INSIGHTS IN MOLECULAR DIAGNOSTICS AND THERAPEUTICS: 2021

EDITED BY: William C. Cho and Matteo Becatti

PUBLISHED IN: Frontiers in Molecular Biosciences

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-83250-704-9
DOI 10.3389/978-2-83250-704-9

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

INSIGHTS IN MOLECULAR DIAGNOSTICS AND THERAPEUTICS: 2021

Topic Editors:
William C. Cho, QEH, Hong Kong, SAR China
Matteo Becatti, University of Firenze, Italy

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.

Citation: Cho, W. C., Becatti, M., eds. (2022). Insights in Molecular Diagnostics and Therapeutics: 2021. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83250-704-9

Table of Contents

05 ***Editorial: Insights in Molecular Diagnostics and Therapeutics: 2021***
Matteo Becatti and William C. Cho

07 ***Experimental and Meta-Analytic Validation of RNA Sequencing Signatures for Predicting Status of Microsatellite Instability***
Maksim Sorokin, Elizaveta Rabushko, Victor Efimov, Elena Poddubskaya, Marina Sekacheva, Alexander Simonov, Daniil Nikitin, Aleksey Drobyshev, Maria Suntsova and Anton Buzdin

19 ***The Impact of Oxidative Stress in Male Infertility***
Amanda Mannucci, Flavia Rita Argento, Eleonora Fini, Maria Elisabetta Coccia, Niccolò Taddei, Matteo Becatti and Claudia Fiorillo

28 ***Toll-Like Receptors Serve as Biomarkers for Early Diagnosis and Prognosis Assessment of Kidney Renal Clear Cell Carcinoma by Influencing the Immune Microenvironment: Comprehensive Bioinformatics Analysis Combined With Experimental Validation***
Xiong Zou, Bingqian Guo, Qiang Ling and Zengnan Mo

46 ***Validated Impacts of N6-Methyladenosine Methylated mRNAs on Apoptosis and Angiogenesis in Myocardial Infarction Based on MeRIP-Seq Analysis***
Yingjie Zhang, Wenjie Hua, Yini Dang, Yihui Cheng, Jiayue Wang, Xiu Zhang, Meiling Teng, Shenrui Wang, Min Zhang, Zihao Kong, Xiao Lu and Yu Zheng

61 ***Complex Involvement of the Extracellular Matrix, Immune Effect, and Lipid Metabolism in the Development of Idiopathic Pulmonary Fibrosis***
Weiping Qian, Shu Xia, Xiaoyun Yang, Jiaying Yu, Bingpeng Guo, Zhengfang Lin, Rui Wei, Mengmeng Mao, Ziyi Zhang, Gui Zhao, Junye Bai, Qian Han, Zhongfang Wang and Qun Luo

71 ***Comprehensive Analysis of the Prognostic and Immunological Role of PAFAH1B in Pan-Cancer***
Yixiao Yuan, Xiulin Jiang, Lin Tang, Juan Wang and Lincan Duan

91 ***Systematic Analysis and Validation of the Prognosis, Immunological Role and Biology Function of the Ferroptosis-Related lncRNA GSEC/miRNA-101-3p/CISD1 Axis in Lung Adenocarcinoma***
Xiulin Jiang, Yixiao Yuan, Lin Tang, Juan Wang, Dahang Zhang and Lincan Duan

107 ***The Prognostic Value of Lysine Acetylation Regulators in Hepatocellular Carcinoma***
Liying Sun, Jian Zhang, Kai Wen, Shenglan Huang, Dan Li, Yongkang Xu and Jianbing Wu

120 ***Gene Expression-Based Signature Can Predict Sorafenib Response in Kidney Cancer***
Alexander Gudkov, Valery Shirokorad, Kirill Kashintsev, Dmitriy Sokov, Daniil Nikitin, Andrey Anisenko, Nicolas Borisov, Marina Sekacheva, Nurshat Gaifullin, Andrew Garazha, Maria Suntsova, Elena Koroleva, Anton Buzdin and Maksim Sorokin

132 ***Long Non-Coding RNA AP000695.2 Acts as a Novel Prognostic Biomarker and Regulates the Cell Growth and Migration of Lung Adenocarcinoma***
Chunyan Wang, Jishu Guo, Rongyan Jiang, Chenyang Wang, Chenglong Pan, Zhi Nie and Xiulin Jiang

146 ***Pulmonary Microbial Composition in Sepsis-Induced Acute Respiratory Distress Syndrome***
Peng Zhang, Baoyi Liu, Weihao Zheng, Yantang Chen, Zhentao Wu, Yuting Lu, Jie Ma, Wenjie Lu, Mingzhu Zheng, Wanting Wu, Zijie Meng, Jinhua Wu, Yan Zheng, Xin Zhang, Shuang Zhang and Yanming Huang

162 ***Tumor Microenvironment Heterogeneity-Based Score System Predicts Clinical Prognosis and Response to Immune Checkpoint Blockade in Multiple Colorectal Cancer Cohorts***
Hufei Wang, Zhi Li, Suwen Ou, Yanni Song, Kangjia Luo, Zilong Guan, Lei Zhao, Rui Huang and Shan Yu

179 ***Systematic Analysis of MCM3 in Pediatric Medulloblastoma* via *Multi-omics Analysis***
Liangliang Cao, Yang Zhao, Zhuangzhuag Liang, Jian Yang, Jiajia Wang, Shuangwei Tian, Qinhua Wang, Baocheng Wang, Heng Zhao, Feng Jiang and Jie Ma

frontiers | Frontiers in Molecular Biosciences

TYPE Editorial
PUBLISHED 21 October 2022
DOI 10.3389/fmolb.2022.1014416

OPEN ACCESS

EDITED BY
Satyendra Chandra Tripathi,
All India Institute of Medical Sciences Nagpur, India

REVIEWED BY
Theam Soon Lim,
Universiti Sains Malaysia (USM), Malaysia
Martin Richter,
Bundesanstalt für Risikobewertung (BfR), Germany

*CORRESPONDENCE
Matteo Becatti,
matteo.becatti@unifi.it
William C. Cho,
chocs@ha.org.hk

SPECIALTY SECTION
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 08 August 2022
ACCEPTED 21 September 2022
PUBLISHED 21 October 2022

CITATION
Becatti M and Cho WC (2022), Editorial: Insights in molecular diagnostics and therapeutics: 2021.
Front. Mol. Biosci. 9:1014416.
doi: 10.3389/fmolb.2022.1014416

Editorial: Insights in molecular diagnostics and therapeutics: 2021

Matteo Becatti[1]* and William C. Cho[2]*

[1]Department of Experimental and Clinical Biomedical Sciences "Mario Serio", University of Firenze, Firenze, Italy, [2]Department of Clinical Oncology, Queen Elizabeth Hospital, Kowloon, Hong Kong SAR, China

KEYWORDS

biomedical research, molecular diagnostic, therapeutics, biomarkers, precision medicine

Editorial on the Research Topic
Insights in molecular diagnostics and therapeutics: 2021

The most recent advances and discoveries in biomedical research, together with technological progress, have made a paradigm shift from traditional medicine towards more precise and predictable health care, customized for an individual patient at a specific time. In this era of precision medicine, having validated biomarkers to inform clinical decision-making is crucial. Biomarkers have various applications including disease diagnosis and monitoring, prognosis, and prediction of response to treatment. This Research Topic collects some of the latest molecular diagnostic and therapeutic advances and their studies in precision medicine.

The interplay between basic and applied research is essential for the study of the molecular and cellular mechanisms that regulate biological systems. The correlation between the integrated "omics" data, such as the genome, transcriptome, proteome, and other "omics" information, should be considered in a patient-specific, rather than symptom-specific, approach with precision medicine (Abdelhalim et al., 2022). This approach must be adapted to each individual's unique omics leading to personalized management of diseases. While most existing studies analyze the omics data separately, data integration is crucial on the horizon of precision medicine by using machine learning or artificial intelligence approaches (Hasanzad et al., 2021). However, the use of "omics" technologies generates enormous amounts of data that require complex computational analyses and interpretations. In this Research Topic, five original research articles (Jiang et al.; Cao et al.; Sun et al.; Wang et al.; Yuan et al.) presented the interesting "omics" studies that identify new diagnostic and prognostic biomarkers using bioinformatic tools and public databases in diverse types of cancer. Technologies such as array-based hybridization assays and next-generation sequencing promote the identification of new molecular targets and the development of prognostic and predictive biomarkers for their use in precision medicine (Wang and Zheng, 2021). Sorokin et al. used public databases to compare and validate gene signatures of microsatellite instability status by RNA-Seq as diagnostic biomarkers in solid tumors. Moreover, RNA expression-

based signature is used by Gudkov et al. to predict sorafenib response in kidney cancer. The identification of new prognostic and predictive biomarkers is essential for cancer drug development and therapeutic decision-making for individual patients. One of the greatest challenges today is to develop prognostic and predictive biomarkers that translate the genomic information of individual tumors into a personalized therapeutic approach. In this regard, Zou et al. investigated the role of Toll-like receptor (TLR) subtypes expression in kidney renal clear cell carcinoma (KIRC). Bioinformatics analysis and experimental data demonstrate that the occurrence and development of KIRC are closely related to TLRs and TLRs can be early diagnostic and prognostic biomarkers of KIRC. In another way, Wang et al. explored the diagnostic and prognostic values of long non-coding RNA AP000695.2 in lung adenocarcinoma. Using a data mining approach in The Cancer Genome Atlas and the Gene Expression Omnibus databases, they demonstrated that a higher expression of AP000695.2 correlated with aggressive clinicopathological characteristics and AP000695.2 was an independent prognostic indicator for the overall survival, disease-free survival, and progression-free survival in patients with lung adenocarcinoma. Bioinformatics methods are also used by Zhang et al. to study the impacts of N6-methyladenosine (m^6A) methylated mRNAs on epigenetic changes after myocardial infarction. In this Research Topic, another interesting retrospective study analyzes new potential biomarkers of idiopathic pulmonary fibrosis (IPF) by mRNA-Seq analysis of IPF lung tissue obtained from surgical lung biopsy and lung transplantation (Qian et al.). Through bioinformatics analyses using Gene Ontology and Kyoto Encyclopedia of Genes and Genomes public databases, the authors identified differentially expressed transcripts, suggesting the synergic role of extracellular matrix remodeling, lipid metabolism, and immune effects in the early development of IPF.

Indeed, bioinformatic tools are essential for data management in modern biology, life sciences, and medicine. In an interesting work, Zhang et al. through metagenomic next-generation sequencing and bioinformatics analysis, explored the changes in the lung microbiome before and after treatment in acute respiratory distress syndrome (ARDS) patients demonstrating the key role of respiratory tract microbiome in the pathogenesis and development of ARDS.

Last but not the least, Mannucci et al. reviewed the recent advances in the role of oxidative stress in the pathogenesis of male infertility, underlining the clinical use of redox biomarkers and the new treatment of oxidative-stress-mediated male infertility.

Together, the articles comprising this Research Topic shed some light on the current status of molecular diagnostic and therapeutic advances. Each report raises questions and indicates aspects that require further attention and scientific inquiry. The progress of molecular diagnostics will continue to grow in the race to enhance care for individuals using genomic and metagenomic information followed by artificial intelligence data evaluation and machine learning algorithms. This approach should combine traditional clinical data with patients' biological profiles including various omics-based datasets to create a new and exciting path of personalized medicine.

Author contributions

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

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

References

Abdelhalim, H., Berber, A., Lodi, M., Jain, R., Nair, A., Pappu, A., et al. (2022). Artificial intelligence, healthcare, clinical genomics, and pharmacogenomics approaches in precision medicine. *Front. Genet.* 13, 929736. doi:10.3389/fgene.2022.929736

Hasanzad, M., Sarhangi, N., Ehsani Chimeh, S., Ayati, N., Afzali, M., Khatami, F., et al. (2021). Precision medicine journey through omics approach. *J. Diabetes Metab. Disord.* 21 (1), 881–888. doi:10.1007/s40200-021-00913-0

Wang, Y., and Zheng, D. (2021). The importance of precision medicine in modern molecular oncology. *Clin. Genet.* 100 (3), 248–257. doi:10.1111/cge.13998

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 23 November 2021
doi: 10.3389/fmolb.2021.737821

Experimental and Meta-Analytic Validation of RNA Sequencing Signatures for Predicting Status of Microsatellite Instability

Maksim Sorokin[1,2,3]*, *Elizaveta Rabushko*[1,4], *Victor Efimov*[2,5,6], *Elena Poddubskaya*[5], *Marina Sekacheva*[5], *Alexander Simonov*[5,6], *Daniil Nikitin*[6,7], *Aleksey Drobyshev*[1], *Maria Suntsova*[2,5] and *Anton Buzdin*[2,3,5,7]

[1]Laboratory For Clinical and Genomic Bioinformatics, I.M. Sechenov First Moscow State Medical University, Moscow, Russia, [2]Moscow Institute of Physics and Technology, Dolgoprudny, Russia, [3]OmicsWay Corp., Walnut, CA, United States, [4]Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia, [5]World-Class Research Center "Digital Biodesign and Personalized Healthcare", Sechenov First Moscow State Medical University, Moscow, Russia, [6]Oncobox Ltd., Moscow, Russia, [7]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia

OPEN ACCESS

Edited by:
William C. Cho,
QEH, Hong Kong, SAR China

Reviewed by:
Xiaosheng Wang,
China Pharmaceutical University, China
Ximing Xu,
Renmin Hospital of Wuhan University, China

***Correspondence:**
Maksim Sorokin
sorokin@oncobox.com

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 07 July 2021
Accepted: 19 October 2021
Published: 23 November 2021

Citation:
Sorokin M, Rabushko E, Efimov V, Poddubskaya E, Sekacheva M, Simonov A, Nikitin D, Drobyshev A, Suntsova M and Buzdin A (2021) Experimental and Meta-Analytic Validation of RNA Sequencing Signatures for Predicting Status of Microsatellite Instability. Front. Mol. Biosci. 8:737821. doi: 10.3389/fmolb.2021.737821

Microsatellite instability (MSI) is an important diagnostic and prognostic cancer biomarker. In colorectal, cervical, ovarian, and gastric cancers, it can guide the prescription of chemotherapy and immunotherapy. In laboratory diagnostics of susceptible tumors, MSI is routinely detected by the size of marker polymerase chain reaction products encompassing frequent microsatellite expansion regions. Alternatively, MSI status is screened indirectly by immunohistochemical interrogation of microsatellite binding proteins. RNA sequencing (RNAseq) profiling is an emerging source of data for a wide spectrum of cancer biomarkers. Recently, three RNAseq-based gene signatures were deduced for establishing MSI status in tumor samples. They had 25, 15, and 14 gene products with only one common gene. However, they were developed and tested on the incomplete literature of The Cancer Genome Atlas (TCGA) sampling and never validated experimentally on independent RNAseq samples. In this study, we, for the first time, systematically validated these three RNAseq MSI signatures on the literature colorectal cancer (CRC) (n = 619), endometrial carcinoma (n = 533), gastric cancer (n = 380), uterine carcinosarcoma (n = 55), and esophageal cancer (n = 83) samples and on the set of experimental CRC RNAseq samples (n = 23) for tumors with known MSI status. We found that all three signatures performed well with area under the curve (AUC) ranges of 0.94–1 for the experimental CRCs and 0.94–1 for the TCGA CRC, esophageal cancer, and uterine carcinosarcoma samples. However, for the TCGA endometrial carcinoma and gastric cancer samples, only two signatures were effective with AUC 0.91–0.97, whereas the third signature showed a significantly lower AUC of 0.69–0.88. Software for calculating these MSI signatures using RNAseq data is included.

Keywords: microsatellite instability, RNA sequencing, NGS, RNAseq, gene signatures, experimental validation

INTRODUCTION

Microsatellite instability (MSI) results from and is a marker of defective DNA mismatch repair (dMMR). Tumors accumulate multiple mutations across the genome (Ryan et al., 2017). Short tandem repeats are particularly frequent targets to mismatch errors, and dMMR-linked mutations are prone to be present in microsatellite regions (tandem repeats of up to six nucleotides short stretches of DNA) (Johansen et al., 2019). Detectable expansion or shrinkage of microsatellite repeats is referred to as MSI (Marcus et al., 2019).

MSI was the second clinically approved predictive biomarker for the PD1-specific immunotherapy in adult and pediatric advanced cancer patients. In 2017, the approval of the PD1-specific checkpoint inhibitor antibody pembrolizumab for patients with high MSI was based on the evidence of clinical efficacy from five clinical trials (Marcus et al., 2019). This was the first time when a cancer drug was approved based on a general, not a tumor type-specific biomarker.

Tumors with dMMR also have more mutations in non-microsatellite DNA and thus have more neoantigens. For example, an average figure of ~1,800 mutations and ~580 neoantigens was detected in colorectal cancers (CRCs) with dMMR compared with only ~70 mutations and ~20 predicted neoantigens in CRCs with normal MMR (Le et al., 2015). An increased amount of neoantigens in dMMR tumors promotes tumor infiltration by lymphocytes (Dudley et al., 2016; Giannakis et al., 2016), which may cause a more effective response to immunotherapy (Luchini et al., 2019). This provides a theoretical basis for MSI/dMMR biomarker effectiveness for the treatment response to immune checkpoint inhibitors targeting PD-1, PD-L1, and CTLA-4 proteins (Le et al., 2015).

The Food and Drug Administration did not specify which assay should be used to measure MSI. Currently, there are three basic options available for determining MSI status in clinical practice: immunohistochemistry (IHC) for testing dMMR, polymerase chain reaction (PCR), and genomic/exome/panel sequencing for detecting MSI (Ryan et al., 2017; Baretti and Le, 2018; Waalkes et al., 2018).

IHC test interrogates expressions of four proteins: MLH1, MSH2, MSH6, and PMS2. dMMR is diagnosed when there is detected loss of expression of one or more such proteins (Danaher et al., 2019). IHC tests for dMMR/MSI is simple and cost-effective, but it has a downside of relatively low analytic accuracy due to technical inconsistencies such as tissue fixation issues (Engel and Moore, 2011) and biological reasons such as missense mutations in MMR genes that can functionally inactivate protein without altering its IHC-tested expression level (Shia, 2008).

Alternatively, several PCR MSI panels have been designed, and two are most frequently used in practice: (1) two mononucleotide (*BAT-25* and *BAT-26*) and three dinucleotide (*D5S346*, *D2S123*, and *D17S250*) repeat panel (Boland et al., 1998) and (2) five poly-A mononucleotide (*BAT-25*, *BAT-26*, *NR-21*, *NR-24*, and *NR-27*) repeat panel. The latter has greater sensitivity and specificity compared with the (1) panel (Suraweera et al., 2002). Moreover, unlike (1), panel (2) has no requirement of having both tumors and paired healthy tissue for the test (Shemirani et al., 2011). If at least two biomarkers in either panel lose stability, the tumor is diagnosed as MSI-positive.

As PCR testing is based on a limited number of specific microsatellite sites, this approach cannot capture full microsatellite profiles and thus cannot detect ~0.3–10% of MSI cases (16). Furthermore, MSI prevalence and type are markedly different across the different cancer types. For example, lung, breast, and prostate cancers have only ~1–2% MSI incidence (Luchini et al., 2019; Marcus et al., 2019). This proportion is higher for gastric, ovarian, and cervical cancers and is maximal for CRC. These observations are reflected in specific diagnostic guidelines, and MSI testing is not routinely recommended for most tumor types. These factors limit the use of the PCR MSI test on a broad scale (Wang et al., 2021).

DNA sequencing tests use either whole-exome sequencing (WES) or cancer gene panels. For targeted gene panels, the number of genes varies from around 200 to >5,000 genes (Waalkes et al., 2018). Thus, the analytic sites for testing MSI are strongly different among the different targeted panels, whereas the WES approach can provide more objective data, as evidenced by ~100% agreement with gold standard IHC and PCR MSI testing methods for 130 CRC patients when using the MSI sensor method (Johansen et al., 2019).

As opposed to IHC- or PCR-based MSI testing, which are most suitable for CRC and other cancers belonging to the spectrum of Lynch syndrome, the sequencing MSI approach can be used for more tumor types. It can provide an advantage of combining MSI analysis with mutation screening and tumor mutation burden analysis (Wang et al., 2021). However, genomic deep sequencing-based testing has major challenges of high cost and lack of wide availability (Waalkes et al., 2018).

On the other hand, RNA sequencing (RNAseq) can provide another type of data for MSI assessment. In turn, the RNAseq approach has several serious advantages that make it another candidate for an emerging method of choice for MSI testing. RNAseq is a well-established technology for tumor specimens, including formalin-fixed, paraffin-embedded (FFPE) tissue samples (Buzdin et al., 2020). Typically, one RNAseq analysis is less expensive than for WES or panel genomic sequencing (Bossel Ben-Moshe et al., 2018). It can be informative for the assessment of IHC biomarkers (Sorokin et al., 2020c; 2020b), expression of cancer drug target genes (Buzdin et al., 2020; Sorokin et al., 2020d), tumor-specific molecular pathway activation (Buzdin et al., 2018; Borisov et al., 2020a), for personalized modeling of tumor drug response (Kim et al., 2020; Tkachev et al., 2020), and even for tumor mutation burden assessment (DiGuardo et al., 2021). Furthermore, RNAseq data that inform on total gene expression profiles can also be applicable for generating MSI gene signatures. Three such signatures were recently developed (Danaher et al., 2019; Pačínková and Popovici, 2019; Li et al., 2020) based on TCGA project (Tomczak et al., 2015) publicly available RNAseq data for CRC samples annotated with MSI status by gold standard IHC and/or PCR methods. A signature established by Li et al. (2020) includes 25 genes, a signature by Pačínková and Popovici (2019)

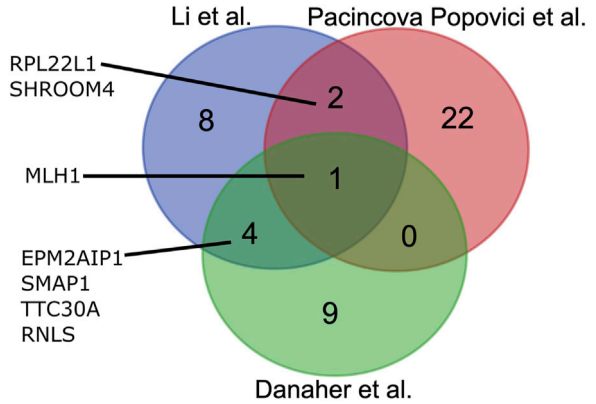


FIGURE 1 | Overlap between gene composition of MSI expression signatures developed by Li et al. (2020), Pačínková and Popovici (2019), and Danaher et al. (2019).

includes 15 genes, and a double signature by Danaher et al. (2019)—14 genes. Interestingly, those signatures are mostly different by gene content and have only one common gene (**Figure 1**).

