation (ASCT) and novel agents, including proteasome inhibitors, immunomodulatory drugs, and monoclonal antibodies (Kumar et al., 2008). Despite its overall improvement, a group of high-risk patients with clinically aggressive behaviors (PFS less than 18 months, OS less than 1.5–3 years) continues to suffer. Current therapies rarely consider molecular information to personalized MM therapy. In order to continually improve MM patients' outcomes, information regarding the genomic abnormalities leading to these heterogenous outcomes should be incorporated into personal clinical care (Pawlyn and Davies, 2019). Recently, high-throughput sequencing has brought about personalized treatment according to the specific genetic composition and molecular phenotype of patients. Notably, in this study, a majority of known MM driver genes previously reported did not actually exist in this individual's MM genome sequences (Robiou du Pont et al., 2017), suggesting that MM is a highly heterogeneous malignant disease, and personal omics is crucial in clinical management. This is conducive in developing individualized targeted medication and identifying potential personal therapeutic targets. In this case, mutations may be classified into susceptible SNVs and resistant SNVs for each of P1's treatment courses. Disappearance of susceptible somatic mutations and emergence of resistant somatic mutations may provide guidance for personalized-therapy. Accordingly, agents effective against these recurring mutations during the earlier stages may be re-administrated. The alternate use of agents not only conforms to the theory of adaptive therapy, but also provides a therapeutic solution for multi-line recurrence MM patients.

In regard to MATH and TME, dynamic imbalance was found to not be conducive to stabilizing malignant MM cells. In 2009, Gatenby et al. (2009) first put forward the adaptive therapy theory and presented powerful mathematical models to represent their findings. They pointed out that resistant populations that are present before therapy would rapidly grow with treatments designed to kill the majority of malignant cells, which is a result of removal of the inhibitory effect that the sensitive population has on the resistant population, in conjunction with the disturbed balance between both populations. Adaptive therapy takes advantage of the inhibitory effect of the sensitive population on the resistant population, leading to the slow growth of entire cancer cells. A study on the spatial and temporal clonal evolution of intrahepatic cholangiocarcinoma also provided a theoretical basis for adaptive therapy (Dong et al., 2018). Recently, An et al. (2020) reported that persistent cytogenetic abnormalities were detected in residual PCs in the majority of MM patients, which were even present in the minimal residual disease (MRD) negative cohort. These findings may further provide evidence for the potential effectiveness of adaptive therapy in MM treatment.

Moreover, understanding the trajectories of the temporal consecutive changes of PCs may serve as a powerful tool in estimating risk of progression and could bring about profound implications in clinical management. According to Charlotte Pawlyn and Gareth J Morgan (Pawlyn and Morgan, 2017),

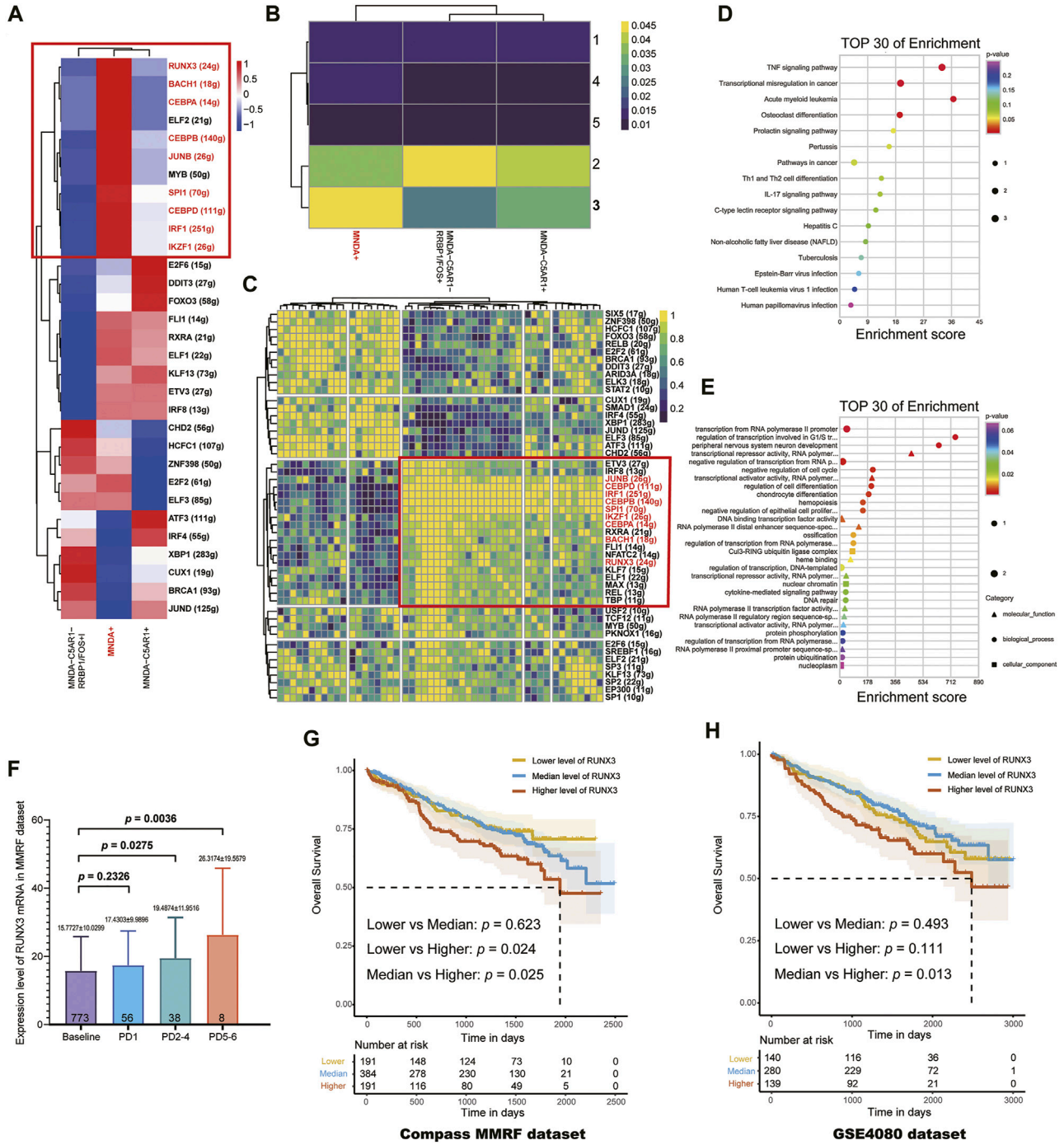


FIGURE 3 | (A) The regulon activity score (RAS) for MNDA + sub-cluster, C5AR1+ MNDA- sub-cluster and MNDA-C5AR1-RRBP1/FOS + sub-cluster, respectively. **(B)** Five regulon modules were identified for the three sub-groups by the Connection Specificity Index. **(C)** Inter-gene expression correlations and specific involved genes in each module. **(D)** Enrichment analysis using KEGG of 9 TFs. **(E)** Enrichment analysis using GO of 9 TFs. **(F)** Progressive increase of RUNX3 mRNA expression level with times of PD (Baseline vs. PD 2–4: $p = 0.0275$, Baseline vs. PD 5-6: $p = 0.0036$) using the MMRF-CoMMpass datasets. **(G)** RUNX3 mRNA expression level was found to be the only poor prognostic factor among the 9 TFs (OS, low vs. high: $p = 0.024$, median vs. high: $p = 0.025$) in 766 newly diagnosed MM. **(H)** The survival effect of RUNX3 gene was validated in GSE24080 dataset (median vs. high: $p = 0.013$).

subsequent driver lesions often occur in a sub-clonal PCs in order to facilitate MM progression. In the patient presented in this study, driver genes were observed to be scattered along the phylogenetic tree, suggesting that the force to drive branching-pattern evolution of malignant PCs was sustained. In addition, a total of four identical passenger mutations were present in all samples: FOS, RASAL1, CYB5R1, and DNAH11. Accordingly, we hypothesized that the four mutations may be closely associated

TABLE 2 | Multivariate analysis of OS in MMRF compass dataset.

Variable	HR	95% CI	*p*-value
RUNX3	1.028	1.003–1.053	0.027
Age	1.047	1.028–1.067	0.000
ECOG	1.364	1.102–1.689	0.004
R_ISS	2.191	1.571–3.057	0.000

Abbreviations: CI, confidence interval; HR, hazard ratio; ECOG: Eastern Cooperative Oncology Group performance status; R-ISS: the Revised International Staging System.

TABLE 3 | Multivariate analysis of OS in GSE24080 dataset.

Variable	HR	95% CI	*p*-value
Age	1.015	0.998–1.031	0.083
RUNX3	1.254	0.982–1.601	0.070
ALB	0.698	0.558–0.873	0.002
β2M	1.067	1.047–1.088	0.000
Cyto abnor	2.096	1.548–2.838	0.000

Abbreviations: CI, confidence interval; HR, hazard ratio; Cyto abnor: the detection of cytogenetic abnormalities; ALB: Albumin; β2M: Beta-2 microglobulin.

with the proliferation of MM cells. FOS, Fos proto-oncogene or AP-1 transcription factor subunit, belongs to the Fos gene family and encodes leucine zipper proteins that dimerize with proteins of the JUN family (Cannarile et al., 2019). FOS proteins have established roles as regulators of cell proliferation, differentiation, and apoptosis. According to Youg Raj Thaker et al., Rasal1, a type of GTPase-activating protein, associates with ZAP-70 of the TCR and negatively regulates T-cell activation and tumor immunity (Thaker et al., 2019). GO annotations associated with CYB5R1 include oxidoreductase activity and cytochrome-b5 reductase activity, which act on NAD(P)H. Moreover, NADH-cytochrome b5 reductases are involved in the desaturation and elongation of fatty acids, cholesterol biosynthesis, and drug metabolism. GO annotations associated with DNAH11 include ATPase activity and microtubule motor activity.

This study also demonstrated that CIN and HRD notably increased in the PD samples, especially in the period of extramedullary plasmacytoma. Cells with genomic instability, a hallmark of cancer, have an increased tendency of underdoing genomic alterations during cell division (Negrini et al., 2010). As a result, increased genomic alterations in MM may lead to more neoantigens being expressed on the surface of PCs. Unfortunately, the immune system is naive to these novel peptides, escaping detection (Neuse et al., 2020).

Furthermore, increasing evidence in other cancers have suggested that clones with polyploid and aneuploid may be enriched at recurrence periods and mediate therapeutic resistance (Mirzayans et al., 2018; Pienta et al., 2021). These polyploid clones always exhibit a protected cellular phenotype that is largely resilience to environmental disruption. On an evolutionary timescale, polyploid programs could provide increased fitness. Upon removal of chemotherapy stress, these polyploid clones may undergo depolyploidization and generate resistant progeny. These theories may provide explanations for the results that CIN significantly increased in the PD samples in this study.

Despite the clinical success of the proteasome inhibitor bortezomib and immunomodulatory imide drugs (IMiDs), treating RRMM remains a challenge. According to the results of the WES evolutionary tree in this study, from the level of single-cell transcriptomes, malignant PCs of P1 in the final state (PD9) were further divided into three subgroups, ultimately suggesting the potential importance of the RUNX3 gene. According to literature, RUNX3 belongs to the family of RUNXs (RUNX1, RUNX2, RUNX3). In 2019, Zhou et al. (2019) demonstrated that in myeloma cells, RUNX3 and RUNX1 could interact with Ikaros family zinc finger protein 1 and 3 (IKZF1 and IKZF3). Aa a result, myeloma cell lines and primary tumors may become refractory to CRBN-dependent ubiquitylation and degradation induced by IMiDs. Moreover, inhibition of RUNX proteins resulted in enhanced sensitivity of MM cells to IMiDs. These results may improve the understanding of the complex pathophysiology of MM and provide an alternative approach for applications in personalized medicine.

This study has certain limitations. First, since the total quantity of each BM sample was stored at −80°C for a relatively long time, it was insufficient for integrated omics data, such as epigenomics and proteomics, resulting in a lack of combing data. Second, all temporal consecutive samples were obtained from only one RRMM patient, which may introduce potential bias in the interpretation of the results. Further studies with increased sample sizes could provide more definitive evidence. Third, adaptive therapy and recycle therapy require further validation in large-scale preclinical and clinical studies. In summary, this study demonstrated that temporal consecutive samples from one RRMM patient can assess the diversity of sub-clones as well as the clonal evolution trajectory driven by multi-line therapies. Moreover, distribution of driver mutations and passenger mutations were observed to be scattered in the trunk and non-trunk of the branching-pattern phylogenetic tree. Personal high-throughput sequencing that reveals specific genetic compositions and molecular phenotypes may contribute to more robust personal biomarkers and serve as guidance for personalized therapy. Furthermore, adaptive and recycle therapy that are grounded in evolutionary law may prevent a rise in the resistant PCs population and ser as a potential strategy for RRMM treatment.

DATA AVAILABILITY STATEMENT

The clinical data of this study are restricted by the ethics committee in Zhongshan Hospital in order to protect patients' privacy. Requests to access these datasets should be directed to (Peng Liu, liu.peng@zs-hospital.sh.cn). The WES datasets used and/or analyzed during the current study are available from PRJNA690173 in SRA (https://www.ncbi.nlm.nih.gov/sra/PRJNA690173). The scRNA-seq are available from GSE188632 in GEO (https://www.ncbi.nlm.nih.gov/geo/info/linking.html).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethics committee of Fudan University,

Zhongshan Hospital (B2017-031R). The patients/participants provided their written informed consent to participate in this study. 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

PL designed research; JX performed research, analyzed data and wrote the paper; ZW, JZ, SG, YZ and CC collected clinical data and samples; YW contributed analytical tools; YS, JJ, YZ and PL interpreted data and helped perform the analysis with constructive discussions. LR and YW checked the data and revised the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China under Grant 81570123 and 82100215, the National Key New Drug Creation Special Programs under Grant 2017ZX09304-021, and Academic Pacesetters Program of Shanghai Healthcare System under Grant 2017BR033.

ACKNOWLEDGMENTS

We pleased to acknowledge Beijing Bethune Charitable Foundation for their assistance with the article processing fee. We pleased to acknowledge Lian Li, Yincong Gu and Yongbing Ba for their assistance throughout the preparation of the original manuscript. The authors would like to thank the National Natural Science Foundation of China under Grant 81570123 and 82100215, the National Key New Drug Creation Special Programs under Grant 2017ZX09304-021, and Academic Pacesetters Program of Shanghai Healthcare System under Grant 2017BR033.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.794144/full#supplementary-material

REFERENCES

An, G., Yan, Y., Xu, Y., Mao, X., Liu, J., Fan, H., et al. (2020). Monitoring the Cytogenetic Architecture of Minimal Residual Plasma Cells Indicates Therapy-Induced Clonal Selection in Multiple Myeloma. *Leukemia* 34 (2), 578–588. doi:10.1038/s41375-019-0590-x

Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappo, J., Schleiermacher, G., et al. (2012). Control-FREEC: A Tool for Assessing Copy Number and Allelic Content Using Next-Generation Sequencing Data. *Bioinformatics* 28 (3), 423–425. doi:10.1093/bioinformatics/btr670

Cannarile, L., Delfino, D. V., Adorisio, S., Riccardi, C., and Ayroldi, E. (2019). Implicating the Role of GILZ in Glucocorticoid Modulation of T-Cell Activation. *Front. Immunol.* 10, 1823. doi:10.3389/fimmu.2019.01823

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: An Ultra-fast All-In-One FASTQ Preprocessor. *Bioinformatics* 34 (17), i884–i890. doi:10.1093/bioinformatics/bty560

Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., et al. (2013). Sensitive Detection of Somatic Point Mutations in Impure and Heterogeneous Cancer Samples. *Nat. Biotechnol.* 31 (3), 213–219. doi:10.1038/nbt.2514

Davies, H., Glodzik, D., Morganella, S., Yates, L. R., Staaf, J., Zou, X., et al. (2017). HRDetect Is a Predictor of BRCA1 and BRCA2 Deficiency Based on Mutational Signatures. *Nat. Med.* 23 (4), 517–525. doi:10.1038/nm.4292

Dong, L.-Q., Shi, Y., Ma, L.-J., Yang, L.-X., Wang, X.-Y., Zhang, S., et al. (2018). Spatial and Temporal Clonal Evolution of Intrahepatic Cholangiocarcinoma. *J. Hepatol.* 69 (1), 89–98. doi:10.1016/j.jhep.2018.02.029

Egan, J. B., Shi, C.-X., Tembe, W., Christoforides, A., Kurdoglu, A., Sinari, S., et al. (2012). Whole-Genome Sequencing of Multiple Myeloma from Diagnosis to Plasma Cell Leukemia Reveals Genomic Initiating Events, Evolution, and Clonal Tides. *Blood* 120 (5), 1060–1066. doi:10.1182/blood-2012-01-405977

Favero, F., Joshi, T., Marquard, A. M., Birkbak, N. J., Krzystanek, M., Li, Q., et al. (2015). Sequenza: Allele-Specific Copy Number and Mutation Profiles from Tumor Sequencing Data. *Ann. Oncol.* 26 (1), 64–70. doi:10.1093/annonc/mdu479

Gatenby, R. A., Silva, A. S., Gillies, R. J., and Frieden, B. R. (2009). Adaptive Therapy. *Cancer Res.* 69 (11), 4894–4903. doi:10.1158/0008-5472.can-08-3658

Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F., and Theis, F. J. (2016). Diffusion Pseudotime Robustly Reconstructs Lineage Branching. *Nat. Methods* 13 (10), 845–848. doi:10.1038/nmeth.3971

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

Keats, J. J., Chesi, M., Egan, J. B., Garbitt, V. M., Palmer, S. E., Braggio, E., et al. (2012). Clonal Competition with Alternating Dominance in Multiple Myeloma. *Blood* 120 (5), 1067–1076. doi:10.1182/blood-2012-01-405985

Kumar, S. K., Rajkumar, S. V., Dispenzieri, A., Lacy, M. Q., Hayman, S. R., Buadi, F. K., et al. (2008). Improved Survival in Multiple Myeloma and the Impact of Novel Therapies. *Blood* 111 (5), 2516–2520. doi:10.1182/blood-2007-10-116129

Kumar, S. K., and Rajkumar, S. V. (2018). The Multiple Myelomas - Current Concepts in Cytogenetic Classification and Therapy. *Nat. Rev. Clin. Oncol.* 15 (7), 409–421. doi:10.1038/s41571-018-0018-y

Kumar, S., Paiva, B., Anderson, K. C., Durie, B., Landgren, O., Moreau, P., et al. (2016). International Myeloma Working Group Consensus Criteria for Response and Minimal Residual Disease Assessment in Multiple Myeloma. *Lancet Oncol.* 17 (8), e328–e346. doi:10.1016/s1470-2045(16)30206-6

Li, H., and Durbin, R. (2009). Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* 25 (14), 1754–1760. doi:10.1093/bioinformatics/btp324

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

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

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: A MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data. *Genome Res.* 20 (9), 1297–1303. doi:10.1101/gr.107524.110

Mirzayans, R., Andrais, B., and Murray, D. (2018). Roles of Polyploid/Multinucleated Giant Cancer Cells in Metastasis and Disease Relapse Following Anticancer Treatment. *Cancers (Basel)* 10 (4), 118. doi:10.3390/cancers10040118

Müeller-Pillasch, F., Lacher, U., Wallrapp, C., Micha, A., Zimmerhackl, F., Hameister, H., et al. (1997). Cloning of a Gene Highly Overexpressed in

Cancer Coding for a Novel KH-Domain Containing Protein. *Oncogene* 14 (22), 2729–2733. doi:10.1038/sj.onc.1201110

Negrini, S., Gorgoulis, V. G., and Halazonetis, T. D. (2010). Genomic Instability - an Evolving Hallmark of Cancer. *Nat. Rev. Mol. Cel Biol* 11 (3), 220–228. doi:10.1038/nrm2858

Neuse, C. J., Lomas, O. C., Schliemann, C., Shen, Y. J., Manier, S., Bustoros, M., et al. (2020). Genome Instability in Multiple Myeloma. *Leukemia* 34, 2887–2897. doi:10.1038/s41375-020-0921-y

Pawlyn, C., and Davies, F. E. (2019). Toward Personalized Treatment in Multiple Myeloma Based on Molecular Characteristics. *Blood* 133 (7), 660–675. doi:10.1182/blood-2018-09-825331

Pawlyn, C., and Morgan, G. J. (2017). Evolutionary Biology of High-Risk Multiple Myeloma. *Nat. Rev. Cancer* 17 (9), 543–556. doi:10.1038/nrc.2017.63

Pienta, K. J., Hammarlund, E. U., Brown, J. S., Amend, S. R., and Axelrod, R. M. (2021). Cancer Recurrence and Lethality Are Enabled by Enhanced Survival and Reversible Cell Cycle Arrest of Polyaneuploid Cells. *Proc. Natl. Acad. Sci. U S A.* 118 (7), e2020838118. doi:10.1073/pnas.2020838118

Robiou du Pont, S., Cleynen, A., Fontan, C., Attal, M., Munshi, N., Corre, J., et al. (2017). Genomics of Multiple Myeloma. *J. Clin. Oncol.* 35 (9), 963–967. doi:10.1200/jco.2016.70.6705

Roth, A., Khattra, J., Yap, D., Wan, A., Laks, E., Biele, J., et al. (2014). PyClone: Statistical Inference of Clonal Population Structure in Cancer. *Nat. Methods* 11 (4), 396–398. doi:10.1038/nmeth.2883

Thaker, Y. R., Raab, M., Strebhardt, K., and Rudd, C. E. (2019). GTPase-Activating Protein Rasal1 Associates with ZAP-70 of the TCR and Negatively Regulates T-Cell Tumor Immunity. *Nat. Commun.* 10 (1), 4804. doi:10.1038/s41467-019-12544-4

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., et al. (2014). The Dynamics and Regulators of Cell Fate Decisions Are Revealed by Pseudotemporal Ordering of Single Cells. *Nat. Biotechnol.* 32 (4), 381–386. doi:10.1038/nbt.2859

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., and Kinzler, K. W. (2013). Cancer Genome Landscapes. *Science* 339 (6127), 1546–1558. doi:10.1126/science.1235122

Walker, B. A., Mavrommatis, K., Wardell, C. P., Ashby, T. C., Bauer, M., Davies, F. E., et al. (2018). Identification of Novel Mutational Drivers Reveals Oncogene Dependencies in Multiple Myeloma. *Blood* 132 (6), 587–597. doi:10.1182/blood-2018-03-840132

Walker, B. A., Wardell, C. P., Melchor, L., Hulkki, S., Potter, N. E., Johnson, D. C., et al. (2012). Intraclonal Heterogeneity and Distinct Molecular Mechanisms Characterize the Development of T(4;14) and T(11;14) Myeloma. *Blood* 120 (5), 1077–1086. doi:10.1182/blood-2012-03-412981

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data. *Nucleic Acids Res.* 38 (16), e164. doi:10.1093/nar/gkq603

Xu, J., Sun, Y., Jiang, J., Xu, Z., Li, J., Xu, T., et al. (2020). Globular C1q Receptor (gC1qR/p32/HABP1) Suppresses the Tumor-Inhibiting Role of C1q and Promotes Tumor Proliferation in 1q21-Amplified Multiple Myeloma. *Front. Immunol.* 11, 1292. doi:10.3389/fimmu.2020.01292

Zhou, N., Gutierrez-Uzquiza, A., Zheng, X. Y., Chang, R., Vogl, D. T., Garfall, A. L., et al. (2019). RUNX Proteins Desensitize Multiple Myeloma to Lenalidomide via Protecting IKZFs from Degradation. *Leukemia* 33 (8), 2006–2021. doi:10.1038/s41375-019-0403-2

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: 01 February 2022
doi: 10.3389/fcell.2021.764481

Anhydroicaritin Inhibits EMT in Breast Cancer by Enhancing GPX1 Expression: A Research Based on Sequencing Technologies and Bioinformatics Analysis

Feifei Li[1†], Youyang Shi[1†], Xiaojuan Yang[1†], Zhanyang Luo[1], Guangtao Zhang[1], Kui Yu[2], Feng Li[2], Lixin Chen[2], Youkang Zhao[1], Ying Xie[1], Yuanyuan Wu[3], Jianfeng Yang[2]*, Xiqiu Zhou[2]* and Sheng Liu[1]**

[1]Longhua Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China, [2]Department of Surgery, Pudong Branch of Longhua Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China, [3]Department of Gastroenterology, Naval Medical Center of PLA, Naval Military Medical University, Shanghai, China

Background: Breast cancer (BC) is the leading cause of cancer-related deaths among women worldwide. The application of advanced technology has promoted accurate diagnosis and treatment of cancer. Anhydroicaritin (AHI) is a flavonoid with therapeutic potential in BC treatment. The current study aimed to determine AHI's mechanism in BC treatment *via* RNA sequencing, comprehensive bioinformatics analysis, and experimental verification.

Methods: Network pharmacology and MTT (3-(4,5)-dimethylthiazolyl-3,5-diphenyltetrazolium bromide) experiments were conducted to first confirm AHI's anti-BC effect. RNA sequencing was performed to identify the genes affected by AHI. Differential expression analysis, survival analysis, gene set enrichment analysis, and immune infiltration analysis were performed *via* bioinformatics analysis. Western blot analysis, reverse transcription–polymerase chain reaction (RT-PCR) experiment, molecular docking, and drug affinity responsive target stability (DARTS) experiments were also performed to confirm AHI's direct effect on glutathione peroxidase 1 (GPX1) expression. Confocal immunofluorescence analysis was conducted to verify AHI's effect on the occurrence and development of epithelial–mesenchymal transition (EMT). Finally, BC nude mouse xenografts were established, and AHI's molecular mechanism on BC was explored.

Results: Network pharmacology results demonstrated that AHI's therapeutic targets on BC were related to the proliferation, invasion, and metastasis of BC cells. AHI significantly inhibited the proliferation of 4T1 and MDA-MB-231 BC cells in the MTT experiments. RNA sequencing results showed that AHI upregulated the GPX1 expression in the 4T1 and MDA-MB-231 BC cells. Next, bioinformatics analysis revealed that GPX1 is less expressed in BC than in normal breast tissues. Patients with high GPX1 expression levels tended to have prolonged overall survival and disease-free survival than patients with low GPX1

OPEN ACCESS

Edited by:
Geng Chen,
GeneCast Biotechnology Co., Ltd., China

Reviewed by:
Chen Jue,
Yangzhou University, China
Hailiang Zhang,
Fudan University, China
Ling Liu,
Third Affiliated Hospital of Zhengzhou University, China

***Correspondence:**
Yuanyuan Wu
wyyshutcm@163.com
Jianfeng Yang
yangjianfeng@shutcm.edu.cn
Xiqiu Zhou
zhouxiqiu1970@aliyun.com
Sheng Liu
lslhtcm@163.com

[†]These authors 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: 25 August 2021
Accepted: 10 December 2021
Published: 01 February 2022

Citation:
Li F, Shi Y, Yang X, Luo Z, Zhang G, Yu K, Li F, Chen L, Zhao Y, Xie Y, Wu Y, Yang J, Zhou X and Liu S (2022) Anhydroicaritin Inhibits EMT in Breast Cancer by Enhancing GPX1 Expression: A Research Based on Sequencing Technologies and Bioinformatics Analysis. Front. Cell Dev. Biol. 9:764481. doi: 10.3389/fcell.2021.764481

expression levels in BC. Western blot and RT-PCR experiments revealed that AHI increased the protein and mRNA levels of GPX1. Molecular docking and DARTS experiments confirmed the direct binding combination between AHI and GPX1. After the evaluation of the EMT scores of 1,078 patients with BC, we found a potential anti-BC role of GPX1 possibly *via* suppression of the malignant EMT. The confocal immunofluorescence analysis showed that AHI increased E-cadherin expression levels and reduced vimentin expression levels in BC cells. Animal experiments showed that AHI significantly inhibited tumor growth. AHI also inhibited EMT by enhancing GPX1 and caspase3 cleavage, hence inhibiting EMT markers (i.e., N-cadherin and vimentin) and Ki-67.

Conclusion: GPX1 plays a critical role in BC, which may be a biomarker for the prognosis. In addition, AHI suppressed EMT by increasing GPX1 expression, which may serve as a potential therapy for BC treatment.

Keywords: RNA sequencing, bioinformatics analysis, anhydroicaritin, breast cancer, GPx1, epithelial to mesenchymal transformation

INTRODUCTION

The global cancer statistics report predicted that new cases of breast cancer (BC) would reach 2.3 million in 2020, accounting for 11.7% of all new cancers. BC has become the most common cancer overall (Sung et al., 2021). The treatment of BC has developed into a multidisciplinary comprehensive diagnosis and treatment model, which includes surgery, chemotherapy, radiotherapy, endocrine therapy, targeted therapy, and immunotherapy.

The advantages of traditional Chinese medicine (TCM) have gradually emerged with its core ideas of holistic view and syndrome differentiation-based treatment. TCM has played a critical role in BC prevention and treatment. The previous studies of our team have confirmed that TCM has apparent advantages in improving the quality of life of patients, reducing the toxicity of chemotherapy, and prolonging disease-free survival (DFS) (Hao et al., 2020; Wang Y. et al., 2020). Anhydroicaritin (AHI) is a flavonoid with antitumor biological and pharmacological effects (Nguyen et al., 2017). AHI is the primary metabolic product of *Epimedium brevicornum* after digestion and absorption (Zhou et al., 2021). Glutathione peroxidase 1 (GPX1) is a vital antioxidant enzyme that can catabolize the hydrogen oxide entering the human body and reduce peroxide's damage to the body (Huang et al., 2018). As one of the important cell biology processes, epithelial–mesenchymal transition (EMT) plays a key role in multiple links of tumor metastasis (Wilson et al., 2021; Yang et al., 2021). A previous study also revealed that GPX1 inhibits EMT progression in cancer and plays an anticancer role (Meng et al., 2018). However, research on the application of AHI and GPX1 in BC treatment remains limited, especially at the transcriptome level and the molecular mechanism.

The present study is an in-depth research on the anti-BC mechanism of AHI *via* RNA sequencing, bioinformatics analysis, network pharmacology, molecular docking, and *in vitro* and *in vivo* experiments. Network pharmacology and MTT (3-(4,5)-dimethylthiazolyl-3,5- diphenyltetrazolium bromide) experiments demonstrated the anti-BC potential of AHI. RNA sequencing revealed that AHI enhanced GPX1 in 4T1 and MDA-MB-231 BC cells. Comprehensive bioinformatics analysis showed that GPX1 is closely related to the prognosis of BC patients. Western blot analysis, reverse transcription–polymerase chain reaction (RT-PCR), molecular docking, and drug affinity responsive target stability (DARTS) experiments confirmed that AHI directly affects GPX1 expression. Animal experiments showed the anti-BC effect of AHI, possibly through the enhancement of GPX1 expression and thus EMT inhibition. The detailed strategy of this study is shown in **Figure 1**.

MATERIALS AND METHODS

Network Pharmacology Analysis

The targets of AHI were searched in the TCMSP (http://tcmspw.com/tcmsp.php) (Ru et al., 2014), Swiss target prediction (http://www.swisstargetprediction.ch/) (Daina et al., 2019), PubChem, TargetMol, and BioCrick databases. The names of the target proteins were uniformly converted into gene names using UniProt database (http://www.uniprot.org/) (Rovatsos et al., 2019). BC-related targets were collected and screened *via* GeneCards database (https://www.genecards.org/). DAVID database (https://david.ncifcrf.gov/home.jsp) was used to analyze Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment (Jiao et al., 2012). The AHI–targets–pathways network was constructed using Cytoscape 3.7.2.

Cell Cultures

Two BC cell lines, including murine 4T1 cells (TNBC, highly metastatic) and MDA-MB-231 (TNBC), were obtained from Chinese Academy of Sciences (Shanghai, China). 4T1 cells

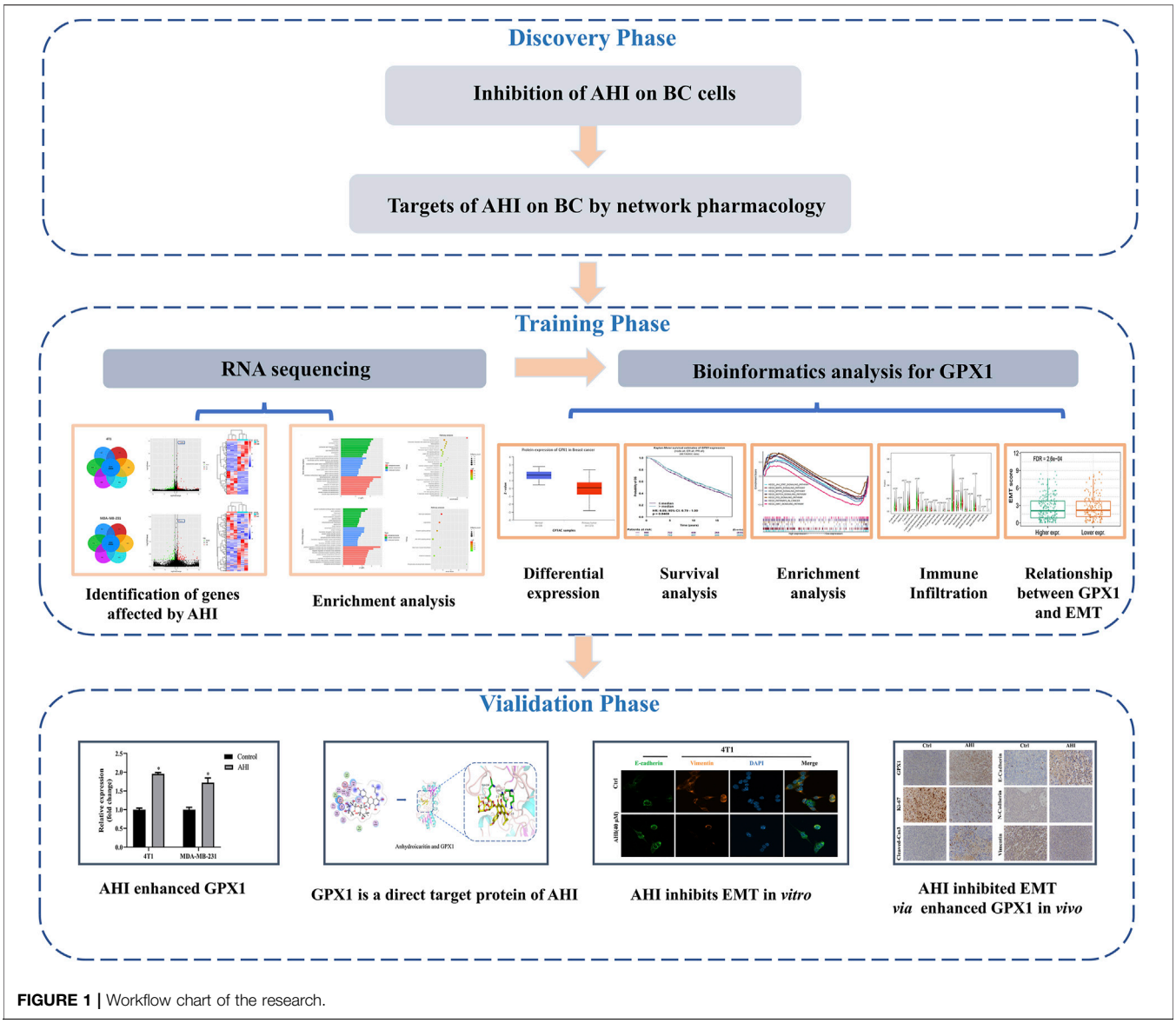


FIGURE 1 | Workflow chart of the research.

were maintained in RPMI 1640 medium supplemented with 10% (vol/vol) fetal bovine serum (FBS), 100 U/mL penicillin, and 100 mg/L streptomycin. MDA-MB-231 cells were maintained in Dulbecco modified eagle medium (DMEM) supplemented with 10% FBS, 100 U/mL penicillin, and 100 mg/L streptomycin. The cells were incubated at 37°C in a 5% CO_2 incubator with saturated humidity.

Cell Proliferation Assays

AHI (**Figure 2A**) was purchased from Shanghai Weiao Medical Technology Co., Ltd. (Shanghai, China). 4T1 and MAD-MB-231 cells (5×10^3/well) were separately seeded into 96-well plates. After 24 h, the culture medium was replaced with fresh medium containing various concentrations of AHI (0, 5, 10, 20, 40, 80, and 160 μM). After incubation for 24 h, cell viability was detected by MTT assay. Values of OD490 nm were measured using a BioTek instrument (Winooski, VT, USA). Data are expressed as the mean ± SD of at least three independent experiments.

RNA Extraction and Sequencing

After treatment of the 4T1 and MDA-MB-231 cells with AHI (40 μM), RNA from the 4T1 and MDA-MB-231 cells treated with the optimal concentration of AHI was extracted using Trizol reagent. To ensure that the samples were qualified and appropriate for sequencing, Nanodrop, Qubit 2.0, and Agilent 2100 were used to detect the purity, concentration, and integrity of the RNA samples, respectively. The sample's optical density 260/280 must be greater than 1.8, and the sample has to be free of protein or visible impurities. Library construction and RNA sequences were performed according to the manufacturer's instructions. RNA sequencing was performed on the 4T1 and MDA-MB-231 cells using NovaSeq. Construction of the

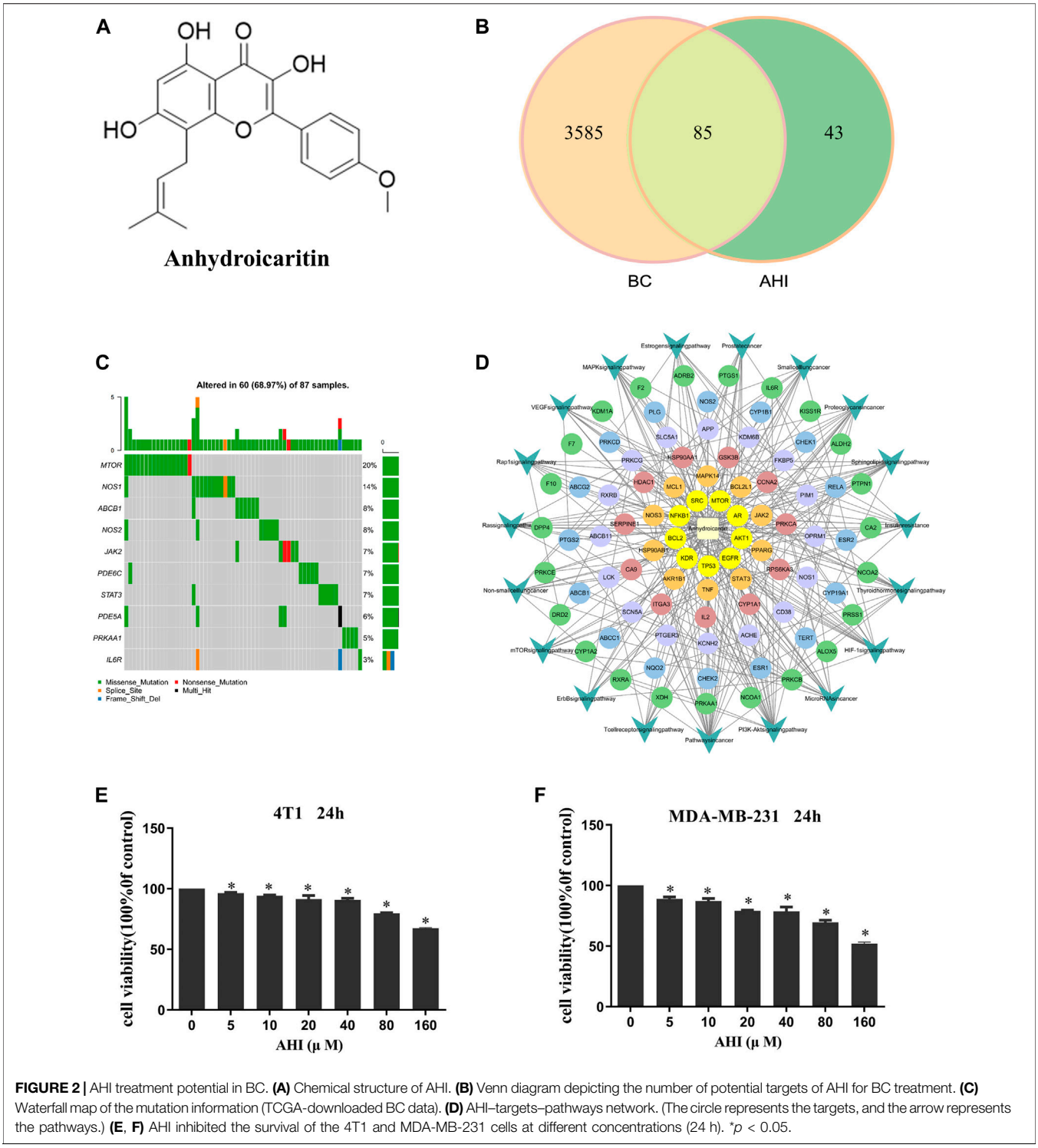


FIGURE 2 | AHI treatment potential in BC. **(A)** Chemical structure of AHI. **(B)** Venn diagram depicting the number of potential targets of AHI for BC treatment. **(C)** Waterfall map of the mutation information (TCGA-downloaded BC data). **(D)** AHI–targets–pathways network. (The circle represents the targets, and the arrow represents the pathways.) **(E, F)** AHI inhibited the survival of the 4T1 and MDA-MB-231 cells at different concentrations (24 h). *$p < 0.05$.

sequencing library and RNA sequencing were performed by Sangon (Shanghai, China) using Illumina No-voseq Platform. Differentially expressed genes (DEGs) were calculated *via* DESeq version 1.30.0. The p value was adjusted according to the false discovery rate (FDR). $p < 0.05$ and fold change ≥1.5 were set to be statistically significant. Gene Ontology (GO) functional enrichment and KEGG pathway analyses were performed for the bioinformatics analysis.

Analysis of Gene Expression Differences and Prognostic Value

The protein expression of GPX1 in BC was evaluated through CPTAC analysis (http://ualcan.path.uab.edu/) (Chen et al., 2019). The prognostic value of GPX1 was assessed using Breast Cancer Gene-Expression Miner version 4.7r (http://bcgenex.centregauducheau.fr/BC-GEM) (Jézéquel et al., 2012).

Protein Interaction Relationship and Gene Set Enrichment Analysis

The upstream and downstream protein regulatory relationship of GPX1 was predicted by STRING database (https://string-db.org/cgi/input.pl) (Szklarczyk et al., 2019). The enrichment pathways of the gene in each phenotype were ranked using the gene set enrichment analysis (GSEA) calculation method (Subramanian et al., 2007). The FDR value and normalized enrichment score (NES) were used to evaluate whether the pathway enrichment is significant (FDR <0.05 and NES >0.25).

Immune Infiltration Analysis

The gene expression profiles of 1,053 BC tissues and 111 normal breast tissues were downloaded from The Cancer Genome Atlas (TCGA) database (Blum et al., 2018). CiberSort algorithm was used to evaluate the correlation between the gene expression and the immune infiltration of 22 immune cell subtypes. The relationships among gene and B cells, $CD4^+$ T cells, $CD8^+$ T cells, neutrophils, macrophages, and dendritic cells in BC were evaluated through the deconvolution method of TIMER database (Li et al., 2017). The gene module used the relationship between the gene expression and the abundance of immune infiltration.

RT-PCR

Total RNA was isolated using Trizol reagent. The first strand of cDNA was synthetized using a reverse transcription kit (PrimeScript™ Synthesis Kit; Takara Bio, Inc., Dalian, China). RT-PCR experiment was performed using SYBR Premix Ex Taq Kit (Takara Bio, Inc.) on an Applied Biosystems 7500 Real-Time PCR System (Applied Biosystems, White Plains, NY, USA). The GAPDH was used as internal control. Primers for GPX1 and GAPDH are presented in **Supplementary Table S1**.

Western Blot Analysis

After the 4T1 and MDA-MB-231 cells were treated with AHI (0, 10, 20, or 40 μM) for 24 h, total cell protein lysates were extracted using RIPA lysis buffer that contained protease and phosphatase inhibitor cocktails. As determined *via* BCA analysis (Beyotime Biotechnology, Shanghai, China), protein lysates (20 μg) were then subjected to Western blot analysis. The primary antibody used in the analyses was GPX1 (1:1,000; ABclonal Technology, Wuhan, China). The secondary antibody was a goat anti-rabbit secondary antibody coupled to horseradish peroxidase (1:5,000; Proteintech, China). The target protein bands were visualized using a chemiluminescent kit (Beyotime Biotechnology) and analyzed using Image Lab software.

Molecular Docking

The energy of the compound downloaded from the PubChem database was minimized through Chem3D. We then converted the downloaded compound to the mol2 format. The target's crystal structure was downloaded from PDB database (https://www.rcsb.org/) (Burley et al., 2021). The additional process was undertaken using Pymol 2.1 and AutoDockTools 1.5.6. The treated compound was used as a small molecule ligand, and the target was used as a receptor. Molecular docking was conducted using AutoDockTools.

DARTS Analysis

Cells were exposed to AHI (40 μM) and DMEM for 3 h and then lysed using an NP-40 lysis buffer containing phosphatase and complete protease inhibitors. The protein concentration was measured using a BCA protein assay reagent. Pronase was then added with the ratios of 1:2,000 and 1:3,000 (pronase/total protein), and the mixtures were then incubated at room temperature for 30 min. Subsequently, the reaction products were analyzed *via* Western blot assays.

Analysis of the Relationship Between GPX1 and EMT

Tumor RNA-seq data for the 1,078 patients with BC were downloaded and normalized from the TCGA and gene expression omnibus databases. Student *t* test was performed to compare the statistical difference between the high and low GPX1 expression groups using ggplot2 R package. The EMT scores for individual cases were assessed as the average expression of genes upregulated in EMT minus the average expression of genes downregulated in EMT using 17 EMT signatures (i.e., ZEB1, ZEB2, Snai1, Snail2, Twist1, Twist2, FOXC2, VIM, FN1, SOX10, MMP2, MMP3, CDH1, CLDN3, CLDN4, CLDN7, and DSP) [PMID: 26245467]. Spearman correlation analysis was performed to evaluate the correlation between GPX1 expression and the EMT signatures without a normal distribution using ggstatsplot R package. The MCPCOUNTER algorithm was used to score 1,097 BC samples for immune cell infiltration. Statistical analyses were performed using R software version 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria). $p < 0.05$ was considered statistically significant.

Confocal Immunofluorescence Assay

When the 4T1 and MDA-MB-231 cells seeded in the confocal plates grew into 60% confluence, they were washed with phosphate-buffered saline (PBS), fixed in 4% paraformaldehyde for 5 min, permeabilized with 0.5% Triton X-100 for 5 min, and blocked with bovine serum albumin for 1 h at room temperature. The cells were then incubated with primary antibodies at 4°C overnight, followed by incubation with fluorophore-conjugated secondary antibody for 1 h. After washing, the samples were stained with DAPI and imaged using a confocal microscope. The primary antibodies were E-cadherin (1:100; Cell Signaling Technology 14472) and vimentin (1:100, A19607; ABclonal Technology).

Establishment of Animal Xenografts

All animal experiments were performed according to protocols approved by the Animal Care and Use Committee of Shanghai University of TCM. Six-week-old female nude mice (SLAC Laboratory Animal Co., Ltd., Shanghai, China) were anesthetized, and the human TNBC cell line MDA-MB-231 (1×10^7) premixed with Matrigel at a ratio of 1:1 was subcutaneously injected into the fourth pair of breast fat pads on the left side of each mouse. When the tumor approximately grew to 5 × 5 mm, the mice were randomized into the control group and AHI group (20 mg/kg) (n = 5). Mice received AHI treatment by intraperitoneal injection once every 2 days for 4 weeks. Tumor volumes were calculated using formula $V = 0.5 \times a \times b^2$, where V denotes tumor volume, a denotes maximum tumor diameter, and b denotes minimum tumor diameter. To assess the animals' overall health, their body weight was measured once a week as an indicator. The mice were euthanized at the experiments' termination. Tumors were isolated, imaged, weighed, and fixed with paraformaldehyde for further immunohistochemistry (IHC) evaluation.

Hematoxylin–Eosin Staining

To better evaluate AHI's anticancer effects, hematoxylin–eosin (H&E) staining was performed using a commercial H&E staining kit (Beyotime Biotechnology) following the manufacturer's instruction. Paraffin-embedded tumor sections with thickness of 4 μm were fixed onto poly-L-lysine–coated slides dewaxed twice with xylene for 10 min each, followed by gradual rehydration in 100%–70% ethanol, and then immersion in distilled water. Subsequently, 10% hematoxylin was applied to stain nuclear for 5 min followed by counterstaining cytoplasm with eosin for 2 min. After dehydration, hyalinization, and sealing with neutral gum, the sections were baked, and pictures were taken and analyzed.

IHC Analysis

For IHC analysis, the tumor tissues' slides were deparaffinized twice with xylene for 10 min and rehydrated with 100%–75% ethanol for 10 min. After washing with PBS for three times, the slices were boiled in 10-mm sodium citrate buffer solution (pH 6.0; Solarbio, Beijing, China) for 8 min to perform antigen repair. To eliminate endogenous peroxidase activity, sections were permeabilized with 3% hydrogen peroxide dissolved in methanol at room temperature in the dark and then blocked by 10% goat serum to reduce nonspecific binding. The samples were then washed with PBS three times and incubated with a 1: 100 diluted primary antibodies including Ki67 (CST, USA), cleaved caspase3 (CST, USA), GPX1 (ABclonal Technology, China), E-cadherin (CST, USA), N-cadherin (CST, USA), and vimentin (CST, USA) in a humid chamber at 4°C overnight, followed by incubation with a 1:200 dilution of biotinylated secondary antibodies. Immediately thereafter, 3,3-diaminobenzidine substrate (DAB, ZSGB-BIO, Beijing, China) was applied for color development, and a counterstain with Mayer's hematoxylin was performed. Digital images of stained sections were taken using a BX46 Olympus microscope (Olympus, Center Valley, PA, USA).

RESULTS

AHI's Therapeutic Potential in BC Treatment

A total of 128 AHI-related targets were screened through the TCMSP database, Swiss target prediction, PubChem, TargetMol, and BioCrick databases. A total of 3,670 BC-related targets were screened through the GeneCards database. Taking the intersection, we obtained 85 potential targets of AHI intervention in BC treatment (**Figure 2B**). The maftools package of R language was used to count the mutation data downloaded from TCGA. The waterfall map shows that the mutation rates of target genes were generally low (**Figure 2C**). The network of AHI-targeted genes and signal pathways was constructed through Cytoscape (**Figure 2D**). The MTT experiment showed that the survival rates of the 4T1 and MDA-MB-231 cells gradually decreased with the increased concentration of AHI (**Figures 2E,F**). The IC_{50} values of AHI in the MDA-MB-231 cells were 278.68 μM at 24 h. The IC_{50} values for the 4T1 cells were 319.83 μM at 24 h. To avoid AHI's cytotoxicity, the concentration of 40 μM, which is less than IC_{50}, was used in the succeeding experiments.

Transcriptome Profile of AHI-Treated BC Cells by RNA-Seq

To further identify genes affected by AHI in BC treatment, the RNA of the 4T1 and MDA-MB-231 cells treated with or without 40 μM AHI were extracted for RNA sequencing. A total of 19,535 genes were identified as core genes and used in the analysis of the DEGs in the 4T1 cells (**Figure 3A**). A total of 135 DEGs were subsequently identified between the AHI-treated group and the control group according to the criteria of fold change ≥1.5 and $p < 0.05$. Compared with the control group, 59 genes were significantly upregulated, whereas 76 genes were significantly downregulated (**Figure 3B**). For the MDA-MB-231 cells, 20,500 genes were identified as core genes (**Figure 3D**). A total of 250 DEGs were subsequently identified between the AHI-treated group and the control group (**Figure 3E**). Compared with the control group, 144 genes were significantly upregulated, whereas 106 genes were significantly downregulated. AHI-treated 4T1 and MDA-MB-231 cells possessed genes with hierarchical clustering that are distinct from those in the control cells (**Figures 3C,F**). Regarding potential genes involved in AHI's antitumor property, we found that GPX1 was the overlapping DEG in the 4T1 and MDA-MB-231 cells.

Identification of Potential Biomarkers of AHI in Treating TNBC

To further explore the biological process and KEGG pathway affected by AHI in BC treatment, functional enrichment analysis was conducted for genes affected by AHI in BC treatment. GO functional enrichment was performed to explore the potential molecular or biological functions affected by AHI treatment. All the DEGs were categorized into the following three major categories: cellular components, molecular functions, and

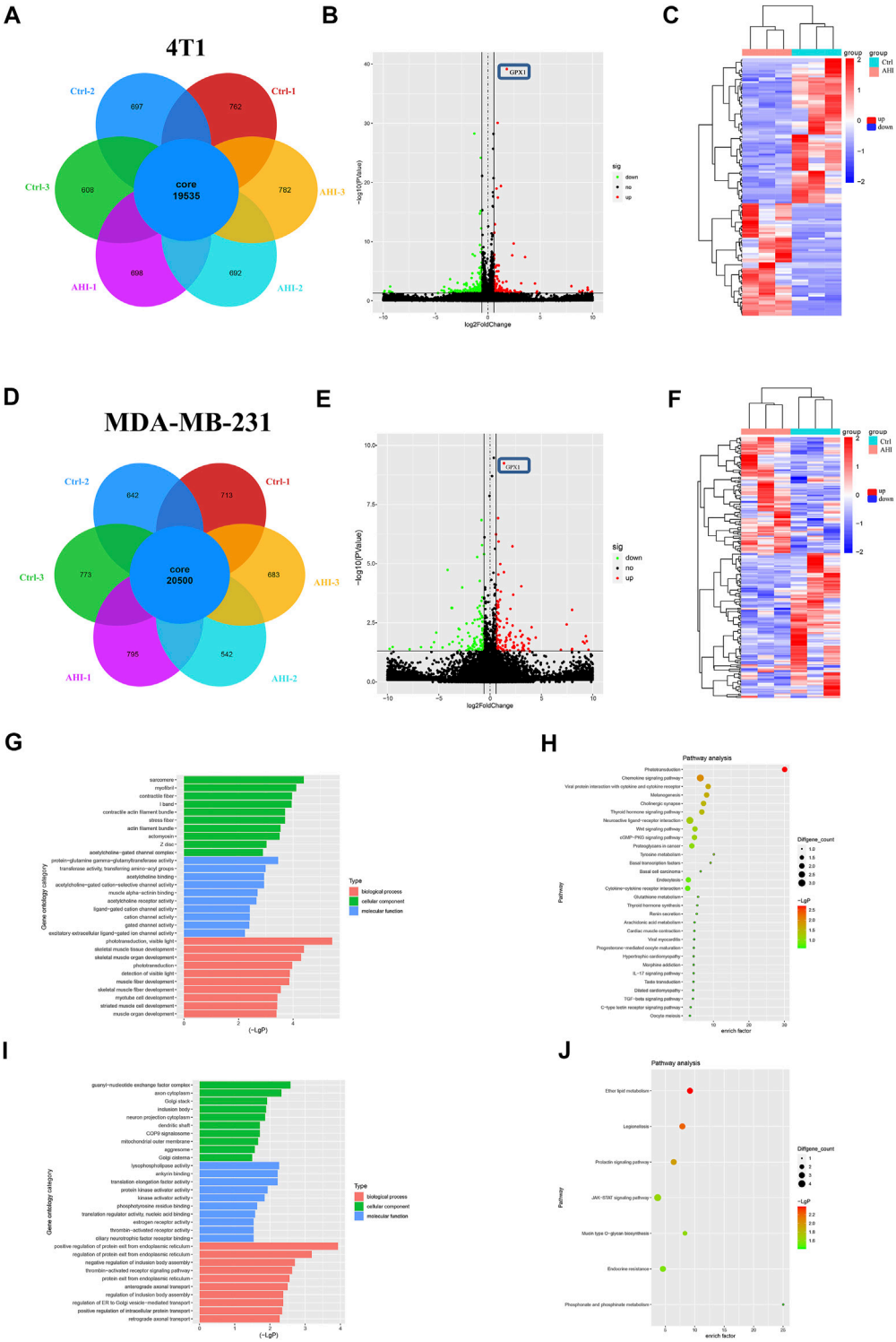


FIGURE 3 | Screening of DEGs by RNA-seq and enrichment analysis. **(A, D)** Analysis of DEGs after AHI treatment. **(B, E)** Identification of DEGs. **(C, F)** Volcano plot of the DEGs with FDR ≤0.05 and fold change ≥1.5. (The red dots indicate upregulated DEGs, green ones indicate downregulated DEGs, and black ones indicate no significant difference. The heatmap was drawn to show the DEGs.) **(G, I)** GO functional enrichment and **(H, J)** KEGG pathway enrichment of the DEGs.

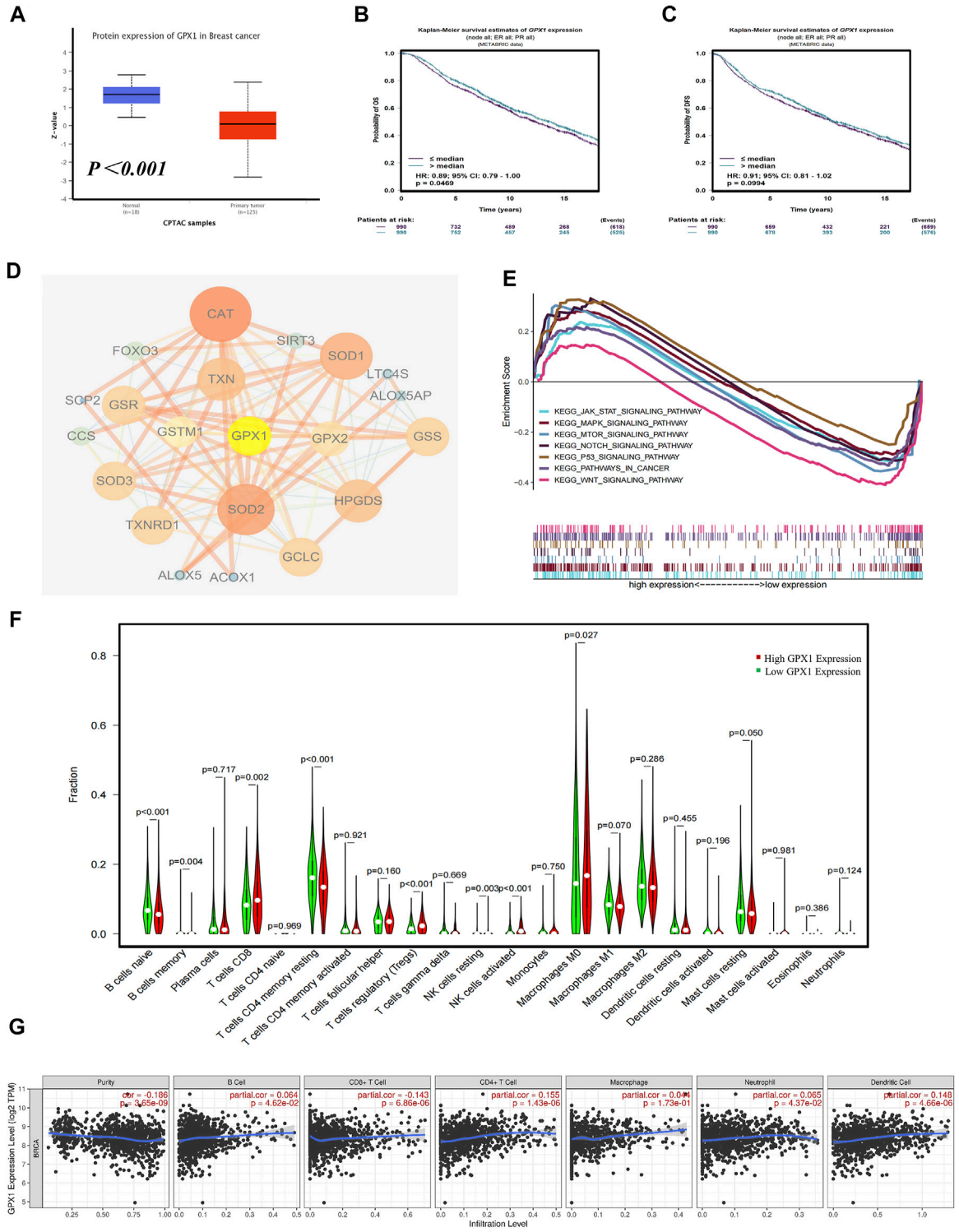


FIGURE 4 | Comprehensive bioinformatics analysis of GPX1. **(A)** Analysis of the protein expression of GPX1 between normal breast tissue and BC tissue. **(B)** The OS and **(C)** DFS of patients in the high and low GPX1 expression groups. **(D)** The reciprocal proteins of GPX1. **(E)** Related pathways of GPX1 by GSEA. **(F)** The infiltration levels of immune cells in the high and low GPX1 expression groups in patients with BC. **(G)** Correlation between GPX1 and the abundance of immune infiltration.

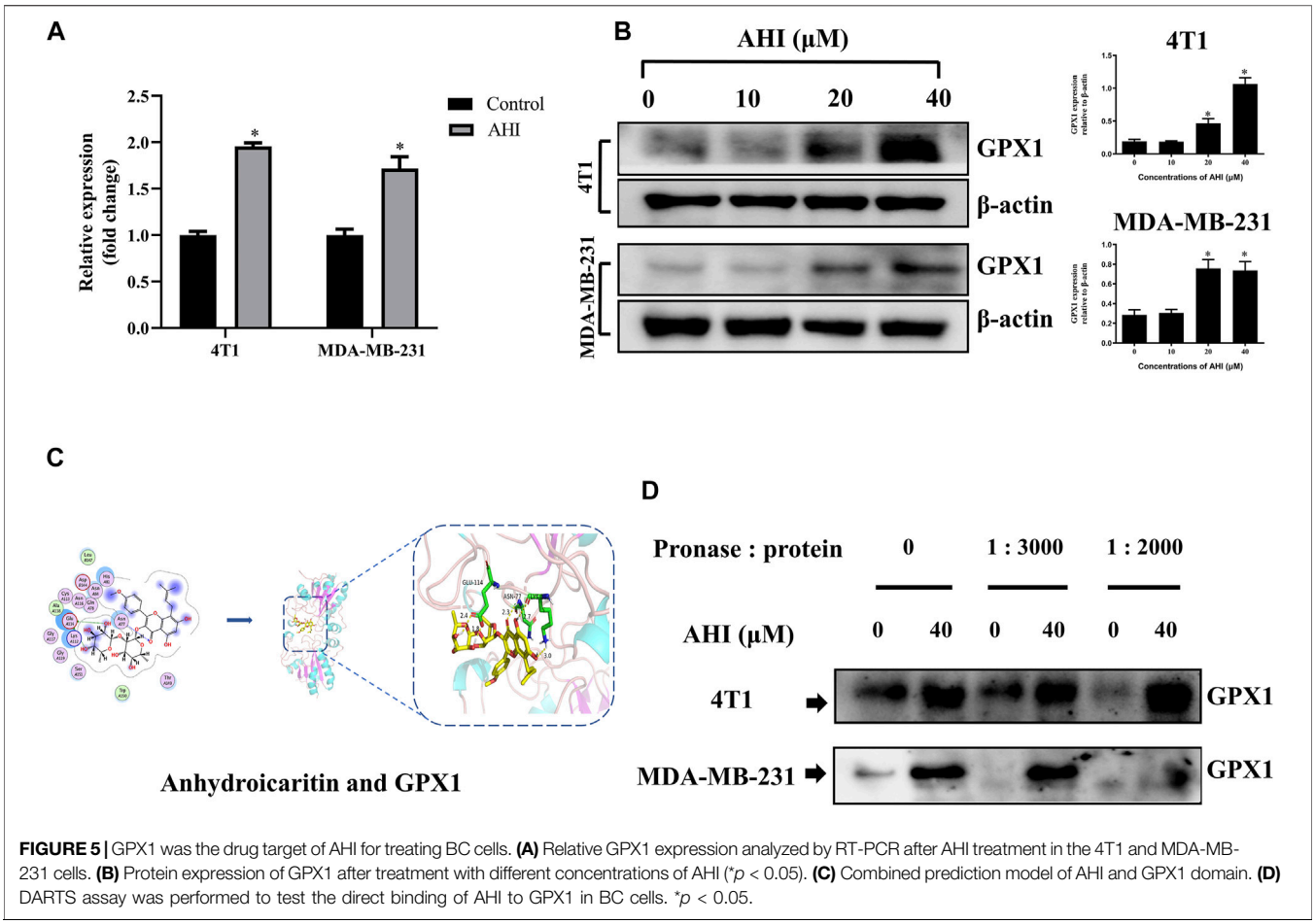


FIGURE 5 | GPX1 was the drug target of AHI for treating BC cells. **(A)** Relative GPX1 expression analyzed by RT-PCR after AHI treatment in the 4T1 and MDA-MB-231 cells. **(B)** Protein expression of GPX1 after treatment with different concentrations of AHI (*$p < 0.05$). **(C)** Combined prediction model of AHI and GPX1 domain. **(D)** DARTS assay was performed to test the direct binding of AHI to GPX1 in BC cells. *$p < 0.05$.

biological process. The DEGs in the AHI-treated group *versus* the control group in the 4T1 cells are significantly involved in sarcomere, myofibril, and contractile fiber, which are associated with tumor invasion and metastasis (**Figure 3G**), and their related signaling pathways, such as the Wnt signaling pathways and basal cell carcinoma (**Figure 3H**). Similarly, the GO biological processes significantly enriched by the AHI treatment in the MDA-MB-231 cells included axon cytoplasm, dendritic shaft (**Figure 3I**), and their related signaling pathways, such as JAK-STAT signaling pathways and mucin type O–glycan biosynthesis (**Figure 3J**). Considering the fundamental role of metastasis in BC pathogenesis, the current study inferred that GPX1 and signaling pathways associated with metastasis were AHI's mechanism in BC treatment.

Comprehensive Bioinformatics Analysis of GPX1

We conducted a multidimensional analysis of GPX1 through bioinformatics analysis. Compared with normal breast tissue, the protein expression of GPX1 is lower in BC ($p < 0.001$) (**Figure 4A**). The overall survival (OS) of BC patients in the high GPX1 expression group was longer than those in the low GPX1 expression group (hazard ratio [HR] = 0.89, *95% CI* 0.79–1.00, $p = 0.0469$) (**Figure 4B**), indicating a prognostic value of GPX1. However, the high GPX1 expression did not exhibit considerable advantages in DFS (HR = 0.91, *95% CI* 0.81–1.02, $p = 0.0994$) (**Figure 4C**). The proteins that interacted with GPX1 mainly include SOD1, SOD2, SOD3, GCLC, TXN, and CAT (**Figure 4D**). The GSEA data showed that GPX1 may play roles in the JAK-STAT, MAPK, mTOR, NOTCH, P53, and WNT signaling pathways (**Figure 4E**). The above signaling pathways are similar to our previous transcriptome results.

The immune infiltration analysis also showed that tumors with high GPX1 expression levels had more infiltrated CD8+T, Tregs, M0 macrophages, and natural killer cells. By contrast, naive B cells and CD4+T cells were significantly reduced ($p < 0.05$) (**Figure 4F**). The expression of GPX1 in BC is related to the abundance of immune infiltration and the immune score (**Figure 4G**, **Supplementary Figure S2**).

GPX1 is a Direct Target Protein of AHI in BC Cells

To confirm the GPX1 identified by RNA-Seq, RT-PCR validation was performed with the 4T1 and MDA-MB-231

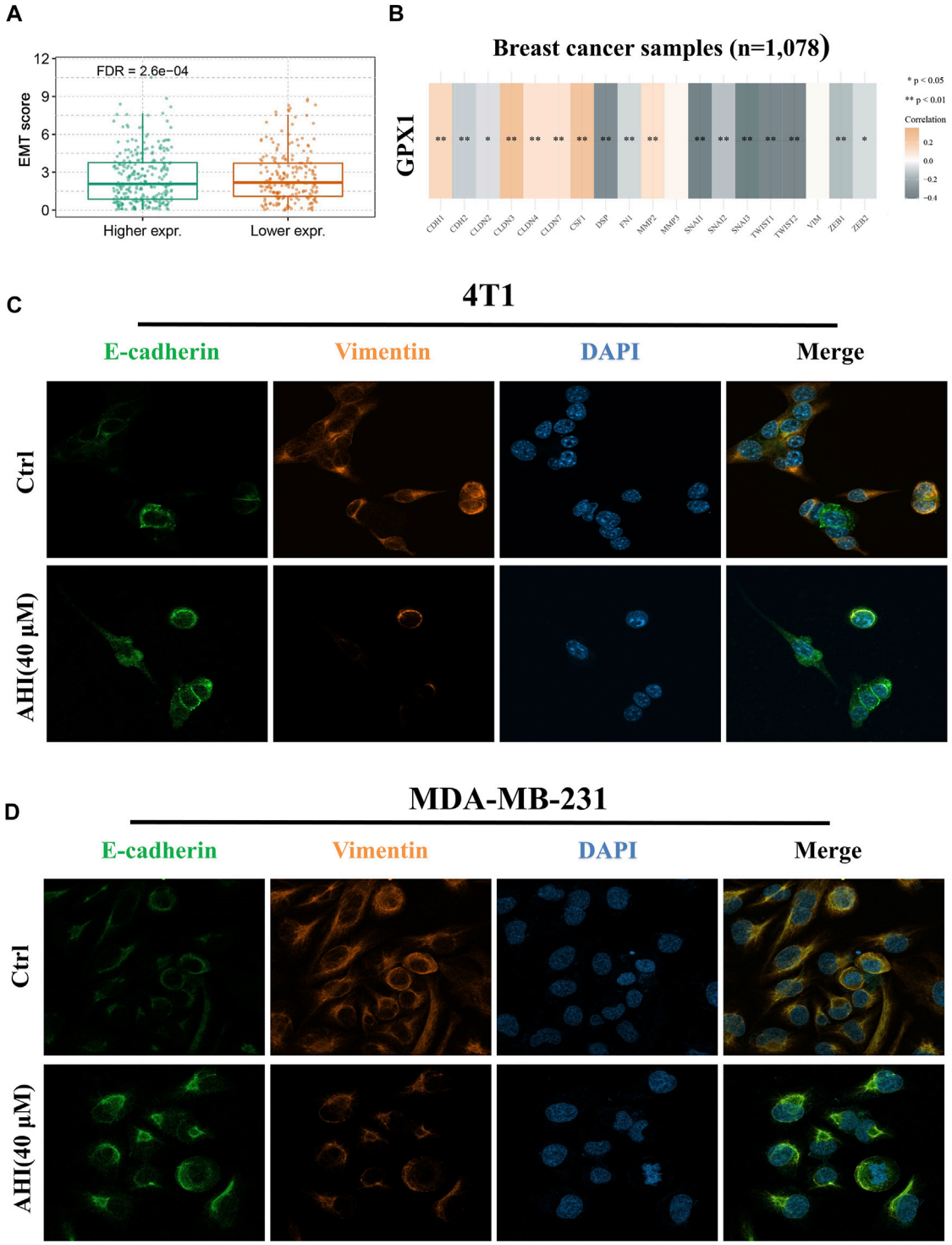


FIGURE 6 | Identification of AHI-inhibited EMT by enhancing GPX1 expression. **(A, B)** Relationship between GPX1 expression and EMT. **(C, D)** AHI significantly increased E-cadherin (green fluorescence) and decreased vimentin (orange fluorescence) expression (×630) by confocal immunofluorescence analysis.

cell lines. After 24 h of AHI (40 μM) treatment, the GPX1 expression levels were significantly higher than those of the control (**Figure 5A**). Similarly, the protein expression of GPX1 gradually increased with the increased AHI concentration (**Figure 5B**). Molecular docking and DARTS analyses were conducted to further confirm whether the GPX1 was the

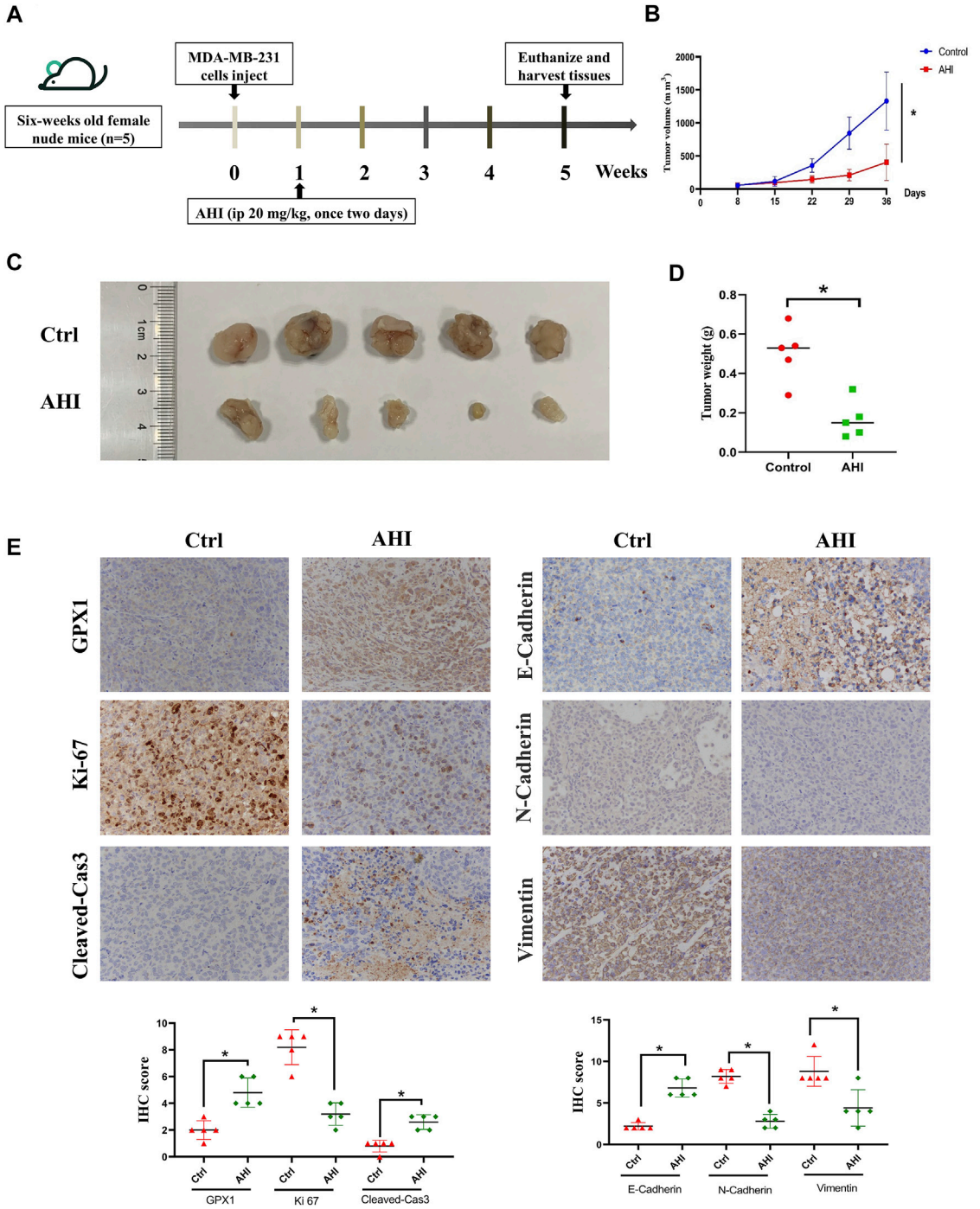


FIGURE 7 | AHI enhances GPX1 expression and inhibits EMT in the BC xenograft mouse model. **(A)** Diagram showing the scheme for tumor implantation and AHI treatment. **(B–D)** BC proliferation after AHI treatment in the mouse model. **(E)** IHC analysis of GPX1, proliferation-related, and EMT-related proteins after AHI treatment in mice. *$p < 0.05$.

solid drug target of AHI. The molecular docking results suggested a high affinity for docking between GPX1 and AHI (affinity = −7.62 kcal/mol) (**Figure 5C**, **Supplementary Table S2**). The DARTS experiment suggested that after AHI treatment, the GPX1 showed more stable property with pronase compared with that in the control group (**Figure 5D**). These results suggested that GPX1 may serve as a potential AHI target.

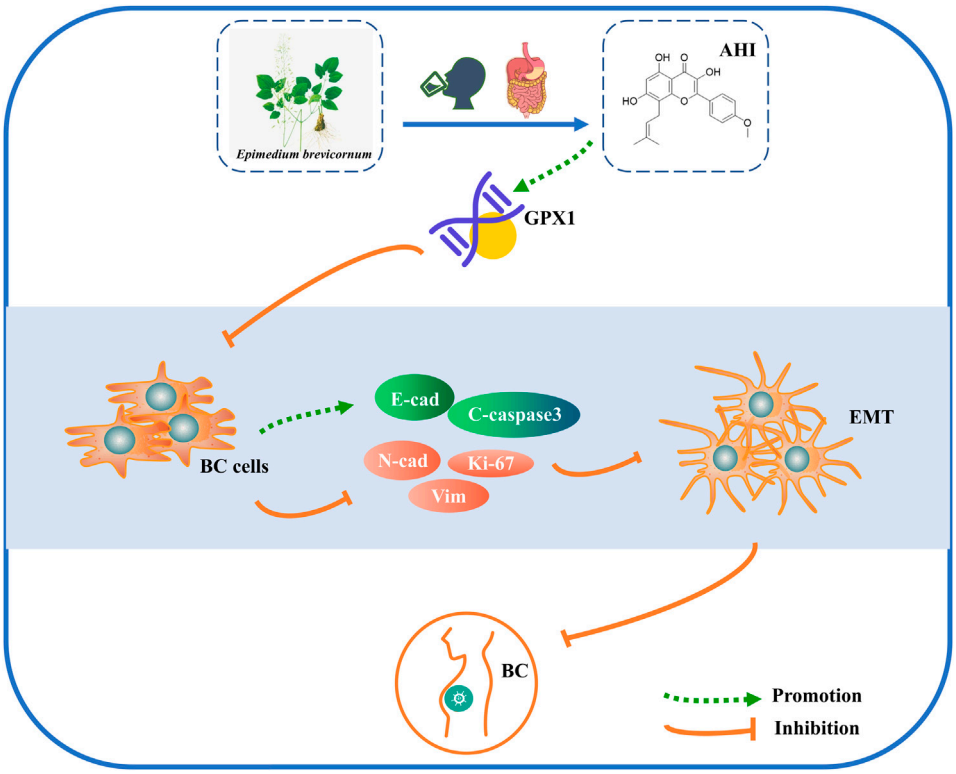


FIGURE 8 | Schematic diagram of the mechanism possibly involved in AHI-mediated BC.

GPX1 Is Negatively Correlated to the Malignant EMT in BC

To further verify the relationship between EMT and hub gene, the GSVA method was used to evaluate the EMT pathway score variation. After the evaluation of the EMT scores of 1,078 patients with BC, we found significantly decreased EMT scores in the high GPX1 expression group compared with those in the low GPX1 expression group (**Figure 6A**). Subsequently, we first quantified GPX1's EMT induction ability to comprehensively identify EMT-related genes. This evaluation suggested a significantly positive correlation between GPX1 expression and E-cadherin, CLDN3, CLDN4, CLND7, CSF1, MMP2, and MMP3. GPX1 also showed significantly negative correlation with N-cadherin, Snail family, TWIST1, TWIST2, ZEB1, and ZEB2. These findings implied the potential antitumor role of GPX1 possibly *via* suppression of the malignant EMT in BC (**Figure 6B**).

AHI Inhibits EMT in BC Cells

Confocal immunofluorescence analysis indicated that AHI markedly increased E-cadherin expression and decreased vimentin expression (**Figures 6C,D**). E-cadherin (green fluorescence) and vimentin (orange fluorescence) are mainly located in the cytoplasm, and DAPI (blue fluorescence) is located in the nucleus. The E-cadherin protein expression level in the AHI group was significantly upregulated, whereas the vimentin expression level was significantly downregulated compared with those in the control group. This result indicated that AHI may significantly inhibit the occurrence and development of EMT.

AHI Enhances GPX1 Expression and Inhibits EMT in the BC Xenograft Mouse Model

To assess AHI's anti-BC effect, a xenograft model of the MDA-MB-231 cells was established in BALB/c nude mice (**Figure 7A**). After 28 days of treatment, the average tumor size of the control group was $1{,}332 \pm 438\ mm^3$ compared with $405 \pm 275\ mm^3$ of the AHI-treated group, which resulted in a significant decrease in tumor volume and weight (**Figures 6B–D**). AHI has no toxicity to the livers and kidneys of BC nude mice (**Supplementary Figure S1**). The IHC analysis revealed that AHI significantly downregulated the expression level of the proliferation-related proteins Ki-67 and upregulated the expression level of GPX1 and the proapoptotic proteins cleaved caspase3. In addition, the classical EMT-associated biomarker levels revealed that the expression level of the epithelial biomarker E-cadherin, which mediates cell-to-cell homogenous adhesion, significantly increased, whereas the expression levels of the interstitial markers, namely, N-cadherin and vimentin, which mediate cell-to-cell matrix heterogeneous adhesion, markedly decreased (**Figure 6E**).

DISCUSSION

BC has become one of the main risk factors leading to death in females (Ferlay et al., 2015). The incidence of BC gradually tends to

be in the young population (Merino Bonilla et al., 2017). Endocrine therapy, targeted therapy, and chemotherapy are major treatments that significantly improve the prognosis of BC (Joseph et al., 2018; Zhao et al., 2020). However, BC is a complex heterogeneous disease with a large proportion of patients with progression, metastasis, and drug resistance (Echavarria et al., 2018).

This condition brings the current treatment into a bottleneck. We need to propose concepts of precise treatment to formulate an individualized BC treatment plan. The advances in genomics have made the establishment of individualized treatment options possible. Valuable methods to achieve targeted and individualized treatment plans for patients with BC include dimensional exploration of its biological nature, development of new targets based on omics data, and molecular biological characteristics.

GPX is a peroxidase containing selenocysteine. GPX consists of eight antioxidant enzymes, and their main function is to defend the organism against oxidative stress. A previous study revealed that GPXs have antitumor effects because they block reactive oxygen species (ROS) and regulate the redox signaling system, which plays a key role in tumor growth (Wu et al., 2021). Hydroperoxides stimulate cell proliferation and migration, but excessive hydroperoxides lead to cell apoptosis. The redox status of tumor cells therefore has a dual role. Downregulation of GPX1 expression levels leads to a decrease in the ability to respond to ROS, which is conducive to the accumulation of oxidative damage and promotes tumor progression (Méplan et al., 2013). GPX1 expression levels were found to be significantly downregulated by 7.4% in the tumorous breast tissue when compared with the nonmalignant one (Król et al., 2018). Lutein significantly inhibits the viability of BC cells, and the inhibition roles may be related to the upregulation of GPX1 expression and the downregulation of oxidative stress, thus blocking the nuclear factor κB signaling pathway (Chang et al., 2018).

As an antioxidant gene, the inhibitory effect of GPX1 on EMT has also been reported previously. GPX1 inhibits the EMT by regulating the Akt/GSK3β/Snail signaling axis in pancreatic adenocarcinoma (PDAC). Moreover, low GPX1 expression levels correlated with a poor survival rate in patients with PDAC (Meng et al., 2018). SEPP1 may inhibit the proliferation of HCC cells, accompanied by a decrease in ROS production and an increase in GPX1 expression (Wang J. et al., 2020). Glutathione metabolism and glutathione itself could be positive targets to prevent EMT in non–small cell lung carcinoma (Matadamas-Guzman et al., 2020). Changes in GPX1 expression may be related to EMT in CC cells (Yoo et al., 2009). Research on this matter is limited, considering that GPX1 serves as an oncogene in cancers (Gong et al., 2021; Lee et al., 2021). However, few studies have been conducted on GPX1 expression and EMT in BC.

E. brevicornum is a Chinese herbal medicine whose main functions are to nourish kidney yang, strengthen muscles and bones and dispel rheumatism. The main active ingredient of *E. brevicornum* is icariin. AHI is the main metabolite after icariin is taken orally into the body and decomposed by intestinal bacteria (Liu et al., 2008). AHI has antiosteoporosis (Zhai et al., 2011), antitumor (Nguyen et al., 2017), and antioxidant properties (Bao et al., 2012). AHI also significantly inhibits the activation of mouse macrophages and is a potential immunosuppressant (Lai et al., 2012).

Here, we found that the inhibition rates of 4T1 and MDA-MB-231 cells gradually decreased with varying concentrations of AHI. This finding indicates that AHI inhibits the survival of BC cells. By comparing the survival rates of 4T1 and MDA-MB-231 cells treated by different concentrations of AHI, 40 μM was identified to be the optimal concentration. 4T1 and MDA-MB-231 cells treated with this concentration were extracted for RNA sequencing. The result showed that AHI significantly enhanced GPX1 expression. The role of GPX1 expression in BC and its relationship with EMT remain unclear.

We thus performed a comprehensive biological analysis of GPX1 expression. We found that GPX1 expression relatively decreased in BC, and GPX1 expression was positively correlated with the survival duration. Analysis of GPX1 upstream and downstream regulatory proteins revealed that SOD1, SOD2, SOD3 (Yi et al., 2017; Li et al., 2018), GCLC (Matadamas-Guzman et al., 2020), and TXN (Cao et al., 2020) are related to EMT and cell migration and invasion. Network pharmacology analysis revealed that AHI's main targets are SRC (Qin et al., 2021), NFKB1 (Xu et al., 2021), and EGFR (Jin et al., 2021), which have close relationships with EMT. The gene set variation analysis results showed significantly decreased EMT scores in the high GPX1 expression group compared with those in the low GPX1 expression group. The significant correlation between GPX1 and EMT-related genes was detected through comprehensive identification.

We then confirmed the promotion of AHI on GPX1 from protein and gene expression *via* Western blot analysis and RT-PCR experiment. Molecular docking and DARTS experiments also proved that AHI and GPX1 bind well. The EMT pathway score variation results show that GPX1 expression possibly suppresses the malignant EMT in BC. Through the above research, we proposed a preliminary conjecture that AHI can inhibit EMT by enhancing GPX1 expression and play an anti-BC effect. Further animal experiments showed that the final tumor volume and weight of the AHI group were smaller compared with those in the control group. The IHC analysis showed that AHI could reduce the expression of Ki-67, N-cadherin, and vimentin, while increasing the expression of GPX1, E-cadherin, and cleaved caspase3.

Immunotherapy has become an emerging therapy in cancer treatment. Programmed cell death protein-1 and programmed cell death protein 1 ligand inhibitors can benefit patients with metastatic BC (Page et al., 2019). However, the immunotherapy of BC remains a major challenge to medical research (Jiang et al., 2021; Tabana et al., 2021). In the present study, we found that GPX1 is related to the immune infiltration of BC. Whether GPX1 can be used as a BC-related immune checkpoint remains to be further verified.

CONCLUSION

The present study determined a potential mechanism of AHI against BC by conducting bioinformatics and experimental

studies. AHI plays an anti-BC effect by inhibiting EMT progression. The potential mechanism may be related with enhancing the expression of GPX1, E-cadherin, and cleaved caspase3 and inhibiting the expression of N-cadherin, vimentin, and Ki-67. AHI may significantly upregulate GPX1 expression in BC, thus effectively suppressing EMT (**Figure 8**).

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: NCBI (accession: SAMN21399181-SAMN21399192).

ETHICS STATEMENT

The animal study was reviewed and approved by Ethics Committee of Shanghai University of Traditional Chinese Medicine.

AUTHOR CONTRIBUTIONS

FFL, YS, and XY made substantial contributions to the research design, acquisition, analysis, and interpretation of data. ZL was responsible for the layout of the pictures. GZ performed the DARTS experiments. YX was responsible for writing the manuscript. LC conducted the MT experiments. FL and YZ performed the Western blot and RT-PCR experiments. YW, JY, SL, and XZ provided substantial contributions to the conceptualization and design of the study. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (81774308) and the Longhua Hospital Science and Technology Innovation Project (012-6).

ACKNOWLEDGMENTS

The authors would like to appreciate all the participating centers and members.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.764481/full#supplementary-material

REFERENCES

Bao, Y., Yang, J. X., and Sun, R. G. (2012). *In Vitro* Antioxidant Study of Icariin and Dehydrated Icariin. *J. Shaanxi Normal Univ.Nat. Sci. Ed.* 40, 67–70. doi:10.15983/j.cnki.jsnu.2012.03.008

Blum, A., Wang, P., and Zenklusen, J. C. (2018). SnapShot: TCGA-Analyzed Tumors. *Cell* 173, 530. doi:10.1016/j.cell.2018.03.059

Burley, S. K., Bhikadiya, C., Bi, C., Bittrich, S., Chen, L., Crichlow, G. V., et al. (2021). RCSB Protein Data Bank: Powerful New Tools for Exploring 3D Structures of Biological Macromolecules for Basic and Applied Research and Education in Fundamental Biology, Biomedicine, Biotechnology, Bioengineering and Energy Sciences. *Nucleic Acids Res.* 49, D437–D451. doi:10.1093/nar/gkaa1038

Cao, M. Q., You, A. B., Cui, W., Zhang, S., Guo, Z. G., Chen, L., et al. (2020). Cross Talk between Oxidative Stress and Hypoxia via Thioredoxin and HIF-2α Drives Metastasis of Hepatocellular Carcinoma. *FASEB j.* 34, 5892–5905. doi:10.1096/fj.202000082R

Chang, J. Z., Wang, C., Li, Y. C., Liu, Q., Shen, Y. J., Lu, K., et al. (2018). The Effect of Lutein on the Viability of Breast Cancer Cells. *Chin. J. Pathophysiol.* 34, 930–933. doi:10.3969/j.issn.1000-4718.2018.05.0

Chen, F., Chandrashekar, D. S., Varambally, S., and Creighton, C. J. (2019). Pan-cancer Molecular Subtypes Revealed by Mass-Spectrometry-Based Proteomic Characterization of More Than 500 Human Cancers. *Nat. Commun.* 10, 5679. doi:10.1038/s41467-019-13528-0

Daina, A., Michielin, O., and Zoete, V. (2019). SwissTargetPrediction: Updated Data and New Features for Efficient Prediction of Protein Targets of Small Molecules. *Nucleic Acids Res.* 47, W357–W364. doi:10.1093/nar/gkz382

Echavarria, I., López-Tarruella, S., Picornell, A., García-Saenz, J. Á., Jerez, Y., Hoadley, K., et al. (2018). Pathological Response in a Triple-Negative Breast Cancer Cohort Treated with Neoadjuvant Carboplatin and Docetaxel According to Lehmann's Refined Classification. *Clin. Cancer Res.* 24, 1845–1852. doi:10.1158/1078-0432.CCR-17-1912

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., et al. (2015). Cancer Incidence and Mortality Worldwide: Sources, Methods and Major Patterns in GLOBOCAN 2012. *Int. J. Cancer* 136, E359–E386. doi:10.1002/ijc.29210

Gong, T.-T., Guo, Q., Li, X., Zhang, T.-N., Liu, F.-H., He, X.-H., et al. (2021). Isothiocyanate Iberin Inhibits Cell Proliferation and Induces Cell Apoptosis in the Progression of Ovarian Cancer by Mediating ROS Accumulation and GPX1 Expression. *Biomed. Pharmacother.* 142, 111533. doi:10.1016/j.biopha.2021.111533

Hao, W., Shi, Y., Qin, Y., Sun, C., Chen, L., Wu, C., et al. (2020). *Platycodon Grandiflorum* Protects against Anthracycline-Induced Cardiotoxicity in Early Breast Cancer Patients. *Integr. Cancer Ther.* 19, 153473542094501. doi:10.1177/1534735420945017

Huang, J.-Q., Zhou, J.-C., Wu, Y.-Y., Ren, F.-Z., and Lei, X. G. (2018). Role of Glutathione Peroxidase 1 in Glucose and Lipid Metabolism-Related Diseases. *Free Radic. Biol. Med.* 127, 108–115. doi:10.1016/j.freeradbiomed.2018.05.077

Jézéquel, P., Campone, M., Gouraud, W., Guérin-Charbonnel, C., Leux, C., Ricolleau, G., et al. (2012). Bc-GenExMiner: an Easy-To-Use Online Platform for Gene Prognostic Analyses in Breast Cancer. *Breast Cancer Res. Treat.* 131, 765–775. doi:10.1007/s10549-011-1457-7

Jiang, Y.-Q., Wang, Z.-X., Zhong, M., Shen, L.-J., Han, X., Zou, X., et al. (2021). Investigating Mechanisms of Response or Resistance to Immune Checkpoint Inhibitors by Analyzing Cell-Cell Communications in Tumors before and after Programmed Cell Death-1 (PD-1) Targeted Therapy: An Integrative Analysis Using Single-Cell RNA and Bulk-RNA Sequencing Data. *Oncoimmunology* 10, 1908010. doi:10.1080/2162402X.2021.1908010

Jiao, X., Sherman, B. T., Huang, D. W., Stephens, R., Baseler, M. W., Lane, H. C., et al. (2012). DAVID-WS: a Stateful Web Service to Facilitate Gene/protein List Analysis. *Bioinformatics* 28, 1805–1806. doi:10.1093/bioinformatics/bts251

Jin, W., Yin, H., Li, H., Yu, X. J., Xu, H. X., and Liu, L. (2021). Neutrophil Extracellular DNA Traps Promote Pancreatic Cancer Cells Migration and Invasion by Activating EGFR/ERK Pathway. *J. Cel. Mol. Med.* 25, 5443–5456. doi:10.1111/jcmm.16555

Joseph, C., Papadaki, A., Althobiti, M., Alsaleem, M., Aleskandarany, M. A., and Rakha, E. A. (2018). Breast Cancer Intratumour Heterogeneity: Current Status and Clinical Implications. *Histopathology* 73, 717–731. doi:10.1111/his.13642

Król, M. B., Galicki, M., Grešner, P., Wieczorek, E., Jabłońska, E., Reszka, E., et al. (2018). ESR1 and GPX1 Genes Expression Level in Human Malignant and Non-malignant Breast Tissues. *Acta Biochim. Pol.* 65, 51–57. doi:10.18388/abp.2016_1425

Lai, X. Q., Huang, X. Y., and Zeng, Y. Y. (2012). Effects of Dehydrated Icariin on Immune Function of Mouse Macrophages. *J. Cell Mol. Immunol.* 28, 374–376+380. doi:10.13423/j.cnki.cjcmi.006363

Lee, S., Lee, E.-K., Kang, D. H., Lee, J., Hong, S. H., Jeong, W., et al. (2021). Glutathione Peroxidase-1 Regulates ASK1-dependent Apoptosis via Interaction with TRAF2 in RIPK3-Negative Cancer Cells. *Exp. Mol. Med.* 53, 1080–1091. doi:10.1038/s12276-021-00642-7

Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. S., et al. (2017). TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* 77, e108–e110. doi:10.1158/0008-5472.CAN-17-0307

Li, W., Jiang, Z., Xiao, X., Wang, Z., Wu, Z., Ma, Q., et al. (2018). Curcumin Inhibits Superoxide Dismutase-Induced Epithelial-To-Mesenchymal Transition via the PI3K/Akt/NF-κB Pathway in Pancreatic Cancer Cells. *Int. J. Oncol.* 52, 1593–1602. doi:10.3892/ijo.2018.4295

Liu, Z. C., Xu, Y., and Chen, S. Z. (2008). Study on Pharmacokinetics and Tis-Sue Distribution of Icarisidelin Rats by HPLC-UV Assay. *Chin. Pharm. J.* 43, 848–851. doi:10.3321/j.issn:1001-2494.2008.11.013

Matadamas-Guzman, M., Zazueta, C., Rojas, E., and Resendis-Antonio, O. (2020). Analysis of Epithelial-Mesenchymal Transition Metabolism Identifies Possible Cancer Biomarkers Useful in Diverse Genetic Backgrounds. *Front. Oncol.* 10, 1309. doi:10.3389/fonc.2020.01309

Méplan, C., Dragsted, L. O., Ravn-Haren, G., Tjønneland, A., Vogel, U., and Hesketh, J. (2013). Association between Polymorphisms in Glutathione Peroxidase and Selenoprotein P Genes, Glutathione Peroxidase Activity, HRT Use and Breast Cancer Risk. *PloS one* 8, e73316. doi:10.1371/journal.pone.0073316

Meng, Q., Shi, S., Liang, C., Liang, D., Hua, J., Zhang, B., et al. (2018). Abrogation of Glutathione Peroxidase-1 Drives EMT and Chemoresistance in Pancreatic Cancer by Activating ROS-Mediated Akt/GSK3β/Snail Signaling. *Oncogene* 37, 5843–5857. doi:10.1038/s41388-018-0392-z

Merino Bonilla, J. A., Torres Tabanera, M., and Ros Mendoza, L. H. (2017). El cáncer de mama en el siglo XXI: de la detección precoz a los nuevos tratamientos. *Radiología* 59, 368–379. doi:10.1016/j.rx.2017.06.003

Nguyen, V.-S., Shi, L., Wang, S.-C., and Wang, Q.-A. (2016). Synthesis of Icaritin and β-anhydroicaritin Mannich Base Derivatives and Their Cytotoxic Activities on Three Human Cancer Cell Lines. *Acamc* 17, 137–142. doi:10.2174/1871520616666160404111210

Page, D. B., Bear, H., Prabhakaran, S., Gatti-Mays, M. E., Thomas, A., Cobain, E., et al. (2019). Two May Be Better Than One: PD-1/pd-L1 Blockade Combination Approaches in Metastatic Breast Cancer. *NPJ breast cancer* 5, 34. doi:10.1038/s41523-019-0130-x

Qin, L., Chen, J., Lu, D., Jain, P., Yu, Y., Cardenas, D., et al. (2021). Development of Improved SRC-3 Inhibitors as Breast Cancer Therapeutic Agents. *Endocrine-Relat. Cancer* 28, 657–670. doi:10.1530/ERC-20-0402

Rovatsos, M., Altmanová, M., Augstenová, B., Mazzoleni, S., Velenský, P., and Kratochvíl, L. (2019). ZZ/ZW Sex Determination with Multiple Neo-Sex Chromosomes Is Common in Madagascan Chameleons of the Genus *Furcifer* (Reptilia: Chamaeleonidae). *Genes* 10, 1020. doi:10.3390/genes10121020

Ru, J., Li, P., Wang, J., Zhou, W., Li, B., Huang, C., et al. (2014). TCMSP: a Database of Systems Pharmacology for Drug Discovery from Herbal Medicines. *J. Cheminform.* 6, 13. doi:10.1186/1758-2946-6-13

Subramanian, A., Kuehn, H., Gould, J., Tamayo, P., and Mesirov, J. P. (2007). GSEA-P: a Desktop Application for Gene Set Enrichment Analysis. *Bioinformatics (Oxford, England)* 23, 3251–3253. doi:10.1093/bioinformatics/btm369

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, 209–249. doi:10.3322/caac.21660

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING V11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-wide Experimental Datasets. *Nucleic Acids Res.* 47, D607–D613. doi:10.1093/nar/gky1131

Tabana, Y., Okoye, I. S., Siraki, A., Elahi, S., and Barakat, K. H. (2021). Tackling Immune Targets for Breast Cancer: Beyond PD-1/pd-L1 Axis. *Front. Oncol.* 11, 628138. doi:10.3389/fonc.2021.628138

Wang, J., Shen, P., Liao, S., Duan, L., Zhu, D., Chen, J., et al. (2020a). Selenoprotein P Inhibits Cell Proliferation and ROX Production in HCC Cells. *PloS one* 15, e0236491. doi:10.1371/journal.pone.0236491

Wang, Y., Li, J.-W., Qin, Y.-N., Sun, C.-P., Chen, J.-J., Ruan, Y.-Y., et al. (2020). Clinical Observation on the Effect of Chinese Medicine-"TCM Formula" Intervention on Recurrence and Metastasis of Triple Negative Breast Cancer. *Complement. Ther. Med.* 52, 102456. doi:10.1016/j.ctim.2020.102456

Wilson, M. M., Callens, C., Le Gallo, M., Mironov, S., Ding, Q., Salamagnon, A., et al. (2021). An EMT-Primary Cilium-GLIS2 Signaling axis Regulates Mammogenesis and Claudin-Low Breast Tumorigenesis. *Sci. Adv.* 7, eabf6063. doi:10.1126/sciadv.abf6063

Wu, W., Li, D., Feng, X., Zhao, F., Li, C., Zheng, S., et al. (2021). A Pan-Cancer Study of Selenoprotein Genes as Promising Targets for Cancer Therapy. *BMC Med. Genomics* 14, 78. doi:10.1186/s12920-021-00930-1

Xu, G., Zhao, H., Xu, J., Zhang, Y., Qi, X., and Shi, A. (2021). Hard Antler Extract Inhibits Invasion and Epithelial-Mesenchymal Transition of Triple-Negative and Her-2+ Breast Cancer Cells by Attenuating Nuclear Factor-κB Signaling. *J. Ethnopharmacol.* 269, 113705. doi:10.1016/j.jep.2020.113705

Yang, Y., Zhao, B., Lv, L., Yang, Y., Li, S., and Wu, H. (2021). FBXL10 Promotes EMT and Metastasis of Breast Cancer Cells via Regulating the Acetylation and Transcriptional Activity of SNAI1. *Cell Death Discov.* 7, 328. doi:10.1038/s41420-021-00722-7

Yi, L., Shen, H., Zhao, M., Shao, P., Liu, C., Cui, J., et al. (2017). Inflammation-mediated SOD-2 Upregulation Contributes to Epithelial-Mesenchymal Transition and Migration of Tumor Cells in Aflatoxin G1-Induced Lung Adenocarcinoma. *Sci. Rep.* 7, 7953. doi:10.1038/s41598-017-08537-2

Yoo, H.-J., Yun, B.-R., Kwon, J.-H., Ahn, H.-S., Seol, M.-A., Lee, M.-J., et al. (2009). Genetic and Expression Alterations in Association with the Sarcomatous Change of Cholangiocarcinoma Cells. *Exp. Mol. Med.* 41, 102–115. doi:10.3858/emm.2009.41.2.013

Zhai, Y. K., Li, Z. Z., Chen, K. M., Zhang, N., Cheng, G. Z., and Zhu, R. Q. (2011). Comparative Study of the Effects of Icariin and Dehydrated Icariin on the Osteogenic Differentiation of Bone Marrow Mesenchymal Stem Cells. *Chin. Pharm. J.* 46, 837–842. doi:10.13863/j.issn1001-4454.2011.06.030

Zhao, S., Ma, D., Xiao, Y., Li, X. M., Ma, J. L., Zhang, H., et al. (2020). Molecular Subtyping of Triple-Negative Breast Cancers by Immunohistochemistry: Molecular Basis and Clinical Relevance. *Oncol.* 25, e1481–e1491. doi:10.1634/theoncologist.2019-0982

Zhou, M., Zheng, W., Sun, X., Yuan, M., Zhang, J., Chen, X., et al. (2021). Comparative Analysis of Chemical Components in Different Parts of Epimedium Herb. *J. Pharm. Biomed. Anal.* 198, 113984. doi:10.1016/j.jpba.2021.113984

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 February 2022
doi: 10.3389/fcell.2021.815895

Strategy for Scanning Peptide-Coding Circular RNAs in Colorectal Cancer Based on Bioinformatics Analysis and Experimental Assays

Zhanghan Chen[1,2†], Zhipeng Qi[1,2†], Dongli He[3†], Jingyi Liu[1,2], Enpan Xu[1,2], Bing Li[1,2], Shilun Cai[1,2], Di Sun[1,2], Yirong Cheng[1,2], Qiang Shi[1,2] and Yunshi Zhong[1,2*‡]*

[1]Endoscopy Center, Zhongshan Hospital of Fudan University, Shanghai, China, [2]Endoscopy Research Institute of Fudan University, Shanghai, China, [3]Department of Internal Medicine of Xuhui Hospital, Affiliated Zhongshan Hospital, Fudan University, Shanghai, China

Background: Colorectal cancer (CRC) is the third most common cause of cancer deaths worldwide. Numerous studies have reported that circular RNAs (circRNAs) have important functions in CRC. It was first thought that circRNAs were non-coding RNA; however, more recently they were discovered to encode peptides and play a pivotal role in cancer development and progression. It was shown that most circRNAs possess coding potential; however, not all of them can truly encode peptides. Therefore, a practical strategy to scan for coding circRNAs is needed.