However, these signatures were developed and validated on the same TCGA samplings and were never validated experimentally on independent RNAseq profiles. In this study, we, for the first time, systematically validated these three RNAseq MSI signatures on the literature CRC (n = 619), endometrial carcinoma (n = 533), gastric cancer (n = 380), uterine carcinosarcoma (n = 55), and esophageal cancer (n = 83) samples and on the set of experimental CRC RNAseq samples (n = 23) for the tumors with known MSI status. As the gold experimental standard, we used seven PCR MSI biomarkers.

We found that all three signatures performed well with area under the curve (AUC) ranges of 0.94–1 for the experimental CRCs and 0.94–1 for the TCGA CRC, esophageal cancer, and uterine carcinosarcoma samples. However, for the TCGA endometrial carcinoma and gastric cancer samples, only two signatures were effective with AUC 0.91–0.97, whereas the third signature showed a significantly lower AUC of 0.69–0.88. Finally, we provide software for calculating these MSI signatures using RNAseq data.

RESULTS

Microsatellite Instability Data Curation and Analysis

For the literature (TCGA) dataset, we extracted MSI statuses for 1,670 available RNAseq samples from the Broad Firehose webpage. These MSI statuses obtained using IHC or PCR profiling were then considered as the gold standards for the assessment of transcriptomic signatures. As only MSI-high tumors are considered for specific therapeutic options, we pooled MSI-low and MSS (microsatellite stable) samples in a single class for further analyses. Totally, we obtained 1,340 MSI-low/MSS and 330 MSI-high profiles. These samples represented

TABLE 1 | Characteristic of literature and experimental cancer patient groups.

Validation set	MSI-high	MSI-low/MSS	Total
Colorectal cancer (CRC)			
Current experimental	6	17	23
Current TCGA	85	534	619
Li TCGA	55	320	375
Pacinkova and Popovici TCGA	35	140	175
Danaher TCGA	27	126	153
Endometrial cancer (UCEC)			
Current TCGA	170	363	533
Li TCGA	123	244	367
Pacinkova and Popovici TCGA	52	64	116
Danaher TCGA	71	176	247
Gastric cancer (STAD)			
Current TCGA	71	309	380
Li TCGA	80	335	415
Pacinkova and Popovici TCGA	54	281	335
Danaher TCGA	64	225	289
Uterine carcinosarcoma (UCS)			
Current TCGA	2	53	55
Li TCGA	2	87	89
Pacinkova and Popovici TCGA	—	—	—
Danaher TCGA	—	—	—
Esophageal cancer (ESCA)			
Current TCGA	2	81	83
Li TCGA	2	54	56
Pacinkova and Popovici TCGA	—	—	—
Danaher TCGA	—	—	—
Control			
Current experimental	1	12	13

CRC, endometrial carcinoma, gastric cancer, uterine cancer, and esophageal cancer (**Table 1**). This was higher than the samplings used previously to validate *Li*, *Pacinkova and Popovici*, and *Danaher* signatures in the original studies (a total of 1,302, 626, and 689 samples, respectively; **Table 1**). We checked RNAseq gene signatures in binary classifier mode.

For the experimental group, we profiled gene expression by RNAseq using FFPE tumor tissue blocks for a total of 23 CRC patients. In addition, we also analyzed a control group of 13 non-CRC tumor blocks to assess MSI signature performance on these samples as well. Among them, five patients had cervical cancer, two had breast cancer, two had gastric cancer, two had glioblastoma, one had ovarian cancer, and one had endometrial carcinosarcoma (**Supplementary Table S1**). In total, the experimental group (n = 36) represented 27 female and nine male patients. The patient age varied from 31 to 84 years; the mean patient age in the experimental group was 60.36 years. More detailed patient information is given in **Supplementary Table S1**.

We performed RNAseq for each tumor sample and obtained ~3.75–78.02 million reads uniquely mapped on known human Ensembl genes (genome version GRCh38 and transcriptome annotation GRCh38.89), on the average ~15.5 million gene-mapped reads per library.

For these samples, "gold standard" MSI statuses were determined by PCR test for seven marker microsatellite loci: BAT25, BAT26, BAT40, NR21, NR24, NR27, and CAT25 that are included in a routinely used clinical panel that requires no healthy tissue control (Suraweera et al., 2002). When there were ≥2 marker loci with detected unstable microsatellite length, these samples were considered MSI-high. Otherwise, the samples were put to the common MSI-low/MSS group. In the experimental group, there were a total of seven MSI-high and 29 MSI-low/MSS samples (**Table 1**, **Supplementary Table S2**).

Performance of Microsatellite Instability RNAseq Gene Signatures

By performing PubMed and Google Scholar literature search with keywords "gene signature," "gene expression," "RNA sequencing," "microsatellite instability,"and "MSI" in March 2021, we extracted 73 hits that were manually processed and returned three recent original publications. These three unrelated research papers authored by Li et al. (2020), Pačínková and Popovici (2019), and Danaher et al. (2019) communicated different gene signatures of MSI status. All these signatures were deduced and initially validated on TCGA CRC samples available at the date of research (**Table 1**). For all the signatures identified, the initial bioinformatic validation cohorts were smaller than those extracted from TCGA in the current study (**Table 1**).

The signatures included 15 genes (Li), 25 genes (Pacincova and Popovici), and 14 genes (Danaher) (**Figure 1**). We compared gene compositions of different signatures and found that they were largely different and shared only one common gene, *MLH1*, which encodes for mutL homolog 1 that can heterodimerize with mismatch repair endonuclease PMS2 to form MutL alpha, part of the DNA mismatch repair system (**Figure 1**). Li signature shared four other genes with Danaher signature: *EPM2AIP1, RNLS, SMAP1*, and *TTC30A*. These genes encode for EPM2A interacting protein 1, renalase, small ArfGAP 1, and tetratricopeptide repeat domain 30A, respectively. Pacincova and Popovici signature also had two other common genes with Li signature: *RPL22L1* and *SHROOM4* encode for ribosomal protein L22 like 1 and shroom family member 4, respectively. Pacincova and Popovici signature had no other common genes with the Danaher signature (**Figure 1**).

The experimental and literature samples were then used to assess the performances of those three signatures. All signature values were calculated as described in the original papers. We created and made publicly available the code for signature calculation at Gitlab: https://gitlab.com/ef.viktor/msi_signatures.

The signatures were validated using TCGA RNAseq datasets for tumor samples annotated by MSI status: CRC (n = 619), endometrial carcinoma (n = 533), gastric cancer (n = 380), uterine carcinosarcoma (n = 55), and esophageal cancer (n = 83) datasets and on the set of experimental CRC RNAseq samples (n = 23) and control experimental dataset for non-CRC cancer samples (n = 13). To assess signature biomarker quality, we used area under the ROC curve (ROC AUC) value as the measure. AUC reflects biomarker robustness and depends on its sensitivity and specificity (Borisov et al., 2020b). It varies between 0.5 and 1, and the typical discrimination threshold is 0.7, where greater values denote high-quality biomarkers and *vice versa* (Boyd, 1997). AUC is often used for scoring different types of molecular biomarkers in oncology (Liu et al., 2018; Tanioka et al., 2018; Chen et al., 2019; Sorokin et al., 2020a). AUC and 95% confidence intervals were calculated using DeLong's method implemented in pROC R-package. The entire experimental

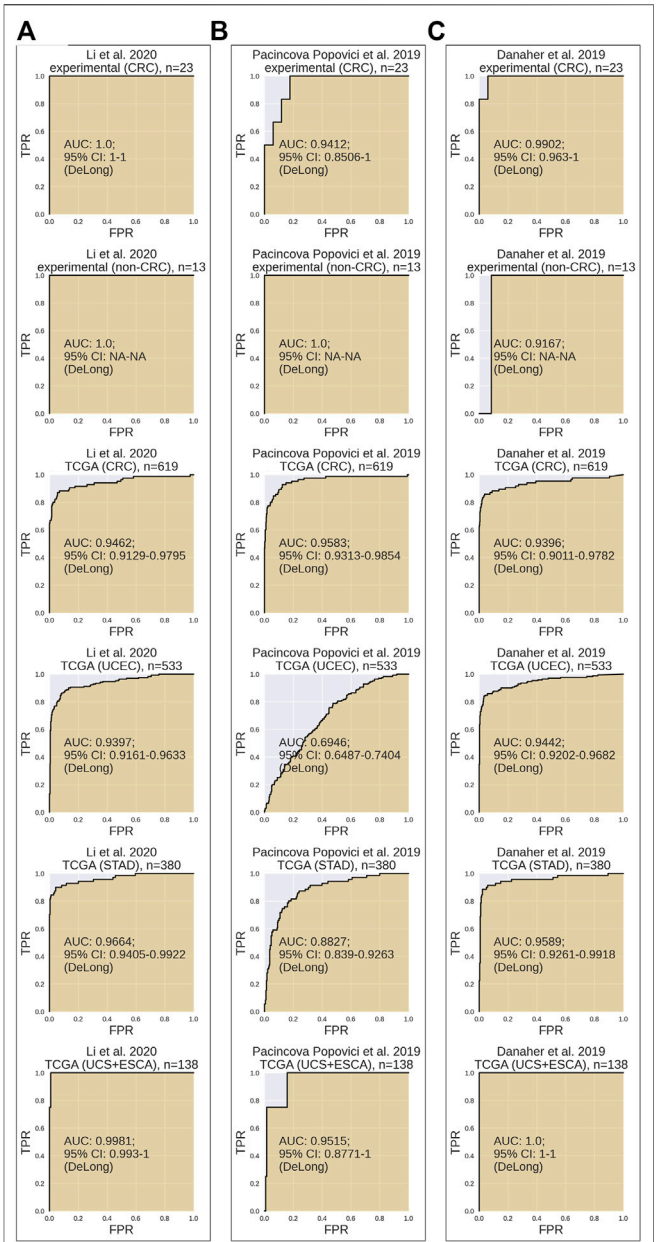


FIGURE 2 | Performance test of MSI RNAseq gene signatures. All signatures were tested for assessment of MSI status on CRC experimental dataset, non-CRC experimental dataset, TCGA CRC dataset, TCGA UCEC dataset, TCGA STAD dataset, and joint TCGA UCS + ESCA dataset. Results for Li et al. (2020) **(A)**, Pačínková and Popovici (2019) **(B)**, and Danaher et al. (2019) **(C)** gene signatures are shown.

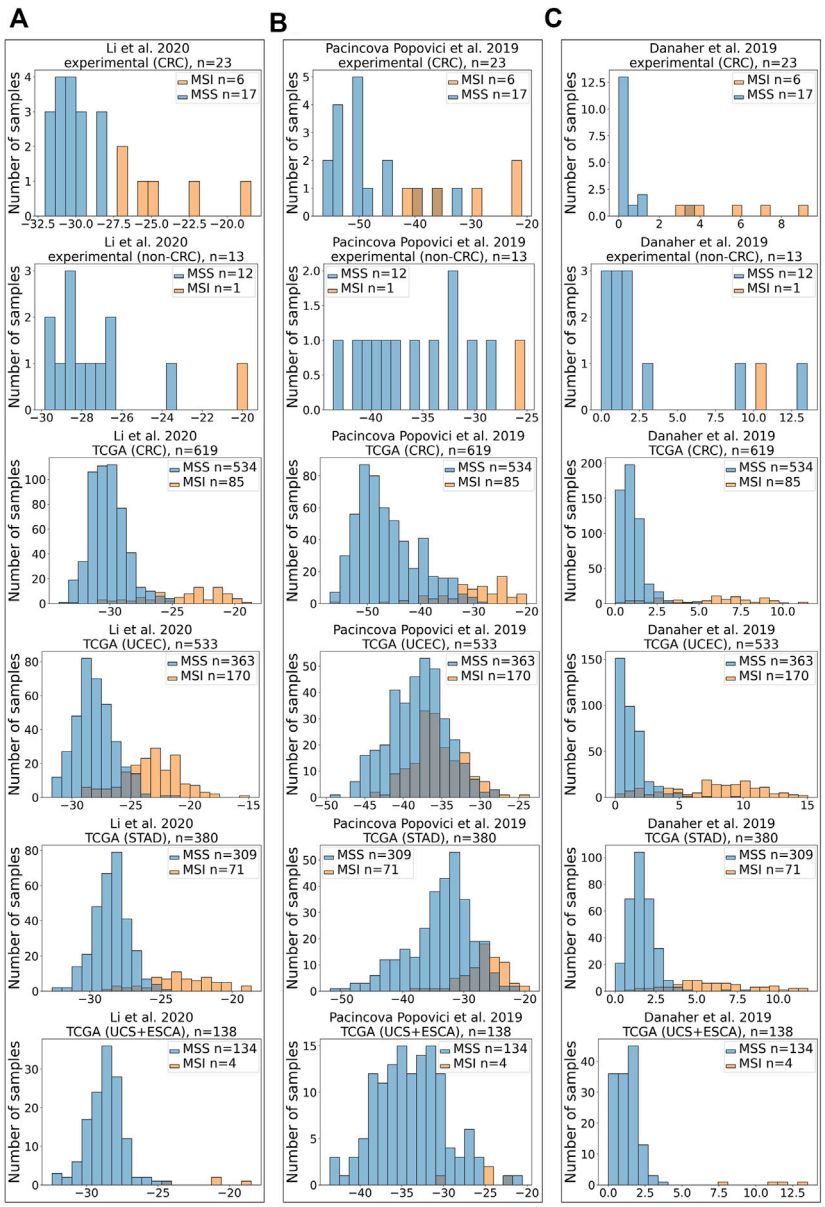


FIGURE 3 | Distribution of scores for MSI RNAseq gene signatures. X-axis shows MSI signature score, Y-axis—number of samples. All signatures were tested for assessment of MSI status on CRC experimental dataset, experimental non-CRC (control) dataset, TCGA CRC dataset, TCGA UCEC dataset, TCGA STAD dataset, and joint TCGA UCS + ESCA dataset. Results for Li et al. (2020) **(A)**, Pačínková and Popovici (2019) **(B)**, and Danaher et al. (2019) **(C)** gene signatures are shown.

dataset contained different cancer types; therefore, AUC was calculated only for the CRC subgroup of the experimental samples.

In our analysis, *Li* MSI signature (**Figure 2A**) scored AUC = 1.0 for the experimental CRC dataset, AUC = 0.9462 for the TCGA CRC, AUC = 0.9397 for the TCGA uterine corpus endometrial carcinoma (UCEC), AUC = 0.9664 for the TCGA STAD dataset, and AUC = 0.9981 for the TCGA joint dataset of UCS + ESCA samples. *Pacincova and Popovici* signature (**Figure 2B**) performed as high as AUC = 0.9412 for the experimental CRC dataset, AUC = 0.9583 for the TCGA CRC dataset, AUC = 0.6946 for the TCGA UCEC, AUC = 0.8827 for the TCGA STAD dataset, and AUC = 0.9515 for the TCGA joint dataset of UCS + ESCA samples. In turn, *Danaher* signature (**Figure 2C**) showed AUC = 0.9902 for the experimental CRC dataset, AUC = 0.9396 for the TCGA CRC dataset, AUC = 0.9442 for the TCGA UCEC, AUC = 0.9589 for the TCGA STAD dataset, and AUC = 1 for the TCGA joint dataset of UCS + ESCA samples dataset.

Similar to variations in AUC metrics for the three signatures tested, their extents related differently to the true-positive or true-negative MSI statuses (**Figures 3A–C**).

TABLE 2 | AUC scores and (95% confidence interval) for three RNAseq MSI gene signatures.

Signature	Li et al. (2020)	Pacinkova and Popovici (2019)	Danaher et al. (2019)
Experimental (CRC), *n* = 23	1.0 (1–1)	0.9412 (0.8506–1)	0.9902 (0.963–1)
TCGA (CRC), *n* = 619	0.9462 (0.9129–0.9795)	0.9583 (0.9313–0.9854)	0.9396 (0.9011–0.9782)
TCGA (UCEC), *n* = 533	0.9397 (0.9161–0.9633)	0.6946 (0.6487–0.7404)	0.9442 (0.9202–0.9682)
TCGA (UCS + ESCA), *n* = 138	0.9981 (0.993–1)	0.9515 (0.8771–1)	1.0 (1–1)
TCGA (STAD), *n* = 380	0.9664 (0.9405–0.9922)	0.8827 (0.839–0.9263)	0.9589 (0.9261–0.9918)

In the experimental CRC group, there were 6 MSI-high and 17 MSI-low samples. However, in the experimental control group that included non-CRC cancers, there was only one MSI-high sample for endometrial carcinosarcoma, whereas all other samples were MSI-low (**Supplementary Table S2**). All three signatures supported the true MSI status of samples in the control group (**Figures 3A–C**).

Assessment of MSI signatures is summarized in **Table 2**. It can be seen that *Li* signature showed the highest AUC in the experimental CRC group, followed by *Danaher* and *Pacincova and Popovici* signatures, respectively (**Table 2**). Also, all three signatures performed accurately on TCGA CRC, esophageal cancer, and uterine carcinosarcoma samples with AUC 0.94-1 and highly overlapping 95% confidence intervals. However, in the endometrial carcinoma (UCEC) cohort of TCGA data, *Pacincova and Popovici* signature showed low AUC below 0.7 threshold, whereas two other signatures showed AUC of at least 0.94. The latter also showed lower performance for TCGA gastric cancer samples (AUC = 0.88 *vs*. 0.96–0.97 in the other two signatures).

Thus, we conclude that in our tests, all three signatures were equally effective for the CRC, esophageal cancer, and uterine carcinosarcoma samples, whereas for the endometrial carcinomas and gastric cancer samples, the *Danaher* and *Li* signatures were found more effective.

We also separately analyzed only early-stage (stages I, IA, and IB) cancer patients from TCGA. In this case, statistical analysis could be performed only for CRC and gastric cancer groups because there were no early-stage MSI-high patients in the other groups. There were 16/13 MSI-high and 89/42 MSI-low samples in CRC and gastric cancer groups, respectively (**Supplementary Figure S1**). All three signatures performed accurately on early-stage TCGA CRC with AUC 0.966–0.997 and highly overlapping 95% confidence intervals (**Supplementary Figure S2**). AUC for *Li* signature was the highest for predicting MSI status in gastric cancer (AUC = 0.956), followed by *Danaher* (AUC = 0.934) and *Pacincova and Popovici* (AUC = 0.919) signatures (**Supplementary Figure S2**).

DISCUSSION

In this study, we, for the first time, systematically compared and validated RNAseq gene signatures of MSI status in human solid tumors. All the signatures performed well on both literature and experimental samplings with the MSI statuses determined using the gold standard techniques routinely used in cancer molecular diagnostics. Interestingly, these three signatures were developed by different teams using different logical rationale and were mostly nonoverlapping with only one common gene, *MLH1*, which protein product heterodimerizes to form MutL alpha (Lindner et al., 2021; Pannafino and Alani, 2021), important actor of the DNA mismatch repair system that is widely associated with the Lynch syndrome known as hereditary nonpolyposis CRC, and MSI (Yamamoto and Imai, 2019; Lindner et al., 2021; Stinton et al., 2021).

However, the functions of most other genes in the three MSI signatures strongly differ. We used Gene Ontology (GO) analysis to identify GO term "biological processes" enriched among the genes forming each signature. Of note, we found 23 enriched biological processes in *Li* gene signature (**Figure 4**), 30 in *Danaher* signature (**Figure 5**), and no significantly enriched processes in *Pacincova and Popovici* signature.

The most significant terms in *Li* signature were associated with meiosis, mismatch repair, and (unexpectedly) with glycogen biosynthesis (**Figure 4**). Interestingly, there were previously only indirect links reported for the glycogen metabolism and Lynch syndrome (Kato, 2020) or MSI (Krausova and Korinek, 2014; Oh et al., 2016), e.g., through the Wnt signaling pathway (Krausova and Korinek, 2014). In *Danaher* signature, the most significant terms were associated with mismatch repair and with somatic hypermutation of immunoglobulin genes and physiologically related processes: somatic diversification of immune receptors and immunoglobulins (**Figure 5**). The latter feature is widely associated with Lynch syndrome and MSI (Anghileri et al., 2021; Mäki-Nevala et al., 2021). Among the signatures by *Li* and *Danaher*, "Mismatch repair" was the only common GO term (highlighted in italic on **Figures 4** and **5**), and mismatch repair deficiency is one of the most obvious reasons for MSI (Jin and Sinicrope, 2021). However, analysis of *Pacincova and Popovici* signature returned no enriched functional terms, thus evidencing that it contains quite a functionally heterogeneous gene set.

We then performed Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment and gene set enrichment (GSEA) analyses. The analyses returned three common statistically significantly enriched pathways for *Danaher* signature: "Mismatch repair," "Platinum drug resistance," and "Colorectal cancer" (**Supplementary Figures S3 and S4**). Thus, GSEA and KEGG analyses confirmed our previous finding that *Danaher* signature is enriched by mismatch repair genes. However, neither KEGG pathway enrichment nor GSEA provided significantly enriched pathways for both *Pacincova and Popovici* and *Li* signatures.

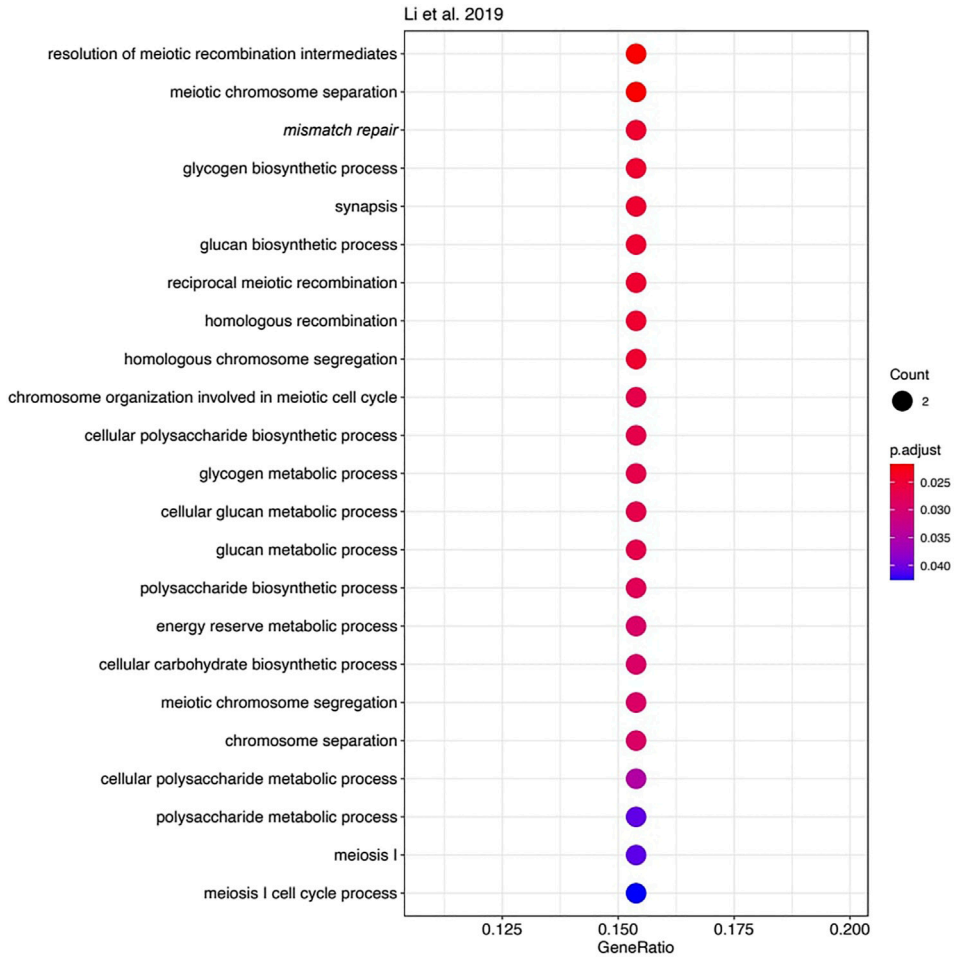


FIGURE 4 | Biological process GO terms for genes included in Li signature. Visualized using R package enrichplot (http://bioconductor.org/packages/release/bioc/html/enrichplot.html). All terms passed Benjamini–Hochberg adjusted *p*-value threshold of 0.05.

This apparent gene content diversity among the signatures demonstrates that MSI can be associated with several or many processes that are not exclusively linked with DNA hypermutation and repair. This gives hope for building next-generation MSI signatures with even better performance/classifier scores.

Our results also imply that the *Li* and *Danaher* signatures may be effective for the CRCs, esophageal cancers, uterine carcinosarcomas, endometrial carcinomas, and gastric cancers. However, the overall effectiveness of *Pacincova and Popovici* signature in our tests was lower and limited to the first three among the cancer types discussed earlier. Moreover, all three signatures performed well for predicting MSI status in early-stage CRC and gastric cancer. Interestingly, the *Li* and *Danaher* signatures that were significantly enriched by genes for certain biological processes (**Figures 3** and **4**) were effective for more cancer types than *Pacincova and Popovici* signature that lacked enriched GO terms.

In addition, the current experimental dataset may serve for validating new such signatures. Finally, we implemented here all the MSI signatures assessed as the free code ready to use with the user RNAseq data. In the future and after careful clinical validation, this may have a practical significance for establishing MSI statuses by screening, when available, RNAseq data for the cancers not necessarily strongly associated with the Lynch syndrome.

MATERIALS AND METHODS

Patients and Samples

In this study, we investigated MSI status-annotated RNAseq profiles for a total of 1,693 cancer samples (one sample per individual patient). Among them, there were 619 literature CRC samples from TCGA cohort, 533 TCGA UCEC samples, 380 TCGA gastric cancer samples, 55 TCGA uterine carcinosarcoma samples, 83 TCGA esophageal cancer samples, and 36 experimental samples profiled by RNA sequencing in this study. TCGA RNAseq samples were extracted from five source datasets: COAD (colon cancer, $n = 389$) and READ (rectal cancer,

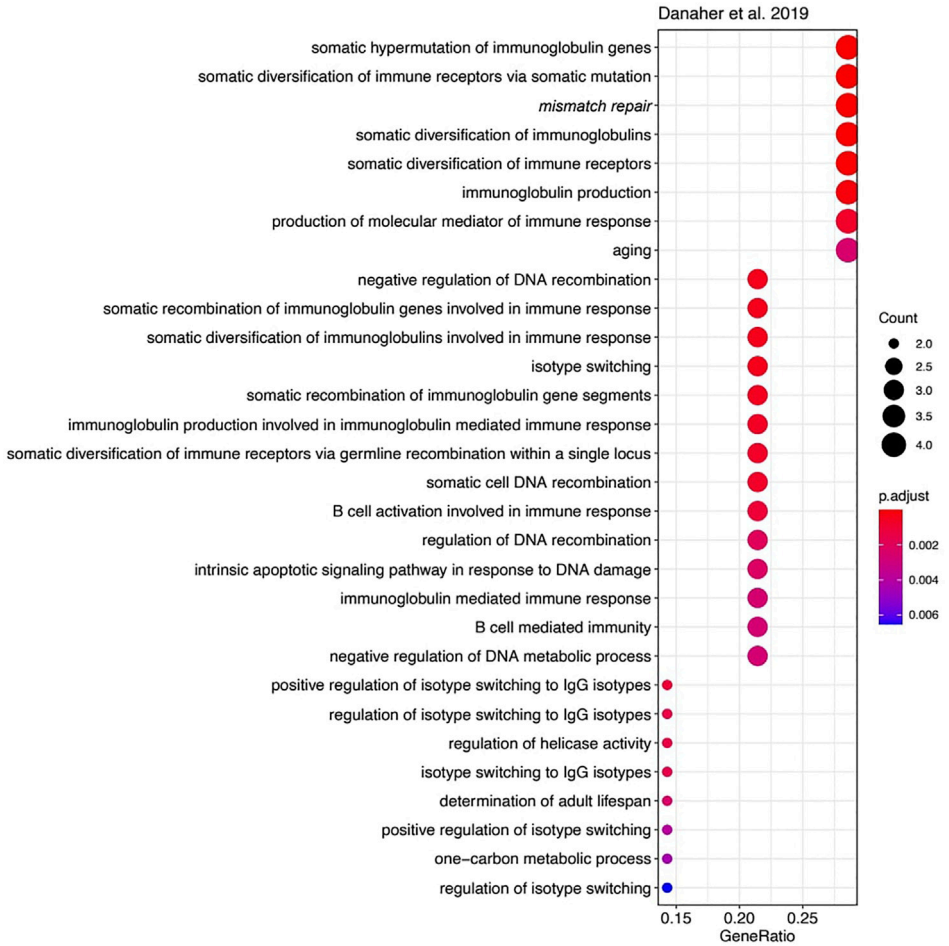


FIGURE 5 | Biological process GO terms for genes included in Danaher signature. Visualized using R package enrichplot (http://bioconductor.org/packages/release/bioc/html/enrichplot.html). All terms passed Benjamini–Hochberg adjusted *p*-value threshold of 0.05.

n = 230) for "CRC," UCEC (endometrial carcinoma, *n* = 533), STAD (gastric cancer, *n* = 380), UCS (uterine carcinosarcoma, *n* = 55), and ESCA (esophageal cancer, *n* = 83). MSI annotated TCGA data were downloaded from https://gdac.broadinstitute.org/.

The experimental dataset included 23 colon cancer samples, five cervical cancer samples, two breast cancer, two gastric cancer samples, two glioblastoma samples, one ovarian cancer sample, and one endometrial carcinosarcoma sample. All experimental specimens were stored in the form of FFPE tissue blocks.

Gene Expression Profiling

To isolate RNA, 10-μM thick paraffin slices were trimmed from each FFPE tissue block using a microtome. RNA preps were extracted using QIAGEN RNeasy FFPE Kit. RNA 6000 Nano or Qubit RNA Assay kits were used to measure RNA concentration. RNA integrity number was measured using Agilent 2100 bio-Analyzer. For depletion of ribosomal RNA and library construction, KAPA RNA Hyper with rRNA erase kit (HMR only) was used. Different adaptors were used for multiplexing samples in one sequencing run. Library concentrations and quality were measured using Qubit ds DNA HS Assay kit (Life Technologies) and Agilent Tapestation (Agilent). RNA sequencing was done using Illumina NextSeq 550 equipment for single-end sequencing, 50-bp read length, for approximately 30 million (mln) raw reads per sample. Data quality check was done on Illumina SAV. De-multiplexing was performed with the Illumina Bcl2fastq2 v 2.17 program. Sequencing data were deposited in National Center for Biotechnology Information Sequencing Read Archive under accession ID PRJNA744404.

Processing of Experimental RNAseq Data

RNAseq FASTQ files were processed with STAR aligner (Dobin et al., 2013) in "GeneCounts" mode with the Ensembl human transcriptome annotation (Build version GRCh38 and transcript annotation GRCh38.89). Ensembl gene IDs were converted to HUGO Gene Nomenclature Committee (HGNC) gene symbols using the Complete HGNC dataset (https://www.genenames.org/, database version from July 13, 2017). Totally, expression levels were established for 36,596 annotated genes with the corresponding HGNC identifiers. Quantile normalization (qnorm python package) was used to normalize gene expression values.

TABLE 3 | Oligonucleotide sequences and fluorescent labels used.

Marker	Gene		Primer sequence and fluorescent labels (5′-3′)	Length (bp)
BAT26	*hMSH2*	Forward	FAM-CTGCGGTAATCAAGTTTTTAG	183
		Reverse	AACCATTCAACATTTTTAACCC	
BAT25	*c-kit*	Forward	R6G-TACCAGGTGGCAAAGGGCA	153
		Reverse	TCTGCATTTTAACTATGGCTC	
NR24	*Zinc finger 2 (ZNF-2)*	Forward	TAMRA-GCTGAATTTTACCTCCTGAC	131
		Reverse	ATTGTGCCATTGCATTCCAA	
NR21	*SLC7A8*	Forward	FAM-GAGTCGCTGGCACAGTTCTA	109
		Reverse	CTGGTCACTCGCGTTTACAA	
NR27	*Inhibitor of apoptosis Protein-1 (IAP1)*	Forward	R6G-AACCATGCTTGCAAACCACT	87
		Reverse	CGATAATACTAGCAATGACC	
BAT40	*3-β-hydroxysteroid dehydrogenase (HSD3B1)*	Forward	ROX-AGTCCATTTTATATCCTCAAGC	145
		Reverse	GTAGAGCAAGACCACCTTG	
CAT25	*Caspase 2*	Forward	ROX-CTTCCCAACTTCCCTGTTCTTT	109
		Reverse	TGAGCTGAGATCGTGCCACT	

Calculating Li et al. Signature Values

MSI RNAseq signature described by Li et al. (2020) was calculated according to the original paper. This signature defines *LYG1*, *MSH4*, and *RPL22L1* genes as "plus"-genes and *DDX27, EPM2AIP1, HENMT1, MLH1, NHLRC1, NOL4L, RNLS, RTFDC1, SHROOM4, SMAP1, TTC30A,* and *ZSWIM3* as "minus"-genes. The final score is a sum of log10-transformed normalized gene expression levels with consideration of each gene sign.

Calculating Pacincova and Popovici Signature

MSI RNAseq signature described by Pacincova and Popovici was calculated according to the original paper (Pačínková and Popovici, 2019). This signature defines *AGR2*, *TNNT1*, *VNN2*, *TNFSF9*, *TRIM7*, and *RPL22L1* genes as "plus"-genes and *ACSL6*, *ARID3A*, *ASCL2*, *AXIN2*, *EPDR1*, *GGT7*, *GNG4*, *KHDRBS3*, *KRT23*, *MLH1*, *NKD1*, *PLAGL2*, *PRR15*, *RUBCNL*, *SHROOM2*, *SHROOM4*, *TFCP2L1*, *TNNC2*, and *VAV3* genes as "minus"-genes. The final score is a sum of log10-transformed gene expression levels with consideration of each gene sign.

Calculating Danaher et al. Signature

MSI RNAseq signature described by Danaher et al. (2019)was calculated according to the original paper. This signature includes *MLH1, MSH2, MSH6,* and *PMS2* genes for calculating MMR loss score (MLS). First, a minimal Z-score (Zmin) of log2-transformed gene expressions was found. The final MLS = (Zmin + 1.03)/0.69, where 1.03 and 0.69 are the theoretical expectation and standard deviation of the minimum of four standard normal random variables, respectively.

Hypermutation predictor score was calculated by multiplying log2-transformed expressions of *EPM2AIP1, TTC30A, SMAP1, RNLS, WNT11, SFXN1, SREBF1, TYMS, EIF5AL1,* and *WDR76* genes by coefficients from the table given in the original article. The final hypermutation predictor score is a Z-score of the calculated value. The resulting MSI predictor score was calculated as follows:

$$\sqrt{min(\boldsymbol{MLS}, 0)^2 + max(\boldsymbol{HPS}, 0)^2}$$

The MSI predictor score is further used as a predictor of MSI-high status.

Functional Gene Set Enrichment Analysis

KEGG and GO analyses were performed using the R clusterProfiler package. EnrichKEGG and enrichGO functions were used to implement enrichment analysis. GSEA analysis was performed using the web service http://www.webgestalt.org. The following non-default parameters were selected: KEGG pathways were used as a functional database, and the minimum number of genes for a category was set to 3. We used Benjamini–Hochberg false discovery rate correction method and applied a *p*-value threshold of 0.05 as a cutoff value for filtering pathways and GO terms.

Experimental Microsatellite Instability Assessment by Polymerase Chain Reaction

Genomic DNA was isolated from FFPE tissue sections using the QIAamp DNA FFPE Tissue Kit (Qiagen, Valencia, CA).

We performed MSI analysis using a set of five so-called "main" mononucleotide repeat markers: *BAT25*, *BAT26*, *NR21*, and *NR24* selected from the revised Bethesda panel (Suraweera et al., 2002) and *NR27* selected from the modified pentaplex panel (Buhard et al., 2006). Two additional mononucleotide repeat markers were also included: *BAT40*, as it was shown to improve the sensitivity of MSI testing in both CRC and extra-colonic tumors (Hartmann et al., 2002; Pagin et al., 2013) and *CAT-25*, which was reported to increase the sensitivity for identifying d*MSH6* tumors (Takehara et al., 2018).

The primer sequences were taken from previous reports (Hartmann et al., 2002; Suraweera et al., 2002; Buhard et al., 2006; Takehara et al., 2018). The sequences of fluorescently labeled oligonucleotides are listed in **Table 3**.

The marker DNA products were PCR amplified using the qPCRmix-HS (Evrogen, Russia). PCR was carried out in a 20-μl final volume containing 1× qPCRmix-HS, 2 *p*moles of each primer, and approximately 20 ng of DNA template.

The marker sets (1) *BAT25*, *BAT26*, *NR21*, and *NR27* and (2) *BAT-40* and *CAT-25* were co-amplified in one PCR tube

per set. The marker *NR-24* was amplified in a separate PCR tube.

PCR conditions for the tetraplex and duplex assays consisted of an initial 2-min denaturation step at 94C, followed by 37 cycles at 94°C for 20 s, 54°C for 10 s, and 72°C for 12 s, with a final extension at 72°C for 2 min. Conditions for monoplex reaction differed in annealing temperature: 53°C.

Amplified PCR products were analyzed by capillary electrophoresis performed on ABI prism 3130 × l System (Applied Biosystems, United States). The microsatellite marker lengths were detected by Sequence Scanner software (Applied Biosystems, United States).

The cutoff for MSI status classification was chosen on the basis of the threshold of approximately 40%, according to Umar A. et al. (2004). Tumors with instability at $\geqslant 2$ of the five main mononucleotide markers were defined as MSI-H. Samples with instability at one main marker were further tested with the additional markers. Tumors with at least one unstable additional marker were defined as MSI-high. Otherwise, tumors were classified as MSI-low/MSS.

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/bioproject/PRJNA744404.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Vitamed Oncological Clinical Center. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AB supervised this study. EP and MSe collected and clinically annotated cancer patient biosamples. ER, AD, and MSu performed MSI experimental PCR analyses. VE and AS implemented computational platforms. MSo and VE analyzed expression data and did statistical calculations. AB, MSo, MSe, ER, and VE wrote the paper. AB is the guarantor of this work and, as such, has full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

FUNDING

This study was financed by the Ministry of Science and Higher Education of the Russian Federation within the framework of state support for the creation and development of World-Class Research Centers "Digital biodesign and personalized healthcare" No 075-15-2020-926.

ACKNOWLEDGMENTS

We thank the OmicsWay research initiative for clinical and technical support. Amazon and Microsoft Azure grants supported Cloud-based computational facilities.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.737821/full#supplementary-material

REFERENCES

Anghileri, E., Di Ianni, N., Paterra, R., Langella, T., Zhao, J., Eoli, M., et al. (2021). High Tumor Mutational burden and T-Cell Activation Are Associated with Long-Term Response to Anti-PD1 Therapy in Lynch Syndrome Recurrent Glioblastoma Patient. *Cancer Immunol. Immunother.* 70, 831–842. doi:10.1007/s00262-020-02769-4

Baretti, M., and Le, D. T. (2018). DNA Mismatch Repair in Cancer. *Pharmacol. Ther.* 189, 45–62. doi:10.1016/j.pharmthera.2018.04.004

Boland, C. R., Thibodeau, S. N., Hamilton, S. R., Sidransky, D., Eshleman, J. R., Burt, R. W., et al. (1998). A National Cancer Institute Workshop on Microsatellite Instability for Cancer Detection and Familial Predisposition: Development of International Criteria for the Determination of Microsatellite Instability in Colorectal Cancer. *Cancer Res.* 58, 5248–5257.

Borisov, N., Sorokin, M., Garazha, A., and Buzdin, A. (2020a). "Quantitation of Molecular Pathway Activation Using RNA Sequencing Data," in *Methods in Molecular Biology*. (Totowa: Humana Press), 189–206.doi:10.1007/978-1-0716-0138-9_15

Borisov, N., Sorokin, M., Tkachev, V., Garazha, A., and Buzdin, A. (2020b). Cancer Gene Expression Profiles Associated with Clinical Outcomes to Chemotherapy Treatments. *BMC Med. Genomics* 13, 1–9. doi:10.1186/s12920-020-00759-0

Bossel Ben-Moshe, N., Gilad, S., Perry, G., Benjamin, S., Balint-Lahat, N., Pavlovsky, A., et al. (2018). mRNA-seq Whole Transcriptome Profiling of Fresh Frozen versus Archived Fixed Tissues. *BMC Genomics* 19, 1–11. doi:10.1186/s12864-018-4761-3

Boyd, J. (1997). Mathematical Tools for Demonstrating the Clinical Usefulness of Biochemical Markers. *Scli* 57, 46–63. doi:10.3109/00365519709168308

Buhard, O., Cattaneo, F., Wong, Y. F., Yim, S. F., Friedman, E., Flejou, J.-F., et al. (2006). Multipopulation Analysis of Polymorphisms in Five Mononucleotide Repeats Used to Determine the Microsatellite Instability Status of Human Tumors. *Jco* 24, 241–251. doi:10.1200/JCO.2005.02.7227

Buzdin, A., Sorokin, M., Garazha, A., Glusker, A., Aleshin, A., Poddubskaya, E., et al. (2020). RNA Sequencing for Research and Diagnostics in Clinical Oncology. *Semin. Cancer Biol.* 60, 311–323. doi:10.1016/j.semcancer.2019.07.010

Buzdin, A., Sorokin, M., Garazha, A., Sekacheva, M., Kim, E., Zhukov, N., et al. (2018). Molecular Pathway Activation - New Type of Biomarkers for Tumor Morphology and Personalized Selection of Target Drugs. *Semin. Cancer Biol.* 53, 110–124. doi:10.1016/j.semcancer.2018.06.003

Chen, L., Zhou, Y., Tang, X., Yang, C., Tian, Y., Xie, R., et al. (2019). EGFR Mutation Decreases FDG Uptake in Non-small C-ell L-ung C-ancer *via* the NOX4/ROS/GLUT1 axis. *Int. J. Oncol.* 54, 370–380. doi:10.3892/ijo.2018.4626

Danaher, P., Warren, S., Ong, S., Elliott, N., Cesano, A., and Ferree, S. (2019). A Gene Expression Assay for Simultaneous Measurement of Microsatellite Instability and Anti-tumor Immune Activity. *J. Immunotherapy Cancer* 7, 15. doi:10.1186/s40425-018-0472-1

DiGuardo, M. A., Davila, J. I., Jackson, R. A., Nair, A. A., Fadra, N., Minn, K. T., et al. (2021). RNA-seq Reveals Differences in Expressed Tumor Mutation

Burden in Colorectal and Endometrial Cancers with and without Defective DNA-Mismatch Repair. *J. Mol. Diagn.* 23, 555–564. doi:10.1016/j.jmoldx.2021.01.008

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

Dudley, J. C., Lin, M.-T., Le, D. T., and Eshleman, J. R. (2016). Microsatellite Instability as a Biomarker for PD-1 Blockade. *Clin. Cancer Res.* 22, 813–820. doi:10.1158/1078-0432.CCR-15-1678

Engel, K. B., and Moore, H. M. (2011). Effects of Preanalytical Variables on the Detection of Proteins by Immunohistochemistry in Formalin-Fixed, Paraffin-Embedded Tissue. *Arch. Pathol. Lab. Med.* 135, 537–543. doi:10.5858/2010-0702-rair.1

Giannakis, M., Mu, X. J., Shukla, S. A., Qian, Z. R., Cohen, O., Nishihara, R., et al. (2016). Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. *Cell ReportsCell Rep* 1517 (4), 8571206–8571865. S2211124716303643. doi:10.1016/j.celrep.2016.10.009

Hartmann, A., Zanardo, L., Bocker-Edmonston, T., Blaszyk, H., Dietmaier, W., Stoehr, R., et al. (2002). Frequent Microsatellite Instability in Sporadic Tumors of the Upper Urinary Tract. *Cancer Res.* 62, 6796–6802.