Method: Sequence analyses included open reading frame (ORF) prediction, coding peptide prediction, and the identification of unique sequences. Then, experimental assays were used to verify the coded peptides, liquid chromatography-tandem mass spectrometry (LC-MS/MS) was introduced to detect sequences of circRNAs with coding potential, and Western blot was used to identify the encoded peptides. Finally, the functions of the circRNAs were primarily explored.

Result: An efficient strategy for searching circRNAs with coding potential was created. We verified this schedule using public databases and LC-MS/MS, then two of these circRNAs were selected for further verification. We used commercial antibodies that can also identify the predicted peptides to test the coded peptides. The functions of the circRNAs were explored primarily, and the results showed that they were mainly involved in the promotion of proliferation and invasion ability.

Discussion: We have constructed an efficient strategy of scanning circRNAs with coding potential. Our strategy helped to provide a more convenient pathway for identifying circRNA-derived peptides, which can be a potential therapeutic target or a diagnostic biomarker.

Keywords: colorectal cancer, circRNAs, short peptides, LC-MS/MS, bioinformatic analysis

OPEN ACCESS

Edited by:
Geng Chen,
GeneCast Biotechnology Co., Ltd., China

Reviewed by:
Faezeh Vakhshiteh,
Iran University of Medical Sciences, Iran
Xiaoxu Shen,
Sichuan Agricultural University, China
Chris Nicholas James Young,
De Montfort University, United Kingdom

***Correspondence:**
Qiang Shi
shiqiangqy@126.com
Yunshi Zhong
zhong.yunshi@zs-hospital.sh.cn

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

[‡]ORCID:
Yunshi Zhong
orcid.org/0000-0002-3128-3168

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

Received: *16 November 2021*
Accepted: *23 December 2021*
Published: *25 February 2022*

Citation:
Chen Z, Qi Z, He D, Liu J, Xu E, Li B, Cai S, Sun D, Cheng Y, Shi Q and Zhong Y (2022) Strategy for Scanning Peptide-Coding Circular RNAs in Colorectal Cancer Based on Bioinformatics Analysis and Experimental Assays. Front. Cell Dev. Biol. 9:815895. doi: 10.3389/fcell.2021.815895

INTRODUCTION

Colorectal cancer (CRC) is now ranked third in terms of incidence and second in terms of mortality of all cancers worldwide (Sung et al., 2021). Even with the popularization of endoscopic surgery and the improvement of CRC treatment, efficient and potent biomarkers for early diagnosis and therapeutic targets are still needed (Siegel et al., 2020; Sung et al., 2021). Circular RNAs (circRNAs), a group of RNAs that possess a cyclic structure, were once believed to be non-coding and were reported to play a pivotal role in tumorigenesis and tumor development (Artemaki et al., 2020). Mechanistically, the circRNAs can act as miRNA sponges and regulate downstream targets. Previous research found that circCAMSAP1 regulated by ESRP1 could sponge miR-328-5p and then target E2F1 to mediate CRC progression (Zhou et al., 2020), and circHIPK3 could promote CRC proliferation and metastasis through the c-Myb/circHIPK3/miR-7 axis (Zeng et al., 2018). Moreover, circRNA can also interact with proteins and function as a tumor suppressor or promoter (Long et al., 2021). Researchers found that circPTK2 could promote the epithelial–mesenchymal transition (EMT) of CRC cells by interacting with the Ser38, Ser55, and Ser82 sites of the vimentin protein (Yang et al., 2020). Another circRNA, circ_cse1l, can downregulate the expression of *PCNA* and inhibit the proliferation of CRC cells by binding to eIF4A3 (Xu et al., 2020). It was not until recent years that circRNAs were discovered to have coding potential. Legnini et al. have identified a protein encoded by circ-ZNF609 in myogenesis (Legnini et al., 2017; Pamudurti et al., 2017). Moreover, researchers have determined the translatomes of 80 human hearts and identified undetected microproteins, which also included peptides encoded by circRNAs (van Heesch et al., 2019). Moreover, Zhang and his colleagues found several circRNAs that could encode peptides that functioned significantly in glioma (Yang et al., 2018; Zhang et al., 2018; Xia et al., 2019; Gao et al., 2021; Wu et al., 2021a). Furthermore, the coding potential of circRNAs can also be observed in liver cancer (Liang et al., 2019; Li et al., 2021), gastric cancer (Jiang et al., 2021), and colon cancer (Zheng et al., 2019; Pan et al., 2020). Therefore, exploring the coding potential and functions of circRNAs is important. In this study, we developed an effective schedule to scan circRNAs that possessed coding potential in CRC and found two circRNAs that were most likely to encode peptides. Preliminarily, we verified the existence of coded peptides and the functions of their derived circRNAs.

MATERIALS AND METHODS

Selection and Profiling of RNA Datasets

All eligible microarray datasets available up to December 2020 were downloaded from GEO (http://www.ncbi.nlm.nih.gov/gds/). The following search words were used: (circular RNA or circRNA) and (CRC or colorectal cancer). The filters were as follows: 1) at least five pairs of normal and cancer tissues, 2) no metastasis, 3) human solid tissues, and 4) arrays should have differentially expressed circRNAs between normal and cancer tissue after differential analysis. Two datasets were finally included in this study: GSE126095 and GSE142837. The R software was used to calibrate, standardize, and $\log_2$ transform the downloaded files.

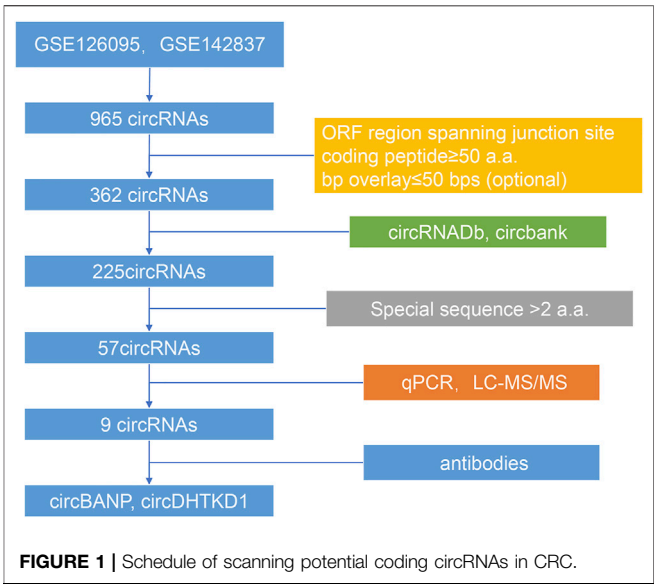


FIGURE 1 | Schedule of scanning potential coding circRNAs in CRC.

Open Reading Frame Prediction of circRNAs

The original sequences of circRNAs were downloaded from the circBase (http://www.circbase.org/). Then the open reading frames (ORFs) were predicted by using the getorf database (http://emboss.bioinformatics.nl/cgi-bin/emboss/getorf) with the sequence set as "circular." After infinite ORFs were excluded (Abe et al., 2015; Mo et al., 2019), the amino acid (a.a.) sequences that spanned the junction sites of the circRNAs were selected.

Identification and Evaluation of the Coding Potential in Differentially Expressed circRNAs

The batch effects between the datasets GSE126095 and GSE142837 were normalized by the sva package. Differential analysis was performed using the Limma package. The differentially expressed circRNAs met the standards: $|\log_2 FC| > 1$ and adjust p-value <0.05. The coding potential of differentially expressed circRNAs was predicted by the following rules: 1) predicted ORF regions spanned the junction sites of circRNAs, 2) the predicted coded peptides had more than 50 a.a., and 3)the bases overlay less than 50 bps, which could make the peptides easier to identify. The databases circRNADb and circbank were introduced to evaluate the overall coding potential of differentially expressed circRNAs. Finally, the circRNA-derived peptides should have a unique sequence with more than two a.a. The flowchart is shown in **Figure 1**.

Cell Culture

The CRC cells HCT116, SW480, SW620, Lovo, HT29, DLD-1, and CACO2 and normal control cells NCM460 were all cultured in Dulbecco's modified essential medium (DMEM) with 10% fetal bovine serum (FBS). Cells were obtained from the Chinese Academy of Sciences (Shanghai, China) and Chuan Qiu Biotechnology (Shanghai, China).

RNA Extraction and Real-Time Quantitative Reverse Transcription PCR Analysis

The TRIzol Reagent (Invitrogen, United States; 15596-026) (1 ml per well) was used to lyse cells, then the RNAs were extracted using RNAsimple Kit (TIANGEN, China; DP419). The concentration of RNAs was tested by DS-11 spectrophotometer (DeNovix, United States). RNA samples (2 μg) were mixed with PrimeScript™ RT reagent Kit with gDNA Eraser (TaKaRa, Tokyo, Japan; RR047A) and then applied to the ProFlex™ PCR system (ABI, United States) to obtain cDNAs. The cDNA samples were mixed with SYBR Premix Ex Taq II (TaKaRa, Tokyo, Japan; RR820A) and quantified by using ABI QuantStudio 5 (ABI, United States); the reaction was initiated at 95°C for 1 min, then at 95°C for 5 s, and 60°C 30 s for a total of 40 cycles. The primers are listed in **Supplementary Table S1**. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was set as the internal reference. The $2^{-\Delta\Delta Ct}$ method was used to process the data.

Western Blot Analysis

Radioimmunoprecipitation assay (RIPA) lysis buffer (Beyotime, China; P0013B) and protease and phosphatase inhibitor cocktail for general use (Beyotime, China; P1045) were used to obtain protein samples. Protein samples were denatured at 100°C for 10 min. Electrophoresis was started at 80 V to concentrate the protein and 120 V to separate the protein, then the transfer was conducted for 1 h under 300 mA. After blocking for 1 h with 5% non-fat milk (BD Biosciences, United States; 232100), the membranes were incubated overnight with primary antibodies BANP (ABclonal, China; A7595) and DHTKD1 (Santa Cruz Biotechnology, Inc., United States; sc-398620) at 4°C. After washing away the primary antibody with Tris-buffered saline Tween-20 (TBST), the secondary antibody (Abmart, China; M21003) was incubated for 1 h at room temperature. The bands were detected in automatic chemiluminescence and fluorescence analysis system (Tanon, China) by adding enhanced chemiluminescent reagent (NCM Biotech, China; P2200).

Liquid Chromatography-Tandem Mass Spectrometry Analysis

To identify the unique a.a. sequence of circRNAs encoded peptides, the total protein was subjected to liquid chromatography-tandem mass spectrometry (LC-MS/MS) sequencing and data analysis by Oebiotech Co., Ltd. (Shanghai, China). In brief, enzymolysis was performed using 0.02 μg/μl trypsin, and then the peptides were desalinated. The sample was analyzed using a Nano-HPLC liquid phase system EASY-NLC1200 and a Q-Exactive mass spectrometer. The acquired data were analyzed using ProteomeDiscover (V2.4, Thermo, United States) software.

Sanger Sequencing

The cDNA from DLD-1 were amplified by PCR using a primer specifically targeting hsa_circ_0000725 and hsa_circ_0008826 and 2×Hieff ®PCR Master Mix (with dye) (Yeasen, China; 10102ES03). To confirm the junction sites, Sanger sequencing was conducted by using the primers listed in **Supplementary Table S1**.

Construction and Transfection of siRNA

To silence the circRNAs, siRNAs were synthesized by RIBOBIO (Guangzhou, China). Lipofectamine™ RNAiMAX Transfection Reagent (Invitrogen, United States; 13778150) was used to transfect the siRNAs, and the efficiency was evaluated by qPCR.

EdU and CCK-8 Cell Proliferation Assays

Cells (2,000 per well) were seeded in 96-well plates. EdU (RIBOBIO, China; C10310-3) and CCK-8 (BBI, China; E606335) reagents were used to estimate cell proliferative ability following the manufacture instructions. The images of EdU were analyzed by ImageJ, and the GraphPad Prism 7.0 software (GraphPad Software Inc., La Jolla, CA, United States) was used to analyze CCK-8 results.

Colony Formation Assay

Cells (2,000 per well) were seeded in each well of a six-well plate. After 14 days of culture, the cells were fixed by 4% paraformaldehyde and stained by crystal violet. The images were analyzed by ImageJ.

Invasion Assay

To evaluate cell invasion ability, each upper chamber (Biofil, China; TCS003024) with a 1:7 concentration of Matrigel membrane (BD Biosciences, United States; 356234) was seeded with 50,000 cells with the serum-free culture medium. The lower chambers were put into complete culture medium. After 72 h, the cells in the chambers were fixed by 4% paraformaldehyde and stained by crystal violet. Then, the upper chamber cells were cleared. The images were analyzed by ImageJ.

Flow Cytometry of Cell Cycle and Apoptosis

In the cell cycle test, the cells were collected with cold phosphate-buffered saline (PBS) and fixed in 70% cold ethanol at −20°C overnight. After staining with propidine iodide (BD Biosciences, United States) for 30 min, the cells were loaded into a flow cytometer (BD Biosciences, United States), then 20,000–30,000 cells per tube were collected, and the results were analyzed by FlowJo (Version 10, United States). In the apoptosis test, the cells were suspended in Annexin-V binding buffer and then dyed using Annexin V combined fluorescein isothiocyanate (BD Biosciences, United States) and propidine iodide (BD Biosciences, United States) simultaneously in the dark for 15 min. The cells were loaded into a flow cytometer (BD Biosciences, United States), then 20,000–30,000 cells per tube were collected, and the results were analyzed by FlowJo (Version 10, United States).

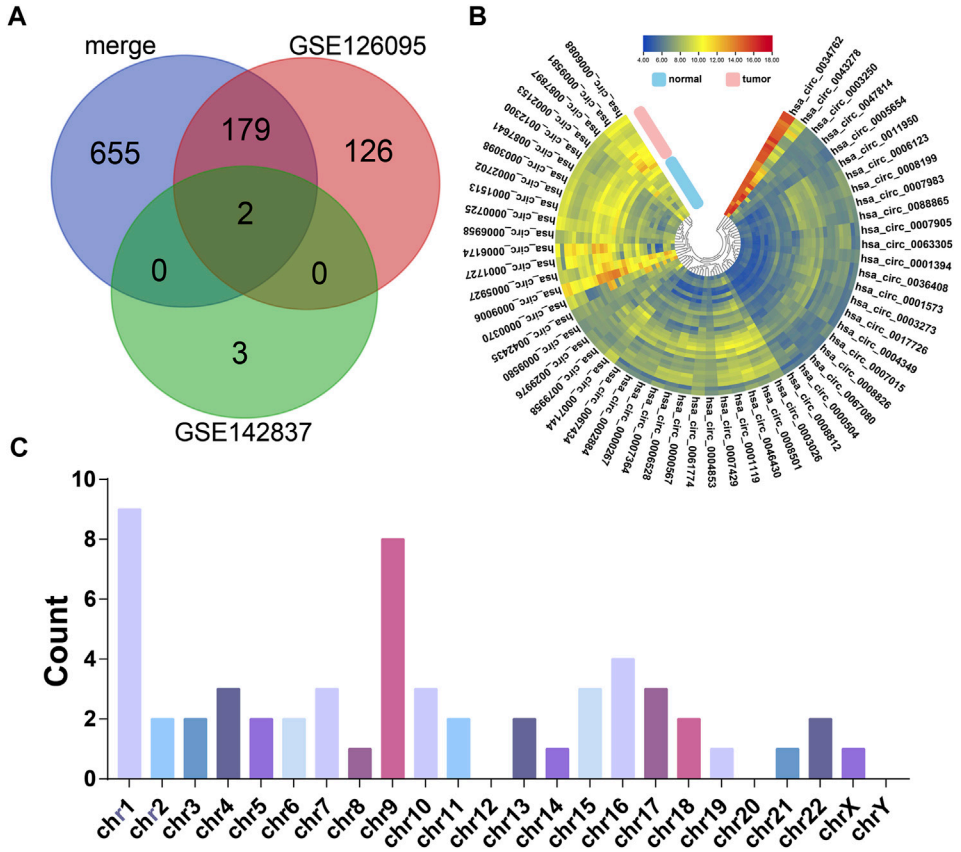


FIGURE 2 | The candidates of coding potential circRNAs. **(A)** Differential analysis of GSE126095 and GSE142837. **(B)** Heat map of the 57 circRNAs with coding potential in GSE126095 and GSE142837 after filtering by sequence and databases. **(C)** Chromosome location summary of 57 circRNAs with coding potential.

Statistical Analyses

The Student's *t*-test was used to evaluate the difference. A two-sided *p*-value <0.05 was considered statistically significant. Statistical analysis and graphs were constructed using GraphPad Prism 7.0 software (GraphPad Software Inc., La Jolla, CA, United States).

RESULTS

Evaluation of the Coding Potential of Differentially Expressed circRNAs in CRC

According to our strategy (**Figure 1**), we scanned the circRNAs with coding potential in CRC. First, to extend the candidate list, we compared the normal and CRC samples in the microarray datasets both together and separately. After deduplicating, the differential analysis found 965 differentially expressed circRNAs, including 836 circRNAs from the merge of GSE126095 and GSE142837 microarray datasets, 307 circRNAs from the GSE126095, and 5 circRNAs from the GSE142837 (**Figure 2A**). Then we analyzed the sequences of these circRNAs; the predicted ORF should span the junction site of their circRNAs, and the predicted peptides should be over 50 a.a. so that the peptides were not too small to be detected. The circular structure of the circRNAs enabled their bps to be used more than one time. However, this repeated use can make the analysis very difficult, so we mainly focused on circRNAs that were predicted to have less than 50 bps for repeated use. This resulted in the exclusion of 603 circRNAs and left 362 circRNAs. Next, the databases circRNADb and circbank were introduced to evaluate the potential coding function, and 225 circRNAs were selected. Special sequences in the coding peptides from circRNAs were the identification codes of circRNAs themselves; we selected circRNAs that had more than two unique a.a. to ensure the quality of the following LC-MS/MS assays because only one unique a.a. can have many deviations, and the more a.a. detected the more convincing the results will be. A total of 57 circRNAs were identified (**Table 1**), containing 46 upregulated circRNAs and 11 downregulated circRNAs **Figure 2B**. Besides, most of these circRNAs were derived from chromosomes 2 and 9, which may indicate their functions (**Figure 2C**).

Expression of Candidate circRNAs in CRC

Experimentally, we tested the basic expression levels of the 57 circRNAs in a panel of seven CRC cell lines and one normal colorectal cell line NCM460 (data available under request). From those circRNAs, hsa_circ_0000725, hsa_circ_0007429,

TABLE 1 | The basic information of 57 coding potential circRNAs.

ID	Log FC	Host gene	Best transcript	Position
hsa_circ_0006174	2.96	RAD23B	NM_002874	chr9:110064315-110068928
hsa_circ_0001727	2.05	ZKSCAN1	NM_003439	chr7:99621041-99621930
hsa_circ_0009581	1.95	RERE	NM_012102	chr1:8555122-8601377
hsa_circ_0046430	1.94	FOXK2	NM_004514	chr17:80521229-80545148
hsa_circ_0002702	1.91	RUSC2	NM_001135999	chr9:35546426-35548532
hsa_circ_0000567	1.81	SETD3	NM_199123	chr14:99924615-99932150
hsa_circ_0003098	1.72	BANP	NM_001173542	chr16:88037900-88071617
hsa_circ_0007429	1.71	RREB1	NM_001003699	chr6:7181356-7189555
hsa_circ_0001119	1.69	NDUFA10	NM_004544	chr2:240929490-240954277
hsa_circ_0006528	1.61	PRELID2	NM_138492	chr5:145197456-145205763
hsa_circ_0003026	1.56	USP10	NM_005153	chr16:84773914-84779279
hsa_circ_0063305	1.55	CSNK1E	NM_152221	chr22:38698865-38699253
hsa_circ_0087641	1.49	CDC14B	NM_033331	chr9:99284787-99327765
hsa_circ_0011950	1.48	HIVEP3	NM_024503	chr1:42041214-42050989
hsa_circ_0007144	1.45	PTPRM	NM_001105244	chr18:8076452-8143777
hsa_circ_0006123	1.43	KCNH1	NM_172362	chr1:211092981-211093411
hsa_circ_0001394	1.41	TBC1D14	NM_001113361	chr4:6925099-6925838
hsa_circ_0000725	1.40	BANP	NM_001173539	chr16:88008653-88052273
hsa_circ_0002884	1.33	PICALM	NM_007166	chr11:85692171-85707972
hsa_circ_0012300	1.31	PIK3R3	NM_003629	chr1:46521466-46546422
hsa_circ_0007983	1.31	GABPB1	NM_005254	chr15:50592985-50596330
hsa_circ_0067434	1.31	RYK	NM_001005861	chr3:133894452-133901915
hsa_circ_0007905	1.30	STX6	NM_005819	chr1:180953812-180962561
hsa_circ_0007364	1.28	PTP4A2	NM_080391	chr1:32381495-32385259
hsa_circ_0001573	1.25	RREB1	NM_001168344	chr6:7176887-7189555
hsa_circ_0007015	1.24	PTPN14	NM_005401	chr1:214625147-214638300
hsa_circ_0017726	1.23	DHTKD1	NM_018706	chr10:12123470-12139995
hsa_circ_0000267	1.22	FAM53B	NM_014661	chr10:126370175-126370948
hsa_circ_0004349	1.21	RAC1	NM_018890	chr7:6426842-6431672
hsa_circ_0061774	1.19	BACE2	NM_012105	chr21:42598192-42629253
hsa_circ_0006958	1.19	ACSF3	NM_001243279	chr16:89164998-89169167
hsa_circ_0008199	1.18	ATXN10	NM_013236	chr22:46085591-46136418
hsa_circ_0001513	1.15	LNPEP	NM_005575	chr5:96314841-96322374
hsa_circ_0008812	1.14	RAD23B	NM_002874	chr9:110062421-110074018
hsa_circ_0079958	1.13	HECW1	NM_015052	chr7:43508518-43540921
hsa_circ_0008501	1.12	RERE	NM_012102	chr1:8601272-8674745
hsa_circ_0088865	1.09	SPTAN1	NM_001130438	chr9:131365821-131367756
hsa_circ_0004853	1.07	CD97	NM_078481	chr19:14513408-14515374
hsa_circ_0067080	1.06	ITGB5	NM_002213	chr3:124560229-124592378
hsa_circ_0008826	1.06	DHTKD1	NM_018706	chr10:12136071-12162266
hsa_circ_0009006	1.05	LDB2	NM_001130834	chr4:16587544-16597498
hsa_circ_0006088	1.05	SPTAN1	NM_001130438	chr9:131362358-131367756
hsa_circ_0002153	1.03	MID2	NM_012216	chrX:107083899-107097934
hsa_circ_0000504	1.02	TUBGCP3	NM_006322	chr13:113170753-113181798
hsa_circ_0003273	1.02	TRIP12	NM_004238	chr2:230723487-230744844
hsa_circ_0036408	1.00	ETFA	NM_000126	chr15:76580186-76588078
hsa_circ_0005654	−1.04	PRDM5	NM_018699	chr4:121675707-121732604
hsa_circ_0087897	−1.15	C9orf5	NM_032012	chr9:111795586-111870850
hsa_circ_0047814	−1.23	ZNF532	NM_018181	chr18:56585502-56587865
hsa_circ_0005927	−1.25	VDAC3	NM_001135694	chr8:42259305-42260979
hsa_circ_0029976	−1.37	NBEA	NM_015678	chr13:35615069-35644989
hsa_circ_0003250	−1.38	MRRF	NM_138777	chr9:125042721-125054119
hsa_circ_0000370	−1.52	FLI1	NM_002017	chr11:128628009-128651918
hsa_circ_0034762	−1.57	MAPKBP1	NM_001128608	chr15:42103080-42105299
hsa_circ_0009580	−1.65	RERE	NM_012102	chr1:8555122-8568734
hsa_circ_0042435	−1.76	SPECC1	NM_001243439	chr17:20149238-20209395
hsa_circ_0043278	−5.12	TADA2A	NM_001488	chr17:35797838-35800763

hsa_circ_0008501, and hsa_circ_0067080 were expressed significantly higher in most CRC cells, whereas hsa_circ_0005654 was expressed significantly lower in most CRC cells, which is the same as the differential analysis. However, the expressions of hsa_circ_0006088, hsa_circ_0007364, hsa_circ_0008199, and hsa_circ_0008826 were slightly varied in different CRC cells (**Figure 3A** and **Supplementary Figure S1A**).

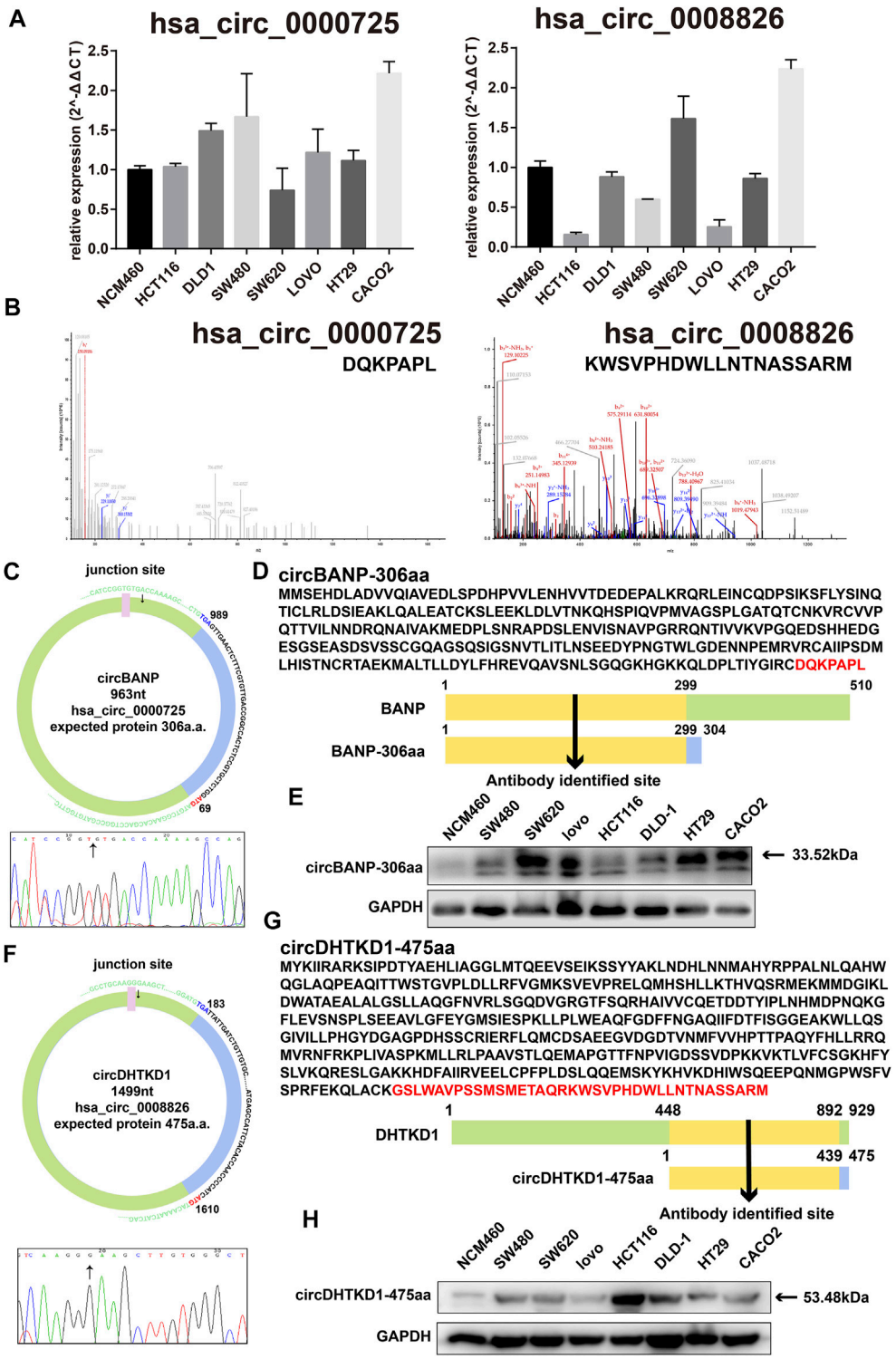


FIGURE 3 | Basic information of hsa_circ_0000725 and hsa_circ_0008826. **(A)** Expression level of hsa_circ_0000725 and hsa_circ_0008826 in CRC cells and normal cells. **(B)** LC-MS/MS found the unique sequence of hsa_circ_0000725 and hsa_circ_0008826. **(C)** The structure and junction site of hsa_circ_0000725. **(D)** Sequence alignment of hsa_circ_0000725 to its host gene. **(E)** The expression of circBANP-306aa in CRC cells. **(F)** The structure and junction site of hsa_circ_0008826. **(G)** Sequence alignment of hsa_circ_0008826 to its host gene. **(H)** The expression of circDHTKD1-475aa in CRC cells.

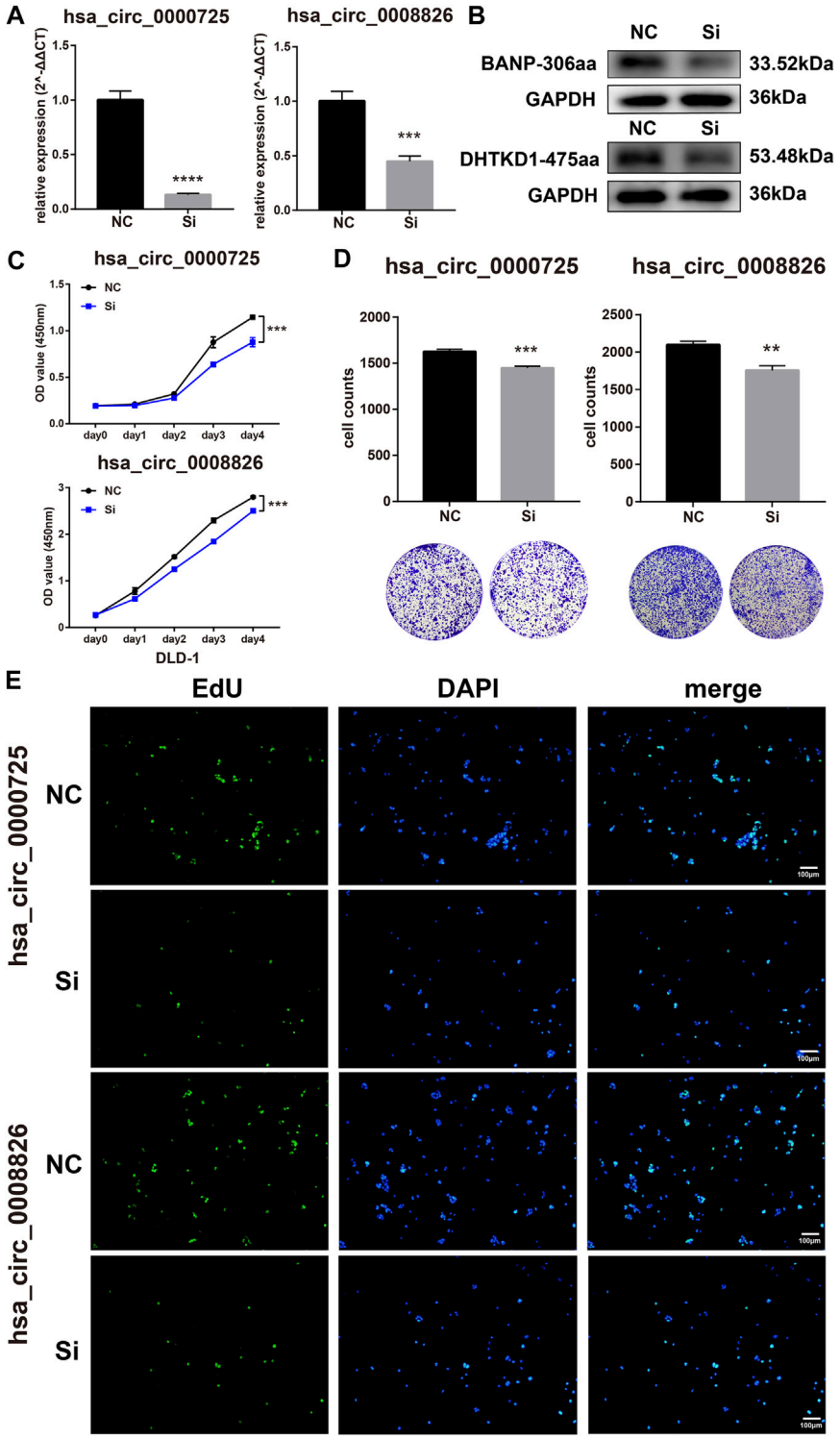


FIGURE 4 | Hsa_circ_0000725 and hsa_circ_0008826 influence the proliferation ability of CRC cells. **(A,B)** efficiency evaluation of siRNAs. **(C)** CCK8 assay showed the silencing of hsa_circ_0000725 and hsa_circ_0008826 decreased the proliferation ability. **(D)** Colony formation assay showed that the silencing of hsa_circ_0000725 and hsa_circ_0008826 weakened colony formation ability. **(E)** EdU assay showed the silencing of hsa_circ_0000725 and hsa_circ_0008826 reduced the proliferation rate of CRC cells.

Coding Potential Evaluation of Candidate circRNAs

Sequencing analysis found the unique a.a. sequence coded by the candidate circRNAs (**Supplementary Table S2**). We also performed LC-MS/MS analysis to identify the 57 circRNAs. Strikingly, the LC-MS/MS results found the special sequences of nine circRNAs, including hsa_circ_0000725, has_circ_0007429, hsa_circ_0008501, hsa_circ_0067080, hsa_circ_0005654, hsa_circ_0006088, hsa_circ_0007364, hsa_circ_0008199, and hsa_circ_0008826 (**Figure 3B** and **Supplementary Figure S1B**). Exploration of the circRNAs showed that the hsa_circ_0000725 contained a complete ORF with 963 nt and might encode a 306-a.a. peptide; hsa_circ_0008826 contained a 1,499 nt ORF with the potential of encoding a 475-a.a. length peptide (**Figures 3C, F**). The ORF structures of hsa_circ_0007429, hsa_circ_0008501, hsa_circ_0067080, hsa_circ_0005654, hsa_circ_0006088, hsa_circ_0007364, and hsa_circ_0008199 are shown in **Supplementary Figure S2**. The a.a. sequence encoded by hsa_circ_0000725 had a 299-a.a. area in common with its host gene BANP; therefore, a commercial BANP antibody that targeted this area (**Figure 3D**) was used to detect the circBANP-306aa, and the Western blot identified a band at about 33.52 kDa (**Figure 3E**). Also, the a.a. encoded by hsa_circ_0008826 had a 439-a.a. area in common with its host gene DHTKD1; a commercial DHTKD1 antibody that targets this area (**Figure 3G**) was used to detect circDHTKD1-475aa, and the Western blot identified a band at about 53.48 kDa area (**Figure 3H**).

Silencing of hsa_circ_0000725 and hsa_circ_0008826 can Inhibit CRC Cell Proliferation

The function of circRNAs-derived coding peptides may significantly relate to their derived circRNA. Therefore, we preliminarily explored the function of hsa_circ_0000725 and hsa_circ_0008826; their siRNAs were constructed and transfected into DLD-1. The siRNAs could efficiently silence the hsa_circ_0000725 and hsa_circ_0008826 in DLD-1 (**Figures 4A, B**). The CCK8 proliferation assay showed that the silencing of hsa_circ_0000725 and hsa_circ_0008826 could significantly inhibit the proliferation ability of DLD-1(**Figure 4C**). The colony formation assay indicated that the silencing of hsa_circ_0000725 and hsa_circ_0008826 could significantly decrease the colony-forming ability of the DLD-1 (**Figure 4D**). The EdU staining also apparently showed the number of cells in a proliferation state decreased significantly when the silenced hsa_circ_0000725 and hsa_circ_0008826 were compared to the normal control in DLD-1 (**Figure 4E**).

Silencing of hsa_circ_0000725 and hsa_circ_0008826 Could Also Influence the Cell Cycle, Apoptosis, and Invasion Ability of CRC Cells

Flow cytometry assays found that the silencing of hsa_circ_0000725 could significantly lead to the accumulation in the G0/G1 stage of CRC cell, while the silencing of hsa_circ_0008826 showed a subtle change of cell cycle stages (**Figures 5A, B**). Meanwhile, apoptosis assays found that the silencing of hsa_circ_0000725 could promote apoptosis; however, the silencing of hsa_circ_0008826 made no difference in apoptosis (**Figures 5C, D**). The invasion assay showed that both hsa_circ_0000725 and hsa_circ_0008826 promoted the invasion ability of DLD-1 (**Figures 5E, F**).

DISCUSSION

As CRC has a high incidence and high mortality, early diagnosis and intervention methods are essential in the successful management of the disease. Thus, seeking efficient biomarkers for diagnosis or specific targets for treatment is an urgent priority. Numerous studies have found novel biomarkers and potential therapeutic targets for CRC (Mai et al., 2020; Feng et al., 2021; Ning et al., 2021; Sun et al., 2021; Wu et al., 2021b). However, few can be introduced truly in clinical use. Therefore, finding stable and detectable molecules that can be used in clinical applications is important. CircRNAs possess a unique circle structure that is stable and easily detected. In the past, they were not believed to have coding ability. However, in recent years, a gradual number of studies have discovered that circRNAs can be cap-independent in translation (Pamudurti et al., 2017), and the translation process can be conducted through other approaches, for example, IRES-mediated and m6A-mediated (Yang et al., 2017).

Before our study, emerging evidence had shown the importance of the coding function of circRNAs. In 2017, Legnini and his colleagues first identified that circ-ZNF609, which specifically controls myoblast proliferation, can be translated into a protein in myogenesis (Legnini et al., 2017). Another study illustrated the translational landscape of the human heart by analyzing the translatomes of 80 human hearts and also discovered the coding potential of circRNAs (van Heesch et al., 2019). Our study has constructed a strategy that can be used to identify circRNAs with coding potential. The coding potential of circRNAs has promising application prospects. Recently, a study used the coding potential of circRNA to produce a circRNA vaccine that encoded the trimeric RBD of the SARS-CoV-2 spike protein with great efficacy, high design flexibility, and fast manufacturing speed compared to the traditional methods of vaccine production (Qu et al., 2021). Peptides encoded by circRNAs also have important functions in other cancers. In CRC, it was reported that peptide circPPP1R12A-73aa encoded by hsa_circ_0000423 can promote tumor pathogenesis and metastasis of CRC through the Hippo pathway (Zheng et al., 2019), and circFNDC3B-218aa encoded by circFNDC3B can regulate Snail to inhibit tumor progression and EMT in CRC (Pan et al., 2020). Zhang and his colleagues found several circRNAs that could encode peptides and function significantly in glioma (Yang et al., 2018; Zhang et al., 2018; Xia et al., 2019; Gao et al., 2021; Wu et al., 2021b). The FBXW7-185aa encoded by circ-FBXW7 can inhibit the proliferation and cell cycle acceleration of glioma (Yang et al., 2018), and the circular SHPRH as well as circular AKT3 could encode

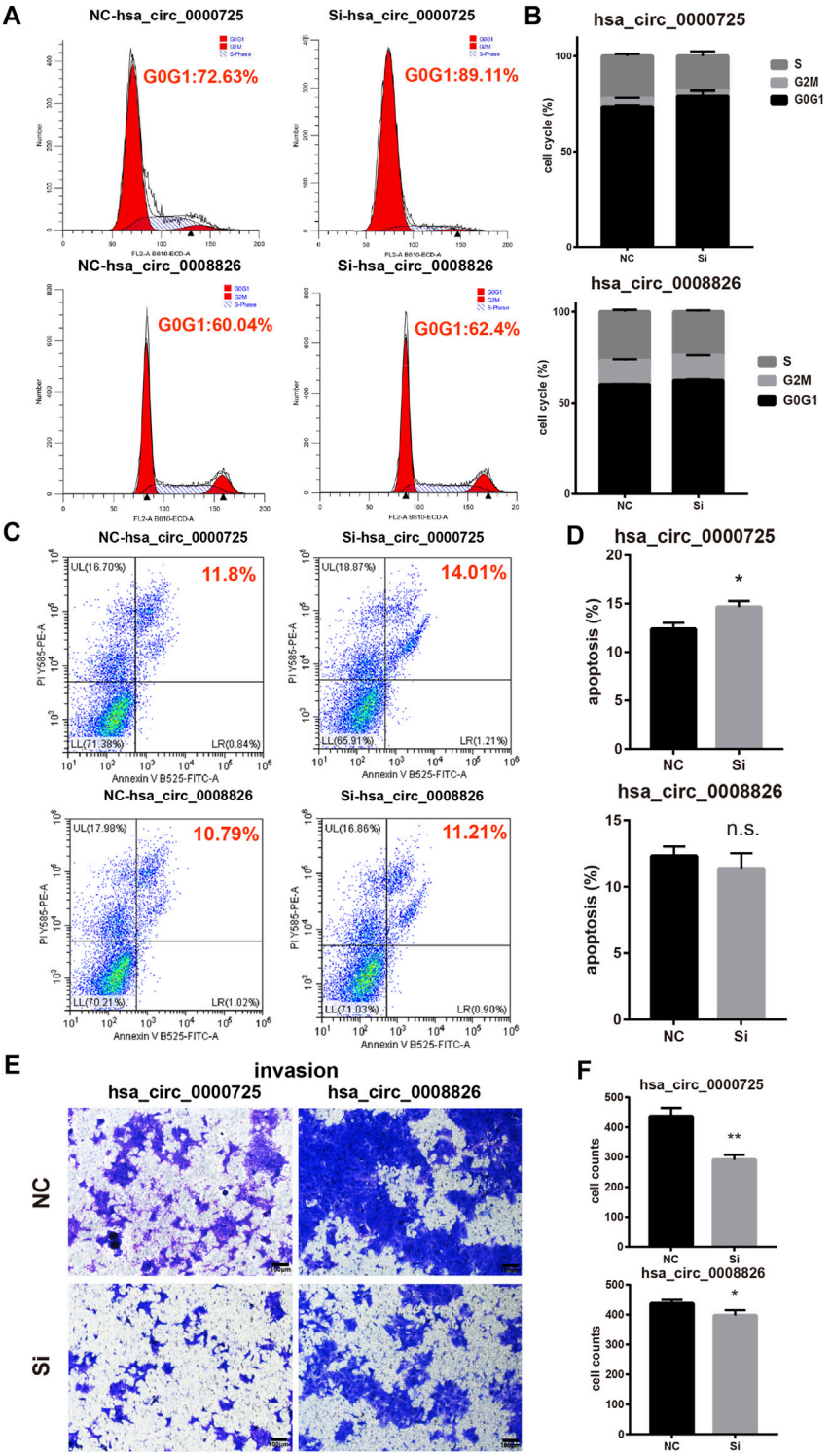


FIGURE 5 | Other function assays of hsa_circ_0000725 and hsa_circ_0008826. **(A,B)** cell cycle analysis found silencing of hsa_circ_0000725 could lead to the accumulation of cells in the G0/G1 stage, while the silencing of hsa_circ_0008826 shows no difference. **(C,D)** Apoptosis analysis found silencing of hsa_circ_0000725 could increase apoptosis, while the silencing of hsa_circ_0008826 shows no difference. **(E,F)** Invasion assay showed the silencing of hsa_circ_0000725 and hsa_circ_0008826 could reduce the invasion ability of CRC cells.

tumor-suppressed peptides in glioblastoma (Zhang et al., 2018; Xia et al., 2019). Meanwhile, a unique 14-a.a. peptide encoded by circ-E-Cad can promote the tumorigenicity of glioblastoma by activating the EGFR-STAT3 signal (Gao et al., 2021). Moreover, the phenomenon of coding circRNAs has also been observed in liver cancer (Liang et al., 2019; Zheng et al., 2019). The Wnt pathway is a classical tumor suppressor pathway; studies have found that circβ-catenin could influence a wide variety of Wnt pathway-related genes and encode a novel 370-a.a. peptide. This peptide can help to stabilize full-length β-catenin by antagonizing GSK3β-induced β-catenin phosphorylation and degradation (Liang et al., 2019). Furthermore, a novel protein, MAPK1-109aa, encoded by circMAPK1, can inhibit the progression of gastric cancer by suppressing the activation of the MAPK pathway (Jiang et al., 2021). Growing evidence has demonstrated the importance of circRNA-derived peptides in tumorigenesis and tumor development, which may be a potential biomarker or therapeutic target for cancer treatment in the future. Our strategy may be a convenient approach for searching the coding potential of circRNAs not only in CRC but also in other cancers.

In this study, we constructed a strategy to efficiently search the coding potential circRNAs in CRC. In total, the candidate list lasts nine circRNAs, including hsa_circ_0000725, has_circ_0007429, hsa_circ_0008501, hsa_circ_0067080, hsa_circ_0005654, hsa_circ_0006088, hsa_circ_0007364, hsa_circ_0008199, and hsa_circ_0008826 (**Figures 3A, B** and **Supplementary Figures S1 and S2**). Owing to the unique sequence of the novel peptides, we focused on those that had sequences identifiable by antibodies in their linear mRNA; two circRNAs were selected. Hsa_circ_0000725 and hsa_circ_0008826 were predicted to encode a 306-a.a. peptide and 475-a.a. peptide, respectively. Western blot analysis showed that hsa_circ_0000725- and hsa_circ_0008826-derived peptides were expressed in multiple CRC cells, which is almost in accordance with the qPCR results. Then we preliminarily explore the function of hsa_circ_0000725 and hsa_circ_0008826; both of them could significantly promote the proliferation and colony formation as well as invasion ability of CRC. Besides, the silencing of hsa_circ_0000725 could lead to the accumulation of cells in the G0/G1 stage as well as the increase of apoptosis rate. In a word, these peptide-coding hsa_circ_0000725 and hsa_circ_0008826 could promote the proliferation as well as invasion ability of DLD-1; however, whether their coded peptides make the most contribution to this phenomenon still needs further validation.

Our findings helped to construct a new schedule to scan potential coding circRNAs, which may be very practical in functional studies on circRNAs. However, our study still had limitations and shortcomings. The function of the circRNA itself could not be excluded, including the competing endogenous RNA (ceRNA) mechanism or RNA-binding protein mechanism. Whether the circRNA-derived peptides functioned in a pivotal role still needs further validation; specific antibodies are needed to identify these peptides, and overexpression vectors are needed to clarify the findings. Therefore, with the limitations considered, our preliminary exploration indicated that hsa_circ_0000725 and hsa_circ_0008826 possessed the ability of coding peptides and could promote the proliferation and invasion of CRC cells, but only hsa_circ_0000725 could influence the cell cycle and apoptosis rate of colorectal cells. Our strategy for scanning potential coding circRNAs was found to be effective. And these potential coding circRNAs may be good choices when seeking CRC therapeutic targets or diagnosis biomarkers.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found here: The data that support the findings of this study are available from the corresponding author upon reasonable request.

AUTHOR CONTRIBUTIONS

ZC, ZQ, and DH contributed equally to this work. ZC, ZQ, and DH performed the experiments and wrote the manuscript. JL, EX, BL, SC, DS, and YC contributed to experiments and data analysis. YZ and QS designed the study. All authors approved the final version of the manuscript.

FUNDING

This work was supported by the National Key R&D Program of China (No. 2019YFC1315800/2019YFC1315802), National Natural Science Foundation of China (Nos. 81902394, 82002515, 81502000), Shanghai Sailing Program (No. 20YF1407200), China Postdoctoral Science Foundation (No. 2020M681177), and Shanghai Municipal Commission of Science and Technology (No. 19140901902, 18DZ1930302, 19140901902).

ACKNOWLEDGMENTS

We sincerely appreciate the guidance of Jianmin Xu and the invaluable assistance of Dexiang Zhu, Meiling Ji and Qi Lin during the current study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.815895/full#supplementary-material

REFERENCES

Abe, N., Matsumoto, K., Nishihara, M., Nakano, Y., Shibata, A., Maruyama, H., et al. (2015). Rolling Circle Translation of Circular RNA in Living Human Cells. *Sci. Rep.* 5, 16435. doi:10.1038/srep16435

Artemaki, P. I., Scorilas, A., and Kontos, C. K. (2020). Circular RNAs: A New Piece in the Colorectal Cancer Puzzle. *Cancers (Basel)* 12 (9). doi:10.3390/cancers12092464

Feng, W., Huang, W., Chen, J., Qiao, C., Liu, D., Ji, X., et al. (2021). CXCL12-mediated HOXB5 Overexpression Facilitates Colorectal Cancer Metastasis through Transactivating CXCR4 and ITGB3. *Theranostics* 11 (6), 2612–2633. doi:10.7150/thno.52199

Gao, X., Xia, X., Li, F., Zhang, M., Zhou, H., Wu, X., et al. (2021). Circular RNA-Encoded Oncogenic E-Cadherin Variant Promotes Glioblastoma Tumorigenicity through Activation of EGFR-STAT3 Signalling. *Nat. Cel. Biol.* 23 (3), 278–291. doi:10.1038/s41556-021-00639-4

Jiang, T., Xia, Y., Lv, J., Li, B., Li, Y., Wang, S., et al. (2021). A Novel Protein Encoded by circMAPK1 Inhibits Progression of Gastric Cancer by Suppressing Activation of MAPK Signaling. *Mol. Cancer* 20 (1), 66. doi:10.1186/s12943-021-01358-y

Legnini, I., Di Timoteo, G., Rossi, F., Morlando, M., Briganti, F., Sthandier, O., et al. (2017). Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. *Mol. Cel.* 66 (1), 22–37. doi:10.1016/j.molcel.2017.02.017

Li, Y., Chen, S., Zhao, J., Qin, Li., Chen, S., Guo, T., et al. (2021). HNRNPL Circularizes ARHGAP35 to Produce an Oncogenic Protein. *Adv. Sci.* 8 (13), 2001701. doi:10.1002/advs.202001701

Liang, W.-C., Wong, C.-W., Liang, P.-P., Shi, M., Cao, Y., Rao, S.-T., et al. (2019). Translation of the Circular RNA Circβ-Catenin Promotes Liver Cancer Cell Growth through Activation of the Wnt Pathway. *Genome Biol.* 20 (1), 84. doi:10.1186/s13059-019-1685-4

Long, F., Lin, Z., Li, L., Ma, M., Lu, Z., Jing, L., et al. (2021). Comprehensive Landscape and Future Perspectives of Circular RNAs in Colorectal Cancer. *Mol. Cancer* 20 (1), 26. doi:10.1186/s12943-021-01318-6

Mai, D., Zheng, Y., Guo, H., Ding, P., Bai, R., Li, M., et al. (2020). Serum piRNA-54265 Is a New Biomarker for Early Detection and Clinical Surveillance of Human Colorectal Cancer. *Theranostics* 10 (19), 8468–8478. doi:10.7150/thno.46241

Mo, D., Li, X., Raabe, C. A., Cui, D., Vollmar, J.-F., Rozhdestvensky, T. S., et al. (2019). A Universal Approach to Investigate circRNA Protein Coding Function. *Sci. Rep.* 9 (1), 11684. doi:10.1038/s41598-019-48224-y

Ning, T., Li, J., He, Y., Zhang, H., Wang, X., Deng, T., et al. (2021). Exosomal miR-208b Related with Oxaliplatin Resistance Promotes Tregs Expansion in Colorectal Cancer. *Mol. Ther.* 29 (9), 2723–2736. doi:10.1016/j.ymthe.2021.04.028

Pamudurti, N. R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., et al. (2017). Translation of CircRNAs. *Mol. Cel.* 66 (1), 9–21. doi:10.1016/j.molcel.2017.02.021

Pan, Z., Cai, J., Lin, J., Zhou, H., Peng, J., Liang, J., et al. (2020). A Novel Protein Encoded by circFNDC3B Inhibits Tumor Progression and EMT through Regulating Snail in colon Cancer. *Mol. Cancer* 19 (1), 71. doi:10.1186/s12943-020-01179-5

Qu, L., Yi, Z., Shen, Y., Xu, Y., Wu, Z., Tang, H., et al. (2021). *Circular RNA Vaccines against SARS-CoV-2 and Emerging Variants*. Biorxiv. doi:10.1101/2021.03.16.435594

Siegel, R. L., Miller, K. D., Goding Sauer, A., Fedewa, S. A., Butterly, L. F., Anderson, J. C., et al. (2020). Colorectal Cancer Statistics, 2020. *CA A. Cancer J. Clin.* 70 (3), 145–164. doi:10.3322/caac.21601

Sun, L., Yang, X., Huang, X., Yao, Y., Wei, X., Yang, S., et al. (2021). 2-Hydroxylation of Fatty Acids Represses Colorectal Tumorigenesis and Metastasis via the YAP Transcriptional Axis. *Cancer Res.* 81 (2), 289–302. doi:10.1158/0008-5472.can-20-1517

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 Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

van Heesch, S., Witte, F., Schneider-Lunitz, V., Schulz, J. F., Adami, E., Faber, A. B., et al. (2019). The Translational Landscape of the Human Heart. *Cell* 178 (1), 242–260. doi:10.1016/j.cell.2019.05.010

Wu, Q. N., Luo, X., Liu, J., Lu, Y.-X., Wang, Y., Qi, J., et al. (2021a). MYC-Activated LncRNA MNX1-AS1 Promotes the Progression of Colorectal Cancer by Stabilizing YB1. *Cancer Res.* 81 (10), 2636–2650. doi:10.1158/0008-5472.can-20-3747

Wu, X., Xiao, S., Zhang, M., Yang, L., Zhong, J., Li, B., et al. (2021b). A Novel Protein Encoded by Circular SMO RNA Is Essential for Hedgehog Signaling Activation and Glioblastoma Tumorigenicity. *Genome Biol.* 22 (1), 33. doi:10.1186/s13059-020-02250-6

Xia, X., Li, X., Li, F., Wu, X., Zhang, M., Zhou, H., et al. (2019). A Novel Tumor Suppressor Protein Encoded by Circular AKT3 RNA Inhibits Glioblastoma Tumorigenicity by Competing with Active Phosphoinositide-dependent Kinase-1. *Mol. Cancer* 18 (1), 131. doi:10.1186/s12943-019-1056-5

Xu, B., Yang, N., Liu, Y., Kong, P., Han, M., and Li, B. (2020). Circ_cse1l Inhibits Colorectal Cancer Proliferation by Binding to eIF4A3. *Med. Sci. Monit.* 26, e923876. doi:10.12659/MSM.923876

Yang, H., Li, X., Meng, Q., Sun, H., Wu, S., Hu, W., et al. (2020). CircPTK2 (Hsa_circ_0005273) as a Novel Therapeutic Target for Metastatic Colorectal Cancer. *Mol. Cancer* 19 (1), 13. doi:10.1186/s12943-020-1139-3

Yang, Y., Fan, X., Mao, M., Song, X., Wu, P., Zhang, Y., et al. (2017). Extensive Translation of Circular RNAs Driven by N6-Methyladenosine. *Cell Res.* 27 (5), 626–641. doi:10.1038/cr.2017.31

Yang, Y., Gao, X., Zhang, M., Yan, S., Sun, C., Xiao, F., et al. (2018). Novel Role of FBXW7 Circular RNA in Repressing Glioma Tumorigenesis. *J. Natl. Cancer Inst.* 110 (3), 304–315. doi:10.1093/jnci/djx166

Zeng, K., Chen, X., Xu, M., Liu, X., Hu, X., Xu, T., et al. (2018). CircHIPK3 Promotes Colorectal Cancer Growth and Metastasis by Sponging miR-7. *Cell Death Dis.* 9 (4), 417. doi:10.1038/s41419-018-0454-8

Zhang, M., Huang, N., Yang, X., Luo, J., Yan, S., Xiao, F., et al. (2018). A Novel Protein Encoded by the Circular Form of the SHPRH Gene Suppresses Glioma Tumorigenesis. *Oncogene* 37 (13), 1805–1814. doi:10.1038/s41388-017-0019-9

Zheng, X., Chen, L., Zhou, Y., Wang, Q., Zheng, Z., Xu, B., et al. (2019). A Novel Protein Encoded by a Circular RNA circPPP1R12A Promotes Tumor Pathogenesis and Metastasis of colon Cancer via Hippo-YAP Signaling. *Mol. Cancer* 18 (1), 47. doi:10.1186/s12943-019-1010-6

Zhou, C., Liu, H.-s., Wang, F.-W., Hu, T., Liang, Z.-X., Lan, N., et al. (2020). circCAMSAP1 Promotes Tumor Growth in Colorectal Cancer via the miR-328-5p/E2F1 Axis. *Mol. Ther.* 28 (3), 914–928. doi:10.1016/j.ymthe.2019.12.008

Conflicts 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: 23 March 2022
doi: 10.3389/fcell.2022.859958

A Potential Diagnostic and Prognostic Biomarker TMEM176B and Its Relationship With Immune Infiltration in Skin Cutaneous Melanoma

Linlan Jiang[1†], Yanyin Yang[2†], Fangming Liu[3], Mingyue Ma[2], Jie Gao[3], Lu Sun[3], Yuwen Chen[4], Zan Shen[1] and Duojiao Wu[4*]*

[1]Department of Oncology, Affiliated Sixth People's Hospital, Shanghai Jiaotong University, Shanghai, China, [2]Department of Endocrinology and Metabolism, Zhongshan Hospital, Key Laboratory of Metabolism and Molecular Medicine, the Ministry of Education, Fudan University, Shanghai, China, [3]Institute of Clinical Science, Zhongshan Hospital, Fudan University, Shanghai, China, [4]Jinshan Hospital Center for Tumor Diagnosis and Therapy, Jinshan Hospital, Fudan University, Shanghai, China

OPEN ACCESS

Edited by:
Geng Chen,
GeneCast Biotechnology Co., Ltd., China

Reviewed by:
Ting Li,
National Center for Toxicological Research (FDA), United States
Liqun He,
Uppsala University, Sweden

***Correspondence:**
Duojiao Wu
wu.duojiao@zs-hospital.sh.cn
Zan Shen
sshenzzan@vip.sina.com

[†]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 January 2022
Accepted: 04 March 2022
Published: 23 March 2022

Citation:
Jiang L, Yang Y, Liu F, Ma M, Gao J, Sun L, Chen Y, Shen Z and Wu D (2022) A Potential Diagnostic and Prognostic Biomarker TMEM176B and Its Relationship With Immune Infiltration in Skin Cutaneous Melanoma. Front. Cell Dev. Biol. 10:859958. doi: 10.3389/fcell.2022.859958

Background: Melanoma is a highly malignant and aggressive tumor. The search for new and effective biomarkers facilitates early diagnosis and treatment, ultimately improving the prognosis of melanoma patients. Although the transmembrane protein TMEM176B has been linked to a number of cancers, its role in cancer immunity remains unknown.

Methods: Expression levels of TMEM176B in normal tissues and several cancers, including Skin Cutaneous Melanoma (SKCM), were collected from TCGA and GTEx. We used Receiver operating characteristic and Kaplan–Meier survival curves and performed regression analysis to elucidate the link between TMEM176B and clinicopathological features of SKCM in order to determine the prognostic significance of TMEM176B in SKCM. We then used the GEPIA and STRING websites to search for proteins and associated top genes that may interact with TMEM176B and enriched them for analysis. The link between TMEM176B and immune cells infiltration was then investigated using TIMER, CIBERSORT algorithm and GSVA package of R (v3.6.3). Finally, animal tests were conducted to confirm the expression of *Tmem176b* and its influence on T-cell immune infiltration.

Results: *TMEM176B* expression was considerably elevated in SKCM compared to normal tissues. Particularly, *TMEM176B* expression was also linked to pathological stage, tumor ulceration and radiation therapy. Patients with elevated *TMEM176B* expression had a better prognosis, according to the survival analysis. The majority of tumor infiltrating lymphocytes (TILs) especially T cells in SKCM was positively linked with *TMEM176B* expression. Our animal experiments also verified that the T-cell infiltration was significantly inhibited in local melanoma tissue of *Tmem176b* knockout mice. At the same time deleting *Tmem176b* accelerated tumor progress and impaired T cells effector function.

Conclusion: Upregulated expression of TMEM176B in SKCM is associated with a better prognosis and it has the potential to serve as a diagnostic and prognostic marker for the

disease. It may serve as a target for SKCM immunotherapy by regulating $CD8^+$ T cells although it requires more evidence.

Keywords: TMEM176B, cancer, $CD8^+$ T cell, tumor immunology, SKCM

INTRODUCTION

Melanoma, which is aggressive, is a malignant tumor that develops from melanocytes (Schadendorf et al., 2015; Shain and Bastian, 2016; Lin and Fisher, 2017). Cutaneous melanoma is most common in Western countries, but the global incidence has gradually increased in recent years (Hollestein et al., 2012). Ultraviolet light (UV) radiation is an important pathogenic factor in the development of melanoma (Schadendorf et al., 2015; Michielin et al., 2019). Early diagnosis is made easier by general clinical and pathological aspects due to the specific color characteristics of melanoma and the fact that it occurs on the skin surface, although diagnostic hurdles still exist (Schadendorf et al., 2015). Patients with advanced metastases have a poor prognosis, despite the fact that early-stage tumors can be surgically excised with a high cure rate (Lin and Fisher, 2017). As a result, new diagnostic and prognostic biomarkers are needed to increase the survival rate of melanoma patients.

To achieve complete control of tumors, the antitumor effector cells in the immune system must coordinate with each other to overcome tumor immune evasion (Huntington et al., 2020). T cells are one of the most important anti-tumor effector cells because they may directly attack cancer cells. Immune checkpoint therapy targeting T cells, for instance, cytotoxic T lymphocyte-associated antigen 4 (CTLA4) and programmed cell death protein 1 (PD1) has shown promising results in patients with melanoma in recent years, but only some tumor types have benefited, the degree of tumor infiltration and immune cell activation, particularly $CD8^+$ T cells, are both factors that influence immune checkpoint therapy efficacy (Kim et al., 2020). Exploring the immune phenotype and interactions in the tumor microenvironment is therefore critical for discovering new immunotherapeutic targets in melanoma.

Some transmembrane proteins have been found to be overexpressed in tumors and to be linked to prognosis, suggesting that they could be useful targets for immunotherapy (Choi et al., 2007; Tomimaru et al., 2015). TMEM176B (Transmembrane protein 176B), which was formerly known as TORID (Tolerance-related and Inducible Transcript) (Louvet et al., 2005), is a gene that is, highly overexpressed in tolerance allografts (Viklicky et al., 2013). Its human homolog was originally found to be expressed in a subset of lung fibroblasts, the mouse homolog Clast1 was upregulated after B cells were activated by CD40 ligand (Lurton et al., 1999). TMEM176B, like its paralog TMEM176A (Transmembrane protein 176A), has four transmembrane domains and belongs to the MS4A family (Eon Kuek et al., 2016). It is mostly expressed in the lungs, kidneys, lymph nodes, and spleen, among different immune cell subtypes, it has been observed that the high expression of *Tmem176b* and *Tmem176a* in BMDCs (Bone Marrow-Derived Dendritic Cells) and cDCs (Classic Dendritic Cells), and is associated with the immature state of DCs (Dendritic Cells) (Condamine et al., 2010).

The expression level of TMEM176B was higher in gastric cancer tissues than in normal tissues, and according to survival analysis, higher expression levels were related to poorer prognosis (Sun et al., 2018). The expression of TMEM176B in lymphoma and its control tissues also differed significantly (Cuajungco et al., 2012; Ma and Li, 2021). The expression of TMEM176B is up-regulated in the tumor blood vessels of human renal cell carcinoma specimens, suggesting that it is implicated in tumor angiogenesis and could be a target of anti-angiogenesis therapy for cancer patients (Otsubo et al., 2014). However, the role of TMEM176B in cancers has yet to be determined, or we do not yet have a complete grasp of its function in malignancies.

In addition, the correlation between TMEM176B and tumor immunity and its role in melanoma is less reported. *Tmem176b* and its homolog *Tmem176a* has been found to be highly expressed in $CD4^+$ Th17 cells (Drujont et al., 2016), this study also found that these two molecules play the same ion channel function and are co-localized near the Golgi apparatus (Drujont et al., 2016). The expression of some costimulatory molecules in T cells can be inhibited by transfection of *Tmem176b* into immature dendritic cells (Louvet et al., 2005). It revealed that suppressing inflammasomes by knocking down *Tmem176b* in mice or utilizing TMEM176B inhibitors can improve the anti-tumor effect of $CD8^+$ T and the efficacy of anti-CTLA-4 and anti-PD1 therapy (Segovia et al., 2019). However, cancer is a very heterogeneous disease and melanoma has its unique immune microenvironment. It is unknown the role of TMEM176B in regulating $CD8^+$ T cells biology and the course of melanoma.

We explored the expression of TMEM176B in SKCM and its relationship with tumor patients' prognosis and immune infiltration. Furthermore, we found the regulatory effect of TMEM176B on tumor-infiltrating $CD8^+$ T cells. Therefore, TMEM176B is a potential diagnostic or prognostic biomarker for melanoma.

METHODS

Gene Expression Analysis

We utilized R package "ggplot2" to analysis the differential expression of *TMEM176B* in various malignancies or cancer subtypes. Extraction of SKCM in TCGA and corresponding normal tissue data in GTEx has been analyzed for differential expression of *TMEM176B* in cutaneous melanoma. Click "repository" in TCGA database and make the following settings: under File directory (select "Gene Expression Quantification" under Data Category module, select "HTseq-FPKM" under Workflow Type module); under Cases (under Primary Site module, select "skin," and under Project, select

"TCGA-SKCM"); under Cases (select "skin" under the Primary Site module and "TCGA-SKCM" under Project) to get the clinical information of SKCM.

Survival Prognosis Analysis

R package "survminer" and "survival" were utilized to analyze the prognostic value of the expression level of *TMEM176B* in different tumors. We collected the data from the TCGA database (Liu et al., 2018) and compared the overall survival (OS) of cancer patients separated by the median expression level of *TMEM176B*. The "pROC" package and the "ggplot2" package were used to extract and analyze the SKCM data of TCGA and the corresponding normal tissue data in GTEx to obtain the receiver operating characteristic (ROC) curves.

Cox Regression

We performed Cox regression analysis to identify the association between SKCM disease characteristics and the expression level of *TMEM176B*.

Survival Analysis in PrognoScan

In PrognoScan, the association between TMEM176B expression and prognosis of cancer patients was analyzed, including distant metastasis-free survival (DMFS), disease-free survival (DFS), relapse-free survival (RFS), OS, and distant recurrence-free survival (DRFS).

Relationship Between TMEM176B Expression and Immune Cell Infiltration

TIMER is a fully functional public platform for analyzing of immune infiltration in tumors. Correlations between *TMEM176B* expression and immune infiltration levels in more than 30 different cancer types in TCGA were obtained in the gene module of TIMER. We also investigated the connection between tumor purity and TMEM176B expression. And we utilized R package "GSVA" to obtain the link among TMEM176B and other immunocytes (Bindea et al., 2013). Next "CIBERSORT algorithm" was used to explore the immune infiltration between high- and low-expression with *TMEM176B* to obtain enriched immunocytes between the two groups.

TMEM176B-Related Gene Enrichment Analysis and Single Gene Co-Expression Analysis

We acquired the top 100 genes associated with *TMEM176B* in SKCM cases using the "Similar Gene Detection" module of GEPIA2 and performed correlation analysis on the top 5 genes using "Correlation Analysis" in GEPIA2. We then obtained protein networks that may interact with TMEM176B in the STRING database. In addition, we subjected the 100 genes obtained in GEPIA2 to GO and KEGG pathway analysis using the "clusterProfiler" and "ggplot2" R packages.

R package "ggplot2" was used to determine the link between the expression level of TMEM176B in SKCM and other genes, including CD274, CTLA4, GZMA, GAMB, IFNG, PRF1, and so on.

Animals

Aged 6–8 weeks *Tmem176b*$^{-/-}$ mice and C57BL/6J mice were purchased from GemPharmatech Co. Ltd. and Shanghai Jie Si Jie Laboratory Animal Co. Ltd., respectively.

Cell Culture

B16-OVA melanoma cells were cultured in DMEM medium, which contains 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin.

Tumor Model

We injected 1×10^6 B16-OVA cells into the right flank of WT or *Tmem176b*$^{-/-}$ mice subcutaneously and used vernier calipers to measure tumor volume every 1–2 days. On the 14th day, the mice were euthanized and their tumor tissues, spleen, lymph nodes, and peripheral blood were taken for further research.

Flow Cytometry

Cells were stained with CD3 (BioLegend, clone:17A2), CD4 (BioLegend, clone: GK1.5), CD8 (BioLegend, clone:53-6.7), GZMB (Biolegend, clone: QA16A02), IFN-γ (Biolegend, clone: XMG1.2), Ki-67 (BioLegend, clone: 16A8), PD-1(Biolegend, clone:29F.1A12), PD-L1 (BioLegend, clone:10F.9G2), CD44 (Biolegend, clone: IM7), CD62L (BioLegend, clone: MEL-14), CD25(BioLegend, clone: 3C7), CD69(BioLegend, H1.2F3), BCL2 (Cell Signaling Technology, clone: 124), PRF(BioLegend, clone: S16009A). For intracellular staining, cells were given PMA/ionomycin (BioLegend, Cat. No. 423303) re-stimulation for 4–5 h, fixed, permeabilized, and then stained. FACS analysis was performed on a BD FACS Aria III flow cytometer and analyzed by FlowJo V.10 software. BD FACS Aria III flow cytometer was used to perform FACS analysis and data analysis was performed in FlowJo V.10 software.

Statistical Analysis

Graphpad Prism (v8) and R language version (3.6.3) were utilized for designing figures and statistical analysis. Wilcoxon rank sum test and Student's t tests were used to compare the expression of TMEM176B in different groups. When p values were less than 0.05, differences were considered significant.

RESULTS

Clinical Characteristics of Patients

We extracted clinical information from the TCGA, including TNM stage, pathologic stage, age, race, weight, gender and whether or not they had received radiation. **Supplementary Table S1** summarizes all the information. The table manifested that the expression of *TMEM176B* is closely related to the TNM stage of SKCM, the pathological stage and radiation therapy.

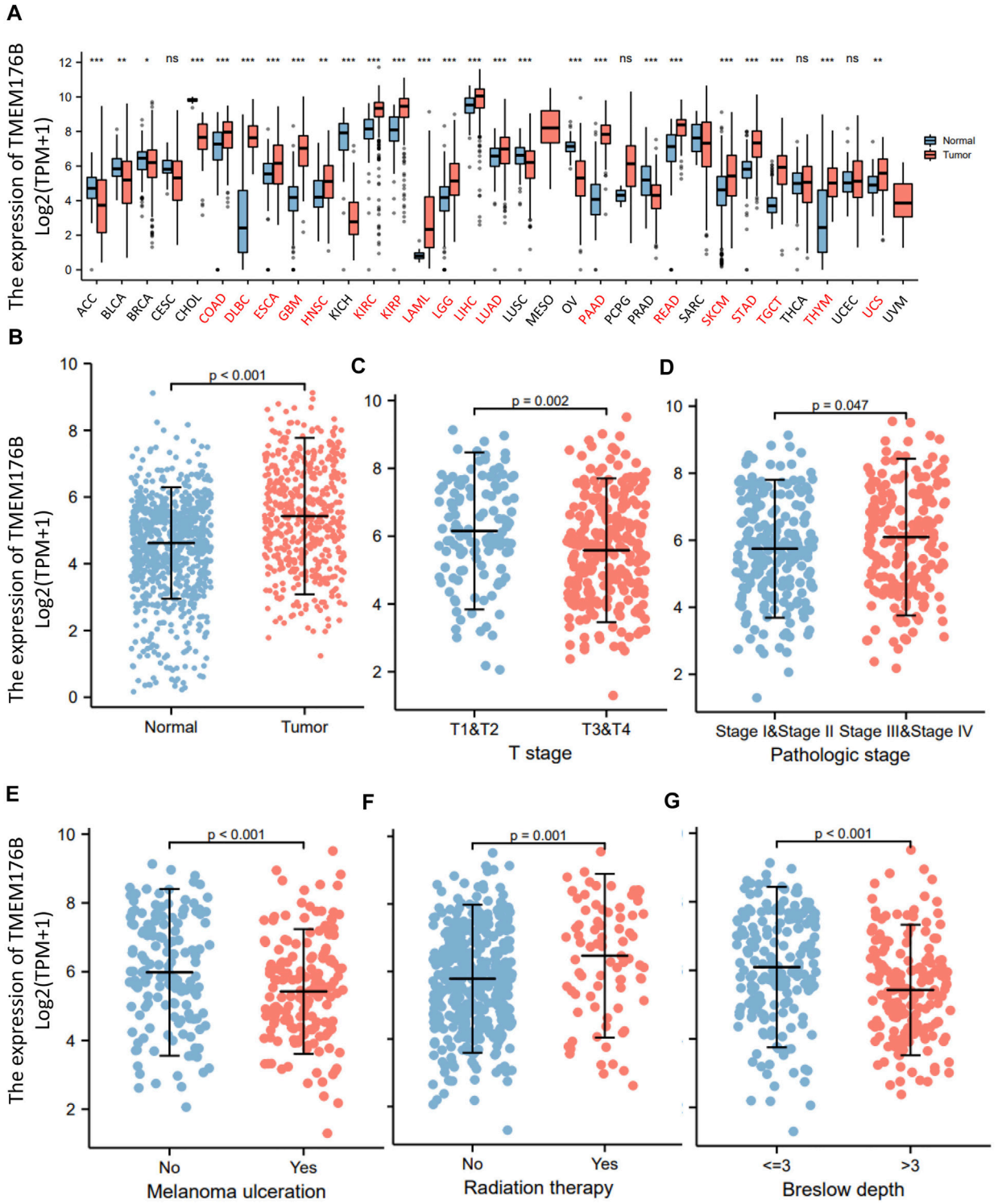
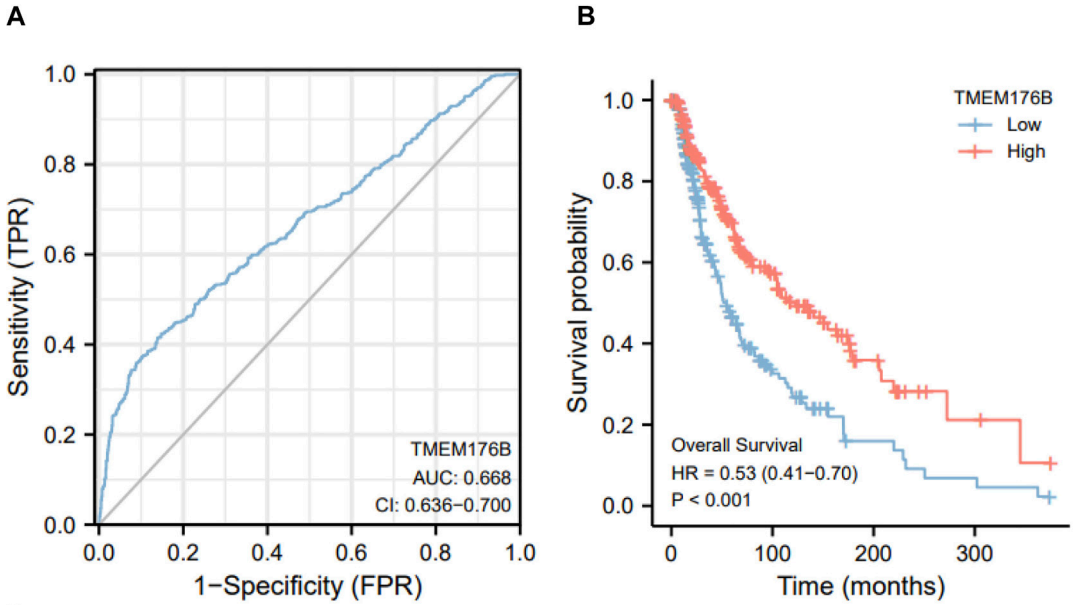


FIGURE 1 | The expression of *TMEM176B* in malignancies. **(A)** The expression status of the *TMEM176B* across various cancer types. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. **(B)** The expression level of *TMEM176B* in SKCM was significantly increased compared to the normal tissue ($p < 0.001$). **(C–G)** Association between *TMEM176B* and several clinical features in SKCM.



C

Characteristics	Total(N)	HR(95% CI)	P value
Gender	456		
Female	172	Reference	
Male	284	1.172 (0.879−1.563)	0.281
Age	456		
<=60	246	Reference	
>60	210	1.656 (1.251−2.192)	<0.001
T stage	361		
T1&T2	118	Reference	
T3&T4	243	2.085 (1.501−2.895)	<0.001
N stage	402		
N0&N1	297	Reference	
N2&N3	105	1.818 (1.310−2.523)	<0.001
M stage	430		
M0	406	Reference	
M1	24	1.897 (1.029−3.496)	0.04
Pathologic stage	410		
Stage I&Stage II	217	Reference	
Stage III&Stage IV	193	1.617 (1.207−2.165)	0.001
TMEM176B	456		
Low	226	Reference	
High	230	0.534 (0.407−0.701)	<0.001

FIGURE 2 | The prognostic value of *TMEM176B*. **(A)** ROC analysis showed that *TMEM176B* was able to identify tumor and normal tissue, the AUC was 0.668. **(B)** Survival evaluation indicators OS, *via* the expression level of the *TMEM176B* gene in SKCM. **(C)** Univariate Cox regression analysis of *TMEM176B* and clinicopathological variables with OS in SKCM.

The Expression of TMEM176B in SKCM Was Increased

We demonstrated the expression status of *TMEM176B* across various types of cancer. *TMEM176B* expression was considerably enhanced in some malignancies, as demonstrated in **Figure 1A**, which further demonstrates that heterogeneity is an important attribute of cancer and a major contributor to tumor progression. Compared to normal tissue, *TMEM176B* expression was higher in SKCM (p <0.001) (**Figure 1B**). The correlation analysis revealed that there was a statistical difference between *TMEM176B* and T stage, pathologic stage (p <0.05), and melanoma ulceration, radiation therapy, breslow depth (p <0.001) (**Figures 1C–G**).

The Prognostic Value of TMEM176B in SKCM

Since the significant differential expression within tumor and normal tissues, we investigated the effect of *TMEM176B* expression on the prognosis of tumor patients. Cancer cases were separated into high- and low-expression groups based on *TMEM176B* median expression levels to evaluate how *TMEM176B* expression influences prognosis of patients in SKCM. The area under the curve can be derived from the ROC curve as 0.668 (**Figure 2A**). High *TMEM176B* level is linked with a better prognosis, according to the Kaplan-Meier survival analysis (**Figure 2B**). We also summarized in **Supplementary Figures 1, 2** that the expression level of *TMEM176B* is closely associated with the prognosis of patients with various malignancies. We next combined clinical and pathological data, and analyzed the association between *TMEM176B* expression and a number of clinical parameters in SKCM patients. For OS, univariate Cox regression showed that *TMEM176B* played a detrimental role in patients with SKCM with the following characteristics: age >60 (n = 210, HR = 1.656, 95% CI from 1.251 to 2.192, p < 0.001), T stage 3 and 4 (n = 243, HR = 2.085, 95% CI from 1.501 to 2.895, p < 0.001), N stage 2 and 3 (n = 105, HR = 1.818, 95% CI from 1.310 to 2.523, p < 0.001), M stage 1 (n = 24, HR = 1.897, 95% CI from 1.029 to 3.496, p < 0.05), and Pathologic stage III and IV (n = 193, HR = 1.617, 95% CI from 1.207 to 2.165, p = 0.001) (**Figure 2C**). It is clear from this that TMEM176B plays an important role and serves as a reliable prognostic marker in SKCM.

TMEM176B-Related Gene Enrichment Analysis

We screened for proteins or related genes that might bind to TMEM176B in order to learn more about how it plays a role in carcinogenesis. GEPIA2 was used to identify the top 100 genes with the strongest relationship to *TMEM176B* in SKCM. The expression level of *TMEM176B* is positively linked with *TMEM176A* (Transmembrane Protein 176A) (R = 0.94), *LAIR1* (Leukocyte Associated Immunoglobulin Like Receptor 1) (R = 0.79), *SLC7A7* (Solute Carrier Family 7 Member 7) (R = 0.78), *ITGB2* (Integrin Subunit Beta 2) (R = 0.77), *HCK* (HCK Proto-Oncogene, Src Family Tyrosine Kinase) (R = 0.77), as illustrated in **Figures 3A–E**. The STRING program was also used to generate a network of proteins that interact with TMEM176B (**Figure 3F**). We subjected the top 100 genes obtained from GEPIA2 to GO/KEGG analysis, **Figures 3G,H** suggest these gene enriched in "regulation of lymphocyte activation" and "phagocytosis" pathway.

Correlation Between TMEM176B and Tumor Immune Infiltration

As can be seen from the above, *TMEM176B* was highly expressed in tumor tissues of SKCM and that higher expression correlated with better prognosis of SKCM patients. A variety of immune cells with different functions are present in the tumor microenvironment, including immunosuppressive cells and cells that promote tumor immunity (Quail and Joyce, 2013). As a result, it is necessary to continually explore the relationship between TMEM176B and the infiltration of immune cells. Data from TIMER suggests statistical significance between *TMEM176B* expression and immune cells, with B cells, $CD4^+$ T cells, $CD8^+$ T cells, macrophages, neutrophils, and dendritic cells were significantly positively correlated with the expression level of *TMEM176B* in SKCM (**Figure 4A**). *TMEM176B* expression levels, on the other hand, were significantly adversely associated with tumor purity (**Figure 4A**). We can also derive the association among *TMEM176B* expression and other immune cells from **Figure 4B**. **Supplementary Figures S3A–C** demonstrated correlation of *TMEM176B* expression and immune infiltration in different cancers of TCGA except SKCM. Among them, high expression of *TMEM176B* is associated with a higher degree of immune infiltration in most tumors. However, in LIHC, COAD, MESO and DLBC, the correlations were negative.

Using the CIBERSORT method, we investigated *TMEM176B* expression profiles to assess the level of 22 different immunocytes (**Figure 4C**). Our findings revealed that 9 kinds of immunocytes, including $CD8^+$ T cells, activated CD4 memory T cells, resting and activated NK cells, Monocytes, Macrophages, resting dendritic cells and eosinophils were influenced by *TMEM176B* expression, among these cells $CD8^+$ T cells and macrophages were significantly different between the *TMEM176B* high and low groups. The results of the high expression group compared to the low expression group showed that $CD8^+$ T cells increased, whereas M0 Macrophages decreased. These findings imply that TMEM176B modulates patient survival via immunological infiltrative interactions in the tumor microenvironment. $CD8^+$ cytotoxic T lymphocytes (CTLs) are essential in antitumor immunity. Therefore, the impact of TMEM176B on $CD8^+$ T cell infiltration level and the prognosis of tumor patients deserves close attention.

The Effects of *Tmem176b* on $CD8^+$ T Cells in the TME

We would like to further study the regulatory relationship between TMEM176B and $CD8^+$ T cells in the tumor microenvironment (TME). We investigated the role of

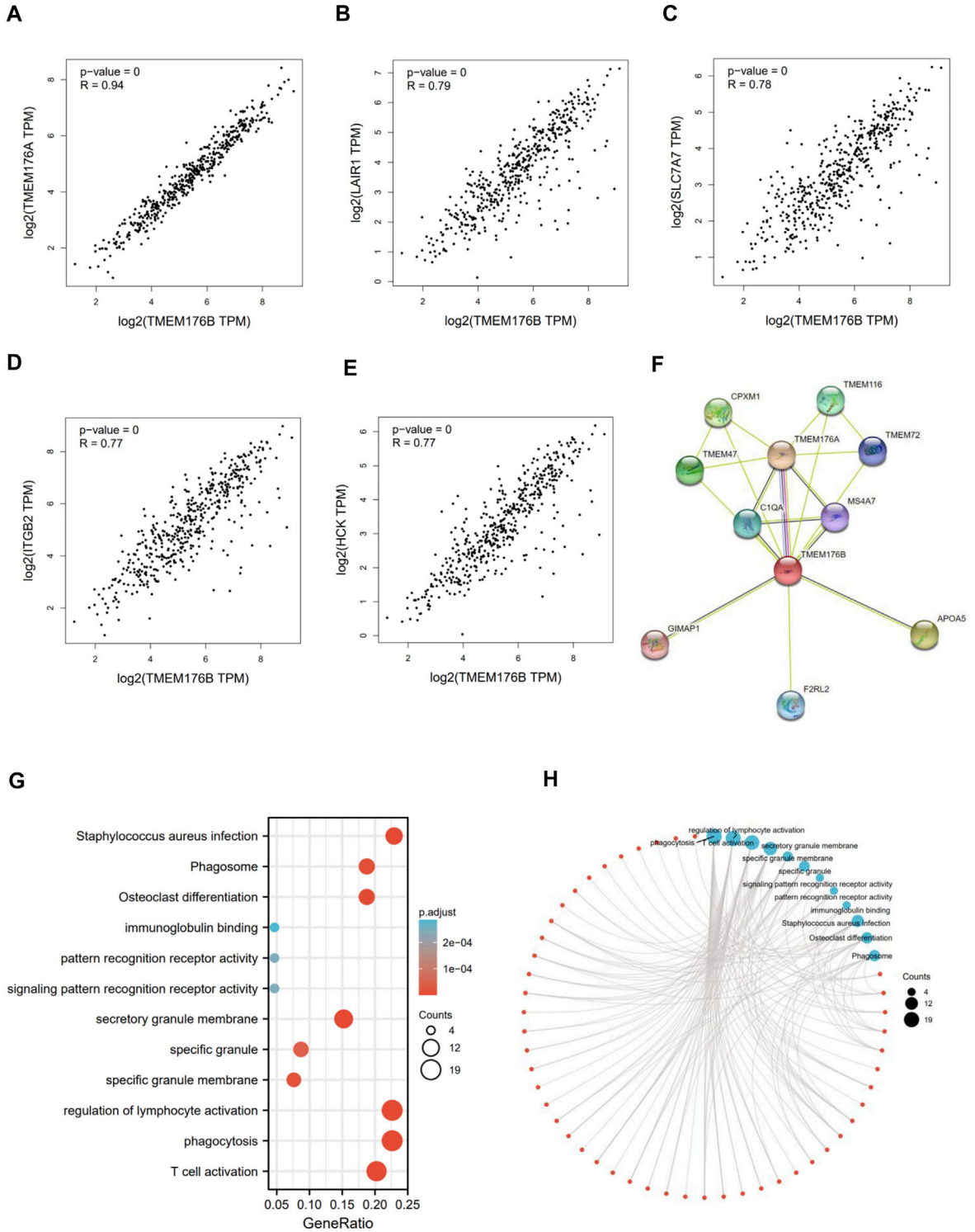


FIGURE 3 | TMEM176B-related gene enrichment analysis. **(A–E)** Using GEPIA2, we identified the top 100 genes related with *TMEM176B* in SKCM and examined the expression connection between *TMEM176B* and five genes, including*TMEM176A*, *LAIR1*, *SLC7A7*, *ITGB2*, and *HCK*. **(F)** Protein networks that may interact with TMEM176B, data from STRING. **(G,H)** Enrichment analysis of *TMEM176B*-related genes.

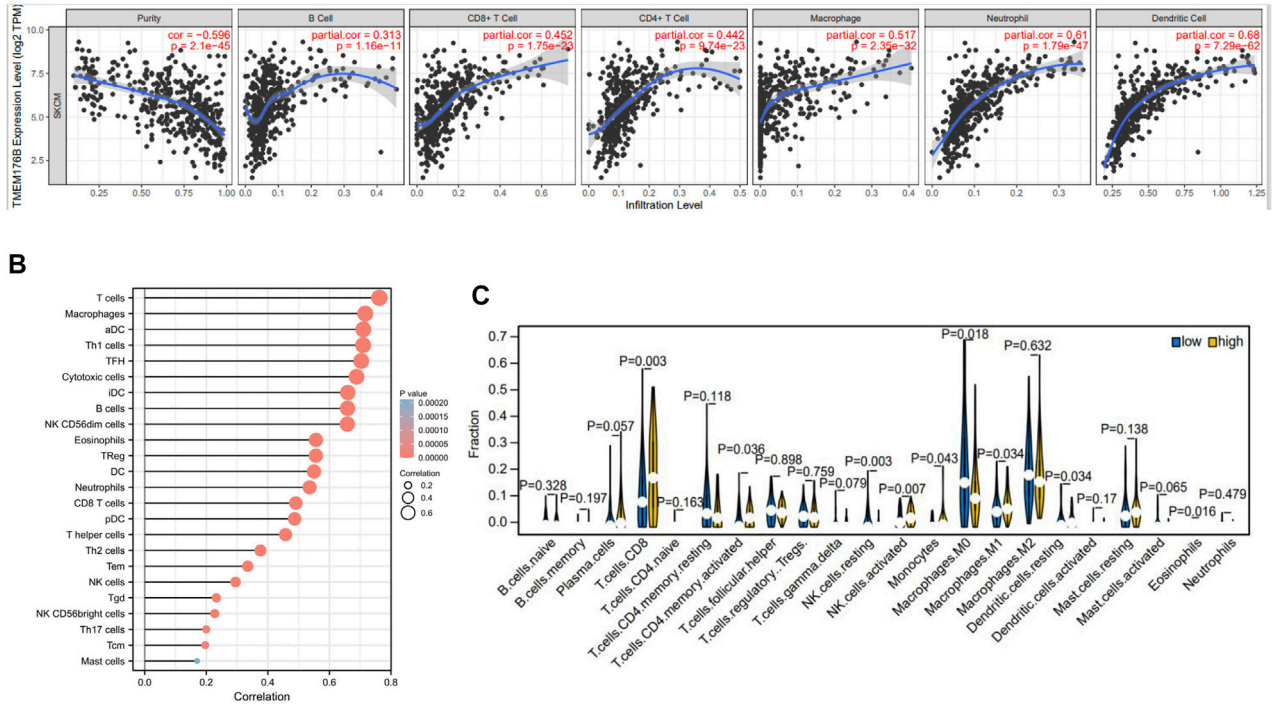


FIGURE 4 | Correlation of *TMEM176B* expression and the level of immune infiltration in SKCM. **(A,B)** In SKCM, the infiltration levels of B cells, $CD8^+$ T cells, $CD4^+$ T cells, macrophages, neutrophils, dendritic cells and other immune cells with *TMEM176B* expression. $p < 0.05$ is considered significant. **(C)** The varied proportions of 22 subtypes of immune cells in high and low *TMEM176B* expression groups in tumor samples.

Tmem176b in a mouse melanoma model and sought to continue to explore the role of *Tmem176b* in the regulation of $CD8^+$ T cells in the tumor microenvironment. We constructed a mouse tumor model (**Figure 5A**) to figure out how *Tmem176b* regulates the differentiation, function, and immune infiltration levels of $CD8^+$ T cells in the tumor microenvironment. After implanting the tumor subcutaneously for 14 days, we euthanized the mice for flow cytometry analysis. The tumor volume of mice following *Tmem176b* knockout was similarly greater and grew quicker than that of WT mice ($p < 0.05$) (**Figure 5B**).

Our findings revealed that the proportion of $CD3^+$ $CD8^+$ T cells in Tumor-infiltrating T lymphocytes (TILS), is dramatically reduced after *Tmem176b* knockout (**Figures 5C,D**), the same trend can be seen in $CD3^+$ $CD4^+$ T cells (**Figure 5H**). This downward trend was also found in peripheral blood, spleen, and lymph nodes, although we did not obtain statistical differences in all tissues (**Supplementary Figure S4A**), and this difference is more pronounced in $CD8^+$ T cells, consistent with our previous results that higher expression of *TMEM176B* in SKCM was associated with higher levels of infiltration of immune cells (**Figure 4**). $CD8^+$ T cells are subdivided into naive T cells (Tnaive), effector T cells (Teff) and memory T cells (Tmem) subsets. In our study, compared to wild type mice, more $CD8^+$ Tnaive and less $CD8^+$ Teff presented in tumor of $Tmem176b^{-/-}$ mice (**Figures 5E,F**), and the same change had also been observed in tumor-infiltrated $CD4^+$ T cells (**Figure 5H**). However, Tmem remained unchanged between groups (**Figures 5G,H**). The influence of *Tmem176b* on the differentiation of T cells was not significant in other non-tumor organs (**Supplementary Figures S4B,C**). It suggested that T cell populations were normal in non-tumor organs and T cell homeostasis was altered in response to immunogenic stress exhibited by the facts that increased Tnaive cell number and decreased Teff cell number were presented in the local tumor of $Tmem176b^{-/-}$ mice.

We also looked at the expression of some immunosuppressive molecules following tumor transplantation. Comparing to WT mice, our findings revealed that the expression of multiple activation or effector function-related molecules were inhibited in tumor-infiltrated $CD8^+$ T cells of $Tmem176b^{-/-}$ mice (**Figure 6**) although the trend was not consistent in blood, spleen or lymph node (**Supplementary Figures S4D, S5**). Our findings showed that after tumor implantation, knocking down *Tmem176b* has effect on T cell activation rather than anti-apoptosis, or proliferation (**Supplementary Figure S5**).

We could conclude from **Figure 6** that the expression levels of cytotoxic molecules such as IFN-γ, PRF1, and GZMB in tumor-infiltrating $CD8^+$ T cells was reduced following *Tmem176b* deletion, indicating that the function of tumor-infiltrating T cells was significantly compromised. It is the possible reason for the larger tumor size in *Tmem176b* knockout mice. To verify our results, we researched the correlation between the expression

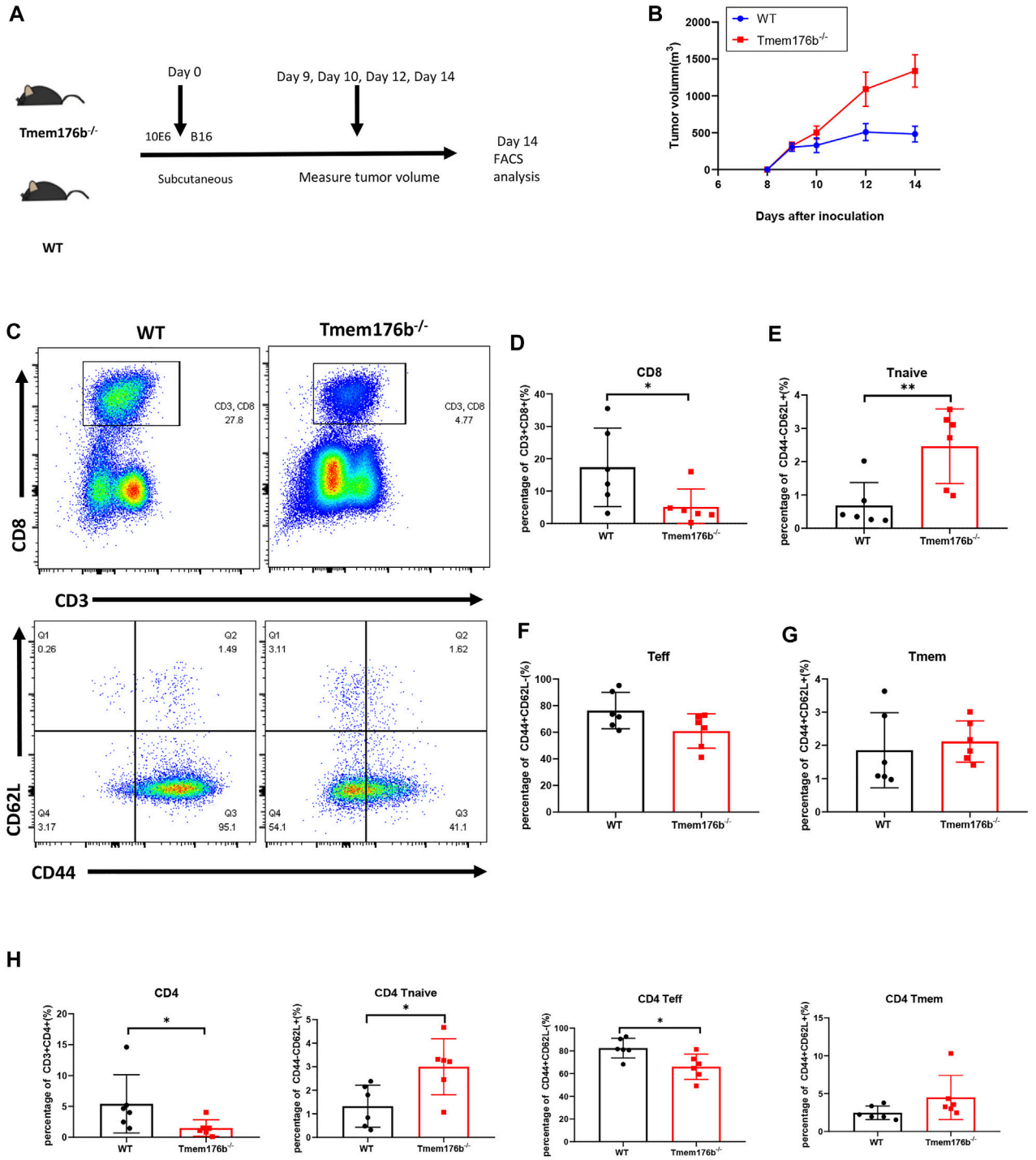


FIGURE 5 | The regulatory effect of *Tmem176b* on T cells after tumor implantation. **(A)** Schematic diagram of mouse tumor model experiment. Six female mice per group. **(B)** Tumor growth curves of WT and *Tmem176b*$^{-/-}$ mice injected subcutaneously with 1 × 10^6 B16-OVA melanoma cells (per group $n = 6$). **(C)** Flow cytometry of tumor-infiltrating lymphocytes TIL CD8$^+$ T cells (top rows) and its subgroup (bottom rows) from WT and *Tmem176b*$^{-/-}$ mice at day 14 after subcutaneous injection with 1 × 10^6 B16-OVA melanoma cells ($n = 6$). **(D–G)** Percentage of tumor-infiltrating CD3$^+$ CD8$^+$ T cells and their subpopulations Tnaive, Teff, Tmem from mice as in **(C)** ($n = 6$). **(H)** Percentage of tumor-infiltrating CD3$^+$ CD4$^+$ T cells and their subpopulations Tnaive, Teff, Tmem from mice as in **(C)** ($n = 6$). Each symbol represents a single mouse; data are mean SD from six independent samples. ns, not significant (two-tailed unpaired Student's t test). *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

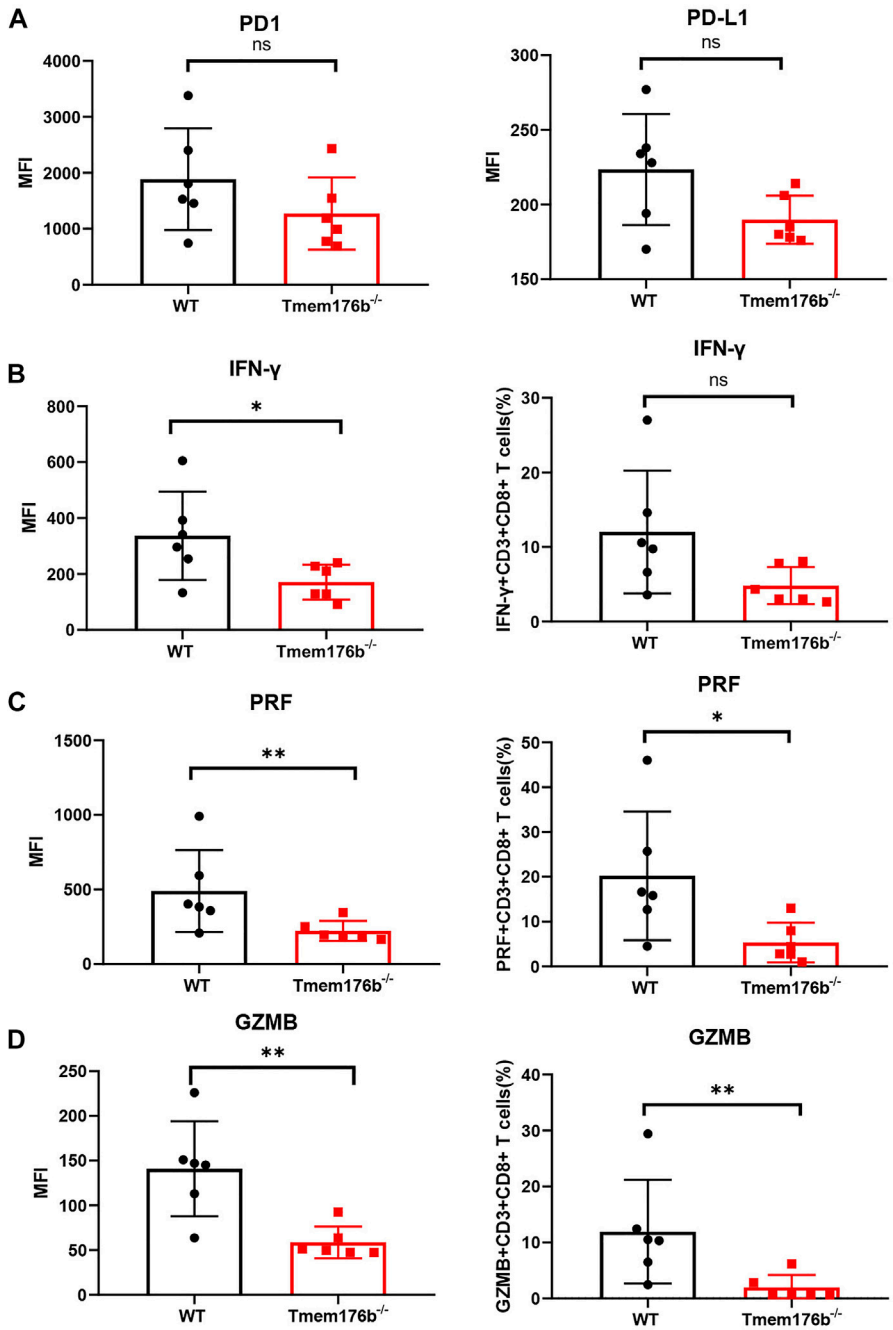


FIGURE 6 | The regulatory effect of Tmem176b on T cells after tumor implantation. **(A)** The expression level of PD1 and PD-L1 of tumor-infiltrating CD8$^+$ T lymphocytes (TIL CD8$^+$ T cells) from WT and *Tmem176b*$^{-/-}$ mice ($n = 6$). **(B–D)** The MFI (left row) level of IFN-γ, PRF and GZMB of TIL CD8$^+$ T cells, and frequency (right row) of IFN-γ$^+$, PRF$^+$ or GZMB$^+$ (among TIL CD8$^+$ T cells) from mice as in **(A)** ($n = 6$). Each symbol represents a single mouse; data are mean SD from six independent samples. ns, not significant (two-tailed unpaired Student's t test). *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

of *TMEM176B* and some molecules in SKCM. As shown in **Supplementary Figure S6**, the expression of *TMEM176B* was significantly positive related to the expression levels of PD-L1 (CD274), CTLA4, LAG3, GZMK, GZMB, GZMA, GZMM, PRF1, TNF and IFNG in SKCM. PD1 is a hallmark of T-cell exhaustion related to co-expression of other checkpoint molecules such as LAG-3 and TIM-3, as well as an activation marker for T-cells including antigen specific responses. The results of this part indicate that *Tmem176b* positively regulates the effector function and activation of tumor-infiltrating CD8$^+$ T cells.

DISCUSSION

Immune cells play an essential role in tumor immunity, however, to date, there are few reports on the association of TMEM176B with immune cells. *Tmem176b* is predominantly expressed in BMDCs and cDC, and is involved in maintaining the immature state of DC (Condamine et al., 2010). In addition, strong expression of *Tmem176b* in a family of retinoic acid-related orphan receptor RORγt$^+$ lymphocytes have been reported (Drujont et al., 2016; Lancien et al., 2021). Our research discovered a link between the expression of *TMEM176B* and the infiltration degree of immunocytes in cancers. *TMEM176B* was found to be substantially connected with the purity of most tumors in TIMER, as well as the infiltration of B cells, CD4$^+$ T cells, CD8$^+$ T cells, macrophages, neutrophils, and dendritic cells. In SKCM, our results indicated that *TMEM176B* was significantly and positively correlated with the infiltration of most immune cells. Robust infiltration of lymphocytes in melanoma correlates with a reduced risk of tumor growth (Galon et al., 2006).

Next, we used an animal model to identify the connection between *Tmem176b* and immune cell infiltration. After tumor implantation, we tested the effect of deleting *Tmem176b* on T lymphocyte development and function in diverse tissues and immunological organs. Lancien, M. et al. reported that *Tmem176a* and *Tmem176b* (*Tmem176a/b*) deficiency has no effect on tumor growth in MCA101-sOVA fibrosarcoma and B16-OVA melanoma (Lancien et al., 2021). However, our results showed that in comparison to wild-type mice, *Tmem176b* knockout mice had a greater volume of tumors. The study has (Segovia et al., 2019) reported that targeting TMEM176B controls tumor growth. The observed differences and inconsistency may due to the fact of tumor heterogeneity in which diverse roles of certain gene in different tumor models.

It was found that deletion of *Tmem176a* and *Tmem176b* does not affect CD8$^+$ T cell responses but plays an irreplaceable role in the optimal initiation of initial CD4$^+$ T cells by DCs (Lancien et al., 2021). Our results showed that the overall proportion of CD3$^+$ CD4$^+$ T and CD3$^+$ CD8$^+$ T cells in tissues was reduced, and tumor-infiltrating T lymphocytes (TIL CD4$^+$ T, TIL CD8$^+$ T) were significantly reduced in *Tmem176b* $^{-/-}$mice, which is consistent with our findings from public databases, where the higher the expression of *TMEM176B*, the higher the degree of T cell infiltration in SKCM. The reduced proportion of CD8$^+$ T cells in *Tmem176b* knockout mouse could explain our finding of the accelerating progress and a poor prognosis in same mice model. Because high levels of CD8$^+$ T cell infiltration have been shown to facilitate tumor treatment in tumors such as breast cancer and uroepithelial carcinoma (Sharma et al., 2007; Mahmoud et al., 2011).

In addition, we observed that the expression of some immunosuppressive molecules such as PD1, PD-L1 in *Tmem176b* knockout mice was lower than WT mice, and this downward trend was more significant at the level of PD-L1 expression. Although sometimes high PD-L1 levels predict suppression of anti-cancer immune responses upon binding to PD-1, PD-L1 expression is not necessarily equivalent to immune escape (Taube et al., 2014). A clinical study has shown that high PD-L1 expression on immune cells in uremic patients is associated with an increased probability of response to atezolizumab (An immune checkpoint blocker targeting PD-L1) (Rosenberg et al., 2016). Therefore, it is necessary to clarify that although high expression of immunosuppressive molecules or their ligands in tumors usually indicates the failure of the immune response that accompanies disease progression, they may also be a marker for the effectiveness of blockade therapies (Galluzzi et al., 2018). It has been explained that transcriptional regulation of PD-L1 may promote immune inflammation in the local tumor microenvironment, thus making tumors more sensitive to immune checkpoint blockade therapies (Gao et al., 2020). In addition, the increased expression of PD-1 in CD8$^+$ T cells does not exactly indicate that the cells are in an exhaustion state, but also symbolize the activation of the cells (Wherry and Kurachi, 2015).

In the study, we found that, in comparison to wild-type animals, *Tmem176b* knockout lowered the expression of IFN-γ, PRF1 and GZMB in TIL CD8$^+$ T cells, indicating that the function of tumor-infiltrating T cells was suppressed. As a result of that enhanced expression of IFN-γ and GZMB in the tumor microenvironment can achieve optimal tumor clearance (Lin et al., 2014). We analyzed the expression of *TMEM176B* in SKCM with the expression of some T cell-related genes to confirm our findings in the mouse tumor model, and we discovered that the expression of *TMEM176B* was significantly and positively linked with the expression of immunosuppressive molecules CD274 (PD-L1), CTLA4, LAG3, and cytotoxic molecules associated with T cell function TNF, PRF1, GZMA, GZMK, and GZMM, and most cytotoxic molecules, have been validated to act mainly as anti-tumoral and anti-infectious factors (Arias et al., 2017).

CONCLUSION

Taken together, as one of the aggressive malignancies, the survival rate of melanoma varies depending on the stage of diagnosis. We discovered that *TMEM176B* expression was higher in SKCM than in normal tissue, and enhanced expression was linked to better OS. Also, TMEM176B was positively linked to immune cell infiltration in SKCM, which may explain that TMEM176B affects the prognosis of tumor patients. These findings lead us to hypothesize that TMEM176B can be employed as a diagnostic and prognostic marker. To validate our hypothesis, we developed melanoma model in *Tmem176b* knockout mice and revealed that T cell infiltration and effector function were significantly inhibited. It suggests that TMEM176B plays an essential role in anti-tumor function of CD8$^+$ T cells, but the underlying mechanisms need to be further investigated.

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 animal study was reviewed and approved by the Zhongshan Hospital, Fudan University Ethical Committee.

AUTHOR CONTRIBUTIONS

DW and ZS gained funding, designed the study and revised the manuscript. LJ and YY performed experiments and wrote the manuscript; FL and MM analyzed the data; JG, LS, and YC performed experiments and drew figures.

FUNDING

This study was supported by the National Natural Science Foundation of China (No. 81771672), "Cross key project of mathematics and medical health" of the National Natural Science Foundation of China (No. 12026608), Open fund project of Shenzhen BGI Institute of Life Science (No. BGIRSZ20200004) and Special Fund for Clinical Research of Zhongshan Hospital, Fudan University (No. 2020ZSLC07), Key Subject Construction Program of Shanghai Health Administrative Authority (ZK 2019B30), and Shanghai Engineering Research Center of Tumor Multi-Target Gene Diagnosis (20DZ2254300).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2022.859958/full#supplementary-material

Supplementary Table 1 | Clinical characteristics of the SKCM patients based on TCGA.

Supplementary Table 2 | The number of cases used in each cancer types.

REFERENCES

Arias, M., Martínez-Lostao, L., Santiago, L., Ferrandez, A., Granville, D. J., and Pardo, J. (2017). The Untold Story of Granzymes in Oncoimmunology: Novel Opportunities with Old Acquaintances. *Trends Cancer* 3 (6), 407–422. doi:10.1016/j.trecan.2017.04.001

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., et al. (2013). Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the Immune Landscape in Human Cancer. *Immunity* 39 (4), 782–795. doi:10.1016/j.immuni.2013.10.003

Choi, Y., Jeon, Y.-H., Kang, J.-H., Chung, J.-K., Schmidt, M., and Kim, a. C.-W. (2007). MIDGE/hNIS Vaccination Generates Antigen-Associated CD8+IFN-γ+ T Cells and Enhances Protective Antitumor Immunity. *Int. J. Cancer* 120 (9), 1942–1950. doi:10.1002/ijc.22567

Condamine, T., Le Texier, L., Howie, D., Lavault, A., Hill, M., Halary, F., et al. (2010). Tmem176B and Tmem176A Are Associated with the Immature State of Dendritic Cells. *J. Leukoc. Biol.* 88 (3), 507–515. doi:10.1189/jlb.1109738

Cuajungco, M. P., Podevin, W., Valluri, V. K., Bui, Q., Nguyen, V. H., and Taylor, K. (2012). Abnormal Accumulation of Human Transmembrane (TMEM)-176A and 176B Proteins Is Associated with Cancer Pathology. *Acta Histochem.* 114 (7), 705–712. doi:10.1016/j.acthis.2011.12.006

Drujont, L., Lemoine, A., Moreau, A., Bienvenu, G., Lancien, M., Cens, T., et al. (2016). RORγt+ Cells Selectively Express Redundant Cation Channels Linked to the Golgi Apparatus. *Sci. Rep.* 6, 23682. doi:10.1038/srep23682

Eon Kuek, L., Leffler, M., Mackay, G. A., and Hulett, M. D. (2016). The MS4A Family: Counting Past 1, 2 and 3. *Immunol. Cel Biol* 94 (1), 11–23. doi:10.1038/icb.2015.48

Galluzzi, L., Chan, T. A., Kroemer, G., Wolchok, J. D., and López-Soto, A. (2018). The Hallmarks of Successful Anticancer Immunotherapy. *Sci. Transl Med.* 10 (459), eaat7807. doi:10.1126/scitranslmed.aat7807

Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pagès, C., et al. (2006). Type, Density, and Location of Immune Cells within Human Colorectal Tumors Predict Clinical Outcome. *Science* 313 (5795), 1960–1964. doi:10.1126/science.1129139

Gao, Y., Nihira, N. T., Bu, X., Chu, C., Zhang, J., Kolodziejczyk, A., et al. (2020). Acetylation-Dependent Regulation of PD-L1 Nuclear Translocation Dictates the Efficacy of Anti-PD-1 Immunotherapy. *Nat. Cel Biol* 22 (9), 1064–1075. doi:10.1038/s41556-020-0562-4

Hollestein, L. M., van den Akker, S. A. W., Nijsten, T., Karim-Kos, H. E., Coebergh, J. W., and de Vries, E. (2012). Trends of Cutaneous Melanoma in The Netherlands: Increasing Incidence Rates Among All Breslow Thickness Categories and Rising Mortality Rates since 1989. *Ann. Oncol.* 23 (2), 524–530. doi:10.1093/annonc/mdr128

Huntington, N. D., Cursons, J., and Rautela, J. (2020). The Cancer-Natural Killer Cell Immunity Cycle. *Nat. Rev. Cancer* 20 (8), 437–454. doi:10.1038/s41568-020-0272-z

Kim, H., Kim, H., Feng, Y., Li, Y., Tamiya, H., Tocci, S., et al. (2020). PRMT5 Control of cGAS/STING and NLRC5 Pathways Defines Melanoma Response to Antitumor Immunity. *Sci. Transl Med.* 12 (551), eaaz5683. doi:10.1126/scitranslmed.aaz5683

Lancien, M., Bienvenu, G., Salle, S., Gueno, L., Feyeux, M., Merieau, E., et al. (2021). Dendritic Cells Require TMEM176A/B Ion Channels for Optimal MHC Class II Antigen Presentation to Naive CD4+ T Cells. *J. Immunol.* 207 (2), 421–435. doi:10.4049/jimmunol.2000498

Lin, R., Chen, L., Chen, G., Hu, C., Jiang, S., Sevilla, J., et al. (2014). Targeting miR-23a in CD8+ Cytotoxic T Lymphocytes Prevents Tumor-dependent Immunosuppression. *J. Clin. Invest.* 124 (12), 5352–5367. doi:10.1172/jci76561

Lin, W. M., and Fisher, D. E. (2017). Signaling and Immune Regulation in Melanoma Development and Responses to Therapy. *Annu. Rev. Pathol. Mech. Dis.* 12, 75–102. doi:10.1146/annurev-pathol-052016-100208

Liu, J., Lichtenberg, T., Hoadley, K. A., Poisson, L. M., Lazar, A. J., Cherniack, A. D., et al. (2018). An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* 173 (2), 400–416. doi:10.1016/j.cell.2018.02.052

Louvet, C., Chiffoleau, E., Heslan, M., Tesson, L., Heslan, J.-M., Brion, R., et al. (2005). Identification of a New Member of the CD20/FcepsilonRIbeta Family Overexpressed in Tolerated Allografts. *Am. J. Transpl.* 5 (9), 2143–2153. doi:10.1111/j.1600-6143.2005.01007.x

Lurton, J., Rose, T. M., Raghu, G., and Narayanan, A. S. (1999). Isolation of a Gene Product Expressed by a Subpopulation of Human Lung Fibroblasts by Differential Display. *Am. J. Respir. Cel Mol Biol* 20 (2), 327–331. doi:10.1165/ajrcmb.20.2.3368

Ma, C., and Li, H. (2021). Hub Gene and its Key Effects on Diffuse Large B-Cell Lymphoma by Weighted Gene Coexpression Network Analysis. *Biomed. Res. Int.* 2021, 8127145. doi:10.1155/2021/8127145

Mahmoud, S. M. A., Paish, E. C., Powe, D. G., Macmillan, R. D., Grainge, M. J., Lee, A. H. S., et al. (2011). Tumor-Infiltrating CD8+ Lymphocytes Predict Clinical Outcome in Breast Cancer. *J. Clin. Oncol.* 29 (15), 1949–1955. doi:10.1200/jco.2010.30.5037

Michielin, O., van Akkooi, A. C. J., Ascierto, P. A., Dummer, R., and Keilholz, U. (2019). Cutaneous Melanoma: ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up. *Ann. Oncol.* 30 (12), 1884–1901. doi:10.1093/annonc/mdz411

Otsubo, T., Hida, Y., Ohga, N., Sato, H., Kai, T., Matsuki, Y., et al. (2014). Identification of Novel Targets for Antiangiogenic Therapy by Comparing the Gene Expressions of Tumor and Normal Endothelial Cells. *Cancer Sci.* 105 (5), 560–567. doi:10.1111/cas.12394

Quail, D. F., and Joyce, J. A. (2013). Microenvironmental Regulation of Tumor Progression and Metastasis. *Nat. Med.* 19 (11), 1423–1437. doi:10.1038/nm.3394

Rosenberg, J. E., Hoffman-Censits, J., Powles, T., van der Heijden, M. S., Balar, A. V., Necchi, A., et al. (2016). Atezolizumab in Patients with Locally Advanced and Metastatic Urothelial Carcinoma Who Have Progressed Following Treatment with Platinum-Based Chemotherapy: A Single-Arm, Multicentre, Phase 2 Trial. *Lancet* 387 (10031), 1909–1920. doi:10.1016/S0140-6736(16)00561-4

Schadendorf, D., Fisher, D. E., Garbe, C., Gershenwald, J. E., Grob, J.-J., Halpern, A., et al. (2015). Melanoma. *Nat. Rev. Dis. Primers* 1, 15003. doi:10.1038/nrdp.2015.3

Segovia, M., Russo, S., Jeldres, M., Mahmoud, Y. D., Perez, V., Duhalde, M., et al. (2019). Targeting TMEM176B Enhances Antitumor Immunity and Augments the Efficacy of Immune Checkpoint Blockers by Unleashing Inflammasome Activation. *Cancer Cell* 35 (5), 767–781. doi:10.1016/j.ccell.2019.04.003

Shain, A. H., and Bastian, B. C. (2016). From Melanocytes to Melanomas. *Nat. Rev. Cancer* 16 (6), 345–358. doi:10.1038/nrc.2016.37

Sharma, P., Shen, Y., Wen, S., Yamada, S., Jungbluth, A. A., Gnjatic, S., et al. (2007). CD8 Tumor-Infiltrating Lymphocytes Are Predictive of Survival in Muscle-Invasive Urothelial Carcinoma. *Proc. Natl. Acad. Sci. U.S.A.* 104 (10), 3967–3972. doi:10.1073/pnas.0611618104

Sun, L., Zhang, Y., and Zhang, C. (2018). Distinct Expression and Prognostic Value of MS4A in Gastric Cancer. *Open Med. (Wars)* 13, 178–188. doi:10.1515/med-2018-0028

Taube, J. M., Klein, A., Brahmer, J. R., Xu, H., Pan, X., Kim, J. H., et al. (2014). Association of PD-1, PD-1 Ligands, and Other Features of the Tumor Immune Microenvironment with Response to Anti-PD-1 Therapy. *Clin. Cancer Res.* 20 (19), 5064–5074. doi:10.1158/1078-0432.ccr-13-3271

Tomimaru, Y., Mishra, S., Safran, H., Charpentier, K. P., Martin, W., De Groot, A. S., et al. (2015). Aspartate-β-Hydroxylase Induces Epitope-Specific T Cell Responses in Hepatocellular Carcinoma. *Vaccine* 33 (10), 1256–1266. doi:10.1016/j.vaccine.2015.01.037

Viklicky, O., Krystufkova, E., Brabcova, I., Sekerkova, A., Wohlfahrt, P., Hribova, P., et al. (2013). B-Cell-Related Biomarkers of Tolerance Are Up-Regulated in Rejection-Free Kidney Transplant Recipients. *Transplantation* 95 (1), 148–154. doi:10.1097/tp.0b013e3182789a24

Wherry, E. J., and Kurachi, M. (2015). Molecular and Cellular Insights into T Cell Exhaustion. *Nat. Rev. Immunol.* 15 (8), 486–499. doi:10.1038/nri3862

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

ACC Adrenocortical carcinoma

BLCA Bladder urothelial carcinoma

BRCA Breast invasive carcinoma

CESC Cervical squamous cell carcinoma and endocervical adenocarcinoma

CHOL Cholangiocarcinoma

COAD Colon adenocarcinoma

DLBC Lymphoid neoplasm diffuse large B-cell lymphoma

ESCA Esophageal carcinoma

GBM Glioblastoma multiforme

GBMLGG Glioma

HNSC Head and neck squamous cell carcinoma

KICH Kidney chromophobe

KIRC Kidney renal clear cell carcinoma

KIRP Kidney renal papillary cell carcinoma

LAML Acute myeloid leukemia

LGG Brain lower grade glioma

LIHC Liver hepatocellular carcinoma

LUAD Lung adenocarcinoma

LUSC Lung squamous cell carcinoma

MESO Mesothelioma

OV Ovarian serous cystadenocarcinoma

PAAD Pancreatic adenocarcinoma

PCPG Pheochromocytoma and paraganglioma

PRAD Prostate adenocarcinoma

READ Rectum adenocarcinoma

SARC Sarcoma

SKCM Skin cutaneous melanoma

STAD Stomach adenocarcinoma

TGCT Testicular germ cell tumors

THCA Thyroid carcinoma

THYM Thymoma

UCEC Uterine corpus endometrial carcinoma

UCS Uterine carcinosarcoma

UVM Uveal melanoma

frontiers | Frontiers in Cell and Developmental Biology

ORIGINAL RESEARCH
published: 11 April 2022
doi: 10.3389/fcell.2022.839821

Identification of KIF23 as a Prognostic Biomarker Associated With Progression of Clear Cell Renal Cell Carcinoma

Zonglong Wu†, Yimeng Song†, Yaqian Wu, Liyuan Ge, Zhuo Liu, Tan Du, Shudong Zhang and Lulin Ma**

Department of Urology, Peking University Third Hospital, Beijing, China

About 3% of adult cancers are caused by renal cell carcinoma (RCC) and its pathogenesis remains elusive. Among RCC, clear cell renal cell carcinoma (ccRCC) is the predominant histological subtype. Resistance to conventional treatments leaves few treatment options for advanced ccRCC. Although the transcriptome profile of primary ccRCC has been comprehensively summarized, the transcriptome profile of metastatic ccRCC is still lacking. In this study we identified a list of metastasis-related genes and constructing a metastasis-associated prognostic gene signature. By analyzing data from GSE85258 and GSE105288 datasets, 74 genes were identified as metastasis-related genes. To construct prognostic features, we downloaded the expression data of ccRCC from the Cancer Genome Atlas (TCGA). Metastasis-associated genes were initially selected through the LASSO Cox regression analysis and 12 metastasis-related were included to construct prognostic model. Transcriptome profile, patient prognosis, and immune cell infiltration characteristics differed between low- and high-risk groups after grouping according to median risk score. Through explored the functions of differentially expressed genes (DEGs) between the two groups. Kinesin family member 23 (KIF23) was identified as a prognostic marker in ccRCC patients. Furthermore, inhibition of KIF23 expression reduced the proliferation, migration and invasion of ccRCC cells. We further demonstrated that KIF23 promote nuclear translocation of β-catenin in ccRCC cells, which provides novel insight into the functions and molecular machinery of KIF23 in ccRCC.

Keywords: renal cell carcinoma, kinesin family member 23, metastasis-associated genes, prognosis signature, Wnt/β-catenin signaling pathway

OPEN ACCESS

Edited by:
Fangqing Zhao,
Beijing Institutes of Life Science (CAS), China

Reviewed by:
Yu-Zheng Ge,
Nanjing Medical University, China
Di Gu,
First Affiliated Hospital of Guangzhou Medical University, China

****Correspondence:***
Shudong Zhang
shootong@163.com
Lulin Ma
malulinpku@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: *20 December 2021*
Accepted: *28 March 2022*
Published: *11 April 2022*

Citation:
Wu Z, Song Y, Wu Y, Ge L, Liu Z, Du T, Zhang S and Ma L (2022) Identification of KIF23 as a Prognostic Biomarker Associated With Progression of Clear Cell Renal Cell Carcinoma. Front. Cell Dev. Biol. 10:839821. doi: 10.3389/fcell.2022.839821

INTRODUCTION

About 3% of adult cancers are caused by renal cell carcinoma (RCC) and its pathogenesis remains elusive. ccRCC is the predominant histological subtype and is prone to metastases. Patients with metastatic ccRCC have a poor prognosis, with a 5-year overall survival (OS) rate is 8–12% (Choueiri and Motzer, 2017). ccRCC is a highly vascularized malignant tumor. Therapies targeting angiogenesis are effective in tumor regression. However, most patients treated with targeted agents often develop resistance within 1 year of treatment, which warrants new treatment strategies (Rini and Atkins, 2009). ccRCC is a highly heterogeneous disease with complex molecular signatures and biological characteristics (Iacovelli et al., 2014). Several biological

factors may affect the biological process and treatment response. An accurate understanding of the characteristics of ccRCC will help in the management of ccRCC patients. Therefore, stratifying the risk of patients to discover their molecular characteristics and predict response to personalized drug treatment is of great significance.

Due to insidious progression, nearly one-third of ccRCC patients have metastases at initial medical attention, approximately 60% of non-metastatic ccRCC patients develop lung, bone or other organ metastases within 2–3 years (Park and Eisen, 2007; Casuscelli et al., 2019). Metastasis is the main reason of ccRCC-related deaths. Although some patients benefit from immunotherapy, ccRCC remains an incurable disease for the majority of patients (Escudier et al., 2020). Therefore, exploring a new biomarker of tumor metastasis is crucial for monitoring metastasis and treatment. Although somatic mutation and transcriptome profiles of primary ccRCC have been comprehensively summarized in the Cancer Genome Atlas (TCGA) database, the transcriptome and genome profiles of metastatic ccRCC are largely lacking. Analyzing the differential gene expression between metastatic and primary tumors helps to determine factors influencing metastasis and predict prognosis in advance.

In this study, we compared transcriptome profiles between metastatic and primary tumors to identify genes associated with ccRCC metastases. Furthermore, prognostic signatures of metastasis-related genes were constructed. After stratifying ccRCC patients according to risk score, transcriptional profiles, patient prognosis, and immune cell infiltration characteristics differed between low- and high-risk groups. Finally, we identified kinesin family member 23 (KIF23) as an important gene that drives ccRCC progression and affects patient prognosis. Knockdown of KIF23 expression in ccRCC cells attenuated cell proliferation migration and invasion, and reduced nuclear translocation of β-catenin.

MATERIALS AND METHODS

Identification of Metastasis-Associated Genes in Clear Cell Renal Cell Carcinoma

Microarray datasets GSE85258 and GSE105288 in the Gene Expression Omnibus (GEO) database were used to analyze the transcriptomic signature of metastatic ccRCC. The GSE85258 dataset contained 15 primary tumors and 16 metastatic tumor samples. The GSE105288 dataset contained nine primary tumors and 26 metastatic tumor samples. By GEO2R tool analysis, genes with fold change (FC) > 1.5 and $p < 0.05$ in metastatic ccRCC compared with primary tumors were defined as metastasis-related genes. Next, we obtained gene expression profiles of 530 ccRCCs from the TCGA data portal and analyzed them by the limma package. The screening criteria for differentially expressed genes (DEGs) were $p < 0.05$ and FC > 2. The Ggplot2 and pheatmap packages were used for volcano plots and heatmaps generation.

Construction of Clinical Prognostic Model

The R package "glmnet" performed LASSO regression analysis to identify key prognostic genes. One standard error of the optimal penalty parameter λ value after cross-validation was used to build the simplest gene feature model. The risk score was determined by calculating the sum of the product of each gene expression level and its corresponding coefficient according to the risk score formula. Next, patients were stratified into two groups based on the median risk score. Differences in survival time between the two groups were estimated by the Kaplan-Meier method. Time-related receiver operating characteristic (ROC) analysis was used to verify the prediction accuracy of the model.

CIBERSORT Analysis

CIBERSORT analysis tool was used to infers the proportion of infiltrating immune cells based on cell-type-specific gene expression profiles from the mixed cell data. Inferred results were considered accurate at $p < 0.05$.

Functional Enrichment Analysis

ClusterProfiler package performed Gene Ontology (GO) and the Kyoto encyclopedia of genes and genomes (KEGG) pathway analysis (Yu et al., 2012). The PPI network was retrieved by the STRING database and reconstructed by Cytoscape software (Shannon et al., 2003; Szklarczyk et al., 2015). Cytohubba was used to analysis and select top 10 hub genes.

Gene Set Enrichment Analysis Analysis

The C2 kegg gene set was used for GSEA enrichment analysis and FDR values <0.01 were considered significantly enriched.

Clear Cell Renal Cell Carcinoma Specimens and Cell Culture

The ccRCC tissues and normal samples of 10 ccRCC patients involved in this study were obtained from Peking University Third Hospital after approval by the Ethics Committee of Peking University Third Hospital. The ccRCC cell lines A498, caki-1, caki-2, 786-O and OS-RC2 and the immortalized human kidney cell line HK-2 were purchased from the American Type Culture Collection. 1640 medium (Biological Industries) containing 10% fetal bovine serum (FBS) (Biological Industries) was used to culture cells.

Synthesis of Small Interfering RNA and Transfection of Cells

Cells were transfected with specific siRNAs targeting KIF23 (si-KIF23). si-KIF23 were purchased from GenePharma Biotech (Shanghai, China). The siRNA sequence is:

si-KIF23-1 sense: 5′-GGUCCCAAACGAACCUUAATT-3′,
antisense: 5′-UUAAGGUUCGUUUGGGACCTT-3′;
si-KIF23-2 sense: 5′- GCUAUUGUUACCGAACCUATT-3′,
antisense: 5′- UAGGUUCGGUAACAAUAGCTT-3′;
si-KIF23-3 sense: 5′-CCUCAUGCCAUCACAGUAUTT-3′,
antisense: 5′- AUACUGUGAUGGCAUGAGGTT-3′;
si-N.C sense: 5′-UUCUCCGAACGUGUCACGUTT-3′,
antisense: 5′-ACGUGACACGUUCGGAGAATT-3′.

Lipofectamine 2000 transfection reagent (Invitrogen, Carlsbad, CA, United States) was used to transfect cells following the manufacturer's instructions.

Cell Proliferation Assay

2×10^3 ccRCC cells were cultured in a 96-well plate, and cell proliferation changes were assessed by a Cell Counting Kit 8 (CCK-8) (Beyotime, Shanghai, China) at 24, 48, and 72 h after si-KIF23 transfection. Ethynyl deoxyuridine (EdU) kit (Ribobio, Guangzhou, China) was used to perform EdU assay following the manufacturer's instructions.

Transwell Migration and Invasion Assays

100 μl of serum-free medium with 2×10^4 ccRCC cells was added into the upper chamber of a transwell migration chamber (8-μm pore size, Costar, New York). The lower chamber was filled with medium containing 20% FBS. After 24 h, cells were fixed with 4% PFA and stained with crystal violet. Cells were counted under a microscope in five random fields after removal of cells on the upper surface. For invasion assay, upper chamber was coated with Matrigel (354480, Corning) before seeded cells.

Wound-Healing Assay

The cells with different treatments were seeded in six-well plates, starved for 24 h, and wounded with a pipette tip. Then replace with fresh medium. After 24 h, the wound closure process was observed and photographed.

Western Blot and Antibodies

RIPA buffer (Beyotime) was used for cell and tissue lysis. 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was used for protein separation. Then, proteins were transferred onto polyvinylidene difluoride (PVDF) membrane (Millipore, Billerica, MA, United States) by electroblotting for 90 min. After blocked in 5% non-fat dry milk for 2 h, the PVDF membranes incubated with primary antibodies at 4°C overnight, followed by the secondary antibody for at least 1 h. The blots were then visualized using a Western Blotting Luminol Reagent. The primary antibodies of KIF23 (1:1000, Proteintech, 28587-1-AP), β-catenin (1:1000, Cell Signaling Technology, 8480), GSK-3β (1:1000, Cell Signaling Technology, 12456), p-GSK-3β (1:1000, Cell Signaling Technology, 5558), c-Myc (1:1000, Cell Signaling Technology, 18583) and GAPDH (1:2000, Proteintech, 10494-1-AP). Cytoplasmic protein extraction kit (Beyotime) was used to extract nuclear proteins following the manufacturer's manual.

Immunohistochemistry and Immunofluorescence Analysis

ccRCC tissue sections were deparaffinized and the antigen was retrieved, followed by blocking with 5% bovine serum albumin. Next, ccRCC tissue sections were incubated with KIF23 primary antibody (1:100, Proteintech, 28587-1-AP) at 4°C overnight, a peroxidase-conjugated secondary antibody was used to detect antigen-antibody complexes. Thereafter, a color reaction was conducted using a 3,3′-Diaminobenzidine (DAB) substrate kit (ZsBio). For immunofluorescence staining, ccRCC cells were seeded on coverslips and fixed in 4% PFA. Then, cells were washed and treated with 0.25% Triton X-100 for 15 min. After blocked by 5% donkey serum at room temperature for 30 min, cells were incubated with β-catenin (1:100, Cell Signaling Technology, 8480) at 4°C overnight. Alexa Fluor 594 conjugated secondary antibody (1:500; Yeasen) was used to detect antigen-antibody complexes. 4′,6-Diamidino-2-phenylindole (DAPI) was used for the staining of nucleus.

Statistical Analysis

R (version 3.6) and GraphPad Prism software were used to execute statistical analysis. Differences between groups were statistically compared using the student's t test or two-way ANOVA. Each experiment was performed in triplicate, and values were expressed as mean ± standard deviation. A p-value <0.05 was regarded statistically significant. Survival curves were obtained by the Kaplan-Meier method and compared by log-rank tests. And levels of significance were presented at $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$ and $^{****}p < 0.0001$ respectively.

RESULT

Identification of Metastasis-Associated Genes

DEGs between metastatic and primary tumors were analyzed by the GEO2R tool with FC > 1.5 and $p < 0.05$ as screening criteria. 693 genes (473 up and 220 down) were identified as DEGs in GSE85258.497 genes (307 up and 190 down) were identified as DEGs in GSE105288 (**Figures 1A–D**). Venn diagrams were used to integrate GSE85258 and GSE105288 datasets, and 74 DEGs were identified as metastasis-related genes (**Figure 2A**). To validate these metastasis-related genes, the ccRCC dataset from TCGA was downloaded and analyzed. The expression levels of these DEGs were visualized using a heatmap (**Figure 2B**). Univariate Cox proportional hazards regression analysis revealed 54 genes significantly associated with OS in ccRCC patients (**Figure 2C**). Forty seven genes with a hazard ratio (HR) greater than one were identified as "high risk factors", and seven genes with HR < 1 were considered protective factors.

Construction of a Prognostic Signature

It is difficult for a single metastasis-associated gene to accurately predict patient prognosis. Thus, development of a multi-gene combined prediction model is essential. To construct the prognostic model, the metastasis-associated genes were used to LASSO Cox regression analysis, and the regression coefficients were calculated as shown in **Figure 3A**. The best performing prognostic model was obtained when 12 genes were included (**Figure 3B**). The risk score for each patient is calculated from the expression levels of 12 genes: risk score = (−0.0072 * PDGFRL) + (0.0012 * C1R) + (0.1278 * TRIP13) + (0.0004 * TIMP1) + (0.0016 * C1S) + (−0.0072 *

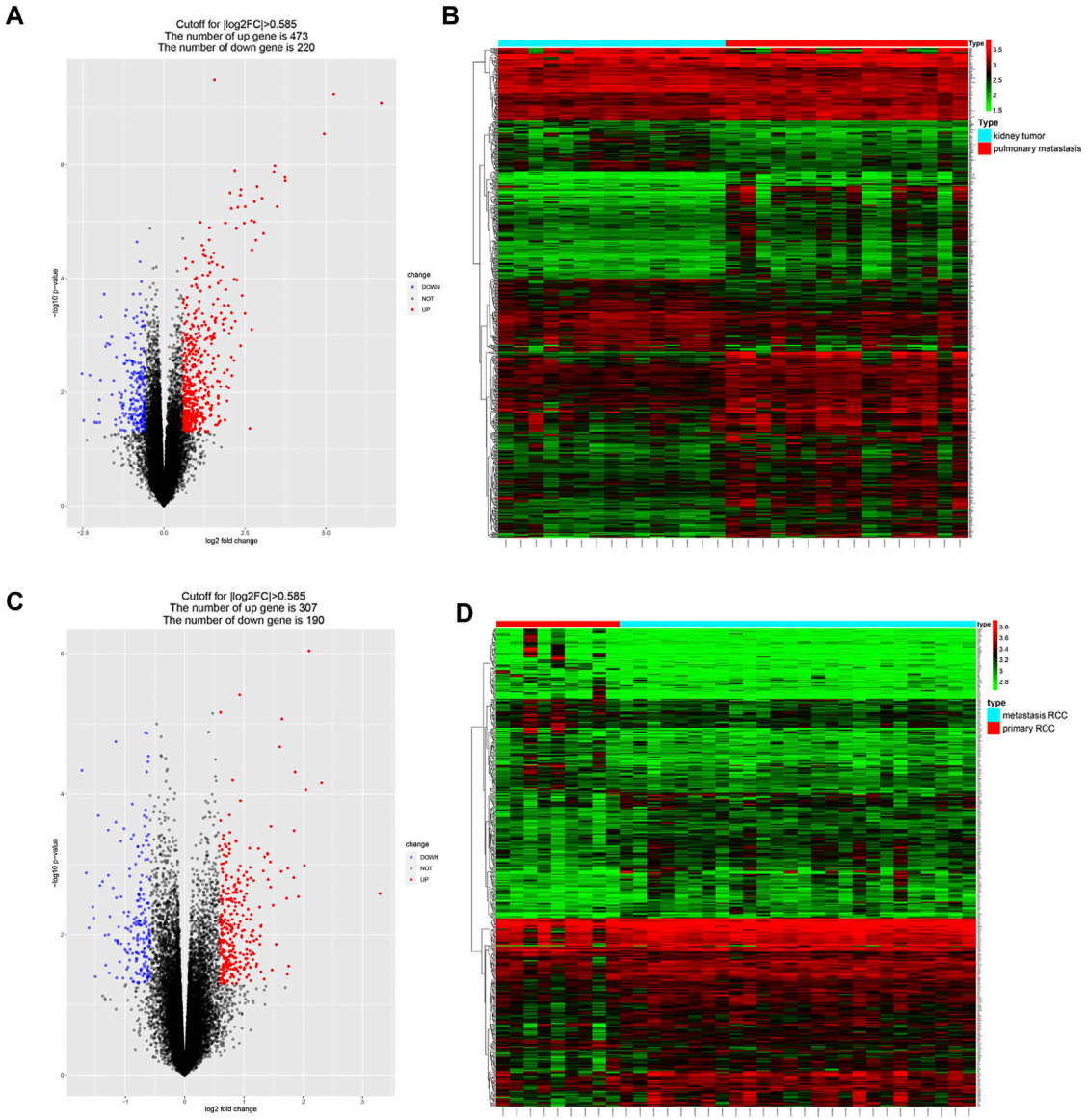


FIGURE 1 | Identification of metastasis-associated genes in ccRCC. **(A–D)** Volcano plots and heatmaps of the DEGs in GSE85258 and GSE105288.

NAPSA) + (0.0011 * SERPINF1) + (−0.06808 * CGN) + (0.07435 * ASPM) + (0.02810 * CEP55) + (0.00039 * COL5A1) + (0.01375 * SCG5). Then, patients were divided into low-risk ($n = 263$) and high-risk ($n = 262$) groups based on their risk scores (**Figure 3C**). Patients in high-risk group have a worse prognosis (**Figure 3D**). The area under curve (AUC) value for metastasis-associated prognostic gene signature was 0.7 and 0.661 for 3- and 5-year survival rates, respectively (**Figures 3E,F**).

Identification of Immune Cell Infiltration Characteristics of Each Group

In this study, CIBERSORT algorithm was used to estimate the proportion of infiltrating immune cells in tumor tissue. Immune cell infiltration signatures of the low-risk and the high-risk groups are showed in **Figures 4A,B**. There were higher proportions of resting memory CD4 T cells, activated natural killer (NK) cells, monocytes, resting dendritic cells (DCs) and resting mast cells in the low-risk group. Plasma cells, regulatory T cells, and M0 macrophages were higher in the high-risk group (**Figure 4C**).

Differences in Gene Expression Between the Low- and High-Risk Groups

Since the two groups showed variations in gene expression profiles and prognosis, gene expression patterns between the groups were analyzed and 588 DEGs were identified (408 and 181 genes were up-regulated in the high- and low-risk groups, respectively, **Figures 5A,B**).

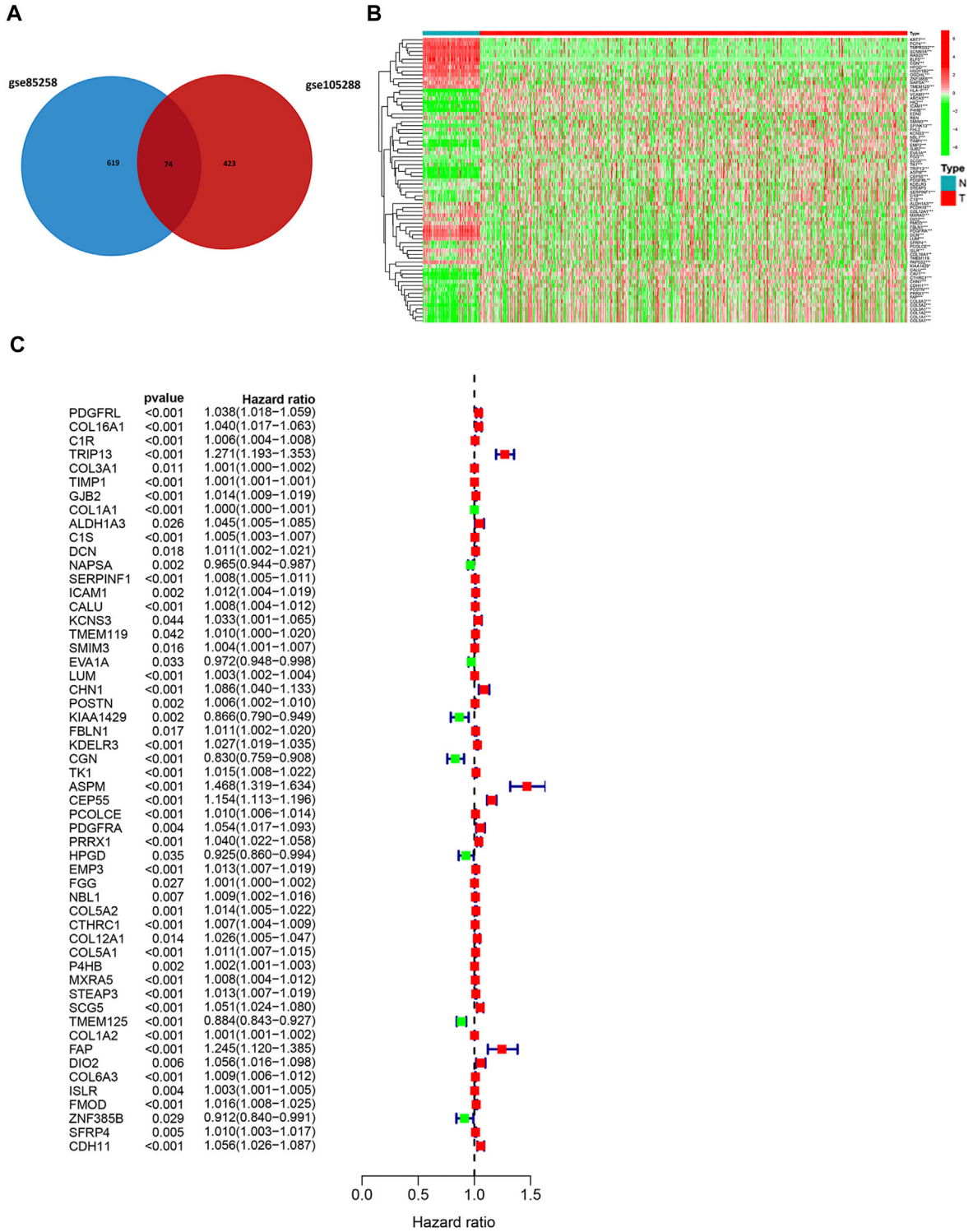


FIGURE 2 | Expression profile of metastasis-associated genes in TCGA dataset. **(A)** Venn diagram of metastasis-associated DEGs. **(B)** Heatmap of metastasis-associated genes in TCGA ccRCC dataset. **(C)** Univariate Cox regression analysis.

Function Enrichment Analysis

GO enrichment analysis showed that DEGs were involved in "extracellular matrix organization", "extracellular structure organization" and "positive regulation of secretion by cell" in the "biological process" category. "Collagen-containing extracellular matrix" was enriched in the "cellular components" category and

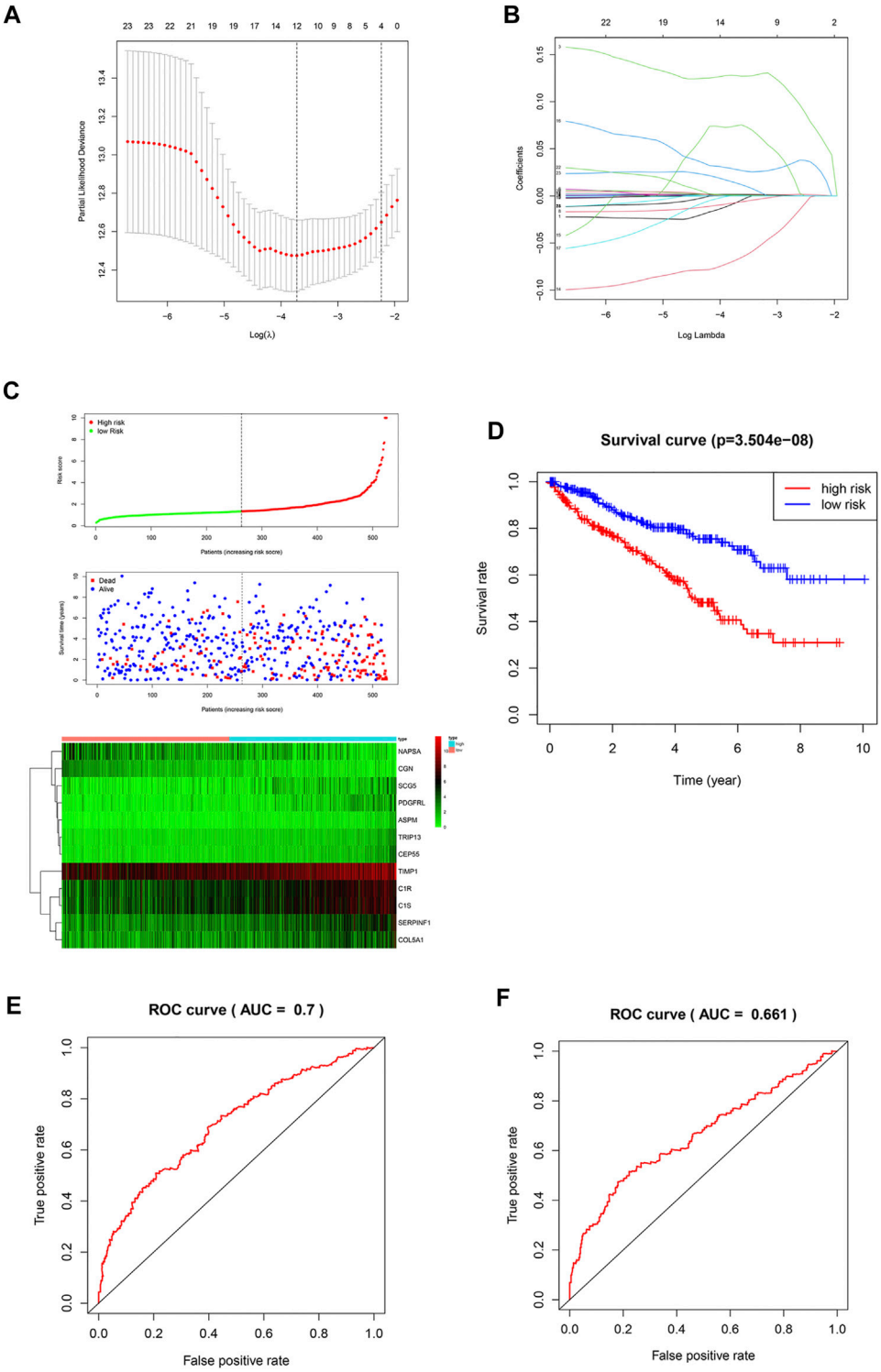


FIGURE 3 | Construction of a prognostic signature. **(A,B)** LASSO analysis with a minimum lambda value. **(C)** Risk score, patient survival status distribution and expression heatmaps for 12 metastasis-associated prognostic risk signatures. **(D)** Kaplan-Meier curves of the low- and high-risk groups ($p < 0.001$). **(E,F)** ROC curve for 3- and 5-year survival predictions in patients with ccRCC based on the risk score.

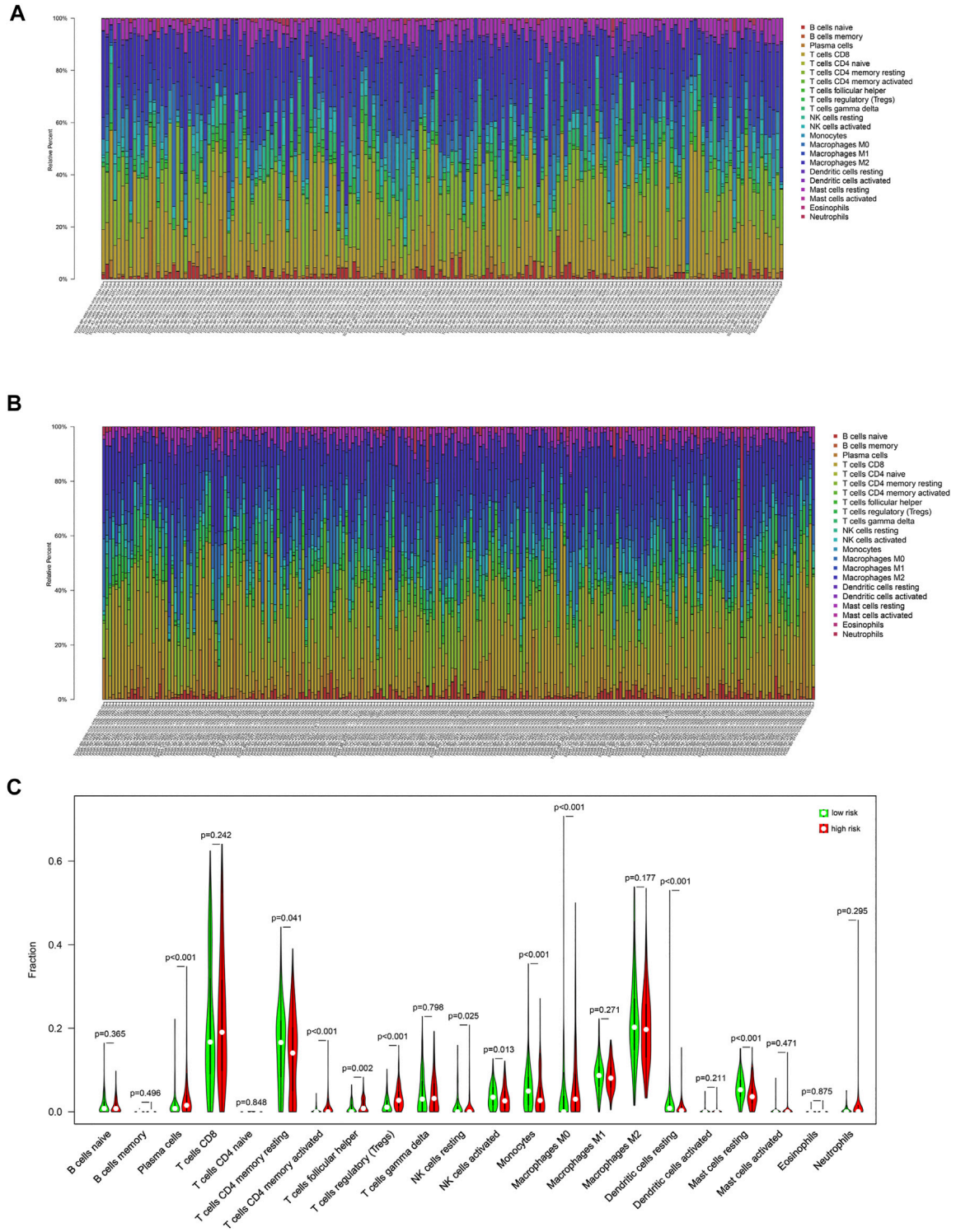


FIGURE 4 | Immune cell infiltration signatures of each group. **(A)** The immune cell infiltration signatures in the low-risk group. **(B)** The immune cell infiltration signatures in the high-risk group. **(C)** Differences in immune cell infiltration characteristics between low-risk and high-risk populations.

"extracellular matrix structural constituent" was significantly enriched in the "molecular function" category. KEGG pathway enrichment analysis results indicated that "cytokine-cytokine receptor interaction" and "NF-kappa B signaling pathway" are significantly enriched (**Figures 5C,D**). Next, we constructed PPI network, and cytoHubba was employed to select the top 10 hub

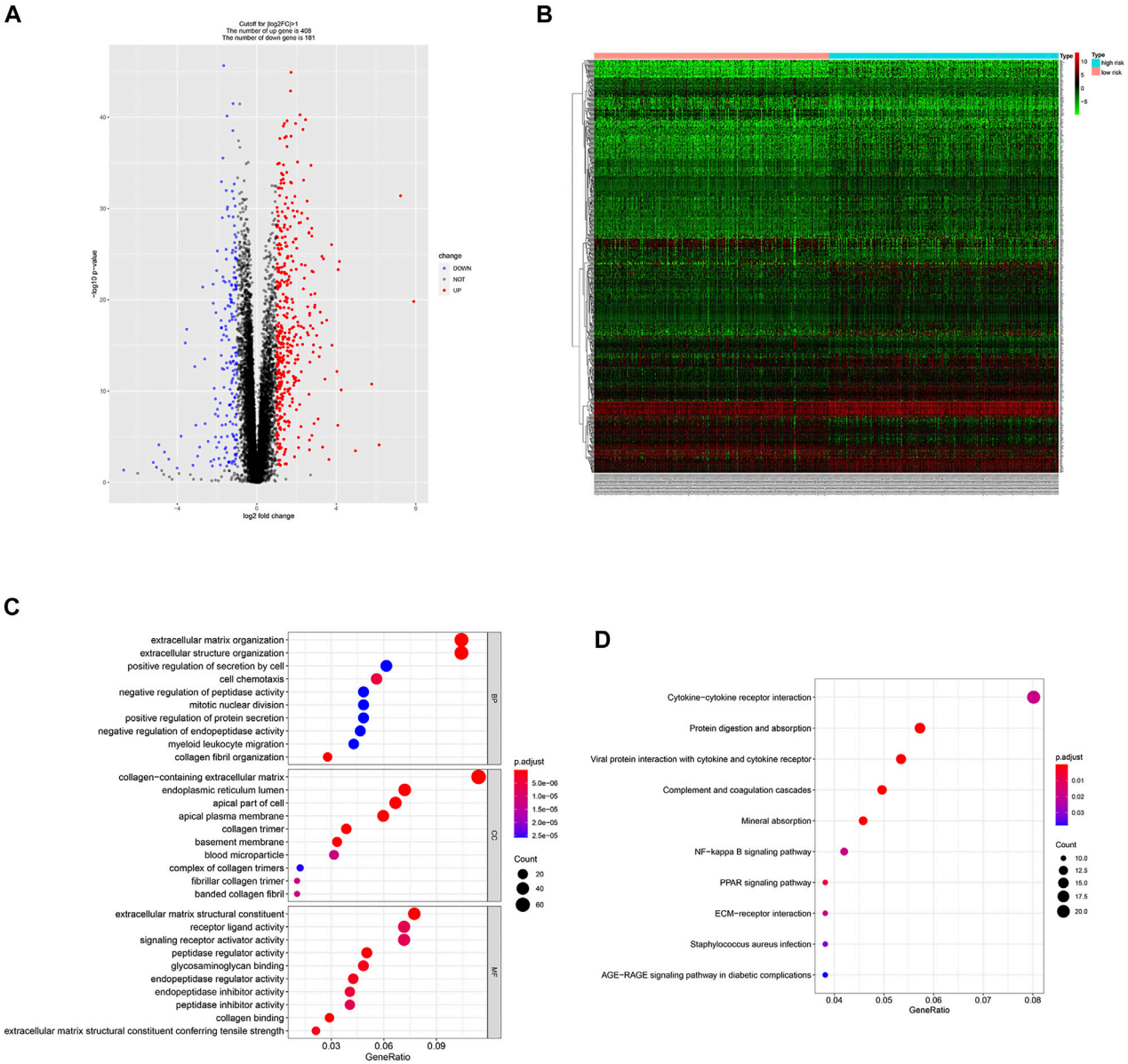


FIGURE 5 | DEGs identification and functional enrichment analysis. **(A,B)** Volcano plot and heatmap of 588 DEGs between low- and high-risk group. **(C,D)** Function enrichment analysis of DEGs.

nodes, including PLK1, DLGAP5, KIF23, BIRC5, AURKB, CDCA8, RRM2, BUB1B, UBE2C, and CDC20 (**Figure 6**). KIF23, a member of kinesin proteins, and its role in tumor development has been demonstrated in many tumors (Gao et al., 2020; Hu et al., 2020), yet its role in ccRCC remains elusive. Therefore, we explored the expression and function of KIF23 in ccRCC.

Kinesin Family Member 23 was Significantly Elevated in Clear Cell Renal Cell Carcinoma

The TCGA database showed that KIF23 mRNA was significantly elevated in ccRCC tissues (**Figure 7A**). And patients with high expression of KIF23 showed worse OS rates (**Figure 7B**). Univariate analysis showed no significant difference in survival based on gender. KIF23 expression level, age, tumor stage and grade were associated with OS in ccRCC patients (**Supplementary Table S1**). Multivariate Cox risk regression analysis showed that KIF23 was an independent factor of OS (HR = 1.919, 95% confidence interval, 1.439–2.560, $p < 0.001$) (**Figure 7C**, **Supplementary Table S2**). Moreover, results from GSEA analysis showed that cell adhesion molecules (CAMs), DNA replication, cell cycle and Wnt signaling pathway were significantly enriched in the KIF23 high-expression group (**Figure 7D**). As anticipated, the protein level of KIF23 was also

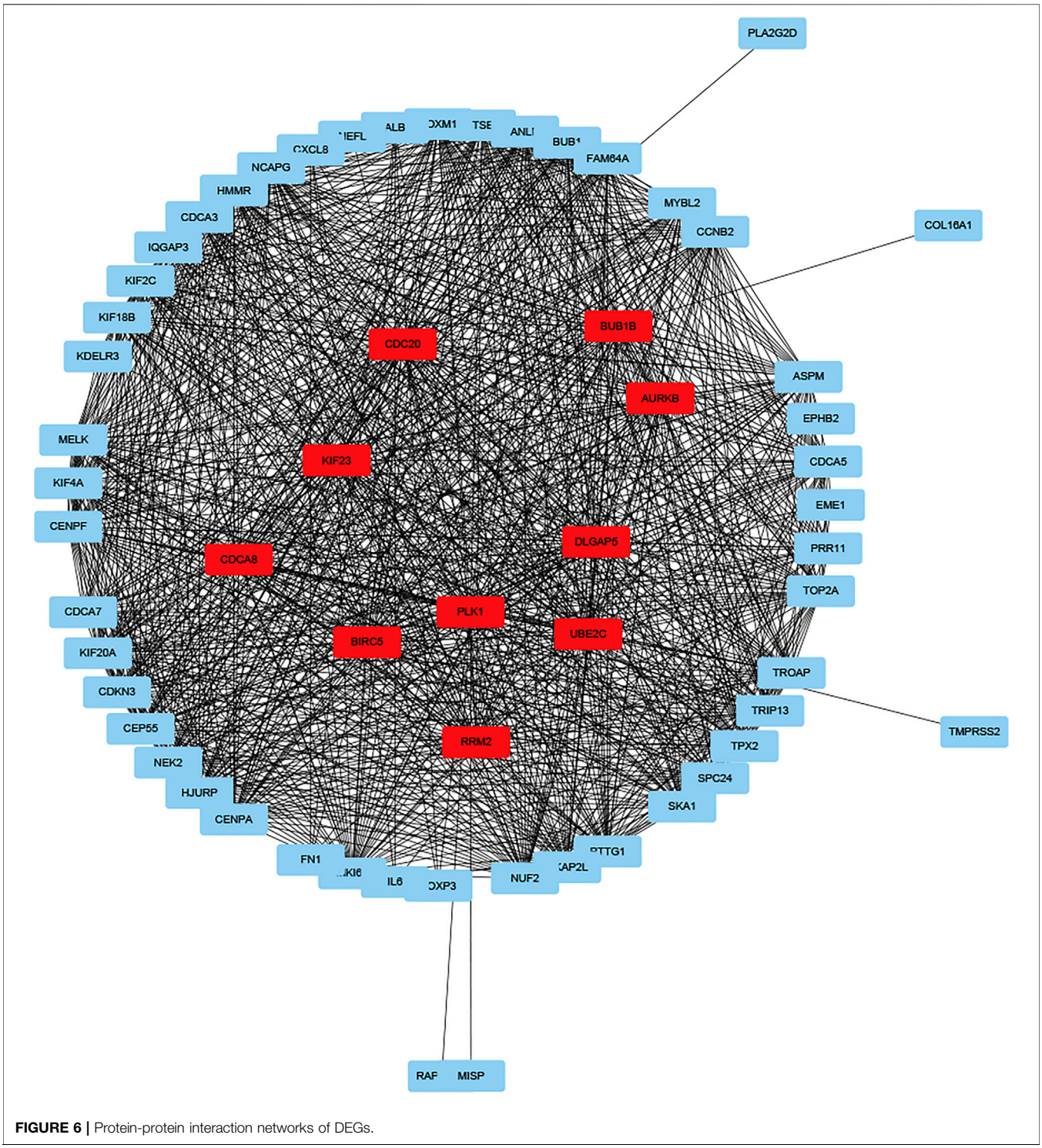


FIGURE 6 | Protein-protein interaction networks of DEGs.

increased in ccRCC tumor tissues (**Figure 8A**). The protein expression level of KIF23 was detected in HK-2 cells and five human ccRCC cell lines. As showed in **Figure 8B**, KIF23 was increased in ccRCC cell lines compared with HK-2 cells. Moreover, the results of immunohistochemistry staining showed that KIF23 expression were higher in tumor specimens (**Figure 8C**).

Kinesin Family Member 23 Promotes the Malignant Behavior of Clear Cell Renal Cell Carcinoma Cells

We further explored the role of KIF23 on the proliferation, migration and invasion of ccRCC cells. Silencing of KIF23 expression using siRNA resulted in down-regulation of KIF23 protein in 786-O and

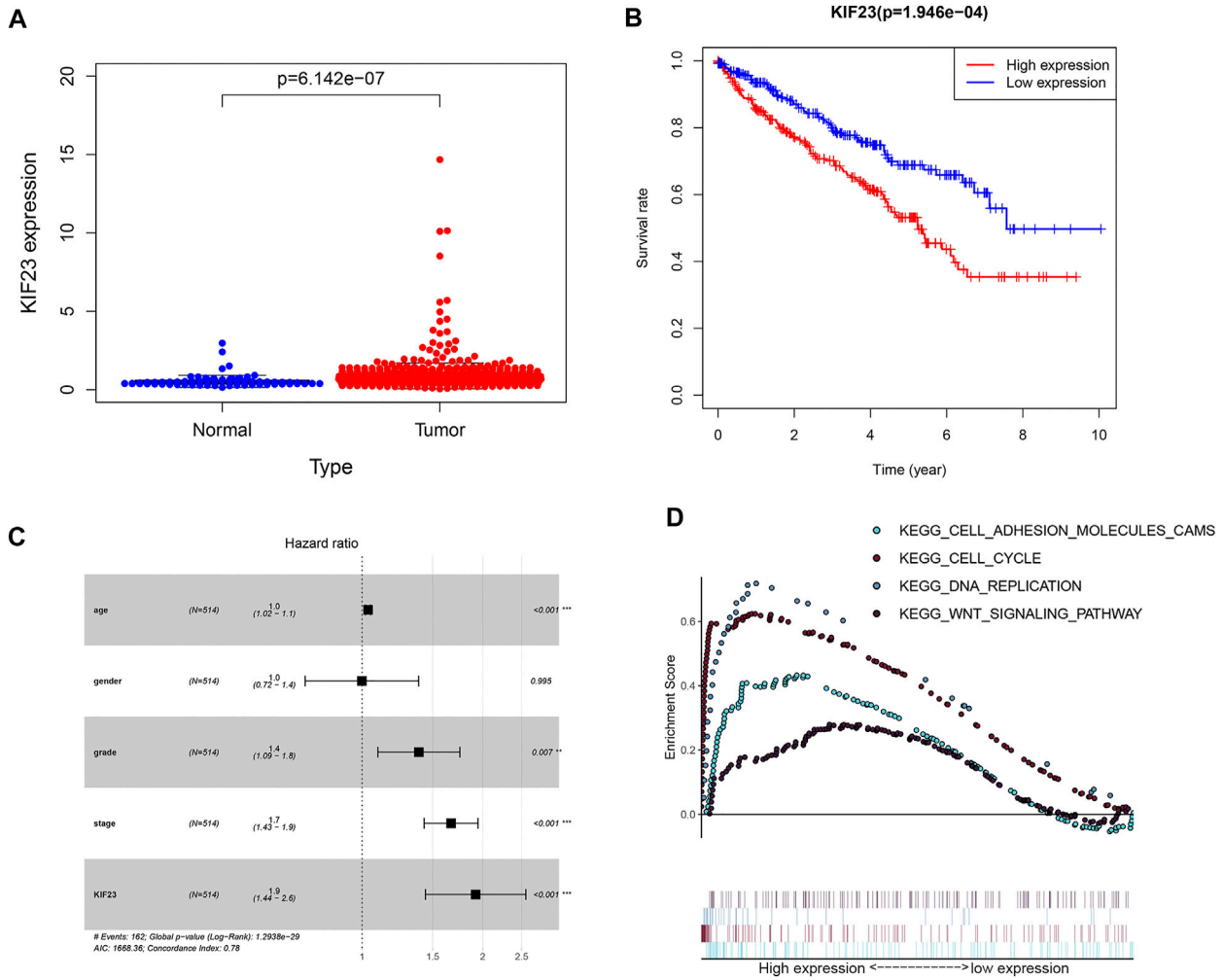


FIGURE 7 | The expression data of KIF23 in ccRCC from TCGA database. **(A)** Human KIF23 expression levels in normal and ccRCC tissues from the TCGA database. **(B)** Kaplan-Meier plotters of KIF23 in ccRCC. **(C)** Multivariate Cox hazards regression analyses. **(D)** GSEA analysis.

A498 (**Figure 9A**). Compared with the negative control (NC) group, knockdown of KIF23 expression inhibited the proliferation of ccRCC cells as measured by CCK-8 and EdU assays (**Figures 9B,C**). The migration and invasion ability were further measured using wound-healing assay and transwell assays. The results showed that interfering with KIF23 expression reduced the migration and invasion ability of ccRCC cells (**Figures 9D,E**). To further elucidate the mechanisms underlying KIF23-promoted ccRCC proliferation and migration, GSEA analysis was used to predict the downstream pathways, and Wnt signaling pathways were enriched in the KIF23 high-expression group (**Figure 7D**). We found that KIF23 knockdown decreased the expression levels of molecules on the Wnt signaling pathway include p-GSK3-β, β-catenin and c-Myc (**Figure 10A**). Moreover, analysis of subcellular protein fraction showed that knockdown of KIF23 reduced nuclear translocation of β-catenin (**Figure 10B**). Immunofluorescence staining was used to assess β-catenin translocation, and it was found that knockdown of KIF23 expression reduced nuclear localization of β-catenin in 786-O and A498 cells (**Figures 10C,D**).

DISCUSSION

Advanced RCC has a poor prognosis due to its resistance to conventional radiation therapy and chemotherapy (Santoni et al., 2013). With the increasing understanding of RCC biology, many therapies have been introduced. Tyrosine kinase monoclonal antibodies or inhibitors targeting vascular endothelial growth factor (VEGF) have showed a better therapeutic effect on RCC (Motzer and Molina, 2009). mTOR inhibitors also have a remarkable survival advantage (Hudes et al., 2007). Despite the use of novel targeted therapeutic agents, metastatic RCC will eventually progress owing to primary or secondary drug resistance (Iacovelli et al., 2014). Metastasis is the most important reason affecting patient prognosis in ccRCC patients, and analyzing transcriptome differences of metastatic tumors will

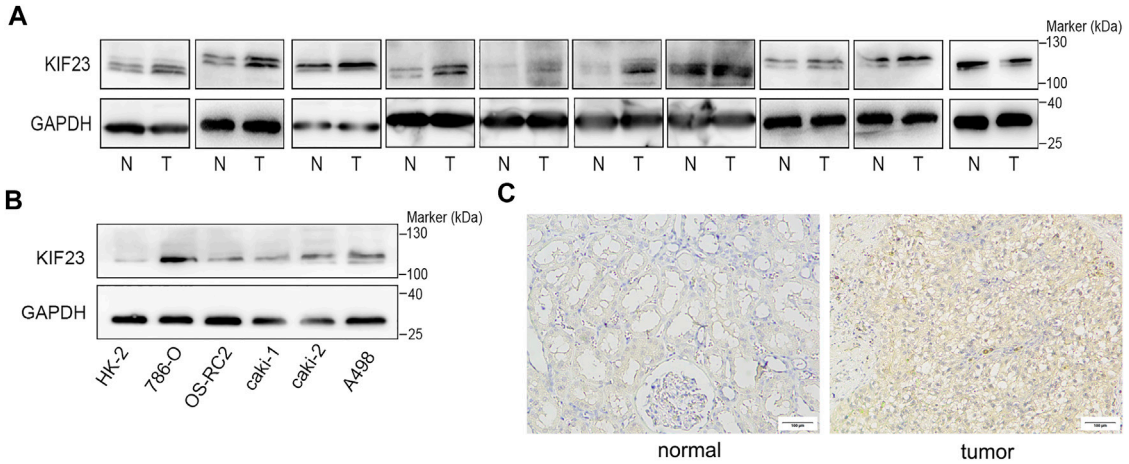


FIGURE 8 | KIF23 is overexpressed in ccRCC cell lines and tissues. **(A)** The expression level of KIF23 in 10 paired ccRCC samples. **(B)** KIF23 protein levels in ccRCC cell lines (786-O, OS-RC-2, caki1, caki2, and A498) and HK-2 were assessed using Western blot assays. **(C)** Immunohistochemistry staining of KIF23 expression level in ccRCC samples.

help to identify factors affecting metastasis and predict clinical outcomes. The present study sought to address these gaps in knowledge by comparing transcriptome profiles between metastatic and primary tumors, and 74 genes were identified as metastasis-associated genes. Furthermore, 54 genes associated with the prognosis of ccRCC patients were screened.

In a prognostic model of metastasis-related genes based on LASSO Cox regression analysis, we found a shorter survival time in the high-risk score group. Since the tumor immune microenvironment has a non-negligible impact on tumor progression, we analyzed differences in the immune components of the high- and low-risk groups. ccRCC in the low-risk group was associated with higher proportions of resting memory CD4 T cells, activated NK cells, monocytes, resting DCs, and resting mast cells. This is similar to previous studies showing that accumulation of CD8 T lymphocytes and NK cells in tumors implies favorable clinical outcomes (Fridman et al., 2012). DCs are antigen-presenting cells that play an important role in promoting antitumor immune responses. Current studies suggest that DCs infiltration is an important prognostic parameter, which usually depends on cellular localization and cellular maturation (Eisenthal et al., 2001; Sandel et al., 2005). Overall, our results indicate that there are different immunobiological processes and pathways between the two groups. Immune heterogeneity between two groups may be an important factor driving the differences observed in OS.

Gene expression profiles also differed between the two groups, which may affect ccRCC progression, response to treatment and prognosis. DEGs were associated with different biological processes. GO enrichment analysis indicated that DEGs were enriched in "extracellular matrix organization", "extracellular structure organization" and "positive regulation of secretion by cell" in the "biological process" category, Previous studies found that extracellular matrix (ECM) is a complex network providing structural and guiding clues to surrounding cells (Naba et al., 2017). The ECM triggers tumor progression by affecting cell proliferation and metastasis. In addition, abnormal ECM also affects stromal cell behavior, such as inflammation and angiogenesis, which promote the formation of tumorigenic microenvironment (Cukierman and Bassi, 2010; Pickup et al., 2014). ECM protein is also considered to be an important part of the metastatic niche, which can maintain the characteristics of cancer stem cells and promote the growth of metastatic cells (Malanchi et al., 2011; Oskarsson et al., 2011). Tumor ECM induced tumor cell metastasis through extracellular proteases such as matrix metalloproteinases (MMPs) (Wang et al., 2021). PLK1, DLGAP5, KIF23, BIRC5, AURKB, CDCA8, RRM2, BUB1B, UBE2C, and CDC20 were identified as the hub genes. Multiple members of the KIF family have been shown to promote ccRCC metastasis, invasion and drug resistance (Li et al., 2019a; Jin et al., 2019; Ren et al., 2020). As a member of KIF family, KIF23 promotes ovarian tumor proliferation, cell cycle progression and is closely related to metastasis. In glioma, silencing KIF23 inhibits the proliferation of glioma cells (Takahashi et al., 2012; Li et al., 2019b). In ccRCC, we found the expression of KIF23 was significantly elevated in tumor tissues. Its expression level correlates with patient prognosis. In ccRCC tumor biology, knocking down KIF23 expression in ccRCC cells inhibited the malignant behavior of ccRCC cells. Therefore, KIF23 may be an important molecular marker promoting the progression of ccRCC.

Abnormal Wnt signaling is involved in many human tumors (Nwabo Kamdje et al., 2014; Cojocaru et al., 2015; Xu et al., 2016). β-catenin is an executor of the Wnt signaling. β-catenin is phosphorylated by a "degradation complex" in the absence of Wnt ligands, phosphorylated β-catenin will be further degraded by ubiquitination to inhibit the activation of Wnt signaling (Bhanot et al., 1996). When the Wnt signaling is activated, the binding of Wnt protein phosphorylate and inactivate GSK3-β in cells, inhibit the formation of degradation complexes and cause accumulation of β-catenin and promote its nucleus translocation (Behrens et al., 1998). β-catenin forms the β-catenin-TCF4 complex in the nucleus and activates the transcription of multiple oncogenes including c-Myc and cyclin

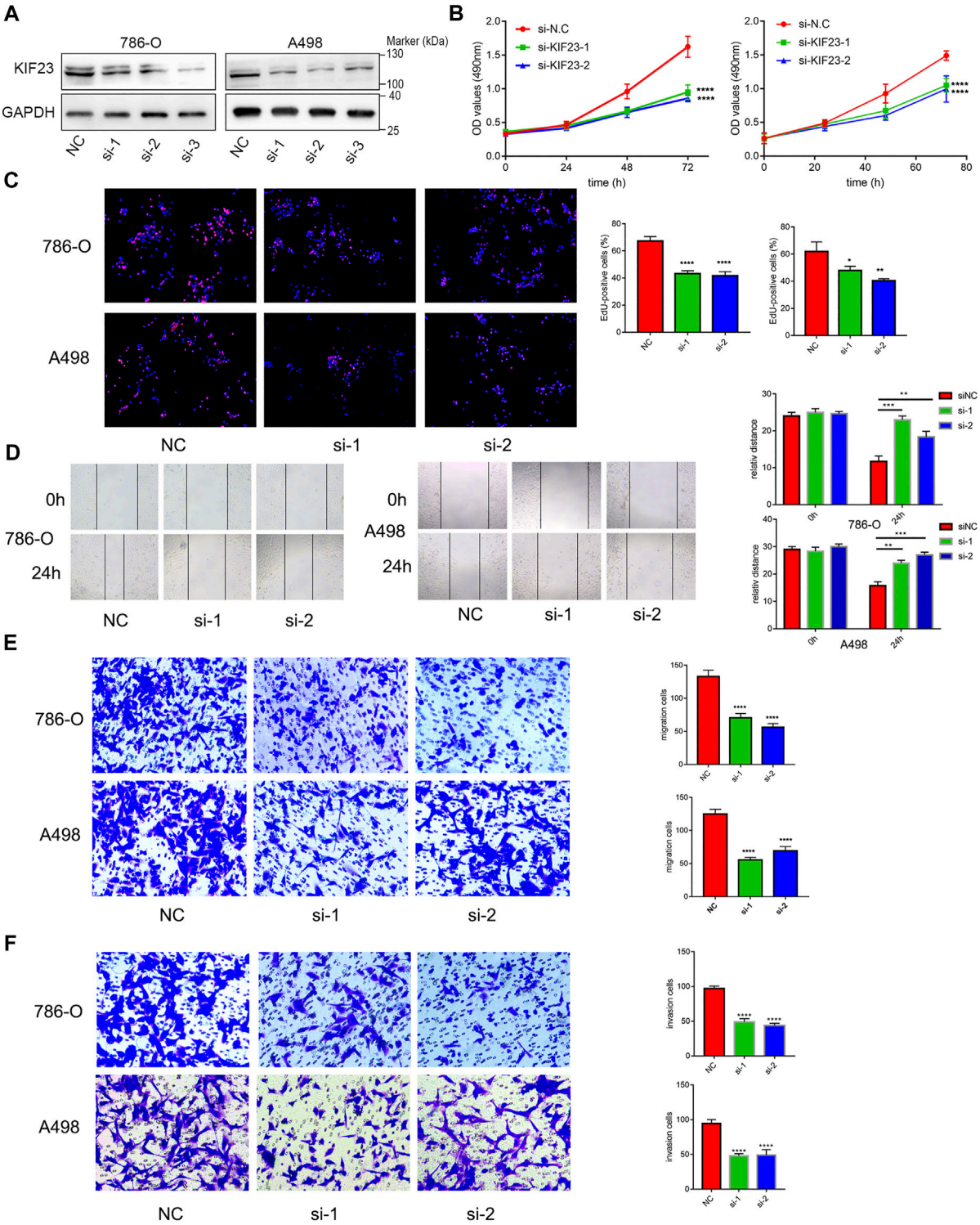


FIGURE 9 | KIF23 promotes proliferation, migration and invasion of ccRCC cells. **(A)** Protein expression levels of KIF23 in 786-O and A498 cells following siRNA interference. **(B,C)** CCK-8, EdU assays. **(D–F)** Wound-healing, transwell migration and invasion assays.

D1 (Morin et al., 1997; Vallée et al., 2017). Our study found that knockdown of KIF23 expression in ccRCC inhibited the phosphorylation of GSK3-β, decreased the expression level of β-catenin and its nuclear translocation and thus inhibited the expression of c-Myc, thereby regulating the malignant behavior of ccRCC cells.

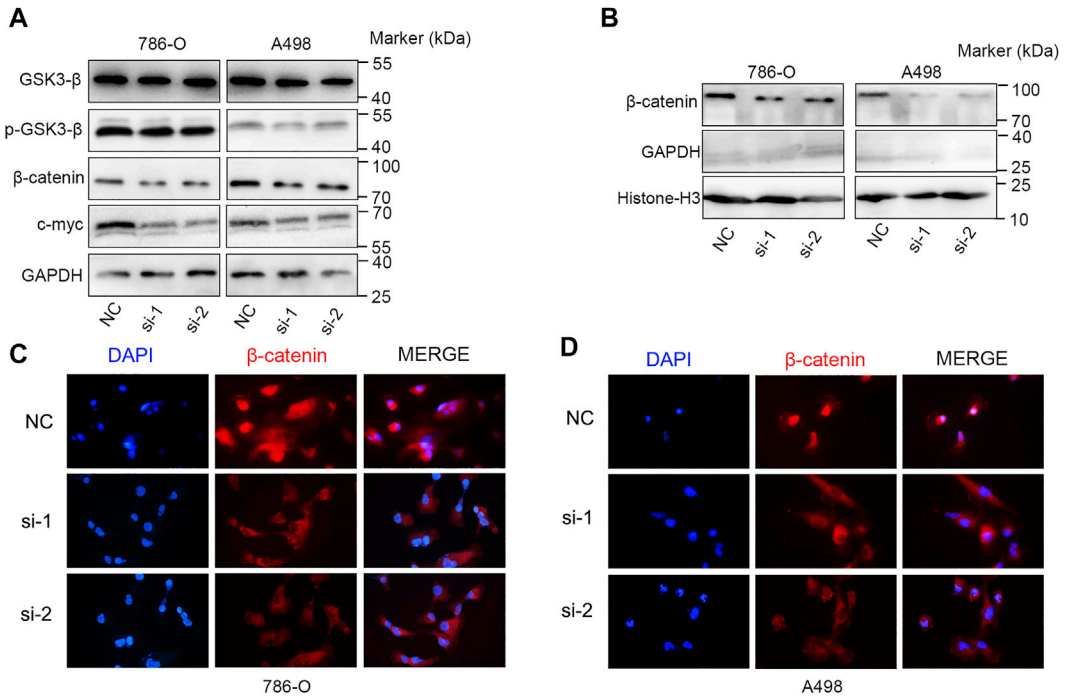


FIGURE 10 | KIF23 knockdown inhibits the Wnt/β-catenin signaling pathway. **(A)** KIF23 knockdown reduced the levels of p-GSK3-β, β-catenin and c-Myc. **(B)** KIF23 knockdown reduced β-catenin levels in the nucleus. **(C,D)** Immunofluorescence staining of β-catenin in 786-O and A498 cells.

In conclusion, based on the construction of prognostic signatures of metastasis-related genes, we identified KIF23 as an important gene that drives ccRCC progression and affects patient prognosis. By knocking down KIF23 expression, we found that KIF23 enhanced the proliferation, migration and invasion capacity of ccRCC cells by regulating the nuclear translocation of β-catenin. Inhibition of KIF23 expression or function is a potential therapeutic approach for ccRCC.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The microarray datasets GSE85258 and GSE105288 were obtained from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database. The gene expression profiles of 530 ccRCC were downloaded from the TCGA data portal (https://tcga-data.nci.nih.gov/tcga/).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Peking University Third Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ZW designed the research, analyzed the data, and wrote the manuscript. YS and YW contributed to the conduction of the experiments. LG, ZL, and TD performed the data analysis and interpretation of the data. SZ and LM assisted with the study design and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Natural Science Foundation of China (Grant Nos 81972381 and 82072828).

ACKNOWLEDGMENTS

We acknowledge TCGA and GEO database for providing their platforms and contributors for uploading their meaningful datasets.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2022.839821/full#supplementary-material

REFERENCES

Behrens, J., Jerchow, B.-A., Würtele, M., Grimm, J., Asbrand, C., Wirtz, R., et al. (1998). Functional Interaction of an Axin Homolog, Conductin, with β-Catenin, APC, and GSK3β. *Science* 280, 596–599. doi:10.1126/science.280.5363.596

Bhanot, P., Brink, M., Samos, C. H., Hsieh, J.-C., Wang, Y., Macke, J. P., et al. (1996). A New Member of the Frizzled Family from Drosophila Functions as a Wingless Receptor. *Nature* 382, 225–230. doi:10.1038/382225a0

Casuscelli, J., Becerra, M. F., Manley, B. J., Zabor, E. C., Reznik, E., Redzematovic, A., et al. (2019). Characterization and Impact of TERT Promoter Region Mutations on Clinical Outcome in Renal Cell Carcinoma. *Eur. Urol. Focus* 5, 642–649. doi:10.1016/j.euf.2017.09.008

Choueiri, T. K., and Motzer, R. J. (2017). Systemic Therapy for Metastatic Renal-Cell Carcinoma. *N. Engl. J. Med.* 376, 354–366. doi:10.1056/NEJMra1601333

Cojocaru, E., Lozneanu, L., Giuşcă, S. E., Căruntu, I. D., and Danciu, M. (2015). Renal Carcinogenesis-Iinsights into Signaling Pathways. *Rom. J. Morphol. Embryol.* 56, 15–19.

Cukierman, E., and Bassi, D. E. (2010). Physico-mechanical Aspects of Extracellular Matrix Influences on Tumorigenic Behaviors. *Semin. Cancer Biol.* 20, 139–145. doi:10.1016/j.semcancer.2010.04.004

Eisenthal, A., Polyvkin, N., Bramante–Schreiber, L., Misonznik, F., Hassner, A., and Lifschitz–Mercer, B. (2001). Expression of Dendritic Cells in Ovarian Tumors Correlates with Clinical Outcome in Patients with Ovarian Cancer. *Hum. Pathol.* 32, 803–807. doi:10.1053/hupa.2001.26455

Escudier, B., Motzer, R. J., Tannir, N. M., Porta, C., Tomita, Y., Maurer, M. A., et al. (2020). Efficacy of Nivolumab Plus Ipilimumab According to Number of IMDC Risk Factors in CheckMate 214. *Eur. Urol.* 77, 449–453. doi:10.1016/j.eururo.2019.10.025

Fridman, W. H., Pagès, F., Sautès-Fridman, C., and Galon, J. (2012). The Immune Contexture in Human Tumours: Impact on Clinical Outcome. *Nat. Rev. Cancer* 12, 298–306. doi:10.1038/nrc3245

Gao, C.-T., Ren, J., Yu, J., Li, S.-N., Guo, X.-F., and Zhou, Y.-Z. (2020). KIF23 Enhances Cell Proliferation in Pancreatic Ductal Adenocarcinoma and Is a Potent Therapeutic Target. *Ann. Transl Med.* 8, 1394. doi:10.21037/atm-20-1970

Hu, Y., Zheng, M., Wang, C., Wang, S., Gou, R., Liu, O., et al. (2020). Identification of KIF23 as a Prognostic Signature for Ovarian Cancer Based on Large-Scale Sampling and Clinical Validation. *Am. J. Transl Res.* 12, 4955–4976.

Hudes, G., Carducci, M., Tomczak, P., Dutcher, J., Figlin, R., Kapoor, A., et al. (2007). Temsirolimus, Interferon Alfa, or Both for Advanced Renal-Cell Carcinoma. *N. Engl. J. Med.* 356, 2271–2281. doi:10.1056/NEJMoa066838

Iacovelli, R., Alesini, D., Palazzo, A., Trenta, P., Santoni, M., De Marchis, L., et al. (2014). Targeted Therapies and Complete Responses in First Line Treatment of Metastatic Renal Cell Carcinoma. A Meta-Analysis of Published Trials. *Cancer Treat. Rev.* 40, 271–275. doi:10.1016/j.ctrv.2013.09.003

Jin, Q., Dai, Y., Wang, Y., Zhang, S., and Liu, G. (2019). High Kinesin Family Member 11 Expression Predicts Poor Prognosis in Patients with clear Cell Renal Cell Carcinoma. *J. Clin. Pathol.* 72, 354–362. doi:10.1136/jclinpath-2018-205390

Li, G., Xie, Z. K., Zhu, D. S., Guo, T., Cai, Q. L., and Wang, Y. (2019a). KIF20B Promotes the Progression of clear Cell Renal Cell Carcinoma by Stimulating Cell Proliferation. *J. Cel. Physiol.* 234, 16517–16525. doi:10.1002/jcp.28322

Li, T., Li, Y., Gan, Y., Tian, R., Wu, Q., Shu, G., et al. (2019b). Methylation-mediated Repression of MiR-424/503 Cluster Promotes Proliferation and Migration of Ovarian Cancer Cells through Targeting the Hub Gene KIF23. *Cell Cycle* 18, 1601–1618. doi:10.1080/15384101.2019.1624112

Malanchi, I., Santamaria-Martínez, A., Susanto, E., Peng, H., Lehr, H.-A., Delaloye, J.-F., et al. (2011). Interactions between Cancer Stem Cells and Their Niche Govern Metastatic Colonization. *Nature* 481, 85–89. doi:10.1038/nature10694

Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B., et al. (1997). Activation of β-Catenin-Tcf Signaling in Colon Cancer by Mutations in β-Catenin or APC. *Science* 275, 1787–1790. doi:10.1126/science.275.5307.1787

Motzer, R. J., and Molina, A. M. (2009). Targeting Renal Cell Carcinoma. *Jco* 27, 3274–3276. doi:10.1200/JCO.2009.21.8461

Naba, A., Pearce, O. M. T., Del Rosario, A., Ma, D., Ding, H., Rajeeve, V., et al. (2017). Characterization of the Extracellular Matrix of Normal and Diseased Tissues Using Proteomics. *J. Proteome Res.* 16, 3083–3091. doi:10.1021/acs.jproteome.7b00191

Nwabo Kamdje, A. H., Seke Etet, P. F., Vecchio, L., Muller, J. M., Krampera, M., and Lukong, K. E. (2014). Signaling Pathways in Breast Cancer: Therapeutic Targeting of the Microenvironment. *Cell Signal.* 26, 2843–2856. doi:10.1016/j.cellsig.2014.07.034

Oskarsson, T., Acharyya, S., Zhang, X. H.-F., Vanharanta, S., Tavazoie, S. F., Morris, P. G., et al. (2011). Breast Cancer Cells Produce Tenascin C as a Metastatic Niche Component to Colonize the Lungs. *Nat. Med.* 17, 867–874. doi:10.1038/nm.2379

Park, W.-H., and Eisen, T. (2007). Prognostic Factors in Renal Cell Cancer. *BJU Int.* 99, 1277–1281. doi:10.1111/j.1464-410X.2007.06828.x

Pickup, M. W., Mouw, J. K., and Weaver, V. M. (2014). The Extracellular Matrix Modulates the Hallmarks of Cancer. *EMBO Rep.* 15, 1243–1253. doi:10.15252/embr.201439246

Ren, X., Chen, X., Ji, Y., Li, L., Li, Y., Qin, C., et al. (2020). Upregulation of KIF20A Promotes Tumor Proliferation and Invasion in Renal clear Cell Carcinoma and Is Associated with Adverse Clinical Outcome. *Aging* 12, 25878–25894. doi:10.18632/aging.202153

Rini, B. I., and Atkins, M. B. (2009). Resistance to Targeted Therapy in Renal-Cell Carcinoma. *Lancet Oncol.* 10, 992–1000. doi:10.1016/S1470-2045(09)70240-2

Sandel, M. H., Dadabayev, A. R., Menon, A. G., Morreau, H., Melief, C. J. M., Offringa, R., et al. (2005). Prognostic Value of Tumor-Infiltrating Dendritic Cells in Colorectal Cancer: Role of Maturation Status and Intratumoral Localization. *Clin. Cancer Res.* 11, 2576–2582. doi:10.1158/1078-0432.CCR-04-1448

Santoni, M., De Tursi, M., Felici, A., Lo Re, G., Ricotta, R., Ruggeri, E. M., et al. (2013). Management of Metastatic Renal Cell Carcinoma Patients with Poor-Risk Features: Current Status and Future Perspectives. *Expert Rev. Anticancer Ther.* 13, 697–709. doi:10.1586/era.13.52

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13, 2498–2504. doi:10.1101/gr.1239303

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING V10: Protein-Protein Interaction Networks, Integrated over the Tree of Life. *Nucleic Acids Res.* 43, D447–D452. doi:10.1093/nar/gku1003

Takahashi, S., Fusaki, N., Ohta, S., Iwahori, Y., Iizuka, Y., Inagawa, K., et al. (2012). Downregulation of KIF23 Suppresses Glioma Proliferation. *J. Neurooncol.* 106, 519–529. doi:10.1007/s11060-011-0706-2

Vallée, A., Lecarpentier, Y., Guillevin, R., and Vallée, J.-N. (2017). Thermodynamics in Gliomas: Interactions between the Canonical WNT/Beta-Catenin Pathway and PPAR Gamma. *Front. Physiol.* 8, 352. doi:10.3389/fphys.2017.00352

Wang, L., Li, H., Shi, L., Li, L., Jia, F., Gao, T., et al. (2022). *In Situ* peptide Self-Assembly on Ionic Nanochannel for Dynamic Monitoring of MMPs in Extracellular Matrix. *Biosens. Bioelectron.* 195, 113671. doi:10.1016/j.bios.2021.113671

Xu, Q., Krause, M., Samoylenko, A., and Vainio, S. (2016). Wnt Signaling in Renal Cell Carcinoma. *Cancers* 8, 57. doi:10.3390/cancers8060057

Yu, G., Wang, L.-G., Han, Q.-Y., and He, Q. Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16, 284–287. doi:10.1089/omi.2011.0118

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: 13 April 2022
doi: 10.3389/fcell.2022.854640

Detection of Structural Variations and Fusion Genes in Breast Cancer Samples Using Third-Generation Sequencing

Taobo Hu[1†], Jingjing Li[2,3†], Mengping Long[4†], Jinbo Wu[1†], Zhen Zhang[5], Fei Xie[1], Jin Zhao[1], Houpu Yang[1], Qianqian Song[6], Sheng Lian[7], Jiandong Shi[7], Xueyu Guo[3], Daoli Yuan[3], Dandan Lang[3], Guoliang Yu[3], Baosheng Liang[6], Xiaohua Zhou[6], Toyotaka Ishibashi[8], Xiaodan Fan[5], Weichuan Yu[7], Depeng Wang[3], Yang Wang[3], I-Feng Peng[3]* and Shu Wang[1]**

[1]Department of Breast Surgery, Peking University People's Hospital, Beijing, China, [2]State Key Laboratory of Genetic Engineering, School of Life Sciences and Human Phenome Institute, Fudan University, Shanghai, China, [3]GrandOmics Inc., Beijing, China, [4]Department of Pathology, Peking University Cancer Hospital, Beijing, China, [5]Department of Statistics, The Chinese University of Hong Kong, Sha Tin, China, [6]Department of Biostatistics, School of Public Health, Peking University, Beijing, China, [7]Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Kowloon, Hong Kong SAR, China, [8]Division of Life Science, Hong Kong University of Science and Technology, Kowloon, Hong Kong SAR, China

Background: Structural variations (SVs) are common genetic alterations in the human genome that could cause different phenotypes and diseases, including cancer. However, the detection of structural variations using the second-generation sequencing was limited by its short read length, which restrained our understanding of structural variations.