Jin, Z., and Sinicrope, F. A. (2021). Prognostic and Predictive Values of Mismatch Repair Deficiency in Non-metastatic Colorectal Cancer. *Cancers* 13, 300–316. doi:10.3390/cancers13020300

Johansen, A. F. B., Kassentoft, C. G., Knudsen, M., Laursen, M. B., Madsen, A. H., Iversen, L. H., et al. (2019). Validation of Computational Determination of Microsatellite Status Using Whole Exome Sequencing Data from Colorectal Cancer Patients. *BMC Cancer* 19, 971. doi:10.1186/s12885-019-6227-7

Kato, N. (2020). Pathology of clear Cell Carcinoma of the Ovary: A Basic View Based on Cultured Cells and Modern View from Comprehensive Approaches. *Pathol. Int.* 70, 591–601. doi:10.1111/pin.12954

Kim, E. L., Sorokin, M., Kantelhardt, S. R., Kalasauskas, D., Sprang, B., Fauss, J., et al. (2020). Intratumoral Heterogeneity and Longitudinal Changes in Gene Expression Predict Differential Drug Sensitivity in Newly Diagnosed and Recurrent Glioblastoma. *Cancers* 12, 520. doi:10.3390/cancers12020520

Krausova, M., and Korinek, V. (2014). Wnt Signaling in Adult Intestinal Stem Cells and Cancer. *Cell Signal.* 26, 570–579. doi:10.1016/j.cellsig.2013.11.032

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, L., Feng, Q., and Wang, X. (2020). PreMSIm: An R Package for Predicting Microsatellite Instability from the Expression Profiling of a Gene Panel in Cancer. *Comput. Struct. Biotechnol. J.* 18, 668–675. doi:10.1016/j.csbj.2020.03.007

Lindner, A. K., Schachtner, G., Tulchiner, G., Thurnher, M., Untergasser, G., Obrist, P., et al. (2021). Lynch Syndrome: Its Impact on Urothelial Carcinoma. *Ijms* 22, 531. doi:10.3390/ijms22020531

Liu, T., Cheng, G., Kang, X., Xi, Y., Zhu, Y., Wang, K., et al. (2018). Noninvasively Evaluating the Grading and IDH1 Mutation Status of Diffuse Gliomas by Three-Dimensional Pseudo-continuous Arterial Spin Labeling and Diffusion-Weighted Imaging. *Neuroradiology* 60, 693–702. doi:10.1007/s00234-018-2021-5

Luchini, C., Bibeau, F., Ligtenberg, M. J. L., Singh, N., Nottegar, A., Bosse, T., et al. (2019). ESMO Recommendations on Microsatellite Instability Testing for Immunotherapy in Cancer, and its Relationship with PD-1/pd-L1 Expression and Tumour Mutational burden: A Systematic Review-Based Approach. *Ann. Oncol.* 30, 1232–1243. doi:10.1093/annonc/mdz116

Mäki-Nevala, S., Ukwattage, S., Olkinuora, A., Almusa, H., Ahtiainen, M., Ristimäki, A., et al. (2021). Somatic Mutation Profiles as Molecular Classifiers of Ulcerative Colitis-associated Colorectal Cancer. *Int. J. Cancer* 148, 2997–3007. doi:10.1002/ijc.33492

Marcus, L., Lemery, S. J., Keegan, P., and Pazdur, R. (2019). FDA Approval Summary: Pembrolizumab for the Treatment of Microsatellite Instability-High Solid Tumors. *Clin. Cancer Res.* 25, 3753–3758. doi:10.1158/1078-0432.CCR-18-4070

Oh, B. Y., Kim, S.-Y., Lee, Y. S., Hong, H. K., Kim, T. W., Kim, S. H., et al. (2016). Twist1-induced Epithelial-Mesenchymal Transition According to Microsatellite Instability Status in colon Cancer Cells. *Oncotarget* 7, 57066–57076. doi:10.18632/oncotarget.10974

Pačínková, A., and Popovici, V. (2019). Cross-platform Data Analysis Reveals a Generic Gene Expression Signature for Microsatellite Instability in Colorectal Cancer. *Biomed. Res. Int.* 2019, 1–9. doi:10.1155/2019/6763596

Pagin, A., Zerimech, F., Leclerc, J., Wacrenier, A., Lejeune, S., Descarpentries, C., et al. (2013). Evaluation of a New Panel of Six Mononucleotide Repeat Markers for the Detection of DNA Mismatch Repair-Deficient Tumours. *Br. J. Cancer* 108, 2079–2087. doi:10.1038/bjc.2013.213

Pannafino, G., and Alani, E. (2021). Coordinated and Independent Roles for MLH Subunits in DNA Repair. *Cells* 10, 948. doi:10.3390/cells10040948

Ryan, E., Sheahan, K., Creavin, B., Mohan, H. M., and Winter, D. C. (2017). The Current Value of Determining the Mismatch Repair Status of Colorectal Cancer: A Rationale for Routine Testing. *Crit. Rev. Oncology/Hematology* 116, 38–57. doi:10.1016/j.critrevonc.2017.05.006

Shemirani, A. I., Haghighi, M. M., Zadeh, S. M., Fatemi, S. R., Taleghani, M. Y., Zali, N., et al. (2011). Simplified MSI Marker Panel for Diagnosis of Colorectal Cancer. *Asian Pac. J. Cancer Prev.* 12, 2101–2104.

Shia, J. (2008). Immunohistochemistry versus Microsatellite Instability Testing for Screening Colorectal Cancer Patients at Risk for Hereditary Nonpolyposis Colorectal Cancer Syndrome. *J. Mol. Diagn.* 10, 293–300. doi:10.2353/jmoldx.2008.080031

Sorokin, M., Ignatev, K., Barbara, V., Vladimirova, U., Muraveva, A., Suntsova, M., et al. (2020a). Molecular Pathway Activation Markers Are Associated with Efficacy of Trastuzumab Therapy in Metastatic HER2-Positive Breast Cancer Better Than Individual Gene Expression Levels. *Biochem. Mosc.* 85, 758–772. doi:10.1134/S0006297920070044

Sorokin, M., Ignatev, K., Poddubskaya, E., Vladimirova, U., Gaifullin, N., Lantsov, D., et al. (2020b). RNA Sequencing in Comparison to Immunohistochemistry for Measuring Cancer Biomarkers in Breast Cancer and Lung Cancer Specimens. *Biomedicines* 8, 114. doi:10.3390/BIOMEDICINES8050114

Sorokin, M., Kholodenko, I., Kalinovsky, D., Shamanskaya, T., Doronin, I., Konovalov, D., et al. (2020c). RNA Sequencing-Based Identification of Ganglioside GD2-Positive Cancer Phenotype. *Biomedicines* 8, 142. doi:10.3390/BIOMEDICINES8060142

Sorokin, M., Poddubskaya, E., Baranova, M., Glusker, A., Kogoniya, L., Markarova, E., et al. (2020d). RNA Sequencing Profiles and Diagnostic Signatures Linked with Response to Ramucirumab in Gastric Cancer. *Cold Spring Harb. Mol. Case Stud.* 6, a004945. doi:10.1101/MCS.A004945

Stinton, C., Jordan, M., Fraser, H., Auguste, P., Court, R., Al-Khudairy, L., et al. (2021). Testing Strategies for Lynch Syndrome in People with Endometrial Cancer: Systematic Reviews and Economic Evaluation. *Health Technol. Assess.* 25, 1–216. doi:10.3310/hta25420

Suraweera, N., Duval, A., Reperant, M., Vaury, C., Furlan, D., Leroy, K., et al. (2002). Evaluation of Tumor Microsatellite Instability Using Five Quasimonomorphic Mononucleotide Repeats and Pentaplex PCR. *Gastroenterology* 123, 1804–1811. doi:10.1053/gast.2002.37070

Takehara, Y., Nagasaka, T., Nyuya, A., Haruma, T., Haraga, J., Mori, Y., et al. (2018). Accuracy of Four Mononucleotide-Repeat Markers for the Identification of DNA Mismatch-Repair Deficiency in Solid Tumors. *J. Transl. Med.* 16, 5. doi:10.1186/s12967-017-1376-4

Tanioka, M., Fan, C., Parker, J. S., Hoadley, K. A., Hu, Z., Li, Y., et al. (2018). Integrated Analysis of RNA and DNA from the Phase III Trial CALGB 40601 Identifies Predictors of Response to Trastuzumab-Based Neoadjuvant Chemotherapy in HER2-Positive Breast Cancer. *Clin. Cancer Res.* 24, 5292–5304. doi:10.1158/1078-0432.CCR-17-3431

Tkachev, V., Sorokin, M., Garazha, A., Borisov, N., and Buzdin, A. (2020). "Oncobox Method for Scoring Efficiencies of Anticancer Drugs Based on Gene Expression Data," in *Methods in Molecular Biology* (Totowa: Humana Press), 235–255. doi:10.1007/978-1-0716-0138-9_17

Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015). Review the Cancer Genome Atlas (TCGA): an Immeasurable Source of Knowledge. *wo* 1A, 68–77. doi:10.5114/wo.2014.47136

Waalkes, A., Smith, N., Penewit, K., Hempelmann, J., Konnick, E. Q., Hause, R. J., et al. (2018). Accurate Pan-Cancer Molecular Diagnosis of Microsatellite Instability by Single-Molecule Molecular Inversion Probe Capture and High-Throughput Sequencing. *Clin. Chem.* 64, 950–958. doi:10.1373/clinchem.2017.285981

Wang, Y., Tong, Z., Zhang, W., Zhang, W., Buzdin, A., Mu, X., et al. (2021). FDA-approved and Emerging Next Generation Predictive Biomarkers for Immune Checkpoint Inhibitors in Cancer Patients. *Front. Oncol.* 11, 683419. doi:10.3389/fonc.2021.683419

Yamamoto, H., and Imai, K. (2019). An Updated Review of Microsatellite Instability in the Era of Next-Generation Sequencing and Precision Medicine. *Semin. Oncol.* 46, 261–270. doi:10.1053/j.seminoncol.2019.08.003

Conflict of Interest: Authors MSo and AB were employed by the company OmicsWay Corp. Authors VE, AS, and DN were employed by the company Oncobox 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: 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 Molecular Biosciences

MINI REVIEW
published: 05 January 2022
doi: 10.3389/fmolb.2021.799294

The Impact of Oxidative Stress in Male Infertility

Amanda Mannucci[1], *Flavia Rita Argento*[1], *Eleonora Fini*[1], *Maria Elisabetta Coccia*[2], *Niccolò Taddei*[1], *Matteo Becatti*[1]*† *and Claudia Fiorillo*[1]†

[1]Department of Experimental and Clinical Biomedical Sciences "Mario Serio", University of Florence, Florence, Italy, [2]Assisted Reproductive Technology Centre, Careggi Hospital, University of Florence, Florence, Italy

At present infertility is affecting about 15% of couples and male factor is responsible for almost 50% of infertility cases. Oxidative stress, due to enhanced Reactive Oxygen Species (ROS) production and/or decreased antioxidants, has been repeatedly suggested as a new emerging causative factor of this condition. However, the central roles exerted by ROS in sperm physiology cannot be neglected. On these bases, the present review is focused on illustrating both the role of ROS in male infertility and their main sources of production. Oxidative stress assessment, the clinical use of redox biomarkers and the treatment of oxidative stress-related male infertility are also discussed.

Keywords: oxidative stress, ROS, reactive oxygen species, male infertility, spermatozoa, semen parameters

OPEN ACCESS

Edited by:
Mahendra Pratap Kashyap, University of Alabama at Birmingham, United States

Reviewed by:
Arianna Vignini, Marche Polytechnic University, Italy

***Correspondence:**
Matteo Becatti
matteo.becatti@unifi.it

†*These authors share senior authorship*

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *21 October 2021*
Accepted: *16 December 2021*
Published: *05 January 2022*

Citation:
Mannucci A, Argento FR, Fini E, Coccia ME, Taddei N, Becatti M and Fiorillo C (2022) The Impact of Oxidative Stress in Male Infertility. Front. Mol. Biosci. 8:799294. doi: 10.3389/fmolb.2021.799294

INTRODUCTION

Infertility is a multifactorial disease affecting 15% of couples and defined as the inability to achieve spontaneous pregnancy after 12 months or more of regular unprotected sexual intercourse (Sharlip et al., 2002; Aitken, 2020). Male factor is responsible for almost 50% of infertility cases, contributing equally as female factor (Athayde et al., 2007; Wagner et al., 2018). Male infertility diagnosis is commonly based on standard semen parameters analysis (Nallella et al., 2006), according to the WHO guidelines, nevertheless, a large proportion of infertile males does not receive a clear diagnosis, considering them as idiopathic or unexplained cases (Sharlip et al., 2002).

Many studies suggested oxidative stress, a condition characterized by an imbalance between reactive oxygen species (ROS) production and antioxidant defence systems, as a new emerging factor in unexplained male infertility (Saleh and Agarwal, 2002; Makker et al., 2009; Agarwal et al., 2018; Cito et al., 2020).

At physiological levels, ROS are associated with the development of sperm fertilization properties, promoting chromatin compaction in maturing spermatozoa, motility, chemotaxis, sperm capacitation, hyperactivation, acrosome reaction and oocyte interaction (Kothari et al., 2010; Du Plessis et al., 2015). An excessive ROS production represents an important cause of sperm injury. Indeed, due to the large amount of membrane unsaturated fatty acids and the lack of cytoplasmic antioxidant enzymes, spermatozoa are highly susceptible to oxidation (Agarwal et al., 2017), with consequent detrimental effects on sperm quality/functioning (Aitken and Baker, 2006; Venkatesh et al., 2011; Barati et al., 2020).

Here, we discuss about the different roles of ROS on spermatozoa pathophysiology, paying particular attention to ROS effects on semen parameters. Finally, we focus on the available techniques to assess redox status in biological fluids and the clinical use of redox biomarkers for diagnosis and management of male infertility.

Oxidative Stress

Oxygen has a central role in life, displaying both beneficial and harmful effects on biological systems. The main oxygen involvement is in adenosine-5-triphosphate (ATP) generation via mitochondrial

oxidative phosphorylation (Burton and Jauniaux, 2011; Lushchak, 2014), a reaction also implicated in ROS and RNS production (Pham-Huy et al., 2008).

At moderate levels, ROS/RNS play an important role in regulating several intracellular signaling pathways, immune and mitogen responses and in maintaining cellular homeostasis (Kruk et al., 2019). On the contrary, higher ROS levels can be responsible for oxidative damages on proteins, lipids and nucleic acids (DNA, RNA), with harmful cellular effects. However, a complex system of antioxidant molecules has been evolved to maintain a redox balance and avoid biological system injury (Burton and Jauniaux, 2011; Kruk et al., 2019).

Several conditions (as environmental factors, excessive physical exercise, deficiencies in antioxidants, immune system dysfunctions, chronic disorders) may alter oxidant/antioxidant balance, leading to oxidative stress (Halliwell, 2007).

Oxidative stress mediates tissue injury and cell death, displaying a pathological role in several disorders including inflammation and aging, cardiovascular and neurodegenerative diseases, autoimmune disorders, cancer and reproductive system alterations (Burton and Jauniaux, 2011; Birben et al., 2012; Kruk et al., 2019).

The Physiological Role of ROS in Spermatozoa

Physiologically, ROS are considered regulators of several intracellular pathways, modulating the activation of different transcription factors (Burton and Jauniaux, 2011). ROS stimulate cyclic adenosine monophosphate (cAMP) in sperms, promoting tyrosine phosphorylation by tyrosine phosphatase inhibition (Wagner et al., 2018). This molecular mechanism results in the activation of several transcription factors involved in intracellular signaling cascades for sperm physiology. Indeed, several studies showed that higher ROS levels stimulate sperm capacitation and hyperactivation, acrosome reaction, motility and chemotaxis and chromatin compaction in maturing spermatozoa (Du Plessis et al., 2015; Wagner et al., 2018). Furthermore, ROS can improve sperm capacity of binding to the zona pellucida, inducing sperm-oocyte fusion (Wagner et al., 2018). By the way, antioxidant molecules may alter spermatozoa maturation, interfering with physiological sperm function. Particularly, it was showed that catalase or superoxide dismutase (SOD) inhibit sperm capacitation or acrosome reaction, supporting the evidence of the central involvement of ROS in spermatozoa functioning (Wagner et al., 2018).

The Pathological Role of ROS in Spermatozoa

Besides to the physiological role of ROS, excessive ROS generation and oxidative stress seem to be associated with harmful effects on spermatozoa, resulting in morphological and dynamic cellular properties alterations and finally in lower fertilization ability.

During recent years, a growing literature has shown that an altered redox balance in seminal fluid may display deleterious effects on sperm homeostasis, leading to male infertility (Agarwal et al., 2008; Makker et al., 2009; Agarwal et al., 2014b; Sabeti et al., 2016; Agarwal et al., 2017; Agarwal et al., 2018; Majzoub and Agarwal, 2018).

Blood and plasma redox status alterations have been reported in infertile men, as recently described in a study (Cito et al., 2020) showing higher blood leukocytes ROS production, increased plasma lipid peroxidation (LPO) and reduced plasma total antioxidant capacity (TAC) in oligoasthenozoospermic men compared to healthy subjects (Cito et al., 2020). In line with this, several findings also suggest that ROS-mediated sperm oxidation may induce cellular dysfunctions, affecting spermatozoa concentration, total number and motility (Agarwal et al., 2008; Agarwal et al., 2014b).

Spermatozoa are particularly susceptible to ROS-induced oxidation due to the presence, in their plasma membrane, of elevated levels of polyunsaturated fatty acids as docosahexaenoic acid containing six double bonds per molecule (Aitken et al., 2014). Indeed, ROS mediate the hydrogen abstraction from the hydrocarbon side-chain of a fatty acid, yielding to a carbon-centered lipid radical (L·) whose interaction with oxygen produces a lipid peroxyl radical (LOO·), able to react with an adjacent fatty acid propagating the process. Following internal molecular rearrangements conjugated dienes and hydroperoxides are generated (Phaniendra et al., 2015; Yoshida et al., 2015).

LPO products can also react with proteins, DNA and phospholipids, generating end-products involved in cellular dysfunction. Particularly, the interaction of LPO products with amino residues can result in protein oxidation, affecting protein structural and functional features (Niki, 2014). In this context, it was observed that LPO products as 4-hydroxy-2-nonenal (4HNE) are able to propagate ROS generation via interaction with proteins of the sperm mitochondrial electron transport chain (Aitken et al., 2014).

Lipid peroxidation is strictly associated with fluidity and permeability membrane alterations, inhibition of membrane-bound enzymes and receptors and activation of apoptotic cascade, supporting oxidative stress involvement in motility and morphology sperms abnormalities (Nowicka-Bauer and Nixon, 2020). Among LPO products, 4HNE seems to be highly responsible for cytotoxic effects on cellular sperm membrane, inducing loss of membrane integrity, motility alterations and compromising sperm-oocyte interactions (Baker et al., 2015; Walters et al., 2018; Nowicka-Bauer and Nixon, 2020) It was observed that 4HNE-mediated effects depends on several factors: cellular differentiation status, amount of substrates for 4HNE attack and antioxidant defense systems (Walters et al., 2018).

ROS can also affect sperm functioning by post-translational oxidative protein modifications (Salvolini et al., 2012; Morielli and O'Flaherty, 2015). The important association between protein oxidation markers, as three nitro-tyrosines (3NT), and sperm motility and morphology in oligoasthenoteratospermia has been reported (Kalezic et al., 2018). In particular, signs of sperm protein S-glutathionylation and tyrosine nitration were

found in infertile men (Salvolini et al., 2012; Morielli and O'Flaherty, 2015). Accordingly, higher peroxynitrite levels in human asthenozoospermic sperm samples, emphasizing their negative impact on sperm motility through the formation of three nitro-tyrosines were reported (Vignini et al., 2006).

Several investigations observed that not all sperm proteins are equally susceptible to ROS or to lipid aldehydes (Nowicka-Bauer and Nixon, 2020). The principal 4HNE target proteins are represented by metabolic enzymes, involved in bioenergetic pathways needed for sperm motility (Moscatelli et al., 2019). Several proteomics studies have been performed on infertile men spermatozoa in this context. A downregulation of proteins involved in bioenergetic pathways in altered spermatozoa of asthenozoospermic men was revealed (Amaral et al., 2014; Moscatelli et al., 2019). Particularly, some authors observed alterations in proteins associated with metabolic pathways as glycolysis, pyruvate metabolism, TCA or beta-oxidation in asthenozoospermic men, supporting that oxidative stress compromises sperm functionality by altering bioenergetic pathways (Elkina et al., 2011; Guo et al., 2019).

It is traditionally accepted that nucleic acids represent another crucial target of oxidative stress. Both nuclear and mitochondrial DNA are vulnerable to hydroxyl radical (OH.) attack, leading to the formation of several biomarkers of oxidative stress. OH. can react with guanine to produce 8-hydroxy-2′-deoxyguanosine (8-OH-G), an important marker of DNA oxidative damage, detectable in several biological samples (Burton and Jauniaux, 2011).

The lack of adequate antioxidant systems makes spermatozoa highly susceptible to DNA oxidation (Agarwal et al., 2003; Aitken et al., 2014). Sperm DNA oxidation is also due to the lack of complete DNA repair strategies in spermatozoa. Indeed, if the 8-oxoguanine glycosylase (OGG1) is able to remove the 8OHdG residue from DNA producing an abasic site, sperms do not possess any base excision repair system for the insertion of a new base (Aitken et al., 2014).