Methods: In this study, we developed a 28-gene panel for long-read sequencing and employed it to Oxford Nanopore Technologies and Pacific Biosciences platforms. We analyzed structural variations in the 28 breast cancer-related genes through long-read genomic and transcriptomic sequencing of tumor, para-tumor, and blood samples in 19 breast cancer patients.

Results: Our results showed that some somatic SVs were recurring among the selected genes, though the majority of them occurred in the non-exonic region. We found evidence supporting the existence of hotspot regions for SVs, which extended our previous understanding that they exist only for single nucleotide variations.

Conclusion: In conclusion, we employed long-read genomic and transcriptomic sequencing to identify SVs from breast cancer patients and proved that this approach holds great potential in clinical application.

Keywords: long-read sequencing, breast cancer, structural variation, fusion gene, sequencing panel

OPEN ACCESS

Edited by:
Geng Chen, GeneCast Biotechnology Co., Ltd., China

Reviewed by:
Andrew Carroll, Google, United States
Ting Li, National Center for Toxicological Research (FDA), United States

****Correspondence:***
Yang Wang
wangyang@grandomics.com
I-Feng Peng
ifengpeng@gmail.com
Shu Wang
shuwang@pkuph.edu.cn

[†]These authors 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: *14 January 2022*
Accepted: *23 March 2022*
Published: *13 April 2022*

Citation:
Hu T, Li J, Long M, Wu J, Zhang Z, Xie F, Zhao J, Yang H, Song Q, Lian S, Shi J, Guo X, Yuan D, Lang D, Yu G, Liang B, Zhou X, Ishibashi T, Fan X, Yu W, Wang D, Wang Y, Peng I-F and Wang S (2022) Detection of Structural Variations and Fusion Genes in Breast Cancer Samples Using Third-Generation Sequencing. Front. Cell Dev. Biol. 10:854640. doi: 10.3389/fcell.2022.854640

Abbreviations: CCS, Circular Consensus Sequence; DCIS, Ductal Carcinoma In Situ; HRR, Homologous Recombination Repair; ICGC, International Cancer Genome Consortium; NGS, next-generation sequencing; ONT, Oxford Nanopore Technologies; RIN, RNA Integrity Number; SINE, Short Interspersed Nuclear Element; SMRT, Single-Molecule Real-Time; SNV, Single Nucleotide Variation; SDS, Sodium Dodecyl Sulphate; SV, Structural Variation; TGS, Third-Generation Sequencing; TNBC, Triple Negative Breast Cancer; UTR, UnTranslated Region.

BACKGROUND

Breast cancer is the most common malignancy in women. Genome instability is a critical molecular characteristic of breast cancer, whereas structural variation directly manifests genome instability (Duijf et al., 2019). Structural variations (SVs), including insertion, deletion, duplication, inversion, and translocation, affect nucleotides on a much larger scale over single nucleotide variations (SNVs) (Sudmant et al., 2015). SVs are common variations in the general population, as shown by the 1,000 genome project (Iafrate et al., 2004; Sebat et al., 2004), where specific variations are known to be responsible for developing a number of genetic diseases and cancers (Feuk et al., 2006; Sharp et al., 2006; Alkan et al., 2011; Li et al., 2020). Previous studies of structural variation influence on gene structure and expression have significantly deepened our understanding of tumorigenesis (Hollox et al., 2021). Many oncogenes have been proven to be the products of chromosomal translocations and can be served as therapeutic targets. However, it remains challenging to identify SVs in the cancer genome due to the limitation of the next-generation sequencing (NGS), i.e., short read-length and sequence preference in PCR, which hinder NGS from detecting complex SVs. Moreover, algorithms trying to identify SVs from NGS data of short read-length showed a high false-negative rate (Sobczak and Krzyzosiak, 2002). Third-generation sequencing (TGS) techniques, including Single-Molecule Real-Time (SMRT) sequencing of Pacific Biosciences (PacBio) and the Nanopore long-read single-molecule sequencing of Oxford Nanopore Technologies (ONT), have shown higher sensitivity and specificity in structural variation detection and have been applied in tumor research including breast cancer research (Nattestad et al., 2018; Vasan et al., 2019; Aganezov et al., 2020).

Even though SVs of breast cancer in SKBR-3 cell line and patient-derived organoids have been widely studied (Zhuang et al., 2021), more proof is needed to illustrate the relationship between SVs and cancer. Nevertheless, the emerging TGS technologies with long-read capability have demonstrated their strengths in cancer study, which allows us to analyze the haploid genome at unprecedented precision. They could provide valuable insights into precision medicine, as in the case of double *in-cis* PIK3CA mutations showing high sensitivity for alpesilib (Vasan et al., 2019).

In this study, we aim to accurately detect DNA structural variations of a 28-gene panel in breast cancer tissue, matched by para-tumor tissues and blood samples via both ONT and PacBio TGS platforms. To the best of our knowledge, this study was the first to comprehensively analyze structural variation in breast cancer tissue directly via multiple TGS technologies.

METHODS

Ethical Approval

The study was approved by the Peking University People's Hospital ethics committee (Reference number 2021PHB227-001).

DNA Extraction

Genomic DNA was extracted from the frozen tissue/blood specimens using the standard phenol/chloroform extraction protocol. Briefly, the tissue specimens were fully ground with liquid nitrogen. For blood, 1 ml whole-blood samples were added with an equal amount of ice-cold cell lysis buffer (1.28 M sucrose, 40 mM Tris hydrochloride, 20 mM $MgCl_2$, 4% Triton X-100 [pH 7.5]), and three volumes of ice-cold distilled water. This mixture was incubated for 10 min on ice, and the nuclear pellet was collected by centrifugation (6,000 rpm, 5 min, 4°C). The nuclei both from tissue and blood samples were suspended in extraction buffer (1 M sodium chloride, 100 mM Tris, and 50 mM EDTA, buffered at pH 8.0) containing 2% sodium dodecyl sulfate (SDS) and proteinase K (2 mg/ml final concentration). The suspended nuclei were incubated at 56°C for 2 h, extracted with phenol-chloroform-isoamyl alcohol (28:24:1 by volume), one more time with chloroform-isoamyl alcohol (24:1 by volume), and precipitated with 0.7 volume of isopropyl alcohol at −20°C for 40 min. The DNA precipitates were washed in ice-cold 80% ethanol twice, collected by centrifugation (12,000 rpm, 15 min, 4°C), dried under vacuum, and finally resuspended in 100 ul of EB (10 mM Tris hydrochloride [pH 8.0]) (#19086, Qiagen). The quantity and quality of DNA samples were measured by NanoDrop One (ND-ONE-W, Thermo Fisher Scientific Inc.) and on 1% agarose gel electrophoresis.

Target Regions Capturing and Sequencing

DNA probes of 120 bases were designed to cover full-length genes of interest as a custom-made DNA-Cap Panel and were synthesized by Boke Biotechnologies (Wuxi, Jiangsu, China). During the design of the probes, the Repeat Masker dataset was used to remove probes corresponding to repetitive sequences in the human genome. Capture and enrichment of regions of interest were performed following the manufacturer's protocol. Briefly, 3 ugs genomic DNA was sheared to around 5–6 kb fragments by a g-TUBE (#520079, Covaris, Woburn, MA, United States) centrifugation (15,000 g, 2 min, twice). End-repair and dA-tailing of DNA fragments according to protocol recommendations were performed using the Ultra II End Prep module (#E7546, NEB) through pre-capture amplification. Targeted sequence capture was conducted by pooling indexed PCR products and hybridizing them with custom-made probes. Captured DNA fragments were amplified by PCR again using universal primer. After purification, the prepared target DNA was sequenced using the Pacific Biosciences (PacBio, Menlo Park, CA, United States) SMRT sequencing technology according to protocol recommendations. The PacBio SMRT Bell™ sequencing library was constructed using a SMRTbell Express Template Prep Kit 2.0 (#100-938-900, PacBio). Finally, sequencing was performed on the PacBio Sequel II platform according to the manufacturer's instructions.

Data Quality Control and Detection/ Annotation of SVs

Raw sequencing data (also called raw polymerase reads) were first tested in a standard quality control protocol by using the

SMRTlink 8.0 (PacBio) to remove low-quality reads and adapters resulting in subreads. The minimum polymerase reads accuracy was 0.75. The read quality (RQ) was marked as 0.8 if passed the quality control or as 0 if failed in the filtering. Subreads were obtained by the above filtering. Circular Consensus Sequence (CCS) was used to get CCS reads, and Lima was used for barcode splitting. PBMarkDUP (PacBio) was used to remove potential copies in CCS reads, and PBMM2 (PacBio) was used to compare CCS reads to the reference genome hg38. PBSV (V9.0, https://www.pacb.com/support/software-downloads/) was used to detect SVs, and DeepVariant (Poplin et al., 2018) (V1.0.0, https://github.com/google/deepvariant) was used to detect SNP and InDel. Detected mutations were be annotated by Annovar (Wang et al., 2010) (http://nar.oxfordjournals.org/content/38/16/e164) if the following criteria have been met. For SVs, 1) the number of supported reads with mutations ≥2; 2) mutation frequency among tumor samples ≥0.1; 3) mutation frequency = 0 in reference, and 4) screening mutations at interested regions. For SNP and InDel, 1) number of reads covering mutation sites ≥5, 2) the number of reads with mutations ≥2, 3) mutation frequency among samples ≥0.05; 4) the number of reads covering mutation sites ≥ in reference control ≥0; 5) the frequency ratio between reference and tumor samples <0.143, and 6) screening mutations at interested regions.

RNA Sample Preparation, cDNA Library Construction, and Sequencing

Total RNA from each tissue sample was extracted using the RNeasy Plus Mini Kit (Qiagen, Germany). The RNA purity was checked using the NanoDrop™ One (Thermo Fisher Scientific, United States). RNA degradation and contamination were monitored using 1% agarose gels. The RNA concentration was measured using the Qubit® RNA Assay Kit in the Qubit® 3.0 Fluorometer (Life Technologies, CA, United States). The RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2,100 system (Agilent Technologies, CA, United States). The RNA quality criteria for the RNA samples was RNA Integrity Number (RIN) > 8.0 and 2.0 < OD 260/280 < 2.2. Qualified RNAs were used for Nanopore library preparation. First, reverse transcription of qualified RNA, PCR amplification, and adapter ligation were performed using the library preparation kit SQK- PCS109 (Oxford Nanopore Technologies) following the recommended protocol. Then prepared libraries were sequenced on a Nanopore PromethION platform using flowcell R9.4.1.

Preprocessing of Sequencing Reads and Genome Mapping

For the raw sequencing reads, reads of which quality score is lower than seven or length is shorter than 200 bp were discarded using quality control tool Nanofilt (De Coster et al., 2018) (https://github.com/wdecoster/nanofilt). Then full-length reads were identified and oriented from sequencing reads by the pychopper tool (https://github.com/nanoporetech/pychopper) with default parameters. Then full-length reads were aligned to the hg38 reference genome using minimap2 (Li, 2018) (-ax splice -uf -junc-bed). Genome mapping results of full-length reads were visualized using the Integrative Genome Viewer (Robinson et al., 2011).

Prediction of Coding Sequences and Fusion Transcript Identification

Prediction of coding sequences and protein sequence was performed in all novel isoforms using the ANGEL software (Shimizu et al., 2006) (https://github.com/PacificBiosciences/ANGEL). Fusion transcripts were identified using fusion_finder.py from software cDNA_Cupcake (https://github.com/Magdoll/cDNA_Cupcake). Specifically, an identified fusion transcript must meet the following criteria: 1) fusion transcripts map to two or more loci in the genome; 2) each mapped locus must align with at least 95% identity and at least 5% coverage; 3) total aligned coverage of the fusion transcript must be above 99%; 4) each mapped locus must be at least 10 kb apart.

RESULTS

Target Regions Capturing and Coverage

Several approaches have been used to examine the genomic and transcriptional signatures in breast cancer patients. We recruited 19 breast cancer patients and seven control cases in this long-read study during 2019–2020 (**Figure 1**). All experimental designs and procedures abide by the regulations from the Institutional Review Board of Peking University People's Hospital. Multiple subtypes of breast cancer were selected as research subjects in this study, including four invasive subtypes (Luminal A, Luminal B, HER-2 enriched, and Triple Negative Breast Cancer (TNBC) cases previously classified by immunohistochemical staining) and Ductal Carcinoma *in situ* (DCIS) cases (**Table 1**). Three sets of samples (blood, para-tumor, and tumor) were obtained from all patients. Long-read DNA and RNA information was obtained for a 28-gene panel using the PacBio platform and the ONT full-length whole transcriptome platform, respectively (refer to methods). In addition, blood samples from 7 healthy control donors were processed with the same procedures (**Figure 1**).

By the combination of a full-length panel approach and long-read sequencing tools, it was possible to explore not only SNPs but also most SVs within these genes, regardless of their locations at either exons or introns. The panel in our study focuses on two gene types: twenty genes associated with a high risk of breast cancer and also participated in homologous recombination repair (HRR) (Breast Cancer Association et al., 2021; Hu et al., 2021; Yadav et al., 2021), and eight genes involved in the precision medicine during breast cancer treatment (Harbeck et al., 2019; Sparano et al., 2019; Waks and Winer, 2019) (**Supplementary Table S1**). Probes were designed to cover the whole genome regions of these genes, which are about 5 M bases. Our results are shown in **Figure 2**; **Supplementary Table S2** summarized some essential characteristics of this panel plus a long-read approach:

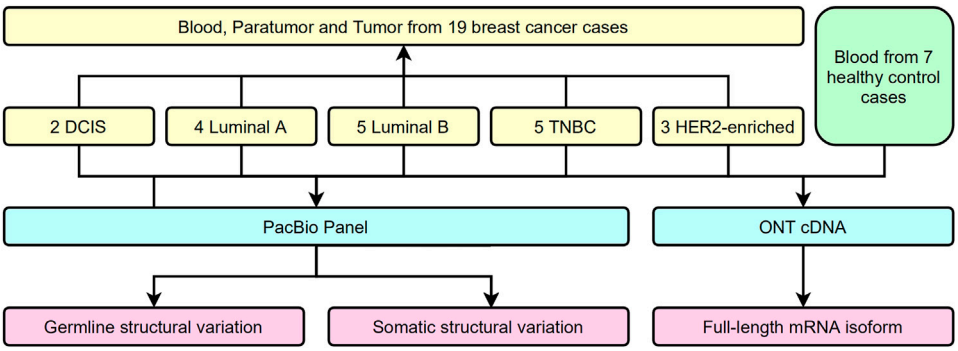


FIGURE 1 | Flow chart of study design.

TABLE 1 | Clinicopathological features of breast cancer patients and healthy controls recruited.

Classification	Patient #	Tumor Size (cm)	Lymph Node Metastasis	Ki-67	Sample Inclusion		
					B	P	T
DCIS	RM65	3*2	0	10%	DNA	DNA & RNA	DNA
	RM80	0.5	0	10%	DNA	DNA	DNA & RNA
HER2	RM62	3*1	0	30%	DNA	—	DNA & RNA
	RM63	3*1.5	0	15%	—	DNA & RNA	DNA & RNA
	RM71	2.5*2	2	30%	DNA	DNA	DNA & RNA
Luminal A	RM66	1.7*1.4	0	10%	—	DNA & RNA	RNA
	RM70	2.4*2	3	10%	DNA	DNA	DNA & RNA
	RM74	4.9*4*2.4	1	10%	DNA	DNA & RNA	DNA & RNA
	RM77	1.7*1.2	0	5%	—	DNA	RNA
Luminal B	RM72	2.5*2.5*1.8	2	40%	DNA	DNA	DNA & RNA
	RM73	3*2.8	6	50%	DNA	DNA & RNA	DNA & RNA
	RM76	9.5*7.5*1.8	1	20%	DNA	DNA & RNA	DNA & RNA
	RM78	0.6	0	40%	DNA	DNA	DNA & RNA
	RM79	4.3*3.7*2.7	2	20%	DNA	—	DNA & RNA
TNBC	RM64	2.7*2.2	0	70%	DNA	DNA & RNA	DNA & RNA
	RM67	1.6*1.5	3	70%	DNA	DNA & RNA	DNA & RNA
	RM68	2.8*2*1.9	1	90%	—	DNA & RNA	RNA
	RM69	1.5*1*1	0	70%	DNA	DNA & RNA	DNA & RNA
	RM75	2.5*2	0	20%	DNA	DNA	DNA & RNA
Healthy Control	RMH3				DNA		
	RMH7				DNA		
	RMH9				DNA		
	RMH12				DNA		
	RMH15				DNA		
	RMH20				DNA		
	RMH25				DNA		

sufficient depth of sequencing, long reads (N50 is around 3,500 bases), and high target coverage (>99.5%). There were no significant differences in these essential characteristics among the three types of samples and no apparent disparity between samples from patients and healthy controls (**Figure 2**; **Supplementary Table S2**).

Analysis of Germline SVs in Breast Cancer Patients

In our panel study, germline SVs were detected in the blood sample of 12 patients (12/19, 63%) against the healthy controls (**Figure 3**). The number of SV carried by a single patient varies from one to six (left subset, **Figure 3A**). Based on their locations, these SVs could be classified into exons, introns, upstream or downstream regions, untranslated regions (UTRs) at 3′ or 5’ side, flanking regions of genes within two kilo-bases, or multiple-hit sites, which means more than one of the aforementioned categories. Only a few SVs were found at exonic regions (6/33, bright blue blocks, upper inset in **Figure 3A**), which agrees with previous studies (Sakamoto et al., 2020). From another perspective, SVs could be found in HRD genes like *RAD51B* and *BRIP1* or treatment-related genes like *ERBB4* and *EGFR4*. The distribution of germline SVs in these genes was sporadic, and

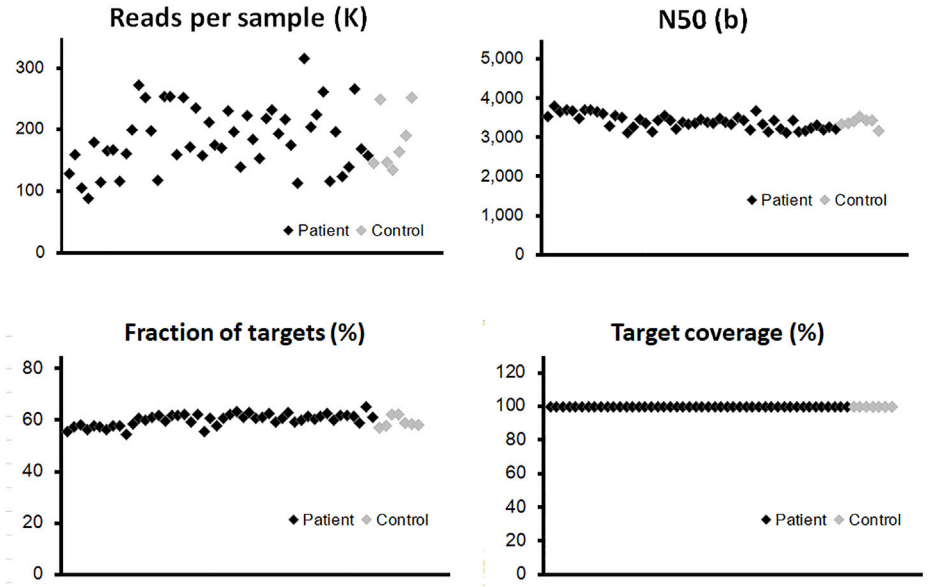


FIGURE 2 | Quality control of long-read sequencing for the panel of 28 genes. The probes were designed to cover the whole genome regions of all panel genes, which are around 5M-base coverage of detection. The vertical axis of each point illustrated the quantitative information from individual blood sample, para-tumor tissue, or tumor tissue. The effective read numbers were around 100–300 kb per sample, and the N50s were around 3,000–4,000. No obvious differences could be detected between patient and control groups (black and gray points, respectively). After the alignment process, the fraction of targets among different samples was around 55% with a slightly fluctuation. The coverage of the target region was above 99% in all tested samples.

no apparent high-frequency genes were counted, presumably due to the relatively small number of samples.

Notably, our long-read plus full-length-gene approach allowed us to detect SVs at locations that were hardly noticed by the conventional short-read techniques (**Figure 3B**). For instance, an about 250-base insertion at 3' UTR of EGFR was seen in patient RM65B, but not in healthy control RMH3. This UTR region is close to the centromere of Chromosome 7 and contains many TA repeats. Meanwhile, among individual reads, the locations of this insertion and its size are slightly different, as shown in **Figure 3B**, which further demonstrates the complexity of this mutation site.

Potential Hotspot of Somatic SVs Revealed in Tumor Tissue

The somatic SVs could be identified by annotating the unique SVs in tumor tissues against those in either blood samples or para-tumor tissues. When comparing SVs detected from para-tumor samples with that from matching blood samples, it was found that most of them were shared by both control samples (43–55 per patient, the upper plot in **Figure 4**, see also **Supplementary Table S3**), implying that these common SVs are background germline variants. Meanwhile, the existence of unique SVs in para-tumor samples (0–10, middle plot) and blood samples (2–8, bottom plot) was possibly caused by loss of heterozygosity (LOH). It indicated that the para-tumor tissue, which was histologically normal, had already been genetically altered in terms of SVs. This is consistent with findings from SNV studies (Hu et al., 2018). Our study used blood samples as a reference for tumor tissues to find cancer-driven SVs. The somatic SVs in tumor tissue affecting the 28 breast cancer-related genes were identified and displayed in **Figure 5** ; **Supplementary Table S4**. Our results showed that each patient carried none or only a few somatic SVs (0–3, 13 out of 19 patients had SVs ≥1, upper inset in **Figure 5A**) in this 28-gene panel study. Meanwhile, somatic SVs were detected in 12 out of 28 genes (12/28, 43%, right inset of **Figure 5A**). SVs had been classified into exonic and intronic according to their locations. In consistent with previous studies, most SVs were identified within the intronic region (Tuzun et al., 2005). Among the 12 genes, *ERBB2* had the highest SV frequency, which was detected in 4 patients and being all intronic, followed by *NF1* and *RAD51B*. **Figure 5B** summarized the four cases of SVs in *ERBB2*: two insertions and two duplications. Noteworthy, three of these cases have starting sites close to each other, resulting in a certain degree of overlap among the following sequences (RM73T, RM75T, and RM80T). These patients were clinically divided into three groups (Luminal B, TNBC, and DCIS). As far as we know, this region was AT-rich and had not been reported to cause disease. However, the relative enrichment of somatic SVs at the same site in *ERBB2* (3 out of 19 independent patients) suggests that this is an interest area in breast cancer and needs further validation with more samples.

Full-Length Transcriptome Analysis of Tumor and Para-tumor

Changes at the transcriptional level could provide support and direct evidence for genomic mutations. The cDNA of para-tumor and tumor were sequenced using the Nanopore

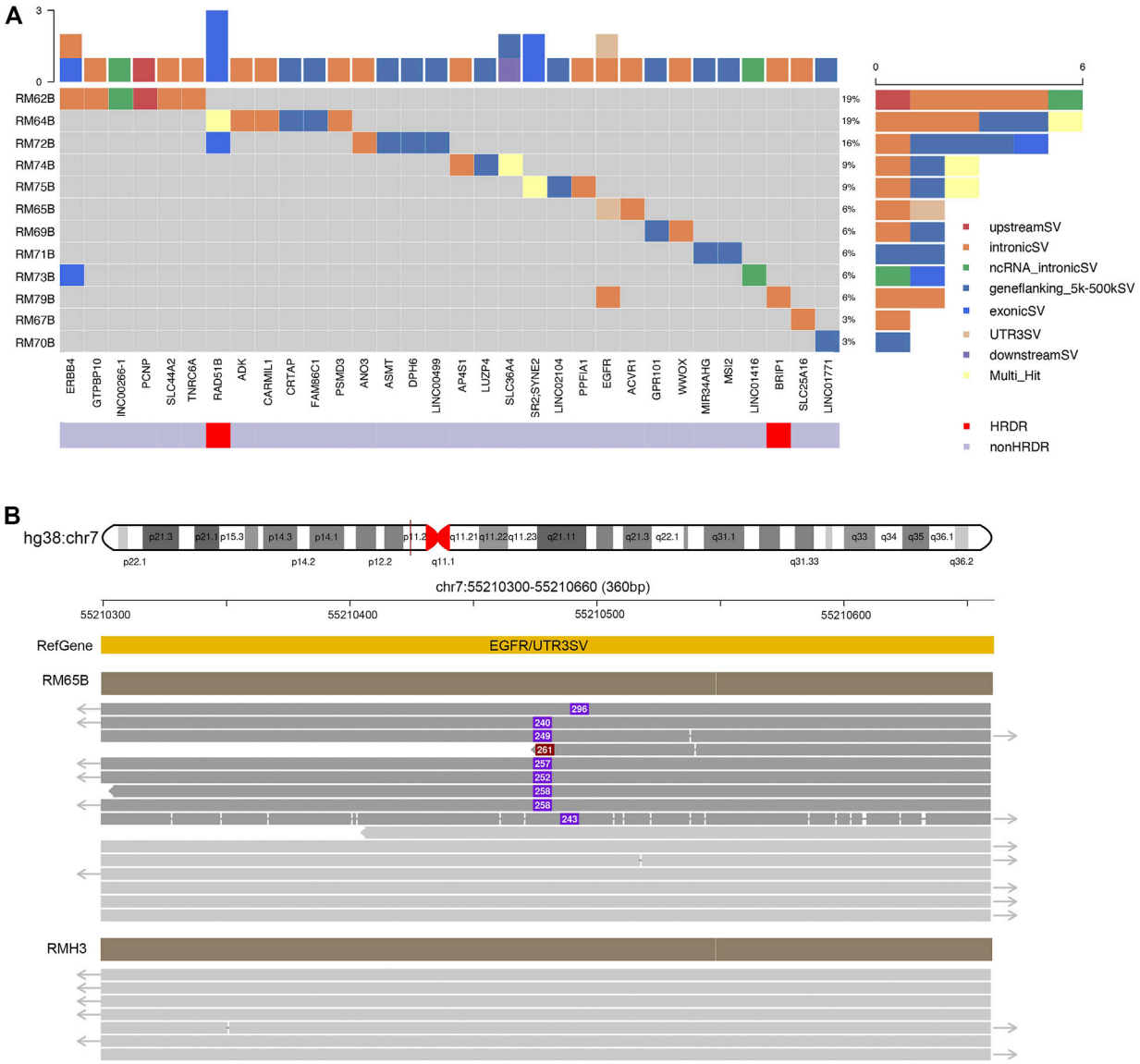


FIGURE 3 | Excessive germline structural variants occurred in non-exonic regions in breast cancer patients. **(A)** Summary of germline SVs in specific genes and patients. Individual patients' blood samples (starting with "RM" labeling, plus patient number, and end with B for blood samples) were examined against to seven healthy samples as control. The calling of SVs would be classified to one of the following categories: exons, introns, upstream or downstream regions, un-translated regions (UTRs) at 3′ or 5′ side, flanking regions of genes within 2 kilo-bases, or multiple-hit sites. The scales in the top and right insets illustrated the cumulative numbers of SVs in particular genes and patients, respectively. Most SVs were located at non-exonic regions. **(B)** SVs identified in EGFR. In RM65B, an insertion (~280 bp) was identified at the UTR 3′ region of EGFR genes. The dark red solid line in hg38:chr7 pointed to a 360-bp region as expanded below. Representative reads from RM65B and RMH3 (control) were aligned accordingly. The purple boxes and inside numbers showed the locations and sizes of this insertion in individual reads. Such insertion was identified only in a part of reads in RM65B (24/49) but not in RMH3 (0/83).

PromethION platform to get the full-length transcriptome data. Data points below 7 in Read Quality (accuracy lower than 85%) were excluded, and the valid points were scattered based on their length in **Figure 6A**. Our results showed that the average read quality was about 10, and the mean and median for read length were 1.3 k and 1.9 k base pairs, respectively. Principal component analysis revealed that the para-tumor and tumor tissues could be efficiently distinguished based on their transcriptomic data (red and green dots, respectively, **Figure 6B**). The density plot of reads per gene per 10,000 reads (RPG10K) showed that the reads per gene of tumor tissues is apparently shorter than that of the para-tumor tissues, but this conclusion can't be drawn at this stage due to the small sample size (**Figure 6C**). With all mentioned characteristics taken together, it suggests that long-read sequencing on transcriptome could potentially be a good candidate technique for diagnostic application in the future.

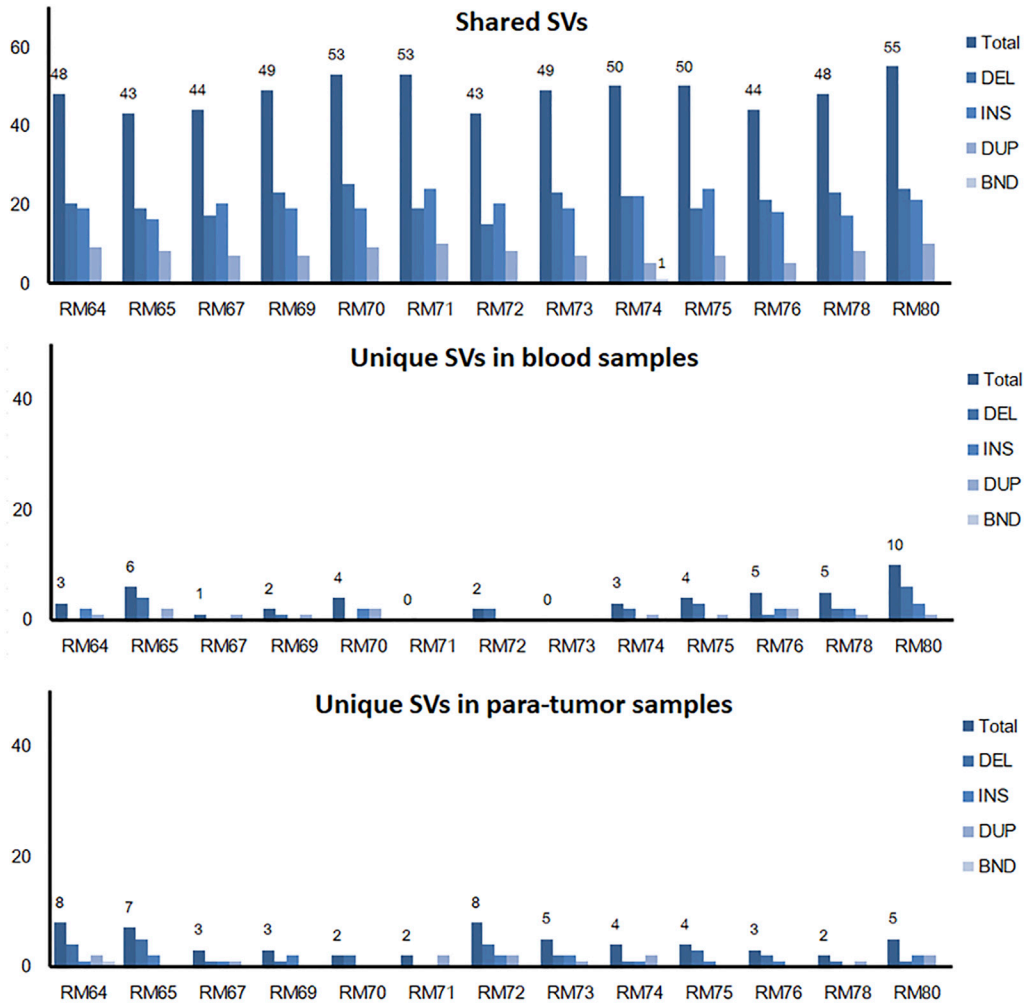


FIGURE 4 | Shared and unique SVs in para-tumor tissue and blood samples. A comparison of shared and unique SVs between two kinds of samples. Numbers above individual bars showed the number of total SVs. DEL, deletion; INS, insertion; DUP, duplication; and BND, Breakpoint notation. Most SVs were found in both tissues, while a few unique SVs were only observed in one of the tissues.

Gene Fusions With Both Genomic and Transcriptomic Evidence

The accumulation of fusion genes is one of the patterns commonly found in tumor tissues (Matsushige et al., 2019). However, how the fusion genes contribute to or are formed during cancer progression is barely documented. Due to their long-read sequencing characteristics, PacBio, and ONT platforms might benefit to studying fusion genes. In a total of 19 cases, we reported that there were seven fusion genes observed in six patients (**Table 2**). One case found in RM64 showed that a fusion gene at *RECQL5* in Chromosome 17 contained two other segments from Chromosomes 8 and 7 (**Figure 7A**), showing a particular case of a three-segment fusion gene. The confidence of this fusion gene is supported by its high coverage (depth >30X, **Figure 7B**) of reads obtained by the high fidelity PacBio HiFi platform. In addition, transcripts of this fusion gene were also obtained from the ONT Platform (**Figure 7C**). It appeared that a certain degree of alternative splicing was processed (indicated by dash lines between **Figures 7B,C**), resulting in a missing of Chromosome-7 segment as well as shorter length in transcripts.

DISCUSSION

A previous study demonstrates that PacBio long reads could detect over 20,000 SVs in a typical whole human genome (Chaisson et al., 2015). However, whole-genome third-generation sequencing is rather expensive and limits its clinical applications. To address this issue, we applied to the best of our knowledge the first clinical TGS panel using the PacBio HiFi platform to breast cancer samples. We conducted a comprehensive analysis on structural variations across 28 breast cancer-related genes through long-read genomic and transcriptomic sequencing of paired breast cancer tissue and blood. We compared the genome of tumor and paratumor tissue in 19 breast cancer patients to identify somatic SVs. We also compared the genome of white-blood-cell from 19 breast

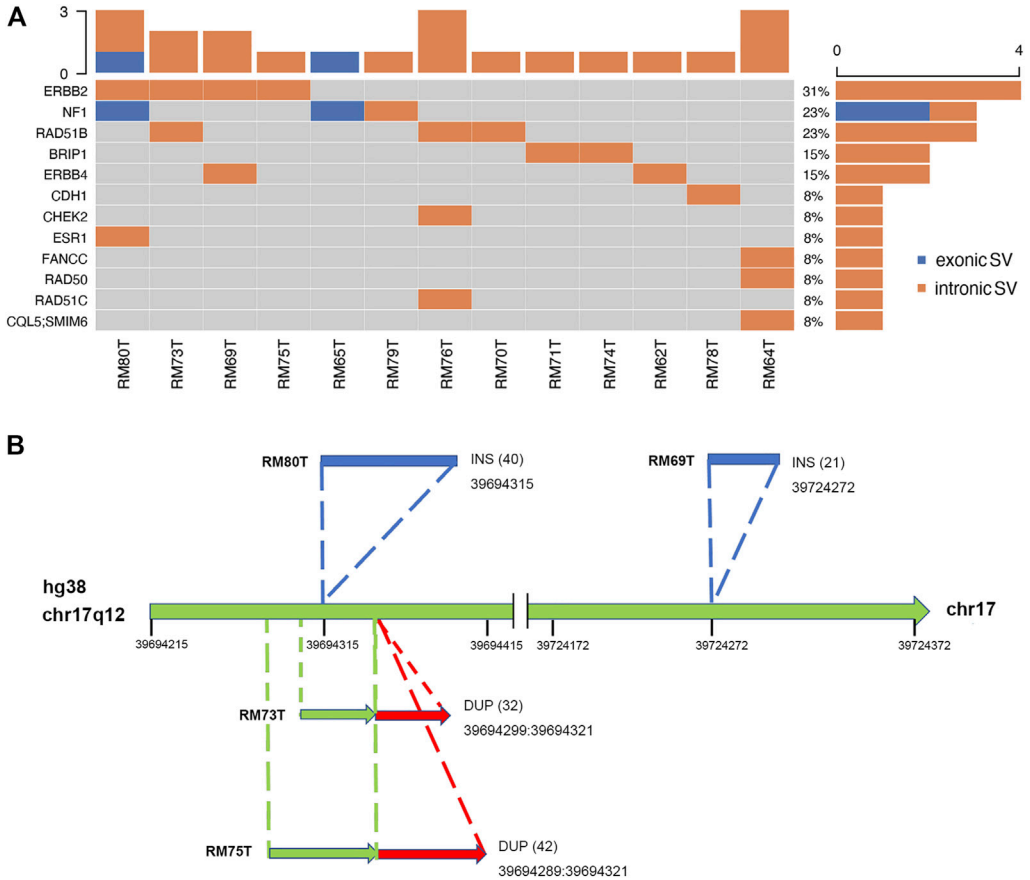


FIGURE 5 | Excessive somatic SVs in non-exonic region. **(A)** A summary of somatic SVs in particular genes and patients. Patients' tumor samples were examined against to their blood samples as control. SVs are sorted into exonic and intronic types based on their locations. Similarly, most SVs were found at non-exonic regions. Among thirteen patients, four of them had been identified to carry SVs at ERBB2. **(B)** Hotspot of SVs in ERBB2. There were two INS SVs (shown in blue) and two DUP SVs (red) found in ERBB2. Numbers inside parentheses indicated the sizes of SVs. A conserved region with SV occurrence was exposed among three independent patients (RM80T, RM73T, and RM75T).

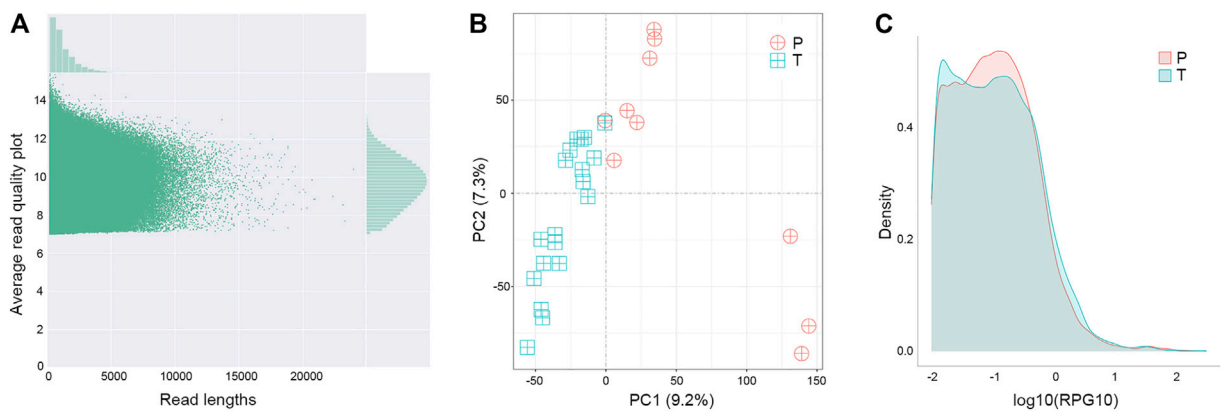


FIGURE 6 | Distinguishable transcriptomes of tumor samples from others in long-read sequence. **(A)** Quality control of transcriptome analysis. Transcriptome was built-up based on their cDNA library construction followed by ONT nanopore sequencing. Plot showed the quality of individual data points (main figure) as well as their distributions (top and left insets). Data points below 7 in Average Read Quality (i.e., lower than 85% accuracy from ONT manual instruction) were excluded from analysis. **(B)** Principal component (PC) analyses. Each symbol represented one clinical sample from para-tumor tissues (P) and tumor tissues (T). Clearly the majority of samples were quite distinguishable from ones from the other group. **(C)** Density of reads per gene10 (RPG10K) plot. Differential patterns from different types of samples revealed that tumor tissues had smaller reads per genes than para-tumor tissues.

TABLE 2 | Fusion genes detected in tumor samples.

Sample	Chr	Start	SV type	SV ID	Gene	Location
RM64	chr17	75646925	BND	pbsv.BND.chr17:75646925-chr7:9488789_1	RECQL5; SMIM6	intronic
RM64	chr9	95119452	BND	pbsv.BND.chr9:95119452-chr19:36870553_1	FANCC	intronic
RM71	chr17	61860680	BND	pbsv.BND.chr17:61860680-chr17:72564876_1	BRIP1	intronic
RM76	chr17	58714578	BND	pbsv.BND.chr17:58714578-chr17:57731280_1	RAD51C	intronic
RM78	chr16	68793666	BND	pbsv.BND.chr16:68793666-chr16:72915551_1	CDH1	intronic
RM79	chr17	31350606	BND	pbsv.BND.chr17:3,1350606-chr1:248935729_1	NF1	intronic
RM80	chr6	151943280	BND	pbsv.BND.chr6:151943280-chr17:38647479_1	ESR1	intronic

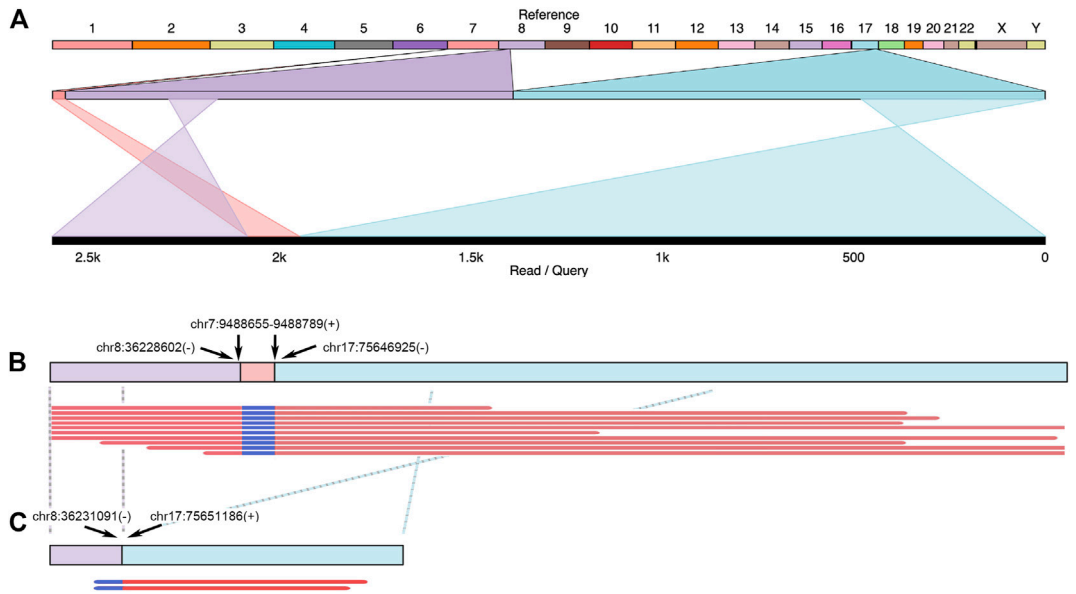


FIGURE 7 | Example of fusion genes and transcripts observed in patient tumor tissues. In Patient RM64, several structural variants (including DEL and BND) were identified at FANCC, RAD50, and RECQL5 regions. A three-segment fusion gene observed at RECQL5 was illustrated. **(A)** Top line: reference genome from hg38. Numbers and letters indicated individual chromosomes. Middle line: expansion of sequenced regions from chromosomes 7, 8, and 17. Bottom line: Illustration of one genomic structure in RM64T samples, containing regions from chromosomes 8, 7, and 17. Crossed projection lines from middle to bottom lines represented reversions occurred during the fusion process. **(B)** Top line: a plot of this fusion gene at genomic DNA. Breakpoint notation locations were labeled by arrows. Plus and minus symbols showed forward and reverse directions, respectively. Bottom lines: representative data selected from individual reads. **(C)** Transcripts observed in this area. Top line: a cartoon demonstration of corresponding mRNA. Crossed dash lines showed a reversion observed after a comparison to reference. Bottom lines: two reads continuously from chr8 to chr17 were shown. The chr7 segment in this fusion gene did not have detectable mRNA reads.

cancer patients and seven healthy controls to identify possible pathogenic SVs. Our results suggested that germline and somatic SVs were common in the selected genes among breast cancer patients, though the majority of them occurred in the non-exonic region. We also identified a potential hotspot region for somatic SVs. Taking together, our results demonstrated that SVs are potentially important in the tumorigenesis of breast cancer. Indeed, the International Cancer Genome Consortium (ICGC) previously showed that driver SVs are more prevalent than point mutations in breast adenocarcinomas (6.4 SVs compared with 2.2 point mutations on average) (Consortium, 2020).

The traditional NGS platforms have a poor mapping to repetitive elements, including tandem repeats and interspersed repeats, which has made a substantial fraction of most genomes inaccessible and limited its ability to detect SVs (Sedlazeck et al., 2018). One a representative type of interspersed repeats is Alu element which accounted for 11% of the human genome sequences on average, it belongs to a class of retroelements termed short interspersed nuclear elements (SINEs) and often causes SVs through homologous recombination (Doronina et al., 2021). An important reason that we developed the 28-gene TGS panel for illuminating the full landscape of SVs in breast cancer is to overcome the limitations of NGS in detecting SVs around repetitive elements. The repetitive elements are abundant in the 28-gene panel which contains most of the breast cancer-related genes, for instance, the *BRCA1* gene has around 40% of Alu family repetitive elements in its DNA sequences (Sobczak and Krzyzosiak, 2002; Ewald et al., 2009).

In this paper, by acquiring paired blood, paratumor and tumor tissue from patients, we delineated germline and somatic mutations which were both reported to be responsible for

carcinogenesis. Interestingly, we found a potential somatic SV hotspot in the AT-rich region of *ERBB2* gene. Although this region does not belong to interspersed repeats which often causes SVs through homologous recombination, there are proofs in previous studies that SV hotspots could exist in regions other than SINE elements and DNA transposons (Lin and Gokcumen, 2019). Hence, our method of fine-scale characterization of genomic structural variations using TGS holds great potential to elucidate the full landscape of SV in breast cancer.

We have also systematically examined the paratumor tissues which was used as control samples to identify somatic mutations in tumor. During the process of carcinogenesis, somatic mutations continuously accumulated within the tumor tissue, turning the genomic structure different from its surrounding paratumor tissues (Shoshani et al., 2021). It is essential to figure out to which extent the paratumor is different from the blood and tumor. We have shown that most SVs were the same in both blood and paratumor tissues, but different from those in the breast cancer tissues. This is in accordance with a previous study that demonstrated paratumor and tumor show very different occurrence patterns in copy number variations (Hu et al., 2018).

Our 28-gene TGS panel also showed great promise in identify casual SVs of breast cancer. *NF1* is one of the 12 breast cancer predisposition genes identified to date, however, virtually all previous studies have focused on evaluating breast cancer risk associated with putative pathogenic SNVs and small InDels (Chen et al., 2021; Hu et al., 2021). We have successfully identified two exonic SVs in two breast tumor tissues, which proves that our TGS panel is useful for detecting cancer-related SVs. Moreover, our TGS panel is robust in identifying SVs, as indicated by the concordance of most fusion genes identified by long-read genomic and transcriptomic sequencing.

Nevertheless, there are still a few hurdles that remain to address. The first one would be to study more cases. The main roadblock to this purpose was the high cost of TGS at the time of writing; hopefully, as the TGS market continues to grow, the price could eventually drop to a more reasonable level. The second one would be to substantiate the hypothesis of LOH occurrence in some SV regions. The challenge lies in the read-length limitation of the contemporary panel sequencing. We believe the advance of technologies would breakthrough in this section, as we have seen the industry pushing the read-length longer every few years. Last but not least, we plan to further evaluate the effectiveness of our TGS analysis pipeline by running it against public datasets side by side with most of the analysis methods available in the market.

CONCLUSION

In conclusion, we found that somatic SVs were abundant in the breast cancer genome, which suggests that they may play an important role in the process of tumorigenesis and cancer development. This is especially important for breast cancer, since the pan-cancer studies conducted by ICGC found that the driver SVs was most evidently prevalent in breast cancer compared to driver point mutations (Consortium, 2020). Taking together, our clinical TGS panel shown here is an accurate and robust method to detect SVs in breast cancer, which is both important for breast cancer research and holds great potential for further clinical application.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The name of the repository and accession number can be found below: National Genomics Data Center (https://ngdc.cncb.ac.cn/); PRJCA008780.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Peking University People's Hospital ethics committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

TH, YW, I-FP, and SW conceived and initiated the study; TH, ML, JW, FX, JZ, HY, and SW organized and collected the clinical samples and data; TH, JL, ZZ, XG, DY, QS, SL, JS, BL, XZ, DL, GY, TI, and SW analyzed the data; and TH, XF, WY, DW, YW, I-FP, and SW wrote the paper.

FUNDING

This work was supported by the National Key Research and Development Program of China (Grant No. 2021YFE0203200), the National Natural Science Foundation of China (Grant Nos. 92059105, 82002979), the Beijing Municipal Natural Science Foundation (Grant No. 7202212), the Research and Development Funds of Peking University People's Hospital (Grant Nos. RDX 2021-05, RDY 2020-16) and the Young Investigator Program of Peking University Health Science Center (Grant No. BMU2021PYB013).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2022.854640/full#supplementary-material

REFERENCES

Aganezov, S., Goodwin, S., Sherman, R. M., Sedlazeck, F. J., Arun, G., Bhatia, S., et al. (2020). Comprehensive Analysis of Structural Variants in Breast Cancer Genomes Using Single-Molecule Sequencing. *Genome Res.* 30, 1258–1273. doi:10.1101/gr.260497.119

Alkan, C., Coe, B. P., and Eichler, E. E. (2011). Genome Structural Variation Discovery and Genotyping. *Nat. Rev. Genet.* 12, 363–376. doi:10.1038/nrg2958

Breast Cancer Association, C., Dorling, L., Carvalho, S., Allen, J., González-Neira, A., Luccarini, C., et al. (2021). Breast Cancer Risk Genes - Association Analysis in More Than 113,000 Women. *N. Engl. J. Med.* 384, 428–439. doi:10.1056/NEJMoa1913948

Chaisson, M. J., Huddleston, J., Dennis, M. Y., Sudmant, P. H., Malig, M., Hormozdiari, F., et al. (2015). Resolving the Complexity of the Human Genome Using Single-Molecule Sequencing. *Nature* 517, 608–611. doi:10.1038/nature13907

Chen, Z., Guo, X., Long, J., Ping, J., Li, B., Fadden, M. K., et al. (2021). Discovery of Structural Deletions in Breast Cancer Predisposition Genes Using Whole Genome Sequencing Data from > 2000 Women of African-Ancestry. *Hum. Genet.*

Consortium, I. T. P.-C. a. O. W. G. (2020). Pan-cancer Analysis of Whole Genomes. *Nature* 578, 82–93.

De Coster, W., D'hert, S., Schultz, D. T., Cruts, M., and Van Broeckhoven, C. (2018). NanoPack: Visualizing and Processing Long-Read Sequencing Data. *Bioinformatics* 34, 2666–2669. doi:10.1093/bioinformatics/bty149

Doronina, L., Reising, O., and Schmitz, J. (2021). Gene Conversion Amongst Alu SINE Elements. *Genes (Basel)* 12. doi:10.3390/genes12060905

Duijf, P. H. G., Nanayakkara, D., Nones, K., Srihari, S., Kalimutho, M., and Khanna, K. K. (2019). Mechanisms of Genomic Instability in Breast Cancer. *Trends Mol. Med.* 25, 595–611. doi:10.1016/j.molmed.2019.04.004

Ewald, I. P., Ribeiro, P. L., Palmero, E. I., Cossio, S. L., Giugliani, R., and Ashton-Prolla, P. (2009). Genomic Rearrangements in BRCA1 and BRCA2: A Literature Review. *Genet. Mol. Biol.* 32, 437–446. doi:10.1590/s1415-47572009005000049

Feuk, L., Carson, A. R., and Scherer, S. W. (2006). Structural Variation in the Human Genome. *Nat. Rev. Genet.* 7, 85–97. doi:10.1038/nrg1767

Harbeck, N., Penault-Llorca, F., Cortes, J., Gnant, M., Houssami, N., Poortmans, P., et al. (2019). Breast Cancer. *Nat. Rev. Dis. Primers* 5, 66. doi:10.1038/s41572-019-0111-2

Hollox, E. J., Zuccherato, L. W., and Tucci, S. (2021). Genome Structural Variation in Human Evolution. *Trends Genet.*

Hu, C., Hart, S. N., Gnanaolivu, R., Huang, H., Lee, K. Y., Na, J., et al. (2021). A Population-Based Study of Genes Previously Implicated in Breast Cancer. *N. Engl. J. Med.* 384, 440–451. doi:10.1056/nejmoa2005936

Hu, T., Kumar, Y., Shazia, I., Duan, S. J., Li, Y., Chen, L., et al. (2018). Forward and Reverse Mutations in Stages of Cancer Development. *Hum. Genomics* 12, 40. doi:10.1186/s40246-018-0170-6

Iafrate, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y., et al. (2004). Detection of Large-Scale Variation in the Human Genome. *Nat. Genet.* 36, 949–951. doi:10.1038/ng1416

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

Li, Y., Roberts, N. D., Wala, J. A., Shapira, O., Schumacher, S. E., Kumar, K., et al. (2020). Patterns of Somatic Structural Variation in Human Cancer Genomes. *Nature* 578, 112–121. doi:10.1038/s41586-019-1913-9

Lin, Y. L., and Gokcumen, O. (2019). Fine-scale Characterization of Genomic Structural Variation in the Human Genome Reveals Adaptive and Biomedically Relevant Hotspots. *Genome Biol. Evol.* 11, 1136–1151. doi:10.1093/gbe/evz058

Matsushige, T., Kuwamoto, S., Matsushita, M., Oka Wardhani, L., Horie, Y., Hayashi, K., et al. (2019). Detection of Disease-specific Fusion Genes of Soft Tissue Tumors Using Formalin-Fixed Paraffin-Embedded Tissues; its Diagnostic Usefulness and Factors Affecting the Detection Rates. *Yonago Acta Med.* 62, 115–123. doi:10.33160/yam.2019.03.016

Nattestad, M., Goodwin, S., Ng, K., Baslan, T., Sedlazeck, F. J., Rescheneder, P., et al. (2018). Complex Rearrangements and Oncogene Amplifications Revealed by Long-Read DNA and RNA Sequencing of a Breast Cancer Cell Line. *Genome Res.* 28, 1126–1135. doi:10.1101/gr.231100.117

Poplin, R., Chang, P. C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., et al. (2018). A Universal SNP and Small-Indel Variant Caller Using Deep Neural Networks. *Nat. Biotechnol.* 36, 983–987. doi:10.1038/nbt.4235

Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative Genomics Viewer. *Nat. Biotechnol.* 29, 24–26. doi:10.1038/nbt.1754

Sakamoto, Y., Sereewattanawoot, S., and Suzuki, A. (2020). A new era of Long-Read Sequencing for Cancer Genomics. *J. Hum. Genet.* 65, 3–10. doi:10.1038/s10038-019-0658-5

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., et al. (2004). Large-scale Copy Number Polymorphism in the Human Genome. *Science* 305, 525–528. doi:10.1126/science.1098918

Sedlazeck, F. J., Lee, H., Darby, C. A., and Schatz, M. C. (2018). Piercing the Dark Matter: Bioinformatics of Long-Range Sequencing and Mapping. *Nat. Rev. Genet.* 19, 329–346. doi:10.1038/s41576-018-0003-4

Sharp, A. J., Cheng, Z., and Eichler, E. E. (2006). Structural Variation of the Human Genome. *Annu. Rev. Genomics Hum. Genet.* 7, 407–442. doi:10.1146/annurev.genom.7.080505.115618

Shimizu, K., Adachi, J., and Muraoka, Y. (2006). ANGLE: a Sequencing Errors Resistant Program for Predicting Protein Coding Regions in Unfinished cDNA. *J. Bioinform Comput. Biol.* 4, 649–664. doi:10.1142/s0219720006002260

Shoshani, O., Brunner, S. F., Yaeger, R., Ly, P., Nechemia-Arbely, Y., Kim, D. H., et al. (2021). Chromothripsis Drives the Evolution of Gene Amplification in Cancer. *Nature* 591, 137–141. doi:10.1038/s41586-020-03064-z

Sobczak, K., and Krzyzosiak, W. J. (2002). Structural Determinants of BRCA1 Translational Regulation. *J. Biol. Chem.* 277, 17349–17358. doi:10.1074/jbc.m109162200

Sparano, J. A., Gray, R. J., Ravdin, P. M., Makower, D. F., Pritchard, K. I., Albain, K. S., et al. (2019). Clinical and Genomic Risk to Guide the Use of Adjuvant Therapy for Breast Cancer. *N. Engl. J. Med.* 380, 2395–2405. doi:10.1056/nejmoa1904819

Sudmant, P. H., Rausch, T., Gardner, E. J., Handsaker, R. E., Abyzov, A., Huddleston, J., et al. (2015). An Integrated Map of Structural Variation in 2,504 Human Genomes. *Nature* 526, 75–81. doi:10.1038/nature15394

Tuzun, E., Sharp, A. J., Bailey, J. A., Kaul, R., Morrison, V. A., Pertz, L. M., et al. (2005). Fine-scale Structural Variation of the Human Genome. *Nat. Genet.* 37, 727–732. doi:10.1038/ng1562

Vasan, N., Razavi, P., Johnson, J. L., Shao, H., Shah, H., Antoine, A., et al. (2019). Double PIK3CA Mutations in Cis Increase Oncogenicity and Sensitivity to PI3Kalpha Inhibitors. *Science* 366, 714–723. doi:10.1126/science.aaw9032

Waks, A. G., and Winer, E. P. (2019). Breast Cancer Treatment: A Review. *JAMA* 321, 288–300. doi:10.1001/jama.2018.19323

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data. *Nucleic Acids Res.* 38, e164. doi:10.1093/nar/gkq603

Yadav, S., Hu, C., Nathanson, K. L., Weitzel, J. N., Goldgar, D. E., Kraft, P., et al. (2021). Germline Pathogenic Variants in Cancer Predisposition Genes Among Women with Invasive Lobular Carcinoma of the Breast. *J. Clin. Oncol.* 39, 3918–3926. doi:10.1200/jco.21.00640

Zhuang, Y., Grainger, J. M., Vedell, P. T., Yu, J., Moyer, A. M., Gao, H., et al. (2021). Establishment and Characterization of Immortalized Human Breast Cancer Cell Lines from Breast Cancer Patient-Derived Xenografts (PDX). *NPJ Breast Cancer* 7, 79. doi:10.1038/s41523-021-00285-x

Conflicts of Interest: JL, XG, DY, DL, GY, DW, YW and I-FP were employed by company GrandOmics Inc.

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

frontiers | Frontiers in Cell and Developmental Biology

ORIGINAL RESEARCH
published: 23 May 2022
doi: 10.3389/fcell.2022.893546

A Novel DNA Repair Gene Signature for Immune Checkpoint Inhibitor-Based Therapy in Gastric Cancer

Binbin Yuan[1†], Chengfei Jiang[1†], Lingyan Chen[1†], Lihui Wen[2†], Jinlong Cui[3], Min Chen[1], Shu Zhang[1], Lin Zhou[1], Yimeng Cai[4], Jian-Hua Mao[5], Xiaoping Zou[1], Bo Hang[5]* and Pin Wang[1]**

[1]Department of Gastroenterology, Affiliated Drum Tower Hospital, Medical School of Nanjing University, Nanjing, China, [2]Department of Rheumatology and Immunology, The Affiliated Drum Tower Hospital of Nanjing University Medical School, Nanjing, China, [3]Berkeley-Nanjing Research Center, Nanjing, China, [4]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, United States, [5]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, United States

OPEN ACCESS

Edited by:
D. P. Kreil,
Boku University Vienna, Austria

Reviewed by:
Taobo Hu,
Peking University People's Hospital, China
Alexandre Maréchal,
Université de Sherbrooke, Canada

***Correspondence:**
Xiaoping Zou
13770771661@163.com
Bo Hang
Bo_hang@lbl.gov
Pin Wang
pinwang729@126.com

[†]These authors have contributed equally to this work

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

Received: 10 March 2022
Accepted: 27 April 2022
Published: 23 May 2022

Citation:
Yuan B, Jiang C, Chen L, Wen L, Cui J, Chen M, Zhang S, Zhou L, Cai Y, Mao J-H, Zou X, Hang B and Wang P (2022) A Novel DNA Repair Gene Signature for Immune Checkpoint Inhibitor-Based Therapy in Gastric Cancer. Front. Cell Dev. Biol. 10:893546. doi: 10.3389/fcell.2022.893546

Gastric cancer is a heterogeneous group of diseases with only a fraction of patients responding to immunotherapy. The relationships between tumor DNA damage response, patient immune system and immunotherapy have recently attracted attention. Accumulating evidence suggests that DNA repair landscape is a significant factor in driving response to immune checkpoint blockade (ICB) therapy. In this study, to explore new prognostic and predictive biomarkers for gastric cancer patients who are sensitive and responsive to immunotherapies, we developed a novel 15-DNA repair gene signature (DRGS) and its related scoring system and evaluated the efficiency of the DRGS in discriminating different molecular and immune characteristics and therapeutic outcomes of patients with gastric adenocarcinoma, using publicly available datasets. The results demonstrated that DRGS high score patients showed significantly better therapeutic outcomes for ICB compared to DRGS low score patients ($p < 0.001$). Integrated analysis of multi-omics data demonstrated that the patients with high DRGS score were characteristic of high levels of anti-tumor lymphocyte infiltration, tumor mutation burden (TMB) and PD-L1 expression, and these patients exhibited a longer overall survival, as compared to the low-score patients. Results obtained from HPA and IHC supported significant dysregulation of the genes in DRGS in gastric cancer tissues, and a positive correlation in protein expression between DRGS and PD-L1. Therefore, the DRGS scoring system may have implications in tailoring immunotherapy in gastric cancers. A preprint has previously been published (Yuan et al., 2021).

Keywords: gastric cancer, immunotherapy, immune checkpoint blockade, immune checkpoint inhibitors, DNA repair gene signature, prognostic biomarker, score system

INTRODUCTION

Gastric cancer is one of the most common cancers worldwide, with more than one million new cases in 2020, ranking fifth in incidence and third in mortality worldwide (Abbott et al., 2016). Up to now, the prognosis of this disease at advanced stages remains dismal. Gastric cancer is a heterogeneous group of diseases with variable responsiveness to treatment such as chemotherapy and immunotherapy, and new biomarkers are needed to identify patients with gastric cancer for sensitivity toward such therapies.

Chemotherapy is used as standard treatment for gastric cancer, either as preoperative or postoperative therapy, at all pathological stages of the disease. Some regimens showed better 5-years survival (OS) or relapse-free survival (RFS) rates in certain patients. However, some regimens demonstrated considerable toxicity and mortality in reports (Harada et al., 2018; Joshi and Badgwell, 2021).

In recent years, the approval of multiple immune checkpoint inhibitors (ICIs) and the promising results from clinical trials of tumor immunotherapies have led to the development of tumor immunotherapy (Harada et al., 2018; Joshi and Badgwell, 2021). In ICI therapies, anti-PD-1 and anti-PD-L1 antibodies and anti-CTLA4 antibodies are highly effective in patients with microsatellite instability-high (MSI-H) subtype or high expression of PD-L1 (Chao et al., 2020; Fuchs et al., 2021). Most recently, the effect of immunotherapy combined with chemotherapy or targeted therapy on gastric cancer has been studied. For example, in CheckMate-649, the largest global randomized phase III clinical study in gastric cancer (Moehler et al., 2020), nivolumab plus chemotherapy as first-line treatment for previously untreated and advanced gastric cancer demonstrated superior OS and RFS compared to chemotherapy alone, reducing the risk of death by 20%. Some conflicting results have also been obtained with combined therapy, which may be due to different chemotherapy regimens and heterogeneity of patients. In summary, even though its impact on the outcome of gastric cancer has not been clearly defined, today, immunotherapy plays an important role in the treatment of gastric cancer, and combinational therapy is becoming a new trend.

The association between genetic instability or DNA repair defects with tumor susceptibility to immunotherapy has been observed in certain types of tumors. DNA damage and genomic instability have been found to affect the anti-tumor immune response. There are several major repair pathways for the removal of different types of exogenous and endogenous DNA lesions, including direct reversal, base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR) and double-strand break (DSB) repair that includes homologous recombination (HR), non-homologous end joining (NHEJ) and the Fanconi anemia pathway (Hang, 2004; Friedberg et al., 2015; Rodríguez and D'Andrea, 2017). Deficiency in such repair is associated with reduced DNA repair capacity and increased genetic instability, thus promoting cancer development. It can also provide an opportunity or benefit for cancer therapy, as the efficacy of certain anticancer drugs or therapies is highly influenced by cellular DNA repair capacity, for example, small-molecule inhibitors of DNA repair have been combined with conventional chemotherapy drugs (Helleday et al., 2008). As mentioned above, the noticeable efficacy of ICIs in cancers with MMR-deficient and its characteristic genetic signature MSI has recently been discovered, and a large-scale analysis showed that mutations of mismatch repair genes and DNA polymerases account for 13.5% of high tumor mutation burden (TMB) tumors (Chalmers et al., 2017; Le et al., 2017). There is also evidence of potential implications for immunotherapy in cancers with homologous recombination deficiency (HRD) and somatic changes in the NER pathway (Bever and Le, 2018). Taken together, accumulating evidence suggests that a systemic understanding of DNA repair landscape in tumor may help assess the tumor susceptibility to immunotherapy.

There are still many challenges in the application of tumor immunotherapy, for instance, only a small portion of tumors are susceptible, the overall clinical response rate is low and it remains difficult to accurately predict treatment efficacy and response. Such limitations warrant a search for new immunotherapy biomarkers, such as those based on PD-L1 expression, tumor-infiltrating lymphocytes, TMB, deficient MMR, immune gene signatures, and multiplex immunohistochemistry (Gibney et al., 2016). More recently, a role for DNA repair in the selection of patients for immunotherapy has emerged. In this study, we aimed to explore the prognostic role of DNA repair gene expression in gastric cancer in relation to the prediction of response to ICI-based immunotherapy. We identified a 15-DNA repair gene signature (DRGS) and developed a scoring system for clinical utility. In addition, the association of the DRGS score with the molecular and immune profiles was further investigated in the same patients. The results suggest that our DRGS is a promising biomarker for tailoring immunotherapy.

MATERIALS AND METHODS

Datasets Used in This Study

The next-generation sequencing data and clinicopathological information for 407 patients with gastric cancer were downloaded from The Cancer Genome Atlas Stomach Adenocarcinoma (TCGA-STAD) data collection (https://portal.gdc.cancer.gov). The microarray data for 60 patients with gastric cancer in GSE30727 were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE30727). We also used a dataset with immunotherapy information [the R package "IMvigor210CoreBiologies (version 1.0.0)"] (Balar et al., 2017) to evaluate the prediction efficiency of our scoring system. Drug sensitivity data between different gastric cancer cells *in vitro* were downloaded from Genomics of Drug Sensitivity in Cancer (GDSC, http://www.cancerrxgene.org/downloads).

Construction of the 15-DNA Repair Gene Signature and Scoring System

From the Human DNA Repair Genes website (https://www.mdanderson.org/documents/Labs/Wood-Laboratory/human-dna-repair-genes.html) (Wood et al., 2005) which was last modified by Drs. R. Wood and M. Lowery on 10th June 2020, a list of 219 DNA repair genes were defined. We used the DEseq2

(version 1.28.1) package in R (version 4.0.5) to analyze the absolute counts for differential gene expression in TCGA-STAD and employed the online platform GEO2R to assess whether these DNA repair genes are differentially expressed between tumor and normal tissues (adjusted $p < 0.05$, $|\log2FC| > 1$) (**Supplementary Table S1**). Fifteen differentially expressed DNA repair genes were commonly identified in TCGA-STAD and GSE30727, which are: PRKDC, FANCI, LIG1, RECQL4, FANCA, BRCA2, FANCD2, UBE2T, POLQ, PARPBP, EXO1, XRCC2, RAD54L, EME1 and FANCB. These genes constituted our DNA repair gene signature (DRGS). **Supplementary Table S2** showed expression level of the 15 genes in the DRGS.

Before the analysis, we applied the Z-score normalization to the expression level of the 15 genes from each patient sample. The principal component analysis (PCA) is a proven technique, which can reduce dimension, improve interpretability and minimize information loss of large datasets at the same time. We employed the PCA as a dimensionality reduction method to obtain the 1st dimension correlation coefficients of the 15 genes (see **Supplementary Figure S1** and **Supplementary Table S3**).

The formula for the DRGS scoring system was as follows:

$$\textbf{DRGS score} = \sum (\textbf{correlation coefficients for gene i}) * (\textbf{normalized gene i expression level}) \quad (1)$$

Comparison of the Immunocyte Infiltration

To determine the relationship between the DRGS scoring system and immunocyte infiltration in the gastric cancer tissue, we performed immunocyte infiltration analysis through the online platform Immune Cell Abundance Identifier (ImmuCellAI) (http://bioinfo.life.hust.edu.cn/ImmuCellAI#!/) using TCGA-STAD samples (n = 375) (Miao et al., 2020). Then, we calculated DRGS scores for the same TCGA-STAD samples and used the median score as the cut-point to divide these samples into high-score subtype and low-score subtype (**Supplementary Table S4**). The Mann Whitney U test was used to compare the results of multiple immune cell infiltration between these two subtype groups. We applied hierarchical clustering to immune cells with significantly different infiltration scores between the two groups. The Mann Whitney U test was used to compare the infiltration score between high-score subtype and low-score subtype. We also used ImmuCellAI to predict the patient response to ICB therapy (**Supplementary Table S5**).

Comprehensive Analysis of Histologic and Molecular Characteristics

Samples from the two groups mentioned above (TCGA-STAD, High score $n = 187$, Low score $n = 188$) were also used to explore the relationship between the DRGS score and various tumor characteristics. We compared the Lauren classification, molecular subtype and MSI status using the chi-square test (**Supplementary Table S6**). Tumor mutation levels were calculated through R package "Maftools" (Mayakonda et al., 2018). We used the Mann Whitney U test to compare TMB and PD-L1 expression between the two groups. In light that the TMB and PD-L1 expression are strongly related to the ICI therapy, we performed the Spearman's correlation to analyze the relationship between the DRGS score and the two immunotherapeutic biomarkers (**Supplementary Table S7**).

Relationship Between the 15-DNA Repair Gene Signature Score and Immunotherapy Outcomes

The public dataset IMvigor210CoreBiologies (version 1.0.0) contains a total of 348 advanced urothelial bladder cancer patients who received atezolizumab treatment. We applied the DRGS score to the dataset to investigate its predictive ability of immunotherapy. We performed normalization and PCA analysis with the same method as mentioned above, and all the 348 patients were divided into two equal groups according to their DRGS score (High score $n = 174$, Low score $n = 174$). We then compared their response to immunotherapy and the immunotypes of tumor by the Chi-square test (**Supplementary Table S8**). Also, the Kaplan-Meier analysis of OS in the two groups was performed, and log-rank test was used to compare survival curves.

Prognostic Ability of the 15-DNA Repair Gene Signature Score

We also analyzed the prognostic ability of the DRGS scoring system in TCGA-STAD samples, through Kaplan-Meier analysis of OS and the log-rank test for comparison. The independence of the DRGS score system was verified by univariate and multivariate Cox regression. Three independent gastric cancer datasets, GSE26901, GSE15459, and GSE26899, which were downloaded from the GEO database as validation sets and analyzed with the same method described above.

Association of the 15-DNA Repair Gene Signature Score With Drug Sensitivity

We downloaded the drug sensitivity data from GDSC combined with the RNA-seq data from 23 types of gastric cancer cells. We performed the Pearson correlation analysis to calculate the correlation between the drug sensitivity level of different chemotherapeutic agents and the DRGS score. The criteria for significant correlation were $p < 0.05$ and correlation > 0.4.

Immunohistochemistry Analysis

After obtaining the consent of 10 gastric cancer patients, tissue sections were obtained from the Department of Pathology, the Affiliated Drum Tower Hospital of Nanjing University Medical School. After blocking with endogenous peroxide and protein, the sections were then incubated with diluted specific PD-L1 and DRGS protein antibodies at 4°C overnight, respectively. The next day, the sections were incubated with a secondary antibody at 37°C for 1 h. The sections were then stained with 3,3-diaminobenzidine (DAB) solution for 3 min and counterstained with hematoxylin, and photographed under a microscope. Two experienced pathologists scored these

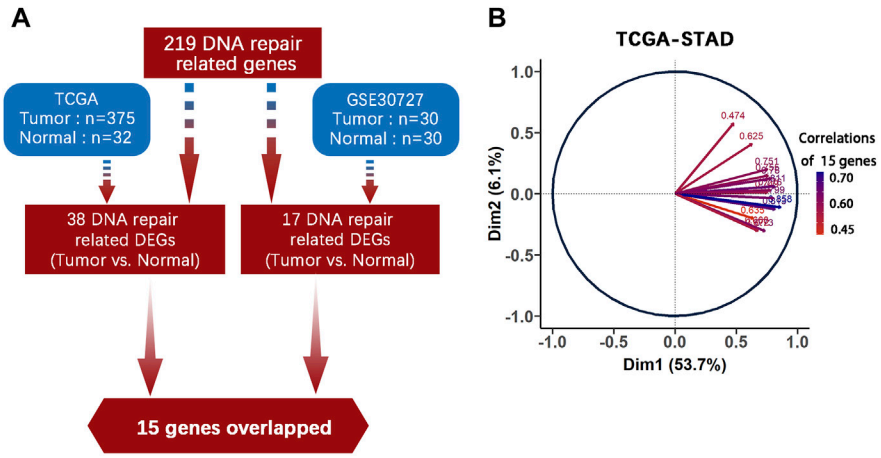


FIGURE 1 | DNA repair genes and construction of the DRGS scoring system. **(A)** Identification of 15 common genes among the 219 DNA repair genes that are differentially expressed in both TCGA-STAD (Tumor *n* = 375, Normal *n* = 32) and GSE30727 (Tumor *n* = 30, Normal *n* = 30); **(B)** PCA analysis was used as a dimensionality reduction method to obtain the 1st dimension correlation coefficients of the 15 repair genes. DEGs: Differentially expressed genes.

samples according to the percentage of staining intensity cells and staining intensity score (1 = low positive; 2 = positive; 3 = high positive). The formula for the IHC score was as follows:

$$\text{IHC score} = \sum (\textit{percentage of staining intensity cells}) * (\textit{staining intensity score}). \quad (2)$$

Verification of Prognosis-Related 15-DNA Repair Gene Signature and Programmed Cell Death Ligand 1 Expression

Data from our IHC results and the Human Protein Atlas (HPA) (https://www.proteinatlas.org/) were analyzed to verify the protein expression of DRGS and PD-L1 in tumor tissues and normal tissues, and to determine whether the expression differences and the correlation between DRGS and PD-L1 were consistent with the previous mRNA results from TCGA. (We performed the Pearson correlation analysis to calculate the correlation between protein expression of DRGS and PD-L1, and the difference was considered significant if the *p*-value is < 0.05.)

Statistical Analysis

Statistical methods were described in different sections above. The significance level of these statistical tests was two-sided *p* value < 0.05. The software used for analysis is R (version 4.0.5).

RESULTS

Identification of a 15-DNA Repair Gene Signature and Construction of the Scoring System

219 DNA repair genes were obtained from the Human DNA Repair Genes website as described above. Next, we used DEseq2 package in R to process the raw data from TCGA-STAD and used the online platform GEO2R to find genes differentially expressed between tumor (*n* = 375) and normal (*n* = 32) tissues (adjusted *p* value < 0.05, |log2FC| > 1). Finally, we identified 15 common genes among 219 DNA repair gene, which were differentially expressed in both TCGA-STAD and GSE30727 (Tumor *n* = 30, Normal *n* = 30) (**Figure 1A**, **Table 1**). We then used PCA analysis as a dimensionality reduction method to obtain the first dimensional correlation coefficients of the 15 genes (**Figure 1B**). Based on these coefficients, we created a 15-DNA repair gene signature (DRGS) scoring system (for details, please refer to the methods section). The DRGS score was then used to divide patients into two groups (high and low score groups).

TABLE 1 | The 15 DNA repair genes in the DRGS and their respective values.

Gene symbol	TCGA-STAD		GSE30727	
	logFC	adj.P.Val	logFC	adj.P.Val
PRKDC	1.081	1.71E-16	1.042	9.93E-04
FANCI	1.825	1.40E-31	1.012	3.08E-05
LIG1	1.012	4.80E-13	1.009	1.70E-02
RECQL4	2.112	5.64E-28	1.007	2.28E-03
FANCA	1.647	5.54E-23	1.412	1.19E-05
BRCA2	1.740	2.90E-22	1.091	3.30E-03
FANCD2	1.582	3.12E-25	1.196	1.04E-03
UBE2T	1.618	2.96E-21	1.036	1.01E-03
POLQ	1.906	5.96E-22	1.182	8.10E-04
PARPBP	1.386	4.54E-18	1.255	2.17E-03
EXO1	2.152	3.02E-32	1.136	6.97E-03
XRCC2	2.177	5.83E-37	1.016	1.33E-03
RAD54L	1.735	4.55E-23	1.290	4.36E-04
EME1	2.160	5.61E-32	1.030	1.05E-03
FANCB	1.789	9.37E-25	1.057	9.79E-04

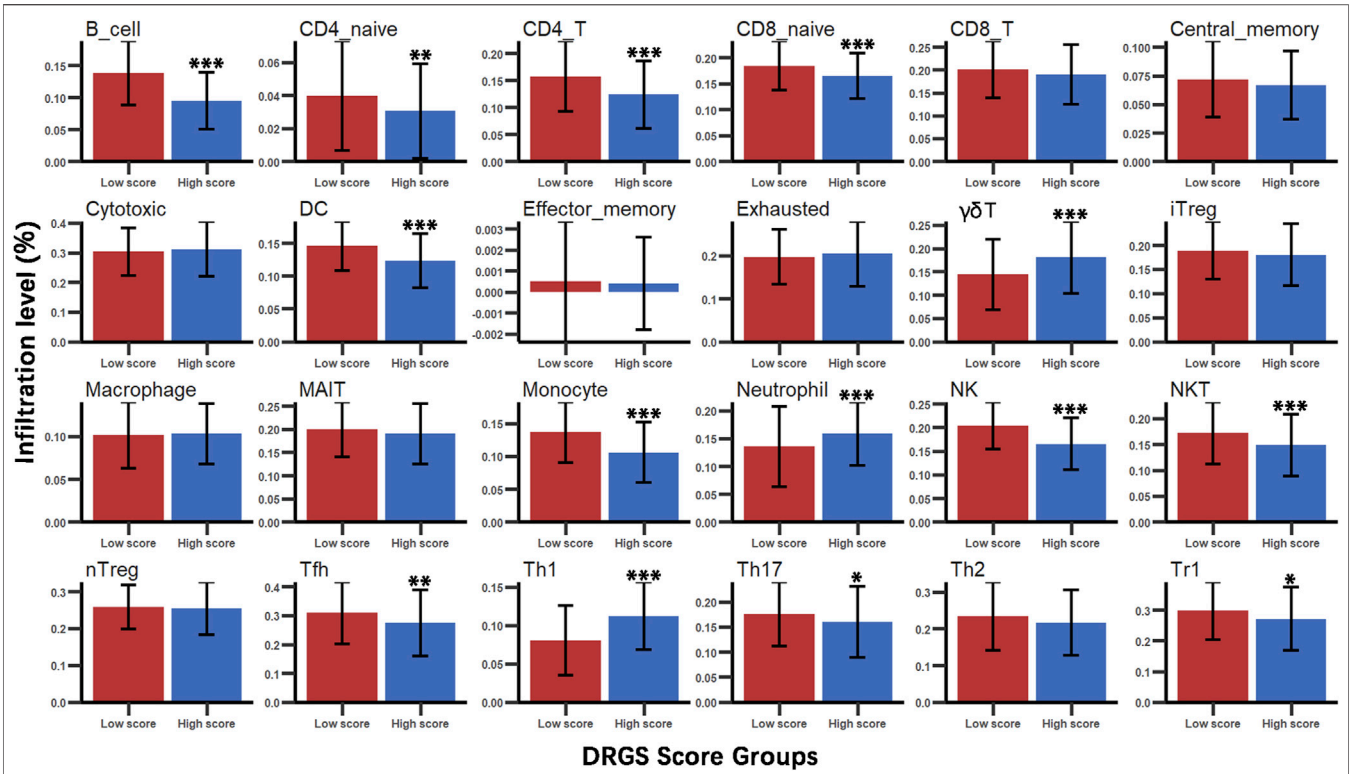


FIGURE 2 | Immune cell types infiltration in DRGS high (n = 187) and low (n = 188) score subgroups. Mann Whitney U test was used to compare the distribution of multiple immune cell infiltrations between the two groups (*p < 0.05, ** p < 0.01, *** p < 0.001).

Immune Characteristics of Different 15-DNA Repair Gene Signature Score Groups

We identified the immune characteristics of 375 TCGA-STAD samples through ImmuCellAI to estimate the specific fractions of 24 types of immune cells in each gastric cancer sample (**Supplementary Table S4**). To analyze the composition of immune cells in different score subgroups (High score n = 187, Low score n = 188), we used the Mann Whitney U test to compare the distribution of immune cell types in the two score groups. The result showed significant differences in the infiltration rates of 14 immune cells between the two groups. We found that anti-tumor cells such as γδ T cells, neutrophil cells and Th1 cells were more abundant in the DRGS high score group (p < 0.001), while immunosuppressive or tumor enhancement cells such as B cells (p < 0.001), CD4 T cells (p < 0.001), monocytes (p < 0.001), Th17 cells (p = 0.031), Tfh cells (p = 0.002) and Tr1 cells (p = 0.012) were more abundant in the DRGS low score group (**Figure 2**, **Figure 3A**). In our study, infiltration score of samples in the DRGS low score group was significantly higher than those in the high score group (p < 0.001, Mann-Whitney U test, **Figure 3B**). Furthermore, we used ImmuCellAI to predict the response of ICB therapy (**Supplementary Table S5**) and found that there would be 53.5% of patients in the DRGS high score group and fewer patients (37.8%) in the DRGS low score group, in response to ICB therapy (p = 0.003, Chi-square test, **Figure 3C**).

Relationship Between 15-DNA Repair Gene Signature Score Groups and Other Histologic and Molecular Classifications

To examine the relationship between our score grouping and other histologic and molecular classifications, we analyzed clinical data of 372 samples of gastric cancer patients from the TCGA-STAD database (**Supplementary Table S6**). According to the Lauren classification, gastric cancers can be divided into 3 subtypes, i.e., intestinal, diffuse and mixed. In our study, the DRGS high score group (n = 115) was characterized by a very high proportion of intestinal subtype samples (86.1%), while diffuse and mixed subtypes samples were only 7.8 and 6.1%, respectively. On the other hand, the DRGS low score group (n = 115) is comprised of 51.3% intestinal, 40.0% diffuse and 8.7% mixed samples. Statistically, there were significantly more intestinal and fewer diffuse type gastric cancer samples in the DRGS high score group as compared to the low score group (p < 0.001, Chi-square test, **Figure 4A**).

TCGA proposed 4 subtypes of gastric adenocarcinoma: 1) EBV-positive, 2) microsatellite instability (MSI), 3) genomically stable (GS), and 4) chromosomal instability (CIN) (The Cancer Genome Atlas Research Network, 2014). In our study, the DRGS high score group (n = 169) had a higher percentage of EBV-positive (11.2%) and MSI (27.8%) subtype samples. In contrast, the DRGS low score group (n = 169) had a higher percentage of

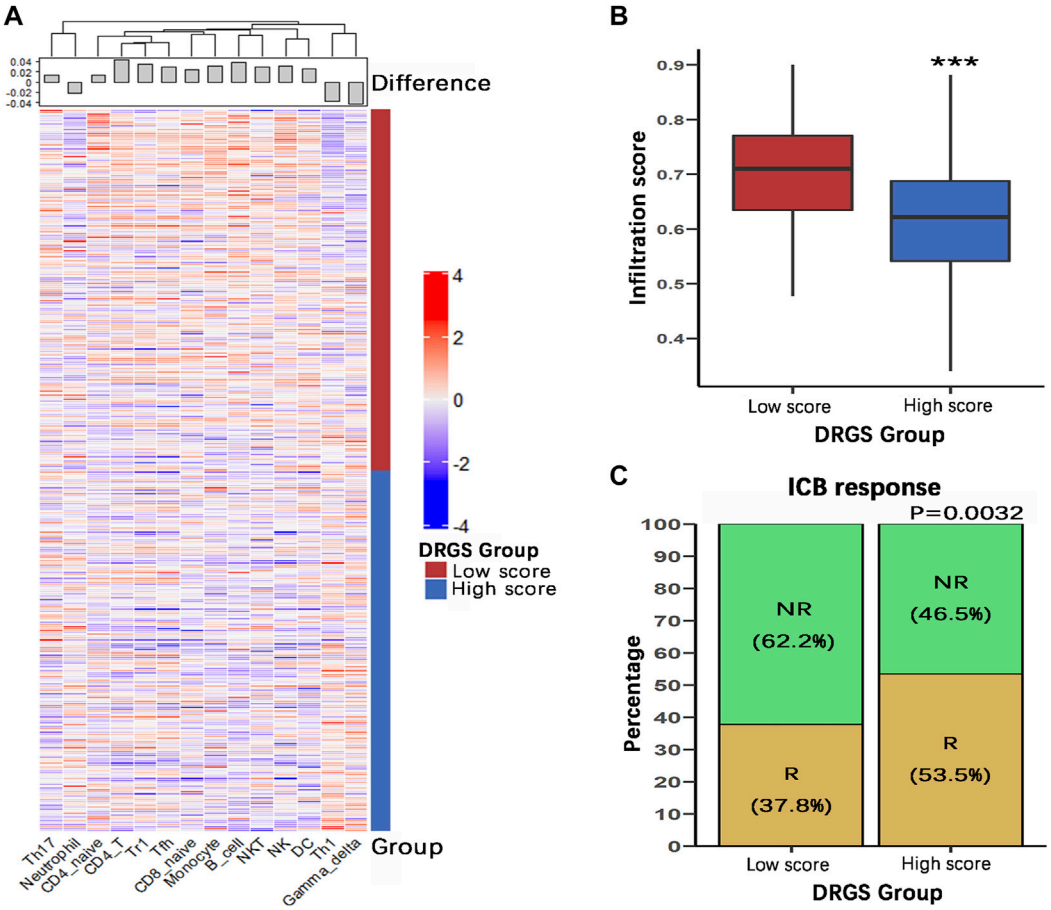


FIGURE 3 | Immune characteristics of the DRGS score subgroups (High score n = 187, Low score *n* = 188). **(A)** Hierarchical clustering of various types of immune cells with significant differences in infiltration scores between the two score groups; **(B)** The overall infiltration score of low score group was significantly higher than the high score group; **(C)** ImmuCellAI was used to predict the response of ICB therapy. R: response; NR: no response. (*$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

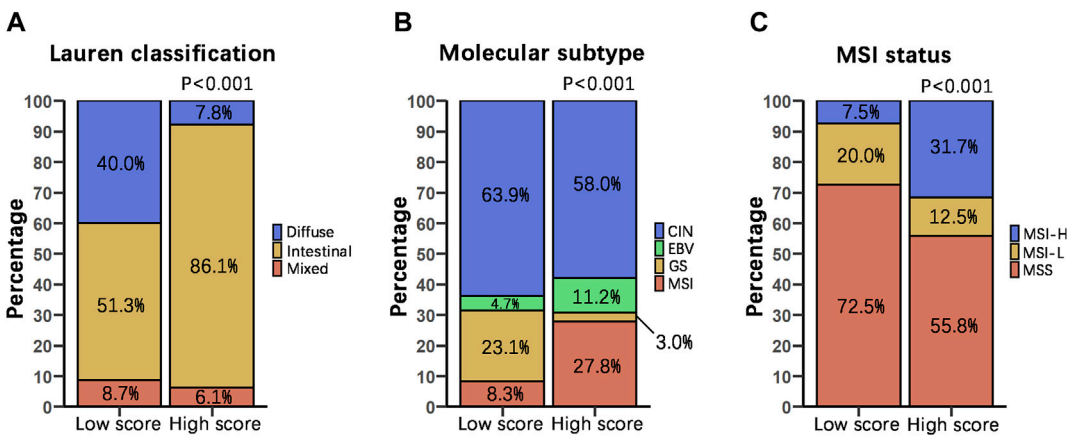


FIGURE 4 | Relationship between score grouping and other histologic and molecular classifications. We compared the percentage distribution of the Lauren classification (High score *n* = 115, Low score *n* = 115) **(A)**, molecular subtype (High score *n* = 169, Low score *n* = 169) **(B)**, and MSI status proportion (High score *n* = 120, Low score *n* = 120) **(C)** of patients with DRGS high and low score using Chi-square test. In all cases, p is less than 0.001.

TABLE 2 | Univariate and multivariate Cox regression confirm the independence of the DRGS scoring system.

Variables	HR	95% CI for HR*		*p* values
		Lower	Upper	
Clinical factors				
Age	1.041	1.020	1.062	<0.001
Stage				0.190
Stage II vs. Stage I	0.916	0.319	2.629	0.871
Stage III vs. Stage I	0.830	0.210	3.280	0.790
Stage IV vs. Stage I	1.778	0.431	7.330	0.426
T grade				0.260
T2 vs. T1	4.571	0.588	35.547	0.146
T3 vs. T1	6.761	0.795	57.478	0.080
T4 vs. T1	5.603	0.636	49.379	0.121
N grade				0.079
N1 vs. N0	1.236	0.583	2.621	0.580
N2 vs. N0	1.823	0.734	4.523	0.196
N3 vs. N0	2.517	1.016	6.236	0.046
M grade	1.788	0.761	4.206	0.183
DRGS	0.972	0.946	0.998	0.038

**HR, hazard ratio; CI, confidence interval.*

GS (23.1%) and CIN (63.9%) subtype samples ($p < 0.001$, Chi-square test, **Figure 4B**).

On the basis of the frequency of mutations in microsatellite markers, gastric cancer can be classified as MSS, MSI-H, and MSI-L (Rhyu et al., 1994; Yoon et al., 2013). In our study, the percentage of MSI-H subtype patients (31.7%) was higher in the DRGS high score group ($n = 120$) than that in the low score group (7.5%) ($n = 120$). In contrast, there were more MSS subtype patients (72.5%) in the DRGS low score group than in the high score group (55.8%). ($p < 0.001$, Chi-square test, **Figure 4C**).

Prognostic Ability of the 15-DNA Repair Gene Signature Score

To investigate whether the prognostic impact of the DRGS score is independent of clinical factors that could be associated with patient outcomes, we first performed univariate Cox regression analysis on all available clinical parameters in the TCGA-STAD dataset ($n = 372$). Clinical parameters that demonstrated significant prognostic impact ($p < 0.05$ by the Wald test) were selected for multivariate Cox regression analysis along with the DRGS. The result indicated that the DRGS is an independent prognostic factor ($p = 0.038$, **Table 2**). Taking the median score as the cut-off value, patients with DRGS high score had a better OS than those with DRGS low score ($p = 0.007$, log-rank test, **Figure 5A**). Then, the prognostic value of scoring system was validated using the GSE26901 ($n = 109$), GSE15459 ($n = 192$) and GSE26899 ($n = 93$) gastric cancer datasets. As shown in **Figure 5**, the patients in GSE26901 in the DRGS high score group had a significantly better prognosis than those in the low score group ($p = 0.011$, log-rank test, **Figure 5B**), consistent with the results from the TCGA dataset. However, there was no significant OS difference in GSE15459 ($p = 0.77$, **Figure 5C**) or GSE26899 ($p = 0.25$, **Figure 5D**) between the two DRGS groups.

Molecular Characteristics of Different 15-DNA Repair Gene Signature Score Groups

To identify the relationship between TMB and the DRGS score, we analyzed the somatic mutation data of 365 gastric adenocarcinoma samples from TCGA-STAD (**Supplementary Table S7**). The result indicated that the DRGS score was positively associated with TMB (rho = 0.50, $p < 0.001$, **Figure 6A**). TMB of samples in DRGS high score group ($n = 183$) was significantly higher than that in DRGS low score group ($n = 182$) ($p < 0.001$, Mann-Whitney U test, **Figure 6B**). All mutation events were divided into different categories, among which missense mutations accounted for the largest proportion in the classification of variation (**Supplementary Figure S2A, B** in the Supplementary Material), single nucleotide polymorphisms (SNPs) occurred much more frequently than insertion or deletion, and C > T was the most general of single nucleotide variants (SNV) in all TCGA-STAD samples. Notably, median variant per sample of DRGS high score group (121) was higher than that of DRGS low score group (59.5). Furthermore, mutation events of the top 10 mutated genes in each sample were shown in the waterfall plot. Amongst DRGS high score group, the mutation frequency of TTN was the highest (59%), followed by TP53 (54%) and MUC16 (39%, **Supplementary Figure S2B**). Amongst DRGS low score group, the highest mutation frequency gene was TP53 (34%), followed by TTN (33%) and MUC16 (21%, **Supplementary Figure S2A**).

PD-L1 expression is another important biomarker of susceptibility to PD-L1 blockade. We found that PD-L1 expression in DRGS high score group ($n = 183$) was significantly higher than in DRGS low score group ($n = 182$) ($p = 0.002$, **Figure 6D**, **Supplementary Table S7**). The DRGS score is positively correlated to PD-L1 expression (rho = 0.17, $p < 0.001$, **Figure 6C**).

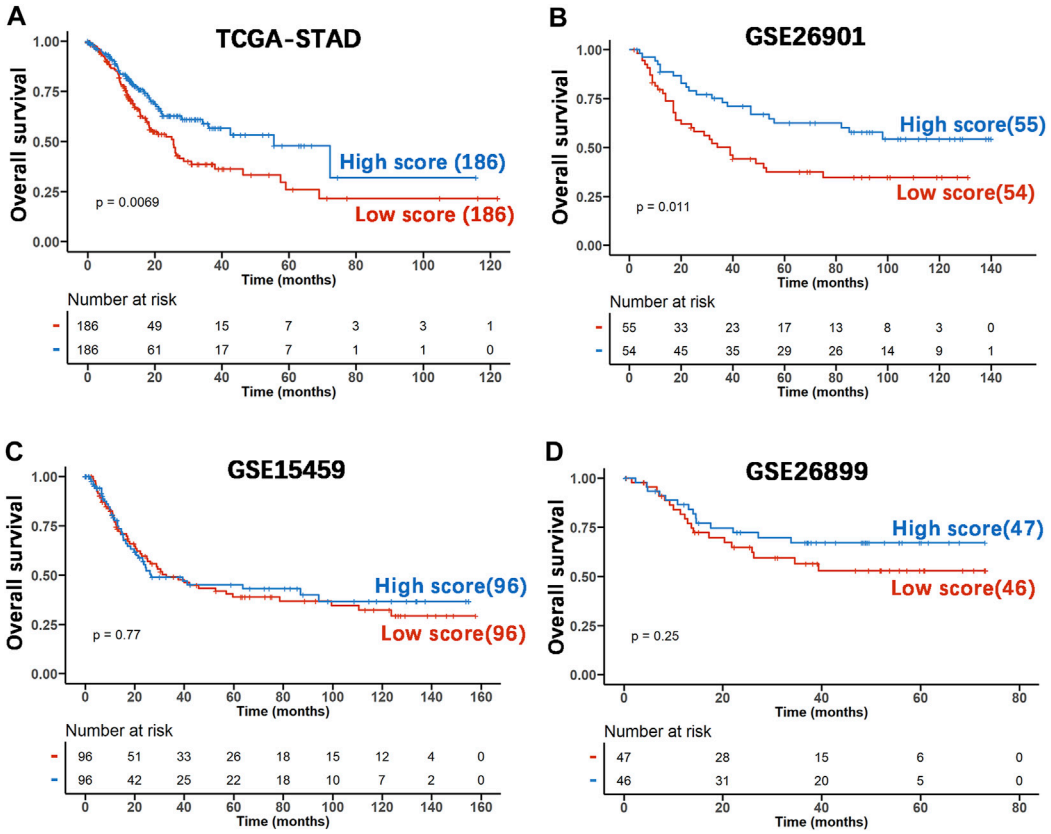


FIGURE 5 | The prognostic capability of the DRGS score system using various publicly available gastric cancer datasets. **(A)** Kaplan-Meier analysis and log-rank test of TCGA-STAD samples were used to compare the OS rate of the two score groups (High score $n = 186$, Low score $n = 186$); **(B)** The independent gastric cancer dataset GSE26901 (High score $n = 55$, Low score $n = 54$) as the validation set was analyzed using the same method as above; **(C)** the dataset GSE15459 (High score $n = 96$, Low score $n = 96$) was used as the validation set; **(D)** GSE26899 (High score $n = 47$, Low score $n = 46$) was used as the validation set.

Benefits of Immunotherapy With Programmed Cell Death Ligand 1 Blockers in Different 15-DNA Repair Gene Signature Score Groups

In view of the association between our score system and the immune microenvironment of the tumor, as described above, we tested the ability of the DRGS score system to predict the response of patients to immunotherapy. This analysis was based on the IMvigor210CoreBiologies cohort ($n = 348$), a large phase 2 trial investigating the clinical activity of PD-L1 blockade with atezolizumab in metastatic urothelial carcinoma (mUC) (Gibney et al., 2016) (**Supplementary Table S8**). We found that DRGS high score patients ($n = 174$) exhibited prominent prolonged overall survival ($p = 0.021$, **Figure 7A**). In addition, 298 patients in this cohort showed varying degrees of response to anti-PD-L1 blockers. DRGS high score patients ($n = 149$) comprised 14.8% complete response (CR), 18.8% partial response (PR), 18.8% stable disease (SD) and 47.6% progressive disease (PD), while DRGS low score patients ($n = 149$) comprised 2.0% CR, 10.1% PR, 23.5% SD and 64.4% PD. The Chi-squared test performed between DRGS high and low score groups also showed significantly better therapeutic outcomes in DRGS high score patients than the low score patients ($p < 0.001$, **Figure 7B**).

We also analyzed the three immune subtypes (immune inflamed, immune excluded and immune desert) of IMvigor210CoreBiologies in two DRGS group patients. DRGS high score patients ($n = 142$) comprised 33.1% inflamed, 45.8% excluded and 21.1% desert, while DRGS low score patients ($n = 142$) comprised 19.0% inflamed, 48.6% excluded and 32.4% desert. Therefore, the DRGS high score patients had a significantly higher percentage of "immune inflamed" tumors and a lower percentage of "immune desert" and "immune excluded" tumors compared to the low score patients ($p = 0.012$, Chi-square test, **Figure 7C**).

15-DNA Repair Gene Signature Score and its Potential Chemotherapeutic Value

To explore the effect of the DRGS score system on drug response in chemotherapy, we evaluated the association between the DRGS score and the response to drugs in gastric cancer cell lines ($n = 23$) in the Genomics of Drug Sensitivity in Cancer (GDSC) database. We identified 5 significantly correlated pairs between the DRGS score and drug sensitivity, all of which demonstrated drug

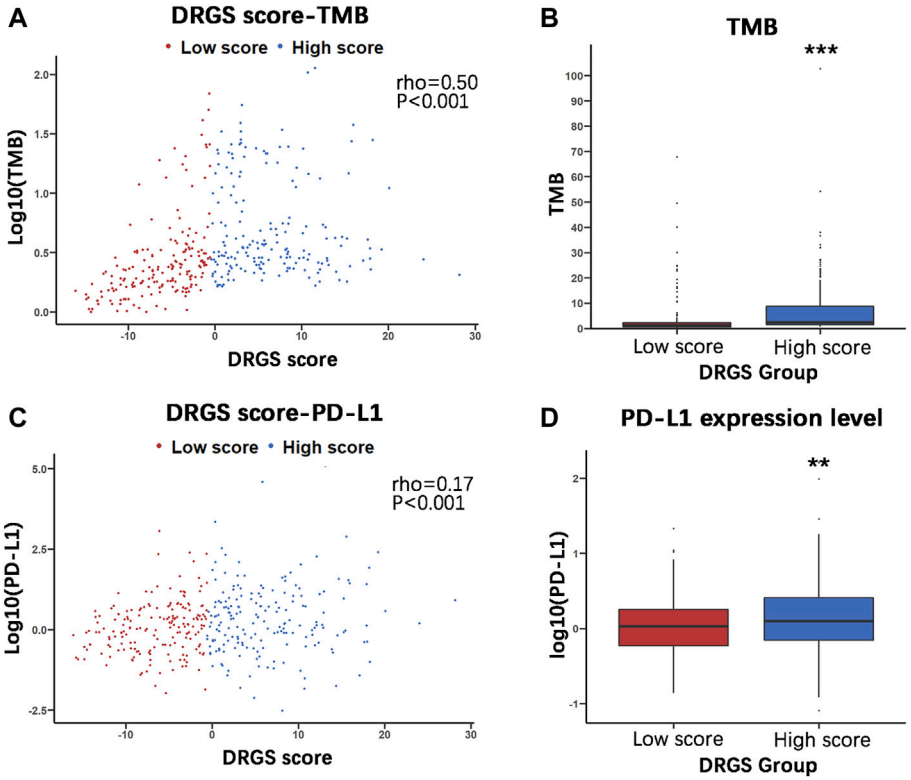


FIGURE 6 | Molecular characteristics of the two DRGS score groups (High score n = 183, Low score n = 182). **(A)** and **(C)** The Spearman's correlation was performed to analyze the relationship between the DRGS score and two important immunotherapy biomarkers, TMB and PD-L1; **(B)** and **(D)** The Mann Whitney U test was used to compare TMB and PD-L1 expression in high and low score groups in TCGA-STAD gastric cancer samples. (*p < 0.05, ** p < 0.01, *** p < 0.001).

resistance correlated with the DRGS score, including insulin-like growth factor receptor IGF1R_3801 (Rs = −0.53, p = 0.019), AKT inhibitor ipatasertib (Rs = −0.46, p = 0.028), surviving inhibitor sepantronium bromide (Rs = −0.54, p = 0.007), ULK1 protein kinase ULK1_4989 (Rs = −0.47, p = 0.043) and AKT inhibitor uprosertib (Rs = −0.50, p = 0.021) (**Figure 8A**). In addition, we found that these drugs mostly target PI3K/MTOR, IGF1R and apoptosis regulation signaling pathways (**Figure 8B**).

Immunohistochemistry Verification of Expression of Genes in 15-DNA Repair Gene Signature and Relationship With Programmed Cell Death Ligand 1 Expression

Through IHC analysis, we found in gastric tumor tissues (n = 10) that the protein expression of most of the genes in the DRGS, except for PARPBP, EME1, and RAD54L, was significantly increased (**Figure 9A**). In the HPA data, compared with normal tissues, the expression of DRGS except FANCA, UBE2T, POLQ, EXO1, and XRCC2 (They were not included) was remarkably upregulated (**Figure 9B**).

We also analysed the relationship between the DRGS and PD-L1 protein expression. As shown in **Figures 10A,B**, based on IHC analysis (n = 10), the protein expression of the genes in the DRGS, except for PARPBP, EME1, and RAD54L, was positively correlated to PD-L1 expression (rho>0.7, p < 0.05, **Figures 10A,B**). In the HPA data, the protein level of PRKDC, FANCI, LIG1, RECQL4, FANCA, PARPBP, RAD54L, and FANCD2 (The rest of the 15 genes were not included) was positively related to that of PD-L1 (**Figure 10C**). Thus, the above results verified the results obtained from bioinformatics analysis.

DISCUSSION

In the emerging immunotherapy strategies such as the ICB therapy in various types of cancer, the interaction of tumor DNA damage and repair landscape with patient immune system and related therapy has recently been revealed as a complex biological process (Fridman et al., 2012; Gentles et al., 2015; Mouw et al., 2017; Bever and Le, 2018; Joshi and Badgwell, 2021; Zhang et al., 2021). Nevertheless, many prognostic and predictive biomarkers are being evaluated clinically to identify criteria for establishing customized therapeutic approaches. This has been verified in cancers with MMR deficiency, which leads to microsatellite instability (MSI) phenotype, in which ICIs were found to be more efficient (Le et al., 2017; Bever and Le, 2018; Joshi and Badgwell, 2021).

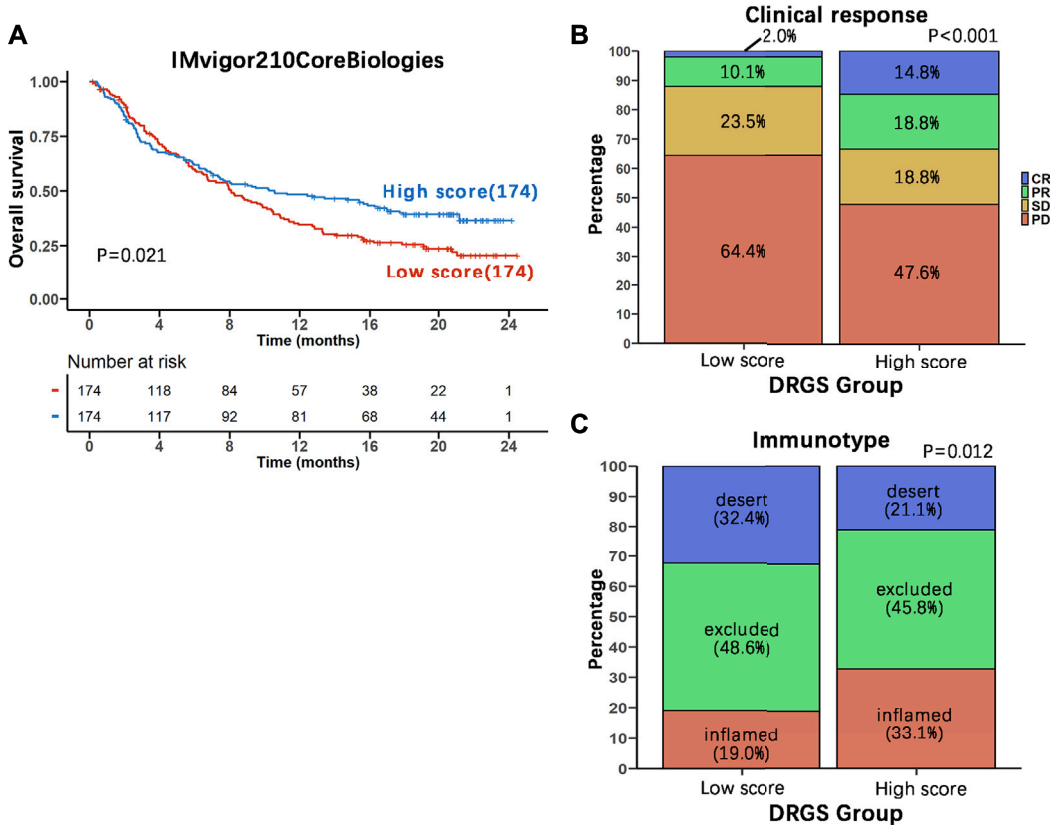


FIGURE 7 | Benefits of immunotherapy with PD-L1 blockers in the two DRGS score groups. **(A)** The 348 patients in IMvigor210CoreBiologies were equally divided into two groups based on their DRGS scores. Kaplan-Meier analysis of overall survival in the two groups was performed and the survival curves were compared using the log-rank test (High score *n* = 174, Low score *n* = 174); **(B)** Comparison of the clinical response to immunotherapy between the two DRGS groups (High score *n* = 149, Low score *n* = 149); **(C)** Comparison of the three immunotypes between the two DRGS groups of patients using the Chi-square test (High score *n* = 142, Low score *n* = 142).

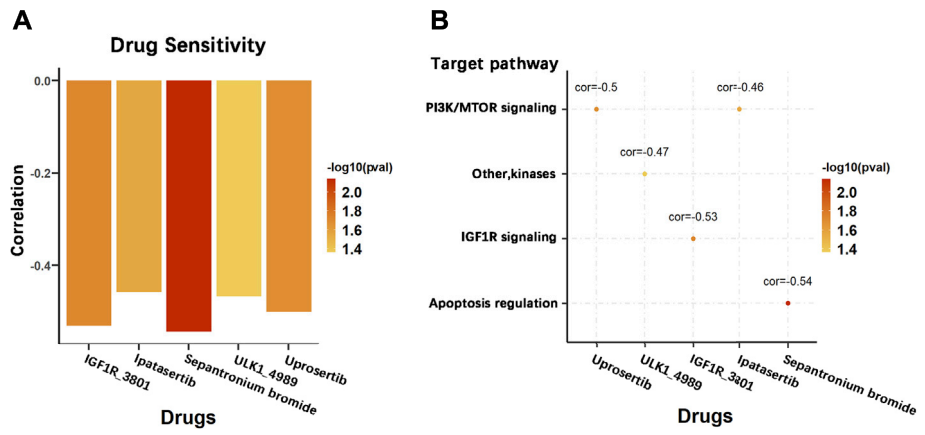


FIGURE 8 | Potential therapeutic value of the DRGS score in chemotherapy. **(A)** Pearson correlation analysis was used to calculate the correlation between the sensitivity level of different chemotherapeutic agents and the DRGS score of 23 types of gastric cancer cells; **(B)** Signalling pathways targeted by the drugs that are correlated with the DRGS score.

In this study, based on the known association of DNA damage and repair with immunotherapy response, we identified a 15-DNA repair gene signature (DRGS) using the gene expression data in TCGA-STAD and GSE 30727 datasets and developed the scoring system with the median score as the cut-point to divide gastric cancer patients into high and low score subtypes. We used

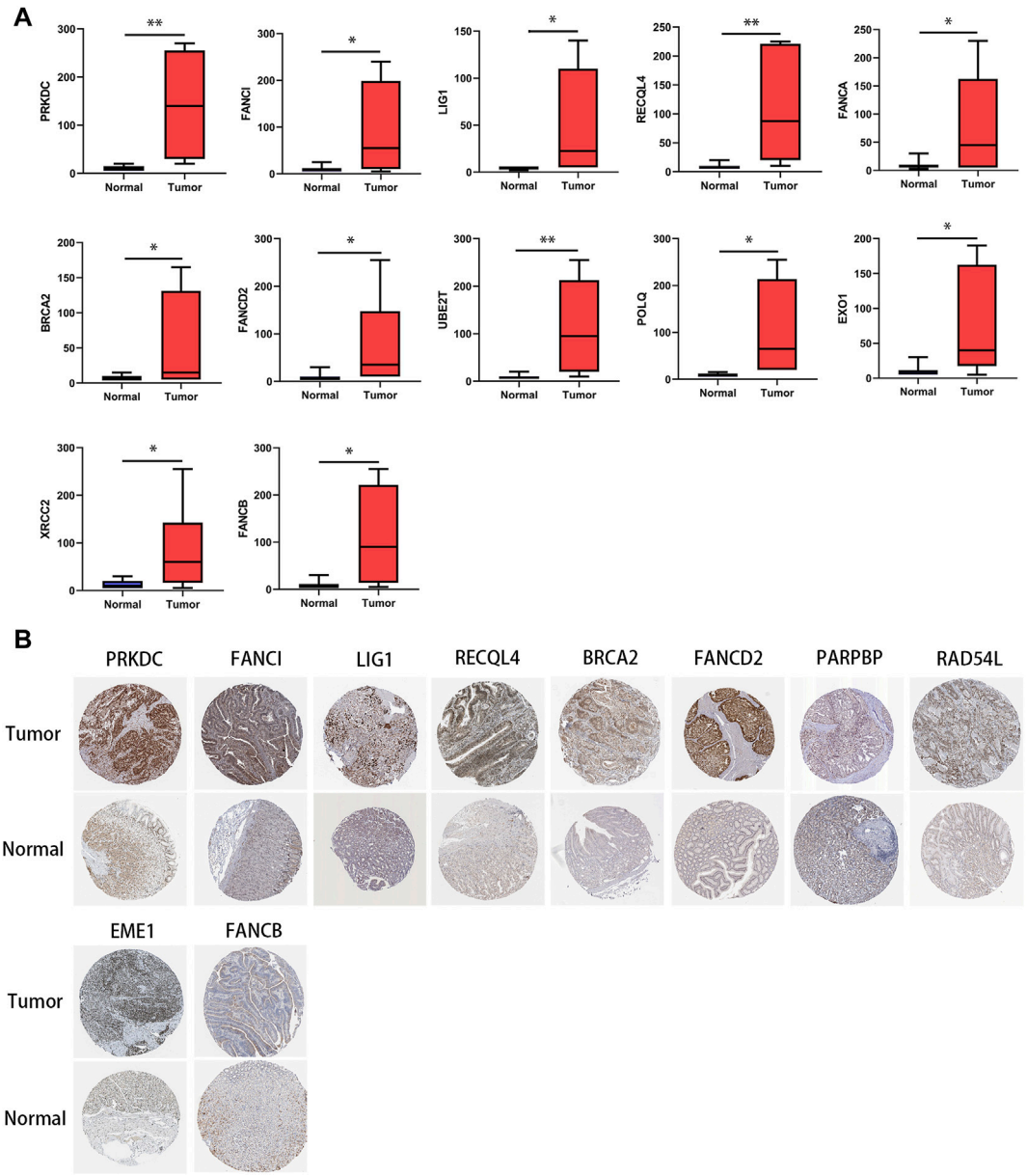


FIGURE 9 | Protein level of genes in the DRGS in gastric cancer tissues and normal tissues. **(A)** IHC (Tumor $n = 10$, Normal $n = 10$); (*$p < 0.05$, **$p < 0.01$). **(B)** HPA database.

the IMvigor210CoreBiologies cohort to assess the ability of the DRGS score to predict patient response to ICB therapy. The results demonstrated that DRGS high score patients showed significantly better therapeutic outcomes compared to the low score patients ($p < 0.001$), suggesting that the DRGS is a promising predictor of patient response to immune therapies.

The 15 genes in DRGS were screened from those that are significantly differentially expressed between tumor and normal tissues in the TCGA-STAD and GSE30727 datasets. Of the DRGS genes, six are associated with tolerance and repair of DNA crosslinks, including six genes from the FA pathway: FANCI, FANCA, FANCD2, FANCB, BRCA2 (FANCD1), and XRCC2. This fact may indicate a damage response to DNA crosslinks formed in the gastric cancer, although no information is available on a possibly higher level of such lesions in this disease. The FA pathway has also been linked to cancer susceptibility, with either sensitivity or resistance to chemotherapeutic agents. Also, four genes in the DRGS are in the HR pathway for the repair of DNA double-strand breaks, i.e., XRCC2, EME1, RAD54L, and BRCA2. This pathway is among the most well-studied of DNA repair defects in tumor immunotherapy response (Strickland et al., 2016; Mouw et al., 2017). It has been suggested that HR-deficient (HRD) tumors may possess enhanced immunogenicity, and thus, become more susceptible to checkpoint inhibitors (van Wilpe et al., 2021). The

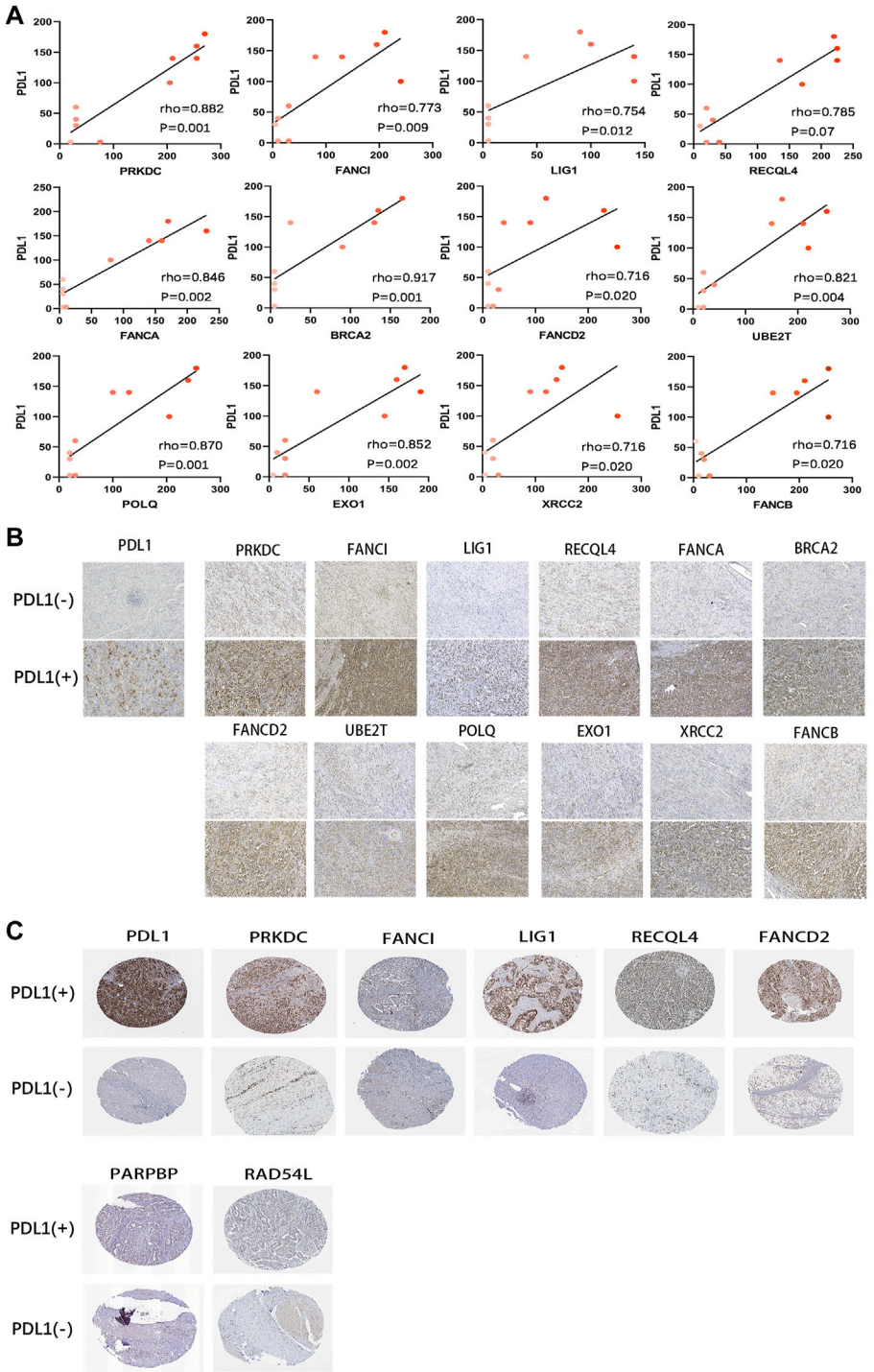


FIGURE 10 | Correlation of protein expression between PD-L1 and DRGS. **(A)** The Spearman's correlation was performed to analyze the relationship between PD-L1 and DRGS in IHC gastric cancer samples (n = 10). The color depth of the points represents the IHC score of DRGS; **(B)** Protein expression (IHC) of DRGS in PD-L1 (–) patients and PD-L1 (+) patients; **(C)** Data from the HPA database.

HPA database and IHC analysis were used to successfully verify the differential expression of these genes in human gastric cancer tissues. It should also be noted that in the DRGS high score group, the high expression level of the DNA repair genes does not mean an increased cellular repair capacity, rather, it could well be a damage response following the increased levels of DNA damage inside tumor cells. Therefore, a better understanding of the DNA damage in the tumor will help to know the underlying mechanism of the DRGS as a

biomarker for selection patients for immunotherapy. In addition, changes in gene expression levels and protein levels do not correlate well in most of the cases, mainly due to the regulation control at different levels.

It is known that the tumor immune microenvironment has great implications in gastric cancer progression and susceptibility to immunotherapy (Lazăr et al., 2018). Given that a multifaceted condition may affect gastric cancer primed for response to immunotherapy, we analyzed several important factors or potential immunotherapy response biomarkers in relation to the DRGS high and low score groups using TCGA-STAD, including tumor-infiltrating lymphocytes, immune subtypes, TMB, PD-L1 expression and other histologic and molecular classifications (detailed in Results). In TCGA-STAD samples analyzed by ImmuCellAI, we found that anti-tumor cells including γδ T cells, neutrophils and Th1 cells were more abundant in the DRGS high score samples, while several types of immunosuppressive or tumor enhancement cells were more abundant in the low-score patients. Next, we analyzed the somatic mutation and PD-L1 expression data in TCGA-STAD and found that both were positively associated with the score, suggesting that DRGS high score patients are more likely to respond to immunotherapy. We also analyzed the three immune subtypes of IMvigor210CoreBiologies and found that the DRGS high score patients were associated with a significantly higher percentage of "immune inflamed" subtype, which has been shown to be infiltrated by a number of subtypes of immune cells (Chen and Mellman, 2017). Finally, the DRGS high score group was shown to have a higher percentage of MSI-H phenotype in the TCGA subtypes. MSI-H is characterized by a better survival (Yamamoto et al., 1999) and a high number of microsatellite mutations (Yoon et al., 2013). Recent clinical trials have confirmed that MSI-positive gastric tumors are sensitive to ICB therapy (Le et al., 2015). In summary, our DRGS high-score gastric cancer patients are characterized by "immune inflamed", "intestinal", and "MSI-H" phenotypes, which indicate an abundant immune cell infiltration, high expression of PD-L1, and high number of microsatellite mutations. All of which are associated with higher sensitivity and better response to immunotherapy. The verification of correlation between DRGS and PD-L1 protein expression with HPA database and our IHC results were consistent with the above observations.

There were also limitations in this work. We used the data from a different type of tumor (IMvigor210CoreBiologies cohort on locally advanced or metastatic urothelial bladder cancer) to evaluate the ability of the DRGS score to predict PD-L1 blockade-based immunotherapy response, because of a lack of such a dataset in gastric cancer. Future studies will use more data from the gastric cancer patients, including those from our own hospital.

CONCLUSION

The DRGS is a promising predictor of patient response to immune checkpoint inhibitor (ICI)-based therapy. Further studies are needed to evaluate the clinical utility of the DRGS for tailoring immunotherapy.

DATA AVAILABILITY STATEMENT

The data used for analysis in this study are available from TCGA-STAD, GEO, GDSC, CPTAC, HPA and IMvigor210CoreBiologies databases. All the original codes have been uploaded to Zenodo website. The link is https://doi.org/10.5281/zenodo.6466144.

AUTHOR CONTRIBUTIONS

BY, CJ, LC, and LW contributed equally to this study. Conceptualization and design, PW, BH, XZ, and JM; Data curation and analysis, BY, CJ, LC, LW, JC, LZ, YC, JM, BH and PW; Funding acquisition, PW; Project administration, LC; Supervision, PW and XZ; Validation, BY, YC, and JM; Writing, BY, BH; Editing, PW, JC, MC, SZ, YC, JM and XZ. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research was funded by 2019 "Six Talent Peaks" Project of Jiangsu Province to P.W. (grant no. WSN-136).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2022.893546/full#supplementary-material

Supplementary Figure S1 | PCA analysis was used as a dimensionality reduction method to obtain the first-dimension correlation coefficients of 15 repair genes in IMvigor210CoreBiologies, GSE26901, GSE15459, GSE26899, and gastric cancer cell lines (from A to E).

Supplementary Figure S2 | Overview of TCGA-STAD mutations. Statistics and visual display of mutation type, SNV classification, TMB and mutation classification in DRGS low score subtype **(A)** versus high score subtype **(B)** samples. The Waterfall plot of the top 10 mutant genes, which classifies the mutations of each gene in each sample.

REFERENCES

Abbott, B. P., Abbott, R., Abbott, T. D., Abernathy, M. R., Acernese, F., Ackley, K., et al. (2016). Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.* 116 (6), 061102. doi:10.1103/PhysRevLett.116.061102

Balar, A. V., Galsky, M. D., Rosenberg, J. E., Powles, T., Petrylak, D. P., Bellmunt, J., et al. (2017). Atezolizumab as First-Line Treatment in Cisplatin-Ineligible Patients with Locally Advanced and Metastatic Urothelial Carcinoma: a Single-Arm, Multicentre, Phase 2 Trial. *Lancet* 389 (10064), 67–76. doi:10.1016/s0140-6736(16)32455-2

Bever, K. M., and Le, D. T. (2018). DNA Repair Defects and Implications for Immunotherapy. *J. Clin. Invest.* 128 (10), 4236–4242. doi:10.1172/jci122010

Chalmers, Z. R., Connelly, C. F., Fabrizio, D., Gay, L., Ali, S. M., Ennis, R., et al. (2017). Analysis of 100,000 Human Cancer Genomes Reveals the Landscape of

Tumor Mutational Burden. *Genome Med.* 9 (1), 34. doi:10.1186/s13073-017-0424-2

Chao, J., Fuchs, C. S., Shitara, K., Tabernero, J., Muro, K., Van Cutsem, E., et al. (2020). Pembrolizumab (Pembro) in Microsatellite Instability-High (MSI-H) Advanced Gastric/gastroesophageal Junction (G/GEJ) Cancer by Line of Therapy. *Jco* 38 (4_Suppl. l), 430. doi:10.1200/jco.2020.38.4_suppl.430

Chen, D. S., and Mellman, I. (2017). Elements of Cancer Immunity and the Cancer-Immune Set Point. *Nature* 541 (7637), 321–330. doi:10.1038/nature21349

Fridman, W. H., Pagès, F., Sautès-Fridman, C., and Galon, J. (2012). The Immune Contexture in Human Tumours: Impact on Clinical Outcome. *Nat. Rev. Cancer* 12 (4), 298–306. doi:10.1038/nrc3245

Friedberg, E. C., Walker, G. C., and Si Ede, W. (2015). A History of the DNA Repair and Mutagenesis Field. *Dna Repair* 33 (2), 35–42. doi:10.1016/j.dnarep.2015.06.007

Fuchs, C. S., Özgüroğlu, M., Bang, Y.-J., Di Bartolomeo, M., Mandala, M., Ryu, M.-H., et al. (2021). Pembrolizumab versus Paclitaxel for Previously Treated PD-L1-Positive Advanced Gastric or Gastroesophageal Junction Cancer: 2-year Update of the Randomized Phase 3 KEYNOTE-061 Trial. *Gastric Cancer* 25, 197–206. doi:10.1007/s10120-021-01227-z

Gentles, A. J., Newman, A. M., Liu, C. L., Bratman, S. V., Feng, W., Kim, D., et al. (2015). The Prognostic Landscape of Genes and Infiltrating Immune Cells across Human Cancers. *Nat. Med.* 21 (8), 938–945. doi:10.1038/nm.3909

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

Hang, B. (2004). Repair of Exocyclic DNA Adducts: Rings of Complexity. *Bioessays* 26 (11), 1195–1208. doi:10.1002/bies.20130

Harada, K., Baba, H., and Ajani, J. A. (2018). Recent Trend in Gastric Cancer Treatment in the USA. *Jcmt* 4, 18. doi:10.20517/2394-4722.2017.74

Helleday, T., Petermann, E., Lundin, C., Hodgson, B., and Sharma, R. A. (2008). DNA Repair Pathways as Targets for Cancer Therapy. *Nat. Rev. Cancer* 8 (3), 193–204. doi:10.1038/nrc2342

Joshi, S. S., and Badgwell, B. D. (2021). Current Treatment and Recent Progress in Gastric Cancer. *CA A Cancer J. Clin.* 71 (3), 264–279. doi:10.3322/caac.21657

Lazăr, D. C., Avram, M. F., Romoşan, I., Cornianu, M., Tăban, S., and Goldiş, A. (2018). Prognostic Significance of Tumor Immune Microenvironment and Immunotherapy: Novel Insights and Future Perspectives in Gastric Cancer. *Wjg* 24 (32), 3583–3616. doi:10.3748/wjg.v24.i32.3583

Le, D. T., Durham, J. N., Smith, K. N., Wang, H., Bartlett, B. R., Aulakh, L. K., et al. (2017). Mismatch Repair Deficiency Predicts Response of Solid Tumors to PD-1 Blockade. *Science* 357 (6349), 409–413. doi:10.1126/science.aan6733

Le, D. T., Uram, J. N., Wang, H., Bartlett, B. R., Kemberling, H., Eyring, A. D., et al. (2015). PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N. Engl. J. Med.* 372 (26), 2509–2520. doi:10.1056/NEJMoa1500596

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

Miao, Y. R., Zhang, Q., Lei, Q., Luo, M., Xie, G. Y., Wang, H., et al. (2020). ImmuCellAI: A Unique Method for Comprehensive T-Cell Subsets Abundance Prediction and its Application in Cancer Immunotherapy. *Adv. Sci.* 7 (7), 1902880. doi:10.1002/advs.201902880

Moehler, M., Shitara, K., Garrido, M., Salman, P., Shen, L., Wyrwicz, L., et al. (2020). LBA6_PR Nivolumab (Nivo) Plus Chemotherapy (Chemo) versus Chemo as First-Line (1L) Treatment for Advanced Gastric Cancer/gastroesophageal Junction Cancer (GC/GEJC)/esophageal Adenocarcinoma (EAC): First Results of the CheckMate 649 Study - ScienceDirect. *Ann. Oncol.* 31. doi:10.1016/j.annonc.2020.08.2296

Mouw, K. W., Goldberg, M. S., Konstantinopoulos, P. A., and D'Andrea, A. D. (2017). DNA Damage and Repair Biomarkers of Immunotherapy Response. *Cancer Discov.* 7 (7), 675–693. doi:10.1158/2159-8290.Cd-17-0226

Rhyu, M. G., Park, W. S., and Meltzer, S. J. (1994). Microsatellite Instability Occurs Frequently in Human Gastric Carcinoma. *Oncogene* 9 (1), 29–32.

Rodríguez, A., and D'Andrea, A. (2017). Fanconi Anemia Pathway. *Curr. Biol.* 27 (18), R986–r988. doi:10.1016/j.cub.2017.07.043

Strickland, K. C., Howitt, B. E., Shukla, S. A., Rodig, S., Ritterhouse, L. L., Liu, J. F., et al. (2016). Association and Prognostic Significance of BRCA1/2-Mutation Status with Neoantigen Load, Number of Tumor-Infiltrating Lymphocytes and Expression of PD-1/pd-L1 in High Grade Serous Ovarian Cancer. *Oncotarget* 7 (12), 13587–13598. doi:10.18632/oncotarget.7277

The Cancer Genome Atlas Research Network (2014). Comprehensive Molecular Characterization of Gastric Adenocarcinoma. *Nature* 513 (7517), 202–209. doi:10.1038/nature13480

van Wilpe, S., Tolmeijer, S. H., Koornstra, R. H. T., de Vries, I. J. M., Gerritsen, W. R., Ligtenberg, M., et al. (2021). Homologous Recombination Repair Deficiency and Implications for Tumor Immunogenicity. *Cancers* 13 (9), 2249. doi:10.3390/cancers13092249

Wood, R. D., Mitchell, M., and Lindahl, T. (2005). Human DNA Repair Genes, 2005. *Mutat. Research/Fundamental Mol. Mech. Mutagen.* 577, 275–283. doi:10.1016/j.mrfmmm.2005.03.007

Yamamoto, H., Perezpiteira, J., Yoshida, T., Terada, M., Itoh, F., Imai, K., et al. (1999). Gastric Cancers of the Microsatellite Mutator Phenotype Display Characteristic Genetic and Clinical Features☆, ☆☆. *Gastroenterology* 116 (6), 1348–1357. doi:10.1016/s0016-5085(99)70499-3

Yoon, K., Lee, S., Han, T.-S., Moon, S. Y., Yun, S. M., Kong, S.-H., et al. (2013). Comprehensive Genome- and Transcriptome-wide Analyses of Mutations Associated with Microsatellite Instability in Korean Gastric Cancers. *Genome Res.* 23 (7), 1109–1117. doi:10.1101/gr.145706.112

Yuan, B., Jiang, C., Chen, L., Cui, J., Chen, M., Zhang, S., et al. (2021). A 15-DNA Repair Gene Prognostic Signature for Immunotherapy in Gastric Cancer. *Preprints* 100297, 2021. doi:10.20944/preprints202110.0297.v1

Zhang, C., Li, D., Yu, R., Li, C., Song, Y., Chen, X., et al. (2021). Immune Landscape of Gastric Carcinoma Tumor Microenvironment Identifies a Peritoneal Relapse Relevant Immune Signature. *Front. Immunol.* 12, 651033. doi:10.3389/fimmu.2021.651033

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

BER base excision repair
CIN chromosomal instability
CR complete response
DRGS 15-DNA repair gene signature
DSB double-strand break
GDSC Genomics of Drug Sensitivity in Cancer
GS genomically stable
HR homologous recombination
HRD homologous recombination deficiency
ICB immune checkpoint blockade
ICI immune checkpoint inhibitors
IHC Immunohistochemistry
MMR mismatch repair
MSI microsatellite instability
MSI-H microsatellite instability-high
MSI-L microsatellite instability-low
MSS Microsatelite Instability-Stable
mUC metastatic urothelial carcinoma
NER nucleotide excision repair
NHEJ non-homologous end joining
OS overall survival
PCA principal component analysis
PD progressive disease
PD-1 programmed cell death protein 1
PD-L1 programmed cell death ligand 1
PR partial response
RFS relapse-free survival
SD stable disease
SNV single nucleotide variants
TCGA-STAD The Cancer Genome Atlas Stomach Adenocarcinoma
TMB tumor mutation burden

frontiers | Frontiers in Cell and Developmental Biology

ORIGINAL RESEARCH
published: 23 May 2022
doi: 10.3389/fcell.2022.840171

Integrated Analysis Reveals the Gut Microbial Metabolite TMAO Promotes Inflammatory Hepatocellular Carcinoma by Upregulating POSTN

Yonglin Wu[1†], Xingyu Rong[1†], Miaomiao Pan[1†], Tongyao Wang[1], Hao Yang[1], Xiejiu Chen[1], Zhenming Xiao[1] and Chao Zhao[1,2]*

[1]Key Laboratory of Medical Molecular Virology (MOE/NHC/CAMS), School of Basic Medical Sciences, Shanghai Medical College and National Clinical Research Center for Aging and Medicine, Shanghai Medical College, Huashan Hospital, Fudan University, Shanghai, China, [2]Shanghai Frontiers Science Center of Pathogenic Microbes and Infection, Shanghai, China

Liver cancer has a high mortality rate. Chronic inflammation is one of the leading causes of hepatocellular carcinoma. Recent studies suggested high levels of trimethylamine N-oxide (TMAO) may correlate with increased risk of inflammatory-induced liver cancer. However, the mechanisms by which TMAO promotes liver cancer remain elusive. Here, we established a model of inflammatory-induced liver cancer by treating Hepa1-6 cells and Huh7 cells with TNF-α. TMAO synergistically increased the proliferation, migration and invasion of Hepa1-6 cells and Huh7 cells in the presence of TNF-α. We conducted bulk RNA-Seq of the TMAO-treated cell model of inflammatory Hepatocellular carcinoma (HCC) and evaluated the influence of the differentially expressed genes (DEGs) on clinical prognosis using Kaplan-Meier Plotter Database and Gene Expression Profiling Interactive Analysis (GEPIA) database. Univariate and multivariate Cox regression analyses of tumor microenvironment and DEGs were performed using Timer2.0. Upregulation of *POSTN*, *LAYN* and *HTRA3* and downregulation of *AANAT* and *AFM* were positively related to poorer overall survival in human liver cancer. Moreover, higher expression of *POSTN* and *HTRA3* positively correlated with infiltration of neutrophils, which can promote tumor progression. *In vitro* experiments showed TMAO activates ILK/AKT/mTOR signaling via *POSTN*, and knocking down *POSTN* significantly reduced ILK/AKT/mTOR signaling and the tumorigenicity of Hepa1-6 cells and Huh7 cells. Collectively, our results suggest the gut microbial metabolite TMAO and *POSTN* may represent potential therapeutic targets for liver cancer.

Keywords: microbial metabolite, TMAO, inflammatory, hepatocellular carcinoma, RNA-seq, database reanalysis

OPEN ACCESS

Edited by:
Lu Xie,
Shanghai Center for Bioinformation Technology, China

Reviewed by:
Shamir Cassim,
Institut Bergonié, France
Zhiqiang Qin,
University of Arkansas for Medical Sciences, United States

***Correspondence:**
Chao Zhao
czhao@fudan.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: *20 December 2021*
Accepted: *12 January 2022*
Published: *23 May 2022*

Citation:
Wu Y, Rong X, Pan M, Wang T, Yang H, Chen X, Xiao Z and Zhao C (2022) Integrated Analysis Reveals the Gut Microbial Metabolite TMAO Promotes Inflammatory Hepatocellular Carcinoma by Upregulating POSTN. Front. Cell Dev. Biol. 10:840171. doi: 10.3389/fcell.2022.840171

INTRODUCTION

Although liver cancer is the sixth most common cancer and has the third highest cancer mortality rate (Sung et al., 2021). Hepatocellular carcinoma (HCC) is the most common type of primary liver cancer (PLC), accounting for 70%–80% of cases (Siegel et al., 2018). The first-line treatment option for HCC is liver transplantation. Local ablation and external radiation are the standard of care for patients who are not suitable for surgery. Sorafenib and lenvatinib are the standard first-line systemic therapies for HCC, but are only indicated for patients with well-preserved liver function (EASL, 2018).

HCC is strongly associated with alcohol consumption, viral hepatitis such as hepatitis B virus (HBV) or hepatitis C virus (HCV), and nonalcoholic fatty liver disease (NAFLD) (Ghouri et al., 2017; Rong et al., 2019). All of these factors alter the hepatic immune system and induce severe chronic liver inflammation via multiple pathways and cytokine signaling molecules (Cui et al., 2017; Yang YM. et al., 2019). Tumor necrosis factor-alpha (TNF-α) is a pro-inflammatory Th1 cytokine that plays critical roles in the inflammatory mechanisms implicated in HCC (Jing et al., 2018). A recent study reported that TNF-α induces activation of the TNFR2-hnRNPK-YAP signaling axis and promotes primary liver tumorigenesis, confirming the important role of TNF-α in the pathogenesis of malignancies and inflammatory autoimmune diseases (Meng et al., 2021). Thus, TNF-α functions as a critical regulator of the induction of inflammatory HCC.

Trimethylamine N-oxide (TMAO) is an oxidation product of trimethylamine (TMA) and a pro-atherosclerotic metabolite. The role of TMAO in the relationship between metabolism of dietary compounds by the microbiota and cardiovascular disease has been investigated. When choline and other trimethylamine-containing species are ingested, TMA is initially formed by initial catabolism by intestinal microbes, and then efficiently metabolized by the hepatic flavin monooxygenase (FMO) family of enzymes to form TMAO (Wang et al., 2011). TMAO is also strongly implicated in development of the chronic inflammatory disease atherosclerosis (AS). Elevated TMAO levels induce activation of the NF-kappa B (NF-κB) pathway, which contributes to the regulation of many AS-related genes (Baker et al., 2011; Ma et al., 2017; Cheng et al., 2019). In addition, TMAO also increases the expression of pro-inflammatory genes, such as inflammatory factors, adhesion molecules and chemokines, in various models (Rohrmann et al., 2016; Chen ML. et al., 2017). TMAO can also activate the NLRP3 inflammasome and induce oxidative stress (Sun et al., 2016; Boini et al., 2017). Thus, TMAO undeniably contributes to the occurrence of chronic inflammation.

Numerous investigations have strongly linked TMAO to colorectal cancer and other malignant diseases (Bae et al., 2014; Chen et al., 2016; Shan et al., 2017). However, few studies have explored the association between TMAO and HCC. A recent case-control study indicated a relationship between TMAO and its precursor choline with HCC, with higher TMAO concentrations associated with increased risk of HCC (Liu et al., 2018). Here, used Hepa1-6 and Huh7 cells to establish the inflammatory hepatocellular carcinoma model promoted by TNF-α to explore the relationship between TMAO and the pathogenesis of inflammatory HCC promoted by TNF-α.

MATERIALS AND METHODS

Cell Lines

The mouse HCC cell lines Hepa1-6 was purchased from American type culture collection (ATCC) and human HCC cell line Huh7 was purchased from JCRB cell bank which were both maintained in Dulbecco's modified Eagle's medium (DMEM) with 10% fetal bovine serum (FBS).

TNF-α and TMAO

Recombinant mouse TNF-α and recombinant human TNF-α were purchased from PeproTech and dissolved in FBS. TMAO was purchased from Sigma-Aldrich and dissolved in DMEM with 10% FBS.

Cell Viability Assay

Cell viability was evaluated using the Cell Counting Kit-8 (CCK-8; Dojindo, Shanghai, China) in accordance with the manufacturer's instructions. Briefly, cells (2×10^3) were seeded into 96-well plates and cultured for 24, 72 or 120 h. To generate a growth curve, 10 μL CCK-8 solution was added to each well, incubated at 37°C protected from light for 4 h, and optical density was measured at 450 nm.

Wound Healing Assay

Cells (2×10^5) were seeded into 12-well plates and cultured to produce confluent monolayers. Wound areas were scraped using 200 μL pipette tips, washed three times with PBS to remove cellular debris and then the cells were cultured in DMEM with 10% FBS. Wound closure was observed and photographed between 0 and 48 h under an inverted microscope.

Migration Assays

The migration assay was performed using Transwell chambers (cat. no. 3422; Corning Incorporated, Shanghai, China). Briefly, 8×10^4 cells were centrifuged, resuspended in 250 μL serum-free media and placed into the upper chamber, and 750 μL of media containing 10% FBS was added to the lower chamber. The plates were incubated at 37°C for 24 h, then the culture media in the upper chamber was discarded, and the cells in the lower chamber were fixed with 600 μL of 4% paraformaldehyde for 20 min, washed with PBS, and stained in 400 μL of 0.1% crystal violet for 10 min. The numbers of cells that had migrated through the polycarbonate membrane at 6 h and invaded at 48 h were observed in at least five randomly selected fields under an inverted light microscope.

Transcriptome Sequencing

Total RNA was extracted, and mRNA was isolated using Oligo Magnetic Beads and fragmented for cDNA synthesis. RNA-seq cDNA libraries were prepared and sequenced on an Illumina sequencing platform in 2*150 sequencing mode.

Gene Correlation Analysis in GEPIA

The online database Gene Expression Profiling Interactive Analysis (GEPIA) (http://gepia.cancer-pku.cn/index.html) is an interactive web resource that includes RNA sequencing and patient outcome data for 9,736 tumor and 8,587 normal samples from the TCGA and GTEx projects. GEPIA was used to generate overall survival (OS) based on gene expression levels in HCC; the curves were compared using the log-rank test and Mantel-Cox test. Gene expression correlation analysis was also

performed for selected TCGA expression datasets, using the tumor and normal tissue datasets.

Kaplan-Meier Plotter Database Analysis

Kaplan-Meier plotter can be used to assess the effects of 54,675 genes on patient survival in 10,461 cancer samples (http://kmplot.com/analysis/). The correlations between *POSTN*, *NAPB*, *LAYN*, *HTRA3*, *AFM* and *AANAT* expression and patient survival in liver cancer were analyzed with Kaplan-Meier plotter; using the best cut-off analysis for stratification of HCC patients; the hazard ratios (HR) with 95% confidence intervals and log-rank p-values were computed.

TIMER 2.0 Database Analysis

TIMER 2.0 is a comprehensive resource for systematic analysis of immune infiltrates across diverse cancer types (https://cistrome.shinyapps.io/timer/). TIMER 2.0 applies deconvolution via a previously published statistical method to infer the abundance of tumor-infiltrating immune cells (TIICs) from gene expression profiles. The TIMER 2.0 database includes 10,897 samples across 32 cancer types from The Cancer Genome Atlas (TCGA) and can be used to estimate the abundance of immune infiltrates. We analyzed the correlations between the levels of infiltration of neutrophils, Tregs, activated NK cells, and $CD8^+$ T cells and the expression levels of *POSTN, NAPB, LAYN, HTRA3, AANAT,* and *AFM* in liver cancer.

Quantitative Real-Time PCR

Cellular and exosomal RNAs were isolated using Eastep® Super Total RNA Extraction Kit (Promega, Madison, Wisconsin United States). First-strand cDNA was synthesized with random primers using Eastep® RT Master Mix (Promega). QPCR was performed with GoTaq® qPCR Master Mix (Promega) on a CFX384 Real-Time PCR Detection System (Bio-Rad, Hercules, California, United States). The primers were synthesized by BGI Genomics (Shanghai, China) and are listed in Additional file: **Supplementary Table S1**. Relative quantification was performed with the $2^{-\Delta\Delta Ct}$ method.

Western Blotting

Total cellular proteins or exosomes were extracted with RIPA buffer (Sigma-Aldrich, St. Louis, Missouri, United States) and protein concentrations were quantified using the BCA assay (Pierce, Rockford, IL, United States). Total proteins were separated on polyacrylamide gels, transferred onto PVDF membranes, blocked with TBS-T solution containing 5% BSA for 1 h, then the membranes were incubated with primary antibodies against POSTN (sc-398631, Santa Cruz; 1:500), NAPB (A18223, ABclonal; 1:1000), HTRA3 (NB600-1151, Novus Biologicals; 1:1000), LAYN (ab192610, Abcam; 1:1000), ILK (3862, Cell Signaling Technology; 1:1000), AKT (4691, Cell Signaling Technology; 1:1000), p-AKT (4060, Cell Signaling Technology; 1:1000), mTOR (2983, Cell Signaling Technology; 1:1000) and p-mTOR (5536, Cell Signaling Technology; 1:1000). Membranes were incubated with secondary antibodies including Anti-rabbit IgG (7074, Cell Signaling Technology; 1:10000) and Goat Anti-Mouse IgG H&L (ab6789, Abcam; 1:10000) for 1 h at room temperature. The western blots were visualized using an enhanced chemiluminescence system.

Production of Postn-Knockdown Recombinant Lentivirus Particles

The sh-*postn*-containing transfer vectors sh-*postn*-1, sh-*postn*-2 and sh-*postn*-3 were purchased from Shanghai Genechem Co., LTD. and co-transfected into 293T cells using Lipofectamine™ 3000, together with the pMD2G and psPAX2 plasmids. Forty-eight hours later, the supernatant was collected and replaced with fresh media, and the supernatant was collected again 72 h later. The supernatants containing the lentivirus particles were gently mixed, centrifuged at 4,000 g for 5 min at 4°C, and stored at −70°C for use in subsequent experiments.

Knockdown of Postn in Hepa1-6 Cells and Huh7 Cells

Hepa1-6 cells (2×10^5) and Huh7 cells were seeded into 12-well plates and cultured to 70%–90% confluence, then the virus supernatant containing the lentivirus particles was added and the cells were incubated for 24 h Hepa1-6 cells and Huh7 cells infected with the *Postn*-knockdown lentivirus were selected by culture in puromycin for 2 weeks.

Statistical Analysis

The Student's *t*-test was used to compare groups for normally distributed data; otherwise, the nonparametric Mann-Whitney test was adopted. One-way ANOVA was applied to compare the differences among three groups. For survival analysis, univariate analysis was conducted by the Kaplan-Meier method with the log-rank test, and multivariate analysis was performed by the stepwise Cox multivariate proportional hazard regression model (Forward LR, likelihood ratio). All *in vitro* experiments were replicated three times. Statistical analyses were performed using Graphpad prism 8.0 software, all the tests are two-sided; p values <0.05 were considered statistically significant.

RESULTS

TMAO Synergistically Enhances the Tumorigenicity of Hepa1-6 Cells in the Presence of TNF-α

TNF-α is a proinflammatory cytokine that has been linked to the initiation and development of HCC. Indeed, we found that TNF-α dose-dependently promoted the proliferation of Hepa1-6 mouse hepatoma cells, with the highest proliferation observed at 50 ng/ml TNF-α (**Figure 1A**). To investigate whether TNF-α promotes Hepa1-6 cell proliferation in an inflammatory state, we measured the levels of the inflammatory hepatocellular carcinoma-related cytokine IL-6 in Hepa1-6 cells treated with 50 ng/ml TNF-α. The levels of IL-6 were higher in Hepa1-6 cells treated with 50 ng/ml TNF-α compared to untreated cells (**Figure 1B**).

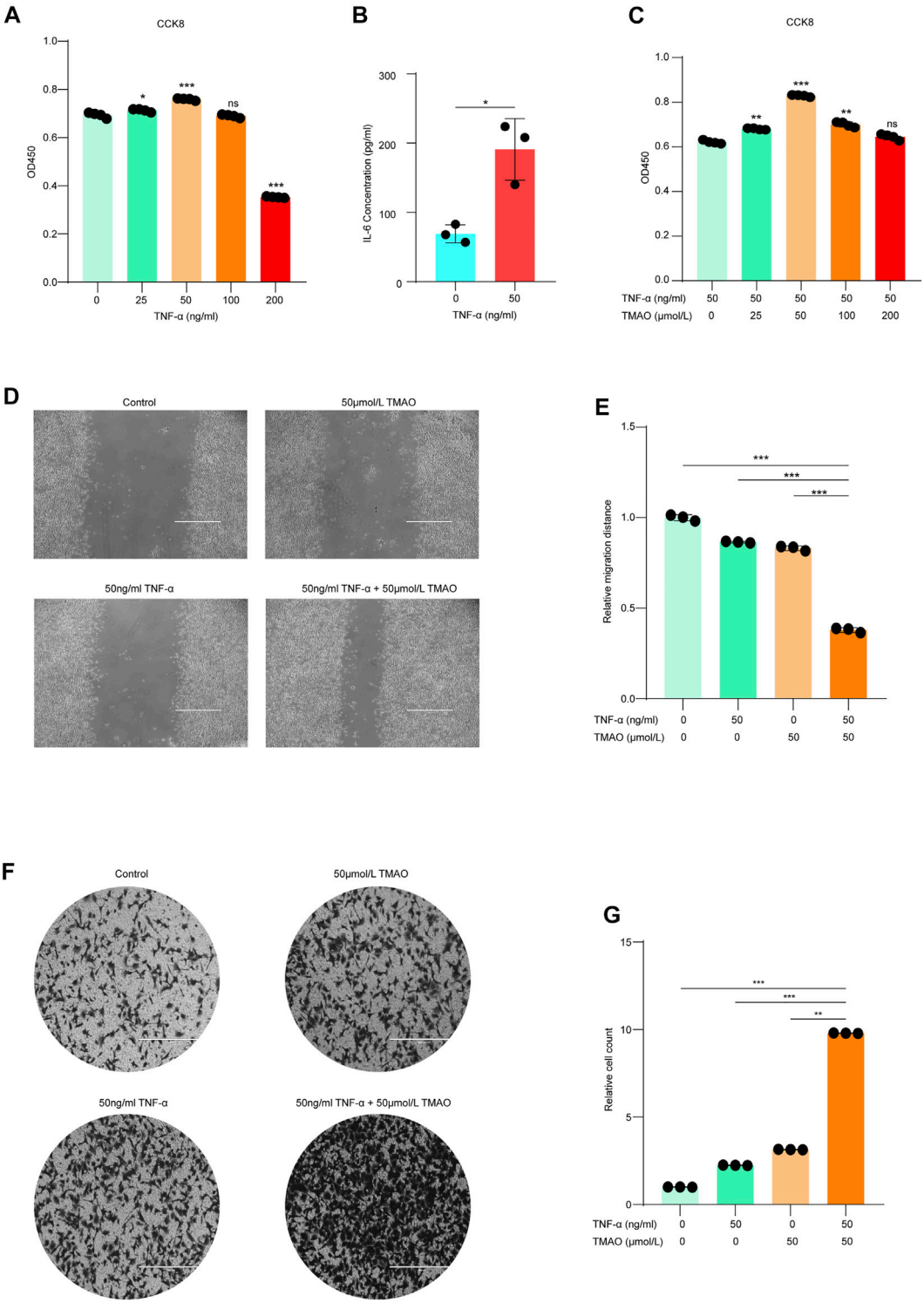


FIGURE 1 | TMAO synergistically enhances the tumorigenicity of Hepa1-6 cells in the presence of TNF-α. **(A)** TNF-α dose-dependently promoted the proliferation of Hepa1-6 cells; maximum proliferation was observed at 50 ng/ml TNF-α; results are mean ± SD (n = 3 per group); one-way ANOVA with Bonferroni's test; ***, p < 0.001. **(B)** Hepa1-6 cells treated with 50 ng/ml TNF-α expressed higher levels of IL-6 compared to untreated cells; results are mean ± SD (n = 3 per group); Student's t-test; ***, p < 0.001. **(C)** The combination of 50 ng/ml TNF-α and 50 μM TMAO led to the maximal synergistic increase in cell proliferation; results are mean ± SD (n = 3 per group); one-way ANOVA with Bonferroni's test; ***, p < 0.001. **(D,E)** Wound healing assay confirmed that 50 ng/ml TNF-α alone and 50 μM TMAO alone slightly promoted the migration of Hepa1-6 cells compared to untreated control cells, while the combination of 50 ng/ml TNF-α and 50 μM TMAO synergistically and

(*Continued*)

FIGURE 1 | significantly promoted the migration of Hepa1-6 cells. Scale bars = 1,000 μm; data were analyzed by ImageJ; data are mean ± SD (n = 3 per group); one-way ANOVA with Bonferroni's test. **(F, G)** Migration assay confirmed that 50 ng/ml TNF-α alone and 50 μM TMAO alone slightly promoted the migration of Hepa1-6 cells compared to untreated control cells, while the combination of 50 ng/ml TNF-α and 50 μM TMAO synergistically and significantly promoted the migration of Hepa1-6 cells. Scale bars = 400 μm; data were analyzed by ImageJ; data are mean ± SD (n = 3 per group); one-way ANOVA with Bonferroni's test.

Interestingly, we also found that 50 ng/ml TNF-α synergistically promoted the proliferation of Hepa1-6 cells in combination with specific concentrations of trimethylamine N-oxide (TMAO), a co-metabolite of the liver and gut microbiota (**Figure 1C**). After screening the optimal doses of TNF-α and TMAO, four treatment groups were established to further investigate the effects of TNF-α and TMAO on the function of hepatoma Hepa1-6 cells: no treatment control, 50 ng/ml TNF-α alone, 50 μM TMAO alone, and the combination of 50 ng/ml TNF-α and 50 μM TMAO.

Wound healing and migration assays confirmed that either 50 ng/ml TNF-α or 50 μM TMAO alone slightly promoted the migration of Hepa1-6 cells compared with control cells; however, the combination of 50 ng/ml TNF-α and 50 μM TMAO synergistically and significantly promoted the migration of Hepa1-6 cells (**Figures 1D–G**).

Transcriptomic Alterations in TMAO-Induced Primary Liver Cancer Cell Lines

In order to explore whether specific signaling pathways are regulated by TNF-α and TMAO in Hepa1-6 cells, we analyzed the transcriptomic alterations in the four treatment groups of Hepa1-6 cells after treatment for 24 h. The RNA sequencing data revealed 3741, 398, 3018 and 2674 differentially expressed genes (DEGs; FC > 2 and $p < 0.05$) in the untreated group (Group Ctrl), cells treated with 50 μM TMAO alone (Group 50T), cells treated with 50 ng/ml TNF-α alone (Group 50α), and cells treated with 50 ng/ml TNF-α and 50 μM TMAO in combination (Group 50T + 50α), respectively. Clustering identified 35 DEGs that were common to all four groups and the expression of 35 DEGs in four groups are shown in the heat map B (**Figures 2A,B**). In addition, to investigate genes involved in the ability of TMAO to amplify hepatocellular carcinoma cell proliferation and migration in the presence of TNF-α, we analyzed the expression differences in the 35 DEGs between Group 50α and Group 50T + 50α. The genes with the greatest differences were *POSTN*, *NAPB*, *LAYN*, *HTRA3*, *AANAT*, *ZFP850* and *AFM* (**Figure 2C**). Then, a volcano plot was generated to identify the most significantly differential genes in Group 50T + 50α compared with Group 50α (**Figure 2D**), which confirmed the previous heat map analysis.

Associations Between the Identified DEGs and Survival Outcomes in Human Liver Cancer

Next, we investigated whether the expression of the identified DEGs correlates with the outcomes of patients with liver cancer using the Kaplan-Meier plotter database. Kaplan Meier plotter can be used to assess the effects of 54,000 genes (mRNA, miRNA, protein) on survival in several cancers based on microarray data; we focused on the overall survival (OS) and disease-specific survival (DSS) rates for liver cancer in this database. Interestingly, higher expression of *POSTN* (OS HR = 1.63, 95% CI = 1.13–2.33, p = 0.0076; DSS HR = 1.8, 95% CI = 1.11–2.93, p = 0.016; **Figure 3A**), *LAYN* (DSS HR = 1.61, 95% CI = 1.02–2.55, p = 0.039; **Figure 3B**) and *HTRA3* (OS HR = 1.53, 95% CI = 1.08–2.17, p = 0.015) were associated with poorer outcomes in liver cancer (**Figure 3C**). Moreover, lower expression of *AANAT* (OS HR = 0.5, 95% CI = 0.35–0.71, p = 7.6e-05; DSS HR = 0.34, 95% CI = 0.22–0.54, p = 8.2e-07; **Figure 3D**) and *AFM* (OS HR = 0.48, 95% CI = 0.33–0.69, P = 6e-05; DSS HR = 0.36, 95% CI = 0.2–0.64, p = 0.00028; **Figure 3E**) were strongly associated with poorer outcomes in liver cancer. In contrast, upregulation of *NAPB* was not associated with OS (HR = 1.28, 95% CI = 0.9–1.81, p = 0.16) or DSS (HR = 1.39, 95% CI = 0.89–2.18, p = 0.14) in liver cancer (**Figure 3F**). Moreover, upregulation of *LAYN* and upregulation of *HTRA3* were not significantly associated with OS (HR = 1.42, 95% CI = 0.97–2.08, p = 0.066) or DSS (HR = 1.56, 95% CI = 0.96–2.52, p = 0.069), respectively in liver cancer (**Figures 3B,C**). Overall, the survival analysis based on the DEGs suggests that the expression of *POSTN, HTRA3*, *LAYN, AFM* and *AANAT* are associated with the outcomes of patients with liver cancer.

In addition, the RNA sequencing data in the TCGA database were used to further analyze the associations of the identified DEGs with the outcomes of patients with liver cancer via GEPIA. High expression of *POSTN*, *LAYN*, *HTRA3* and *NAPB* were associated with poorer OS in liver hepatocellular carcinoma (LIHC; **Figures 3A–C,F**). Moreover, low *AANAT* expression and low *AFM* expression were associated with poorer OS in LIHC (**Figures 3D,E**). These results confirm that the *POSTN*, *NAPB*, *HTRA3*, *LAYN*, *AFM* and *AANAT* genes may contribute to tumorigenicity and lead to poorer patient outcomes in liver cancer.

Associations Between the Identified DEGs and Neutrophil/Treg Infiltration in HCC

To investigate the relationship between the DEGs identified in this study and the diverse variety of infiltrating immune cells in liver cancer, we explored the associations between the DEGs and immune marker sets for various immune cells, including neutrophils, regulatory T cells (Tregs), NK cells and $CD8^+$ T cells, in LIHC using the TIMER database, using LIHC as the control. The immunosuppressed HCC tumor microenvironment was positively correlated with neutrophil and Treg infiltration and negatively correlated with activated NK cell and $CD8^+$ T cell infiltration. After screening in the database, we found the expression levels of *POSTN*, *HTRA3*

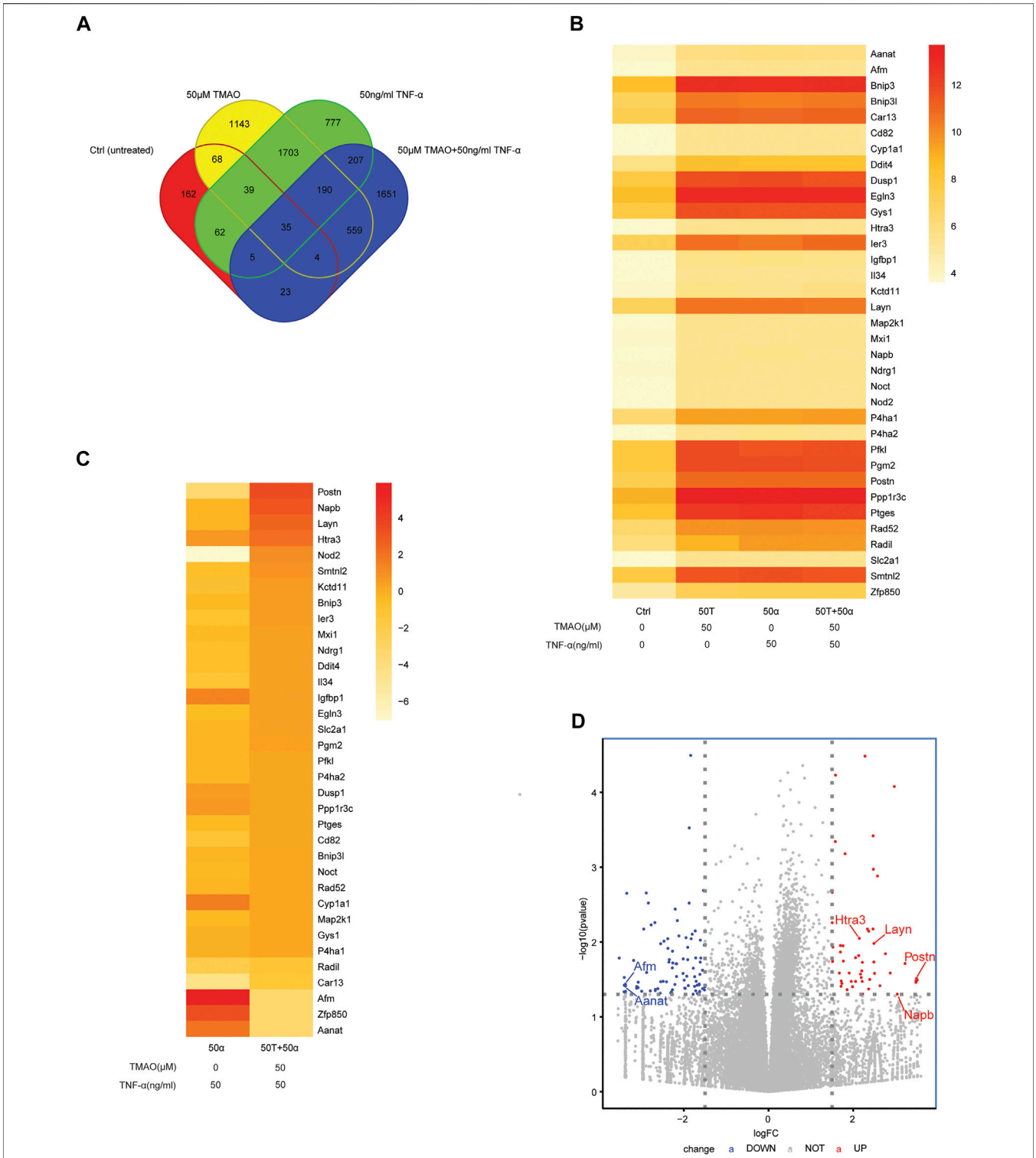


FIGURE 2 | Transcriptomic alterations in TMAO-induced primary liver cancer cells. **(A)** Venn diagrams showing the crossover between the differentially expressed genes (DEGs) in untreated Hepa1-6 cells (Group Ctrl), and Hepa1-6 cells treated with 50 μM TMAO alone (Group 50T), 50 ng/ml TNF-α alone (Group 50α) or 50 ng/ml TNF-α and 50 μM TMAO in combination (Group 50T + 50α). **(B)** Heatmap showing the DEGs common to the four treatment groups; FC > 2 and $p < 0.05$. **(C)** Heatmap showing the DEGs common to Group 50α and Group 50T + 50α; FC > 2 and $p < 0.05$. **(D)** Volcano map showing the significant differentially expressed genes between Group 50α and Group 50T + 50α; FC > 2 and $p < 0.05$; FC: Fold change.

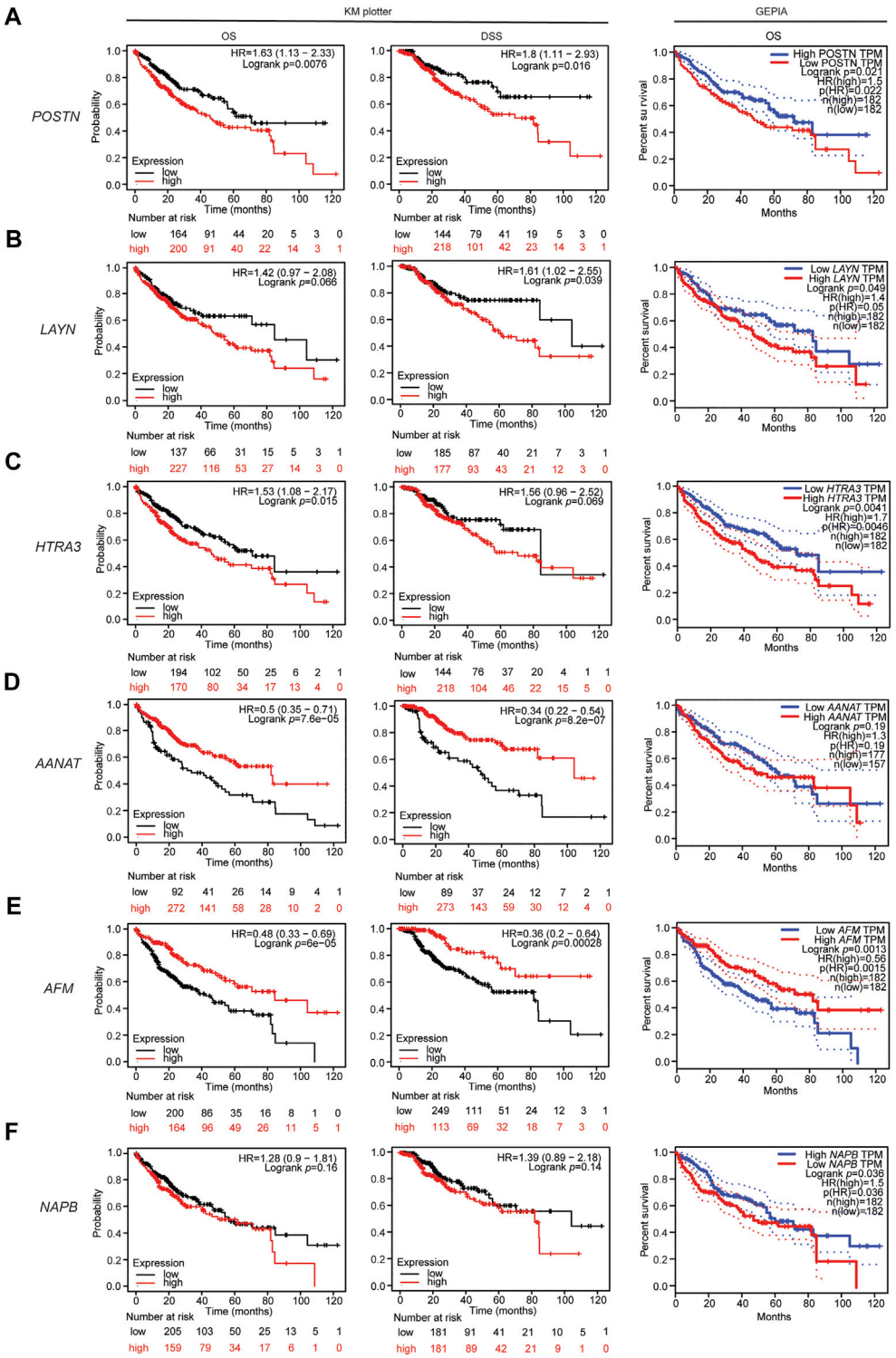


FIGURE 3 | Associations between the DEGs and patient outcomes in liver cancer. Overall survival curve for patients with liver cancer in the GEPIA database (right) and overall survival and disease specific survival curves (left, center) for patients with liver cancer in the Kaplan-Meier plotter databases stratified by expression of **(A)** *POSTN*, **(B)** *LAYN*, **(C)** *HTRA3*, **(D)** *AANAT*, **(E)** *AFM*, and **(F)** *NAPB*.

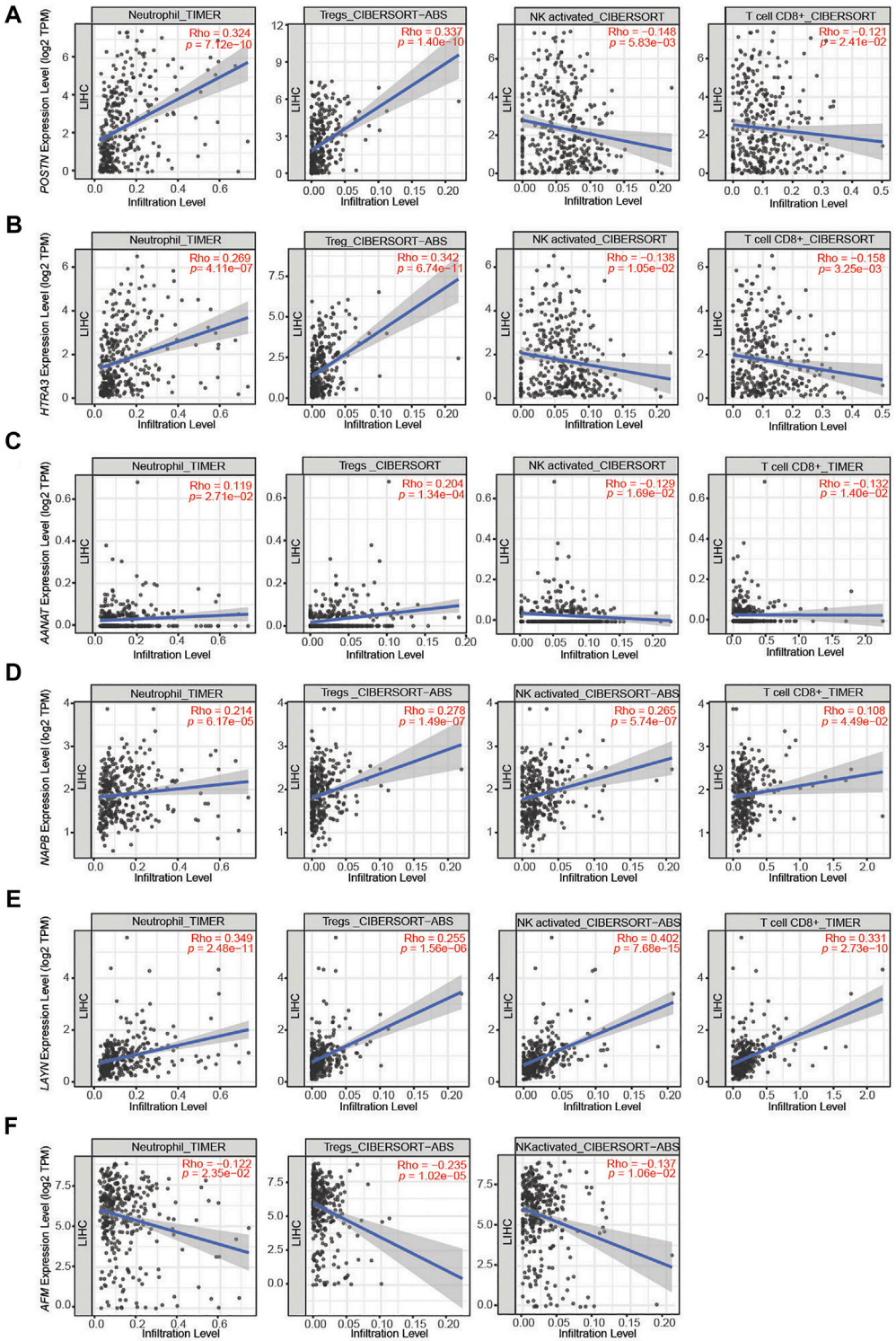


FIGURE 4 | Correlations between the DEGs and immune cell infiltration in liver hepatocellular carcinoma (LIHC). **(A–C)** The expression levels of *POSTN*, *HTRA3* and *AANAT* significantly positively correlated with the levels of neutrophil and Treg infiltration and negatively correlated with the levels of activated NK cell and $CD8^+$ T cell infiltration. **(D,E)** The expression levels of *NAPB* and *LAYN* significantly positively correlated with neutrophil, Treg, activated NK cell and $CD8^+$ T cell infiltration. **(F)** The expression levels of *AFM* negatively correlated with the levels of neutrophil, Treg and activated NK cell infiltration.

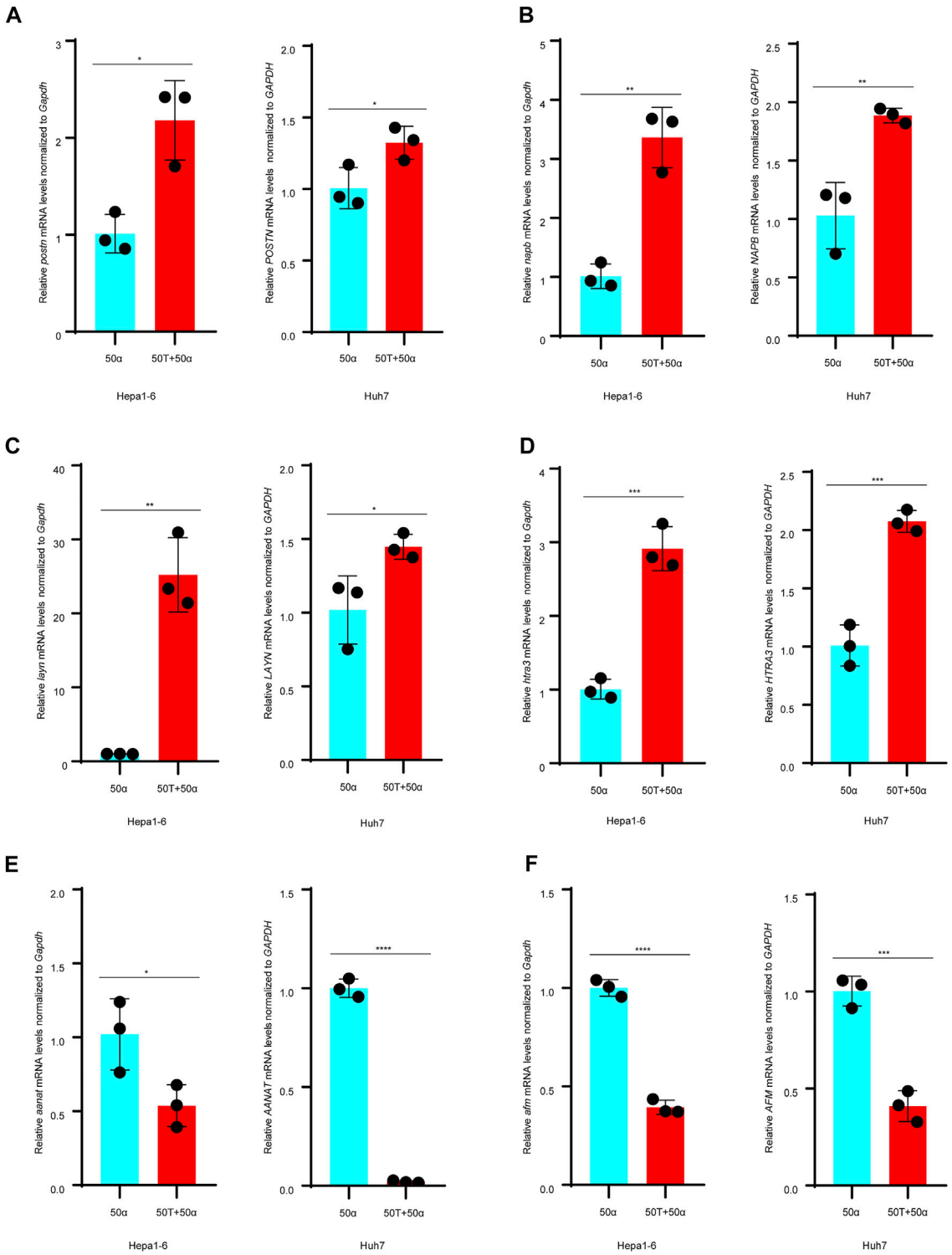


FIGURE 5 | TMAO and TNF-α alter expression of the DEGs in Hepa1-6 cells and Huh7 cells. **(A–D)** Expression of the *POSTN*, *NAPB*, *LAYN* and *HTRA3* mRNAs were significantly upregulated in Hepa1-6 cells and Huh7 cells treated with 50 ng/ml TNF-α and 50 μM TMAO in combination compared to cells treated with 50 ng/ml TNF-α alone; results are mean ± SD (n = 3 per group); Student's t-test. **(E,F)** On the contrary, 50 ng/ml TNF-α and 50 μM TMAO in combination significantly downregulated the expression of *AANAT* and *AFM* compared to cells treated with 50 ng/ml TNF-α alone; results are mean ± SD (n = 3 per group); Student's t-test.

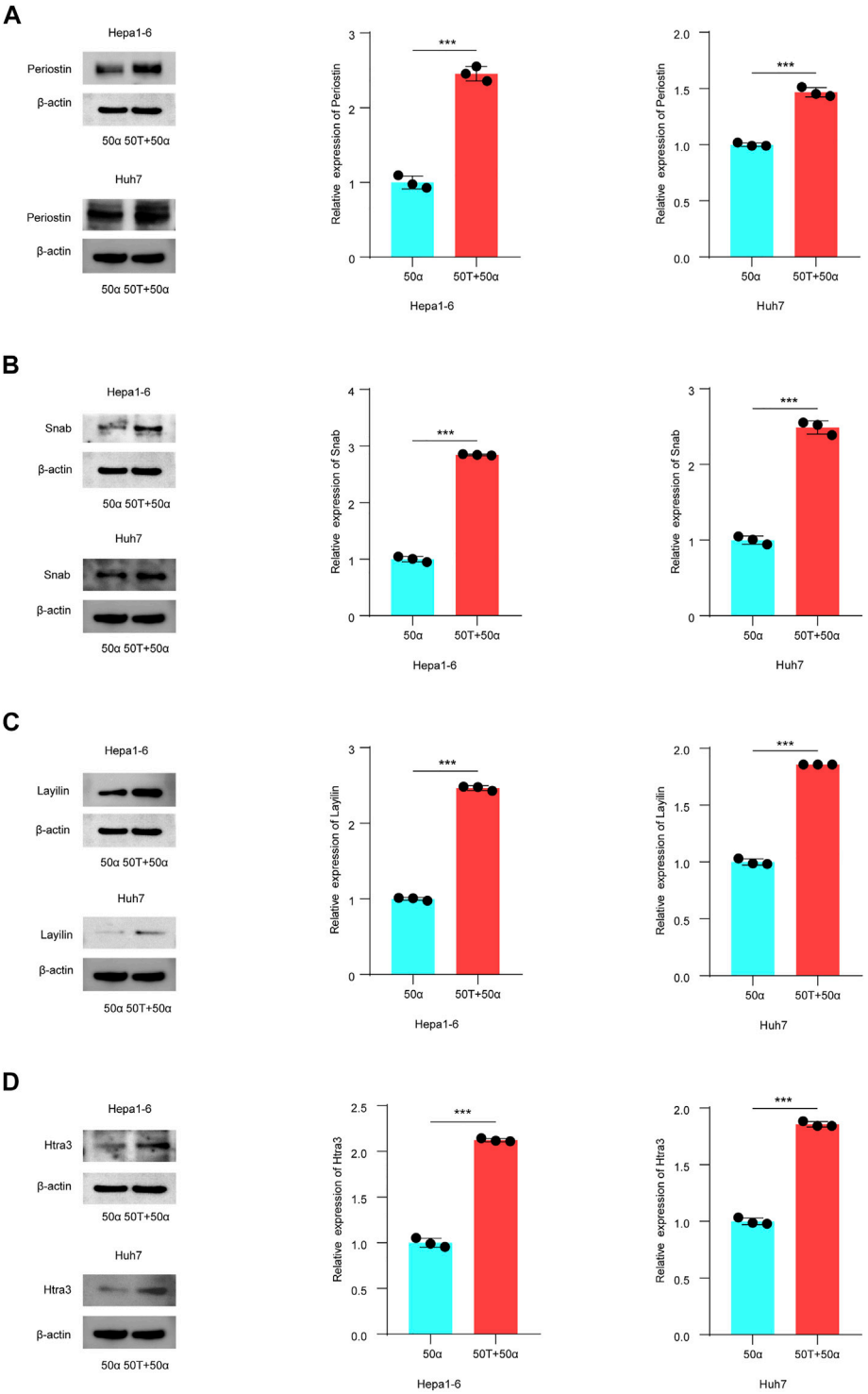


FIGURE 6 | TMAO regulates the expression of DEGs related to patient outcome in TNF-α induced inflammatory HCC. **(A–D)** Western blotting was used to analyze the expression level of Periostin, Snab, Layilin and Htra3 in Hepa1-6 and Huh7 cells treated with 50 ng/ml TNF-α combined with 50 μM TMAO or 50 ng/ml TNF-α alone. Grayscale values were analyzed; data are mean ± SD (n = 3 per group); Student's t-test; ***, $p < 0.001$.

and *AANAT* significantly positively correlated with neutrophil and Treg infiltration and negatively correlated with activated NK cell and $CD8^+$ T cell infiltration in LIHC (**Figures 4A–C**). Moreover, the expression levels of *NAPB* and *LAYN* significantly positively correlated with neutrophil, Treg, activated NK cell and $CD8^+$ T cell infiltration (**Figures 4D,E**),

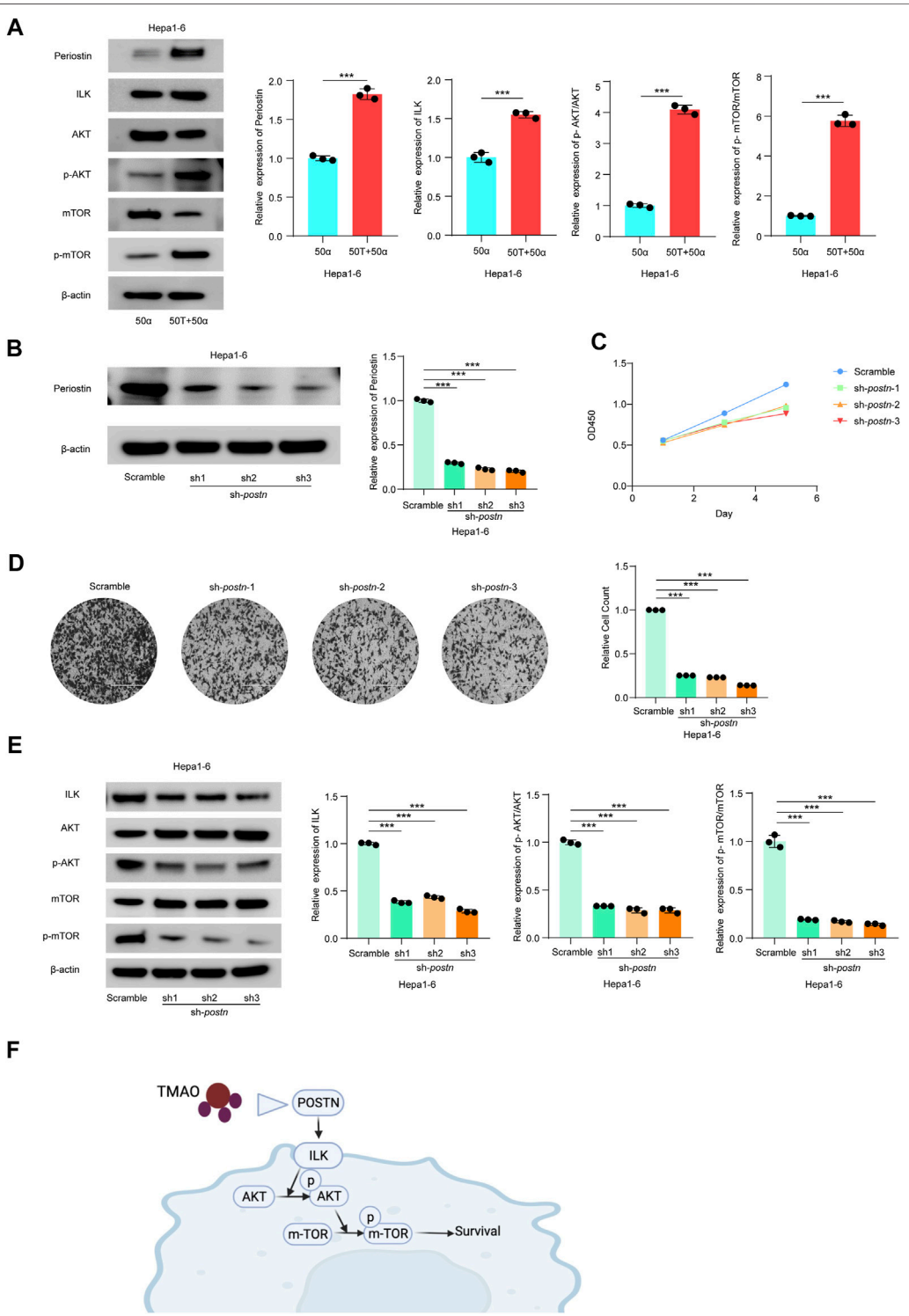


FIGURE 7 | TMAO activates ILK/AKT/mTOR signaling via *Postn* in Hepa1-6 cells. **(A)** The translation levels of POSTN, ILK, AKT, p-AKT, mTOR and p-mTOR measured by western blotting in Hepa1-6 treated with 50 ng/ml TNF-α combined with 50 μM TMAO or 50 ng/ml TNF-α alone. Grayscale values were analyzed; data are mean ± SD ($n = 3$ per group); Student's t-test; ***, $p < 0.001$. **(B)** POSTN translation levels in Hepa1-6 cells infected with pseudovirus expressing sh-*postn* (sh1, sh2, sh3) plasmids were compared with the Hepa1-6 cells infected with pseudovirus expressing scramble plasmids measured by western blotting. Grayscale values were analyzed; data are mean ± SD ($n = 3$ per group); one-way ANOVA with Bonferroni's test; ***, $p < 0.001$. **(C)** CCK-8 assay of the effect of knocking down *Postn* on the proliferation of Hepa1-6 cells; data are mean ± SD ($n = 3$ per group); one-way ANOVA with Bonferroni's test; ***, $p < 0.001$. **(D)** Transwell assays of the effect of knocking

(*Continued*)

FIGURE 7 | down *Postn* on the invasive abilities of Hepa1-6 cells. Scale bars = 100 μm; Data were analyzed by ImageJ; data are mean ± SD (n = 3 per group); one-way ANOVA with Bonferroni's test; ***, $p < 0.001$. **(E)** Western blots of the effects of knocking down *Postn* on the protein levels of Postn, ILK, AKT, p-AKT, mTOR and p-mTOR in Hepa1-6 cells. Grayscale values were analyzed; data are mean ± SD (n = 3 per group); one-way ANOVA with Bonferroni's test; ***, $p < 0.001$. **(F)** Schematic diagram of the role of TMAO in activation of the ILK/AKT/mTOR signaling pathway via POSTN.

whereas the expression of *AFM* negatively correlated with neutrophil, Treg and activated NK cell infiltration (**Figure 4F**). These results indicate that *POSTN* and *HTRA3* are positively associated with an inflammatory tumor microenvironment, which has been proven to promote tumor progression.

TMAO Regulates the Expression of Prognosis-Related Genes in TNF-α Induced Inflammatory HCC

To explore whether TMAO affects the expression of the DEGs identified to affect patient outcomes, Hepa1-6 and Huh7 cell were treated with 50 ng/ml TNF-α either alone or in combination with 50 μM TMAO. The expression of the genes and their products were quantified at the transcriptional and translational levels by RT–qPCR and western blotting. RT–qPCR suggested that TNF-α together with TMAO significantly upregulated the mRNA level of Periostin (*Postn*) compared to cells treated with TNF-α alone in Hepa1-6 cells; these changes were confirmed by western blotting (**Figure 5A**, **6A**). Moreover, the mRNA level of Snab (*Napb*) was significantly increased by the combination of TNF-α and TMAO, with obvious upregulation at the transcriptional level (**Figures 5B**, **6B**). TMAO in combination with TNF-α also significantly upregulated Layilin (*Layn*) and Htra3 (*Htra3*) at both the transcriptional and translational levels (**Figures 5C**,D, **6C**,D). Moreover, afamin (*Afm*) and serotonin N-acetyltransferase (*Aanat*) were downregulated in both cell lines after treatment with TMAO and TNF-α (**Figures 5E,F**). Overall, the changes observed in the mouse cell lines mirrored the trends demonstrated in the transcriptomic analysis of human tumor samples in **Figures 2C,D**, and confirm that TMAO significantly upregulates Periostin, Snab, Layilin and Htra3 and downregulates afamin and serotonin N-acetyltransferase in liver cancer in the presence of TNF-α.

TMAO Activates ILK/AKT/mTOR Signaling *via* Postn

Postn, which encodes Periostin, was the most significantly altered DEG in our transcriptomic profiling and the gene most significantly associated with the immunosuppressed tumor microenvironment in HCC. Previous studies confirmed that *Postn* is closely related to cancer progression and may promote tumor invasion through the ILK (Integrin-linked protein kinase)/RAC-alpha serine (AKT)/Serine/threonine-protein kinase mTOR (mTOR) pathway (Zhou et al., 2015; Chen et al., 2020; Jia et al., 2021). The ILK/AKT and mTOR signaling pathway regulates various aspects of cell development and is also implicated in the progression of human carcinomas (Ma et al., 2011; Jia et al., 2021; Pappa et al., 2021).

Thus, we investigated the effects of TMAO and TNF-α on activation of the ILK/AKT and mTOR signaling pathway *in vitro*. Exposure to 50 ng/ml TNF-α combined with 50 μM TMAO significantly increased the levels of ILK, phosphorylated AKT and mTOR phosphorylation, and also increased expression of Postn compared to cells treated with 50 ng/ml TNF-α alone. Conversely, TMAO and TNF-α reduced the levels of AKT, and significantly decreased mTOR expression (**Figures 7A**, **8A**).

To further study the potential oncogenic activity of Periostin in Hepa1-6 cells, knockdown studies were performed using three short hairpin RNAs (shRNAs) with different sequences targeting *Postn*. Transfection of sh-*postn* (sh1, sh2, sh3) into Hepa1-6 cells and Huh7 cells effectively reduced the expression of *Postn* (**Figures 7B**, **8B**). The CCK-8 cell proliferation assay showed that knockdown of *Postn* significantly reduced the proliferation of Hepa1-6 cells at 72 h compared to cells transfected with the scrambled control shRNA (**Figure 7C**). Moreover, Transwell assays demonstrated that the migratory capacity of Hepa1-6 cells was significantly reduced by knockdown of *Postn* (**Figure 7D**). The same situation also exists in the Huh7 cell line (**Figures 8C,D**). Collectively, these results suggest that *Postn* promotes the proliferation and migration of Hepa1-6 cells and Huh7 cells.

In addition, knockdown of *Postn* downregulated ILK, phosphorylated AKT and mTOR phosphorylation in Hepa1-6 cells and Huh7 cells, and upregulated AKT and mTOR expression (**Figures 7E**, **8E**). Overall, these results indicate that TMAO promotes cell proliferation and migration by activating the ILK/AKT/mTOR pathway via Periostin in Hepa1-6 cells and Huh7 cells (**Figure 7F**).

DISCUSSION

The cometabolite TMAO plays an important role in inflammatory diseases. However, the relationship between TMAO and HCC has received little study. Here, we report that TMAO plays an important role in promoting TNF-α-induced inflammatory liver cancer by enhancing hepatocyte proliferation, migration and invasion. Moreover, TMAO upregulates *POSTN*, *NAPB*, *LAYN*, and *HTRA3* in liver cancer, and high expression levels of these genes are closely associated with suppression of the immune microenvironment and patient survival outcomes in liver cancer. Furthermore, we found that TMAO upregulates *POSTN* and activates the ILK/AKT/mTOR pathway, which induces tumor proliferation and migration. These results provide the first evidence of the mechanism of action of TMAO in TNF-α-induced inflammatory liver cancer.

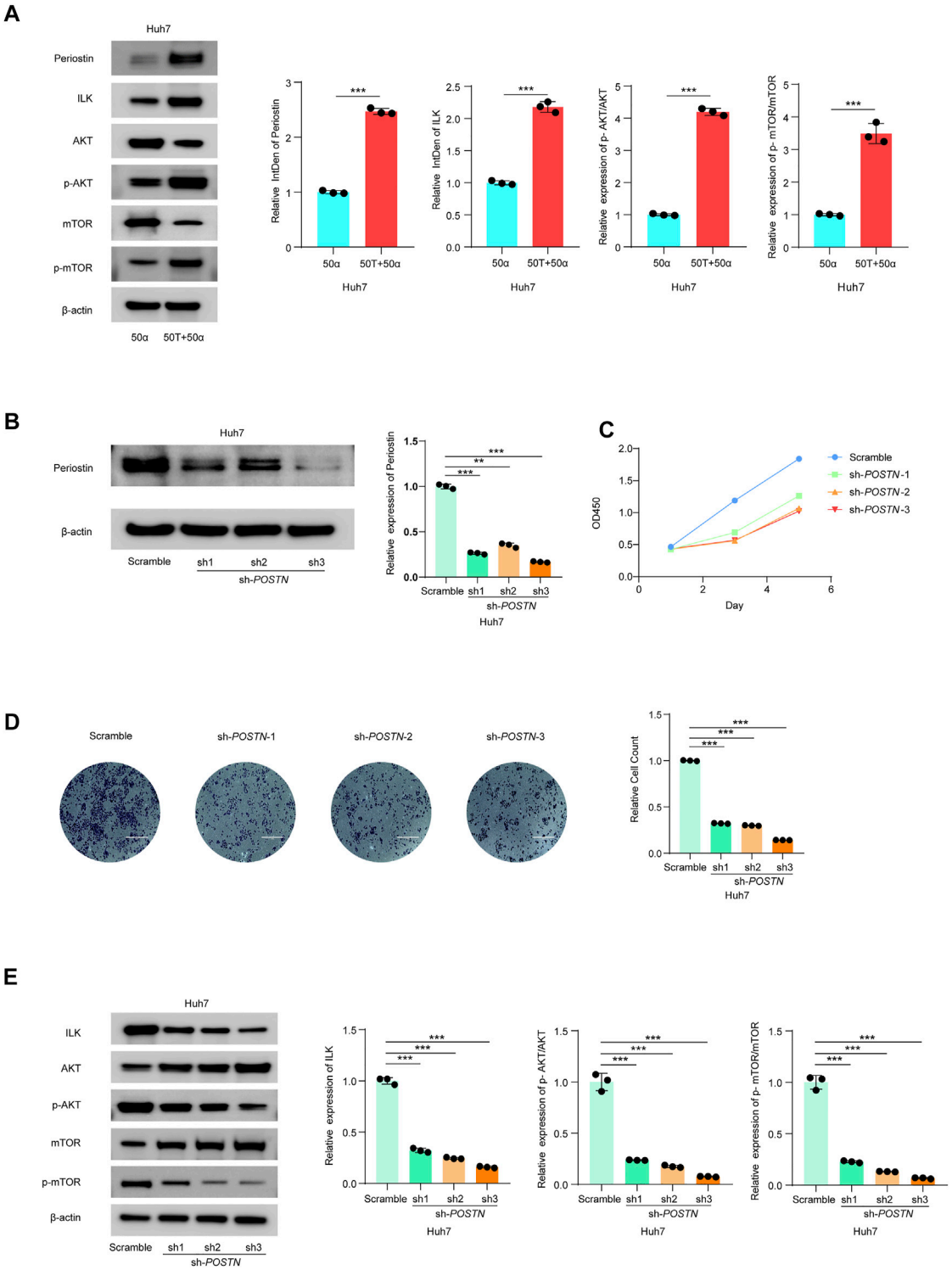


FIGURE 8 | TMAO activates ILK/AKT/mTOR signaling via *Postn* in Huh7 cells. **(A)** The translation levels of POSTN, ILK, AKT, p-AKT, mTOR and p-mTOR measured by western blotting in Huh7 cells treated with 50 ng/ml TNF-α combined with 50 μM TMAO or 50 ng/ml TNF-α alone. Grayscale values were analyzed; data are mean ± SD (n = 3 per group); Student's t-test; ***, $p < 0.001$. **(B)** *POSTN* translation levels in Huh7 cells infected with pseudovirus expressing sh-*POSTN* (sh1, sh2, sh3) plasmids were compared with the Huh7 cells infected with pseudovirus expressing scramble plasmids measured by western blotting. Grayscale values were analyzed; data are mean ± SD (n = 3 per group); one-way ANOVA with Bonferroni's test; ***, $p < 0.001$. **(C)** CCK-8 assay of the effect of knocking down *POSTN* on the proliferation of Huh7 cells; data are mean ± SD (n = 3 per group); one-way ANOVA with Bonferroni's test; ***, $p < 0.001$. **(D)** Transwell assays of the effect of knocking down *POSTN* on the invasive abilities of Huh7 cells. Scale bars = 100 μm; Data were analyzed by ImageJ; data are mean ± SD (n = 3 per group); one-way ANOVA with Bonferroni's test; ***, $p < 0.001$. **(E)** Western blots of the effects of knocking down *POSTN* on the protein levels of Postn, ILK, AKT, p-AKT, mTOR and p-mTOR in Huh7 cells. Grayscale values were analyzed; data are mean ± SD (n = 3 per group); one-way ANOVA with Bonferroni's test; ***, $p < 0.001$.