Several studies indicated that ROS generation is associated with DNA fragmentation and poor chromatin packaging, promoting apoptosis with relevant consequences on sperm count (Aitken et al., 2014; Iommiello et al., 2015). Patients with asthenozoospermia show enhanced mtDNA copy number and reduced mtDNA integrity that are associated with higher ROS generation (Bonanno et al., 2016). Accordingly, other reports underlined the significant association between NO and 8-OHdG levels and semen parameters abnormalities (Gholinezhad et al., 2020), supporting redox status assessment for helping male infertility diagnosis and monitoring.

MAIN SOURCES OF ROS

It is largely accepted that several exogenous factors may contribute to inflammation and redox status alterations, promoting male infertility. Environmental pollution, lifestyle factors as smoke, alcohol, obesity, varicocele, bacterial/viral infections, microorganism mutations or sexual transmitted disorders are actively involved (Iommiello et al., 2015; Agarwal et al., 2018).

However, seminal fluid oxidative stress is mostly due to leukocytes -that produce 1,000 more times ROS than normal spermatozoa- and to immature spermatozoa (Agarwal et al., 2003; Iommiello et al., 2015; Agarwal et al., 2018).

Leukocytospermia. According to WHO guidelines, leukocytospermia, defined as peroxidase-positive leukocytes concentration >1 × 106 per mL of semen, has been found in about 10–20% of infertile men (Saleh et al., 2002; Agarwal et al., 2018). Granulocytes and macrophages are the main cellular types found in the ejaculate and are responsible for ROS generation which is largely associated to glucose-6-phosphate dehydrogenase (G6PDH) activity, producing high amount of NADPH that, in turn, strongly stimulates NADPH oxidase, one of the major ROS sources (Agarwal et al., 2003; Agarwal et al., 2018). New emerging observations revealed that seminal WBC could improve sperm ability to generate ROS in a direct manner or by soluble products released in sperm microenvironment (Saleh et al., 2002). However, the clinical significance of leukocytospermia and its role in sperm quality is still under debate.

Higher seminal WBC levels were observed in infertile men compared to healthy controls and leukocytospermia was significantly correlated with alterations in sperm number, motility and morphology (Wolff, 1995). Moreover, *in vitro* experiments showed that WBC damaged sperm function and hamster ovum penetration, representing important prognostic factors for Assisted Reproductive Technologies (ART) success rate (Wolff, 1995). In line with this evidence, further investigations supported WBC as a trigger factor for spermatozoa ROS generation, leading to reduced sperm quality and sperm DNA damage (Saleh et al., 2002; Agarwal et al., 2014a). Leukocytospermia was associated with alterations in sperm concentration, motility and morphology in leukocytospermic patients respect to nonleukospermic patients or healthy subjects. *In vitro* experiments also underlined that ROS levels remained increased in pure sperms suspensions of leukocytospermic patients also after WBC removal or phorbol 12-myristate 13-acetate (PMA)-induced ROS stimulation. Similar results were obtained after sperm incubation with WBC (Saleh et al., 2002).

Moreover, semen WBC, even at low concentrations, resulted positively correlated with oxidative stress, suggesting that semen WBC removal could be useful to reduce oxidative stress in samples used for ART (Sharma et al., 2001; Agarwal et al., 2014a).

Immature spermatozoa. When spermatogenesis is defective, alterations in cytoplasmic extrusion mechanisms are observed and spermatozoa are released with an excess of residual cytoplasm (cytoplasmic droplets) (Agarwal et al., 2003). Immature spermatozoa are associated with higher ROS generation, via G6PDH and higher creatine phosphokinase (CK) levels (Cayli et al., 2004). The clinical significance of CK in sperm maturity and quality is controversial (Hallak et al., 2001; Cayli et al., 2004; Muratori et al., 2015). Some reports described higher CK levels in oligozoospermic men than in healthy subjects and a significant association between CK levels and semen

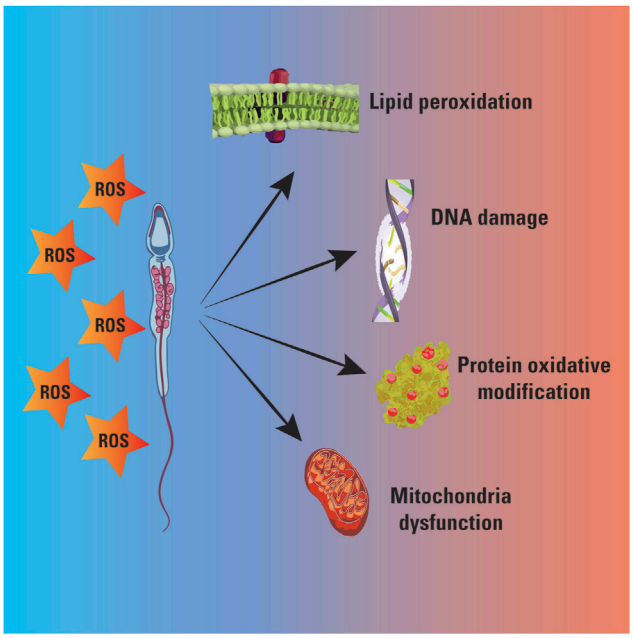


FIGURE 1 | Oxidative stress negatively affects sperm cells causing mitochondrial injury and alterations in lipids, nucleic acids and proteins.

parameters (concentration, motility and morphology), suggesting this marker as a good predictor of sperm quality in the follow-up of patients treated for male infertility (Hallak et al., 2001). Other authors found no difference in CK amount between cells with or without DNA fragmentation, showing no involvement of immature spermatozoa in DNA damage (Muratori et al., 2015). In this context, it was observed that spermatozoa at different stages of maturation are characterized by variations in ROS levels, membrane lipid content, chromatin compaction, morphology and motility. Immature spermatozoa showed higher ROS generation and DNA damage and could be considered an important cause of male infertility, inducing oxidation in mature sperm cells during sperm migration from the seminiferous tubules to the epididymis (Ollero et al., 2001).

Mithocondria. Another potential ROS source in spermatozoa is represented by mitochondria. Indeed, factors as electromagnetic radiation, polyunsaturated fatty acids or apoptotic factors may alter the electron transport chain on mitochondrial membrane, resulting in excessive ROS generation. Several reports indicate sperm mitochondrial dysfunction and oxidative stress as potential factors involved in asthenozoospermia (Nowicka-Bauer et al., 2018). Particularly, interferences in the mitochondrial electron flow at complexes I and III may trigger ROS generation and cause sperm tail oxidation, leading to DNA damage and motility aberrations (Koppers et al., 2008). Sperm mitochondrial dysfunctions enhance ROS production and are associated with sperm quality impairment and loss of fertilization potential. Particularly, a significant correlation between sperm mitochondrial functioning and sperm motility was reported (Cassina et al., 2015).

OXIDATIVE STRESS EFFECTS ON SEMEN PARAMETERS

During these years, the potential correlation between spermatozoa ROS production and semen parameters has been largely investigated (Athayde et al., 2007; Hosseinzadeh Colagar et al., 2013; Agarwal et al., 2014a; Bonanno et al., 2016; Aitken, 2017; Dobrakowski et al., 2017; Dorostghoal et al., 2017). The detrimental effects of ROS on sperm motility and morphology has been repeatedly reported. *In vitro* experiments demonstrated that lipid aldheydes addiction to spermatozoa promoted loss motility in human sperm cells (Agarwal et al., 2014a) (**Figure 1**).

Accordingly, seminal fluid LPO and TAC levels were significantly correlated with sperm motility, morphology and sperm count in astheno- and oligoastheno-teratospermic men (Khosrowbeygi and Zarghami, 2007; Hosseinzadeh Colagar et al., 2013).

The key role of oxidative stress in spermatozoa alterations is also supported by evidence of beneficial effects of therapeutic supplementation with antioxidants on semen quality in infertile men (Gambera et al., 2019). In particular, therapeutic Coenzyme Q10 treatment improved semen parameters (sperm concentration and motility), redox status and sperm DNA fragmentation in idiopathic male infertility (Alahmar et al., 2021). Interestingly, an improvement in sperm concentration and motility after vitamin D supplementation in vitamin D deficient infertile male with oligoasthenozoospermia was observed (Wadhwa et al., 2020). The positive effects of an antioxidant therapy (Gambera et al., 2019) on semen quality has been suggested as a useful tool to improve successful conception rate in patients with oligoasthenozoospermia undergoing intracytoplasmic sperm injection (ICSI).

On the contrary, other authors reported no correlation between ROS levels and sperm motility, underling that it is still unclear if reduced sperm functional performances are due to lower sperm number or to a direct ROS effect (Whittington et al., 1999).

In this context, the usefulness of a new blood diagnostic tool to evaluate sperm morphological and/or functional abnormalities, supporting male infertility diagnosis and management, is increasingly evident.

In this regards, blood SOD and GSH levels were found to positively correlate with sperm count and motility, while enhanced MDA levels were associated with altered sperm morphology (Shamsi et al., 2010). In line with this, signs of oxidative stress in seminal fluid and reduced plasma TAC in infertile men were described. Particularly, plasma TAC significantly and positively correlated both with seminal fluid TAC and with semen parameters (Benedetti et al., 2012), indicating that plasma redox status reflects the redox status of seminal fluid microenvironment and sperm quality.

In agreement, it has been shown that higher MDA and Nitric Oxide (NO) levels in plasma and seminal fluid of infertile men correlated with semen parameters, supporting that blood redox status is associated with semen parameters (Taken et al., 2016).

However, reports about the existing association between blood and seminal fluid oxidative stress are still limited and

controversial, potentially due to different strategies and applied methodologies. Indeed, no correlation was found between blood and seminal fluid oxidative status, suggesting the independence of seminal fluid redox homeostasis from systemic microenvironment and external factors (Guz et al., 2013).

OXIDATIVE STRESS ASSESSMENT AND CLINICAL USE OF REDOX BIOMARKERS IN MALE INFERTILITY

The analysis of semen parameters according to the WHO guidelines represents, currently, the gold standard for male infertility diagnosis. However, several studies showed that ROS-induced sperm oxidation can result in sperm quality alterations, leading to a decrease in sperm fertilizing potential (Agarwal and Majzoub, 2017; Dutta et al., 2019; Martins and Agarwal, 2019). Based on this evidence, new tests aimed to evaluate male fertility by monitoring oxidative stress status are needed.

Assays for oxidative stress detection may suggest new biochemical approaches to improve male infertility diagnosis and management, using simple, fast and less expensive techniques (Agarwal and Majzoub, 2017; Agarwal et al., 2019).

Oxidative stress can be evaluated in different biological samples (plasma, serum, urine, follicular/peritoneal/seminal fluid), obtaining an accurate picture of redox status and eventually planning a therapeutic supplementation with antioxidants where it's needed.

Different oxidative stress assays exist, focusing on ROS generation, lipid peroxidation products and total antioxidant capacity. ROS measurement in semen include different methods as chemiluminescence, nitro blue tetrazolium (NBT) test, cytochrome c reduction test and electron spin resonance (Dutta et al., 2019; Martins and Agarwal, 2019).

However, several reports underlined the central use of cytometry to assess intracellular ROS production in blood cells as erythrocytes and leucocytes, in spermatozoa as well as in other cellular categories by incubating cells with the fluorescent probe H2DCF-DA (2.5 μM) (Invitrogen, Carlsbad, CA, United States) (Becatti et al., 2016a; Becatti et al., 2017a; Becatti et al., 2018; Cito et al., 2020). Due to its susceptibility to ROS-induced oxidation by hydrogen peroxide, peroxynitrite, hydroxyl radicals and also by superoxide anions, H2DCF-DA is now considered among the principal methods for measuring intracellular ROS levels, sensing redox status variations and cellular oxidative stress (Eruslanov and Kusmartsev, 2010).

Cytofluorimetric analysis can be also employed for the assessment of membrane lipid oxidation, using the fluorescent probe BODIPY 581/591 C11. This approach was proposed to investigate redox status alterations both in erythrocytes of RVO (Becatti et al., 2016b) and SSNHL (Becatti et al., 2017b) patients.

Moreover, fluorescent anisotropy of cellular membranes, a new method to evaluate membrane fluidity, could be a future innovation for further investigations about sperm motility defects (Becatti et al., 2016b; Becatti et al., 2017b).

Oxidative stress assessment is also performed by evaluating LPO and TAC levels in biological fluids (Martins and Agarwal, 2019). LPO levels can be detected by measuring lipid oxidation end products as MDA, 4HNE, isoprostanes with spectrofluorimetric or immunochemical assays (Agarwal and Majzoub, 2017). Thiobarbituric Acid (TBA) Assay or the ALDETECT Assay are the mostly used tests for LPO assessment. Highly sensitive high pressure liquid chromatography (HPLC) is promoted for low MDA concentrations (Grotto et al., 2007), whereas commercial immunoassays or mass spectrometry represent an alternative method to evaluate lipid peroxidation end products as isoprostanes (Morrow, 2005).

Parallelly, TAC level can be measured using enhanced-chemiluminescence or colorimetric techniques (Martins and Agarwal, 2019). Among chemiluminescent methodologies, Oxygen Radical Absorbance Capacity (ORAC) Assay is based on the intensity fluorescence decay of a fluorescent probe, fluorescein, consequent to its oxidation by free radical species (particularly peroxyl radical), generated after the thermal decomposition of 2,2′-azobis (2-amidinopropane) dihydrochloride (AAPH) azo-compound. Colorimetric methods evaluate the antioxidant capacity of samples to inhibit the oxidation of 2,2′-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid) (ABTS) to ABTS + by metmyoglobin (Martins and Agarwal, 2019).

Based on redox biomarkers alterations in infertile men, several studies emphasized the elaboration of a specific global parameter/index able to discriminate fertile from infertile men better than ROS or TAC alone (Roychoudhury et al., 2016). Particularly, ROS-TAC score, derived from both ROS levels and antioxidant capacity in a given set of patients, was proposed as a new tool to investigate redox status in male infertility. Infertile men with male factor or idiopathic diagnoses had significantly different ROS-TAC scores than controls (Sharma et al., 1999; Vatannejad et al., 2017). Particularly, the potential use of ROS-TAC score for predicting the oxidative damage of semen samples in asthenozoospermic men was proposed (Vatannejad et al., 2017).

New emerging data have also shown oxidation reduction potential (ORP) measurement as a new fast, easy and reproducible method to assess oxidative stress in seminal fluid (Agarwal and Majzoub, 2017; Martins and Agarwal, 2019).

ORP indicates the ratio between oxidant and antioxidant molecules, evaluating the potential for electrons to move from a chemical specie to another. ORP is assessed by MiOSYS test, that measures electron transfer from antioxidants to oxidants in presence of a low voltage reducing current. The obtained data represent oxidant and antioxidant activity in a sample: particularly, high ORP levels indicate enhanced oxidant activity and therefore a condition of oxidative stress (Agarwal and Majzoub, 2017).

Some evidence reports a good association between ORP level and semen parameters (Agarwal et al., 2017; Majzoub et al., 2018; Homa et al., 2019) being found higher ORP levels in infertile men than in healthy controls (Agarwal et al., 2017). Moreover, a negative correlation between ORP value and semen parameters (sperm concentration and total count, motility and morphology)

was observed suggesting ORP as a further predictor for male infertility diagnosis and management (Agarwal et al., 2017). In line with this, further investigations confirmed the significant association of ORP both with semen parameters and DNA fragmentation in infertile men (Majzoub et al., 2018; Homa et al., 2019). Importantly, it was also shown that ORP is a more accurate tool for investigating redox status in male infertility than chemiluminescent ROS assessment (Homa et al., 2019).

TREATMENT OF OXIDATIVE STRESS RELATED MALE INFERTILITY

Currently, defined guidelines for treatment of oxidative stress-related male infertility are still lacking, in partly due to the unknown etiology of this condition (Agarwal et al., 2019). However, during these years several clinical trials have been developed to investigate the effects of antioxidant supplementation (as L-carnitine, selenium, Coenzyme Q10, ubiquinol, vitamin C and E) on seminal fluid oxidative stress and semen parameters (Majzoub and Agarwal, 2018; Dutta et al., 2019). Many of them reported promising effects of antioxidants on sperm concentration, motility, morphology and DNA fragmentation (Gambera et al., 2019; Alahmar et al., 2021). Twenty clinical trials focused on antioxidant therapy effects on seminal oxidative stress were analyzed. Nineteen of them revealed an improvement in sperm redox status and semen parameters and a good correlation with pregnancy outcome (Gharagozloo and Aitken, 2011).

However, the role of antioxidant therapy in male infertility is still controversial. In a randomized clinical trial, it was showed that 3 months of antioxidant treatment did not improve semen parameters and DNA fragmentation in infertile men and no beneficial effect on pregnancy or live birth rates was observed (Steiner et al., 2020).

These observations indicate that evidence to support the use of antioxidants in male infertility are still uncertain. However, traditional semen analysis together with oxidative stress assessment display a great potential to perform accurate evaluation of infertile patients (Agarwal et al., 2019).

CONCLUSION

Oxidative stress is centrally involved in sperm dysfunctions and represent a new pathological mechanism of male infertility (Agarwal et al., 2008; Hosseinzadeh Colagar et al., 2013; Agarwal et al., 2018). Based on previously reported investigations and results, new methods and diagnostic approaches for male reproductive disorders are needed. Together with seminal fluid oxidative stress assessment, blood redox status monitoring and leukocytes ROS levels, could represent a new potential and less invasive practice for clinicians to evaluate sperm cells quality and fertilization ability. The considered redox parameters may therefore be useful to develop new therapeutic strategies based on antioxidant supplementation in order to reduce systemic oxidative stress in infertile men, improving male infertility diagnosis and management and ART success rate.

AUTHOR CONTRIBUTIONS

AM, FA, EF, MB, and CF conceived the structure of the manuscript and drafted the manuscript. MC, NT, MB, and CF critically revised the manuscript. All authors approved the final version of the manuscript.

FUNDING

This study was supported by "Fondi di Ateneo" research funding from the University of Florence to MB (MATTEOBECATTIRICATEN21).

REFERENCES

Agarwal, A., and Majzoub, A. (2017). Laboratory Tests for Oxidative Stress. *Indian J. Urol.* 33 (3), 199–206. doi:10.4103/iju.IJU_9_17

Agarwal, A., Saleh, R. A., and Bedaiwy, M. A. (2003). Role of Reactive Oxygen Species in the Pathophysiology of Human Reproduction. *Fertil. Sterility* 79 (4), 829–843. doi:10.1016/s0015-0282(02)04948-8

Agarwal, A., Makker, K., and Sharma, R. (2008). Review Article: Clinical Relevance of Oxidative Stress in Male Factor Infertility: An Update. *Am. J. Reprod. Immunol.* 59 (1), 2–11. doi:10.1111/j.1600-0897.2007.00559.x

Agarwal, A., Mulgund, A., Alshahrani, S., Assidi, M., Abuzenadah, A. M., Sharma, R., et al. (2014). Reactive Oxygen Species and Sperm DNA Damage in Infertile Men Presenting with Low Level Leukocytospermia. *Reprod. Biol. Endocrinol.* 12, 126. doi:10.1186/1477-7827-12-126

Agarwal, A., Virk, G., Ong, C., and du Plessis, S. S. (2014). Effect of Oxidative Stress on Male Reproduction. *World J. Mens Health* 32 (1), 1–17. doi:10.5534/wjmh.2014.32.1.1

Agarwal, A., Roychoudhury, S., Sharma, R., Gupta, S., Majzoub, A., and Sabanegh, E. (2017). Diagnostic Application of Oxidation-Reduction Potential Assay for Measurement of Oxidative Stress: Clinical Utility in Male Factor Infertility. *Reprod. BioMedicine Online* 34 (1), 48–57. doi:10.1016/j.rbmo.2016.10.008

Agarwal, A., Rana, M., Qiu, E., AlBunni, H., Bui, A. D., and Henkel, R. (2018). Role of Oxidative Stress, Infection and Inflammation in Male Infertility. *Andrologia* 50 (11), e13126. doi:10.1111/and.13126

Agarwal, A., Parekh, N., Panner Selvam, M. K., Henkel, R., Shah, R., Homa, S. T., et al. (2019). Male Oxidative Stress Infertility (MOSI): Proposed Terminology and Clinical Practice Guidelines for Management of Idiopathic Male Infertility. *World J. Mens Health* 37 (3), 296–312. doi:10.5534/wjmh.190055

Aitken, R. J., and Baker, M. A. (2006). Oxidative Stress, Sperm Survival and Fertility Control. *Mol. Cell Endocrinol.* 250 (1-2), 66–69. doi:10.1016/j.mce.2005.12.026

Aitken, R., Smith, T., Jobling, M., Baker, M., and De Iuliis, G. (2014). Oxidative Stress and Male Reproductive Health. *Asian J. Androl.* 16 (1), 31–38. doi:10.4103/1008-682X.122203

Aitken, R. J. (2017). Reactive Oxygen Species as Mediators of Sperm Capacitation and Pathological Damage. *Mol. Reprod. Dev.* 84 (10), 1039–1052. doi:10.1002/mrd.22871

Aitken, R. J. (2020). Impact of Oxidative Stress on Male and Female Germ Cells: Implications for Fertility. *Reproduction* 159 (4), R189–R201. doi:10.1530/REP-19-0452

Alahmar, A. T., Calogero, A. E., Sengupta, P., and Dutta, S. (2021). Coenzyme Q10 Improves Sperm Parameters, Oxidative Stress Markers and Sperm DNA Fragmentation in Infertile Patients with Idiopathic Oligoasthenozoospermia. *World J. Mens Health* 39 (2), 346–351. doi:10.5534/wjmh.190145

Amaral, A., Paiva, C., Attardo Parrinello, C., Estanyol, J. M., Ballescà, J. L., Ramalho-Santos, J., et al. (2014). Identification of Proteins Involved in Human Sperm Motility Using High-Throughput Differential Proteomics. *J. Proteome Res.* 13 (12), 5670–5684. doi:10.1021/pr500652y

Athayde, K. S., Cocuzza, M., Agarwal, A., Krajcir, N., Lucon, A. M., Srougi, M., et al. (2007). Development of normal Reference Values for Seminal Reactive Oxygen Species and Their Correlation with Leukocytes and Semen Parameters in a fertile Population. *J. Androl.* 28 (4), 613–620. doi:10.2164/jandrol.106.001966

Baker, M. A., Weinberg, A., Hetherington, L., Villaverde, A.-I., Velkov, T., Baell, J., et al. (2015). Defining the Mechanisms by Which the Reactive Oxygen Species By-Product, 4-Hydroxynonenal, Affects Human Sperm Cell Function1. *Biol. Reprod.* 92 (4), 108. doi:10.1095/biolreprod.114.126680

Barati, E., Nikzad, H., and Karimian, M. (2020). Oxidative Stress and Male Infertility: Current Knowledge of Pathophysiology and Role of Antioxidant Therapy in Disease Management. *Cell. Mol. Life Sci.* 77 (1), 93–113. doi:10.1007/s00018-019-03253-8

Becatti, M., Emmi, G., Silvestri, E., Bruschi, G., Ciucciarelli, L., Squatrito, D., et al. (2016). Neutrophil Activation Promotes Fibrinogen Oxidation and Thrombus Formation in Behçet Disease. *Circulation* 133 (3), 302–311. doi:10.1161/CIRCULATIONAHA.115.017738

Becatti, M., Marcucci, R., Gori, A. M., Mannini, L., Grifoni, E., Alessandrello Liotta, A., et al. (2016). Erythrocyte Oxidative Stress Is Associated with Cell Deformability in Patients with Retinal Vein Occlusion. *J. Thromb. Haemost.* 14 (11), 2287–2297. doi:10.1111/jth.13482

Becatti, M., Mannucci, A., Barygina, V., Mascherini, G., Emmi, G., Silvestri, E., et al. (2017). Redox Status Alterations during the Competitive Season in Élite Soccer Players: Focus on Peripheral Leukocyte-Derived ROS. *Intern. Emerg. Med.* 12 (6), 777–788. doi:10.1007/s11739-017-1653-5

Becatti, M., Marcucci, R., Mannucci, A., Gori, A., Giusti, B., Sofi, F., et al. (2017). Erythrocyte Membrane Fluidity Alterations in Sudden Sensorineural Hearing Loss Patients: The Role of Oxidative Stress. *Thromb. Haemost.* 117 (12), 2334–2345. doi:10.1160/TH17-05-0356

Becatti, M., Fucci, R., Mannucci, A., Barygina, V., Mugnaini, M., Criscuoli, L., et al. (2018). A Biochemical Approach to Detect Oxidative Stress in Infertile Women Undergoing Assisted Reproductive Technology Procedures. *Int. J. Mol. Sci.* 19 (2), 592. doi:10.3390/ijms19020592

Benedetti, S., Tagliamonte, M. C., Catalani, S., Primiterra, M., Canestrari, F., Stefani, S. D., et al. (2012). Differences in Blood and Semen Oxidative Status in fertile and Infertile Men, and Their Relationship with Sperm Quality. *Reprod. BioMedicine Online* 25 (3), 300–306. doi:10.1016/j.rbmo.2012.05.011

Birben, E., Sahiner, U. M., Sackesen, C., Erzurum, S., and Kalayci, O. (2012). Oxidative Stress and Antioxidant Defense. *World Allergy Organ. J.* 5 (1), 9–19. doi:10.1097/WOX.0b013e3182439613

Bonanno, O., Romeo, G., Asero, P., Pezzino, F. M., Castiglione, R., Burrello, N., et al. (2016). Sperm of Patients with Severe Asthenozoospermia Show Biochemical, Molecular and Genomic Alterations. *Reproduction* 152 (6), 695–704. doi:10.1530/REP-16-0342

Burton, G. J., and Jauniaux, E. (2011). Oxidative Stress. *Best Pract. Res. Clin. Obstet. Gynaecol.* 25 (3), 287–299. doi:10.1016/j.bpobgyn.2010.10.016

Cassina, A., Silveira, P., Cantu, L., Montes, J. M., Radi, R., and Sapiro, R. (2015). Defective Human Sperm Cells Are Associated with Mitochondrial Dysfunction and Oxidant Production1. *Biol. Reprod.* 93 (5), 119. doi:10.1095/biolreprod.115.130989

Cayli, S., Sakkas, D., Vigue, L., Demir, R., and Huszar, G. (2004). Cellular Maturity and Apoptosis in Human Sperm: Creatine Kinase, Caspase-3 and Bcl-XL Levels in Mature and Diminished Maturity Sperm. *Mol. Hum. Reprod.* 10 (5), 365–372. doi:10.1093/molehr/gah050

Cito, G., Becatti, M., Natali, A., Fucci, R., Picone, R., Cocci, A., et al. (2020). Redox Status Assessment in Infertile Patients with Non-obstructive Azoospermia Undergoing Testicular Sperm Extraction: A Prospective Study. *Andrologia* 8 (2), 364–371. doi:10.1111/andr.12721

Dobrakowski, M., Kasperczyk, S., Horak, S., Chyra-Jach, D., Birkner, E., and Kasperczyk, A. (2017). Oxidative Stress and Motility Impairment in the Semen of fertile Males. *Andrologia* 49 (10), e12783. doi:10.1111/and.12783

Dorostghoal, M., Kazeminejad, S. R., Shahbazian, N., Pourmehdi, M., and Jabbari, A. (2017). Oxidative Stress Status and Sperm DNA Fragmentation in fertile and Infertile Men. *Andrologia* 49 (10), e12762. doi:10.1111/and.12762

Du Plessis, S. S., Agarwal, A., Halabi, J., and Tvrda, E. (2015). Contemporary Evidence on the Physiological Role of Reactive Oxygen Species in Human Sperm Function. *J. Assist. Reprod. Genet.* 32 (4), 509–520. doi:10.1007/s10815-014-0425-7

Dutta, S., Majzoub, A., and Agarwal, A. (2019). Oxidative Stress and Sperm Function: A Systematic Review on Evaluation and Management. *Arab J. Urol.* 17 (2), 87–97. doi:10.1080/2090598X.2019.1599624

Elkina, Y. L., Atroshchenko, M. M., Bragina, E. E., Muronetz, V. I., and Schmalhausen, E. V. (2011). Oxidation of Glyceraldehyde-3-Phosphate Dehydrogenase Decreases Sperm Motility. *Biochem. Mosc.* 76 (2), 268–272. doi:10.1134/s0006297911020143

Eruslanov, E., and Kusmartsev, S. (2010). Identification of ROS Using Oxidized DCFDA and Flow-Cytometry. *Methods Mol. Biol.* 594, 57–72. doi:10.1007/978-1-60761-411-1_4

Gambera, L., Stendardi, A., Ghelardi, C., Fineschi, B., and Aini, R. (2019). Effects of Antioxidant Treatment on Seminal Parameters in Patients Undergoing *In Vitro* Fertilization. *Arch. Ital. Urol. Androl.* 91 (3), 187. doi:10.4081/aiua.2019.3.187

Gharagozloo, P., and Aitken, R. J. (2011). The Role of Sperm Oxidative Stress in Male Infertility and the Significance of Oral Antioxidant Therapy. *Hum. Reprod.* 26 (7), 1628–1640. doi:10.1093/humrep/der132

Gholinezhad, M., Aliarab, A., Abbaszadeh-Goudarzi, G., Yousefnia-Pasha, Y., Samadaian, N., Rasolpour-Roshan, K., et al. (2020). Nitric Oxide, 8-Hydroxydeoxyguanosine, and Total Antioxidant Capacity in Human Seminal Plasma of Infertile Men and Their Relationship with Sperm Parameters. *Clin. Exp. Reprod. Med.* 47 (1), 54–60. doi:10.5653/cerm.2020.00423

Grotto, D., Santa Maria, L. D., Boeira, S., Valentini, J., Charão, M. F., Moro, A. M., et al. (2007). Rapid Quantification of Malondialdehyde in Plasma by High Performance Liquid Chromatography-Visible Detection. *J. Pharm. Biomed. Anal.* 43 (2), 619–624. doi:10.1016/j.jpba.2006.07.030

Guo, Y., Jiang, W., Yu, W., Niu, X., Liu, F., Zhou, T., et al. (2019). Proteomics Analysis of Asthenozoospermia and Identification of Glucose-6-Phosphate Isomerase as an Important Enzyme for Sperm Motility. *J. Proteomics* 208, 103478. doi:10.1016/j.jprot.2019.103478

Guz, J., Gackowski, D., Foksinski, M., Rozalski, R., Zarakowska, E., Siomek, A., et al. (2013). Comparison of Oxidative Stress/DNA Damage in Semen and Blood of fertile and Infertile Men. *PLoS One* 8 (7), e68490. doi:10.1371/journal.pone.0068490

Hallak, J., Sharma, R. K., Pasqualotto, F. F., Ranganathan, P., Thomas, A. J., Jr, and Agarwal, A. (2001). Creatine Kinase as an Indicator of Sperm Quality and Maturity in Men with Oligospermia. *Urology* 58 (3), 446–451. doi:10.1016/s0090-4295(01)01224-9

Halliwell, B. (2007). Biochemistry of Oxidative Stress. *Biochem. Soc. Trans.* 35 (Pt 5), 1147–1150. doi:10.1042/BST0351147

Homa, S., Vassiliou, A., Stone, J., Killeen, A., Dawkins, A., Xie, J., et al. (2019). A Comparison between Two Assays for Measuring Seminal Oxidative Stress and Their Relationship with Sperm DNA Fragmentation and Semen Parameters. *Genes* 10 (3), 236. doi:10.3390/genes10030236

Hosseinzadeh Colagar, A., Karimi, F., and Jorsaraei, S. G. A. (2013). Correlation of Sperm Parameters with Semen Lipid Peroxidation and Total Antioxidants Levels in Astheno- and Oligoasheno- Teratospermic Men. *Iran Red Crescent Med. J.* 15 (9), 780–785. doi:10.5812/ircmj.6409

Iommiello, V. M., Albani, E., Di Rosa, A., Marras, A., Menduni, F., Morreale, G., et al. (2015). Ejaculate Oxidative Stress Is Related with Sperm DNA Fragmentation and Round Cells. *Int. J. Endocrinol.* 2015, 321901. doi:10.1155/2015/321901

Kalezic, A., Macanovic, B., Garalejic, E., Korac, A., Otasevic, V., and Korac, B. (2018). Level of NO/Nitrite and 3-Nitrotyrosine in Seminal Plasma of Infertile Men: Correlation with Sperm Number, Motility and Morphology. *Chemico-Biological Interactions* 291, 264–270. doi:10.1016/j.cbi.2018.07.002

Khosrowbeygi, A., and Zarghami, N. (2007). Levels of Oxidative Stress Biomarkers in Seminal Plasma and Their Relationship with Seminal Parameters. *BMC Clin. Pathol.* 7, 6. doi:10.1186/1472-6890-7-6

Koppers, A. J., De Iuliis, G. N., Finnie, J. M., McLaughlin, E. A., and Aitken, R. J. (2008). Significance of Mitochondrial Reactive Oxygen Species in the Generation of Oxidative Stress in Spermatozoa. *J. Clin. Endocrinol. Metab.* 93 (8), 3199–3207. doi:10.1210/jc.2007-2616

Kothari, S., Thompson, A., Agarwal, A., and du Plessis, S. S. (2010). Free Radicals: Their Beneficial and Detrimental Effects on Sperm Function. *Indian J. Exp. Biol.* 48 (5), 425–435.

Kruk, J., Aboul-Enein, H. Y., Kładna, A., and Bowser, J. E. (2019). Oxidative Stress in Biological Systems and its Relation with Pathophysiological Functions: The Effect of Physical Activity on Cellular Redox Homeostasis. *Free Radic. Res.* 53 (5), 497–521. doi:10.1080/10715762.2019.1612059

Lushchak, V. I. (2014). Free Radicals, Reactive Oxygen Species, Oxidative Stress and its Classification. *Chemico-Biological Interactions* 224, 164–175. doi:10.1016/j.cbi.2014.10.016

Majzoub, A., and Agarwal, A. (2018). Systematic Review of Antioxidant Types and Doses in Male Infertility: Benefits on Semen Parameters, Advanced Sperm Function, Assisted Reproduction and Live-Birth Rate. *Arab J. Urol.* 16 (1), 113–124. doi:10.1016/j.aju.2017.11.013

Majzoub, A., Arafa, M., Mahdi, M., Agarwal, A., Al Said, S., Al-Emadi, I., et al. (2018). Oxidation-reduction Potential and Sperm DNA Fragmentation, and Their Associations with Sperm Morphological Anomalies Amongst fertile and Infertile Men. *Arab J. Urol.* 16 (1), 87–95. doi:10.1016/j.aju.2017.11.014

Makker, K., Agarwal, A., and Sharma, R. (2009). Oxidative Stress & Male Infertility. *Indian J. Med. Res.* 129 (4), 357–367.

Martins, A. D., and Agarwal, A. (2019). Oxidation Reduction Potential: A New Biomarker of Male Infertility. *Panminerva Med.* 61 (2), 108–117. doi:10.23736/S0031-0808.18.03529-2

Morielli, T., and O'Flaherty, C. (2015). Oxidative Stress Impairs Function and Increases Redox Protein Modifications in Human Spermatozoa. *Reproduction* 149 (1), 113–123. doi:10.1530/REP-14-0240

Morrow, J. D. (2005). Quantification of Isoprostanes as Indices of Oxidant Stress and the Risk of Atherosclerosis in Humans. *Arterioscler Thromb. Vasc. Biol.* 25 (2), 279–286. doi:10.1161/01.ATV.0000152605.64964.c0

Moscatelli, N., Lunetti, P., Braccia, C., Armirotti, A., Pisanello, F., De Vittorio, M., et al. (2019). Comparative Proteomic Analysis of Proteins Involved in Bioenergetics Pathways Associated with Human Sperm Motility. *Int. J. Mol. Sci.* 20 (12), 3000. doi:10.3390/ijms20123000

Muratori, M., Tamburrino, L., Marchiani, S., Cambi, M., Olivito, B., Azzari, C., et al. (2015). Investigation on the Origin of Sperm DNA Fragmentation: Role of Apoptosis, Immaturity and Oxidative Stress. *Mol. Med.* 21 (1), 109–122. doi:10.2119/molmed.2014.00158

Nallella, K. P., Sharma, R. K., Aziz, N., and Agarwal, A. (2006). Significance of Sperm Characteristics in the Evaluation of Male Infertility. *Fertil. Sterility* 85 (3), 629–634. doi:10.1016/j.fertnstert.2005.08.024

Niki, E. (2014). Biomarkers of Lipid Peroxidation in Clinical Material. *Biochim. Biophys. Acta (Bba) - Gen. Subjects* 1840 (2), 809–817. doi:10.1016/j.bbagen.2013.03.020

Nowicka-Bauer, K., and Nixon, B. (2020). Molecular Changes Induced by Oxidative Stress that Impair Human Sperm Motility. *Antioxidants* 9 (2), 134. doi:10.3390/antiox9020134

Nowicka-Bauer, K., Lepczynski, A., Ozgo, M., Kamieniczna, M., Fraczek, M., Stanski, L., et al. (2018). Sperm Mitochondrial Dysfunction and Oxidative Stress as Possible Reasons for Isolated Asthenozoospermia. *J. Physiol. Pharmacol.* 69 (3), 403–417. doi:10.26402/jpp.2018.3.05

Ollero, M., Gil-Guzman, E., Lopez, M. C., Sharma, R. K., Agarwal, A., Larson, K., et al. (2001). Characterization of Subsets of Human Spermatozoa at Different Stages of Maturation: Implications in the Diagnosis and Treatment of Male Infertility. *Hum. Reprod.* 16 (9), 1912–1921. doi:10.1093/humrep/16.9.1912

Pham-Huy, L. A., He, H., and Pham-Huy, C. (2008). Free Radicals, Antioxidants in Disease and Health. *Int. J. Biomed. Sci.* 4 (2), 89–96.

Phaniendra, A., Jestadi, D. B., and Periyasamy, L. (2015). Free Radicals: Properties, Sources, Targets, and Their Implication in Various Diseases. *Ind. J. Clin. Biochem.* 30 (1), 11–26. doi:10.1007/s12291-014-0446-0

Roychoudhury, S., Sharma, R., Sikka, S., and Agarwal, A. (2016). Diagnostic Application of Total Antioxidant Capacity in Seminal Plasma to Assess Oxidative Stress in Male Factor Infertility. *J. Assist. Reprod. Genet.* 33 (5), 627–635. doi:10.1007/s10815-016-0677-5

Sabeti, P., Pourmasumi, S., Rahiminia, T., Akyash, F., and Talebi, A. R. (2016). Etiologies of Sperm Oxidative Stress. *Int. J. Reprod. Biomed.* 14 (4), 231–240. doi:10.29252/ijrm.14.4.231

Saleh, R. A., and Agarwal, A. (2002). Oxidative Stress and Male Infertility: from Research Bench to Clinical Practice. *J. Androl.* 23 (6), 737–752. doi:10.1002/j.1939-4640.2002.tb02324.x

Saleh, R. A., Agarwal, A., Kandirali, E., Sharma, R. K., Thomas, A. J., Nada, E. A., et al. (2002). Leukocytospermia Is Associated with Increased Reactive Oxygen Species Production by Human Spermatozoa. *Fertil. Sterility* 78 (6), 1215–1224. doi:10.1016/s0015-0282(02)04237-1

Salvolini, E., Buldreghini, E., Lucarini, G., Vignini, A., Di Primio, R., and Balercia, G. (2012). Nitric Oxide Synthase and Tyrosine Nitration in Idiopathic Asthenozoospermia: an Immunohistochemical Study. *Fertil. Sterility* 97 (3), 554–560. doi:10.1016/j.fertnstert.2011.12.022

Shamsi, M. B., Venkatesh, S., Kumar, R., Gupta, N. P., Malhotra, N., Singh, N., et al. (2010). Antioxidant Levels in Blood and Seminal Plasma and Their Impact on Sperm Parameters in Infertile Men. *Indian J. Biochem. Biophys.* 47 (1), 38–43.

Sharlip, I. D., Jarow, J. P., Belker, A. M., Lipshultz, L. I., Sigman, M., Thomas, A. J., et al. (2002). Best Practice Policies for Male Infertility. *Fertil. Sterility* 77 (5), 873–882. doi:10.1016/s0015-0282(02)03105-9

Sharma, R. K., Pasqualotto, F. F., Nelson, D. R., Thomas, A. J., Jr, and Agarwal, A. (1999). The Reactive Oxygen Species-Total Antioxidant Capacity Score Is a New Measure of Oxidative Stress to Predict Male Infertility*. *Hum. Reprod.* 14 (11), 2801–2807. doi:10.1093/humrep/14.11.2801

Sharma, R. K., Pasqualotto, A. E., Nelson, D. R., Thomas, A. J., Jr, and Agarwal, A. (2001). Relationship between Seminal white Blood Cell Counts and Oxidative Stress in Men Treated at an Infertility Clinic. *J. Androl.* 22 (4), 575–583. doi:10.1002/j.1939-4640.2001.tb02217.x

Steiner, A. Z., Hansen, K. R., Barnhart, K. T., Cedars, M. I., Legro, R. S., Diamond, M. P., et al. (2020). The Effect of Antioxidants on Male Factor Infertility: the Males, Antioxidants, and Infertility (MOXI) Randomized Clinical Trial. *Fertil. Sterility* 113 (3), 552–560. e3. doi:10.1016/j.fertnstert.2019.11.008

Taken, K., Alp, H. H., Eryilmaz, R., Donmez, M. I., Demir, M., Gunes, M., et al. (2016). Oxidative DNA Damage to Sperm Cells and Peripheral Blood Leukocytes in Infertile Men. *Med. Sci. Monit.* 22, 4289–4296. doi:10.12659/MSM.898631

Vatannejad, A., Tavilani, H., Sadeghi, M. R., Amanpour, S., Shapourizadeh, S., and Doosti, M. (2017). Evaluation of ROS-TAC Score and DNA Damage in Fertile Normozoospermic and Infertile Asthenozoospermic Males. *Urol. J.* 14 (1), 2973–2978.

Venkatesh, S., Shamsi, M. B., Deka, D., Saxena, V., Kumar, R., and Dada, R. (2011). Clinical Implications of Oxidative Stress & Sperm DNA Damage in Normozoospermic Infertile Men. *Indian J. Med. Res.* 134 (3), 396–398.