Periostin (encoded by *POSTN*) is a matricellular protein that functions as critical factor in the development of hepatic inflammation, fibrosis, liver cirrhosis and liver cancer (Kudo, 2011; Huang et al., 2015; Chen et al., 2020). Upregulation of POSTN promotes cell migration and invasion and is also associated with the epithelial-mesenchymal transition (EMT) in clinical HCC tissues (Chen et al., 2019). We found that the proliferation and migratory capacity of Hepa1-6 cells were significantly decreased after knockdown of *Postn*. In addition, downregulation of *POSTN* reduced the tumor-forming ability of HCC cell lines in xenograft mouse models (Chen et al., 2021). Moreover, *POSTN* mRNA expression positively correlated in tumor tissues, but not in non-tumor tissues (Kongkavitoon et al., 2018). We also confirmed that *POSTN* was the most significantly altered differential gene in the inflammatory hepatocellular carcinoma environment (**Figure 2C**). This evidence strongly suggests POSTN is involved in the complex mechanisms that lead to liver cancer.

LAYN is a key gene that regulates T cell function. Overexpression of *LAYN* in human blood $CD8^+$ T cells significantly decreased IFN-γ production, which indicates that *LAYN* represses $CD8^+$ T cell function (Zheng et al., 2017). Additionally, *LAYN* is associated with the levels of immune infiltration ($CD8^+$ T cells, $CD4^+$ T cells, macrophages and neutrophils) as well as patient prognosis in various cancers, especially colon cancer and gastric cancer (Pan et al., 2019). We found a significant positive correlation between neutrophils and Tregs infiltration and *LAYN* expression level. The high expression of *LAYN* was associated with poor prognosis in HCC patients. This suggests that *LAYN* is fully involved in the immune infiltration of liver cancer and plays an important role in the development of tumor. A study of the extracellular matrix in the HCC microenvironment suggested that *HTRA3* and four other genes were significantly associated with immune cell infiltration and could be used as genetic markers for a prognostic score (Liu et al., 2020). Similarly, we demonstrated a positive association between HTRA3 and the inflammatory tumor microenvironment. Taken together, these two genes have great potential as new therapeutic targets for liver cancer.

Traditional surgical treatments or drug therapies for HCC have strict indications and limited therapeutic effects. Therefore, a new therapeutic strategy is needed to improve the prognosis of patients with HCC. With the discovery of immune checkpoint mechanisms, the application of immunotherapy in cancer is gradually being explored, and liver cancer has a higher immunotherapeutic potential as an inflammatory-driven disease (Okusaka and Ikeda, 2018). Tumor cell proliferation, invasion and metastasis are regulated by many factors. The interaction of multiple immune cells in complex pathways leads to the occurrence of tumors. We found that DEGs with significant upregulation of TMAO were positively correlated with upregulation of neutrophils and Tregs in the inflammatory hepatocellular carcinoma environment. Neutrophils represent the majority of inflammatory cells in solid tumors (Shen et al., 2014). Tumor-associated neutrophils participate in the tumor microenvironment (TME) by producing cytokines and chemokines that influence the growth of tumor cells. Autophagy is enhanced by activation of ERK1/2, P38, and NF-κB signaling after neutrophil infiltration. This strongly increases neutrophil survival and exerts tumor-promoting effects in HCC (Li et al., 2015; Masucci et al., 2019). Additionally, neutrophil-conditioned medium was shown to increase the migration of HCC cells or mouse macrophages and Tregs. It showed that neutrophils recruit macrophages and Tregs into HCC to promote tumor growth and progression (Zhou et al., 2016). Tregs are recruited into the TME and enable tumor cells to evade immune surveillance, and excessive Treg activity can lead to cancer by inhibiting the anti-tumor immune response (Lee, 2017; Tanaka and Sakaguchi, 2019). Therefore, neutrophils and Tregs promote tumorigenesis through complex mechanisms. TMAO may promote the development of inflammatory liver cancer through these pathways. Therapeutic regulation of Treg function has become a promising treatment approach (Wang et al., 2013). Immunotherapy targeting the TMAO pathway is an effective clinical treatment for HCC patients without surgical indications.

ILK regulates cell growth, motility and differentiation, and plays biological functions through a variety of signaling pathways (Hannigan et al., 1996; Riaz et al., 2013; Jackson et al., 2015; Shen et al., 2016; Piedra-Quintero et al., 2018; Huang et al., 2020). ILK is widely overexpressed in different cancers and its upregulation is closely related to tumor grade and survival (Zheng et al., 2019). Moreover, the ability of ILK to promote tumor proliferation, invasion and metastasis has been widely studied (Oloumi et al., 2010; Taylor et al., 2011). The present study demonstrated that TMAO significantly increased ILK levels, phosphorylated AKT and phosphorylation of mTOR. *POSTN* expression was positively correlated with ILK pathway activation. These results suggest that ILK and related pathways play an important role in the pathophysiological process of liver cancer. ILK is closely related to tumor grade and survival and its up-regulation is associated with poor prognosis. Therefore, it is an important indicator in cancer diagnosis and prognosis. In conclusion, periostin activates the proliferation and migration of tumor cells by the ILK/AKT/mTOR pathway. TMAO promotes the occurrence of inflammatory hepatocellular carcinoma by upregulating *POSTN*.

Studies have shown a direct mechanistic link between the metabolism of dietary nutrients such as choline and the production of TMAO. The precursors of TMAO include carnitine, lecithin, choline and betaine, which are abundant in red meat, liver, fish, milk, wheat bran and spinach (Hamlin et al., 2013). Dairy intake positively correlates with the plasma TMAO concentration (Rohrmann et al., 2016). The levels of TMAO can be easily and effectively reduced by controlling the types and amounts of food consumed, which may provide an effective, low-cost method among patients with liver cancer. However, microorganisms in the gut must first convert nutrients into TMA before TMAO can be produced (Koeth et al., 2013), thus reducing the TMA-targeting pathway may also lower TMAO levels. One study identified a structural analogue of choline, 3,3-dimethyl-1-butanol (DMB), inhibited formation

of TMA by cultured microorganisms and in various microbial lyases, and reduced the level of TMAO in mice fed a high choline or carnitine diet (Wang et al., 2015; Chen K. et al., 2017). DMB has been widely studied and shown to reverse or prevent various heart and brain disorders and problems in other tissues (Yang W. et al., 2019; Mao et al., 2021). However, the potential of DMB as a treatment for liver cancer needs to be further explored. Furthermore, the important role of *POSTN* in HCC makes it a potential therapeutic target and targeted therapy for *POSTN* will provide new ideas for immunotherapy for HCC.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the 'GEO database' repository, accession number GSE185586.

AUTHOR CONTRIBUTIONS

CZ, YW and XR participated in study conception and design. CZ, YW and MP conducted the study and prepared the manuscript. TW, HY and XC provided support of statistical and bioinformatic analysis. ZX helped prepare the figures and provided informatics support. All authors contributed to the manuscript and agreed to submit manuscript.

FUNDING

This work was supported by the National Key Research and Development Program of China (2018YFC2000500/03) and Shanghai Natural Science Foundation (21ZR1409200). CZ was supported by the original exploration program of Fudan University (2021).

ACKNOWLEDGMENTS

Thanks to everyone who helped this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2022.840171/full#supplementary-material

REFERENCES

Bae, S., Ulrich, C. M., Neuhouser, M. L., Malysheva, O., Bailey, L. B., Xiao, L., et al. (2014). Plasma Choline Metabolites and Colorectal Cancer Risk in the Women's Health Initiative Observational Study. *Cancer Res.* 74 (24), 7442–7452. doi:10.1158/0008-5472.can-14-1835

Baker, R. G., Hayden, M. S., and Ghosh, S. (2011). NF-κB, Inflammation, and Metabolic Disease. *Cel Metab.* 13 (1), 11–22. doi:10.1016/j.cmet.2010.12.008

Boini, K. M., Hussain, T., Li, P.-L., and Koka, S. S. (2017). Trimethylamine-N-Oxide Instigates NLRP3 Inflammasome Activation and Endothelial Dysfunction. *Cell Physiol Biochem* 44 (1), 152–162. doi:10.1159/000484623

Chen, G., Wang, Y., Zhao, X., Xie, X.-z., Zhao, J.-g., Deng, T., et al. (2021). A Positive Feedback Loop between Periostin and TGFβ1 Induces and Maintains the Stemness of Hepatocellular Carcinoma Cells via AP-2α Activation. *J. Exp. Clin. Cancer Res.* 40 (1), 218. doi:10.1186/s13046-021-02011-8

Chen, K., Li, Z., Zhang, M., Wang, B., Peng, T., Shen, Y., et al. (2020). miR-876 Inhibits EMT and Liver Fibrosis via POSTN to Suppress Metastasis in Hepatocellular Carcinoma. *Biomed. Res. Int.* 2020, 1964219. doi:10.1155/2020/1964219

Chen, K., Zheng, X., Feng, M., Li, D., and Zhang, H. (2017a). Gut Microbiota-Dependent Metabolite Trimethylamine N-Oxide Contributes to Cardiac Dysfunction in Western Diet-Induced Obese Mice. *Front. Physiol.* 8, 139. doi:10.3389/fphys.2017.00139

Chen, L., Tian, X., Gong, W., Sun, B., Li, G., Liu, D., et al. (2019). Periostin Mediates Epithelial-Mesenchymal Transition through the MAPK/ERK Pathway in Hepatoblastoma. *Cancer Biol. Med.* 16 (1), 89–100. doi:10.20892/j.issn.2095-3941.2018.0077

Chen, M. L., Zhu, X. H., Ran, L., Lang, H. D., Yi, L., and Mi, M. T. (2017b). Trimethylamine-N-Oxide Induces Vascular Inflammation by Activating the NLRP3 Inflammasome Through the SIRT3-SOD2-mtROS Signaling Pathway. *J. Am. Heart Assoc.* 6 (9). doi:10.1161/JAHA.117.006347

Chen, Y.-m., Liu, Y., Zhou, R.-f., Chen, X.-l., Wang, C., Tan, X.-y., et al. (2016). Associations of gut-flora-dependent Metabolite Trimethylamine-N-Oxide, Betaine and Choline with Non-alcoholic Fatty Liver Disease in Adults. *Sci. Rep.* 6, 19076. doi:10.1038/srep19076

Cheng, X., Qiu, X., Liu, Y., Yuan, C., and Yang, X. (2019). Trimethylamine N-Oxide Promotes Tissue Factor Expression and Activity in Vascular Endothelial Cells: A New Link between Trimethylamine N-Oxide and Atherosclerotic Thrombosis. *Thromb. Res.* 177, 110–116. doi:10.1016/j.thromres.2019.02.028

Cui, X.-X., Yang, X., Wang, H.-J., Rong, X.-Y., Jing, S., Xie, Y.-H., et al. (2017). Luteolin-7-O-Glucoside Present in Lettuce Extracts Inhibits Hepatitis B Surface Antigen Production and Viral Replication by Human Hepatoma Cells *In Vitro*. *Front. Microbiol.* 8, 2425. doi:10.3389/fmicb.2017.02425

EASL (2018). EASL Clinical Practice Guidelines: Management of Hepatocellular Carcinoma. *J. Hepatol.* 69 (1), 182–236. doi:10.1016/j.jhep.2018.03.019

Ghouri, Y. A., Mian, I., and Rowe, J. H. (2017). Review of Hepatocellular Carcinoma: Epidemiology, Etiology, and Carcinogenesis. *J. Carcinog* 16, 1. doi:10.4103/jcar.JCar_9_16

Hamlin, J. C., Pauly, M., Melnyk, S., Pavliv, O., Starrett, W., Crook, T. A., et al. (20132013). Dietary Intake and Plasma Levels of Choline and Betaine in Children with Autism Spectrum Disorders. *Autism Res. Treat.* 2013, 578429. doi:10.1155/2013/578429

Hannigan, G. E., Leung-Hagesteijn, C., Fitz-Gibbon, L., Coppolino, M. G., Radeva, G., Filmus, J., et al. (1996). Regulation of Cell Adhesion and anchorage-dependent Growth by a New β1-integrin-linked Protein Kinase. *Nature* 379 (6560), 91–96. doi:10.1038/379091a0

Huang, Y., Gao, J., Zhou, Y., Wu, S., Shao, Y., Xue, H., et al. (2020). Therapeutic Effect of Integrin-Linked Kinase Gene-Modified Bone Marrow-Derived Mesenchymal Stem Cells for Streptozotocin-Induced Diabetic Cystopathy in a Rat Model. *Stem Cel Res Ther* 11 (1), 278. doi:10.1186/s13287-020-01795-4

Huang, Y., Liu, W., Xiao, H., Maitikabili, A., Lin, Q., Wu, T., et al. (2015). Matricellular Protein Periostin Contributes to Hepatic Inflammation and Fibrosis. *Am. J. Pathol.* 185 (3), 786–797. doi:10.1016/j.ajpath.2014.11.002

Jackson, B. C., Ivanova, I. A., and Dagnino, L. (2015). An ELMO2-RhoG-ILK Network Modulates Microtubule Dynamics. *MBoC* 26 (14), 2712–2725. doi:10.1091/mbc.e14-10-1444

Jia, Y.-y., Yu, Y., and Li, H.-j. (2021). POSTN Promotes Proliferation and Epithelial-Mesenchymal Transition in Renal Cell Carcinoma through ILK/AKT/mTOR Pathway. *J. Cancer* 12 (14), 4183–4195. doi:10.7150/jca.51253

Jing, Y., Sun, K., Liu, W., Sheng, D., Zhao, S., Gao, L., et al. (2018). Tumor Necrosis Factor-α Promotes Hepatocellular Carcinogenesis through the Activation of

Hepatic Progenitor Cells. *Cancer Lett.* 434, 22–32. doi:10.1016/j.canlet.2018.07.001

Koeth, R. A., Wang, Z., Levison, B. S., Buffa, J. A., Org, E., Sheehy, B. T., et al. (2013). Intestinal Microbiota Metabolism of L-Carnitine, a Nutrient in Red Meat, Promotes Atherosclerosis. *Nat. Med.* 19 (5), 576–585. doi:10.1038/nm.3145

Kongkavitoon, P., Butta, P., Sanpavat, A., Bhattarakosol, P., Tangtanatakul, P., Wongprom, B., et al. (2018). Regulation of Periostin Expression by Notch Signaling in Hepatocytes and Liver Cancer Cell Lines. *Biochem. Biophysical Res. Commun.* 506 (3), 739–745. doi:10.1016/j.bbrc.2018.10.144

Kudo, A. (2011). Periostin in Fibrillogenesis for Tissue Regeneration: Periostin Actions inside and outside the Cell. *Cell. Mol. Life Sci.* 68 (19), 3201–3207. doi:10.1007/s00018-011-0784-5

Lee, G. R. (2017). Phenotypic and Functional Properties of Tumor-Infiltrating Regulatory T Cells. *Mediators Inflamm.* 2017, 5458178. doi:10.1155/2017/5458178

Li, X.-F., Chen, D.-P., Ouyang, F.-Z., Chen, M.-M., Wu, Y., Kuang, D.-M., et al. (2015). Increased Autophagy Sustains the Survival and Pro-tumourigenic Effects of Neutrophils in Human Hepatocellular Carcinoma. *J. Hepatol.* 62 (1), 131–139. doi:10.1016/j.jhep.2014.08.023

Liu, X., Niu, X., and Qiu, Z. (2020). A Five-Gene Signature Based on Stromal/Immune Scores in the Tumor Microenvironment and its Clinical Implications for Liver Cancer. *DNA Cel Biol.* 39 (9), 1621–1638. doi:10.1089/dna.2020.5512

Liu, Z.-Y., Tan, X.-Y., Li, Q.-J., Liao, G.-C., Fang, A.-P., Zhang, D.-M., et al. (2018). Trimethylamine N-Oxide, a Gut Microbiota-dependent Metabolite of Choline, Is Positively Associated with the Risk of Primary Liver Cancer: a Case-Control Study. *Nutr. Metab. (Lond)* 15, 81. doi:10.1186/s12986-018-0319-2

Ma, B., Sen, T., Asnaghi, L., Valapala, M., Yang, F., Hose, S., et al. (2011). βA3/A1-Crystallin Controls Anoikis-Mediated Cell Death in Astrocytes by Modulating PI3K/AKT/mTOR and ERK Survival Pathways through the PKD/Bit1-signaling axis. *Cell Death Dis* 2 (10), e217. doi:10.1038/cddis.2011.100

Ma, G., Pan, B., Chen, Y., Guo, C., Zhao, M., Zheng, L., et al. (2017). Trimethylamine N-Oxide in Atherogenesis: Impairing Endothelial Self-Repair Capacity and Enhancing Monocyte Adhesion. *Biosci. Rep.* 37 (2). doi:10.1042/BSR20160244

Mao, J., Zhao, P., Wang, Q., Chen, A., Li, X., Li, X., et al. (2021). Repeated 3,3-Dimethyl-1-Butanol Exposure Alters Social Dominance in Adult Mice. *Neurosci. Lett.* 758, 136006. doi:10.1016/j.neulet.2021.136006

Masucci, M. T., Minopoli, M., and Carriero, M. V. (2019). Tumor Associated Neutrophils. Their Role in Tumorigenesis, Metastasis, Prognosis and Therapy. *Front. Oncol.* 9, 1146. doi:10.3389/fonc.2019.01146

Meng, Y., Zhao, Q., An, L., Jiao, S., Li, R., Sang, Y., et al. (2021). A TNFR2-hnRNPK Axis Promotes Primary Liver Cancer Development via Activation of YAP Signaling in Hepatic Progenitor Cells. *Cancer Res.* 81 (11), 3036–3050. doi:10.1158/0008-5472.can-20-3175

Okusaka, T., and Ikeda, M. (2018). Immunotherapy for Hepatocellular Carcinoma: Current Status and Future Perspectives. *ESMO Open* 3 (Suppl. 1), e000455. doi:10.1136/esmoopen-2018-000455

Oloumi, A., Maidan, M., Lock, F. E., Tearle, H., Mckinney, S., Muller, W. J., et al. (2010). Cooperative Signaling between Wnt1 and Integrin-Linked Kinase Induces Accelerated Breast Tumor Development. *Breast Cancer Res.* 12 (3), R38. doi:10.1186/bcr2592

Pan, J.-h., Zhou, H., Cooper, L., Huang, J.-l., Zhu, S.-b., Zhao, X.-x., et al. (2019). LAYN Is a Prognostic Biomarker and Correlated With Immune Infiltrates in Gastric and Colon Cancers. *Front. Immunol.* 10, 6. doi:10.3389/fimmu.2019.00006

Pappa, T., Ahmadi, S., Marqusee, E., Johnson, H. L., Nehs, M. A., Cho, N. L., et al. (2021). Oncogenic Mutations in PI3K/AKT/mTOR Pathway Effectors Associate with Worse Prognosis in BRAFV600E-Driven Papillary Thyroid Cancer Patients. *Clin. Cancer Res.* 27 (15), 4256–4264. doi:10.1158/1078-0432.ccr-21-0874

Piedra-Quintero, Z. L., Serrano, C., Villegas-Sepúlveda, N., Maravillas-Montero, J. L., Romero-Ramírez, S., Shibayama, M., et al. (2018). Myosin 1F Regulates M1-Polarization by Stimulating Intercellular Adhesion in Macrophages. *Front Immunol.* 9, 3118. doi:10.3389/fimmu.2018.03118

Riaz, A., Ilan, N., Vlodavsky, I., Li, J.-P., and Johansson, S. (2013). Characterization of Heparanase-Induced Phosphatidylinositol 3-Kinase-AKT Activation and its Integrin Dependence. *J. Biol. Chem.* 288 (17), 12366–12375. doi:10.1074/jbc.m112.435172

Rohrmann, S., Linseisen, J., Allenspach, M., Von Eckardstein, A., and Müller, D. (2016). Plasma Concentrations of Trimethylamine-N-Oxide Are Directly Associated with Dairy Food Consumption and Low-Grade Inflammation in a German Adult Population. *J. Nutr.* 146 (2), 283–289. doi:10.3945/jn.115.220103

Rong, X., Wang, H., Ma, J., Pan, S., Wang, H., Jing, S., et al. (2019). Chronic Hepatitis B Virus Infection Is Associated with a Poorer Prognosis in Diffuse Large B-Cell Lymphoma: a Meta-Analysis and Systemic Review. *J. Cancer* 10 (15), 3450–3458. doi:10.7150/jca.31033

Shan, Z., Sun, T., Huang, H., Chen, S., Chen, L., Luo, C., et al. (2017). Association between Microbiota-dependent Metabolite Trimethylamine-N-Oxide and Type 2 Diabetes. *Am. J. Clin. Nutr.* 106 (3), 888–894. doi:10.3945/ajcn.117.157107

Shen, H., Ma, J.-L., Zhang, Y., Deng, G.-L., Qu, Y.-L., Wu, X.-L., et al. (2016). Integrin-linked Kinase Overexpression Promotes Epithelial-Mesenchymal Transitionvianuclear Factor-Kb Signaling in Colorectal Cancer Cells. *Wjg* 22 (15), 3969–3977. doi:10.3748/wjg.v22.i15.3969

Shen, M., Hu, P., Donskov, F., Wang, G., Liu, Q., and Du, J. (2014). Tumor-associated Neutrophils as a New Prognostic Factor in Cancer: a Systematic Review and Meta-Analysis. *PLoS One* 9 (6), e98259. doi:10.1371/journal.pone.0098259

Siegel, R. L., Miller, K. D., and Jemal, A. (2018). Cancer Statistics, 2018. *CA: A Cancer J. Clinicians* 68 (1), 7–30. doi:10.3322/caac.21442

Sun, X., Jiao, X., Ma, Y., Liu, Y., Zhang, L., He, Y., et al. (2016). Trimethylamine N-Oxide Induces Inflammation and Endothelial Dysfunction in Human Umbilical Vein Endothelial Cells via Activating ROS-TXNIP-NLRP3 Inflammasome. *Biochem. Biophys. Res. Commun.* 481 (1-2), 63–70. doi:10.1016/j.bbrc.2016.11.017

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

Tanaka, A., and Sakaguchi, S. (2019). Targeting Treg Cells in Cancer Immunotherapy. *Eur. J. Immunol.* 49 (8), 1140–1146. doi:10.1002/eji.201847659

Taylor, C. J., Qiao, J., Colon, N. C., Schlegel, C., Josifi, E., and Chung, D. H. (2011). Integrin-linked Kinase Regulates Phosphatase and Tensin Homologue Activity to Promote Tumorigenesis in Neuroblastoma Cells. *Surgery* 150 (2), 162–168. doi:10.1016/j.surg.2011.05.007

Wang, Q., Yu, T., Yuan, Y., Zhuang, H., Wang, Z., Liu, X., et al. (2013). Sorafenib Reduces Hepatic Infiltrated Regulatory T Cells in Hepatocellular Carcinoma Patients by Suppressing TGF-Beta Signal. *J. Surg. Oncol.* 107 (4), 422–427. doi:10.1002/jso.23227

Wang, Z., Klipfell, E., Bennett, B. J., Koeth, R., Levison, B. S., Dugar, B., et al. (2011). Gut flora Metabolism of Phosphatidylcholine Promotes Cardiovascular Disease. *Nature* 472 (7341), 57–63. doi:10.1038/nature09922

Wang, Z., Roberts, A. B., Buffa, J. A., Levison, B. S., Zhu, W., Org, E., et al. (2015). Non-lethal Inhibition of Gut Microbial Trimethylamine Production for the Treatment of Atherosclerosis. *Cell* 163 (7), 1585–1595. doi:10.1016/j.cell.2015.11.055

Yang, W., Zhang, S., Zhu, J., Jiang, H., Jia, D., Ou, T., et al. (2019a). Gut Microbe-Derived Metabolite Trimethylamine N-Oxide Accelerates Fibroblast-Myofibroblast Differentiation and Induces Cardiac Fibrosis. *J. Mol. Cell Cardiol.* 134, 119–130. doi:10.1016/j.yjmcc.2019.07.004

Yang, Y. M., Kim, S. Y., and Seki, E. (2019b). Inflammation and Liver Cancer: Molecular Mechanisms and Therapeutic Targets. *Semin. Liver Dis.* 39 (1), 26–42. doi:10.1055/s-0038-1676806

Zheng, C. C., Hu, H. F., Hong, P., Zhang, Q. H., Xu, W. W., He, Q. Y., et al. (2019). Significance of Integrin-Linked Kinase (ILK) in Tumorigenesis and its Potential Implication as a Biomarker and Therapeutic Target for Human Cancer. *Am. J. Cancer Res.* 9 (1), 186–197.

Zheng, C., Zheng, L., Yoo, J.-K., Guo, H., Zhang, Y., Guo, X., et al. (2017). Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. *Cell* 169 (7), 1342–1356. e1316. doi:10.1016/j.cell.2017.05.035

Zhou, S.-L., Zhou, Z.-J., Hu, Z.-Q., Huang, X.-W., Wang, Z., Chen, E.-B., et al. (2016). Tumor-Associated Neutrophils Recruit Macrophages and T-Regulatory Cells to Promote Progression of Hepatocellular Carcinoma and Resistance to Sorafenib. *Gastroenterology* 150 (7), 1646–1658. e1617. doi:10.1053/j.gastro.2016.02.040

Zhou, W., Ke, S. Q., Huang, Z., Flavahan, W., Fang, X., Paul, J., et al. (2015). Periostin Secreted by Glioblastoma Stem Cells Recruits M2 Tumour-Associated Macrophages and Promotes Malignant Growth. *Nat. Cel Biol* 17 (2), 170–182. doi:10.1038/ncb3090

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.2022.899353

Evaluation of the EdgeSeq Precision Immuno-Oncology Panel for Gene Expression Profiling From Clinical Formalin-Fixed Paraffin-Embedded Tumor Specimens

Yang Shi, Xiaopeng Ma, Wei Shen, Tengfei Liu, Liang Liang, Silu Liu, Zhirong Shen, Yun Zhang and Pei Zhang**

BeiGene (Beijing) Co., Ltd., Beijing, China

OPEN ACCESS

Edited by:
Fangqing Zhao,
Beijing Institutes of Life Science (CAS), China

Reviewed by:
Bo Wang,
Sun Yat-sen Memorial Hospital, China
Alexander Deutsch,
Medical University of Graz, Austria

***Correspondence:**
Yun Zhang
yun.zhang@beigene.com
Pei Zhang
pei.zhang@beigene.com

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

Received: *18 March 2022*
Accepted: *28 April 2022*
Published: *27 May 2022*

Citation:
Shi Y, Ma X, Shen W, Liu T, Liang L, Liu S, Shen Z, Zhang Y and Zhang P (2022) Evaluation of the EdgeSeq Precision Immuno-Oncology Panel for Gene Expression Profiling From Clinical Formalin-Fixed Paraffin-Embedded Tumor Specimens. Front. Cell Dev. Biol. 10:899353. doi: 10.3389/fcell.2022.899353

Characterizing the tumor microenvironment (TME) of archived clinical tissues requires reliable gene expression profiling (GEP) of formalin-fixed paraffin-embedded (FFPE) samples. The EdgeSeq Precision Immuno-oncology Panel (PIP) is a targeted GEP assay designed for TME characterization but lacks widespread technical validation on a large cohort of clinical samples. Here, we evaluated its performance by exploring its concordance with multiple orthogonal platforms using 1,220 FFPE samples across various cancer types. Quantitative comparisons with RNA-seq and NanoString showed strong correlations at the sample level (median $\rho = 0.73$ and 0.81) and moderate correlations at the single-gene level (median $\rho = 0.49$ and 0.57). Gene signature analysis revealed high concordance with RNA-seq on widely used signatures for TME characterization and immune checkpoint inhibitor (ICI) efficacy prediction, though some genes in these signatures are not targeted by EdgeSeq PIP. From a histopathological viewpoint, the tumor/immune abundances derived from hematoxylin and eosin (H & E) staining were well recapitulated by the transcriptomic profiles assessed by EdgeSeq PIP. Furthermore, the mRNA level of PD-L1 assessed by EdgeSeq PIP was moderately correlated with the PD-L1 score ($\rho = 0.65$) estimated by immunohistochemistry (IHC); the mRNA level of CD8A aligned well ($\rho = 0.55$) with the IHC-derived abundance of $CD8^+$ T cells. Overall, our results showed that EdgeSeq PIP generated well-correlated data with independent approaches at mRNA, protein, and histological levels, thus providing strong technical support for further using EdgeSeq PIP in biomarker studies and companion diagnostic (CDx) development.

Keywords: gene expression profile, tumor microenvironment, immunotherapy, platform evaluation, edgeseq PIP

INTRODUCTION

Over the past years, immune checkpoint inhibitor (ICI)-based therapies have transformed the treatment landscape of cancer (Sanmamed and Chen, 2018). ICIs such as humanized monoclonal antibodies against cytotoxic T lymphocyte antigen 4 (CTLA4), programmed cell death protein 1 (PD-1) and programmed death ligand 1 (PD-L1) have demonstrated impressive efficacy and have

been approved as first-line or second-line therapies for an ever-growing list of malignancies (Vaddepally et al., 2020). However, only a small fraction of patients benefits from ICI-based therapy, and there are urgent needs to identify the mechanisms driving response or resistance and develop new biomarkers to guide personalized therapy (Havel et al., 2019). Unraveling the landscape of immune cell subpopulations in the tumor microenvironment (TME) and investigating their interactions with tumor and stromal cells is a critical step in this process (Chen and Mellman, 2017).

TME characterization heavily relies on precise and comprehensive gene expression profiling (GEP). Among multiple GEP methods, RNA-seq has been established as the gold standard and is widely used in cancer research (Wang et al., 2009), such as The Cancer Genome Atlas (TCGA). However, in clinical practice, RNA-seq (and derived methods, such as RNA Exome) has several limitations. On the one hand, it requires large amounts of high-quality samples, which is usually not feasible in clinical trials in which most samples are formalin-fixed and paraffin-embedded (FFPE) and exhibit moderate-to-severe degradation (Evers et al., 2011). This requirement also contradicts the scarcity of tumor tissues in most clinical trials, especially when other typical biomarker assays, such as tumor mutation burden (TMB) and immunohistochemistry (IHC) assays are in competition (Aisner et al., 2016). On the other hand, RNA-seq is untargeted and thus not cost-effective for the development of companion diagnosis assay, as only a small proportion of genes are of strong interest (Thorsson et al., 2018).

Given these limitations of RNA-seq and the increasing demands for profiling immuno-oncology-related genes in clinical trials, targeted assays such as EdgeSeq PIP (HTG Molecular, Tucson, AZ), PanCancer IO 360 (NanoString, Seattle, WA) and Oncomine IRRA (Thermo Fisher Scientific, Waltham, MA), have emerged in recent years. These assays significantly reduced requirements for sample quality and quantity, thus mitigating the challenges of sample acquisition and making them clinically feasible. EdgeSeq PIP, which focuses on ~1,300 key genes related to immuno-oncology, has the highest coverage among all these assays. By utilizing quantitative nuclease protection chemistry, EdgeSeq PIP quantifies RNAs *via* an extraction-free approach, which eliminates the risk of extraction bias induced by the removal of short or fragmented RNAs (Martel et al., 2002; Qi et al., 2016). The data generated from FFPE samples has good concordance with that from fresh-frozen samples. Moreover, the extraction-free approach circumvents the loss of RNA from sample during extraction; thus, less tissue input is required to generate an equivalent amount of RNA. Last, EdgeSeq PIP has the unique advantage of utilizing samples previously subjected to hematoxylin and eosin (H & E) or IHC staining, which further expands its clinical utility when samples are extremely limited (Qi et al., 2019). Although successfully used in several clinical trials (Wang et al., 2018; Desai et al., 2020; Garg et al., 2020; Martin-Broto et al., 2020; Song et al., 2021), the reliability of EdgeSeq PIP in real clinical settings and its concordance with other platforms have not been well studied. Previous studies either compared only EdgeSeq PIP with RNA-seq at the single-gene level or were limited by a small number of samples to draw solid conclusions (Anguiano et al., 2020; Ran et al., 2020). In this study, using 1,220 FFPE samples across several cancer types from clinical trials, we performed comprehensive comparisons of EdgeSeq PIP and multiple platforms and confirmed its fidelity at the RNA, protein, and histological levels.

MATERIALS AND METHODS

Patient Cohort

Baseline FFPE samples were collected from the following seven clinical studies on tislelizumab monotherapy or tislelizumab combined with chemotherapy or anti-PD-L1 therapy: A317-001 (NCT02407990), A317-102 (NCT04068519), 900-101 (NCT03379259), A317-204 (NCT04004221), A317-205 (NCT03469557), A317-206 (NCT03432598), RATIONALE 307 (NCT03594747), and RATIONALE 309 (NCT03924986). The major cancer types in each study were summarized in **Supplementary Table S1**.

Sample Preparation, Library Construction and Sequencing

All archived FFPE blocks were prepared as previously described. After confirming the presence of malignant cells by histological H & E staining, the samples were processed *via* standardized procedures for biomarker investigation. In general, a tumor content of at least 20% and sufficient tumor area (>20 mm^2 for RNA-seq and >2.5 mm^2 for EdgeSeq PIP) were required for sample inclusion.

For EdgeSeq PIP, sample processing, library construction and sequencing were performed in accordance with OP-00034, OP-00035, OP-00079 (HTG EdgeSeq instrument method). Briefly, tissues were scraped and lysed using lysis buffer from HTG Molecular Dianostics. Next, Proteinase K was added to digest proteins and remove potential contaminations. Gene-specific nuclease protection probes were then added to the lysed samples to form the probe-target RNA heteroduplexes, after which S1 nuclease was added to degrade non-hybridized molecules. Then samples were individually barcoded using a 16-cycle PCR, purified using Agencourt AMPure XP beads and loaded into Illumina MiSeq (Illumina, San Diego, CA) for 50 bp single-end sequencing.

For RNA-seq, RNA was extracted using the AllPrep DNA/RNA FFPE Kit (QIAGEN, Hilden, Germany). The amount of RNA was quantitated by the fluorescence method using Qubit RNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA), and the quality was assessed by Agilent 2,100 Bioanalyzer System (Agilent Technologies, Santa Clara, CA). Only samples with RNA >40 ng and DV200 >20% were included for downstream steps. rRNA depletion, cDNA synthesis and NGS library preparation were performed using the TruSeq RNA Exome (Illumina, San Diego, CA). The libraries were then loaded into HiSeq X Ten instrument (Illumina, San Diego, CA) for 150 bp paired-end sequencing.

For NanoString, all processing steps were performed according to the manufacturer's instructions. In brief, RNA was extracted using RNeasy Mini Kit (QIAGEN, Hilden, Germany). The amount of RNA was quantitated by the fluorescence method using Qubit RNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA), and the quality was assessed by Agilent 2,100 Bioanalyzer System (Agilent Technologies, Santa Clara, CA). After hybridization with the PanCancer IO 360 Panel, sample analysis was performed on a nCounter Digital Analyzer (NanoString Technologies, Seattle, WA).

EdgeSeq Precision Immuno-Oncology Panel Data Processing

Demultiplexed FASTQ files from the Illumina MiSeq were parsed by the EdgeSeq parser (HTG Molecular Diagnostics, Inc.). Three post-sequencing quality control metrics were derived and used for filtering samples: 1) >15% percentage of reads allocated to the positive process control probe; 2) >1.5 million reads; and 3) a relative standard deviation (RSD) of reads allocated to each probe >0.10. After removing samples not meeting these QC requirements, the raw read counts for each sample were transformed to the log2 counts per million (CPM) scale.

RNA-Seq Data Processing

Adapters in raw reads were detected and trimmed by Trimmomatic (0.36) (Bolger et al., 2014). After trimming, reads shorter than 50 bp were removed. These reads were then mapped to the human genome (GRCh38) using STAR (2.7.10a) (Dobin et al., 2013). Gene expression was quantified using the RSEM workflow (1.3.3) with default parameters (Li and Dewey, 2011). Only samples with >1.5 million reads and >80% reads confidently mapped to the transcriptome were retained. Then, transcripts per million (TPM) values were log2-transformed for downstream analysis.

Differential Expression and Gene Set Enrichment Analysis

Differentially expressed genes were identified following the limma-voom workflow (3.50) (Law et al., 2014; Ritchie et al., 2015). Gene set variation analysis (GSVA), single-sample gene set enrichment analysis (ssGSEA) and Z scores for each signature were calculated using the R package GSVA (1.42) by switching the "method" parameter (Hanzelmann et al., 2013). In addition to the signatures collected from specific studies, gene sets from the MSigDB-C2-Canonical Pathway (KEGG, BioCarta, PID, Reactome, WikiPathways) were also included in the signature analysis (Liberzon et al., 2011).

Programmed Death Ligand 1 Immunohistochemistry

PD-L1 expression was assessed by the VENTANA PD-L1 (SP263) IHC assay (Ventana Medical Systems, Oro Valley, AZ, United States). The level of PD-L1 was then scored by the percentage of PD-L1 membranous staining on tumor cells (TC).

CD8 Immunofluorescence

CD8 immunofluorescence analysis of FFPE samples was performed with CD8A antibody (SP57, Ventana 790-4460) in a College of American Pathologist (CAP)-controlled area within the Oncology and Immunology Unit of WuXi AppTec using the IF 6-colorWJJ-CD30 protocol on a Leica BOND Rx platform. Whole-slide images were acquired by Leica Aperio VERSA 8. Z1. Image analysis was performed using the HALO software package (Indica Labs, United States).

Statistical Analysis

Correlation between two continuous variables was assessed by Spearman correlation. The interpretation of correlation was defined as: negligible (0–0.09), weak (0.10–0.39), moderate (0.4–0.69), strong (0.70–0.89) and very strong (0.90–1.00) (Schober et al., 2018). Differences in medians or continuous variables between two groups were assessed by non-parametric Wilcoxon rank-sum tests. The alpha level for all comparisons was 0.05 unless indicated otherwise. All statistical analyses and visualizations were performed with R (v.4.0.2).

RESULTS

High Concordance Between EdgeSeq Precision Immuno-Oncology Panel and Other GEP Platforms at Gene Level

We first calculated Spearman correlation coefficients for all common samples between EdgeSeq PIP and RNA-seq (**Supplementary Table S2**). The strength of correlation was defined according to a widely used guiding rule for correlation interpretation in medical research (Materials and Methods) (Schober et al., 2018). Of the 395 samples, most had strong correlation (median $\rho = 0.73$) (**Figure 1A**). When assessing the concordance of the two platforms using a gene-by-gene approach, we found a moderate correlation (median $\rho = 0.49$) (**Figure 1B** and **Supplementary Table S3**), which was comparable with the result from a previous study comparing RNA-seq with NanoString (Kwong et al., 2018). We presumed that the weak correlation of some genes was partially attributed to their low expression levels. By stratifying genes according to their relative expression in RNA-seq, we found that the correlation coefficients significantly decreased as the expression level decreased (**Figure 1C**). The genes with expression levels in the lowest 25% only had a median correlation of 0.26 (**Figure 1C**). In addition, the dynamic range of gene expression measured by the median absolute deviation (MAD) also associated with the cross-platform correlation (**Supplementary Figure S1A**; **Supplementary Table S3**). This might indicate that correlation analysis was not suitable for numbers with low variances. After excluding genes with low expression (<1 TPM) and dynamic range (MAD <0.98, 25% quartile), the median correlation improved from 0.49 to 0.60 (**Supplementary Figure S1B**). Furthermore, considering the intended use of EdgeSeq PIP for profiling immune statuses in tumors, we evaluated the expression of several key immune markers and found all genes had moderate

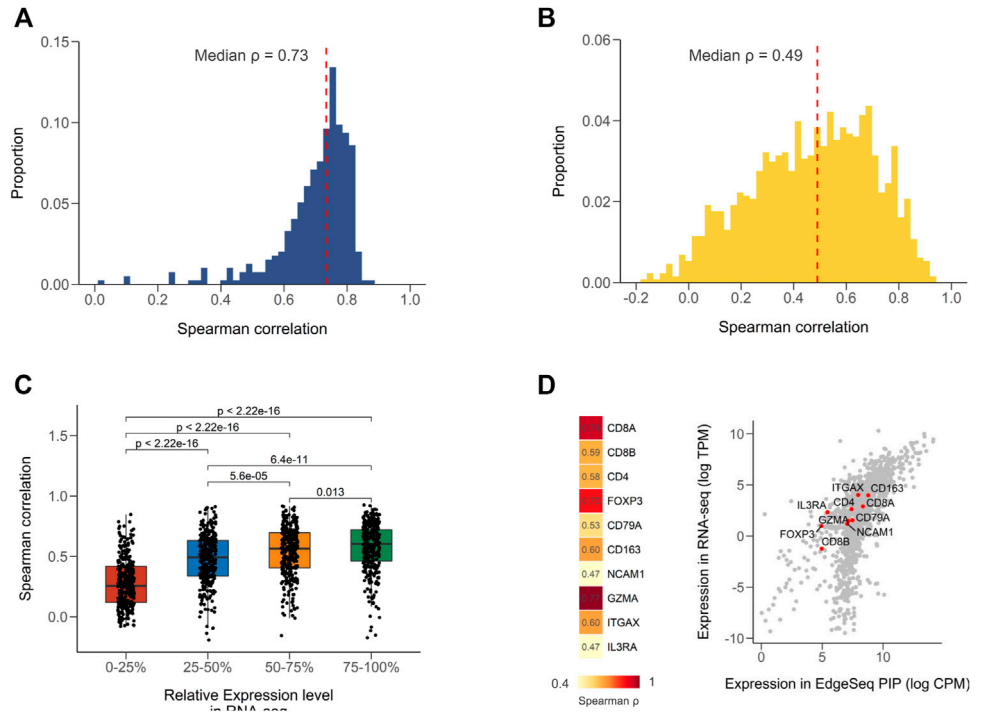


FIGURE 1 | Concordance of EdgeSeq PIP with RNA-seq at gene level. **(A)** Distribution of sample-wise Spearman correlation coefficients between EdgeSeq PIP and RNA-seq. Dashed line represents the median value. **(B)** Distribution of gene-wise Spearman correlation coefficients between EdgeSeq PIP and RNA-seq. Dashed line represents the median value. **(C)** Boxplot showing the difference of Spearman correlation coefficients across genes stratified by their relative expression level (quartile) in RNA-seq. **(D)** Heatmap showing the Spearman correlation of key immune cell markers (left) and scatter plot showing their expression levels (right) in EdgeSeq PIP (*x*-axis) and RNA-seq (*y*-axis).

to strong correlations (ρ range from 0.47 to 0.77) (**Figure 1D** and **Supplementary Figure S1C**).

We also performed comparisons of EdgeSeq PIP and NanoString in another 17 samples and found strong correlations (median ρ = 0.81) using the sample-by-sample approach and moderate correlations (median ρ = 0.57) using the gene-by-gene approach (**Supplementary Figures S2A,B**, **Supplementary Table S4**). Similar to the results in RNA-seq, the immune markers assessed by EdgeSeq PIP also showed a high level of concordance with NanoString (6/9 genes had strong correlations with ρ > 0.7) (**Supplementary Figure S2C**). Taken together, our findings showed that EdgeSeq PIP had a high degree of agreement with RNA-seq and NanoString at the single-gene level.

High Concordance Between EdgeSeq Precision Immuno-Oncology Panel and RNA-Seq for Characteristic Tumor Microenvironment Gene Signatures

Gene signature, a set of genes with similar expression patterns or biological functions, have been widely used in TME profiling (Liberzon et al., 2011). Compared to single gene, signature-based approach has many significant advantages, including dimension reduction and greater biological interpretability. Here, we used 29 signatures that successfully classified the TMEs of tumors from TCGA and investigated the applicability of EdgeSeq PIP for TME subtyping (Bagaev et al., 2021) (**Supplementary Table S5**). First, we assessed the concordance of gene set variation analysis (GSVA) scores derived from EdgeSeq PIP and RNA-seq (Materials and Methods). Most signatures had moderate positive correlations (median ρ = 0.61), which persisted even after samples were divided into smaller groups according to cancer types (**Figure 2A**). Nevertheless, five signatures: neutrophil, Th1, Th2, Treg, and Treg traffic had negligible correlations (median ρ = 0.01) between the two platforms (**Figure 2A**). This result may have been due to the incomplete coverage of gene list in certain signatures by EdgeSeq PIP or to the involvement of genes with low correlations between the two platforms. We first explored whether the percentage of missed genes in each signature influenced the correlation. Surprisingly, the signatures with poor correlation did not have significantly higher rates of missing genes than the others (**Supplementary Figure S3B**). In contrast, most genes in these five signatures with worse correlations had lower RNA-seq expression levels and thus weaker correlations with EdgeSeq PIP at the single-gene level (**Figure 2B** and **Supplementary Figure S3C**). This result suggested that not the incompletion of genes in the signature but rather the correlation at single-gene level drove the concordance of signatures. After removing these weakly correlated signatures, the remaining signatures had significantly higher correlations as well as lower variations

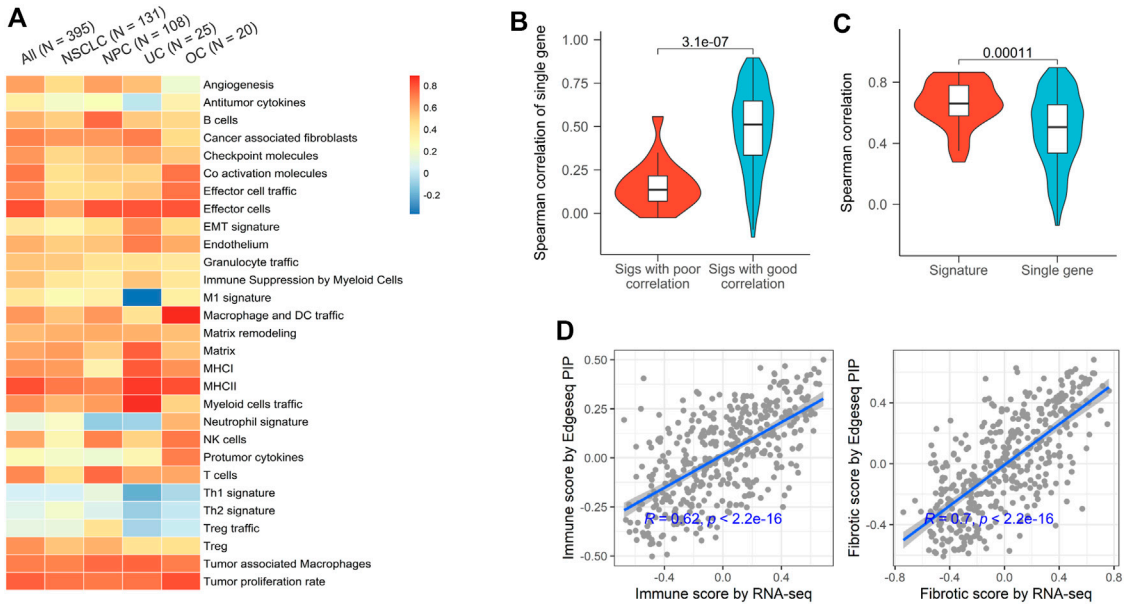


FIGURE 2 | Concordance of EdgeSeq PIP and RNA-seq for TME characterizing signatures. **(A)** Heatmap showing the Spearman correlation coefficients of 29 TME-characterizing signatures between EdgeSeq PIP and RNA-seq. Columns represent the overall correlation and correlation in each major cancer type. **(B)** Violin plots showing the gene-wise Spearman correlation coefficients of genes in poor and well correlated signatures. **(C)** Violin plots showing the distribution of signature-derived and gene-derived Spearman correlation coefficients (between EdgeSeq PIP and RNA-seq). **(D)** Scatterplot showing the correlation of immune scores (left) and fibrotic scores (right) estimated by EdgeSeq PIP and RNA-seq data.

than individual genes (**Figure 2C**), which indicated that signature-based analysis had the property of noise reduction.

In the previous study, the TME of TCGA tumors were classified into 4 subgroups and annotated according to their enrichment of immune or fibroblast components (Bagaev et al., 2021). By averaging the scores of corresponding signatures (**Supplementary Table S5**), we derived "immune enrichment" and "fibroblast enrichment" score for each sample and found the scores generated by EdgeSeq PIP and RNA-seq correlated well (ρ = 0.62 and 0.70 separately) (**Figure 2D**). Taken together, the data from EdgeSeq PIP generated reliable signature scores and TME subtype results.

High Concordance Between EdgeSeq Precision Immuno-Oncology Panel and Other Gene Expression Profiling Platforms Regarding Potential Immune Checkpoint Inhibitor-Predictive Gene Signatures

We chose six well-known ICI-predictive signatures and studied the fidelity of calculating them using EdgeSeq PIP data (**Supplementary Table S6**) (Denkert et al., 2015; Cristescu et al., 2018; Socinski et al., 2018; Motzer et al., 2020; Roelands et al., 2020; Sangro et al., 2020). By calculating the correlation between signature scores derived from EdgeSeq PIP and RNA-seq, we found a high degree of concordance for all signatures (median ρ = 0.84) (**Figure 3A** and **Supplementary Figure S4A**). As expected, most genes in these signatures were well-correlated between the two platforms (median ρ = 0.57) (**Supplementary Figure S4B**).

Among these signatures, tumor inflammation signature (TIS) is originally derived from the NanoString platform using predefined algorithm and weight (Cristescu et al., 2018). Using the 17 samples with both EdgeSeq PIP and NanoString data, we applied GSVA on EdgeSeq PIP data to get "TIS-PIP score" and then compared it with the "TIS-NanoString score" derived from its original algorithm. We found a moderate correlation (ρ = 0.61) of TIS score despite of different platforms and algorithms (**Figure 3B**). In addition to GSVA, we also tested other methods for EdgeSeq PIP, such as ssGSEA and Z-score (Materials and Methods) and found consistent results (**Supplementary Figures S5A,B**). We reasoned that in addition to the cross-platform concordance at the single-gene level (**Supplementary Figure S5C**), the strong intra-signature correlation of genes could also account for the high correlation coefficient. Indeed, we found that except for *CD276*, all other 17 genes were moderately correlated with each other (median ρ = 0.46) (**Figure 3C** and **Supplementary Figure S5D**).

Additionally, we assessed tumor immune dysfunction and exclusion (TIDE), a computational framework integrating multiple signatures of tumor evasion that successfully predicts the responses of melanoma and lung cancer patients to ICIs (Jiang et al., 2018). We found that the TIDE scores generated by EdgeSeq PIP had moderate correlations with those generated by RNA-seq (ρ = 0.55) (**Figure 3D**). Overall, our results suggested that data from EdgeSeq PIP can be reliably used for most well-known ICI efficacy prediction algorithms.

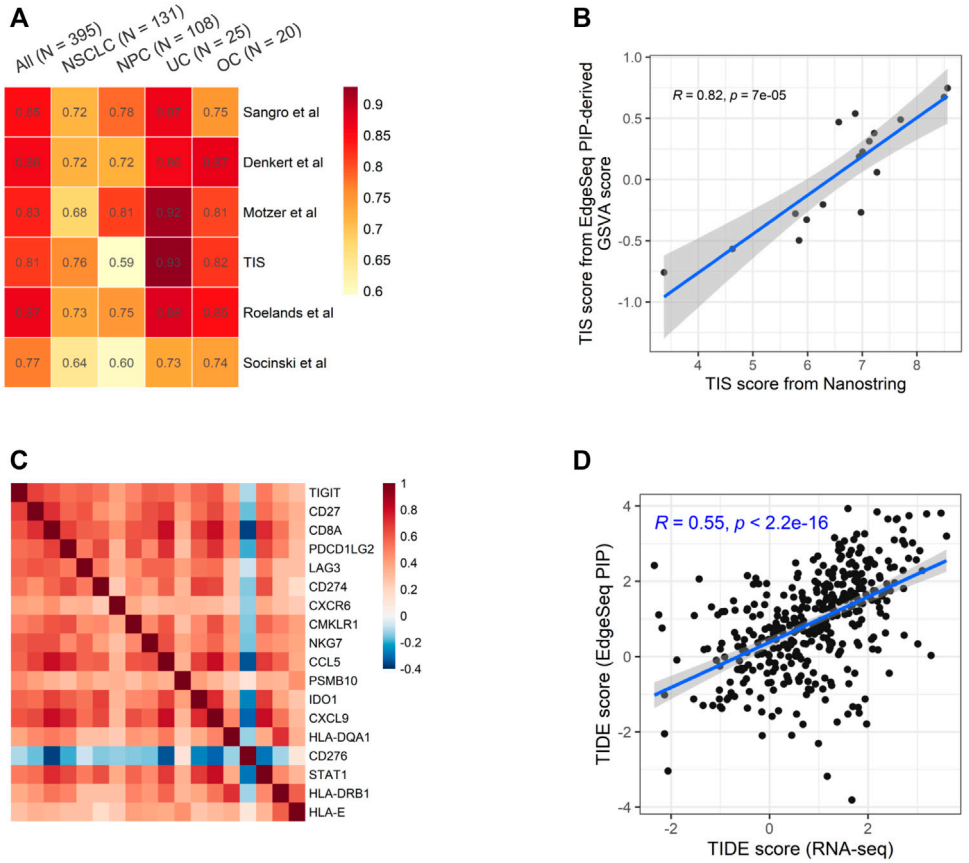


FIGURE 3 | Concordance of EdgeSeq PIP and RNA-seq for potential ICI-predictive signatures **(A)** Heatmap showing the correlation coefficients of six ICI-predictive signatures between EdgeSeq and RNA-seq. Columns represent the overall correlation and correlation in each major cancer type. **(B)** Scatter plot showing the correlation of GSVA score derived from EdgeSeq PIP data with the official TIS scores from NanoString IO360 assay. **(C)** Heatmap showing the co-correlation of genes within TIS signature using RNA-seq data. **(D)** Scatter plot showing the correlation of TIDE score calculated from RNA-seq data and EdgeSeq PIP data.

High Consistency Between EdgeSeq Precision Immuno-Oncology Panel and H & E Staining Results

For 1,174 samples assessed by EdgeSeq PIP, H & E staining was also performed to evaluate the percentages of tumor cells and immune cells. Here, we assessed whether the H&E staining results could be recapitulated by EdgeSeq PIP at mRNA level. First, we used ESTIMATE (Yoshihara et al., 2013) to calculate the tumor purity score for each sample based on EdgeSeq PIP data. We found consistent result with H & E staining: the samples with a high tumor percentage (>70%, assessed by H & E) had significant higher tumor purity scores than those with low tumor percentage (<70%) (**Figure 4A**). In addition, we identified the differentially expressed genes between these two groups and performed gene set enrichment analysis (GSEA) (Materials and Methods) (**Supplementary Table S7**). Our results showed that cell cycle-related pathways were enriched among the genes upregulated in samples with >70% tumor percentage (**Figure 4B** and **Supplementary Figures S6A,B**). This finding aligned with the H & E result as most tumor cells had an extremely high proliferation rate.

Similarly, we studied the concordance of EdgeSeq PIP with the percentage of infiltrated immune cells estimated by H & E. On the one hand, tumors with >5% immune cell percentage (by H & E) had significant higher ImmunoScores (calculated by ESTIMATE) than others (**Figure 4C**). On the other hand, immune-related pathways and representative immune marker genes such as *CD8A*, *STAT1* and *GZMK* were up-regulated in tumors with >5% immune percentage (**Supplementary Figures S6C,D**, **Supplementary Table S8**). Overall, this evidence suggested that the abundance of tumor or immune cells assessed by histological methods was well reflected by EdgeSeq PIP.

High Consistency Between EdgeSeq Precision Immuno-Oncology Panel and Immunohistochemistry

PD-L1 IHC score has been widely used for predicting responses to ICI-based therapies. A previous study showed that the mRNA level of PD-L1 assessed RNA-seq had good concordance with its protein level assessed by IHC (Conroy et al., 2019). Therefore, the correlation with PD-L1 IHC could be used to evaluate the fidelity of NGS-based approaches. In 369 samples with PD-L1 IHC data

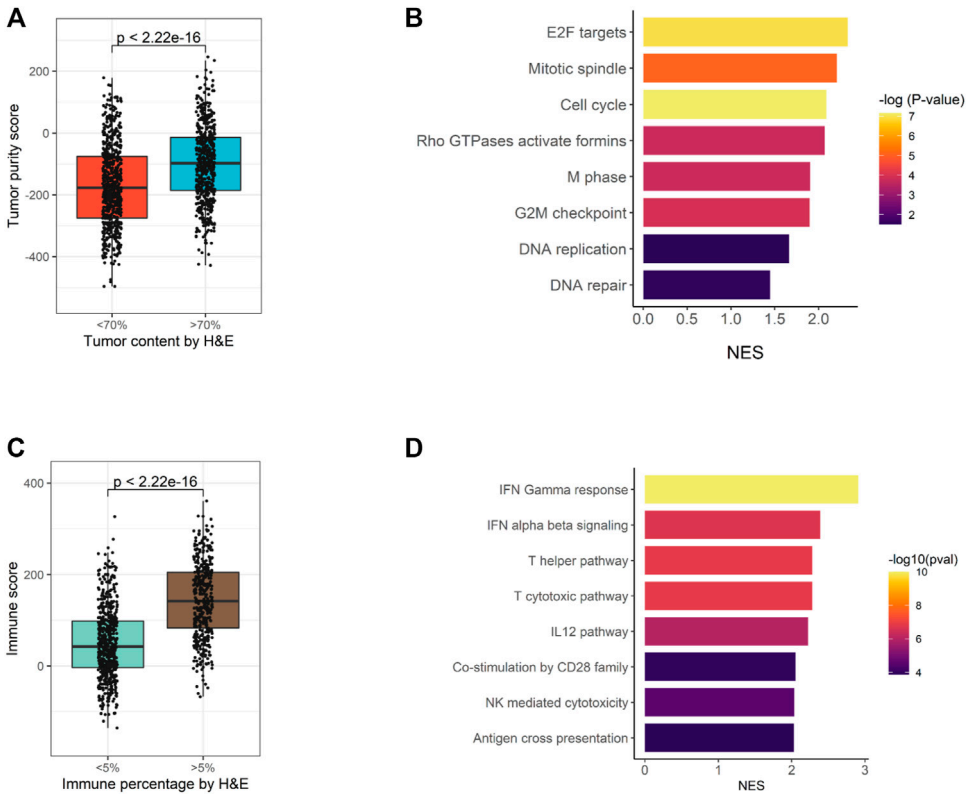


FIGURE 4 | Consistency of EdgeSeq PIP data with H&E staining. **(A)** Boxplot showing the difference in tumor purity scores between samples with high (>70%) tumor percentage and low (<70%) tumor percentage. **(B)** Barplot showing the top eight pathways enriched in tumors with high (>70%) tumor percentage. **(C)** Boxplot showing the difference in immune scores between samples with high (>5%) immune percentage and low (<5%) immune percentage. **(D)** Barplot showing the top eight pathways enriched in tumors with high (>5%) immune percentage.

(**Supplementary Figure S8**), we found that the percentage of PD-L1 positive tumor cells had a moderate correlation ($\rho = 0.65$) with the mRNA level of PD-L1 (*CD274*) assessed by EdgeSeq PIP (**Figure 5A**), which was comparable to its association with RNA-seq ($\rho = 0.69$). It suggested that the remaining lack of correlation was mainly due to the discrepancy between mRNA and protein, rather than technical issues specific to platform. In addition, we calculated the correlation of PD-L1 IHC with every gene and found that *CD274* had the highest correlation coefficient among all genes (**Figure 5B**), which further suggested the fidelity of EdgeSeq PIP data.

To further investigate the concordance of EdgeSeq PIP with IHC, we assessed another important immuno-oncology marker, the fraction of $CD8^+$ T cells, by CD8A IHC in another 98 tumors (**Supplementary Figure S9**). We investigated the correlation of CD8 abundance estimated by IHC with the expression level of 4 traditional $CD8^+$ T cell markers (*CD8A*, *CD8B*, *PRF1*, *GZMA*) from EdgeSeq PIP and the EdgeSeq PIP-derived GSVA scores of 4 well-known $CD8^+$ T signatures (**Supplementary Table S9**) (Rooney et al., 2015; Danaher et al., 2017; Davoli et al., 2017; Prat et al., 2017). All estimates for $CD8^+$ T cells had significant positive correlations with the IHC results (**Figure 5C**). Interestingly, *CD8A* outperformed the other genes ($\rho = 0.55$) and even the gene signatures. This result suggested that at least for data generated by EdgeSeq PIP, the expression level of *CD8A* is sufficient for estimating the fraction of $CD8^+$ T cells in tumors. Overall, our results indicated that the data generated by EdgeSeq PIP were comparable to IHC data and could be used as a complementary method for evaluating the protein levels of PD-L1 and the $CD8^+$ T cell fraction.

DISCUSSION

Gene expression profiling plays an important role in immuno-oncology related biomarker studies, as it helps to elucidate the TME landscape, and many predictive biomarkers have been derived for ICI-based therapies. As clinical specimens in certain cancers are obtained from core needle biopsies of small size and suboptimal quality, tissue availability restricts the clinical application of traditional GEP methods (e.g., RNA-seq). EdgeSeq PIP, an assay focused on well-studied genes related to immuno-oncology, uses an extraction-free technology to minimize sample requirements. This method mitigates the challenge for sample acquisition and has been widely used in clinical trials.

In this study, we provided evidence that EdgeSeq PIP is a robust assay with a high degree of concordance with multiple platforms, including RNA-seq, tumor/immune cell fraction

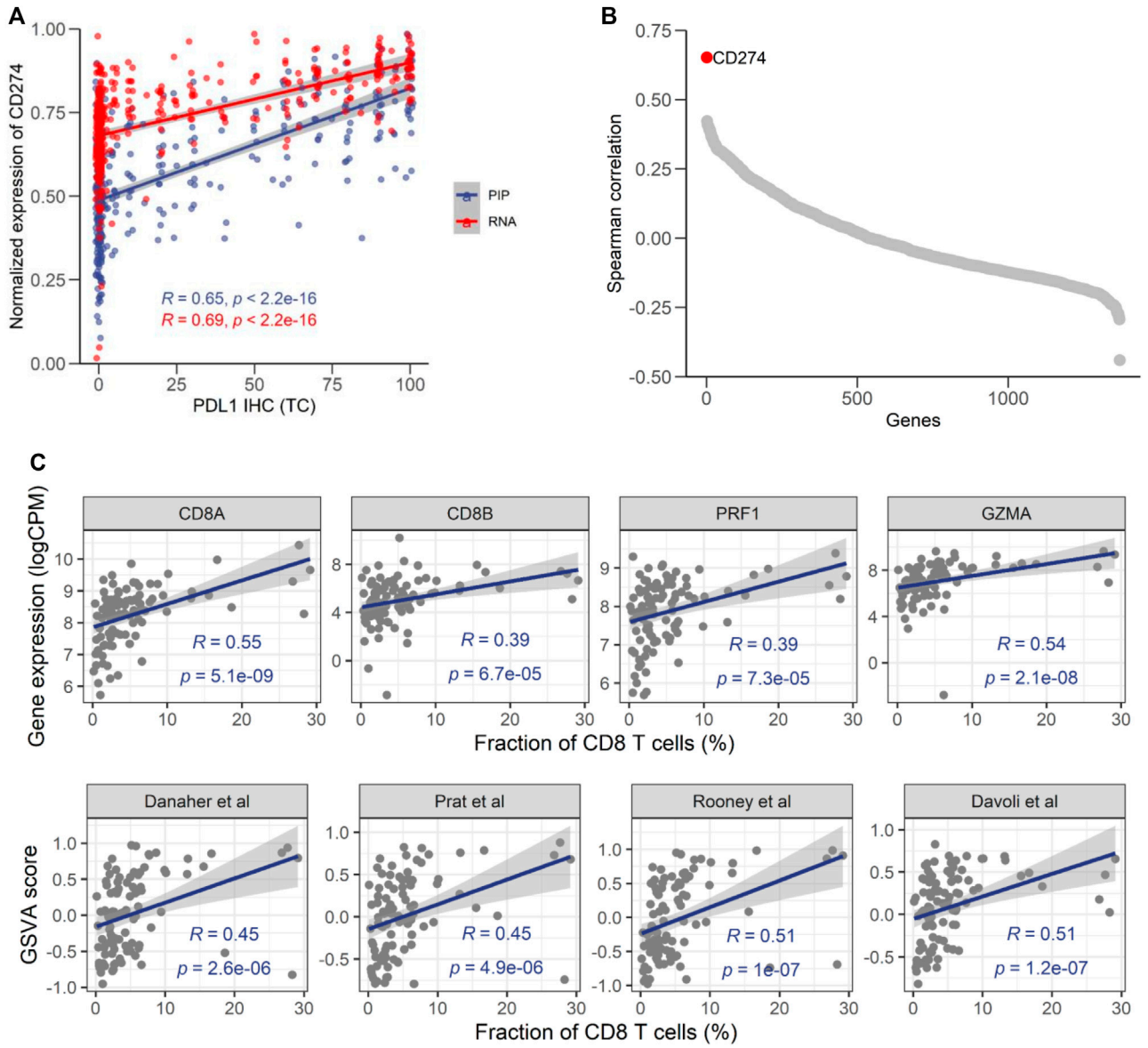


FIGURE 5 | Consistency of EdgeSeq PIP with PD-L1 IHC and CD8 IHC. **(A)** Correlation of the CD274 mRNA expression level from EdgeSeq PIP (blue) and RNA-seq (red) with the PD-L1 TC score. **(B)** Spearman correlation coefficients (*y*-axis) of mRNA expression level assessed by EdgeSeq PIP with the PD-L1 level assessed by IHC. Each dot represents a gene and genes were ordered according to their Spearman correlation coefficients (*x*-axis). **(C)** Scatterplot showing the correlation of CD8 related genes (top) and signatures (bottom) with $CD8^+$ T cell fraction estimated by IHC (%).

estimation by H & E staining, PD-L1 scoring and $CD8^+$ T cell abundance estimation by IHC. By analyzing data from 395 clinical FFPE samples, we found that EdgeSeq PIP showed good correlation with RNA-seq at the single-gene level, especially after excluding genes with low expression levels or dynamic ranges. Importantly, we demonstrated that EdgeSeq PIP, with incomplete coverage of genes in some signatures, still generated signature scores that were well correlated with those of RNA-seq. This result indicated that EdgeSeq PIP data can be reliably utilized for signature-based analyses of TME subtypes and ICI efficacy prediction. In addition, we found that EdgeSeq PIP aligned well with the overall percentage of tumor/immune cells determined by H & E staining. Furthermore, our results showed that the RNA levels determined by EdgeSeq PIP aligned well with the protein levels determined by the IHC assessment of two important immunotherapy biomarkers, the fraction of PD-L1-positive tumor cells and $CD8^+$ T cells.

As the method of choice for transcriptome probing, RNA-seq has always been used as a reference standard to benchmark other GEP assays. In our results, EdgeSeq PIP correlated well with RNA-seq except for a small proportion of genes with low expression levels. This result was within expectations, as true signals for genes with low expression are hard to distinguish from noise, which would diminish the correlation (Kwong

et al., 2018). In addition, loss of degraded transcripts during RNA extraction, especially for low-expressing genes in archived FFPE samples, could further exacerbate the discordancy between the two platforms. There was evidence that EdgeSeq PIP, using its extraction-free technique, offer superior sensitivity on these "discordant" genes (Ran et al., 2020). Nevertheless, such genes should be used with caution and should even excluded from some analyses involving data from both RNA-seq and EdgeSeq PIP.

In this study, we found that the gene signature score derived from EdgeSeq PIP was strongly correlated with that from RNA-seq, although some genes in given signatures were not covered by EdgeSeq PIP probes. We reasoned that such result was attributed to the coexpression of genes within the signature (e.g., TIS). Such coexpression led to gene "redundancy" and made the signature score calculation robustly susceptible to gene dropout and different algorithms. Though some poorly correlated genes resulted in a low correlation among certain signatures, after their exclusion, most signatures defined based on RNA-seq data could be directly transferred and used on EdgeSeq PIP data. However, we must admit that the signatures used in our study were limited to those related to immuno-oncology, which predominantly consisted of surface markers sharing similar expression patterns. For other signatures, such as cellular signaling pathways, the effect of "gene dropout" needs to be further investigated.

PD-L1 IHC has been used to predict responses to ICI-based therapy. A previous study revealed that the expression of PD-L1 measured by RNA-seq had a strong association with that measured by IHC (Conroy et al., 2019). Here, we found that the mRNA level of PD-L1 generated by EdgeSeq PIP had a high correlation with the PD-L1 IHC level as well. Although a biological gap exists between mRNAs and proteins, EdgeSeq PIP at least provides another layer of PD-L1 quantification that can be exploited for the prediction of ICI efficacy. In addition, EdgeSeq PIP quantifies PD-L1 transcripts without the need for subjective scoring and cell type discrimination by pathologists, which might introduce substantial discrepancy across studies (Hirsch et al., 2017).

In summary, for the first time, we comprehensively benchmarked EdgeSeq PIP with multiple platforms using large-scale clinical FFPE samples, and the results are reflective of its actual performance in clinical practice. Our results showed that EdgeSeq PIP generated data comparable to those generated by classical methods. Considering its low sample requirement and cost efficiency, the technology would get an ever-increasing application for biomarker studies in clinical trials.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by please refer to the clinical trials mentioned in our manuscript. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YS, XM, and PZ conceived the study. YS and XM performed the bioinformatics analysis. WS, SL, and TL analyzed the immunohistochemistry data. YS, XM, WS, and PZ wrote the article. All authors discussed the results and commented on the manuscript.

FUNDING

This study was funded by BeiGene. Funding for open access charge: BeiGene.

ACKNOWLEDGMENTS

We thank Jayesh Desai (Peter MacCallum Cancer Centre, University of Melbourne), Jun Guo (Harbin Medical University Cancer Hospital), Lin Shen (Peking University Cancer Hospital and Institute), Jie Wang (Cancer Institute and Hospital, Chinese academy of Medical Sciences), Yilong Wu (Guangdong Lung Cancer Institute), Jianmin Xu (The fifth Medical Center, Chinese PLA General Hospital), Dingwei Ye (Fudan University) and Li Zhang (Sun Yat-sen University) for leading the clinical trials included in this study. We also thank Mo Liu (BeiGene) for his kind suggestions on this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2022.899353/full#supplementary-material

REFERENCES

Aisner, D. L., Rumery, M. D., Merrick, D. T., Kondo, K. L., Nijmeh, H., Linderman, D. J., et al. (2016). Do More with Less: Tips and Techniques for Maximizing Small Biopsy and Cytology Specimens for Molecular and Ancillary Testing: The University of Colorado Experience. *Arch. Pathol. Lab. Med.* 140, 1206–1220. doi:10.5858/arpa.2016-0156-ra

Anguiano, E., Desai, K., Hayati, S., Szabo, P. M., Pant, S., Wang, L., et al. (2020). Abstract 1997: Comparison of Gene Expression Profiling Platforms: Translatability of Tumor Inflammation Gene Signatures. *Cancer Res.* 80, 1997–1997. doi:10.1158/1538-7445.am2020-1997

Bagaev, A., Kotlov, N., Nomie, K., Svekolkin, V., Gafurov, A., Isaeva, O., et al. (2021). Conserved Pan-Cancer Microenvironment Subtypes Predict Response to Immunotherapy. *Cancer Cell* 39, 845–865. doi:10.1016/j.ccell.2021.04.014

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

Chen, D. S., and Mellman, I. (2017). Elements of Cancer Immunity and the Cancer-Immune Set Point. *Nature* 541, 321–330. doi:10.1038/nature21349

Conroy, J. M., Pabla, S., Nesline, M. K., Glenn, S. T., Papanicolau-Sengos, A., Burgher, B., et al. (2019). Next Generation Sequencing of PD-L1 for Predicting Response to Immune Checkpoint Inhibitors. *J. Immunother. Cancer* 7, 18. doi:10.1186/s40425-018-0489-5

Cristescu, R., Mogg, R., Ayers, M., Albright, A., Murphy, E., Yearley, J., et al. (2018). Pan-tumor Genomic Biomarkers for PD-1 Checkpoint Blockade-Based Immunotherapy. *Science* 362. doi:10.1126/science.aar3593

Danaher, P., Warren, S., Dennis, L., D'Amico, L., White, A., Disis, M. L., et al. (2017). Gene Expression Markers of Tumor Infiltrating Leukocytes. *J. Immunother. cancer* 5, 18. doi:10.1186/s40425-017-0215-8

Davoli, T., Uno, H., Wooten, E. C., and Elledge, S. J. (2017). Tumor Aneuploidy Correlates with Markers of Immune Evasion and with Reduced Response to Immunotherapy. *Science* 355. doi:10.1126/science.aaf8399

Denkert, C., von Minckwitz, G., Brase, J. C., Sinn, B. V., Gade, S., Kronenwett, R., et al. (2015). Tumor-infiltrating Lymphocytes and Response to Neoadjuvant Chemotherapy with or without Carboplatin in Human Epidermal Growth Factor Receptor 2-positive and Triple-Negative Primary Breast Cancers. *Jco* 33, 983–991. doi:10.1200/jco.2014.58.1967

Desai, J., Zhou, Q., Deva, S., Zhao, J., Wang, J., Tan, W., et al. (2020). 76 Potential Mechanisms of Resistance Identified through Analysis of Multiple Biomarkers in Immune Hot Non-responders with Non-small Cell Lung Cancer (NSCLC) Treated with Tislelizumab. *J. Immunother. Cancer* 8, A47–A47.

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

Evers, D. L., Fowler, C. B., Cunningham, B. R., Mason, J. T., and O'Leary, T. J. (2011). The Effect of Formaldehyde Fixation on RNA. *J. Mol. Diagnostics* 13, 282–288. doi:10.1016/j.jmoldx.2011.01.010

Garg, S. K., Welsh, E. A., Fang, B., Hernandez, Y. I., Rose, T., Gray, J., et al. (2020). Multi-Omics and Informatics Analysis of FFPE Tissues Derived from Melanoma Patients with Long/Short Responses to Anti-PD1 Therapy Reveals Pathways of Response. *Cancers (Basel)* 12. doi:10.3390/cancers12123515

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

Havel, J. J., Chowell, D., and Chan, T. A. (2019). The Evolving Landscape of Biomarkers for Checkpoint Inhibitor Immunotherapy. *Nat. Rev. Cancer* 19, 133–150. doi:10.1038/s41568-019-0116-x

Hirsch, F. R., McElhinny, A., Stanforth, D., Ranger-Moore, J., Jansson, M., Kulangara, K., et al. (2017). PD-L1 Immunohistochemistry Assays for Lung Cancer: Results from Phase 1 of the Blueprint PD-L1 IHC Assay Comparison Project. *J. Thorac. Oncol.* 12, 208–222. doi:10.1016/j.jtho.2016.11.2228

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

Kwong, L. N., De Macedo, M. P., Haydu, L., Joon, A. Y., Tetzlaff, M. T., Calderone, T. L., et al. (2018). Biological Validation of RNA Sequencing Data from Formalin-Fixed Paraffin-Embedded Primary Melanomas. *JCO Precis. Oncol.* 2018. doi:10.1200/po.17.00259

Law, C. W., Chen, Y., Shi, W., and Smyth, G. K. (2014). Voom: Precision Weights Unlock Linear Model Analysis Tools for RNA-Seq Read Counts. *Genome Biol.* 15, R29. doi:10.1186/gb-2014-15-2-r29

Li, B., and Dewey, C. N. (2011). RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome. *BMC Bioinforma.* 12, 323. doi:10.1186/1471-2105-12-323

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P., and Mesirov, J. P. (2011). Molecular Signatures Database (MSigDB) 3.0. *Bioinformatics* 27, 1739–1740. doi:10.1093/bioinformatics/btr260

Martel, R. R., Botros, I. W., Rounseville, M. P., Hinton, J. P., Staples, R. R., Morales, D. A., et al. (2002). Multiplexed Screening Assay for mRNA Combining Nuclease Protection with Luminescent Array Detection. *ASSAY Drug Dev. Technol.* 1, 61–71. doi:10.1089/154065802761001310

Martin-Broto, J., Cruz, J., Penel, N., Le Cesne, A., Hindi, N., Luna, P., et al. (2020). Pazopanib for Treatment of Typical Solitary Fibrous Tumours: a Multicentre, Single-Arm, Phase 2 Trial. *Lancet Oncol.* 21, 456–466. doi:10.1016/s1470-2045(19)30826-5

Motzer, R. J., Robbins, P. B., Powles, T., Albiges, L., Haanen, J. B., Larkin, J., et al. (2020). Avelumab Plus Axitinib versus Sunitinib in Advanced Renal Cell Carcinoma: Biomarker Analysis of the Phase 3 JAVELIN Renal 101 Trial. *Nat. Med.* 26, 1733–1741. doi:10.1038/s41591-020-1044-8

Prat, A., Navarro, A., Paré, L., Reguart, N., Galván, P., Pascual, T., et al. (2017). Immune-Related Gene Expression Profiling after PD-1 Blockade in Non-small Cell Lung Carcinoma, Head and Neck Squamous Cell Carcinoma, and Melanoma. *Cancer Res.* 77, 3540–3550. doi:10.1158/0008-5472.can-16-3556

Qi, Z., Wang, L., Desai, K., Cogswell, J., Stern, M., Lawson, B., et al. (2019). Reliable Gene Expression Profiling from Small and Hematoxylin and Eosin-Stained Clinical Formalin-Fixed, Paraffin-Embedded Specimens Using the HTG EdgeSeq Platform. *J. Mol. Diagnostics* 21, 796–807. doi:10.1016/j.jmoldx.2019.04.011

Qi, Z., Wang, L., He, A., Ma-Edmonds, M., and Cogswell, J. (2016). Evaluation and Selection of a Non-PCR Based Technology for Improved Gene Expression Profiling from Clinical Formalin-Fixed, Paraffin-Embedded Samples. *Bioanalysis* 8, 2305–2316. doi:10.4155/bio-2016-0186

Ran, D., Moharil, J., Lu, J., Gustafson, H., Culm-Merdek, K., Strand-Tibbitts, K., et al. (2020). Platform Comparison of HTG EdgeSeq and RNA-Seq for Gene Expression Profiling of Tumor Tissue Specimens. *Jco* 38, 3566–3566. doi:10.1200/jco.2020.38.15_suppl.3566

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, e47. doi:10.1093/nar/gkv007

Roelands, J., Hendrickx, W., Zoppoli, G., Mall, R., Saad, M., Halliwill, K., et al. (2020). Oncogenic States Dictate the Prognostic and Predictive Connotations of Intratumoral Immune Response. *J. Immunother. Cancer* 8. doi:10.1136/jitc-2020-000617

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, 48–61. doi:10.1016/j.cell.2014.12.033

Sangro, B., Melero, I., Wadhawan, S., Finn, R. S., Abou-Alfa, G. K., Cheng, A.-L., et al. (2020). Association of Inflammatory Biomarkers with Clinical Outcomes in Nivolumab-Treated Patients with Advanced Hepatocellular Carcinoma. *J. Hepatology* 73, 1460–1469. doi:10.1016/j.jhep.2020.07.026

Sanmamed, M. F., and Chen, L. (2018). A Paradigm Shift in Cancer Immunotherapy: From Enhancement to Normalization. *Cell* 175, 313–326. doi:10.1016/j.cell.2018.09.035

Schober, P., Boer, C., and Schwarte, L. A. (2018). Correlation Coefficients. *Anesth. Analgesia* 126, 1763–1768. doi:10.1213/ane.0000000000002864

Socinski, M. A., Jotte, R. M., Cappuzzo, F., Orlandi, F., Stroyakovskiy, D., Nogami, N., et al. (2018). Atezolizumab for First-Line Treatment of Metastatic Nonsquamous NSCLC. *N. Engl. J. Med.* 378, 2288–2301. doi:10.1056/nejmoa1716948

Song, Y., Gao, Q., Zhang, H., Fan, L., Zhou, J., Zou, D., et al. (2021). Tislelizumab for Relapsed/refractory Classical Hodgkin Lymphoma: 3-year Follow-Up and Correlative Biomarker Analysis. *Clin. Cancer Res.* 28, 1147–1156. doi:10.1158/1078-0432.CCR-21-2023

Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Ou Yang, T. H., et al. (2018). The Immune Landscape of Cancer. *Immunity* 48, 812–e14. doi:10.1016/j.immuni.2018.03.023

Vaddepally, R. K., Kharel, P., Pandey, R., Garje, R., and Chandra, A. B. (2020). Review of Indications of FDA-Approved Immune Checkpoint Inhibitors Per NCCN Guidelines with the Level of Evidence. *Cancers (Basel)* 12. doi:10.3390/cancers12030738

Wang, L., Saci, A., Szabo, P. M., Chasalow, S. D., Castillo-Martin, M., Domingo-Domenech, J., et al. (2018). EMT- and Stroma-Related Gene Expression and Resistance to PD-1 Blockade in Urothelial Cancer. *Nat. Commun.* 9, 3503. doi:10.1038/s41467-018-05992-x

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-seq: a Revolutionary Tool for Transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi:10.1038/nrg2484

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

Conflict of Interest: YS, XM, WS, TL, LL, SL, ZS, YZ, and PZ were employed by the company BeiGene (Beijing) Co., Ltd.

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.2022.897568

Identification of Hub Genes in Colorectal Adenocarcinoma by Integrated Bioinformatics

Yang Liu[1], *Lanlan Chen*[2], *Xiangbo Meng*[1], *Shujun Ye*[1] *and Lianjun Ma*[1]*

[1]*Endoscopy Center, China-Japan Union Hospital of Jilin University, Changchun, China,* [2]*Department of Hepatobiliary and Pancreatic Surgery, The First Hospital of Jilin University, Changchun, China*

An improved understanding of the molecular mechanism of colorectal adenocarcinoma is necessary to predict the prognosis and develop new target gene therapy strategies. This study aims to identify hub genes associated with colorectal adenocarcinoma and further analyze their prognostic significance. In this study, The Cancer Genome Atlas (TCGA) COAD-READ database and the gene expression profiles of GSE25070 from the Gene Expression Omnibus were collected to explore the differentially expressed genes between colorectal adenocarcinoma and normal tissues. The weighted gene co-expression network analysis (WGCNA) and differential expression analysis identified 82 differentially co-expressed genes in the collected datasets. Enrichment analysis was applied to explore the regulated signaling pathway in colorectal adenocarcinoma. In addition, 10 hub genes were identified in the protein–protein interaction (PPI) network by using the cytoHubba plug-in of Cytoscape, where five genes were further proven to be significantly related to the survival rate. Compared with normal tissues, the expressions of the five genes were both downregulated in the GSE110224 dataset. Subsequently, the expression of the five hub genes was confirmed by the Human Protein Atlas database. Finally, we used Cox regression analysis to identify genes associated with prognosis, and a 3-gene signature (CLCA1–CLCA4–GUCA2A) was constructed to predict the prognosis of patients with colorectal cancer. In conclusion, our study revealed that the five hub genes and CLCA1–CLCA4–GUCA2A signature are highly correlated with the development of colorectal adenocarcinoma and can serve as promising prognosis factors to predict the overall survival rate of patients.

Keywords: colorectal adenocarcinoma, differential gene expression analysis, weighted gene co-expression network analysis, tumor biomarkers, predictive model

OPEN ACCESS

Edited by:
Fangqing Zhao,
Beijing Institutes of Life Science (CAS), China

Reviewed by:
Jinyang Zhang,
Beijing Institutes of Life Science (CAS), China
Dong Sun,
Shandong University, China
Yong Zhang,
Jilin University, China

***Correspondence:**
Lianjun Ma
horsejlm@jlu.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: *16 March 2022*
Accepted: *20 April 2022*
Published: *27 May 2022*

Citation:
Liu Y, Chen L, Meng X, Ye S and Ma L (2022) Identification of Hub Genes in Colorectal Adenocarcinoma by Integrated Bioinformatics. Front. Cell Dev. Biol. 10:897568. doi: 10.3389/fcell.2022.897568

INTRODUCTION

With the increasing incidence rate of colorectal cancer (CRC) worldwide, it is considered as one of the leading causes of cancer deaths (Siegel et al., 2020). It is expected to cause about 53,200 deaths by 2020 (Colorectal Cancer Statistics | How Common Is Colorectal Cancer? 2021). Recently, combined therapies including surgery, chemotherapy, targeted therapy, and radiotherapy have prolonged the overall survival (OS) of patients with colorectal cancer (van der Geest et al., 2015). However, distant metastasis and drug resistance are still the main reasons for the poor prognosis effect of colorectal cancer patients. At present, the exact carcinogenic molecular mechanism of colorectal cancer is not precise, and no effective prognostic biomarkers have been thoroughly investigated. Therefore, it is

necessary to explore the molecular mechanism of the proliferation and progression of colorectal cancer to find promising prognostic biomarkers and formulate effective clinical treatment strategies.

With the rapid development of sequencing technology, bioinformatics is increasingly widely used in gene expression profiling to study the molecular mechanism of diseases and find disease-specific biomarkers (Can, 2014). Weighted gene co-expression network analysis (WGCNA) is an effective tool to construct the related networks and identify hub genes, which is widely used to find tumor biomarkers (Langfelder and Horvath, 2008). Highly related genes may be functionally related and can be clustered into a module by WGCNA. The correlation between modules with clinical characteristics can be quantified and helps to identify modules of interest. In addition, differential gene expression analysis can also provide an essential method for studying the molecular mechanism of genome regulation and revealing the quantitative changes in expression levels between the experimental and control groups, which might help us find new colorectal cancer biomarkers. Therefore, two methods are used, combining the results of WGCNA and differential gene expression analysis to enhance the recognition ability of highly related genes, which is helpful to be used as candidate biomarkers (San Segundo-Val and Sanz-Lozano, 2016).

In this study, WGCNA and differential gene expression analysis was performed to analyze the mRNA expression data of COAD-READ in TCGA and GEO databases, and the differentially co-expressed genes were obtained. We further explored the biological function of these differentially co-expressed genes of COAD-READ using functional enrichment and protein interaction (PPI) analysis combined with survival analysis. This study provides a possible basis to understand the regulating mechanism of COAD-READ by analyzing differential co-expression genes and provides a novel 3-gene signature for clinical diagnosis or treatment.

MATERIALS AND METHODS

Datasets From TCGA and GEO Databases

COAD-READ gene expression profiles were downloaded from TCGA (https://portal.gdc.cancer.gov/) and GEO (https://www.ncbi.nlm.nih.gov/gds) databases. In TCGA database, all COAD-READ data and clinical information can be downloaded through R package TCGA biolinks (Colaprico et al., 2016). A total of 444 COAD-READ samples were collected, including 404 colorectal adenocarcinoma and 40 normal tissues, consisting of the raw read count data of 19,600 genes. According to the suggestion of the edgeR package manual, lowly expressed genes are usually considered noise in differential expression analysis (Robinson et al., 2010). Therefore, only genes with ≥1 CPM (count per million) were retained for downstream analysis. Then, the expression level of 14,091 genes was calculated using the rpkm function in the edgeR package. In addition, the standardized expression profile of GSE25070, another gene expression profile of COAD-READ in the GEO, was obtained by R package GEO query (Davis and Meltzer, 2007). GSE25070 was composed of the expression profiles of 26 tumor samples from patients with COAD-READ and 26 pairs of normal tissues. The GPL6883 platform Illumina HumanRef-8 v3.0 expression BeadChip was used to study GSE25070. The probes were converted into gene symbols according to the annotation file provided by the manufacturer, and the duplicate probes of the same gene were removed by determining the expression median of all corresponding probes. In the end, 18,599 genes were chosen to be further analyzed.

Using WGCNA to Identify Key Co-Expression Modules

The co-expression network facilitates gene screening technology, which can be used to identify potential biomarkers and treatment targets. Our study built the gene expression profiles of TCGA COAD-READ and GSE25070 into a weighted gene co-expression network by WGCNA package (Langfelder and Horvath, 2008). WGCNA was used to analyze the highly correlated gene modules among samples, and the gene modules related to the external traits of samples were discussed. The pickSoftThreshold function is used to help with selection β to ensure a scale-free network (Zhang and Horvath, 2005). Next, the topological overlap matrix (TOM) and the corresponding dissimilarity matrix (1-TOM) are calculated by using the obtained adjacency matrix. In order to further identify the functional modules in the co-expression network, the module feature association and clinical trait information between the modules were calculated according to the previous studies. We calculated the correlation between modules and clinical data to determine the crucial clinical module. Therefore, the module with a high correlation is considered the candidate associated with clinical traits and chosen to analyze later.

Differential Gene Expression Analysis and Interaction With the Important Modules

R package limma was used to identify differentially expressed genes (DEGs) in TCGA COAD-READ dataset and the microarray data of GSE25070. Limma is an R/Bioconductor software package, which provides an integrated solution for analyzing the data of gene expression (Ritchie et al., 2015). Astringent filter of $|\log FC| \geq 1.0$ and $p < 0.05$ was applied to identify reliable DEGs. Then, R package "ggplot2" (http://ggplot2.org) was further used to draw the volcanic map of all DEGs between COAD-READ and the control group, and R package "pheatmap" (https://CRAN.R-project.org/package=pheatmap) was utilized to draw the clustering heatmap of DEGs. Then, the overlapping genes between co-expression genes and DEGs extracted from the co-expression network were aggregated to explore candidate marker genes and plotted using the R package Venn diagram (Chen and Boutros, 2011).

Functional Annotation Analysis of Significant Genes

R package clusterProfiler was used to perform KEGG and Gene Ontology (GO) enrichment analysis to explore the possible

biological function of the selected genes, and $p < 0.05$ was considered statistically significant (Yu et al., 2012). All the molecular functions (MF), cellular components (CC), and biological processes (BP) were analyzed (Gene Ontology Consortium, 2006).

Construction of a PPI Network and Screening of Hub Genes

Our study used the online tool STRING (search tool for interacting genes, version 11.0), which is designed to predict protein–protein interaction (PPI), to construct a PPI network of candidate genes (Szklarczyk et al., 2019). The pairing with the PPI score ≥0.4 was reserved, and the PPI network was constructed using Cytoscape software 3.40 (www.Cytoscape.org). Using a plug-in cytoHubba of Cytoscape, the top 10 hub genes are predicted based on the maximum clique centrality (MCC) algorithm in the co-expression network (Chin et al., 2014).

Analysis of the Potential Prognostic Values of the Screened hub Genes

All hub genes were divided into two groups according to their expression levels. Based on Kaplan–Meier analysis, the OS curve was drawn by the survival package. (http://cran.rproject.org/package=survival). In addition, the online tool GEPIA2 was used to determine the association between disease-free survival (DFS) and hub gene expression in COAD-READ patients (Tang et al., 2019). $p < 0.05$ was considered statistically significant. Subsequently, we analyzed and compared the expression of the survival-related genes in GSE110224 and drew a box plot graph. Moreover, we performed a correlation analysis between the survival-related genes and clinical factors.

Analysis of Survival-Related Hub Gene Protein Expression in the HPA Database

The Human Protein Atlas (HPA) database aims to map all human proteins in cells, tissues, and organs using various omics techniques (HPA; http://www.proteinatlas.org) (Zhang and Horvath, 2005; Uhlén et al., 2015; Thul et al., 2017). The HPA online tool has helped thousands of biomedical and disease researchers. Using the HPA database, protein levels of survival-related genes were detected by immunohistochemistry (IHC), and IHC images were obtained from the HPA database.

Gene Signature Identification and Risk Score Calculation

In order to identify multiple gene features with good prognostic performance, we randomly divided the samples into the training set and the test set. Univariate Cox regression analysis was used to screen differentially expressed genes related to patient OS in the training set for survival-related genes screened in the previous step. Genes with p value < 0.05 are used as candidate variables, and the genes with the lowest Akaike information criterion (AIC) value are retained in the final signature. The risk coefficients of these genes were calculated using a multivariable Cox proportional risk model based on "survival" (https://CRAN.R-project.org/package = survival) and "survminer" (https://github.com/kassambara/survminer) packages. The patients were divided into high-risk and low-risk groups according to the median risk score. The Kaplan–Meier method was used to analyze the OS of the two groups, and we verified the model's predicted value by drawing the receiver operating characteristic curves (ROCs) with 5 years of the training set, the test set, and the entire set. The ROC curve analysis was obtained by using the nearest neighbor estimation (NNE) method in "survivalROC" package (https://CRAN.R-project.org/package=survivalROC).

RESULTS

Weighted Gene Co-Expression Network Analysis of the Collected CRC Datasets

To explore the gene expression landscape in CRC patients, TCGA COAD-READ dataset and the GSE25070 dataset were downloaded and re-analyzed (**Figure 1**, Methods). To find the functional clusters of COAD-READ patients, WGCNA package in R was used to construct the gene co-expression network from TCGA COAD-READ (**Figure 2A**) and the GSE25070 dataset (**Figure 2E**). The research shows that the co-expression network conforms to the scale-free network: the log(k) value of node connectivity K is negatively correlated with the log(P (k)) value of node probability, and the correlation coefficient is greater than 0.85 for TCGA COAD-READ dataset and 0.8 for the GSE25070 dataset. In order to ensure that the network is scale-free, we chose $\beta = 13$ (sft$powerEstimate = 13) for TCGA COAD-READ dataset and $\beta = 10$ for the GSE25070 dataset. According to the adjacency matrix, the topological overlap matrix is generated based on the TOM similarity algorithm. Then the genes are hierarchically clustered based on this algorithm, and the minimum number of genes was set in a single gene network module to 50. After the gene modules are determined by the dynamic cutting method, the eigengenes of each module are calculated, and each module is analyzed by cluster analysis. Then, we cluster the modules and merge the highly correlated modules into a new module using the mergeCloseModules function with cutHeight set to 0.25. After merging similar modules, we can identify nine modules in TCGA COAD-READ and eight modules in GSE25070 (exclude gray modules that are not assigned to any cluster), and each module uses different colors to distinguish among them. In addition, we explored the module-trait relationship and plotted the heatmap. In TCGA dataset, the yellow (389 genes, p value = 2E−139), black (418 genes, p-value = 7E−19), and green (418 genes, p value = 4e−41) modules are relevant modules with normal traits, while the brown (611 genes, p value = 1E−20), tan (389 genes, p value = 2E−139), and gray modules (389 genes, p value = 2E−139) are relevant modules with colorectal adenocarcinoma traits. Furthermore, in the GEO dataset, the turquoise (813 genes, p value = 3E−12), pink (353 genes, p value = 0.002), and blue (689 genes, p value = 0.01) modules are relevant modules with normal traits, while the brown (695 genes, p value = 8E−15), black (212 genes, p value = 4E−7), and green modules (250 genes, p value = 5E−4) are relevant modules with colorectal adenocarcinoma

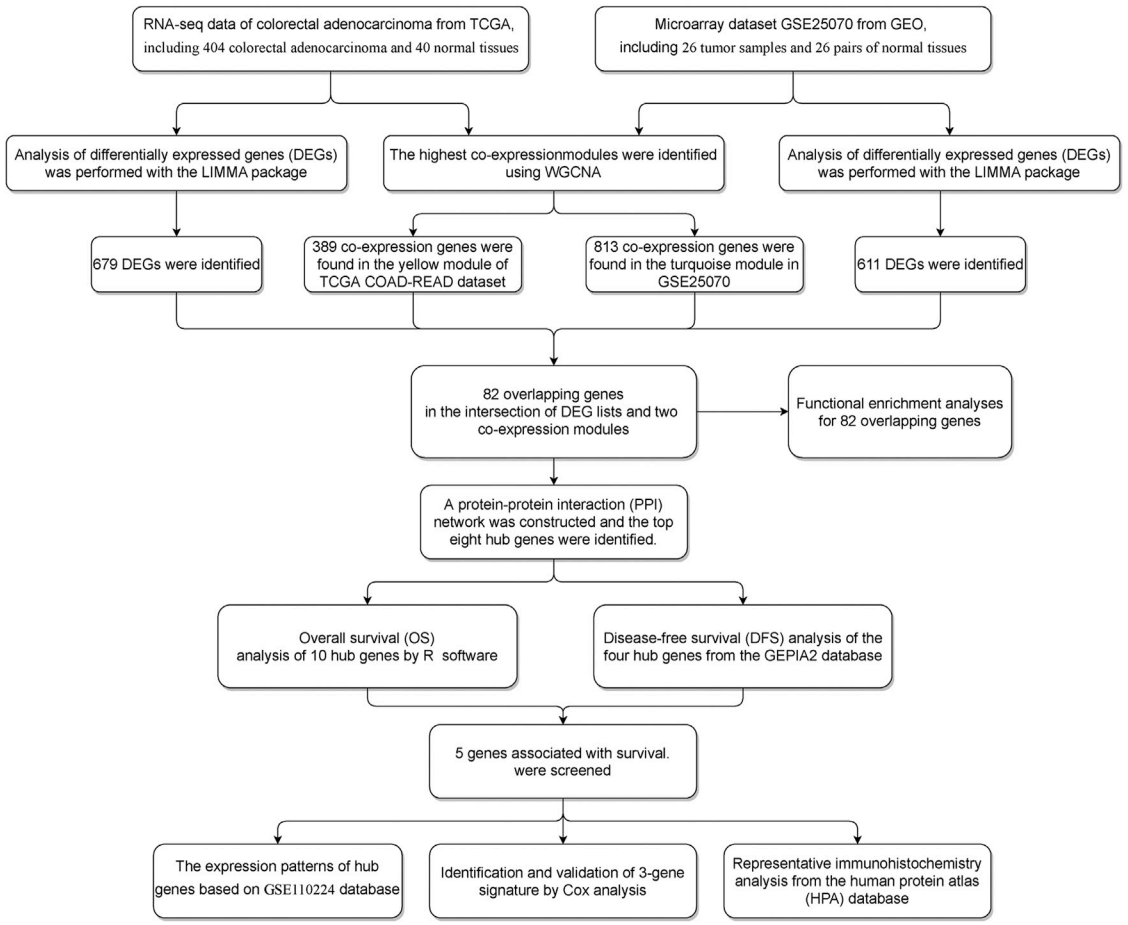


FIGURE 1 | Workflow of the study design.

traits. In particular, the results (**Figures 2B,F**) showed that the yellow module in TCGA COAD-READ and the turquoise module in GSE25070 had the highest negative correlation with colorectal adenocarcinoma (yellow module: r = - 0.87, $p = 2 \times 10^{-139}$ and turquoise module: r = - 0.79, $p = 3 \times 10^{-12}$).

Identification of Genes Between the DEG Lists and Co-Expression Modules

According to the cutoff criteria of $|\log FC| \geq 2.5$ and FDR <0.01 for TCGA dataset and $|\log FC| \geq 1.0$ and FDR <0.05 for the GSE25070 dataset, 679 differentially expressed genes in TCGA dataset (**Figure 2C**) and 611 in the GSE25070 dataset (**Figure 2G**) were identified between COAD-READ and normal tissues. As showcased in **Figure 3A**, 389 and 813 co-expression genes were found in the yellow module of TCGA COAD-READ dataset and the turquoise module in GSE25070, respectively, 82 overlapping genes (**Supplementary Table S1**) were extracted, of which 80 were upregulated, and two were downregulated. Then, the expression pattern of these differentially expressed genes in the COAD-READ and GSE25070 datasets are plotted in **Figures 2D,H**, respectively.

Functional Enrichment Analyses for the 82 Extracted Genes

To further explore the biological function of the 82 differentially expressed genes, the clusterProfiler package (Yu et al., 2012) in R was used to perform GO and KEGG enrichment analyses (Ashburner et al., 2000; Gene Ontology Resource, 2021). The results of the GO enrichment analysis are provided in **Figure 3B**. In the biological processes (BP) category, many terms are related to the metabolism and transport of substances, such as bicarbonate transport and cellular glucuronidation. Moreover, related studies suggest that bicarbonate transporter may change the proliferation rate of cancer cells by regulating the pH value of cancer cell cytoplasm and extracellular space (Gorbatenko et al., 2014). Furthermore, glucuronidation is an essential metabolic pathway of many small endogenous and exogenous lipophilic compounds, including bilirubin, steroid hormones, bile acids, and carcinogens (Hu et al., 2014). Cellular components (CC) showed that these genes were mainly distributed in the apical part of the cell and brush border. The primary enrichment of molecular function (MF) is carbonate dehydratase activity and oxidoreductase activity. In addition, 14 distinct KEGG signaling

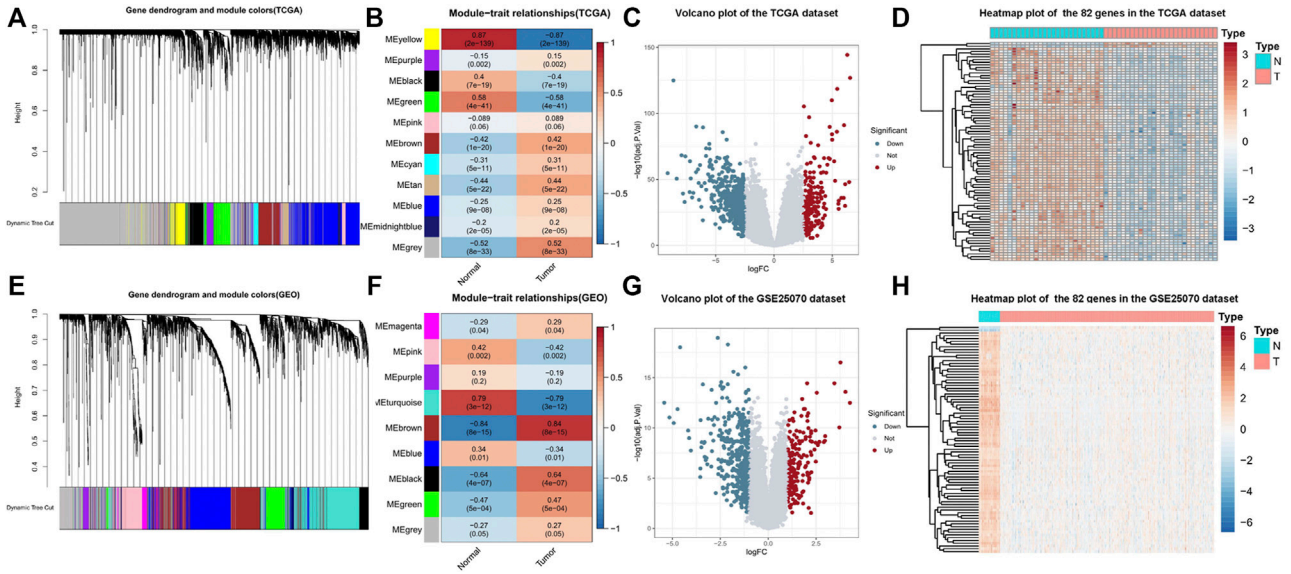


FIGURE 2 | (A) In TCGA COAD-READ dataset, the dendrograms of all differentially expressed genes were clustered based on the measurement of dissimilarity (1-TOM). Each module was assigned a color. **(B)** In TCGA COAD-READ dataset, module-trait relationships. Each row corresponds to a module feature gene (ME), and the column corresponds to a feature. Each cell contains the corresponding correlation and the *p* value. **(C)** Volcano plot of DEGs from TCGA. **(D)** Heatmap of 82 genes in TCGA dataset. **(E)** In the GSE25070 dataset, the dendrograms of all differentially expressed genes were clustered based on the measurement of dissimilarity (1-TOM). Each module was assigned a color. **(F)** In the GSE25070 dataset, module-trait relationships. Each row corresponds to a module feature gene (ME), and the column corresponds to a feature. Each cell contains the corresponding correlation and the *p* value **(G)** Volcano plot of DEGs from the GSE25070 dataset. **(H)** Heatmap of the 82 genes in the GSE25070 dataset.

pathways related to colorectal adenocarcinoma were identified ($p < 0.001$, **Figure 3C**), such as the bile secretion signaling pathway, nitrogen metabolism signaling pathway, steroid hormone biosynthesis signaling pathway, and pancreatic secretion signaling pathway. The bile secretion signaling pathway was the most significantly enriched functional category and had the highest number of enriched genes. Studies have indicated that bile acid-microbiota crosstalk exhibit a vital role in developing colorectal cancer (Louis et al., 2014; Jia et al., 2018). Interestingly, pancreatic secretion, which is also one of the secretions of the digestive glands, appears to play an essential role in the development of colorectal cancer, and some studies suggest hyperinsulinemia and type 2 diabetes mellitus (T2D) are the conditions with an increased risk of CRC (Suh and Yuspa, 2005; Winpenny et al., 2009). Most of the previously mentioned pathways may play a vital role in tumorigenesis.

Construction of a PPI Network and Identification of Hub Genes

The PPI network of overlapping genes was established based on the STRING database, with 82 nodes and 553 edges (**Figure 3D**). The hub gene selected from the PPI network using the cytoHubba plug-in and the MCC algorithm is shown in **Figure 3E**. The top 10 hub genes were screened from the PPI network according to the MCC algorithm, including solute carrier family 26 member 3 (SLC26A3), guanylate cyclase activator 2A (GUCA2A), chloride channel accessory 4 (CLCA4), membrane spanning 4-domains A12 (MS4A12), guanylate cyclase activator 2B (GUCA2B), chloride channel accessory 1 (CLCA1), AQP8 (Aquaporin 8), sucrase-isomaltase (SI), keratin 20 (KRT20), and solute carrier family 4 member 4 (SLC4A4).

Analysis of the Prognostic Values, Expression Patterns, and Protein Expression of Hub Genes

To study the possible clinical relevance of the identified hub gene, survival package and the GEPIA2 database were used to analyze the overall survival (OS) (**Figure 4A**) and disease-free survival (DFS) (**Figure 4B**). Among the ten hub genes, five genes were significantly associated with poor OS of the colorectal adenocarcinoma patients ($p < 0.05$); these are SLC26A3, GUCA2A, CLCA4, CLCA1, and AQP8. In addition, DFS analysis showed that CLCA1 was significantly correlated with DFS. Moreover, the previously mentioned five survival-related genes were further verified in the GSE110224 dataset. As shown in **Figure 4C**, all of the five hub genes were downregulated in colorectal adenocarcinoma tissues compared with adjacent normal tissues. As shown in **Figure 4E**, there was a significant difference in CLCA1 expression between the TNM stage and the N stage. CLCA1 expression in patients with TNM III–IV stages and N 1–2 stages was significantly decreased compared with patients with TNM I–II stages and N 3–4 stages, and SLC26A3 expression in patients with TNM I–II stages were significantly decreased compared with patients with TNM I–II stages. According to the HPA database (**Figure 4D**), the protein levels of SLC26A3, GUCA2A, CLCA4, CLCA1, and AQP8 in tumor

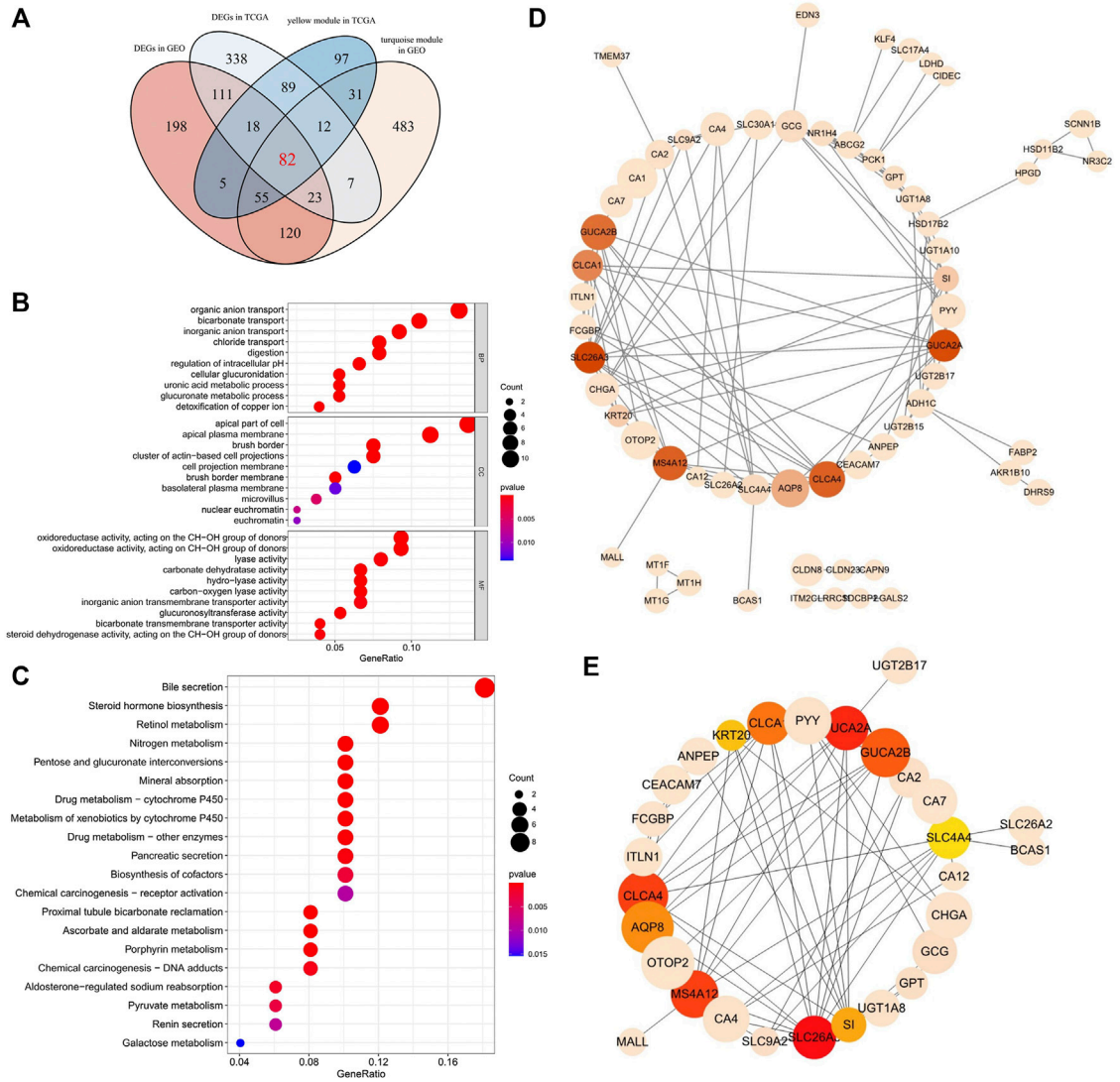


FIGURE 3 | (A) Venn diagram of DEGs and the gene list from the co-expression module. A total of 82 overlapping genes in the intersection of DEGs and two gene lists from the co-expression module. **(B)** Gene Ontology (GO) enrichment analysis for the 82 overlapping genes. The color of each bubble represents the *p* value, and the bubble size represents the number of genes. **(C)** KEGG pathways enrichment analysis for the 82 overlapping genes. The color of each bubble represents the *p* value, and the bubble size represents the number of genes. **(D)** PPI network of the 82 overlapping target genes. There were 82 nodes, where nodes represented genes, and edges were the interactions between two genes. Red is the higher score calculated by the MCC method, followed by yellow, and the size of nodes corresponds to absolute logFC values. **(E)** A total of 10 hub genes were identified from 82 genes *via* MCC algorithm analysis. Network nodes represent proteins; edges represent protein–protein associations. Red is the higher score calculated by the MCC method, followed by yellow, and the size of nodes corresponds to absolute logFC values.

tissues were all significantly lower than that in normal tissues. The aforementioned observations confirmed that the downregulation of SLC26A3, GUCA2A, CLCA4, CLCA1, and AQP8 was associated with poor prognosis and reduced overall survival in colorectal cancer adenocarcinoma patients. Moreover, top-ranking 25% of genes with the highest KME within WGCNA co-expression modules were considered significant genes, and the complete lists of gene sets and kME values are included in **Supplementary Table S2** and **Supplementary Table S3**. We found AQP8, CLCA4, and GUCA2A were significant genes of the yellow module in TCGA dataset, and all five genes were significant genes of the turquoise module in the GSE25070 dataset.

The details of the five hub genes are as follows: chloride channel attachment 1 (CLCA1) and chloride channel attachment 4 (CLCA4) are CLCA proteins. Studies have shown that CLCA protein members affect various biological processes such as cell differentiation, adhesion, apoptosis, and airway inflammation (Elble and Pauli, 2001; Patel et al., 2009; Piirsoo et al., 2009). Both CLCA1 and CLCA4 are expressed in the intestine (Gandhi et al., 1998; Bustin et al., 2001; Jia et al., 2018; Liu et al., 2018), may act as tumor suppressors, and are negatively correlated with tumor formation (Bustin et al., 2001). Recent studies have shown that the expression levels of CLCA proteins, including CLCA1 and CLCA4, are abnormal in many cancer types so it may be a potential cancer

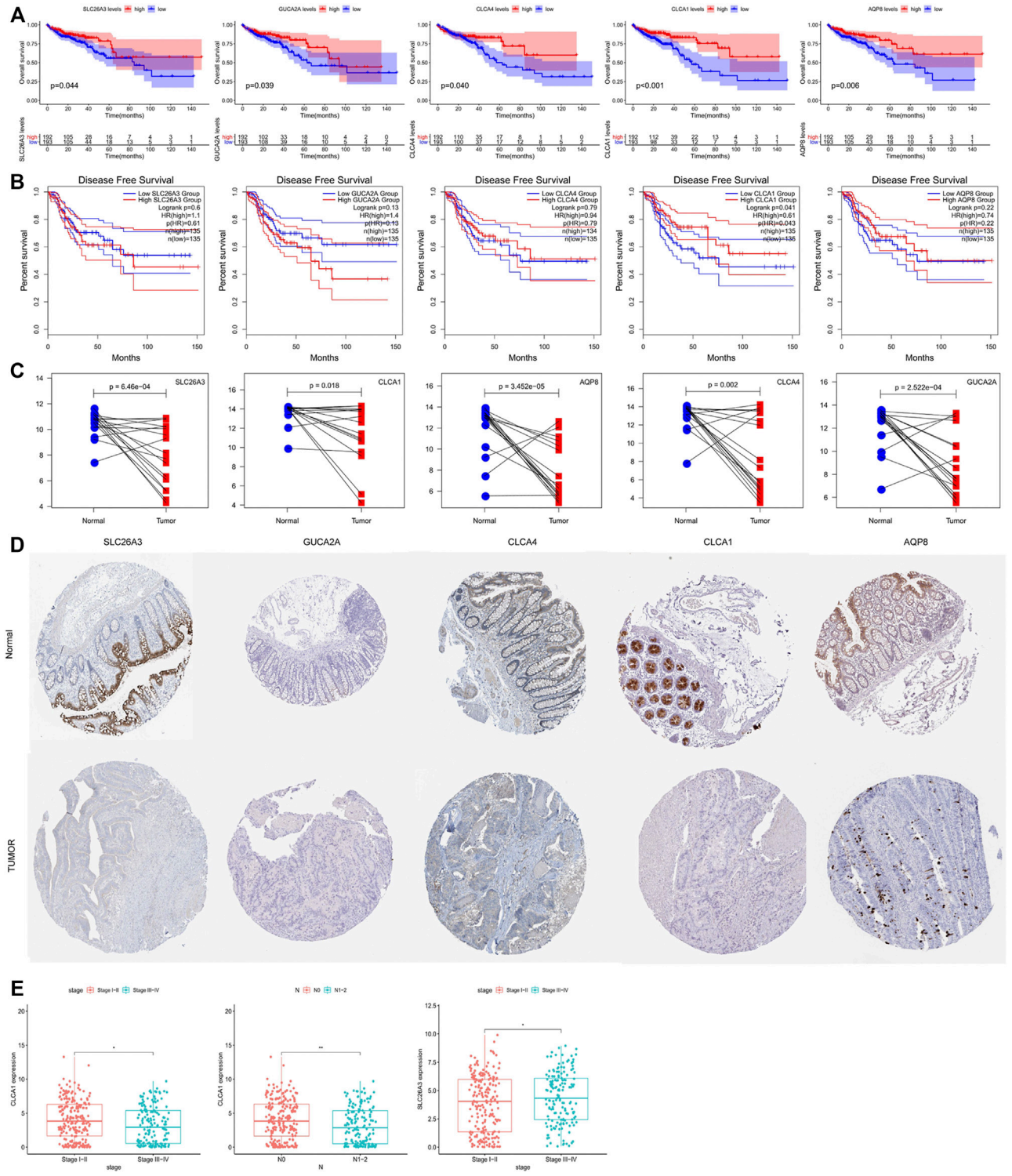


FIGURE 4 | (A) Overall survival (OS) analysis of 10 hub genes in colorectal adenocarcinoma patients by R survival package. Patients were divided into high-level group (red) and low-level group (blue) according to the median. *p* values <0.05 were considered to be statistically significant. **(B)** Disease-free survival (DFS) analysis of the five hub genes (SLC26A3, GUCA2A, CLCA4, CLCA1, and AQP8) in colorectal adenocarcinoma patients from the GEPIA2 database. Patients were divided into high-level group (red) and low-level group (blue) according to the median. *p* values <0.05 were considered to be statistically significant. **(C)** Validation of the expression levels of 10 hub genes in colorectal adenocarcinoma patients and normal tissues in the GSE110224 dataset. **(D)** Expression profiles of the four genes in the colorectal adenocarcinoma patients and normal tissue. Images were taken from the Human Protein Atlas (http://www.proteinatlas.org). **(E)** Correlation analysis between the survival-related genes and clinical factors. *p* values ≤0.05 were considered to be statistically significant.

predictor for patients (Yang et al., 2013; Yu et al., 2013; Hou et al., 2017; Liu et al., 2018). In addition, according to the literature, the mRNA levels of CLCA1 in colorectal cancer (Yang et al., 2013), ovarian cancer (Musrap et al., 2015), and pancreatic ductal adenocarcinoma (Hu et al., 2018) were different from those in normal tissues, and the loss of CLCA4 expression was observed in colorectal cancer (Chen et al., 2019), hepatocellular carcinoma (Liu et al., 2018), breast cancer (Yu et al., 2013), and bladder cancer (Hou et al., 2017). Our data showed that CLCA1 and CLCA4 were significantly downregulated in colorectal cancer compared with normal tissues. Previous studies have shown that the increase of CLCA1 and CLCA4 levels in tumor tissue is related to the excellent prognosis of colorectal cancer patients, consistent with our survival analysis (Chen et al., 2018; Pan et al., 2019a; Pan et al., 2019b; Chen et al., 2019; Wei et al., 2020a; Wei et al., 2020b).

GUCA2A is an endogenous ligand for the guanylate cyclase 2C (GUCY2C) receptor and peptide hormone and is expressed in gut epithelial cells (Blomain et al., 2013; Lima and Fonteles, 2014; Ikpa et al., 2016). Locally, it acts as autocrine and paracrine hormones to regulate GUCY2C signal transduction and humoral electrolyte homeostasis (Steinbrecher et al., 2002; Brierley, 2012). GUCY2C has a protective effect on colorectal tumors (Li et al., 2007). The downregulation of GUCA2A leads to the loss of the GUCY2C signal cascade and promotes tumorigenesis (Li et al., 2007). Previous studies have shown that the effect of the expression of GUCY2C is significant to the occurrence and development of tumors (Pattison et al., 2016; Bashir et al., 2019). This corresponds well with our findings where we showed that the down expression of GUCA2A may also contribute to colorectal adenocarcinoma carcinogenesis.

SLC26A3 is a member of the Slc26 anion transporter and channel family (Mount and Romero, 2004). SLC26A3 plays a vital role in colonic Cl−absorption (Schweinfest et al., 2006; Zhang et al., 2007). Immunohistochemical results showed that SLC26A3 was located in the apical membrane of enterocytes of the surface and crypt in the colon (Worrell et al., 2005; Schweinfest et al., 2006; Soleimani, 2006) and the apical membrane of pancreatic duct (Ishiguro et al., 2007). Consistent with the previous studies (Schweinfest et al., 1993). We found that its expression is downregulated in colorectal adenocarcinoma.

AQP8, an aquaporins (AQPs) member, is a water channel protein. Aquaporins are a family of small integral membrane proteins related to the major intrinsic protein (MIP or AQP0).

Studies show that the overexpression of AQP8 restrained CRC cell proliferation, migration, and invasion capacities *in vitro* (Wu et al., 2018). Inconsistency, AQP8 was found to be downregulated in our study.

Identification and Validation of the 3-Gene Signature

We randomly divided 385 samples and the corresponding clinical data into a training set (n = 193) and a test set (n = 192). We found that all five hub genes were significantly associated with prognosis by univariate survival analysis in the training set. Subsequently, multivariate Cox regression analysis was used to establish a 3-gene signature. The three genes contained in this signature are CLCA1, CLCA4, and GUCA2A (**Supplementary Table S4**). The risk coefficients suggested that all four genes are risk factors for colorectal adenocarcinoma (coef >0). The risk score for each patient was calculated according to the following formula: risk score = (-0.053) × expression value of CLCA1 + (-0.043) × expression value of GUCA2A+ (-0.189) × expression value of CLCA4. The higher the score, the worse the patients' prognosis were. Patients were divided into the high-risk and low-risk groups based on the median risk score. Moreover, the predicted results showed a significant difference in OS between high- and low-risk groups for all three sets ($p < 0.05$). The AUC of the ROC curve was 0.778 in the training set, 0.695 in the test set, and 0.737 in the entire set. Finally, the distribution of the risk score in each patient, survival status, and the expression of three genes are shown in **Figure 5D–F**.

Comparison With Other Published Gene Signatures

We compared our 3-gene signature with other published gene signatures (Zuo et al., 2019; Gong et al., 2020; Li et al., 2020; Hong et al., 2021; Qian et al., 2021; Shao et al., 2021). In **Supplementary Table S5**, the AUC for 5 years OS in our 3-gene signature was 0.778 in the training set, 0.695 in the test set, and 0.737 in the entire set. The C-index of the 18-gene of Qian, 8-gene of Gong, 10-gene of Shao, 16-gene of Meng, 5-gene of Hong, and 6-gene of Zuo are 0.734, 0.713, 0.780, 0.869, 0.675, and 0.683, respectively. Compared with other prognostic models, our model has a substantial predictive value. Furthermore, our model is composed of only three genes, so it is concise and easy to use. Moreover, the three genes were obtained by the multistep screening process, including WGCNA and PPI analysis, so the 3-gene signature was not only associated with the prognosis of CRC but also served an essential role in the CRC initiation and progression.

DISCUSSION

Colorectal adenocarcinoma is one of the most common cancers globally, and its incidence rate is high. Therefore, the research on the tumor metastasis mechanism of colorectal adenocarcinoma is of great significance to explore new targets and improve the therapeutic effect and prognosis of patients with colorectal adenocarcinoma. Compared to using a single analysis or database, this study performed a more detailed and more effective analysis, that is, screening hub genes based on WGCNA and differential gene expression analysis in TCGA and GEO database. Differential gene expression analysis 82 critical genes with the same expression trends were identified in TCGA and GSE25070 databases by bioinformatics analysis. Functional annotation of these genes demonstrated that they were involved in bicarbonate transport and organic anion transport, which can alter the intestinal microenvironment to affect the initiation and progression of tumors.

Furthermore, based on the cytoHubba plug-in in Cytoscape, 10 genes (SLC26A3, GUCA2A, CLCA4, MS4A12, GUCA2B, CLCA1, AQP8, SI, KRT20, and SLC4A4) were screened. The down expression of five genes (SLC26A3, GUCA2A, CLCA4, CLCA1, and AQP8) was significantly associated with poor overall survival in colorectal cancer

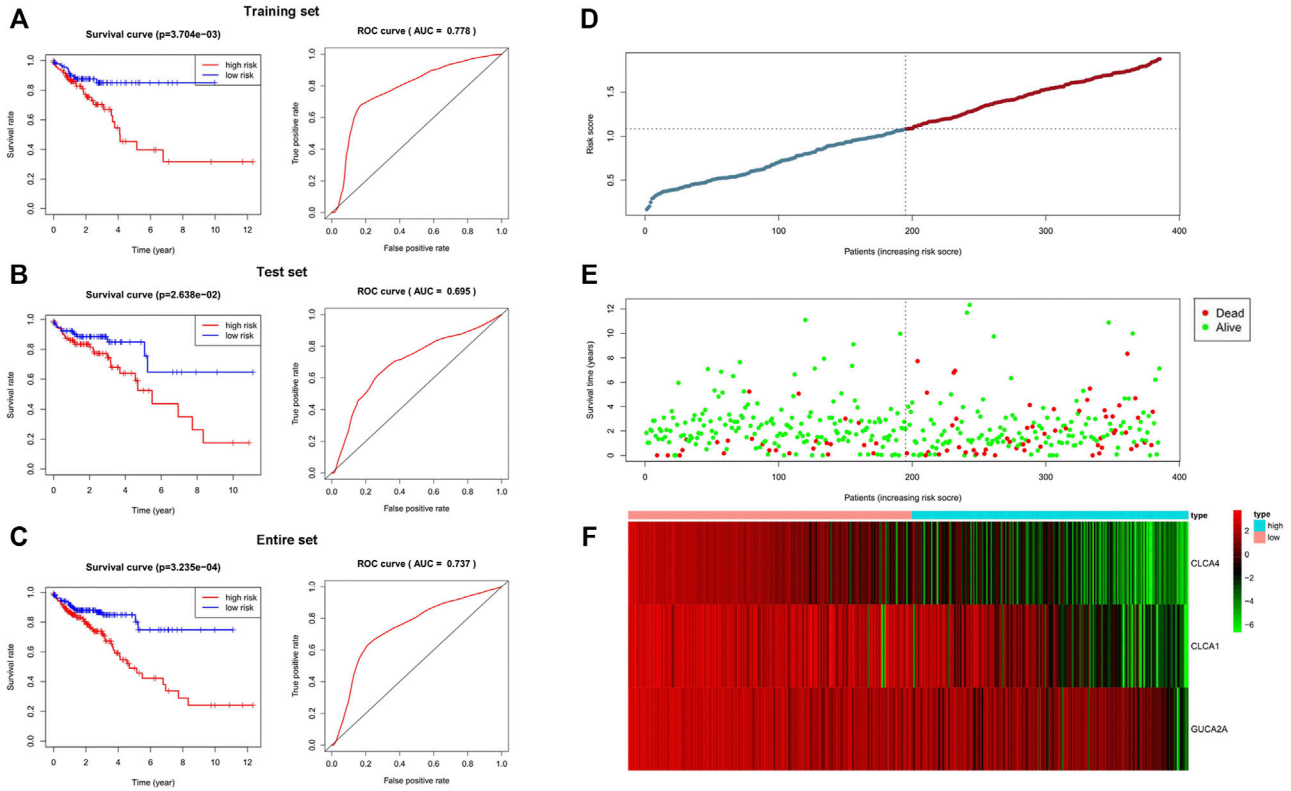


FIGURE 5 | Validation of the 3-gene signature. In all three sets, there were significant differences in OS between the high-risk group and the low-risk group. The area under the ROC was 0.778 in the training set, 0.695 in the test set, and 0.737 in the entire set. The results showed that the 3-gene signature performed well in predicting the prognosis of colorectal adenocarcinoma. **(A)** Training set. **(B)** Test set. **(C)** Entire set. **(D–F)** Distributions of the 3-gene signature, survival status, and expression profiles of the three genes of entire sets.

adenocarcinoma patients, and patients with lower expression levels of CLCA1 had poorer disease-free survival rates. Subsequently, clinical correlation analysis and the immunohistochemical analysis of the five hub genes were analyzed, revealing these genes' potential as novel prognostic biomarkers.

In addition, none of the single biomarkers can be used to detect cancer and achieve the required specificity and sensitivity in the cancer research along (Muinao et al., 2019), and multiple biomarkers have more advantages than a single biomarker. Therefore, we constructed a 3-gene risk model to predict the prognosis of colorectal adenocarcinoma, and subsequent studies suggested the 3-gene risk model has a good prognostic value. Furthermore, the prognostic model showed an excellent predictive efficiency compared with other published models.

Like all studies, our study has limitations in classifying different subtypes of tumors. Our study is conducted without the colorectal adenocarcinoma subtype. In addition, the specific pathogenesis and molecular targets need to be further verified by a series of experiments.

In conclusion, by integrating WGCNA and differential gene expression analysis, our study screened five important survival-related genes (SLC26A3, GUCA2A, CLCA4, CLCA1, and AQP8) and a 3-gene risk model with the potential to predict the prognosis of colorectal adenocarcinoma.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found at: the TCGA database (https://portal.gdc.cancer.gov/) and GEO database (https://www.ncbi.nlm.nih.gov/geo/).

AUTHOR CONTRIBUTIONS

LM conceived the project. YL and LC analyzed the data. XM, SY, and LM wrote the manuscript. The authors read and approved the final manuscript.

FUNDING

This work was supported by grants from the Natural Science Foundation of Jilin Province, China (20210101248JC).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2022.897568/full#supplementary-material

REFERENCES

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene Ontology: Tool for the Unification of Biology. *Nat. Genet.* 25 (1), 25–29. doi:10.1038/75556

Bashir, B., Merlino, D. J., Jeffrey, A. R., Gnass, E., Palazzo, J. P., Feng, Y., et al. (2019). Silencing the GUCA2A-Gucy2c Tumor Suppressor axis in CIN, Serrated, and MSI Colorectal Neoplasia. *Hum. Pathol.* 87. doi:10.1016/j.humpath.2018.11.032

Blomain, E. S., Lin, J. E., Kraft, C. L., Trela, U. T., Rock, J. M., Aing, A. S., et al. (2013). Translating Colorectal Cancer Prevention through the Guanylyl Cyclase C Signaling axis. *Expert Rev. Clin. Pharmacol.* 6 (5), 557–564. doi:10.1586/17512433.2013.827406

Brierley, S. M. (2012). Guanylate Cyclase-C Receptor Activation: Unexpected Biology. *Curr. Opin. Pharmacol.* 12 (6), 632–640. doi:10.1016/j.coph.2012.10.005

Bustin, S. A., Li, S.-R., and Dorudi, S. (2001). Expression of the Ca2+-Activated Chloride Channel Genes CLCA1 and CLCA2 Is Downregulated in Human Colorectal Cancer. *DNA Cell Biol.* 20 (6), 331–338. doi:10.1089/10445490152122442

Can, T. (2014). Introduction to Bioinformatics. *Methods Mol. Biol.* 1107, 51–71. doi:10.1007/978-1-62703-748-8_4

Chen, H., and Boutros, P. C. (2011). VennDiagram: a Package for the Generation of Highly-Customizable Venn and Euler Diagrams in R. *BMC Bioinforma.* 12, 35. doi:10.1186/1471-2105-12-35

Chen, H., Liu, Y., Jiang, C.-J., Chen, Y.-M., Li, H., and Liu, Q.-A. (2019). Calcium-Activated Chloride Channel A4 (CLCA4) Plays Inhibitory Roles in Invasion and Migration through Suppressing Epithelial-Mesenchymal Transition via PI3K/AKT Signaling in Colorectal Cancer. *Med. Sci. Monit.* 25, 4176–4185. doi:10.12659/MSM.914195

Chen, T.-J., He, H.-L., Shiue, Y.-L., Yang, C.-C., Lin, L.-C., Tian, Y.-F., et al. (2018). High Chloride Channel Accessory 1 Expression Predicts Poor Prognoses in Patients with Rectal Cancer Receiving Chemoradiotherapy. *Int. J. Med. Sci.* 15 (11), 1171–1178. doi:10.7150/ijms.26685

Chin, C-H., Chen, S-H., Wu, H-H., Ho, C-W., Ko, M-T., and Lin, C-Y. (2014). cytoHubba: Identifying Hub Objects and Sub-networks from Complex Interactome. *BMC Syst. Biol.* 8 (Suppl. 4Suppl 4). doi:10.1186/1752-0509-8-S4-S11

Colaprico, A., Silva, T. C., Olsen, C., Garofano, L., Cava, C., Garolini, D., et al. (2016). TCGAbiolinks: an R/Bioconductor Package for Integrative Analysis of TCGA Data. *Nucleic Acids Res.* 44 (8), e71. doi:10.1093/nar/gkv1507

Colorectal Cancer Statistics | How Common Is Colorectal Cancer? (2021). Colorectal Cancer Statistics | How Common Is Colorectal Cancer? Available at: https://www.cancer.org/cancer/colon-rectal-cancer/about/key-statistics.html (Accessed March 26.2021

Davis, S., and Meltzer, P. S. (2007). GEOquery: a Bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 23 (14), 1846–1847. doi:10.1093/bioinformatics/btm254

Elble, R. C., and Pauli, B. U. (2001). Tumor Suppression by a Proapoptotic Calcium-Activated Chloride Channel in Mammary Epithelium. *J. Biol. Chem.* 276 (44), 40510–40517. doi:10.1074/jbc.M104821200

Gandhi, R., Elble, R. C., Gruber, A. D., Schreur, K. D., Ji, H.-L., Fuller, C. M., et al. (1998). Molecular and Functional Characterization of a Calcium-Sensitive Chloride Channel from Mouse Lung. *J. Biol. Chem.* 273 (48), 32096–32101. doi:10.1074/jbc.273.48.32096

Gene Ontology Consortium (2006). The Gene Ontology (GO) Project in 2006. *Nucleic Acids Res.* 34 (Database issue), D322–D326. doi:10.1093/nar/gkj021

Gene Ontology Resource (2021). Gene Ontology Resource. Available at: http://geneontology.org/(Accessed May 8, 2021).

Gong, B., Kao, Y., Zhang, C., Sun, F., Gong, Z., and Chen, J. (2020). Identification of Hub Genes Related to Carcinogenesis and Prognosis in Colorectal Cancer Based on Integrated Bioinformatics. *Mediat. Inflamm.* 2020, 1–14. doi:10.1155/2020/5934821

Gorbatenko, A., Olesen, C. W., Boedtkjer, E., and Pedersen, S. F. (2014). Regulation and Roles of Bicarbonate Transporters in Cancer. *Front. Physiol.* 5, 130. doi:10.3389/fphys.2014.00130

Hong, J., Lin, X., Hu, X., Wu, X., and Fang, W. (2021). A Five-Gene Signature for Predicting the Prognosis of Colorectal Cancer. *Cgt* 21 (4), 280–289. doi:10.2174/1566523220666201012151803

Hou, T., Zhou, L., Wang, L., Kazobinka, G., Zhang, X., and Chen, Z. (2017). CLCA4 Inhibits Bladder Cancer Cell Proliferation, Migration, and Invasion by Suppressing the PI3K/AKT Pathway. *Oncotarget* 8 (54), 93001–93013. doi:10.18632/oncotarget.21724

Hu, D., Ansari, D., Zhou, Q., Sasor, A., Hilmersson, K. S., Bauden, M., et al. (2018). Calcium-activated Chloride Channel Regulator 1 as a Prognostic Biomarker in Pancreatic Ductal Adenocarcinoma. *BMC Cancer* 18 (1), 1096. doi:10.1186/s12885-018-5013-2

Hu, D. G., Meech, R., McKinnon, R. A., and Mackenzie, P. I. (2014). Transcriptional Regulation of Human UDP-Glucuronosyltransferase Genes. *Drug Metab. Rev.* 46 (4), 421–458. doi:10.3109/03602532.2014.973037

Ikpa, P. T., Sleddens, H. F. B. M., Steinbrecher, K. A., Peppelenbosch, M. P., de Jonge, H. R., Smits, R., et al. (2016). Guanylin and Uroguanylin Are Produced by Mouse Intestinal Epithelial Cells of Columnar and Secretory Lineage. *Histochem Cell Biol.* 146 (4), 445–455. doi:10.1007/s00418-016-1453-4

Ishiguro, H., Namkung, W., Yamamoto, A., Wang, Z., Worrell, R. T., Xu, J., et al. (2007). Effect of Slc26a6 Deletion on Apical Cl−/HCO3− Exchanger Activity and cAMP-Stimulated Bicarbonate Secretion in Pancreatic Duct. *Am. J. Physiology-Gastrointestinal Liver Physiology* 292 (1), G447–G455. doi:10.1152/ajpgi.00286.2006

Jia, W., Xie, G., and Jia, W. (2018). Bile Acid-Microbiota Crosstalk in Gastrointestinal Inflammation and Carcinogenesis. *Nat. Rev. Gastroenterol. Hepatol.* 15 (2), 111–128. doi:10.1038/nrgastro.2017.119

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, M., Wang, H., Li, W., Peng, Y., Xu, F., Shang, J., et al. (2020). Identification and Validation of an Immune Prognostic Signature in Colorectal Cancer. *Int. Immunopharmacol.* 88, 106868. doi:10.1016/j.intimp.2020.106868

Li, P., Schulz, S., Bombonati, A., Palazzo, J. P., Hyslop, T. M., Xu, Y., et al. (2007). Guanylyl Cyclase C Suppresses Intestinal Tumorigenesis by Restricting Proliferation and Maintaining Genomic Integrity. *Gastroenterology* 133 (2), 599–607. doi:10.1053/j.gastro.2007.05.052

Lima, A. A. M., and Fonteles, M. C. (2014). From *Escherichia coli* Heat-Stable Enterotoxin to Mammalian Endogenous Guanylin Hormones. *Braz J. Med. Biol. Res.* 47 (3), 179–191. doi:10.1590/1414-431X20133063

Liu, Z., Chen, M., Xie, L.-K., Liu, T., Zou, Z.-W., Li, Y., et al. (2018). CLCA4 Inhibits Cell Proliferation and Invasion of Hepatocellular Carcinoma by Suppressing Epithelial-Mesenchymal Transition via PI3K/AKT Signaling. *Aging* 10 (10), 2570–2584. doi:10.18632/aging.101571

Louis, P., Hold, G. L., and Flint, H. J. (2014). The Gut Microbiota, Bacterial Metabolites and Colorectal Cancer. *Nat. Rev. Microbiol.* 12 (10), 661–672. doi:10.1038/nrmicro3344

Mount, D. B., and Romero, M. F. (2004). The SLC26 Gene Family of Multifunctional Anion Exchangers. *Pflugers Archiv Eur. J. Physiology* 447 (5), 710–721. doi:10.1007/s00424-003-1090-3

Muinao, T., Deka Boruah, H. P., and Pal, M. (2019). Multi-biomarker Panel Signature as the Key to Diagnosis of Ovarian Cancer. *Heliyon* 5 (12), e02826. doi:10.1016/j.heliyon.2019.e02826

Musrap, N., Tuccitto, A., Karagiannis, G. S., Saraon, P., Batruch, I., and Diamandis, E. P. (2015). Comparative Proteomics of Ovarian Cancer Aggregate Formation Reveals an Increased Expression of Calcium-Activated Chloride Channel Regulator 1 (CLCA1). *J. Biol. Chem.* 290 (28), 17218–17227. doi:10.1074/jbc.M115.639773

Pan, X., Wang, Q., Chenfei, X., Ling, Y., Pang, S., and Gan, J. (2019). Prognostic Value of Chloride Channel Accessory mRNA Expression in Colon Cancer. *Oncol. Lett.* 18 (3). doi:10.3892/ol.2019.10615

Pan, X., Wang, Q., Xu, C., Yan, L., Pang, S., and Gan, J. (2019). Prognostic Value of Chloride Channel Accessory mRNA Expression in Colon Cancer. *Oncol. Lett.* 18 (3), 2967–2976. doi:10.3892/ol.2019.10615

Patel, A. C., Brett, T. J., and Holtzman, M. J. (2009). The Role of CLCA Proteins in Inflammatory Airway Disease. *Annu. Rev. Physiol.* 71, 425–449. doi:10.1146/annurev.physiol.010908.163253

Pattison, A. M., Merlino, D. J., Blomain, E. S., and Waldman, S. A. (2016). Guanylyl Cyclase C Signaling axis and Colon Cancer Prevention. *Wjg* 22 (36), 8070–8077. doi:10.3748/wjg.v22.i36.8070

Piirsoo, M., Meijer, D., and Timmusk, T. (2009). Expression Analysis of the CLCA Gene Family in Mouse and Human with Emphasis on the Nervous System. *BMC Dev. Biol.* 9, 10. doi:10.1186/1471-213X-9-10

Qian, Y., Wei, J., Lu, W., Sun, F., Hwang, M., Jiang, K., et al. (2021). Prognostic Risk Model of Immune-Related Genes in Colorectal Cancer. *Front. Genet.* 12, 619611. doi:10.3389/fgene.2021.619611

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

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

San Segundo-Val, I., and Sanz-Lozano, C. S. (2016). Introduction to the Gene Expression Analysis. *Methods Mol. Biol.* 1434, 29–43. doi:10.1007/978-1-4939-3652-6_3

Schweinfest, C. W., Henderson, K. W., Suster, S., Kondoh, N., and Papas, T. S. (1993). Identification of a Colon Mucosa Gene that Is Down-Regulated in Colon Adenomas and Adenocarcinomas. *Proc. Natl. Acad. Sci. U.S.A.* 90 (9), 4166–4170. doi:10.1073/pnas.90.9.4166

Schweinfest, C. W., Spyropoulos, D. D., Henderson, K. W., Kim, J.-H., Chapman, J. M., Barone, S., et al. (2006). slc26a3 (Dra)-deficient Mice Display Chloride-Losing Diarrhea, Enhanced Colonic Proliferation, and Distinct Up-Regulation of Ion Transporters in the Colon. *J. Biol. Chem.* 281 (49), 37962–37971. doi:10.1074/jbc.M607527200

Shao, Y., Jia, H., Huang, L., Li, S., Wang, C., Aikemu, B., et al. (2021). An Original Ferroptosis-Related Gene Signature Effectively Predicts the Prognosis and Clinical Status for Colorectal Cancer Patients. *Front. Oncol.* 11, 711776. doi:10.3389/fonc.2021.711776

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

Soleimani, M. (2006). Expression, Regulation and the Role of SLC26 Cl-/HCO3-Exchangers in Kidney and Gastrointestinal Tract. *Novartis Found. Symp.* 273, 91–94.

Steinbrecher, K. A., Wowk, S. A., Rudolph, J. A., Witte, D. P., and Cohen, M. B. (2002). Targeted Inactivation of the Mouse Guanylin Gene Results in Altered Dynamics of Colonic Epithelial Proliferation. *Am. J. Pathology* 161 (6), 2169–2178. doi:10.1016/S0002-9440(10)64494-X

Suh, K., and Yuspa, S. (2005). Intracellular Chloride Channels: Critical Mediators of Cell Viability and Potential Targets for Cancer Therapy. *Cpd* 11 (21), 2753–2764. doi:10.2174/1381612054546806

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING V11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-wide Experimental Datasets. *Nucleic Acids Res.* 47 (D1), D607–D613. doi:10.1093/nar/gky1131

Tang, Z., Kang, B., Li, C., Chen, T., and Zhang, Z. (2019). GEPIA2: an Enhanced Web Server for Large-Scale Expression Profiling and Interactive Analysis. *Nucleic Acids Res.* 47 (W1), W556–W560. doi:10.1093/nar/gkz430

Thul, P. J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., et al. (2017). A Subcellular Map of the Human Proteome. *Science* 356 (6340). doi:10.1126/science.aal3321

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2015). Proteomics. Tissue-Based Map of the Human Proteome. *Science* 347 (6220), 1260419. doi:10.1126/science.1260419

van der Geest, L. G. M., Lam-Boer, J. t., Koopman, M., Verhoef, C., Elferink, M. A. G., and de Wilt, J. H. W. (2015). Nationwide Trends in Incidence, Treatment and Survival of Colorectal Cancer Patients with Synchronous Metastases. *Clin. Exp. Metastasis* 32 (5), 457–465. doi:10.1007/s10585-015-9719-0

Wei, F.-Z., Mei, S.-W., Wang, Z.-J., Chen, J.-N., Shen, H.-Y., Zhao, F.-Q., et al. (2020). Differential Expression Analysis Revealing CLCA1 to Be a Prognostic and Diagnostic Biomarker for Colorectal Cancer. *Front. Oncol.* 10, 573295. doi:10.3389/fonc.2020.573295

Wei, L., Chen, W., Zhao, J., Fang, Y., and Lin, J. (2020). Downregulation of CLCA4 Expression Is Associated with the Development and Progression of Colorectal Cancer. *Oncol. Lett.* 20 (1), 631–638. doi:10.3892/ol.2020.11640

Winpenny, J., Marsey, L., and Sexton, D. (2009). The CLCA Gene Family: Putative Therapeutic Target for Respiratory Diseases. *Iadt* 8 (2), 146–160. doi:10.2174/187152809788462590

Worrell, R. T., Best, A., Crawford, O. R., Xu, J., Soleimani, M., and Matthews, J. B. (2005). Apical Ammonium Inhibition of cAMP-Stimulated Secretion in T84 Cells Is Bicarbonate Dependent. *Am. J. Physiology-Gastrointestinal Liver Physiology* 289 (4), G768–G778. doi:10.1152/ajpgi.00451.2004

Wu, Q., Yang, Z. F., Wang, K. J., Feng, X. Y., Lv, Z. J., Li, Y., et al. (2018). AQP8 Inhibits Colorectal Cancer Growth and Metastasis by Down-Regulating PI3K/AKT Signaling and PCDH7 Expression. *Am. J. Cancer Res.* 8 (2), 266–279.

Yang, B., Cao, L., Liu, B., McCaig, C. D., and Pu, J. (2013). The Transition from Proliferation to Differentiation in Colorectal Cancer Is Regulated by the Calcium Activated Chloride Channel A1. *PLoS One* 8 (4), e60861. doi:10.1371/journal.pone.0060861

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS A J. Integr. Biol.* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Yu, Y., Walia, V., and Elble, R. C. (2013). Loss of CLCA4 Promotes Epithelial-To-Mesenchymal Transition in Breast Cancer Cells. *PLoS One* 8 (12), e83943. doi:10.1371/journal.pone.0083943

Zhang, B., and Horvath, S. (2005). A General Framework for Weighted Gene Co-expression Network Analysis. *Stat. Appl. Genet. Mol. Biol.* 4. doi:10.2202/1544-6115.1128

Zhang, G. H., Zhu, J. X., Xue, H., Fan, J., Chen, X., Tsang, L. L., et al. (2007). Dopamine Stimulates Cl− Absorption Coupled with HCO3− Secretion in Rat Late Distal Colon. *Eur. J. Pharmacol.* 570 (1-3), 188–195. doi:10.1016/j.ejphar.2007.05.038

Zuo, S., Dai, G., and Ren, X. (2019). Identification of a 6-gene Signature Predicting Prognosis for Colorectal Cancer. *Cancer Cell Int.* 19, 6. doi:10.1186/s12935-018-0724-7

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: 08 July 2022
doi: 10.3389/fcell.2022.921320

Establishment of a Necroptosis Related Genes Signature to Predict Prognosis and Therapeutic Response in Colon Cancer

Yuan Wang†, Yongbiao Huang†, Chunya Li, Xi Wang, Mu Yang, Duo Xu, Bo Liu and Xianglin Yuan**

Department of Oncology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Necroptosis, as a form of programmed cell death, is involved in many physiological and pathological processes. However, its role in cancer progression and therapeutic response remains controversial. Colon cancer is one of the leading causes of cancer death and patients' response to immune checkpoint blockade vary to a large degree. In this study, we investigated necroptosis related genes (NRGs) alterations in colon cancer by bioinformatics analysis. Colon cancer patients were classified into two subtypes with distinct clinical and molecular features based on NRGs. After finding differentially expressed genes and lasso regression, a prognostic model based on four necroptosis signature genes was constructed. The necroptosis signature was also a good predictor in the field of chemotherapy and immunotherapy in colon cancer. Altogether, this study illustrates the relationship between necroptosis and colon cancer, and establishes a novel scoring method to predict prognosis and therapeutic response in colon cancer patients.

Keywords: tumor microenvironment, immunotherapy, survival, necroptosis, colon cancer

OPEN ACCESS

Edited by:
Lu Xie,
Shanghai Institute for Biomedical and Pharmaceutical Technologies, China

Reviewed by:
Padhmanand Sudhakar,
KU Leuven, Belgium
Glaucia Maria Machado-Santelli,
Universidade de São Paulo, Brazil

****Correspondence:***
Bo Liu
boliu888@hotmail.com
Xianglin Yuan
yuanxianglin@hust.edu.cn

†These authors have contributed equally to this work

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

Received: *15 April 2022*
Accepted: *21 June 2022*
Published: *08 July 2022*

Citation:
Wang Y, Huang Y, Li C, Wang X, Yang M, Xu D, Liu B and Yuan X (2022) Establishment of a Necroptosis Related Genes Signature to Predict Prognosis and Therapeutic Response in Colon Cancer. Front. Cell Dev. Biol. 10:921320. doi: 10.3389/fcell.2022.921320

INTRODUCTION

Necroptosis was first reported in 2005 as a programmed form of cell death which exhibited features of both necrosis and apoptosis (Degterev et al., 2005). Key necroptosis effector molecules involved in necroptosis were RIPK1, RIPK3, and MLKL. Receptor-interacting serine/threonine protein kinase 1 (RIPK1) was first identified as a regulator of cell death (Hsu et al., 1996). In 2008, it was identified as the target of necrostatin-1 (Nec-1), which suppressed caspase inhibition-mediated cell death (Degterev et al., 2008). RIPK3, another member of the RIPK family, was shown to be crucial for death receptor-triggered necroptosis in 2009 (Cho et al., 2009; He et al., 2009; Zhang et al., 2009). Mixed-lineage kinase domain-like protein (MLKL) was identified to participate in necroptosis after activation of RIPK3 in 2012 (Sun et al., 2012), however, published reports provided conflicting mechanisms on how it led to membrane rupture (Cai et al., 2014; Chen et al., 2014; Dondelinger et al., 2014; Hildebrand et al., 2014; Wang et al., 2014). Necroptosis can be induced by either RIPK1-dependent or RIPK1-independent mechanisms upon diverse stimuli. In RIPK1-dependent necroptosis, the binding of tumor necrosis factor (TNF) to tumor necrosis factor receptor 1 (TNFR1) induces a conformational change in TNFR1 trimers. TNFR1 subsequently leads to the recruitment of downstream proteins, including RIPK1, TRAF2 (TNFR-associated factor 2), TRAF5, TRADD (TNFR-associated death domain), cIAP1 (cellular inhibitor of apoptosis protein 1), and cIAP2. This membrane-bound protein complex is

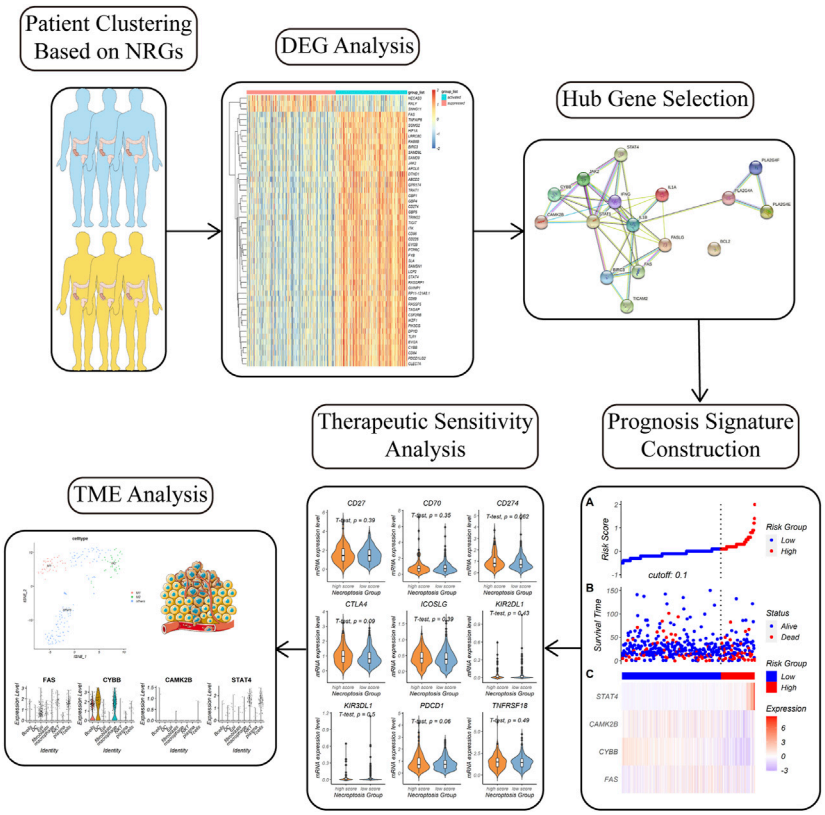


FIGURE 1 | Workflow of the study. Colon cancer patients were divided into two subgroups with distinct molecular and clinicopathologic features based on necroptosis related genes (NRGs). Differential gene analysis, hub gene analysis, lasso regression, and multivariate Cox regression was carried step by step to figure out a necroptosis signature. The necroptosis signature could predict prognosis and therapeutic response in colon cancer patients. We also investigated the relationship between the necroptosis signature and tumor microenvironment (TME) by single-cell analysis.

called complex I (Vandenabeele et al., 2010). A cytosolic death-inducing complex comprised of FADD (FAS-associated death domain protein), RIPK1, caspase-8 and TRADD is formed afterwards, which is known as complex II (Tenev et al., 2011). Deubiquitinated of RIPK1 switches the cell death mode from apoptosis to necroptosis (Vandenabeele et al., 2010). Autophosphorylated RIPK1 interacts with RIPK3 through their RIP homotypic interaction motif (RHIMs) (Li et al., 2012), leading to the formation of the necrosome complex (Li et al., 2012). Mitochondrial reactive oxygen species (ROS) (Zhang et al., 2017) and cylindromatosis (CYLD) (Moquin et al., 2013) was reported to be important for necrosome formation. In necrosomes, RIPK3 phosphorylates its substrate MLKL. MLKL is then oligomerized and translocated to the plasma membrane, leading to the execution of necroptosis. In RIPK1-independent necroptosis, inducers including Toll-like receptor 3 (TLR3), TLR4 and interferons (IFNs) can directly recruit and activate RIPK3 and MLKL (Weinlich et al., 2017). RIPK1 behaves in an inhibitory manner in combination with caspase 8, FADD and FLIP (FLICE-like inhibitory protein) in this case (Weinlich et al., 2017).

Necroptosis is a double-edged sword in many cancer types. On the one hand, necroptosis plays an antitumor role as a form of programmed cell death. On the other hand, necroptosis triggers inflammatory responses and is reported to promote cancer metastasis and immunosuppression (Gong et al., 2019). Colorectal cancer (CRC) is the second most common cause of cancer death in the United States (Siegel et al., 2020). In CRC, although the tumor-suppressing effects of RIPK3 and RIPK1 have been discovered (Moriwaki et al., 2015), RIPK3-mediated inflammation was reported to promote intestinal tumors by inducing an immune-suppressive tumor microenvironment (TME) (Jayakumar and Bothwell, 2019; Liu et al., 2019), and RIPK1 has been shown to interact with mitochondrial Ca^{2+} uniporter (MCU) to promote colorectal oncogenesis (Zeng et al., 2018). Besides, MLKL exhibits a suppressive effect during intestinal tumorigenesis in various researches (Zhao et al., 2019; Zhao et al., 2021). And it was reported that genetic deletion of MLKL had no impact on colon cancer development (Alvarez-Diaz et al., 2021). One possible reason for such differences might be that necroptotic cells can release various regulatory cytokines, which can either facilitate neoplastic progression by stimulating the proliferation of neighboring cells or result in tumor cell elimination by activating cytotoxic $CD8^+$ T lymphocytes (Gong et al., 2019). A growing number of studies have reported the anti-tumor effect of chemotherapy (Oliver Metzig et al., 2016), radiotherapy (Nehs et al., 2011) and

immunotherapy (Van Hoecke et al., 2018) in a necroptosis-dependent manner. These findings shed light on the complexity and importance of necroptosis in cancer. Therefore, it is urgent to systematically analyze the relationship between necroptosis and colon cancer progression and the therapeutic response.

The design of this research is in **Figure 1**. In this study, we downloaded genes related to necroptosis from Gene Ontology (GO) database and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database to obtain a list of necroptosis related genes (NRGs). We then identified two subtypes based on the NRGs in the TCGA-COAD cohort. The two NRGs-based subtypes have distinct clinical features and molecular characteristics. After differentially expressed gene analysis and Lasso regression, a total of four necroptosis signature genes were included to establish a prognostic model. On the basis of the necroptosis scoring model, responses to chemotherapy and immunotherapy were analyzed. Single-cell analysis revealed the difference in tumor microenvironment between the two necroptosis groups. Our study shed light on the essential role of necroptosis in colon cancer, which could be useful in prognosis prediction and guiding therapy in clinical practice.

MATERIALS AND METHODS

Study Population

The RNA transcriptome datasets (HTSeq-counts and HTSeq-FPKM) and the relevant clinical information of 469 primary colon cancer patients from the TCGA Colon Cancer (TCGA-COAD) cohort were downloaded from UCSC Xena website (https://xenabrowser.net/datapages/). Raw counts data was used for identifying differentially expressed genes. For the other analysis, Fragments Per Kilobase of transcript per Million mapped reads (FPKM) was transformed into log2 (FPKM +1). Corresponding somatic mutation profile was also downloaded from UCSC Xena website. Patients in the TCGA-COAD cohort were divided into the training set and validation set randomly. The training set was used for constructing LASSO regression model and the model was validated in the validation set.

Two datasets, GSE28722 and GSE17538, were downloaded using the R package "GEOquery." GSE28722 and GSE17538 contain mRNA and survival profiles of 129 CRC patients and 244 colon cancer patients respectively.

Single-cell transcriptome file and clinical data of 12 CRC samples of GSE166555 was downloaded from the Gene expression omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/).

Patient phenotype information was listed in **Supplementary Table S1**. Baseline characteristics including age, gender, and tumor stage among the four datasets were similar, and also fitted well with previous studies (Siegel et al., 2020).

Acquisition of Necroptosis Related Genes

The necroptosis gene set "hsa04217" contains 159 NRGs and was downloaded from the KEGG database by using the R package "KEGGREST." Another necroptosis gene set "0097528" containing three NRGs was downloaded from the GO database using the R package "GO.db."

Necroptosis Related Genes Based Clustering

We performed K-means consensus clustering with the FPKM matrix of 159 NRGs to identify subgroups in the TCGA-COAD cohort. Consensus clustering was carried out using the function "ExecuteCC" in the R package "CancerSubtypes" (Xu et al., 2017). The number of clusters was determined based on both the clustering results and clinical significance. Heatmap was generated by using the function "drawHeatmap" in the R package "CancerSubtypes". Data were normalized using "max_min" [(value−min)/(max−min)] before drawing the heatmap.

Clinical and Mutational Characteristics of Necroptosis Subgroups

We then compared clinicopathological and molecular characteristics between the two necroptosis subgroups previously identified. Clinical information including pathological and clinical stage were extracted for comparison and the results were presented in bar charts. The Kaplan-Meier (K-M) method was performed to compare overall survival (OS) between the two necroptosis subgroups. The somatic mutation data were further analyzed using the R package "maftools" (Mayakonda et al., 2018). Somatic mutation types that "maftools" could detect include synonymous variant, missense variant, stop-gain, frameshift variant, three prime untranslated region (3′-UTR) variant, intron variant, and multi-hit. It can also provide the concrete basepair substitution information in each individual.

Immune Cell Infiltration Analysis of Necroptosis Subgroups

To examine the relationship between necroptosis and the immune microenvironment, "CIBERSORT" was used to compare the absolute abundance of 22 human hematopoietic cell phenotypes between the two necroptosis subgroups. CIBERSORT is a tool for deconvolution of the expression matrix of human cell subtypes from tissue gene expression profiles based on the principle of linear support vector regression (Newman et al., 2015). Cell types CIBERSORT could identify include seven T-cell subsets, naïve and memory B cells, natural killer (NK) cells, plasma cells and myeloid subtypes. The standard annotation file LM22 containing 547 genes was provided as input.

Differentially Expressed Genes Identification

Differentially expressed genes (DEGs) between the NRGs-activated and NRGs-suppressed subtypes were identified by using the R package "limma." NRGs with | log2 (fold change)

| > 1 and adjusted p value <0.05 were considered as necroptosis subtype specific genes. Gene set enrichment analysis (GSEA) method was carried out to determine the signaling pathways the DEGs involved in with the R package "clusterProfiler." For GSEA analysis using the function "gseKEGG," permutation number was set at 1,000, minimal size of each gene set for analyzing was set at 120, p cutoff value was set at 0.9, and "BH" was chosen for the "pAdjustMethod" parameter. Adjusted p value was set at 0.05 for figuring out significantly up-regulated and downregulated pathways.

Hub Genes Selection

Proteins are executors of biological processes. To extend our research conclusion from genomics to proteomics, we generated a protein-protein interaction (PPI) network including both functional and physical associations by importing the DEGs related to necroptosis into STRING (https://www.string-db.org/) (Szklarczyk et al., 2021). The false discovery rate (FDR) was set at 0.05 and the minimum interaction score was set at 0.4. Then we processed the result with the Cytoscape software (version 3.9.0). Hub genes were identified by the Degree algorithm using the "cytoHubba" plugin (Chin et al., 2014).

Establishment and Validation of a Prognostic Necroptosis Signature

To determine significant prognostic genes among the hub genes, we then applied the Least absolute shrinkage and selection operator (LASSO) method for variable selection in a Cox regression model by using the R packages "lars" and "glmnet." We extracted the hub genes related to necroptosis when the first-rank value of Log(λ) was the minimum likelihood of deviance. A multivariate Cox regression analysis was then used to investigate the correlation between the expression levels of the necroptosis signatures and the overall survival (OS) of patients in the TCGA-COAD cohort using the R package "survival." The necroptosis score could be calculated based on the Cox model using the formula:

$$\text{Necroptosis Score} = \exp\left[\sum_{i=1}^{p} b_i X_i - \sum_{i=1}^{p} b_i \bar{X}_i\right]$$

- The coefficients ($b_1, b_2, \ldots, b_p$) are the coefficients of each gene in the Cox model.
- Xi is the mRNA expression level of the ith gene.
- $\bar{X}_i$ is the mean mRNA expression level of the ith gene.

Based on the necroptosis score, patients in the TCGA-COAD cohort, GSE28722, and GSE17538 datasets were divided into high-necroptosis score and low-necroptosis score subgroups. The function "ComBat" in the R package "sva" (Leek et al., 2012) was used to remove batch effects from the GSE28722 and GSE17538 datasets. The optimal cutoff point for necroptosis score was calculated using the R package "survminer" according to the expression level and the survival information. The K-M method was performed to compare overall survival (OS) between the two subgroups in the GSE28722 and GSE17538 datasets respectively.

Construction and Validation of Nomogram Based on the Necroptosis Signature

With R package "rms," the necroptosis score, age, gender, and tumor stage of the colon cancer patients in the TCGA-COAD cohort were used to set up a nomogram for the 1-, 2-, and 5-year OS. Calibration curves were generated to evaluate the agreement between the actual and predicted survival probabilities at 1-, 2-, and 5-year. Parameters m and B in the function "calibrate" were both set at 100. The 1-, 3-, and 5-year time-dependent receiver operating characteristics (ROC) curves of the model were generated by the R package "survivalROC" (Heagerty and Zheng, 2005).

Prediction of Necroptosis Signature in the Field of Chemotherapy and Immunotherapy

We applied the R package "oncoPredict" to predict clinical response to multiple chemotherapy drugs in the high-necroptosis score and low-necroptosis score groups, which is based on Genomics of Drug Sensitivity in Cancer (GDSC) and Cancer Therapeutics Response Portal (CTRP) (Maeser et al., 2021).

The immunotherapy response prediction of the two subgroups was estimated with tumor immune dysfunction and exclusion (TIDE) score and tumor inflammation signature (TIS) score. TIDE is a computational method that could be used to predict immune checkpoint blockade (ICB) response in cancer patients by computing T cell dysfunction and T cell exclusion (Jiang et al., 2018). We obtained TIDE scores and T cell dysfunction scores from the TIDE web (http://tide.dfci.harvard.edu/). TIS score was calculated as an average value of log2-scale normalized expression of 18 signature genes to predict clinical response to PD-1 blockade (Ayers et al., 2017). Besides, we compared mRNA expression levels of commonly accepted immunotherapy-related genes between the high-necroptosis score and low-necroptosis score subgroups.

Single-Cell Analysis for Necroptosis Heterogeneity Estimation

After filtering out low low-quality cells, single-cell transcriptomic data of cells from 12 primary samples in GSE166555 was used for further analysis (Uhlitz et al., 2021). Cell clusters were annotated based on previously reported cell type-specific signatures and marker genes. R package "Seurat" was used to process the data and generate t-SNE plot for cell types visualization. The function "AddModuleScore" was applied to calculate score of the necroptosis signature. The functions "FindNeighbors" and "FindClusters" were used to identify cell clusters of macrophages.

Statistical Analysis

For the comparison of continuous variables, the unpaired Student's t-test was applied for normally distributed data, and the Wilcoxon test or Kruskal–Wallis test was performed for non-normally distributed data. Two-sided Fisher's exact test was used to measure categorical variables

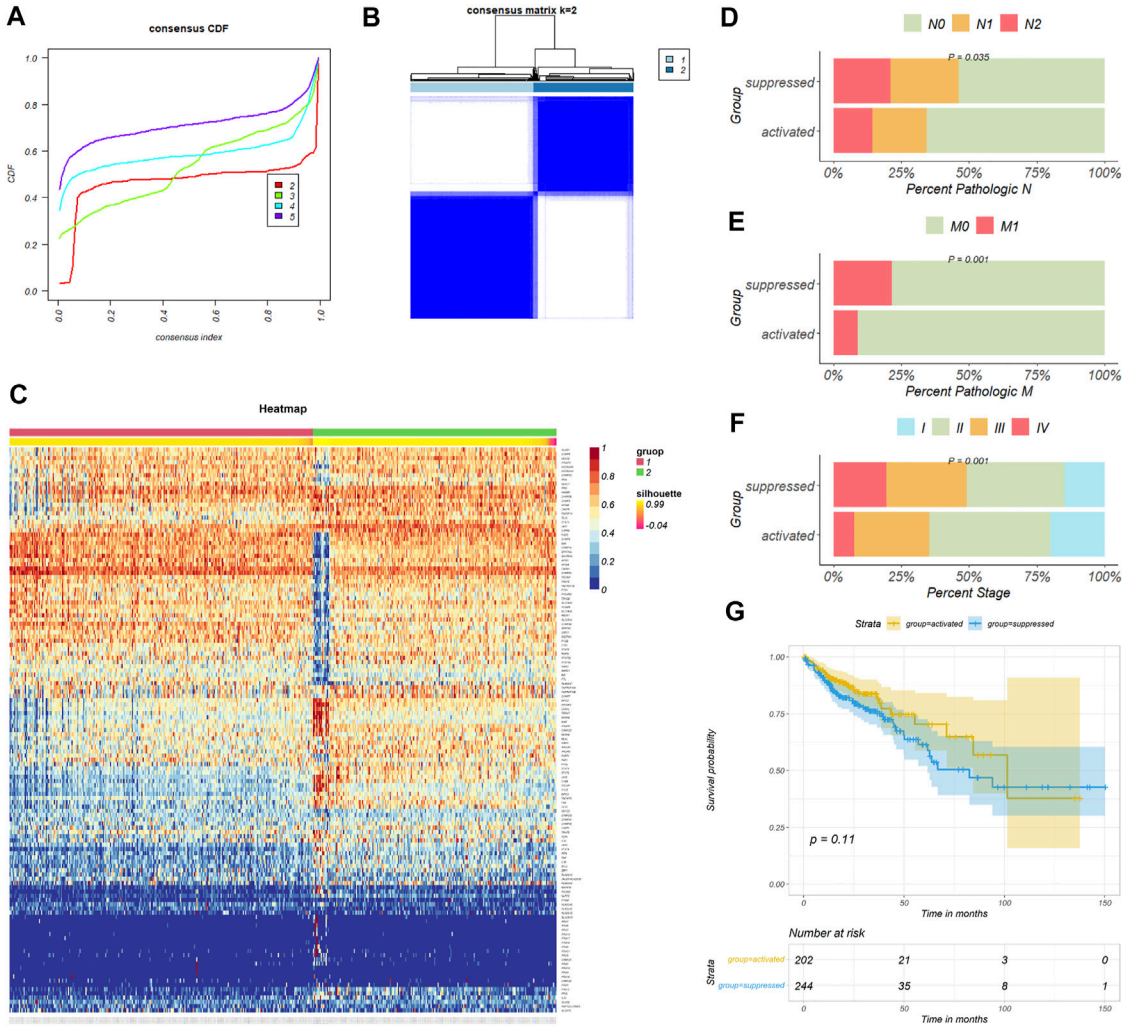


FIGURE 2 | Consensus clustering for necroptosis related genes (NRGs) in colon cancer patients. **(A)** The relative change in area under the CDF curve of K = 2–5. **(B)** The consensus matrix shows patients with two heterogenous necroptosis states in the TCGA-COAD cohort. **(C)** The heatmap shows expression levels of NRGs in the NRGs-activated and NRGs-suppressed group. **(D–F)** Difference of tumor pathological N, M, and clinical stage distribution between the two necroptosis subgroups respectively. **(G)** Kaplan-Meier curves for overall survival based on necroptosis subgroups (Log-rank test) in the TCGA-COAD cohort.

between two groups. $p < 0.05$ was set as a significant difference in all statistical methods, and all p values were two-tailed. R software (version 4.1.1) (http://www.R-project.org) was used for data analysis and generation of figures.

RESULTS

Necroptosis Related Genes Define Subgroups With Different Clinical Characteristics in Colon Cancer Patients

159 necroptosis related genes (NRGs) involved in the pathway "hsa04217" were downloaded from KEGG database, and three NRGs were downloaded from GO database by the accession number "0097528." After removing duplicate genes, a total of 159 NRGs were finally engaged in this study (**Supplementary Table S2**). 469 colon cancer patients from the TCGA-COAD cohort were then divided into distinct subtypes based on 159 NRGs expression profiles. Taking the consensus clustering results and clinical significance into consider (**Figure 2A**), two necroptosis subgroups were identified. Cluster 1 (n = 260, 55.4% of all colon cancer patients) was defined as the NRGs-suppressed subtype (**Figure 2B**), according to the relative downregulation of most NRGs in this cluster (**Figure 2C**). Cluster 2 (n = 209, 44.6% of all colon cancer patients) was therefore defined as the NRGs-activated subtype based on the relative upregulation of NRGs. The two subtypes showed thoroughly heterogenous clinical outcomes (**Table 1**). Patients in the NRGs-activated subgroup had lower pathological N stage (p = 0.035) (**Figure 2D**), pathological M stage (p = 0.001) (**Figure 2E**), and clinical stage (p = 0.001)

TABLE 1 | Clinical characteristics of colon cancer patients in the NRGs-activated and NRGs-suppressed group.

	NRGs-activated group	NRGs-suppressed group	*p* value	Test
Pathologic T stage (%)			0.763	Exact
T1	5 (2.4)	6 (2.3)		
T2	39 (18.7)	41 (15.9)		
T3	137 (65.6)	180 (69.8)		
T4	27 (12.9)	31 (12.0)		
Tis	1 (0.5)	0 (0.0)		
Total	209	258		
Pathologic N stage (%)			0.035	Exact
N0	137 (65.6)	139 (53.9)		
N1	42 (20.1)	65 (25.2)		
N2	30 (14.4)	54 (20.9)		
Total	209	258		
Pathologic M stage (%)			0.001	Exact
M1	16 (8.9)	49 (21.5)		
M2	164 (91.1)	179 (78.5)		
Total	180	228		
Clinical stage (%)			0.001	Exact
I	41 (19.9)	37 (14.8)		
II	92 (44.7)	90 (36.0)		
III	57 (27.7)	74 (29.6)		
IV	16 (7.8)	49 (19.6)		
Total	206	250		

(**Figure 2F**) compared with patients in the NRGs-suppressed subgroup. K-M plots suggested that patients who were divided into the NRGs-activated subgroup tended to have better OS relative to patients in the NRGs-suppressed subgroup, but the results might be biased by limited sample size (**Figure 2G**). Together, these results suggested that our clustering method based on NRGs was reasonable and had clinical significance in colon cancer patients.

Necroptosis Related Genes-Based Subtypes Show Different Mutational and Immunological Characteristics

To identify genomic alterations difference between the two subgroups, we compared NRGs mutation between the NRGs-activated subgroup and the NRGs-suppressed subgroup. The most frequently mutated NRGs in the NRGs-activated subgroup and the NRGs-suppressed subgroup were NLRP3 and GLUD2 (**Figure 3A**). Besides, the NRGs-activated subgroup had a higher probability of necroptosis-related genes mutation than the NRGs-suppressed subgroup (**Figure 3A**).

Immune cell infiltration markedly influenced tumor microenvironment. Therefore, we also explored differences in immune cell infiltrations between two necroptosis subgroups (**Supplementary Table S3**). Notably, M0 macrophages ($p < 0.0001$) and resting dendritic cells ($p < 0.01$) were significantly downregulated in the NRGs-activated subgroup. Neutrophils ($p < 0.0001$), M1 macrophages ($p < 0.001$), activated dendritic cells ($p < 0.001$), M2 macrophages ($p < 0.01$), follicular helper T cells ($p < 0.01$), and plasma cells ($p < 0.01$) were significantly up-regulated in the NRGs-activated subgroup (**Figure 3B**).

Necroptosis Subtype Signature is a Prognostic Indicator for Colon Cancer Patients

To obtain a more practical signature that could be used for necroptosis subtype identification, we next figured out hub DEGs between the NRGs-activated group and the NRGs-suppressed group. Differential gene analysis revealed that 2,060 genes were significantly upregulated in the NRGs-activated subgroup (**Figure 4A**), and 836 genes were significantly downregulated (**Figure 4A**; **Supplementary Table S4**). GSEA results revealed that the DEGs were mainly enriched in the pathways including cytokine-cytokine receptor interaction, chemokine signaling pathway, JAK-STAT signaling pathway, PI3K-Akt signaling pathway, transcriptional misregulation in caner, proteoglycans in cancer, phagosome, NOD-like receptor signaling pathway, natural killer mediated cytotoxicity, focal adhesion, and cell adhesion molecules (**Figure 4B**; **Supplementary Table S5**). Among the 2,896 DEGs, we further chose 16 NRGs to generate a PPI network in STRING. Differentially expressed necroptosis genes between the NRGs-activated and suppressed groups included JAK2, STAT4, BIRC3, CYBB, FAS, IFNG, STAT1, BCL2, FASLG, TICAM2, IL1B, PLA2G4F, PLA2G4E, PLA2G4A, IL1A, and CAMK2B. After importing the PPI network including both functional and physical associations generated by STRING into Cytoscape, we figured out 10 hub genes: IL1B, IFNG, STAT1, JAK2, FASLG, FAS, BIRC3, CYBB, CAMK2B, and STAT4 (**Figure 4C**). To further reduce the dimension of the necroptosis signature, we randomly allocated the patients in the TCGA-COAD cohort into the training set ($n = 233$, 53.8% of all colon cancer patients) and the validation set ($n = 200$, 46.2% of all colon cancer patients), and applied the LASSO Cox regression model to find out the most powerful prognostic

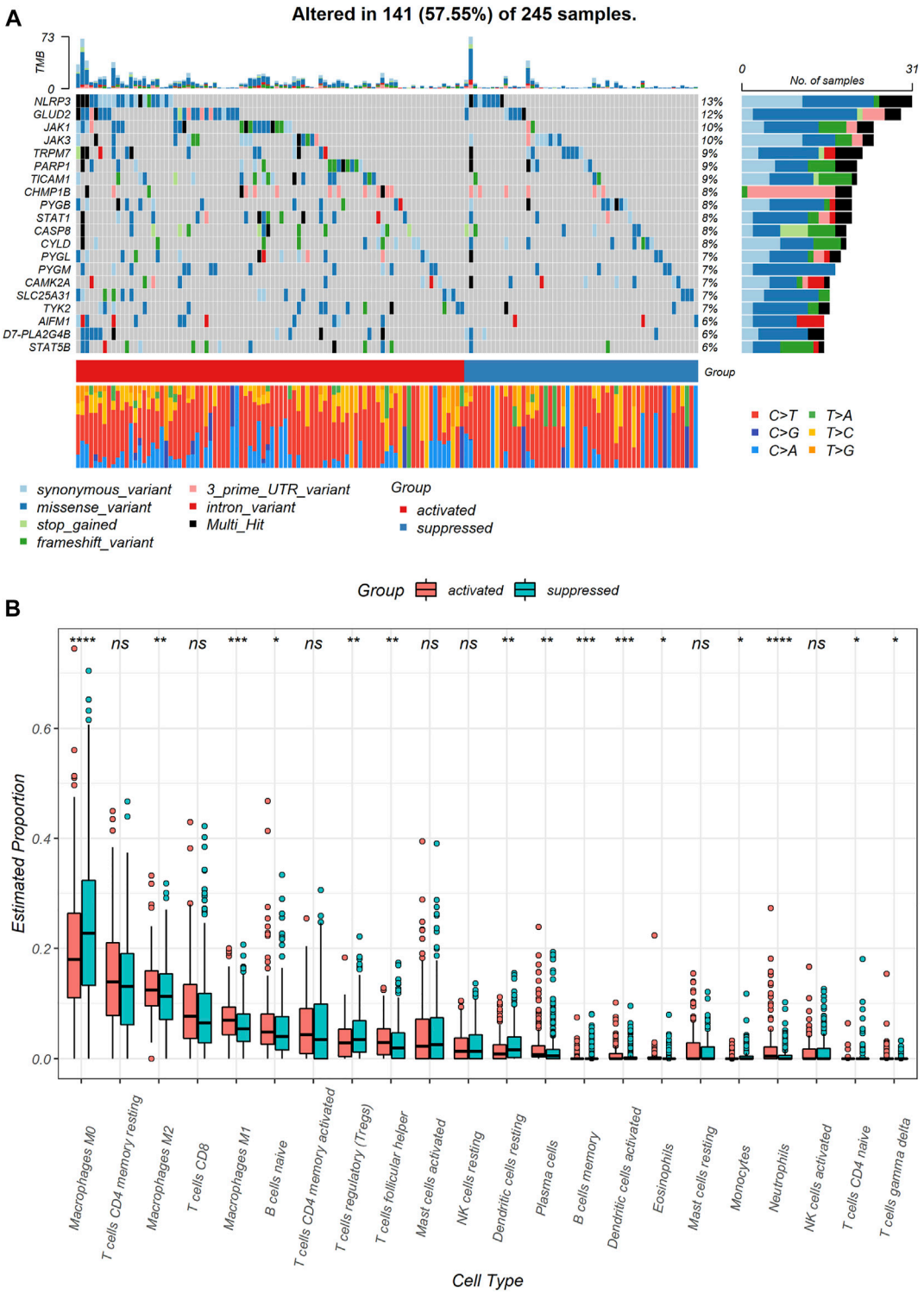


FIGURE 3 | Genomic and immune profile alterations between the NRGs-activated and NRGs-suppressed subgroup. **(A)** Landscape of mutation profiles in colon cancer samples. Mutation information of each gene in each sample is shown in the waterfall plot. Top panel shows individual tumor mutation burden. **(B)** The comparison of infiltration levels of 22 immune cells between the NRGs-activated and NRGs-suppressed groups. * represents $p < 0.05$, ** represents $p < 0.01$, *** represents $p < 0.01$, **** represents $p < 0.0001$, ns represents no significant difference.

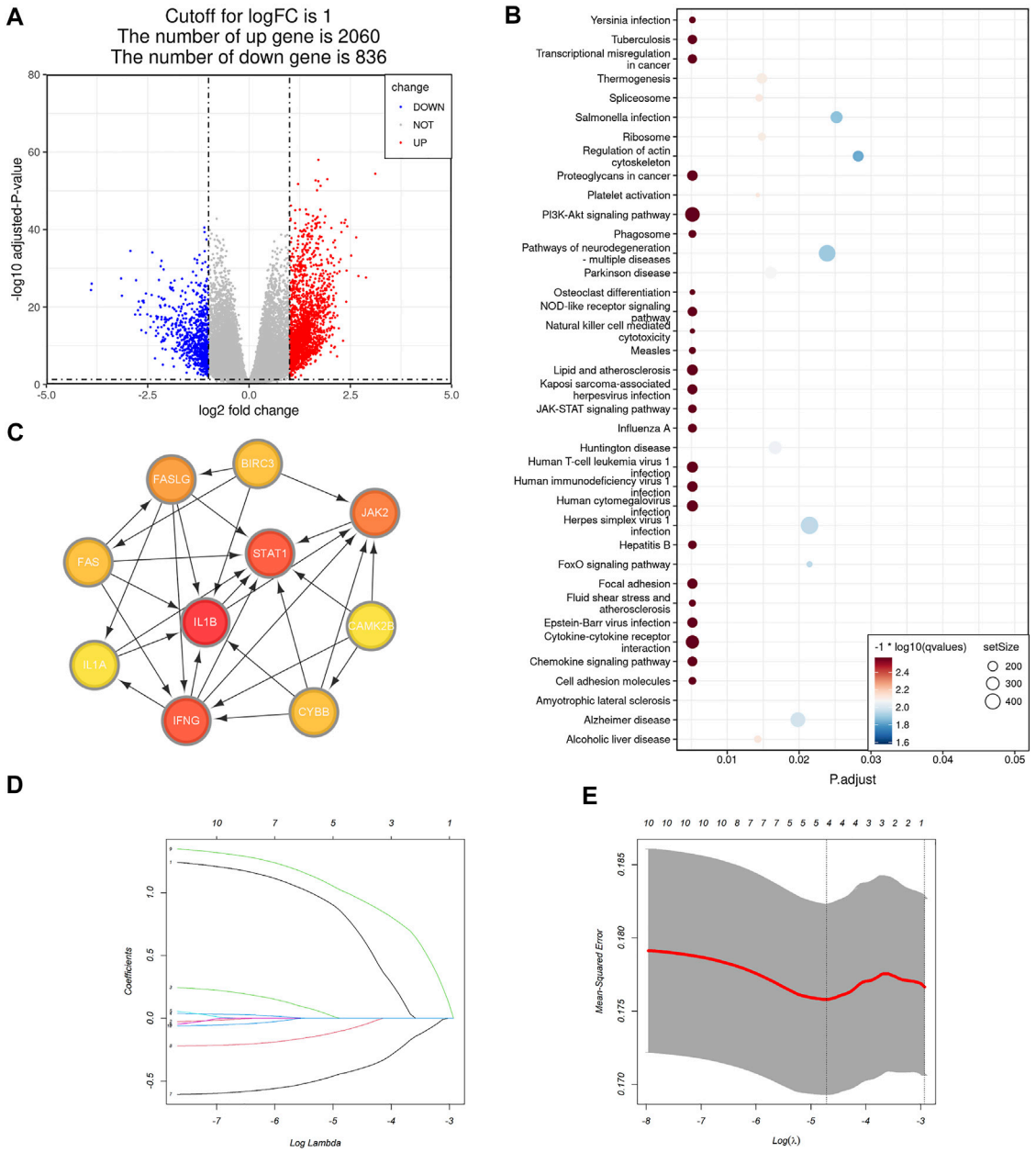


FIGURE 4 | Selection of the necroptosis gene signature in colon cancer. **(A)** Volcano plot of differentially expressed genes (DEGs) between the NRGs-activated group and NRGs-suppressed group. The red dots represent upregulated genes, while the blue dots represent downregulated genes [| log2 (fold change) | > 1 and adjusted *p* value <0.05]. **(B)** Bubble plot of GSEA pathway enrichment results of the DEGs between the NRGs-activated group and NRGs-suppressed group. Size of the circle represents the size of the gene set and color of the circle is based on -log10 (q value). **(C)** Protein-protein interaction (PPI) network including both functional and physical associations of hub necroptosis genes identified in Cytoscape. Color shade of the gene correlates with its score obtained by Degree method using the "cytoHubba" plugin. **(D)** The 10-fold cross-validation for variable selection in the LASSO model. **(E)** The LASSO coefficient profile of the four necroptosis signature genes.

necroptosis genes in the training set. This resulted in a necroptosis signature of four genes: FAS, CYBB, CAMK2B, and STAT4 (**Figures 4D,E**).

A multivariate Cox proportional hazards model was used to find the relationship between the necroptosis signature genes and OS in the TCGA-COAD cohort and the results were displayed in a forest plot (**Figure 5A**). Among all the necroptosis signature genes, the oncogene signal transducer and activator of transcription 4 (STAT4) is a strong predictor of OS in colon cancer patients ($p = 0.01688$, HR = 2.9599, 95%CI: 1.2153–7.209) (**Figure 5A**). Necroptosis scores of each patient in the TCGA-COAD cohort were calculated based on the Cox model as previously mentioned. A nomogram was constructed combining age, gender, and tumor stage with necroptosis

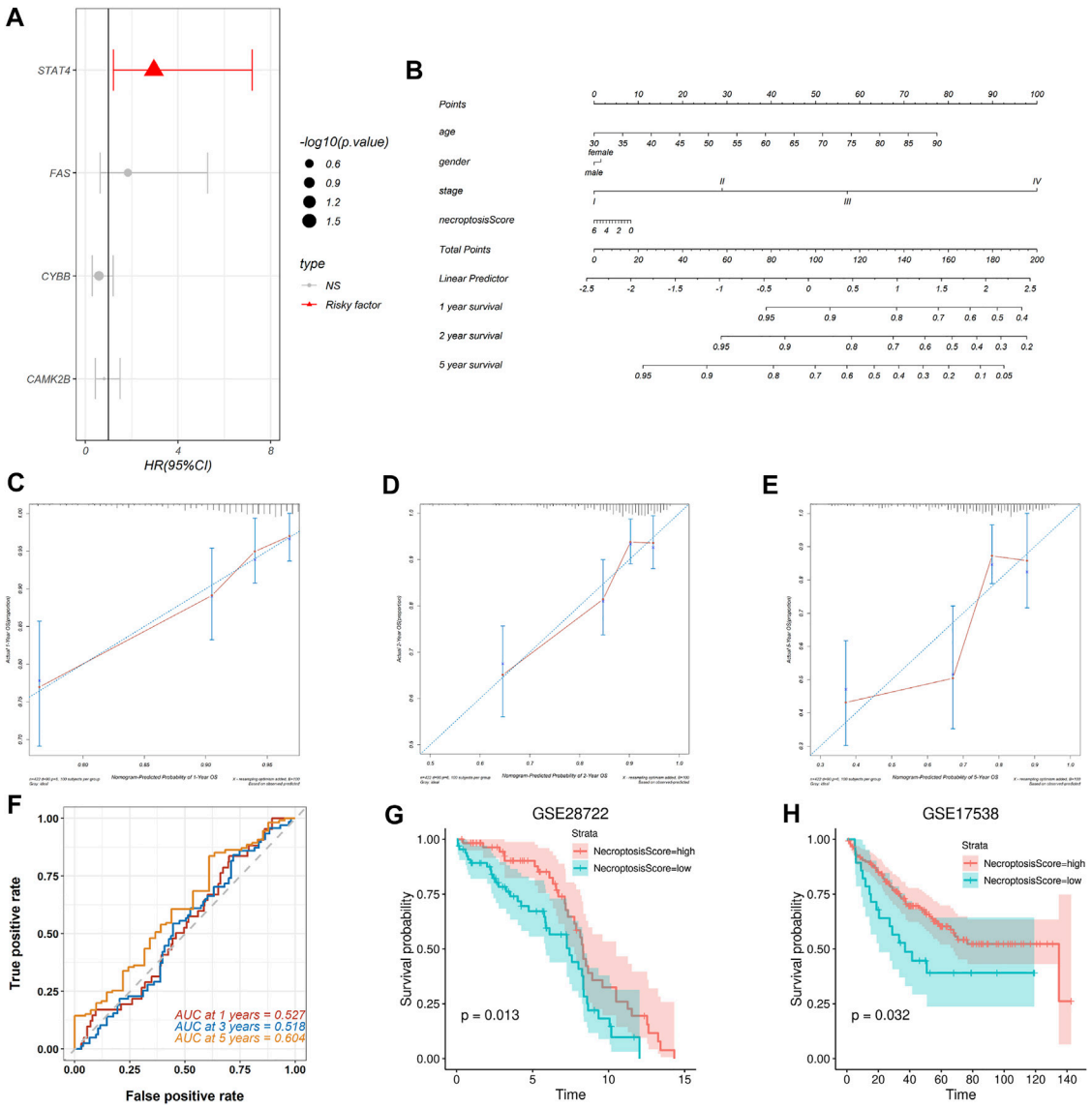


FIGURE 5 | Construction and validation of the necroptosis prognosis model based on the necroptosis gene signature. **(A)** The forest plot shows multi-Cox analysis result between the four necroptosis signature genes and OS. **(B)** The nomogram integrates age, gender, tumor stage, and necroptosis score to predict the probability of the 1-, 2-, and 5-year OS. **(C,D)** The calibration curves for 1-, 2-, and 5-year OS, respectively. **(E)** The 1-, 3-, and 5-year time-dependent ROC for necroptosis signature survival prediction in the TCGA-COAD database, respectively. **(F–H)** Validation of the necroptosis signature in predicting the prognosis of patients in GSE28722 and GSE17538 datasets respectively.

score to offer clinicians a quantitative approach for predicting OS in colon cancer patients (**Figure 5B**). A higher necroptosis signature score is correlated with worse overall survival (**Figure 5B**). The calibration curve showed good concordance between the observations and the predictions at 1-, 2-, and 5-year (**Figures 5C–E**). In addition, time-dependent receiver operating characteristics (ROC) were used to evaluate the sensitivity and specificity of the prognostic model based on the necroptosis signature. The 1-, 3-, and 5-year AUC of the TCGA-COAD cohort were 0.527, 0.518, and 0.604, respectively (**Figure 5F**). Together, these results indicated that the necroptosis signature could be used for prognosis prediction in colon cancer patients.

To validate prognostic predictive power of the necroptosis signature, we assessed its performance in two independent GEO cohorts. We calculated the necroptosis score for each patient and separated the patients in each cohort into the high-necroptosis score group and low-necroptosis score group as previously mentioned. Kaplan-Meier survival analysis revealed that patients in the high-necroptosis score group had better OS than the low-necroptosis score group, both in the dataset GSE28722 (p = 0.013) (**Figure 5G**) and the dataset GSE17538 (p = 0.032) (**Figure 5H**). These results suggested that the necroptosis signature we proposed could be extended to other colon cancer cohorts to predict prognosis.

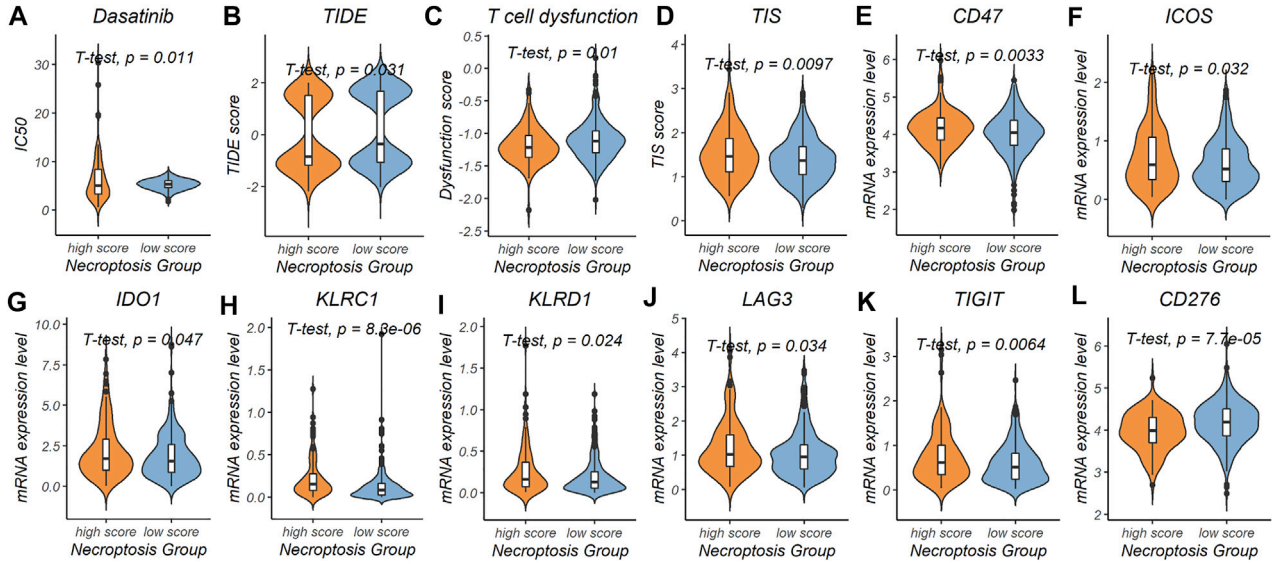


FIGURE 6 | Predictive power of the necroptosis gene signature in chemotherapy and immunotherapy response. **(A)** Prediction of IC50 for dasatinib in the high-necroptosis score and low-necroptosis score group. **(B–D)** Comparison of tumor immune dysfunction and exclusion (TIDE) score, T cell dysfunction score and tumor inflammation signature (TIS) score between the high-necroptosis score and low-necroptosis score group. **(E-L)** mRNA expression level differences of immunotherapy related genes (CD47, ICOS, IDO1, KLRC1, KLRD1, LAG3, TIGIT, and CD276) between the high-necroptosis score and low-necroptosis score group.

Necroptosis Subtype Signature is a Predictor for Chemotherapy and Immunotherapy

To explore the prediction power of the necroptosis subtype signature in the field of chemotherapy and immunotherapy response, patients in the TCGA-COAD cohort were divided into the high-risk subgroup (n = 122, 28.2% of all colon cancer patients) and the low-risk subgroup (n = 311, 71.8% of all colon cancer patients) based on the necroptosis risk score. After evaluating 198 chemotherapy drugs (**Supplementary Table S6**), the low-necroptosis score subgroup was found to have lower half-maximal inhibitory concentration (IC50) for dasatinib compared with the high-necroptosis score subgroup (p = 0.011) (**Figure 6A**). This indicated that colon cancer patients with low necroptosis score might benefit from dasatinib.

We further used the TIDE score and the T cell dysfunction score to evaluate possible response to immunotherapy in the high- and low-necroptosis score subgroup (**Supplementary Table S7**). Patients in the high-necroptosis score subgroup were characterized by a significantly lower TIDE score (p = 1.9e-14) and T cell dysfunction score (p = 1.9e-14), and higher TIS score (p = 0.0097) compared with patients in the low-necroptosis score subgroup (**Figures 6B–D**). This indicated that high-necroptosis score patients seemed to be more sensitive to ICB. We also compared mRNA expression levels of 17 previously reported immunotherapy related genes between the high- and low-necroptosis score subgroup (**Supplementary Table S2**). Interestingly, distinct necroptosis score groups had different immunotherapy-related genes expression levels. CD47 (p = 0.0033), ICOS (p = 0.032), IDO1 (p = 0.047), KLRC1 (p = 8.3e-06), KLRD1 (p = 0.024), LAG3 (p = 0.034), and TIGIT (p = 0.0064) were up-regulated in the high-necroptosis score subgroup (**Figures 6E–K**), while CD276 was up-regulated in the high-necroptosis score subgroup (p = 7.7e-05) (**Figure 6L**). Together, these results revealed that the necroptosis signature we proposed was also helpful in predicting therapeutic response in colon cancer patients.