Vignini, A., Nanetti, L., Buldreghini, E., Moroni, C., Ricciardo-Lamonica, G., Mantero, F., et al. (2006). The Production of Peroxynitrite by Human Spermatozoa May Affect Sperm Motility through the Formation of Protein Nitrotyrosine. *Fertil. Sterility* 85 (4), 947–953. doi:10.1016/j.fertnstert.2005.09.027

Wadhwa, L., Priyadarshini, S., Fauzdar, A., Wadhwa, S. N., and Arora, S. (2020). Impact of Vitamin D Supplementation on Semen Quality in Vitamin D-Deficient Infertile Males with Oligoasthenozoospermia. *J. Obstet. Gynecol. India* 70 (1), 44–49. doi:10.1007/s13224-019-01251-1

Wagner, H., Cheng, J. W., and Ko, E. Y. (2018). Role of Reactive Oxygen Species in Male Infertility: An Updated Review of Literature. *Arab J. Urol.* 16 (1), 35–43. doi:10.1016/j.aju.2017.11.001

Walters, J., De Iuliis, G., Nixon, B., and Bromfield, E. (2018). Oxidative Stress in the Male Germline: A Review of Novel Strategies to Reduce 4-Hydroxynonenal Production. *Antioxidants* 7 (10), 132. doi:10.3390/antiox7100132

Whittington, K., Harrison, S. C., Williams, K. M., Day, J. L., McLaughlin, E. A., Hull, M. G. R., et al. (1999). Reactive Oxygen Species (ROS) Production and the

Outcome of Diagnostic Tests of Sperm Function. *Int. J. Androl.* 22 (4), 236–242. doi:10.1046/j.1365-2605.1999.00174.x

Wolff, H. (1995). The Biologic Significance of white Blood Cells in Semen. *Fertil. Steril* 63 (6), 1143–1157. doi:10.1016/s0015-0282(16)57588-8

Yoshida, Y., Umeno, A., Akazawa, Y., Shichiri, M., Murotomi, K., and Horie, M. (2015). Chemistry of Lipid Peroxidation Products and Their Use as Biomarkers in Early Detection of Diseases. *J. Oleo Sci.* 64 (4), 347–356. doi:10.5650/jos.ess14281

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 Molecular Biosciences

ORIGINAL RESEARCH
published: 21 January 2022
doi: 10.3389/fmolb.2022.832238

Toll-Like Receptors Serve as Biomarkers for Early Diagnosis and Prognosis Assessment of Kidney Renal Clear Cell Carcinoma by Influencing the Immune Microenvironment: Comprehensive Bioinformatics Analysis Combined With Experimental Validation

Xiong Zou[1,2,3,4†], *Bingqian Guo*[2,3,4,5†], *Qiang Ling*[1,2,3,4] *and Zengnan Mo*[1,2,3,4,5*]

[1]*Department of Urology, The First Affiliated Hospital of Guangxi Medical University, Nanning, China,* [2]*Center for Genomic and Personalized Medicine, Guangxi Medical University, Nanning, China,* [3]*Guangxi Collaborative Innovation Center for Genomic and Personalized Medicine, Nanning, China,* [4]*Guangxi Key Laboratory for Genomic and Personalized Medicine, Guangxi Key Laboratory of Colleges and Universities, Nanning, China,* [5]*Collaborative Innovation Center of Regenerative Medicine and Medical BioResource Development and Application, Guangxi Medical University, Nanning, China*

Background: Toll-like receptors (TLRs) are important initiators of innate and acquired immune responses. However, its role in kidney renal clear cell carcinoma (KIRC) remains unclear.

Methods: TLRs and their relationships with KIRC were studied in detail by ONCOMINE, UALCAN, GEPIA, cBioPortal, GeneMANIA, FunRich, LinkedOmics, TIMER and TRRUST. Moreover, we used clinical samples to verify the expressions of TLR3 and TLR4 in early stage of KIRC by real-time fluorescence quantitative polymerase chain reaction (RT-qPCR), flow cytometry (FC) and immunohistochemistry (IHC).

Results: The expression levels of TLRs in KIRC were generally different compared with adjacent normal tissues. Moreover, the expressions of TLR3 and TLR4 elevated significantly in the early stage of KIRC. Overexpressions of TLR1, TLR3, TLR4 and TLR8 in KIRC patients were associated with longer overall survival (OS), while inhibition of TLR9 expression was related to longer OS. Additionally, overexpressions of TLR1, TLR3 and TLR4 in KIRC patients were associated with longer disease free survival (DFS). There were general genetic alterations and obvious co-expression correlation of TLRs in KIRC. The PPI network between TLRs was rather complex, and the key gene connecting the TLRs interaction was MYD88. The GO analysis and KEGG pathway analysis indicated that TLRs were closely related to adaptive immunity, innate immunity and other immune-related processes. RELA, NFKB1, IRF8, IRF3 and HIF1A were key transcription factors regulating the expressions of TLRs. What's more, the expression levels of all TLRs in KIRC were positively correlated with the infiltration levels of dendritic cells, macrophages, neutrophils,

OPEN ACCESS

Edited by:
Matteo Becatti,
University of Firenze, Italy

Reviewed by:
Sriram Chitta,
University of Texas MD Anderson Cancer Center, United States
Yusuke Murakami,
Musashino University, Japan
Madhu Chhanda Mohanty,
National Institute of Virology (ICMR), India

***Correspondence:**
Zengnan Mo
mozengnan@gxmu.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *09 December 2021*
Accepted: *06 January 2022*
Published: *21 January 2022*

Citation:
Zou X, Guo B, Ling Q and Mo Z (2022) Toll-Like Receptors Serve as Biomarkers for Early Diagnosis and Prognosis Assessment of Kidney Renal Clear Cell Carcinoma by Influencing the Immune Microenvironment: Comprehensive Bioinformatics Analysis Combined With Experimental Validation. Front. Mol. Biosci. 9:832238. doi: 10.3389/fmolb.2022.832238

B cells, $CD4^+$ T cells and $CD8^+$ T cells. Finally, the results of RT-qPCR, FC and IHC confirmed that TLR3 and TLR4 were significantly elevated in the early stage of KIRC.

Conclusion: The occurrence and development of KIRC are closely related to TLRs, and TLRs have the potential to be early diagnostic biomarkers of KIRC and biomarkers for judging the prognosis and immune status of KIRC. This study may provide new insights into the selection of KIRC immunotherapy targets.

Keywords: KIRC, TLRs, prognosis, diagnosis, biomarkers

INTRODUCTION

Renal cell carcinoma (RCC) can be divided into different subtypes according to its histological characteristics with unique genetic and molecular alterations, different clinical processes and therapeutic responses (Linehan et al., 2010; Linehan and Ricketts, 2013; Moch et al., 2016). RCC is a tumor originating from renal epithelium (Sanchez-Gastaldo et al., 2017; Zou and Mo, 2021). The most common pathological subtype of RCC is kidney renal clear cell carcinoma (KIRC), accounting for about 75% of RCC (Linehan and Ricketts, 2019). Lipid accumulation and storage are the main pathological features of KIRC (Xiao et al., 2019), and KIRC is the most common cause of death associated with RCC (Hsieh et al., 2017). Local KIRC can be treated by surgical resection, but the treatment of advanced KIRC is still a clinical challenge, and the 5-years overall survival (OS) rate is 0–20% (Petitprez et al., 2021). Patients with recurrent or distant metastasis of KIRC have a poor prognosis and a short median survival (Zheng et al., 2017), moreover, the prognosis of KIRC patients is mainly based on tumor lymph node metastasis (TNM) stage (Pichler et al., 2013; Qin et al., 2013; Zheng et al., 2017), and there is a lack of biomarkers to determine the prognosis of patients, so the identification of early diagnostic biomarkers for KIRC and biomarkers that can assess patient prognosis are critical to the management and treatment of patients with KIRC.

Toll-like receptors (TLRs) are considered to be the key to identify pathogens and control immune response (Marin-Acevedo et al., 2018). TLRs play a crucial role in both innate and subsequent adaptive immunity because of its ability to sense foreign substances, known as pathogen-associated molecular patterns (PAMPs) (Shetab Boushehri and Lamprecht, 2018). In addition to PAMPs, TLRs can also recognize endogenous ligands. When tissue damage or cell death occurs, cells secrete alarmins, also known as danger-associated molecular patterns (DAMPs), but excessive release of the substance is associated with autoimmune diseases and cancer (Chan et al., 2012; Zhao et al., 2014; Urban-Wojciuk et al., 2019). TLRs are expressed in a variety of cells, including immune cells, fibroblasts, and epithelial cells, and their primary role is to protect the host against microbial infection (Kawasaki and Kawai, 2014; Chen et al., 2018). More and more studies have shown that TLRs also play an important role in the occurrence and development of cancer (Wang et al., 2018; Wu et al., 2018), and different TLRs play different roles in different cancers (Dajon et al., 2017). However, the effect of TLRs on KIRC and its mechanism are not clear. In this study, we systematically investigated the expressions of TLRs in KIRC by ONCOMINE, UALCAN and GEPIA databases, and analyzed the relationships between the expressions of TLRs and tumor stage and prognosis in patients with KIRC by GEPIA. In addition, we obtained the genetic alteration information of TLRs and spearman's correlation of co-expression between TLRs through cBioPortal. We explored the protein-protein interaction (PPI) network of TLRs by GeneMANIA, and obtained the most critical gene associated with TLRs by FunRich, and conducted detailed GO analysis and KEGG pathway analysis of TLRs through LinkedOmics database. At the same time, we investigated the key transcription factors regulating TLRs through TRRUST. What's more, we studied the relationships between the expression levels of TLRs in KIRC and the levels of immune cell infiltration by TIMER, and evaluated the effects of TLRs and immune cell infiltration on the survival risk of KIRC. Finally, we used clinical samples to verify the expressions of TLR3 and TLR4 in early stage of KIRC by RT-qPCR, flow cytometry (FC) and immunohistochemistry (IHC). Our study provides new insights into TLRs and their relationships with KIRC, contributing to the research of early diagnosis and therapeutic targets of KIRC.

MATERIALS AND METHODS

ONCOMINE

ONCOMINE is a powerful bioinformatics tool for genome-wide expression analysis (Rhodes et al., 2004). The expressions of TLRs in renal cell carcinoma were evaluated by ONCOMINE, so as to speculate the expressions of TLRs in KIRC. In this study, we used the following criteria: p value as 0.05, gene rank as top 10%, fold change as 2 and data type as all (DNA and mRNA).

UALCAN

UALCAN is a website for mining TCGA and MET500 cohort data. It has a variety of functions, including evaluating the expressions of different genes in different cancers and the effect of gene expression on cancer survival (Chandrashekar et al., 2017). Through the "TCGA Gene Analysis" of the UALCAN, this study explored the expressions of TLRs in KIRC compared with normal tissues.

GEPIA

GEPIA is a website that can be used to analyze RNA expression levels in a variety of tumors and corresponding normal tissues. It also has many functions such as evaluating the effect of different

RNA expression levels on the prognosis of cancer (Tang et al., 2017). In this study, we compared the expressions of TLRs mRNA in KIRC and corresponding normal tissues, as well as the expression levels of TLRs in different stage of KIRC through the "Expression DIY" module of GEPIA. Additionally, we also investigated the effects of TLRs on the survival of patients with KIRC through the "Survival" module.

cBioPortal

cBioPortal is a powerful website for analyzing multidimensional cancer genome data. It can visualize data from cancer tissues and cells into easy-to-understand genetic and gene co-expression events (Gao et al., 2013). Based on the TCGA database of 538 cases of renal clear cell carcinoma (TCGA, Firehose Legacy), the genetic alteration and co-expression of TLRs were obtained from cBioPortal, the threshold of Protein expression z-scores (RPPA) and mRNA expression z scores (RNA Seq V2 RSEM) was set to ±2.0.

GeneMANIA

GeneMANIA is a convenient and versatile website for analyzing PPI, co-expression, pathways and related functions (Warde-Farley et al., 2010). This study studied the related functions of TLRs and its PPI network through GeneMANIA.

FunRich

FunRich (3.1.3 exe), a bioinformatics tool, can perform analysis of multiple genes or proteins data sets provided (Fonseka et al., 2020). Through FunRich, we obtained the most critical gene associated with TLRs, and carried out relevant research on this gene.

LinkedOmics

LinkedOmics is a tool that can be used to analyze the multi-omics data of 32 cancer types from TCGA (Vasaikar et al., 2018). In this study, "GO analysis" and "KEGG Pathway" enrichment analysis of TLRs were carried out using the "Gene Set Enrichment Analysis" tool in "LinkInterpreter" module. We set the "Rank Criteria" as meta *p*-value, the "Minimum Number of Genes Size" as 3 and the "Simulations" as 500. Statistical analysis was conducted using Person Correlation Test.

TIMER

TIMER, a reliable utility, allows users to enter specific parameters to systematically assess immune cell infiltration in different tumors and its impact on clinical outcomes (Li et al., 2017). In this study, the correlations between the expressions of TLRs and the levels of immune cell infiltration in KIRC were evaluated using the "Gene" module, and the correlations between the clinical prognosis of KIRC and the expressions of TLRs and immune cells were evaluated by "Survival" module.

TRRUST

TRRUST is a powerful tool for querying transcription factors that regulate gene expression, and can provide regulatory information on the interaction of many transcription factors in humans and mice (Han et al., 2018). In this study, TRRUST was used to query the transcription factors related to regulating TLRs.

Tissue Collection

Six KIRC tumor tissues and adjacent nontumor tissues were obtained from The First Affiliated Hospital of Guangxi Medical University. The study was approved by the Ethics Committee of The First Affiliated Hospital of Guangxi Medical University (Approval Number: 2021KY-E-182) and all the participants in the experiment gave their informed consent.

Real-Time Fluorescence Quantitative Polymerase Chain Reaction

Total RNA was extracted by Total RNA Kit I (R6834, Omega). According to the instructions for the use of PrimeScript RT reagent kit (RR036a, Takara, Kyoto, Japan). RNA was reverse transcribed into cDNA, and then cDNA was detected by RT-qPCR using FastStart Essential DNA Green Master (06,924,204,001, Roche) and LightCycler® 96 Instrument (Roche). Three repeated assays were set for each sample. Using glyceraldehyde-3-phosphate dehydrogenase (GADPH) as internal reference, the relative expressions of target genes were analyzed by $2^{-\Delta\Delta Ct}$ method. The primer sequences of TLR3 were as follows: 5′-TTGCCTTGTATCTACTTTTGGGG-3' (Forward); 5′-TCAACACTGTTATGTTTGTGGGT-3' (Reverse). The primer sequences of TLR4 were as follows: 5′-AGACCTGTCCCTGAACCCTAT-3' (Forward); 5′-CGATGGACTTCTAAACCAGCCA-3' (Reverse).

Flow Cytometry

Flow cytometry (FC) was used to analyze the expressions of TLR3 and TLR4 in KIRC relative to adjacent nontumor tissues. Fresh human KIRC tissue and adjacent nontumor tissue samples were collected in the First Affiliated Hospital of Guangxi Medical University and prepared into single cell suspension. Firstly, all tissues were cut into small pieces with scissors. After washing twice by D-PBS (311–425-CL, Wisent), the sample was transferred to digestive juice (0.1 mg/ml collagenase I (10,103,578,001, Roche) and 1 mg/ml dnase I (10,104,159,001, Roche) in HBSS (14,025,092, Gibco) solution) and gently shaken at 37°C for 30 min. Digestion was terminated using 10% fetal bovine serum (FBS; SH30070.03, HyClone) in RPMI 1640 and the disaggregated cell suspensions were passed through a 100 μm cell strainer (352,350, Falcon). The cell suspensions were washed thoroughly with D-PBS containing 1% FBS. Red blood cells were eliminated by 1X red blood cell (RBC) lysis buffer (420,301, BioLegend) for 5 min on ice and lysis was terminated by dilution with D-PBS containing 1% FBS, filtered through a 40 μm cell strainer (352,340, Falcon). Then, the cell suspensions were washed thoroughly with D-PBS containing 1% FBS. Finally, cells obtained by centrifugation were resuspended using PBS. Then, single cell suspensions of normal and cancer tissues of kidney were transferred to Fixation Buffer (420,801, Biolegend) for 20 min at room temperature and dark. Precipitation obtained by centrifugation were washed two times by 1× Intracellular Staining Perm Wash Buffer (421,002, Biolegend). Then, about 10^6 cells were incubated with TLR3 (bs-1444R, Bioss) or TLR4 (bs-20594R, bioss) diluted by 1× Intracellular Staining Perm Wash Buffer for 20 min at room temperature. Cells were washed two

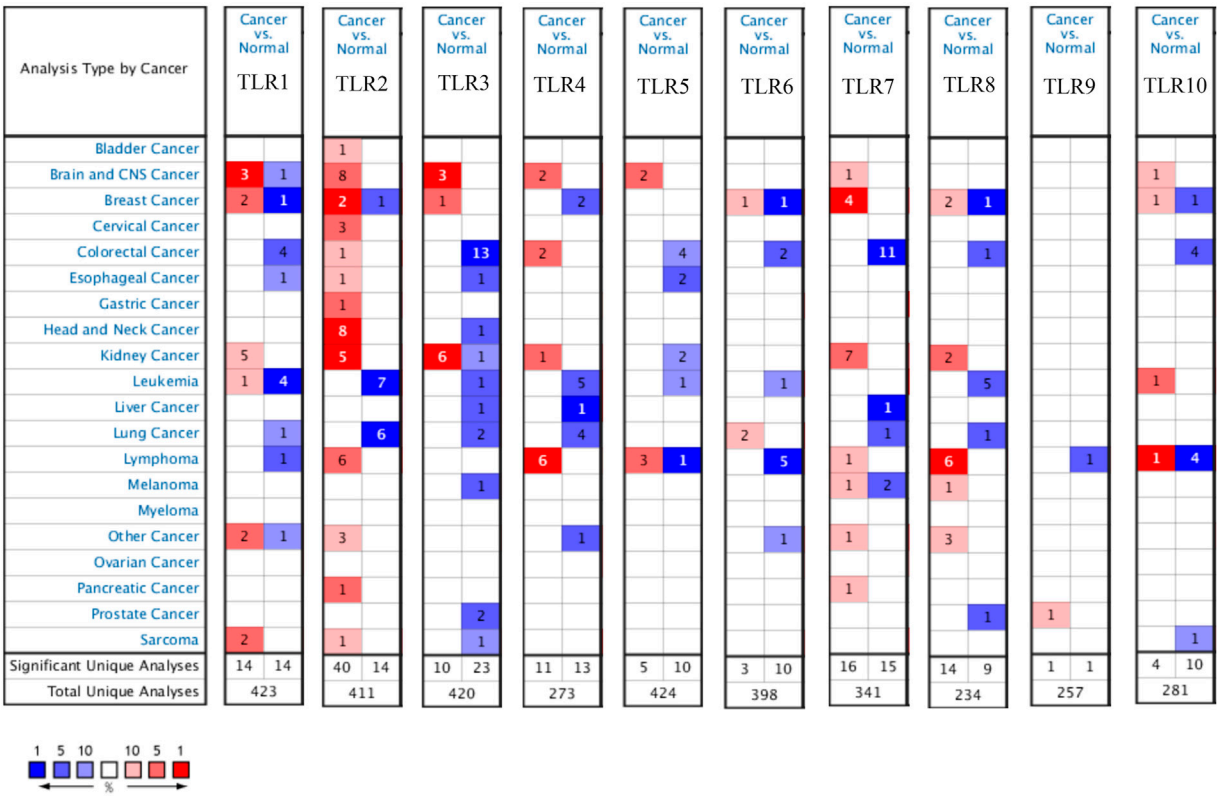


FIGURE 1 | Expression levels of TLRs in kidney cancer. The figure showing expression profiles of TLRs in tumor and paired normal tissue samples from the ONCOMINE database.

times by 1× Intracellular Staining Perm Wash Buffer. The cells were then incubated with Alexa Fluor 488-conjugated donkey anti-rabbit IgG antibodies (1:2000, ab150061, Abcam) for 20 min at room temperature and dark. After washed twice, the cells were resuspended in PBS for flow cytometry analysis. All samples were loaded on a BD C6 Plus for flow cytometry analysis. The data were analyzed using flowjo V10.0.

Immunohistochemistry

Three cases of KIRC tissues and adjacent nontumor tissues from the First Affiliated Hospital of Guangxi Medical University were collected. After fixation with 4% paraformaldehyde for 24 h, paraffin embedding and section were performed. The sections were dewaxed and hydrated using xylene and gradient alcohol. Then, the sections were treated with EDTA at pH 8.5 (C1034, Solarbio) to induce epitope retrieval by heating. After washing three times with PBS, the sections were incubated with an endogenous peroxidase inhibitor (SP-9001, ZSGB-BIO) for 10 min. Then the sections were washed with PBS three times and incubated with normal goat serum blocking solution (SP-9001, ZSGB-BIO) for 15 min. Primary antibodies (TLR3, 1:200; TLR4, 1:200) were incubated overnight at 4°C. After 30 min of room temperature balance the next day, incubated with Biotinylated Second Antibody (SP-9001, ZSGB-BIO) for 15 min. Then the sections were washed with PBS three times and incubated with Streptavidin-Enzyme Conjugate (SP-9001, ZSGB-BIO) for 15 min. After washing three times with PBS, the sections were incubated with DAB chromogenic fluid (ZLI-9018, ZSGB-BIO) for 5 min. Finally, after redyeing with hematoxylin, the slices were fixed with neutral gum. The images were captured using microscope (Olympus, CX23) and then processed with ImageJ software (NIH).

Statistical Analysis

All experimental data were statistically analyzed using GraphPad Prism 7. *t*-test was used to analyze the expressions of TLR3 and TLR4 in KIRC tissues relative to adjacent nontumor tissues. In this paper, $p < 0.05$ was considered statistically significant.