Necroptosis Signature is Heterogeneous in Tumor Immune Microenvironment

As previous GSEA result showed that DEGs between the NRGs-activated group and the NRGs-suppressed group were enriched in proteoglycans in cancer, focal adhesion, cell adhesion molecules, and natural killer mediated cytotoxicity, we further explore the relationship between necroptosis and TME at a single-cell level. Based on the previously reported cell type-specific signatures and marker genes (**Supplementary Table S2**), cells from 12 primary CRC patients were divided into eight types, including B cell, dendritic cell, epithelial cell, fibroblast, macrophage, NKT cell, plasma cell, and T cell (**Figure 7A**). Results from single-cell analysis revealed that expression levels of four necroptosis signature genes varied to a large degree in different cell clusters. The necroptosis signature genes were significant upregulated in some particular cell types, especially macrophages (**Figure 7B**). Among all the necroptosis signatures, CYBB was the most frequently upregulated necroptosis signature in macrophages (**Figure 7C**).

Based on this result, we picked the macrophages cluster out for further exploration. Five clusters were identified using the R package "Seurat" according to gene expression level (**Figure 7D**). We found that the expression of CYBB were quite different among different clusters of macrophages (p =

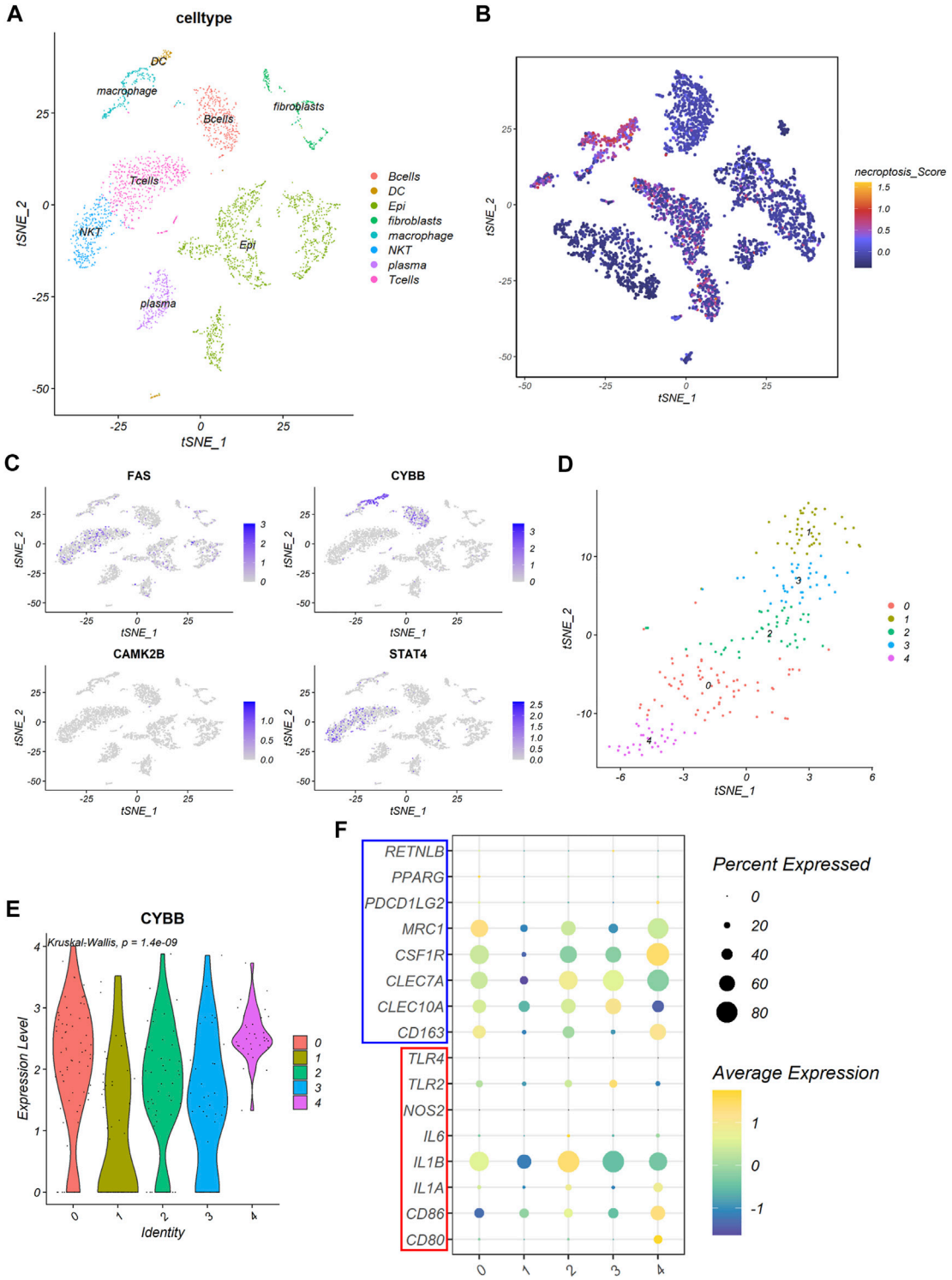


FIGURE 7 | The expression of the necroptosis signature in colon cancer. **(A)** The t-SNE plot shows percentage of eight cell types in colon cancer tissues. **(B)** The t-SNE plot shows necroptosis score of each cell cluster. **(C)** The t-SNE plots shows expression levels of FAS, CYBB, CAMK2B, and STAT4 in each cell cluster. **(D)** The t-SNE plot shows five clusters of macrophages, denoted by 0–4, identified by the R package Seurat. **(E)** The violin plot shows expression levels of CYBB among the five macrophage clusters. **(F)** The bubble plot shows expression of macrophage marker genes among the five macrophage clusters. Gens in the red box are M1 macrophage marker genes and genes in the blue box are M2 macrophage marker genes.

1.4e-09), with cluster 0 and cluster 4 had the highest expression level among all the macrophage clusters (**Figure 7E**). Further analysis revealed that cluster 0 expressed a high level of MRC1, while cluster 4 expressed a high level of CSF1R (**Figure 7F**). This indicated that cluster 0 and cluster 4 all exhibited molecular characteristics of M2 macrophages.

DISCUSSION

Necroptosis is a form of programmed cell death which has attracted more and more attention in recent years. However, conflicting evidence existed in studies about the relationship between necroptosis and different tumor types. Inducing tumor cell necroptosis is a potential treatment strategy but might be a double-edged sword. Colorectal cancer (CRC) is the second most common cause of cancer death in the United States. Therefore, analyses of the relationship between necroptosis and colon cancer were performed to attain a better understanding of the mechanisms of how necroptosis involved in oncogenesis, development, and metastasis of colon cancer. Immunotherapy is proving to be an effective therapeutic approach in a variety of cancers. But only a subset of cancer patients exhibits durable responses, suggesting that a deeper investigation of cancer immunity is required. Thus, we also explored whether different levels of NRGs expression indicated suitable targets for different therapies in colon cancer.

First, we revealed that colon cancer patients could be characterized into two distinct necroptosis statuses with different characteristics. Patients in the NRGs-activated subgroup were characterized by aggressive clinical behavior, such as advanced pathological N, M, and clinical stage. However, no difference in the pathological T stage between the two subgroups was discovered. This indicated that necroptosis might play an important role in the metastasis in colon cancer development. We also identified the mutational characteristics of distinct necroptosis subtypes in colon cancer patients. We found that NLRP3 and GLUD2 were the most common necroptosis-related gene alteration in colon cancer patients. Glutamate dehydrogenase 2 (GLUD2) overexpression was found to inhibit glioblastoma cell growth (Franceschi et al., 2018). Our findings provided novel potential drug targets for the control of colon cancer progression and metastasis. Besides, the NRGs-activated subtype had more somatic mutations of NRGs, suggesting that NRGs mutation burden may predict the clinical and pathological characteristics of colon cancer patients.

Next, we proposed a four gene necroptosis signature for the prediction of prognosis and therapeutic response in colon cancer patients. DEGs between the NRGs-activated subgroup and NRGs-suppressed subgroup were mainly enriched in tumor-related signaling pathways including JAK-STAT signaling pathway and PI3K-Akt signaling pathway, transcriptional misregulation and proteoglycans in cancer. Besides, pathways relating to the regulation of the TME, such as focal adhesion, cell adhesion molecules, and natural killer mediated cytotoxicity, were also involved. It has been reported that chlorpyrifos could induce necroptosis in fish liver cells by regulating the ROS/PTEN/PI3K/AKT axis (Wang et al., 2020). Proteoglycans consist a large proportion of the extracellular matrix (ECM) (Theocharis and Karamanos, 2019), and dysregulation of ECM dynamics leads to the development of cancer (Walker et al., 2018). Cell adhesion contributed a lot to cancer metastasis (Khalili and Ahmad, 2015), and focal adhesion kinase (FAK) was recognized as an anti-cancer target (Dawson et al., 2021). A previous study found that inhibition of cell-surface proteins induced by disintegrin and metalloproteases (ADAMs) disrupted cell adhesion while accelerating necroptosis (Cai et al., 2016). More studies are needed to explore the crosstalk between necroptosis and aforementioned pathways. For chemotherapy, our finding suggested that colon cancer patients with low necroptosis signature score might benefit from the selective tyrosine kinase receptor inhibitor dasatinib. Dasatinib plays an antitumor role in a variety of tumor types by triggering apoptosis of tumoral cells and changing tumor microenvironment (Montero et al., 2011). Our finding favored the combination use of dasatinib with other drugs to obtain a synergistic effect in certain colon cancer patients.

Our results also emphasized the important role of the necroptotic process in cancer immunity. The NRGs-activated subgroup and the NRGs-suppressed subgroup had significantly different tumor microenvironment. Activation of NRGs seemed to be related with inflammation by activating neutrophils, macrophages, dendritic cells, T cells, and B cells. Tumor-associated neutrophils (TANs) can play either pro- or anti-tumor roles depending on the subtypes they are polarized to upon external cues (Powell and Huttenlocher, 2016). Proportion of TANs subtypes needs further exploration to understand the relationship between necroptosis and TANs in the tumor microenvironment. Single-cell analysis highlighted the important relationship between necroptosis and macrophages in colon cancer, and also figured out the possible target turning cold tumors into hot tumors. Tumor-associated macrophages (TAMs) can be polarized toward either a pro-inflammatory (M1) state or an anti-inflammatory (M2) state upon various stimulation (Vitale et al., 2019). CYBB is a subunit of the NADPH oxidase complex 2 (NOX2). NOX2 and IL1B are all important pro-inflammation factors of phagocytes. Our research showed that CYBB (Cytochrome B-245 Beta Chain) was the most apparently upregulated necroptosis signature genes in macrophages, especially in M2 macrophages. This highlighted a novel target on macrophages to remodel TME. However, the underlying mechanism between NRGs and tumor immunity is still poorly understood and warrants further investigation.

In the field of immunotherapy, microsatellite instability-high (MSI-H)/deficient mismatch repair (dMMR) was first identified as a predictive biomarker of PD-1 blockade in a clinical trial (Tabernero et al., 2015). Anti-PD-1 antibody was only effective in dMMR patients (Tabernero et al., 2015). However, the proportion of patients with microsatellite stable (MSS) and proficient MMR (pMMR) metastatic colorectal cancer (mCRC) consists of more than 95% of mCRC patients (Yaeger et al., 2018), which means that the majority of CRC patients could not benefit from immunotherapy. We discovered that patients with high-necroptosis score were potential responders to anti-PD1 and anti-

CTLA4 therapy. This discovery provided theoretical support to the combination use of necroptosis inducers and immunotherapy to achieve a better therapeutic response in cancer patients. Physically induced necroptosis had already been proved to enhance the antitumor response of immune checkpoint blockade therapy (Um et al., 2020). Polyinosinic-polycytidylic acid (PolyI:C), a member of TLR family, is a necroptosis-inducing agent in tumor cells (Takemura et al., 2015). It was reported that PolyI:C could enhance the therapeutic effects of cancer immunotherapy by promoting T cell infiltration (Sultan et al., 2020). However, low-necroptosis score subgroup could benefit from targeting certain molecules, like CD276 on $CD8^+$ T cells. This finding suggested that immunotherapeutic treatment should be customized according to the necroptosis state of colon cancer patients. Our discovery might serve as a useful tool to identify colon cancer patients who might potentially benefit most from precision immunotherapy.

However, as the conclusions of this study were based on bioinformatic analysis using the TCGA and GEO databases, the relationship between necroptosis and colon cancer clinical characteristics needs further validation in prospective studies. The mechanisms underlying the effects of the necroptosis signature genes on colon cancer TME also needs to be verified experimentally.

In conclusion, this study revealed the significant relationship between necroptosis and colon cancer based on bioinformatics analysis. Identifying specific necroptosis state could help in colon cancer clinical management and decision-making process. The necroptosis signature we established could help in predicting prognosis among colon cancer patients and assists in developing more effective therapeutic targets in colon cancer.

DATA AVAILABILITY STATEMENT

Data generated and analyzed in the current study are available from the UCSC TCGA data portal (http://xena.ucsc.edu/public/) and GEO databases (dataset ID: GSE28722, GSE17538 and GSE166555, https://www.ncbi.nlm.nih.gov/).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Huazhong University of Science and Technology. The patients/participants provided their written informed consent to participate in this study. 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

YW, YH, BL, and XY conceived and designed the study; YW, YH, CL, and XW contributed to the data collection; YW, YH, and MY. performed the data analysis and data interpretation; YW, YH, MY, and DX contributed to the figures; YW and YH wrote the manuscript; BL revised the manuscript; XY reviewed the manuscript. All the authors had read and approved the submitted version of the manuscript.

FUNDING

This work was supported by the State Key Program of National Natural Science of China (Grant No. 82130092).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2022.921320/full#supplementary-material

Supplementary Table S1 | Patient phenotype information of the TCGA-COAD, GSE28722, GSE17538, and GSE166555 datasets.

Supplementary Table S2 | Necroptosis related genes, immunotherapy related genes and cell marker genes used in this article.

Supplementary Table S3 | CIBERSORT analysis result in the TCGA-COAD cohort.

Supplementary Table S4 | Differentially expressed genes between the NRGs-activated group and the NRGs-suppressed group.

Supplementary Table S5 | Results of GSEA of DEGs between the NRGs-activated group and the NRGs-suppressed group.

Supplementary Table S6 | OncoPredict result of the high-necroptosis score subgroup (n = 122, 28.2% of all colon cancer patients) and the low-necroptosis score subgroup (n = 311, 71.8% of all colon cancer patients) in the TCGA-COAD cohort.

Supplementary Table S7 | TIDE result of the high-necroptosis score subgroup (n = 122, 28.2% of all colon cancer patients) and the low-necroptosis score subgroup (n = 311, 71.8% of all colon cancer patients) in the TCGA-COAD cohort.

REFERENCES

Alvarez-Diaz, S., Preaudet, A., Samson, A. L., Nguyen, P. M., Fung, K. Y., Garnham, A. L., et al. (2021). Necroptosis Is Dispensable for the Development of Inflammation-Associated or Sporadic Colon Cancer in Mice. *Cell Death Differ.* 28, 1466–1476. doi:10.1038/s41418-020-00673-z

Ayers, M., Lunceford, J., Nebozhyn, M., Murphy, E., Loboda, A., Kaufman, D. R., et al. (2017). IFN-γ-related mRNA Profile Predicts Clinical Response to PD-1 Blockade. *J. Clin. investigation* 127, 2930–2940. doi:10.1172/jci91190

Cai, Z., Jitkaew, S., Zhao, J., Chiang, H.-C., Choksi, S., Liu, J., et al. (2014). Plasma Membrane Translocation of Trimerized MLKL Protein Is Required for TNF-Induced Necroptosis. *Nat. Cell Biol.* 16, 55–65. doi:10.1038/ncb2883

Cai, Z., Zhang, A., Choksi, S., Li, W., Li, T., Zhang, X.-M., et al. (2016). Activation of Cell-Surface Proteases Promotes Necroptosis, Inflammation and Cell Migration. *Cell Res.* 26, 886–900. doi:10.1038/cr.2016.87

Chen, X., Li, W., Ren, J., Huang, D., He, W.-t., Song, Y., et al. (2014). Translocation of Mixed Lineage Kinase Domain-like Protein to Plasma Membrane Leads to Necrotic Cell Death. *Cell Res.* 24, 105–121. doi:10.1038/cr.2013.171

Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., and Lin, C. Y. (2014). cytoHubba: Identifying Hub Objects and Sub-networks from Complex Interactome. *BMC Syst. Biol.* 8 (Suppl. 4), S11. doi:10.1186/1752-0509-8-S4-S11

Cho, Y., Challa, S., Moquin, D., Genga, R., Ray, T. D., Guildford, M., et al. (2009). Phosphorylation-driven Assembly of the RIP1-RIP3 Complex Regulates Programmed Necrosis and Virus-Induced Inflammation. *Cell* 137, 1112–1123. doi:10.1016/j.cell.2009.05.037

Dawson, J. C., Serrels, A., Stupack, D. G., Schlaepfer, D. D., and Frame, M. C. (2021). Targeting FAK in Anticancer Combination Therapies. *Nat. Rev. Cancer* 21, 313–324. doi:10.1038/s41568-021-00340-6

Degterev, A., Hitomi, J., Germscheid, M., Ch'en, I. L., Korkina, O., Teng, X., et al. (2008). Identification of RIP1 Kinase as a Specific Cellular Target of Necrostatins. *Nat. Chem. Biol.* 4, 313–321. doi:10.1038/nchembio.83

Degterev, A., Huang, Z., Boyce, M., Li, Y., Jagtap, P., Mizushima, N., et al. (2005). Chemical Inhibitor of Nonapoptotic Cell Death with Therapeutic Potential for Ischemic Brain Injury. *Nat. Chem. Biol.* 1, 112–119. doi:10.1038/nchembio711

Dondelinger, Y., Declercq, W., Montessuit, S., Roelandt, R., Goncalves, A., Bruggeman, I., et al. (2014). MLKL Compromises Plasma Membrane Integrity by Binding to Phosphatidylinositol Phosphates. *Cell Rep.* 7, 971–981. doi:10.1016/j.celrep.2014.04.026

Franceschi, S., Corsinovi, D., Lessi, F., Tantillo, E., Aretini, P., Menicagli, M., et al. (2018). Mitochondrial Enzyme GLUD2 Plays a Critical Role in Glioblastoma Progression. *EBioMedicine* 37, 56–67. doi:10.1016/j.ebiom.2018.10.008

Gong, Y., Fan, Z., Luo, G., Yang, C., Huang, Q., Fan, K., et al. (2019). The Role of Necroptosis in Cancer Biology and Therapy. *Mol. Cancer* 18, 100. doi:10.1186/s12943-019-1029-8

He, S., Wang, L., Miao, L., Wang, T., Du, F., Zhao, L., et al. (2009). Receptor Interacting Protein Kinase-3 Determines Cellular Necrotic Response to TNF-α. *Cell* 137, 1100–1111. doi:10.1016/j.cell.2009.05.021

Heagerty, P. J., and Zheng, Y. (2005). Survival Model Predictive Accuracy and ROC Curves. *Biometrics* 61, 92–105. doi:10.1111/j.0006-341x.2005.030814.x

Hildebrand, J. M., Tanzer, M. C., Lucet, I. S., Young, S. N., Spall, S. K., Sharma, P., et al. (2014). Activation of the Pseudokinase MLKL Unleashes the Four-Helix Bundle Domain to Induce Membrane Localization and Necroptotic Cell Death. *Proc. Natl. Acad. Sci. U.S.A.* 111, 15072–15077. doi:10.1073/pnas.1408987111

Hsu, H., Huang, J., Shu, H.-B., Baichwal, V., and Goeddel, D. V. (1996). TNF-Dependent Recruitment of the Protein Kinase RIP to the TNF Receptor-1 Signaling Complex. *Immunity* 4, 387–396. doi:10.1016/s1074-7613(00)80252-6

Jayakumar, A., and Bothwell, A. L. M. (2019). RIPK3-Induced Inflammation by I-MDSCs Promotes Intestinal Tumors. *Cancer Res.* 79, 1587–1599. doi:10.1158/0008-5472.can-18-2153

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

Khalili, A., and Ahmad, M. (2015). A Review of Cell Adhesion Studies for Biomedical and Biological Applications. *Ijms* 16, 18149–18184. doi:10.3390/ijms160818149

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, 882–883. doi:10.1093/bioinformatics/bts034

Li, J., McQuade, T., Siemer, A. B., Napetschnig, J., Moriwaki, K., Hsiao, Y.-S., et al. (2012). The RIP1/RIP3 Necrosome Forms a Functional Amyloid Signaling Complex Required for Programmed Necrosis. *Cell* 150, 339–350. doi:10.1016/j.cell.2012.06.019

Liu, Z.-Y., Zheng, M., Li, Y.-M., Fan, X.-Y., Wang, J.-C., Li, Z.-C., et al. (2019). RIP3 Promotes Colitis-Associated Colorectal Cancer by Controlling Tumor Cell Proliferation and CXCL1-Induced Immune Suppression. *Theranostics* 9, 3659–3673. doi:10.7150/thno.32126

Maeser, D., Gruener, R. F., and Huang, R. S. (2021). oncoPredict: an R Package for Predicting *In Vivo* or Cancer Patient Drug Response and Biomarkers from Cell Line Screening Data. *Brief. Bioinform* 22, bbab260. doi:10.1093/bib/bbab260

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, 1747–1756. doi:10.1101/gr.239244.118

Montero, J. C., Seoane, S., Ocaña, A., and Pandiella, A. (2011). Inhibition of SRC Family Kinases and Receptor Tyrosine Kinases by Dasatinib: Possible Combinations in Solid Tumors. *Clin. cancer Res. official J. Am. Assoc. Cancer Res.* 17, 5546–5552. doi:10.1158/1078-0432.ccr-10-2616

Moquin, D. M., McQuade, T., and Chan, F. K.-M. (2013). CYLD Deubiquitinates RIP1 in the TNFα-Induced Necrosome to Facilitate Kinase Activation and Programmed Necrosis. *PLoS One* 8, e76841. doi:10.1371/journal.pone.0076841

Moriwaki, K., Bertin, J., Gough, P. J., Orlowski, G. M., and Chan, F. K. (2015). Differential Roles of RIPK1 and RIPK3 in TNF-Induced Necroptosis and Chemotherapeutic Agent-Induced Cell Death. *Cell Death Dis.* 6, e1636. doi:10.1038/cddis.2015.16

Nehs, M. A., Lin, C.-I., Kozono, D. E., Whang, E. E., Cho, N. L., Zhu, K., et al. (2011). Necroptosis Is a Novel Mechanism of Radiation-Induced Cell Death in Anaplastic Thyroid and Adrenocortical Cancers. *Surgery* 150, 1032–1039. doi:10.1016/j.surg.2011.09.012

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust Enumeration of Cell Subsets from Tissue Expression Profiles. *Nat. Methods* 12, 453–457. doi:10.1038/nmeth.3337

Oliver Metzig, M., Fuchs, D., Tagscherer, K. E., Gröne, H.-J., Schirmacher, P., and Roth, W. (2016). Inhibition of Caspases Primes Colon Cancer Cells for 5-Fluorouracil-Induced TNF-α-dependent Necroptosis Driven by RIP1 Kinase and NF-Kb. *Oncogene* 35, 3399–3409. doi:10.1038/onc.2015.398

Powell, D. R., and Huttenlocher, A. (2016). Neutrophils in the Tumor Microenvironment. *Trends Immunol.* 37, 41–52. doi:10.1016/j.it.2015.11.008

Siegel, R. L., Miller, K. D., Goding Sauer, A., Fedewa, S. A., Butterly, L. F., Anderson, J. C., et al. (2020). Colorectal Cancer Statistics, 2020. *CA A Cancer J. Clin.* 70, 145–164. doi:10.3322/caac.21601

Sultan, H., Wu, J., Fesenkova, V. I., Fan, A. E., Addis, D., Salazar, A. M., et al. (2020). Poly-IC Enhances the Effectiveness of Cancer Immunotherapy by Promoting T Cell Tumor Infiltration. *J. Immunother. Cancer* 8, e001224. doi:10.1136/jitc-2020-001224

Sun, L., Wang, H., Wang, Z., He, S., Chen, S., Liao, D., et al. (2012). Mixed Lineage Kinase Domain-like Protein Mediates Necrosis Signaling Downstream of RIP3 Kinase. *Cell* 148, 213–227. doi:10.1016/j.cell.2011.11.031

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING Database in 2021: Customizable Protein-Protein Networks, and Functional Characterization of User-Uploaded Gene/measurement Sets. *Nucleic acids Res.* 49, D605–D612. doi:10.1093/nar/gkaa1074

Tabernero, J., Yoshino, T., Cohn, A. L., Obermannova, R., Bodoky, G., Garcia-Carbonero, R., et al. (2015). Ramucirumab versus Placebo in Combination with Second-Line FOLFIRI in Patients with Metastatic Colorectal Carcinoma that Progressed during or after First-Line Therapy with Bevacizumab, Oxaliplatin, and a Fluoropyrimidine (RAISE): a Randomised, Double-Blind, Multicentre, Phase 3 Study. *Lancet Oncol.* 16, 499–508. doi:10.1016/s1470-2045(15)70127-0

Takemura, R., Takaki, H., Okada, S., Shime, H., Akazawa, T., Oshiumi, H., et al. (2015). PolyI:C-Induced, TLR3/RIP3-dependent Necroptosis Backs up Immune Effector-Mediated Tumor Elimination *In Vivo*. *Cancer Immunol. Res.* 3, 902–914. doi:10.1158/2326-6066.cir-14-0219

Tenev, T., Bianchi, K., Darding, M., Broemer, M., Langlais, C., Wallberg, F., et al. (2011). The Ripoptosome, a Signaling Platform that Assembles in Response to Genotoxic Stress and Loss of IAPs. *Mol. cell* 43, 432–448. doi:10.1016/j.molcel.2011.06.006

Theocharis, A. D., and Karamanos, N. K. (2019). Proteoglycans Remodeling in Cancer: Underlying Molecular Mechanisms. *Matrix Biol.* 75-76, 220–259. doi:10.1016/j.matbio.2017.10.008

Uhlitz, F., Bischoff, P., Peidli, S., Sieber, A., Trinks, A., Lüthen, M., et al. (2021). Mitogen-activated Protein Kinase Activity Drives Cell Trajectories in Colorectal Cancer. *EMBO Mol. Med.* 13, e14123. doi:10.15252/emmm.202114123

Um, W., Ko, H., You, D. G., Lim, S., Kwak, G., Shim, M. K., et al. (2020). Necroptosis-Inducible Polymeric Nanobubbles for Enhanced Cancer Sonoimmunotherapy. *Adv. Mater* 32, e1907953. doi:10.1002/adma.201907953

Van Hoecke, L., Van Lint, S., Roose, K., Van Parys, A., Vandenabeele, P., Grooten, J., et al. (2018). Treatment with mRNA Coding for the Necroptosis Mediator MLKL Induces Antitumor Immunity Directed against Neo-Epitopes. *Nat. Commun.* 9, 3417. doi:10.1038/s41467-018-05979-8

Vandenabeele, P., Galluzzi, L., Vanden Berghe, T., and Kroemer, G. (2010). Molecular Mechanisms of Necroptosis: an Ordered Cellular Explosion. *Nat. Rev. Mol. Cell Biol.* 11, 700–714. doi:10.1038/nrm2970

Vitale, I., Manic, G., Coussens, L. M., Kroemer, G., and Galluzzi, L. (2019). Macrophages and Metabolism in the Tumor Microenvironment. *Cell metab.* 30, 36–50. doi:10.1016/j.cmet.2019.06.001

Walker, C., Mojares, E., and Del Río Hernández, A. (2018). Role of Extracellular Matrix in Development and Cancer Progression. *Int. J. Mol. Sci.* 19, 3028. doi:10.3390/ijms19103028

Wang, H., Sun, L., Su, L., Rizo, J., Liu, L., Wang, L.-F., et al. (2014). Mixed Lineage Kinase Domain-like Protein MLKL Causes Necrotic Membrane Disruption upon Phosphorylation by RIP3. *Mol. cell* 54, 133–146. doi:10.1016/j.molcel.2014.03.003

Wang, L., Wang, L., Shi, X., and Xu, S. (2020). Chlorpyrifos Induces the Apoptosis and Necroptosis of L8824 Cells through the ROS/PTEN/PI3K/AKT axis. *J. Hazard. Mater.* 398, 122905. doi:10.1016/j.jhazmat.2020.122905

Weinlich, R., Oberst, A., Beere, H. M., and Green, D. R. (2017). Necroptosis in Development, Inflammation and Disease. *Nat. Rev. Mol. Cell Biol.* 18, 127–136. doi:10.1038/nrm.2016.149

Xu, T., Le, T. D., Liu, L., Su, N., Wang, R., Sun, B., et al. (2017). CancerSubtypes: an R/Bioconductor Package for Molecular Cancer Subtype Identification, Validation and Visualization. *Bioinforma. Oxf. Engl.* 33, 3131–3133. doi:10.1093/bioinformatics/btx378

Yaeger, R., Chatila, W. K., Lipsyc, M. D., Hechtman, J. F., Cercek, A., Sanchez-Vega, F., et al. (2018). Clinical Sequencing Defines the Genomic Landscape of Metastatic Colorectal Cancer. *Cancer Cell* 33, 125–136. e123. doi:10.1016/j.ccell.2017.12.004

Zeng, F., Chen, X., Cui, W., Wen, W., Lu, F., Sun, X., et al. (2018). RIPK1 Binds MCU to Mediate Induction of Mitochondrial Ca2+ Uptake and Promotes Colorectal Oncogenesis. *Cancer Res.* 78, 2876–2885. doi:10.1158/0008-5472.can-17-3082

Zhang, D.-W., Shao, J., Lin, J., Zhang, N., Lu, B.-J., Lin, S.-C., et al. (2009). RIP3, an Energy Metabolism Regulator that Switches TNF-Induced Cell Death from Apoptosis to Necrosis. *Science* 325, 332–336. doi:10.1126/science.1172308

Zhang, Y., Su, S. S., Zhao, S., Yang, Z., Zhong, C.-Q., Chen, X., et al. (2017). RIP1 Autophosphorylation Is Promoted by Mitochondrial ROS and Is Essential for RIP3 Recruitment into Necrosome. *Nat. Commun.* 8, 14329. doi:10.1038/ncomms14329

Zhao, Q., Cheng, X., Guo, J., Bi, Y., Kuang, L., Ren, J., et al. (2021). MLKL Inhibits Intestinal Tumorigenesis by Suppressing STAT3 Signaling Pathway. *Int. J. Biol. Sci.* 17, 869–881. doi:10.7150/ijbs.56152

Zhao, Q., Yu, X., Li, M., Liu, Y., Han, Y., Zhang, X., et al. (2019). MLKL Attenuates Colon Inflammation and Colitis-Tumorigenesis via Suppression of Inflammatory Responses. *Cancer Lett.* 459, 100–111. doi:10.1016/j.canlet.2019.05.034

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: 08 July 2022
doi: 10.3389/fcell.2022.930933

AP3S1 is a Novel Prognostic Biomarker and Correlated With an Immunosuppressive Tumor Microenvironment in Pan-Cancer

Gujie Wu[1,2†], *Mianxiong Chen*[1†], *Hefei Ren*[3†], *Xinyu Sha*[2], *Min He*[2], *Kuan Ren*[2], *Juntao Qi*[1*] *and Feng Lin*[1*]

[1]*Department of Urology, Shenzhen Traditional Chinese Medicine Hospital, Guangzhou University of Chinese Medicine, Shenzhen, China,* [2]*Research Center of Clinical Medicine, Affiliated Hospital of Nantong University, Nantong, China,* [3]*Department of Laboratory Medicine, Changzheng Hospital, Naval Medical University, Shanghai, China*

OPEN ACCESS

Edited by:
D. P Kreil,
Boku University Vienna, Austria

Reviewed by:
Eswari Dodagatta-Marri,
University of California, San Francisco, United States
Maulik Vyas,
Massachusetts General Hospital and Harvard Medical School, United States

****Correspondence:***
Juntao Qi
juntaoqi@gmail.com
Feng Lin
lf0237@gzucm.edu.cn

†These authors have contributed equally to this work

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

Received: *28 April 2022*
Accepted: *06 June 2022*
Published: *08 July 2022*

Citation:
Wu G, Chen M, Ren H, Sha X, He M, Ren K, Qi J and Lin F (2022) AP3S1 is a Novel Prognostic Biomarker and Correlated With an Immunosuppressive Tumor Microenvironment in Pan-Cancer. Front. Cell Dev. Biol. 10:930933. doi: 10.3389/fcell.2022.930933

Background: Adaptor-related protein complex 3, sigma one subunit (AP3S1) is one of the encoding subunits of the adaptor complex AP-3. However, its role in various tumor types and relationship with the tumor immune microenvironment (TIME) remains unclear.

Methods: AP3S1 expression was analyzed using datasets from The Cancer Genome Atlas, Genotype-Tissue Expression, UALCAN, and HPA databases. Then, we performed a systematic analysis of the genetic alterations, clinical features, and prognostic value of AP3S1 in pan-cancer. Gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA) were used to identify the signaling pathways associated with AP3S1. The correlation between immune cell infiltration and AP3S1 expression was analyzed using immune cell infiltration data from the ImmuCellAI, TIMER2, and a previous study. Finally, we analyzed the association of AP3S1 with tumor mutational burden (TMB), microsatellite instability (MSI), and immune-related genes.

Results: We found AP3S1 overexpression in most tumors and a significant association with low survival rates. GSEA and GSVA results show that AP3S1 is involved in tumor progression and associated with immune pathways in different tumor types. We also found that AP3S1 expression was positively correlated with the level of infiltration of immunosuppressive cells (tumor-associated macrophages, cancer-associated fibroblasts, Tregs) and negatively correlated with immune killer cells, including NK cells and $CD8^+$ T cells, in pan-cancer. The expression of AP3S1 could affect TMB and MSI in various cancers. In addition, AP3S1 was positively correlated with most immunosuppressive genes, including PD-1, PD-L1, CTLA4, LAG3 and TIGIT in most cancer types.

Conclusion: Our study reveals that AP3S1 is a potential pan-cancer oncogene and plays an essential role in tumorigenesis and cancer immunity. Elevated expression of AP3S1 indicates an immunosuppressive microenvironment and can be used as a potential prognostic biomarker and a target for immunotherapy.

Keywords: AP3S1, pan-cancer, prognostic biomarker, immunosuppressive microenvironment, immunotherapy

INTRODUCTION

Cancer is the leading cause of death globally and seriously impacts people's survival (Siegel et al., 2020). Although tremendous progress has been made in various treatments for tumours in recent years, the prognosis and survival rates for many types of cancers remain unsatisfactory (Kandoth et al., 2013; Zhang et al., 2020; He et al., 2021). Growing evidence suggests that the immunomodulatory mechanisms of tumors have a notable impact on tumor growth, progression, tumor microenvironment and patient prognostic outcome (Lei et al., 2020; Li et al., 2016; Mittal et al., 2018). Therefore, given the promising potential of the tumor immune microenvironment (TIME) in cancer therapy, there is an urgent need to explore the underlying mechanisms of tumors and to identify novel critical biomarkers for tumor patients.

AP3S1 encodes a subunit of the AP-3 complex, an adapter-related complex that is associated with the Golgi apparatus and more peripheral structures (Simpson et al., 1997). To date, only a few studies have reported a role for AP3S1 in the development of tumors. A recent study identified AP3S1 as a tumor driver gene after analyzing 1,145 samples from patients with esophageal squamous cell carcinoma (Zou et al., 2021). Additionally, AP3S1 expression is associated with the progression of rectal and cervical tumors (Petrenko et al., 2006; Nome et al., 2013). Despite these findings, additional studies are needed to reveal more about the role of AP3S1 in tumor development and progression. Our study explores the potential pan-cancer role of AP3S1 through an extensive analysis and its function in the tumour microenvironment (TME).

In this study, we used multi-omics data from 33 cancers to comprehensively analyse the role of AP3S1, including AP3S1 expression, clinical characteristics, prognostic role, DNA methylation, copy number alteration (CNA), and mutational status. GSEA and GSVA results show that AP3S1 is involved in tumor progression and associated with immune pathways in various tumors. Finally, we assessed the correlation between AP3S1 expression, immune cell infiltration levels and immune-related genes. Elevated AP3S1 expression predicts a tumor immunosuppressive microenvironment. Our findings reveal an essential role for AP3S1 in the tumour immunosuppressive microenvironment and AP3S1 could act a potential prognostic and immunotherapeutic biomarker.

MATERIALS AND METHODS

Data Source

Our RNAseq data and clinical information data were derived from The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx). These data are available for download from the UCSC Xena (xenabrowser.net/datapages/) database. cBioPortal database (www.cbioportal.org/) provided us with AP3S1 methylation and CNA data. The Human Protein Atlas database (www.proteinatlas.org/)

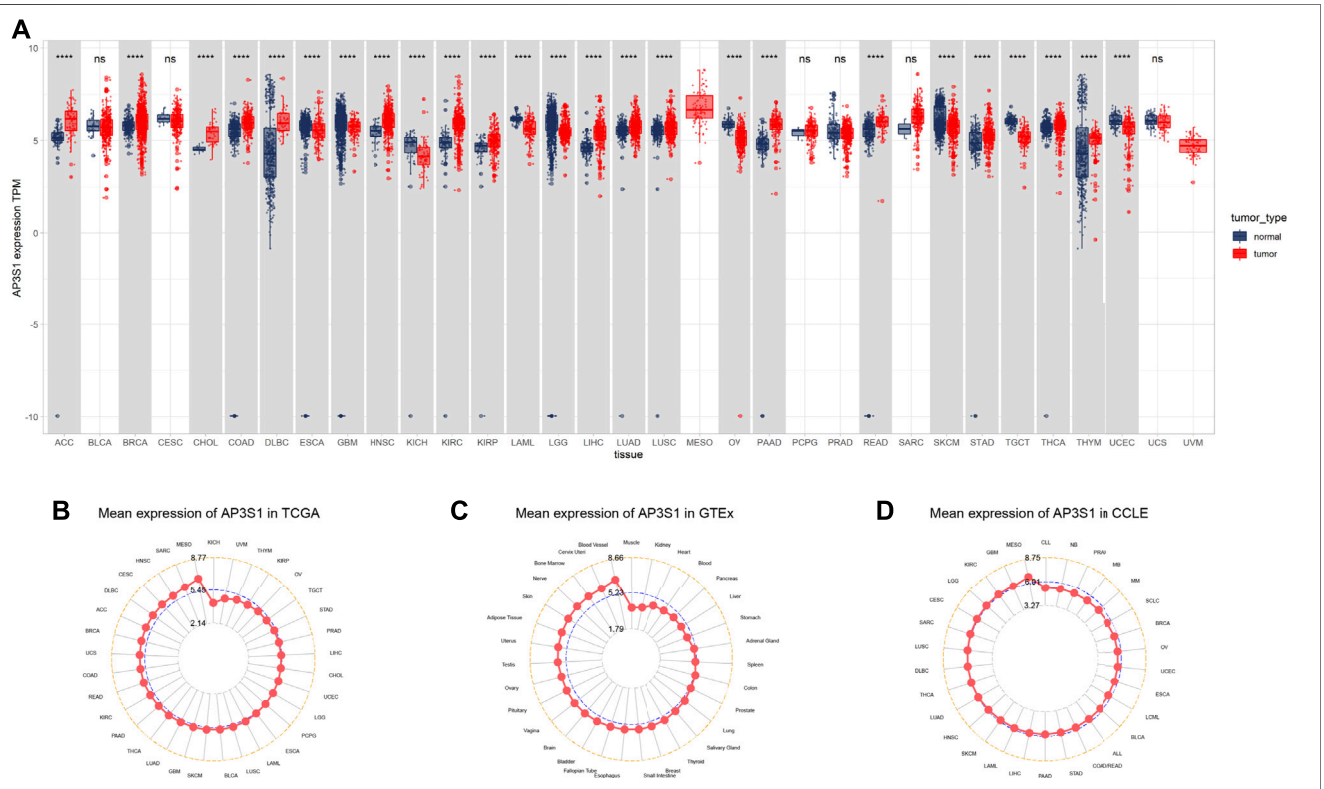


FIGURE 1 | Expression of AP3S1 in pan-cancer. **(A)** pan-cancer expression of AP3S1. **(B)** AP3S1 expression in tumor tissues from TCGA cohort. **(C)** AP3S1 expression in normal tissues from GTEx cohort. **(D)** AP3S1 expression in cancer cell lines from the CCLE cohort.*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

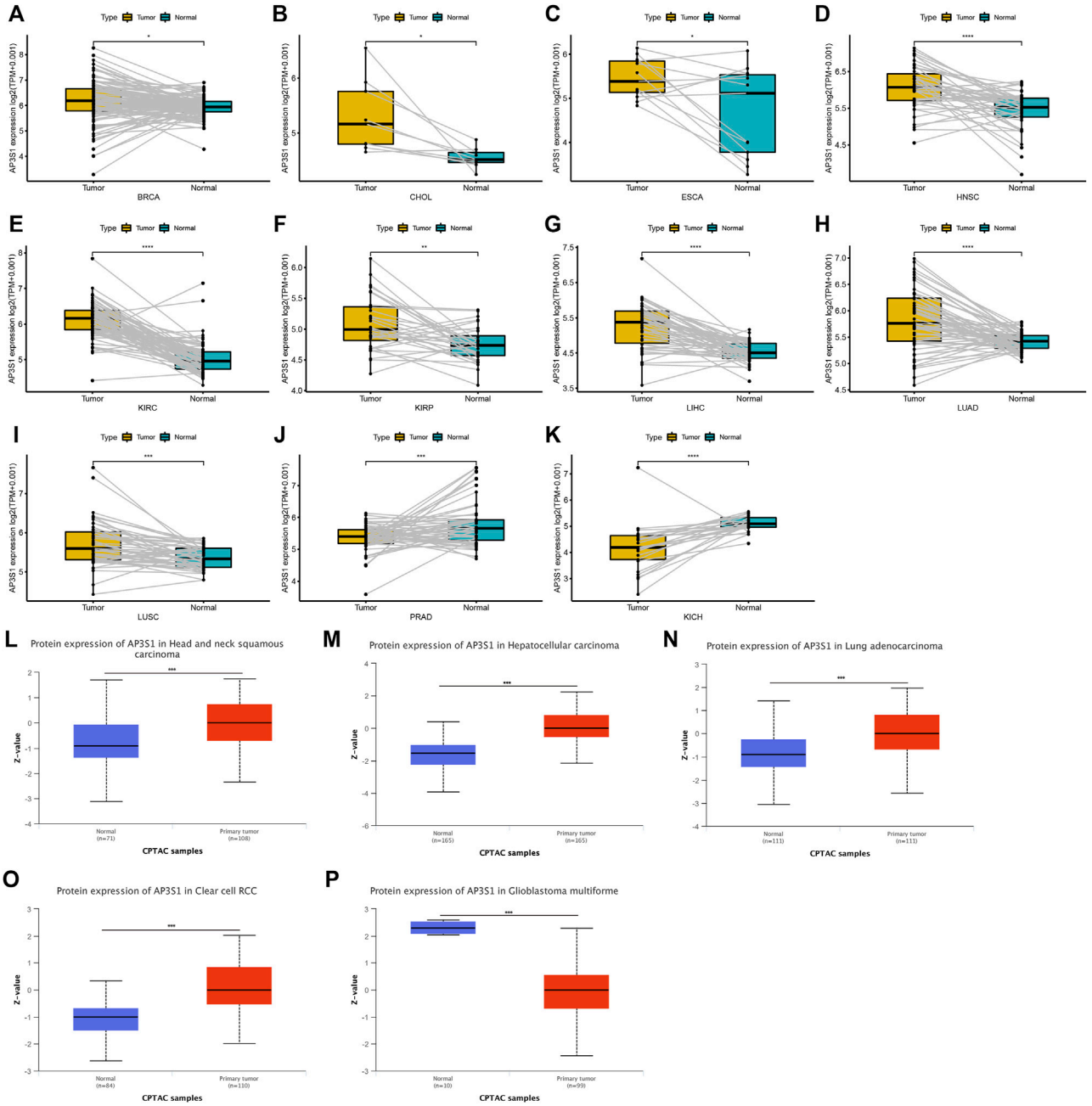


FIGURE 2 | Expression of AP3S1 in paired tumor and adjacent normal tissues. **(A–K)** AP3S1 expression in paired tumor and adjacent normal tissues from TCGA in indicated tumor types. **(L–P)** The protein level of AP3S1 in indicated tumor types using Ualcan database*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

was used to explore the protein level of AP3S1 in human tumors and normal tissues.

Prognostic Analysis

Kaplan-Meier analysis was performed to evaluate the overall survival (OS) of patients from TCGA cohort. Univariate Cox regression analyses were conducted to assess the significance of AP3S1 in predicting OS, disease-specific survival (DSS), the disease-free interval (DFI), and the progression-free interval (PFI) in pan-cancer.

GSEA and GSVA

GSEA analysis was used to evaluate AP3S1 and all associated genes in pan-cancer. Pearson correlation coefficients and the R package "clusterProfiler" were used in the process. GSVA was performed by the "ssGSEA" algorithm and the data were obtained

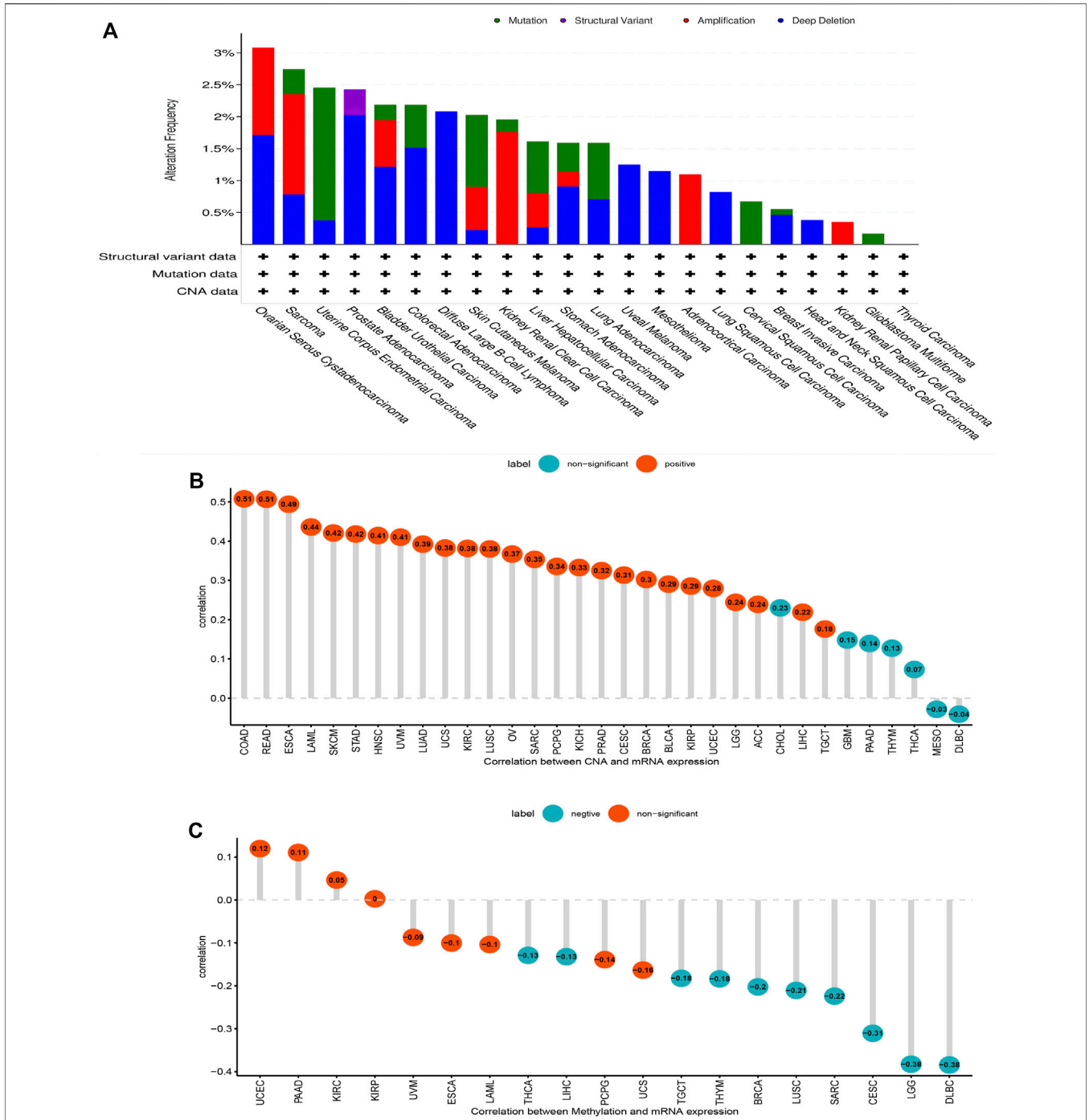


FIGURE 3 | Gene alteration of AP3S1. **(A)** The mutation and CNA status of AP3S1 in TCGA-pan-cancer. **(B)** The correlation between AP3S1 expression and CNA. **(C)** The correlation between AP3S1 expression and DNA methylation.

from the MSigDB database by downloading the landmark pathways.

Immune Infiltration Analysis

The R package "ESTIMATE" was used to calculate the stromal score, immune score, and tumor purity score of each patient in TCGA cohort. The association between AP3S1 expression and these scores was analyzed. The TME-related pathways were obtained and pathway scores were calculated according to the published paper (Zeng et al., 2019). The immune cell infiltration data were downloaded from Immune Cell Abundance Identifier (ImmuCellAI) database (http://bioinfo.life.hust.edu.cn/ImmuCellAI#!/) and TIMER2 database (http://timer.cistrome.org/).

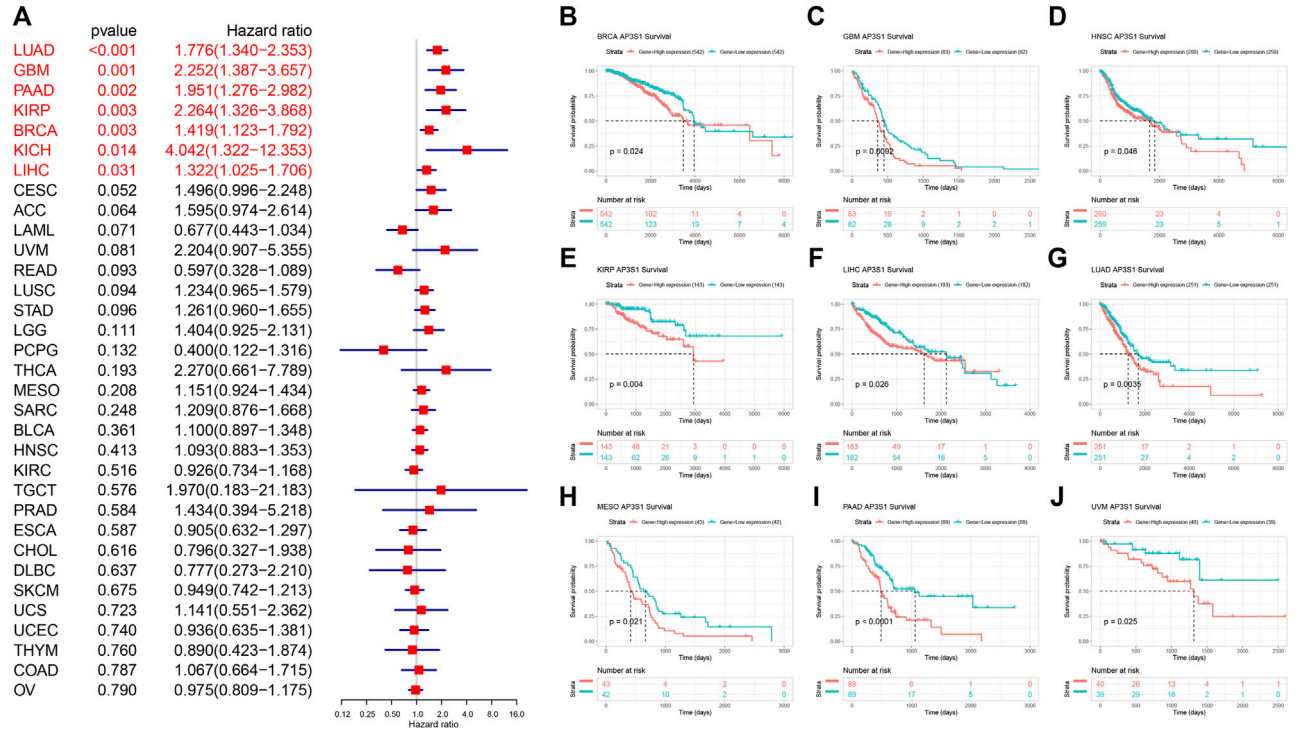


FIGURE 4 | Relationship between AP3S1 level and OS. **(A)** The univariate Cox regression OS analysis of AP3S1 in TCGA pan-cancer. Red color represents significant results ($p < 0.05$). **(B–J)** Kaplan-Meier curves showing OS in pan-cancer.

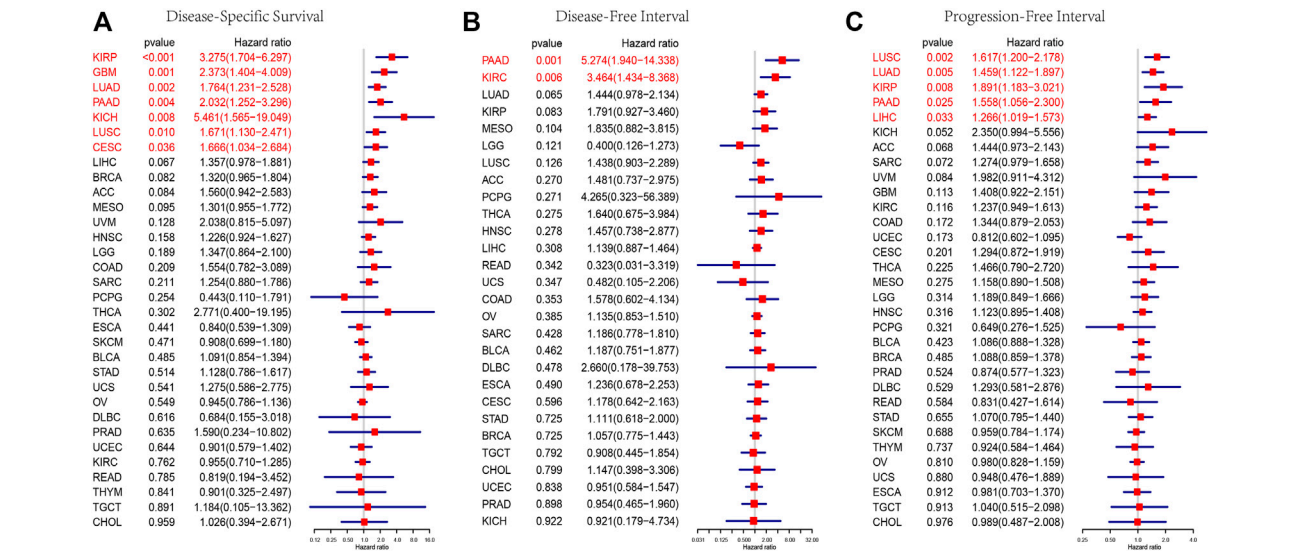


FIGURE 5 | Prognostic value of AP3S1. Forest plots showing results of Univariate Cox Regression analysis for **(A)** DSS, **(B)** DFI, and **(C)** PFI. Red color represents significant results ($p < 0.05$).

Statistical Analysis

All data used in this article are presented as the mean ± standard deviation (SD). The correlation analysis between the two variables used Spearman's or Pearson's test. Statistical analysis was performed using R v4.1.1. A p-value < 0.05 was considered statistically significant.

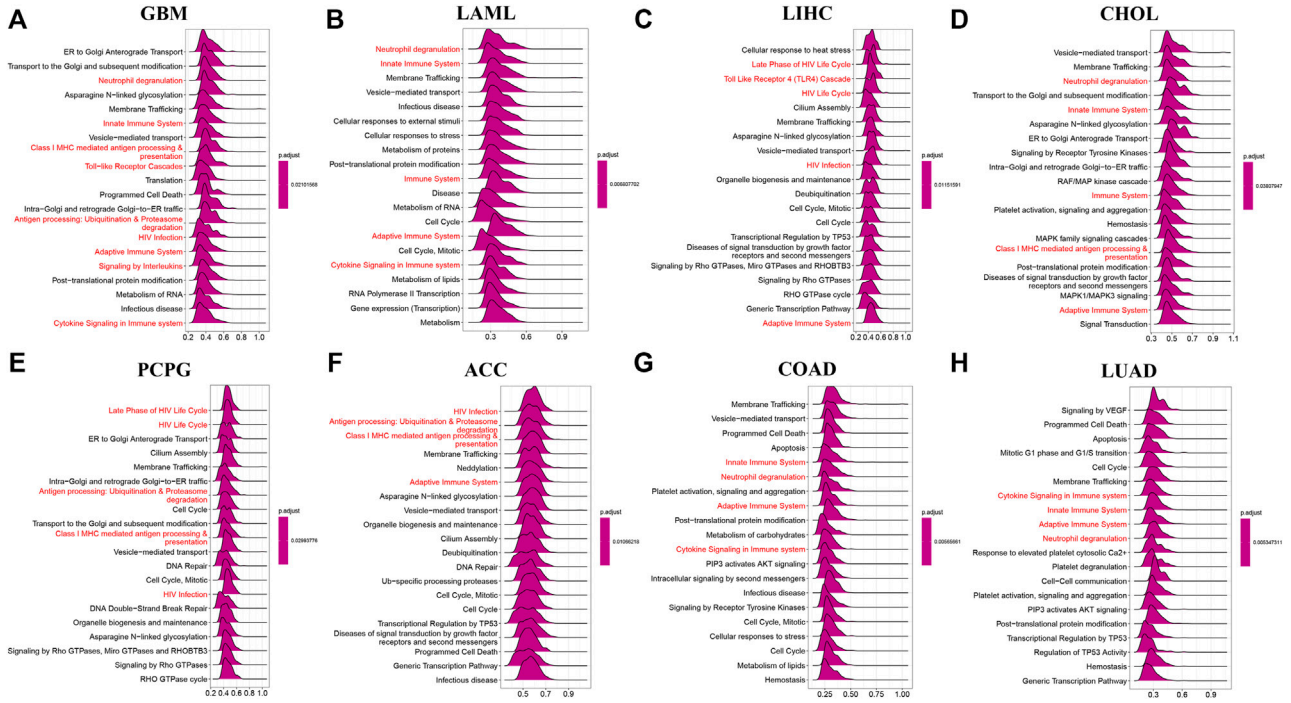


FIGURE 6 | GSEA of AP3S1 in TCGA pan-cancer. **(A–H)** The top 8 significant pathways of AP3S1 GSEA results across the indicated tumor types. Red color represents immune-related pathways.

RESULTS

Expression of AP3S1 in Pan-Cancer

Firstly, we assessed AP3S1 expression using TCGA and GTEx data. We found that AP3S1 was overexpressed in 16 of 33 cancer types, including Adrenocortical carcinoma (ACC), Breast invasive carcinoma (BRCA), Cholangiocarcinoma (CHOL), Colon adenocarcinoma (COAD), Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (DLBC), Head and Neck squamous cell carcinoma (HNSC), Kidney renal clear cell carcinoma (KIRC), Kidney renal papillary cell carcinoma (KIRP), Liver hepatocellular carcinoma (LIHC), Lung adenocarcinoma (LUAD), Lung squamous cell carcinoma (LUSC), Pancreatic adenocarcinoma (PAAD), Rectum adenocarcinoma (READ), Stomach adenocarcinoma (STAD), Thyroid carcinoma (THCA) and Thymoma (THYM). Moreover, the expression of AP3S1 was found to be significantly reduced in Esophageal carcinoma (ESCA), Glioblastoma multiforme (GBM), Kidney Chromophobe (KICH), Acute Myeloid Leukemia (LAML), Lower Grade Glioma (LGG), Ovarian serous cystadenocarcinoma (OV), Skin Cutaneous Melanoma (SKCM), Testicular Germ Cell Tumor (TGCT), Uterine Corpus Endometrial Carcinoma (UCEC) (**Figure 1A** and **Supplementary Table S1**). In order to compare AP3S1 expression in tumor tissues, our findings revealed that AP3S1 expression was highest in MESO and lowest in KICH (**Figure 1B**). In normal tissues from GTEx database, the results revealed that AP3S1 expression was the highest in blood vessel tissues and the lowest in muscles (**Figure 1C**). As for tumor cell lines, we proved that AP3S1 expression was also the highest in MESO cell lines using data from CCLE database (**Figure 1D**).

AP3S1 was also observed to be overexpressed in 10 cancers such as BRCA, CHOL, ESCA, HNSC, KIRC, KIRP, LIHC, LUAD, and LUSC by comparing tumors with adjacent normal tissues in the TCGA cohort (**Figures 2A–I**), with lower expression only in PRAD and KICH (**Figures 2J,K**). After analysis of the UALCAN database, AP3S1 protein was upregulated in tumor tissues of HNSC, LIHC, LUAD, and KIRC and downregulated in GBM compared to normal tissues (**Figures 2L–P**). The HPA database of protein levels of AP3S1 expression between normal and tumor tissues also confirms similar results (**Supplementary Figure S1**). Additionally, we investigated AP3S1 expression at various tumor stages. The results demonstrated that AP3S1 expression was elevated in the relatively worse tumor stages in ACC, CESC, HNSC, KIRP, LUAD, PAAD, and THCA (**Supplementary Figure S2**).

Gene Alteration of AP3S1 in Pan-Cancer

Gene mutations, DNA methylation, and CNA are closely associated with tumor development and progression. Our analysis of AP3S1 gene alterations using cBioPortal showed that the highest alteration frequency of AP3S1 appears for patients in uterine corpus endometrial carcinoma, with "amplification and deep deletion" as the primary types (**Figure 3A**). By correlation analysis between AP3S1 and CNA, we found that AP3S1 expression was positively correlated with

FIGURE 7 | GSVA of AP3S1 in TCGA pan-cancer. GSVA results of 50 hallmark pathways from the MSigDB.

CNA in most tumor types (**Figure 3B**). DNA methylation as an epigenetic modification usually leads to the silencing or inactivation of tumor suppressor genes, which results in cancer development and proliferation. Therefore, we further calculated the correlation levels between AP3S1 and promoter methylation. Gene expression significantly correlated with the methylation status in 19 tumors, with the highest negative correlations in LGG and DLBL (**Figure 3C**).

Prognostic Role of AP3S1

To further explore the impact of AP3S1 expression in pan-cancer on patient prognosis, we performed survival analysis of AP3S1

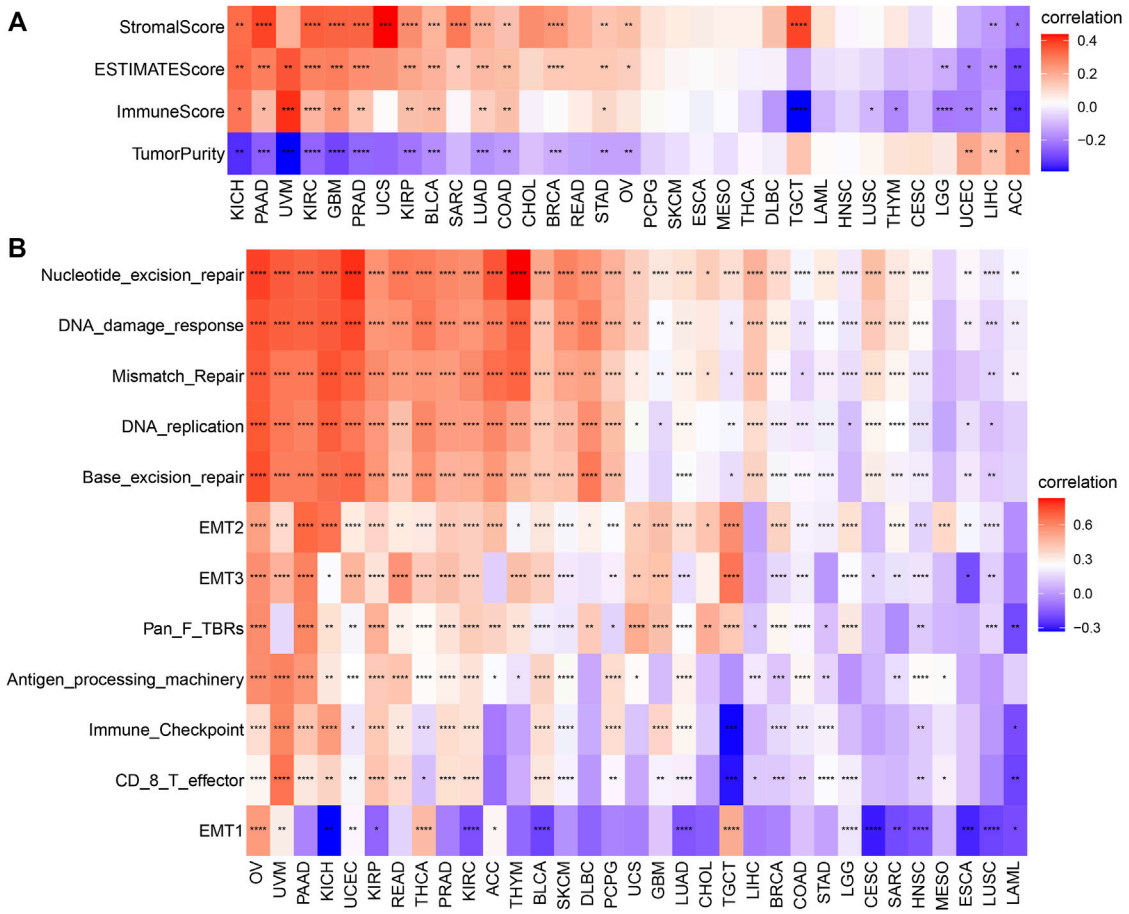


FIGURE 8 | The relationship between AP3S1 and the regulation of the tumour microenvironment. **(A)** Heatmap represents the correlation between AP3S1 expression and TME scores in pan-cancer. **(B)** The relationship between AP3S1 and the tumour microenvironment. Red represents positive correlation, blue represents negative correlation, and the darker the color, the stronger the correlation. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

expression and patient survival data using univariate Cox regression analysis (UniCox) and Kaplan-Meier methods. The analysis revealed that AP3S1 was a risk factor for LUAD, GBM, PAAD, KIRP, BRCA, KICH and LIHC (**Figure 4A**). Kaplan-Meier OS analysis demonstrated that elevated AP3S1 expression predicts BRCA, GBM, HNSC, KIRP, LIHC, LUAD, MESO, PAAD and UVM patients with poor overall survival (OS) (**Figures 4B–J**). Disease-specific survival (DSS) analysis showed AP3S1 as a risk factor in KIRP, GBM, LUAD, PAAD, KICH, LUSC and CESC (**Figure 5A**). Moreover, a high AP3S1 expression predicted a shorter disease-free interval (DFI) in patients with PAAD and KIRC (**Figure 5B**). At the same time, a high AP3S1 expression predicted a worse progression-free interval (PFI) in patients with LUSC, LUAD, KIRP, PAAD and LIHC (**Figure 5C**).

GSEA and GSVA of AP3S1

In order to elucidate the potential biological pathways that AP3S1 might regulate, we performed a GSEA algorithm analysis using the "clusterprofiler" in pan-cancers and then selected 8 tumors with similar results (**Figures 6A–H**). We found that AP3S1 participates in immune regulation-related pathways in pan-cancer, especially for the adaptive and innate immune systems, neutrophil degranulation, HIV infection, cytokine signalling in the immune system, antigen processing and Toll-like receptor cascades. In addition, cycle-related pathways (e.g. "cell cycle," "ER to Golgi anterograde transport" and "vesicle-mediated transport") were closely associated with AP3S1. These results indicated that AP3S1 plays a significant role in tumor development and tumor immunity.

To analyse the potential pathways affected by the AP3S1 expression, we performed a GSVA based on 50 HALLMARK pathways. The association between the AP3S1 expression and GSVA scores in pan-cancer is shown in **Figure 7**. We observed that the AP3S1 overexpression was positively associated with many malignant pathways in pan-cancer, such as IL6 JAK STAT3 signaling, IL2 STAT5 signaling, TGF BETA signaling, interferon gamma response, and interferon alpha response. All these pathways were also closely associated with TIME.

Immune Cell Infiltration Analysis

Subsequently, we further assessed the association between AP3S1 expression and stromal and immune scores using the "ESTIMATE"

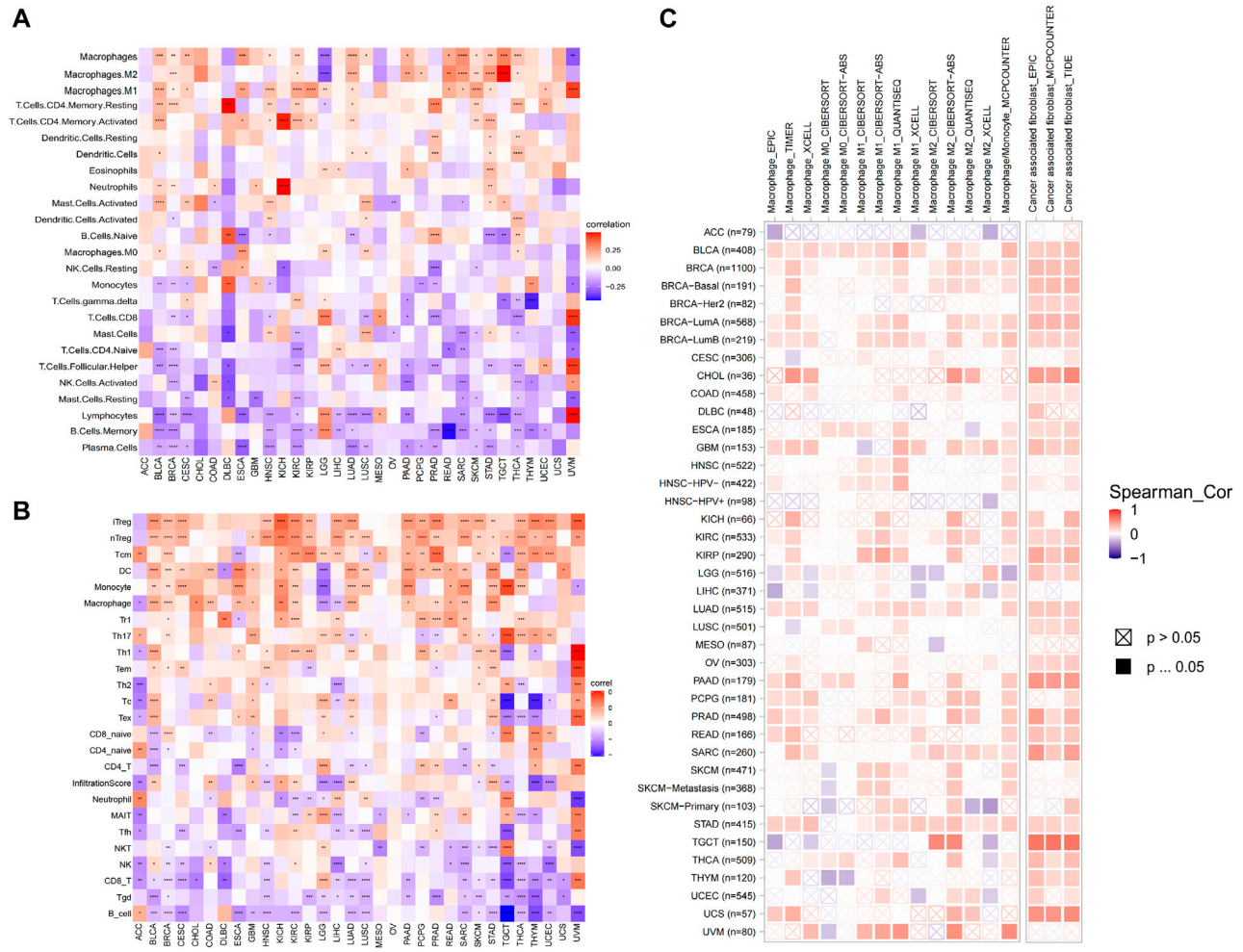


FIGURE 9 | The relationship between AP3S1 and the immune cell infiltration. **(A)** Correlation between AP3S1 and different immune cells from previous study. **(B)** Correlation between AP3S1 expression and different immune cells from ImmuCellAI database. **(C)** Correlation between AP3S1 expression and TAMs/CAFs from TIMER2 database. Red represents positive correlation, blue represents negative correlation, and the darker the color, the stronger the correlation. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

algorithm (**Figure 8A**). The results showed that AP3S1 expression was able to influence immune scores, stromal scores and ESTIMATE scores in most tumors. To validate this finding, we obtained further data and calculated the TME-related pathways involved in AP3S1, including immune-related pathways, stroma-related pathways and DNA repair-related pathways, based on published papers. The results also showed that AP3S1 expression was closely associated with immune-related pathways, including immune checkpoints, antigen processing mechanisms and CD8 T effectors in pan-cancer (**Figure 8B**).

Dysregulation of immune cell infiltration during tumor development and progression can lead to evasion of immune surveillance by tumor cells. Therefore, we used three different methods to assess the relevance of AP3S1 levels to immune cell infiltration in pan-cancer. By analyzing the association between AP3S1 expression and immune cell infiltration using a previous work, we found that AP3S1 expression was positively correlated with infiltration of tumor-associated macrophages (TAMs) and negatively correlated with immune killer cells, such as NK cells and $CD8^{+}$ T cells, in most cancers (**Figure 9A**). Next, we analyzed the relationship between AP3S1 expression and immune cell infiltration using immune data from the ImmuCellAI and TIMER2 databases, respectively (**Figures 9B,C**). The clustering heat map showed a positive correlation between AP3S1 and TAMs, which was consistent with the previous results. In addition, we also found that AP3S1 expression was positively correlated with cancer-associated fibroblasts(CAFs) and Tregs. These results revealed that patients with high AP3S1 expression might be in a relatively immunosuppressive microenvironment.

Immune-Related Genes Analyses

To further explore the relationship between AP3S1 and the immunosuppressive microenvironment, we also performed a correlation analysis between immune-related genes and AP3S1.

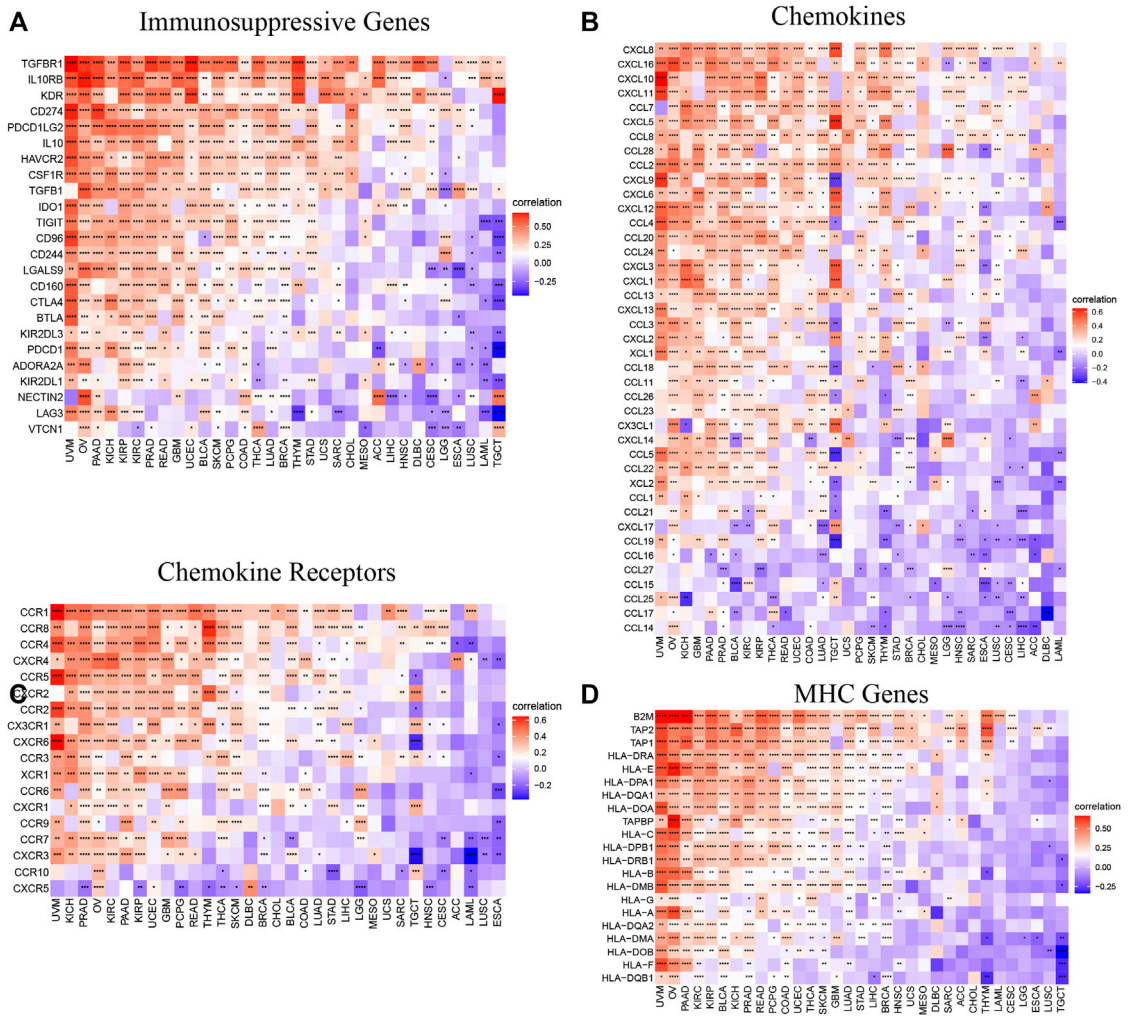


FIGURE 10 | Relationship between AP3S1 expression and that of immune-related genes. **(A)** Immunosuppressive genes. **(B)** Chemokines. **(D)** Chemokine receptors. **(C)** MHC genes. Red represents positive correlation, blue represents negative correlation, and the darker the color, the stronger the correlation. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

AP3S1 expression was positively correlated with most immunosuppressive genes in pan-cancer, such as CD274 (PD-L1), PDCD1 (PD-1), CTLA4, LAG3 and TIGIT (**Figure 10A**). We also found that AP3S1 expression significantly and positively correlated with TGFB1 and IL-10 expression in pan-cancer, which in turn were significantly correlated with TAMs/CAFs, and we speculate that this may be a potential mechanism by which AP3S1 affects infiltration of TAMs/CAFs. Finally, we further demonstrated that AP3S1 was closely associated with immunomodulatory genes, including chemokines (**Figure 10B**), chemokine receptors (**Figure 10C**) and MHC genes (**Figure 10D**). Among them, immune-related genes such as CCL2, CCR2, CXCR4, and CCR5 play a role in the recruitment of TAMs in tumor development and progression. TAMs are also able to mediate the immunosuppressive activity of T cells by releasing immune-related genes such as CCL3, CCL4, CCL5, and CCL22. These results suggest that AP3S1 plays an essential role in the process of immunosuppressive microenvironment.

Tumor Mutational Burden and Microsatellite Instability Analysis

Tumor mutation burden (TMB) or microsatellite instability (MSI) can predict immunotherapeutic response in different tumor types. Patients with tumors with high TMB or MSI tend to have higher sensitivity to immunotherapy. Therefore, we assessed their relationship with AP3S1 expression in pan-cancer. The results showed a clear positive correlation between AP3S1 expression and TMB in four cancers, including STAD, LAML, COAD and SKCM (**Figure 11A**). Similarly, AP3S1 expression was correlated with MSI in 10 cancers, positively correlated with READ, STAD, TGCT, UCEC, COAD, SARC, LUSC, and LGG, and negatively correlated with LUAD and

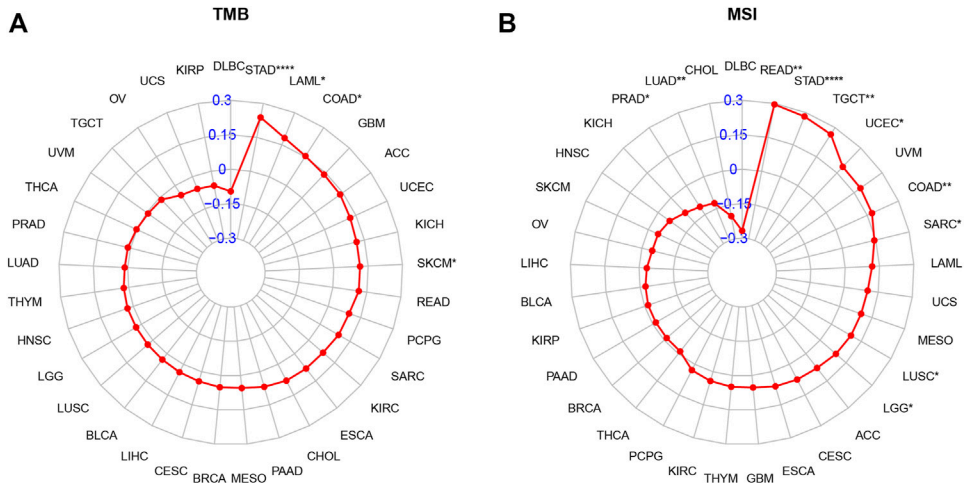


FIGURE 11 | Relationship of AP3S1 expression and tumor mutational burden (TMB), microsatellite instability (MSI). **(A)** Radar map illustrating the relationship between AP3S1 expression and TMB. **(B)** Radar map illustrating the relationship between AP3S1 expression and MSI. The red lines represent correlation coefficients, and blue values represent ranges. Spearman correlation test, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

PRAD (**Figure 11B**). These results suggest a potential role for AP3S1 in immunotherapy.

DISCUSSION

AP3S1 as one of the subunits encoding the adaptor complex AP-3 facilitates the budding of vesicles from the Golgi membrane and may be directly involved in trafficking to lysosomes (Simpson et al., 1997). Although AP3S1 plays an vital role in biological functions, its role in tumors has been neglected in the past. Only a few studies have found a correlation between AP3S1 and tumor progression (Petrenko et al., 2006; Nome et al., 2013; Zou et al., 2021). Here, we extensively analyzed the expression profile and prognostic significance of AP3S1 and explored its potential role in tumor immunology.

Tumor microenvironment (TME) plays a crucial role in tumor development and metastasis (da Silva et al., 2021). A increasing number of evidences support the clinicopathological significance of TME in predicting the survival status and treatment outcome of tumor patients (Binnewies et al., 2018; Thorsson et al., 2018). Recent studies have shown that tumor cells can evade immune surveillance through various mechanisms, especially TAMs and CAFs are remodeled directly or indirectly by tumor cells, thus inhibiting the cytotoxicity of anti-tumor immune cells and subsequently exerting immunosuppressive and tumor promoting effects (Gajewski et al., 2013; Ngambenjawong et al., 2017; Armitage et al., 2021; Cendrowicz et al., 2021). Additionally, tumor cells can also evade immune responses and immunotherapy by utilizing immune checkpoint genes, such as PD-1, PD-L1, and CTLA-4 (Topalian et al., 2012; Barbee et al., 2015).

In our study, we first assessed the expression of AP3S1. The results showed that compared to paracancerous and normal tissues, AP3S1 gene mRNA was highly expressed in 16 types of cancers, including ACC, BRCA, CHOL, COAD, DLBC, HNSC, KIRC, KIRP, LIHC, LUAD, LUSC, PAAD, READ, STAD, THCA, and THYM, whereas low expression was observed in ESCA, GBM, KICH, LAML, LGG, OV, SKCM, TGCT, and UCEC. In addition, gene mutations, DNA methylation, and CNA had important effects on AP3S1 expression in pan-cancer. To determine the prognostic role of AP3S1, we performed UniCox and Kaplan-Meier survival analyses on the TCGA cohort. The UniCox results revealed that AP3S1 was a risk factor for LUAD, GBM, PAAD, KIRP, BRCA, KICH, and LIHC. Kaplan-Meier OS analysis demonstrated that elevated AP3S1 expression predicts that BRCA, GBM, HNSC, KIRP, LIHC LUAD, MESO, PAAD, and UVM patients with a poor OS. These findings suggest that AP3S1 is a potential biomarker to predict the prognosis of tumor patients.

To explore the significance of AP3S1 in cancer immunity, we found that AP3S1 has a broad immunomodulatory function in pan-cancer by using GSEA and GSVA analysis. The correlation of AP3S1 levels with infiltrating immune cells in patients was also assessed. Our results show that AP3S1 expression positively correlated with the level of infiltration of immunosuppressive cells in pan-cancer, such as TAMs and CAFs. In contrast, in most cancers, AP3S1 expression was negatively correlated with immune killer cells, such as NK cells and $CD8^{+}$ T cells. Not only that, we also performed a correlation analysis between immune-related genes and AP3S1. AP3S1 expression was positively correlated with most immunosuppressive genes in pan-cancer, such as the common CD274 (PD-L1), PDCD1 (PD-1), CTLA4, LAG3, and TIGIT. In addition, we found that AP3S1 expression was positively correlated with immunomodulatory genes, including MHC genes, chemokines and chemokine receptors. TAMs are critical players in tumor progression, metastasis and recurrence. They are the most abundant tumor-infiltrating immune cell population in TME. TME is usually classified into two functionally distinct subtypes,

the classically activated M1 and the alternatively activated M2 subtypes (Jayasingam et al., 2019). Upon stimulation by a TLR, microbial substrates, IFNγ and CSF2, M1-TAMs secrete cytokines such as IL6, IL12, IL23 and TNFα and express specific M1 markers like MHCII molecules, CD68, CD80, CD86, iNOS, and pSTAT. However, M2-TAMs are the main urban component of TAMs. Upon stimulation by IL4, IL13, IL10, TGFβ and chemokines, M2-TAMs secrete IL4, IL13, IL10, TGFβ, and PGE2 (Cortez-Retamozo et al., 2012; Movahedi and Van Ginderachter, 2016; Dehne et al., 2017). M2-TAMs are able to participate in inflammatory regression and suppress the cytotoxicity of antitumor immune cells, thus allowing tumor cells to evade immune surveillance (Chen et al., 2019; Badmann et al., 2020). The decrease in immune killer cells such as NK cells and CD8 T cells may also increase the ability of tumors to evade immune attack. In recent years, inhibition of TAMs recruitment or retention at primary tumour and metastatic sites has emerged as a new targeted therapeutic approach (Argyle and Kitamura, 2018). Among them, immune-related genes such as CCL2, CCR2, CXCR4 and CCR5 play a vital role in the recruitment of TAMs in tumorigenesis and progression. These genes can be targeted to inhibit tumor growth by suppressing immunosuppressive mechanisms (Pienta et al., 2013; Bonapace et al., 2014; Peña et al., 2015). In addition, TAMs are also able to mediate the immunosuppressive activity of T cells by releasing immune-related genes such as IL10, TGFβ, CCL3, CCL4, CCL5, and CCL22, and by inhibiting genes encoding granzymes, perforins and cytotoxins (Chen et al., 2003; Thomas and Massagué, 2005; Kim et al., 2011). TMB and MSI can be used as predictors of the efficacy of immune checkpoint inhibitors (Samstein et al., 2019). In this study, we demonstrated that AP3S1 expression was associated with TMB in four cancer types and MSI in ten cancer types. These findings implied that high expression of AP3S1 is associated with an immunosuppressive microenvironment in pan-cancer, providing a potential drug target for tumor immunotherapy.

However, we must acknowledge some limitations in the current study. First, all analyses were performed on data from public databases only, and all samples used in our study were obtained retrospectively. Therefore, inherent case selection bias might have affected the results. Extensive prospective and *in vivo* and *in vitro* experimental studies are needed to confirm our findings. Although our study was analyzed and validated using many patient samples from different databases, it opened new perspectives and outlooks for cancer treatment. Based on this study, researchers could aim to understand the outstanding potential of AP3S1 in tumor immunity, conduct various experiments to explore it in-depth and contribute to cancer treatment.

In conclusion, our study suggests that AP3S1 is a potential prognostic biomarker and therapeutic target for pan-cancer. High expression of AP3S1 may contribute to the formation of an immunosuppressive microenvironment in tumors. Targeting AP3S1 could be a novel approach for tumor immunotherapy.

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

GW, MC and HR designed and wrote this work. XS, MH and KR integrated and analyzed the data. JQ and FL edited and revised the manuscript. All authors approved this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2022.930933/full#supplementary-material

REFERENCES

Argyle, D., and Kitamura, T. (2018). Targeting Macrophage-Recruiting Chemokines as a Novel Therapeutic Strategy to Prevent the Progression of Solid Tumors. *Front. Immunol.* 9, 2629. doi:10.3389/fimmu.2018.02629

Armitage, J. D., Newnes, H. V., McDonnell, A., Bosco, A., and Waithman, J. (2021). Fine-Tuning the Tumour Microenvironment: Current Perspectives on the Mechanisms of Tumour Immunosuppression. *Cells* 10 (1), 56. doi:10.3390/cells10010056

Badmann, S., Heublein, S., Mayr, D., Reischer, A., Liao, Y., Kolben, T., et al. (2020). M2 Macrophages Infiltrating Epithelial Ovarian Cancer Express MDR1: a Feature that May Account for the Poor Prognosis. *Cells* 9, 1224. doi:10.3390/cells9051224

Barbee, M. S., Ogunniyi, A., Horvat, T. Z., and Dang, T.-O. (2015). Current Status and Future Directions of the Immune Checkpoint Inhibitors Ipilimumab, Pembrolizumab, and Nivolumab in Oncology. *Ann. Pharmacother.* 49, 907–937. doi:10.1177/1060028015586218

Binnewies, M., Roberts, E. W., Kersten, K., Chan, V., Fearon, D. F., Merad, M., et al. (2018). Understanding the Tumor Immune Microenvironment (TIME) for Effective Therapy. *Nat. Med.* 24 (5), 541–550. doi:10.1038/s41591-018-0014-x

Bonapace, L., Coissieux, M.-M., Wyckoff, J., Mertz, K. D., Varga, Z., Junt, T., et al. (2014). Cessation of CCL2 Inhibition Accelerates Breast Cancer Metastasis by Promoting Angiogenesis. *Nature* 515, 130–133. doi:10.1038/nature13862

Cendrowicz, E., Sas, Z., Bremer, E., and Rygiel, T. P. (2021). The Role of Macrophages in Cancer Development and Therapy. *Cancers* 13 (8), 1946. doi:10.3390/cancers13081946

Chen, W., Jin, W., Hardegen, N., Lei, K.-j., Li, L., Marinos, N., et al. (2003). Conversion of Peripheral CD4+CD25− Naive T Cells to CD4+CD25+ Regulatory T Cells by TGF-β Induction of Transcription Factor Foxp3. *J. Exp. Med.* 198, 1875–1886. doi:10.1084/jem.20030152

Chen, Y., Song, Y., Du, W., Gong, L., Chang, H., and Zou, Z. (2019). Tumor-Associated Macrophages: An Accomplice in Solid Tumor Progression. *J. Biomed. Sci.* 26 (1), 78. doi:10.1186/s12929-019-0568-z

Cortez-Retamozo, V., Etzrodt, M., Newton, A., Rauch, P. J., Chudnovskiy, A., Berger, C., et al. (2012). Origins of Tumor-Associated Macrophages and Neutrophils. *Proc. Natl. Acad. Sci. U.S.A.* 109, 2491–2496. doi:10.1073/pnas.1113744109

da Silva, P. H. R., Borges, B. C., Uehara, I. A., Soldi, L. R., de Araújo, R. A., and Silva, M. J. B. (2021). Chemokines and the Extracellular Matrix: Set of Targets for Tumor Development and Treatment. *Cytokine* 144, 155548. doi:10.1016/j.cyto.2021.155548

Dehne, N., Mora, J., Namgaladze, D., Weigert, A., and Brüne, B. (2017). Cancer Cell and Macrophage Cross-Talk in the Tumor Microenvironment. *Curr. Opin. Pharmacol.* 35, 12–19. doi:10.1016/j.coph.2017.04.007

Gajewski, T. F., Schreiber, H., and Fu, Y.-X. (2013). Innate and Adaptive Immune Cells in the Tumor Microenvironment. *Nat. Immunol.* 14 (10), 1014–1022. doi:10.1038/ni.2703

He, D., Wang, D., Lu, P., Yang, N., Xue, Z., Zhu, X., et al. (2021). Single-cell RNA Sequencing Reveals Heterogeneous Tumor and Immune Cell Populations in Early-Stage Lung Adenocarcinomas Harboring EGFR Mutations. *Oncogene* 40, 355–368. doi:10.1038/s41388-020-01528-0

Jayasingam, S. D., Citartan, M., Thang, T. H., Mat Zin, A. A., Ang, K. C., and Ch'ng, E. S. (2019). Evaluating the Polarization of Tumor-Associated Macrophages into M1 and M2 Phenotypes in Human Cancer Tissue: Technicalities and Challenges in Routine Clinical Practice. *Front. Oncol.* 9, 1512. doi:10.3389/fonc.2019.01512

Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., et al. (2013). Mutational Landscape and Significance across 12 Major Cancer Types. *Nature* 502, 333–339. doi:10.1038/nature12634

Kim, Y. J., Park, S. J., and Broxmeyer, H. E. (2011). Phagocytosis, a Potential Mechanism for Myeloid-Derived Suppressor Cell Regulation of CD8+ T Cell Function Mediated through Programmed Cell Death-1 and Programmed Cell Death-1 Ligand Interaction. *J. Immunol.* 187, 2291–2301. doi:10.4049/jimmunol.1002650

Lei, X., Lei, Y., Li, J.-K., Du, W.-X., Li, R.-G., Yang, J., et al. (2020). Immune Cells within the Tumor Microenvironment: Biological Functions and Roles in Cancer Immunotherapy. *Cancer Lett.* 470, 126–133. doi:10.1016/j.canlet.2019.11.009

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 (1), 174. doi:10.1186/s13059-016-1028-7

Mittal, S., Brown, N. J., and Holen, I. (2018). The Breast Tumor Microenvironment: Role in Cancer Development, Progression and Response to Therapy. *Expert Rev. Mol. Diagnostics* 18, 227–243. doi:10.1080/14737159.2018.1439382

Movahedi, K., and Van Ginderachter, J. A. (2016). The Ontogeny and Microenvironmental Regulation of Tumor-Associated Macrophages. *Antioxidants Redox Signal.* 25, 775–791. doi:10.1089/ars.2016.6704

Ngambenjawong, C., Gustafson, H. H., and Pun, S. H. (2017). Progress in Tumor-Associated Macrophage (TAM)-targeted Therapeutics. *Adv. Drug Deliv. Rev.* 114, 206–221. doi:10.1016/j.addr.2017.04.010

Nome, T., Thomassen, G. O. S., Bruun, J., Ahlquist, T., Bakken, A. C., Hoff, A. M., et al. (2013). Common Fusion Transcripts Identified in Colorectal Cancer Cell Lines by High-Throughput RNA Sequencing. *Transl. Oncol.* 6 (5), 546–553. doi:10.1593/tlo.13457

Peña, C. G., Nakada, Y., Saatcioglu, H. D., Aloisio, G. M., Cuevas, I., Zhang, S., et al. (2015). LKB1 Loss Promotes Endometrial Cancer Progression via CCL2-dependent Macrophage Recruitment. *J. Clin. Invest.* 125, 4063–4076. doi:10.1172/JCI82152

Petrenko, A. A., Pavlova, L. S., Karseladze, A. I., Kisseljov, F. L., and Kisseljova, N. P. (2006). Downregulation of Genes Encoding for Subunits of Adaptor Complex-3 in Cervical Carcinomas. *Biochem. Mosc.* 71 (10), 1153–1160. doi:10.1134/s0006297906100130

Pienta, K. J., Machiels, J.-P., Schrijvers, D., Alekseev, B., Shkolnik, M., Crabb, S. J., et al. (2013). Phase 2 Study of Carlumab (CNTO 888), a Human Monoclonal Antibody against CC-Chemokine Ligand 2 (CCL2), in Metastatic Castration-Resistant Prostate Cancer. *Invest. New Drugs* 31, 760–768. doi:10.1007/s10637-012-9869-8

Samstein, R. M., Lee, C. H., Shoushtari, A. N., Hellmann, M. D., Shen, R., Janjigian, Y. Y., et al. (2019). Tumor Mutational Load Predicts Survival after Immunotherapy across Multiple Cancer Types. *Nat. Genet.* 51 (2), 202–206. doi:10.1038/s41588-018-0312-8

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

Simpson, F., Peden, A. A., Christopoulou, L., and Robinson, M. S. (1997). Characterization of the Adaptor-Related Protein Complex, AP-3. *J. Cell Biol.* 137, 835–845. doi:10.1083/jcb.137.4.835

Thomas, D. A., and Massagué, J. (2005). TGF-β Directly Targets Cytotoxic T Cell Functions during Tumor Evasion of Immune Surveillance. *Cancer Cell* 8, 369–380. doi:10.1016/j.ccr.2005.10.012

Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Ou Yang, T. H., et al. (2018). The Immune Landscape of Cancer. *Immunity* 48, 812–830.e14. doi:10.1016/j.immuni.2018.03.023

Topalian, S. L., Hodi, F. S., Brahmer, J. R., Gettinger, S. N., Smith, D. C., McDermott, D. F., et al. (2012). Safety, Activity, and Immune Correlates of Anti-PD-1 Antibody in Cancer. *N. Engl. J. Med.* 366 (26), 2443–2454. doi:10.1056/NEJMoa1200690

Zeng, D., Li, M., Zhou, R., Zhang, J., Sun, H., Shi, M., et al. (2019). Tumor Microenvironment Characterization in Gastric Cancer Identifies Prognostic and Immunotherapeutically Relevant Gene Signatures. *Cancer Immunol. Res.* 7 (5), 737–750. doi:10.1158/2326-6066.CIR-18-0436

Zhang, C., Zhang, Z., Zhang, G., Zhang, Z., Luo, Y., Wang, F., et al. (2020). Clinical Significance and Inflammatory Landscapes of a Novel Recurrence-Associated Immune Signature in Early-Stage Lung Adenocarcinoma. *Cancer Lett.* 479, 31–41. doi:10.1016/j.canlet.2020.03.016

Zou, B., Guo, D., Kong, P., Wang, Y., Cheng, X., and Cui, Y. (2021). Integrative Genomic Analyses of 1,145 Patient Samples Reveal New Biomarkers in Esophageal Squamous Cell Carcinoma. *Front. Mol. Biosci.* 8, 792779. doi:10.3389/fmolb.2021.792779

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.

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