RESULTS

Toll-Like Receptors Expression Levels in Renal Cell Carcinoma and Adjacent Nontumor Tissues

Expressions of TLRs in renal cell carcinoma relative to adjacent nontumor tissues were retrieved from ONCOMINE database. The results demonstrated that the expressions of TLR1, TLR2, TLR3, TLR4, TLR7 and TLR8 were significantly elevated, while the expression of TLR5 was significantly decreased in RCC tissues (**Figure 1**). We also

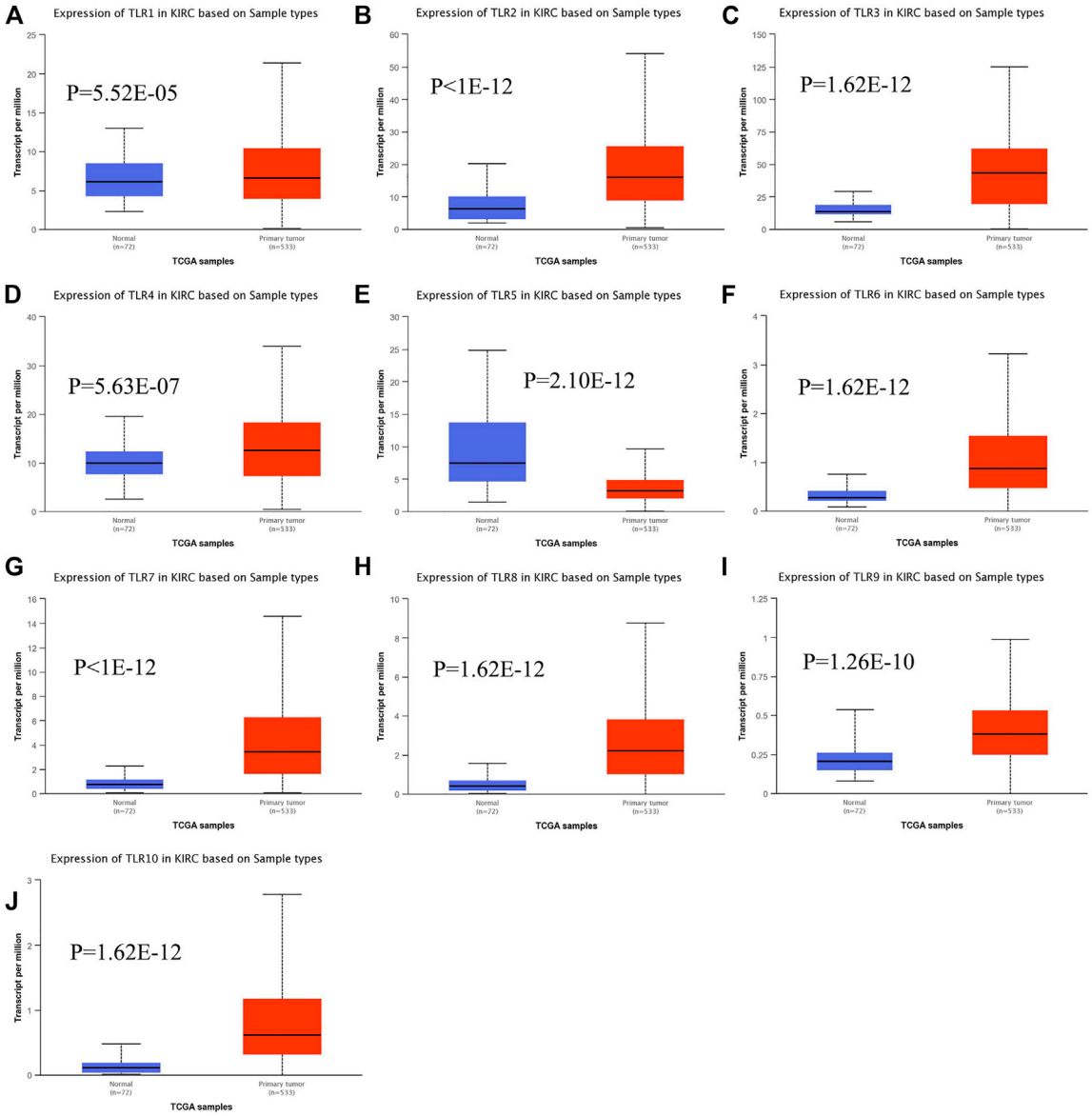


FIGURE 2 | The transcript expression levels of TLRs in KIRC. Box plots showing the transcript expression levels of TLR1 **(A)**, TLR2 **(B)**, TLR3 **(C)**, TLR4 **(D)**, TLR5 **(E)**, TLR6 **(F)**, TLR7 **(G)**, TLR8 **(H)**, TLR9 **(I)** and TLR10 **(J)** in KIRC compared with normal tissues.

evaluated the transcript expression levels of TLRs in KIRC by UALCAN. Compared with nontumor tissues, the expressions of TLR1 (p = 5.52E-05), TLR2 (p < 1E-12), TLR3 (p = 1.62E-12), TLR4 (p = 5.63E-07), TLR6 (p = 1.62E-12), TLR7 (p < 1E-12), TLR8 (p = 1.62E-12), TLR9 (p = 1.62E-10) and TLR10 (p = 1.62E-12) transcripts in KIRC were significantly elevated, while the transcript expression level of TLR5 was significantly decreased (**Figure 2**). Meanwhile, we also used GEPIA to compare the relative expression levels of all TLRs in KIRC tissues. The results showed that the TLR3 expression was the highest compared with other TLRs in KIRC tissues (**Figure 3**).

To further identify TLRs associated with the occurrence, progression, and clinical prognosis of KIRC, we evaluated TLRs expression levels at different pathological stages of KIRC. We found that there were significant correlations of TLR3 (p = 0.008) and TLR4 (p = 0.001) expressions with the pathological stages of KIRC, while there were no significant correlations in the expressions of other TLRs at different pathological stages of KIRC (**Figure 4**). The expressions of TLR3 and TLR4 elevated significantly in the early stage of KIRC, indicating that TLR3 and TLR4 played an important role in the early diagnosis of KIRC. In addition, all these data

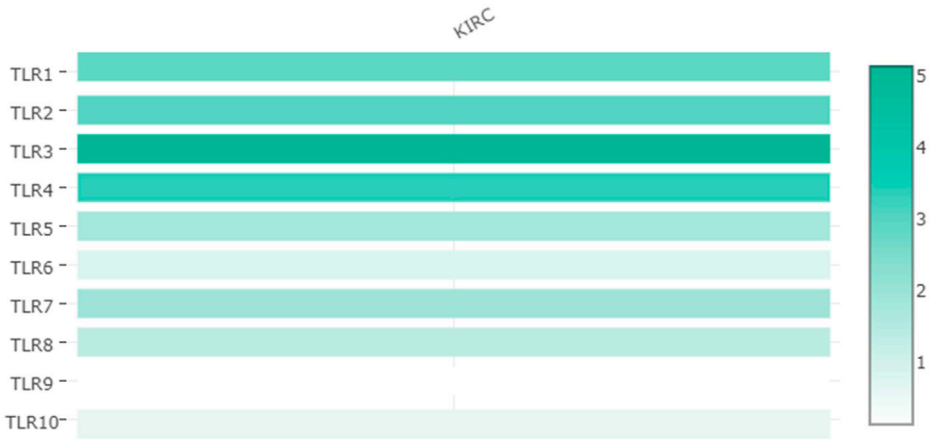


FIGURE 3 | Relative expression levels of TLRs in KIRC.

indicated that TLRs played a momentous influence in the occurrence and progression of KIRC.

The Effects of Toll-Like Receptors on the Prognosis of Kidney Renal Clear Cell Carcinoma

In order to evaluate the effects of TLRs on the prognostic value of KIRC, we used GEPIA to assess the correlations of TLRs with the disease free survival (DFS) and overall survival (OS) of KIRC (**Figures 5**, **6**). High expressions of TLR1 (p = 0.018), TLR3 (p = 2.6e-09), TLR4 (p = 5.4e-05) and TLR8 (p = 0.035) in patients with KIRC were associated with longer OS (**Figures 5A,C,D,H**), while low expression of TLR9 (p = 0.018) in patients with KIRC was associated with longer OS (**Figure 5I**). Moreover, we found that high expressions of TLR1 (p = 0.017), TLR3 (p = 0.00013) and TLR4 (p = 0.00078) in patients with KIRC were associated with longer DFS (**Figures 6A,C,D**), but there were no significant correlations between the expressions of other TLRs and the DFS of KIRC (**Figure 6**).

Analyses of Genetic Altetation, Co-Expression and Protein-Protein Interaction of Toll-Like Receptors in Patients With Kidney Renal Clear Cell Carcinoma

Next, we systematically analyzed the molecular characteristics of TLRs in patients with KIRC. First of all, we analyzed the genetic alterations and co-expression of TLRs in 538 KIRC patients using cBioPortal. In the KIRC samples, the results showed that the altered/profiled ratio of TLR1, TLR2, TLR3, TLR4, TLR5, TLR6, TLR7, TLR8, TLR9 and TL10 is 5, 4, 6, 3, 5, 5, 4, 4, 12 and 5%, respectively (**Figure 7A**). High mRNA expression and deep deletion were the most common changes in these samples. Next, we explored the spearman's correlation of co-expression among TLRs. The results showed that there were general positive correlations between TLRs co-expression (**Table 1**). Not only that, we also analyzed the PPI network between TLRs through GeneMANIA. The functions of these TLRs were mainly related to toll-like receptor signaling pathway, pattern recognition receptor signaling pathway, innate immune response-activating signal transduction, activation of innate immune response, positive regulation of innate immune response, positive regulation of defense response and regulation of innate immune response (**Figure 7B**). Next, we further studied the major genes that interacted with TLRs through FunRich. The results showed that MYD88 was the key gene to connect the interaction between TLRs (**Figure 7C**), indicating that MYD88 plays a crucial role in the expressions of TLRs.

Gene Set Enrichment Analysis of Toll-Like Receptors in Kidney Renal Clear Cell Carcinoma Patients

LinkedOmics was used for gene enrichment analysis of TLRs. We studied TLRs-related GO analysis and KEGG pathway. Many biological processes (BP) of significant enrichment of TLRs were closely related to the occurrence and development of KIRC, including adaptive immune response, regulation of leukocyte activation, immune response-regulating signaling pathway, lymphocyte mediated immunity, leukocyte cell-cell adhesion, positive regulation of cytokine production, interferon-gamma production, regulation of immune effector process, regulation of cell-cell adhesion, positive regulation of defense response, leukocyte differentiation and lymphocyte activation involved in immune response (**Figure 8A**). In addition, side of membrane, secretory granule membrane, tertiary granule, receptor complex, specific granule, endocytic vesicle, membrane region, mast cell granule, cell leading edge, MHC protein complex, protein complex involved in cell adhesion, neuron spine, immunological synapse, PML body, chromosomal region, phagocytic cup and ficolin-1-rich granule were the most obviously enriched projects in the cellular

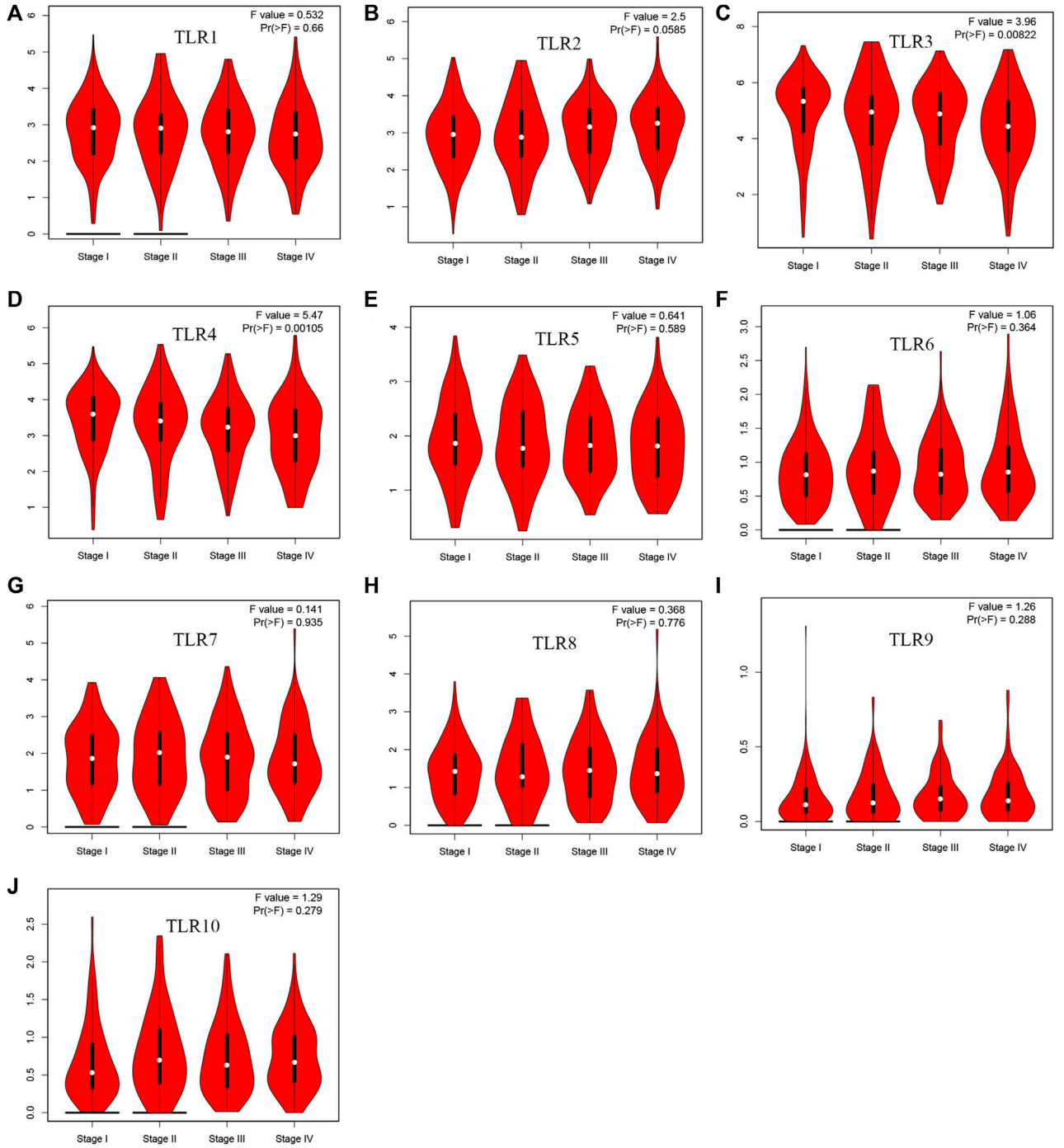


FIGURE 4 | The association of TLRs expression levels with different pathological stages of KIRC. **(A)** TLR1, **(B)** TLR2, **(C)** TLR3, **(D)** TLR4, **(E)** TLR5, **(F)** TLR6, **(G)** TLR7, **(H)** TLR8, **(I)** TLR9, **(J)** TLR10.

components (CC) (**Figure 8B**). The molecular functions (MF) involved in the enrichment of TLRs mainly included cytokine binding, antigen binding, cytokine receptor activity, peptide receptor activity, cytokine receptor binding, MHC protein binding, SH3 domain binding, coreceptor activity, SH2 domain binding, lipopolysaccharide binding, carbohydrate binding and purinergic receptor activity (**Figure 8C**). Among the TLRs-enriched KEGG pathway, Th17 cell differentiation, Th1 and Th2 cell differentiation, Toll-like receptor signaling pathway, TNF signaling pathway, JAK-STAT signaling pathway and cell adhesion molecules were significantly correlated with the tumorigenesis of KIRC (**Figure 8D**).

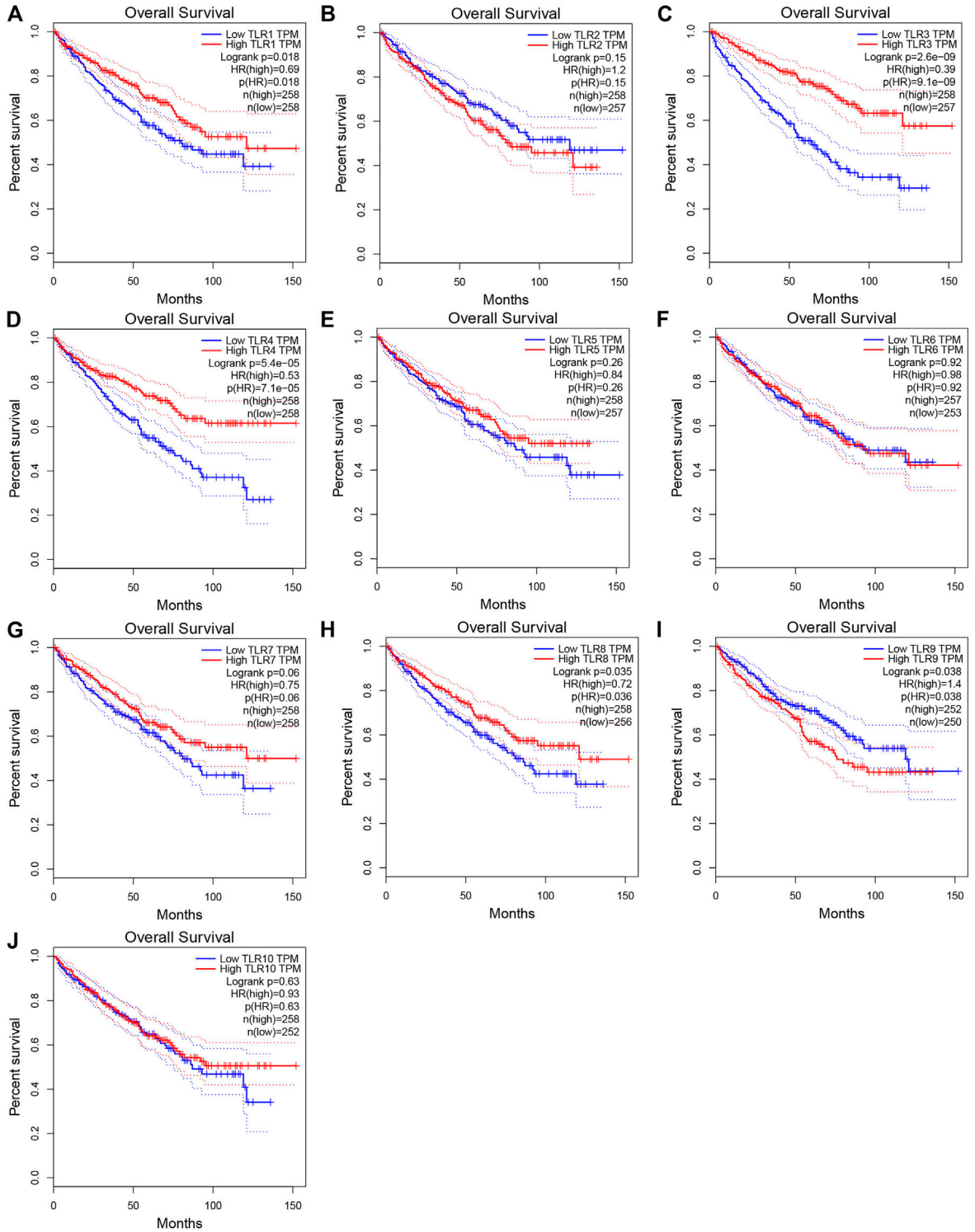


FIGURE 5 | Effects of TLRs on overall survival time of patients with KIRC. **(A)** TLR1, **(B)** TLR2, **(C)** TLR3, **(D)** TLR4, **(E)** TLR5, **(F)** TLR6, **(G)** TLR7, **(H)** TLR8, **(I)** TLR9, **(J)** TLR10.

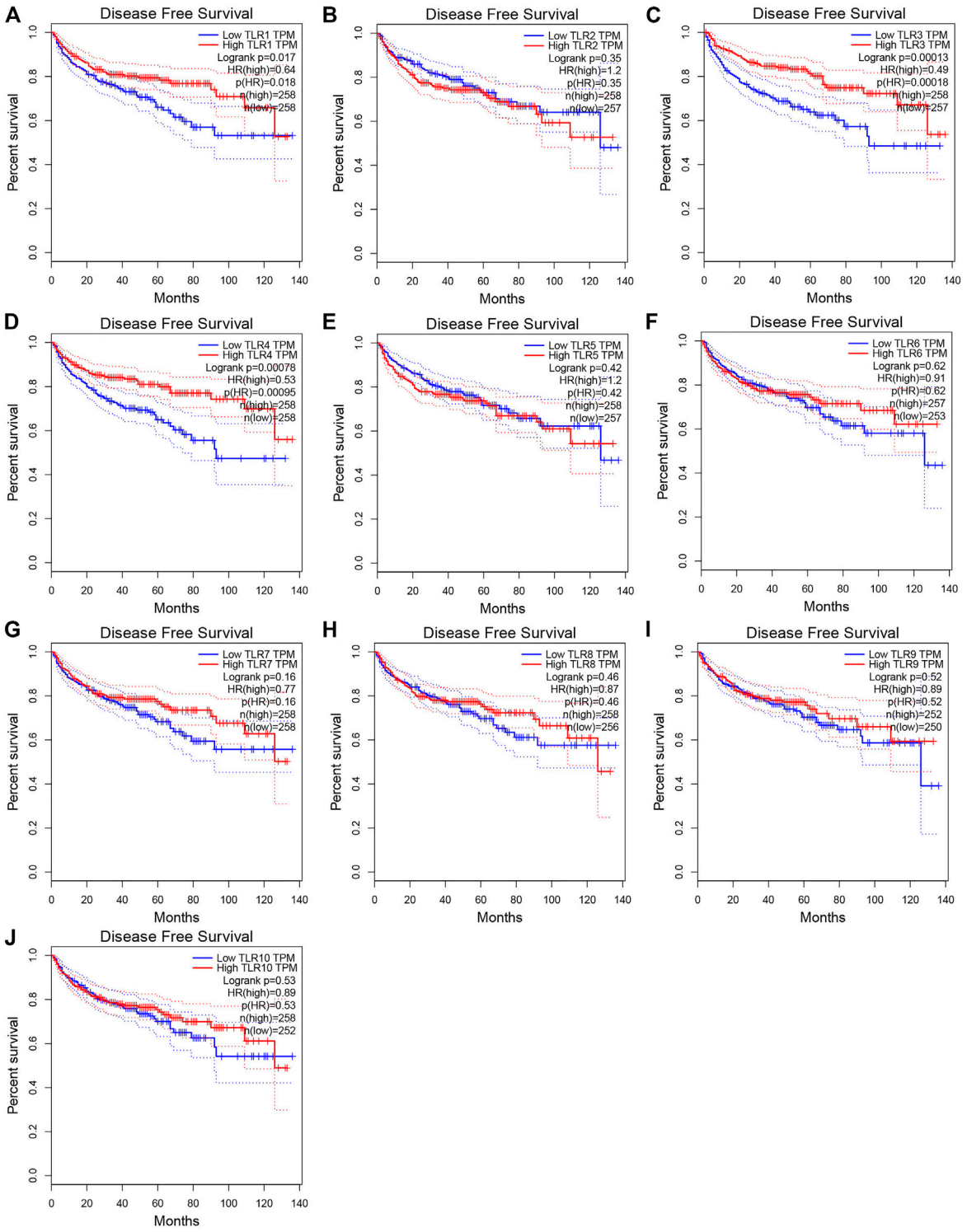


FIGURE 6 | Effects of TLRs on disease free survival time of patients with KIRC. **(A)** TLR1, **(B)** TLR2, **(C)** TLR3, **(D)** TLR4, **(E)** TLR5, **(F)** TLR6, **(G)** TLR7, **(H)** TLR8, **(I)** TLR9, **(J)** TLR10.

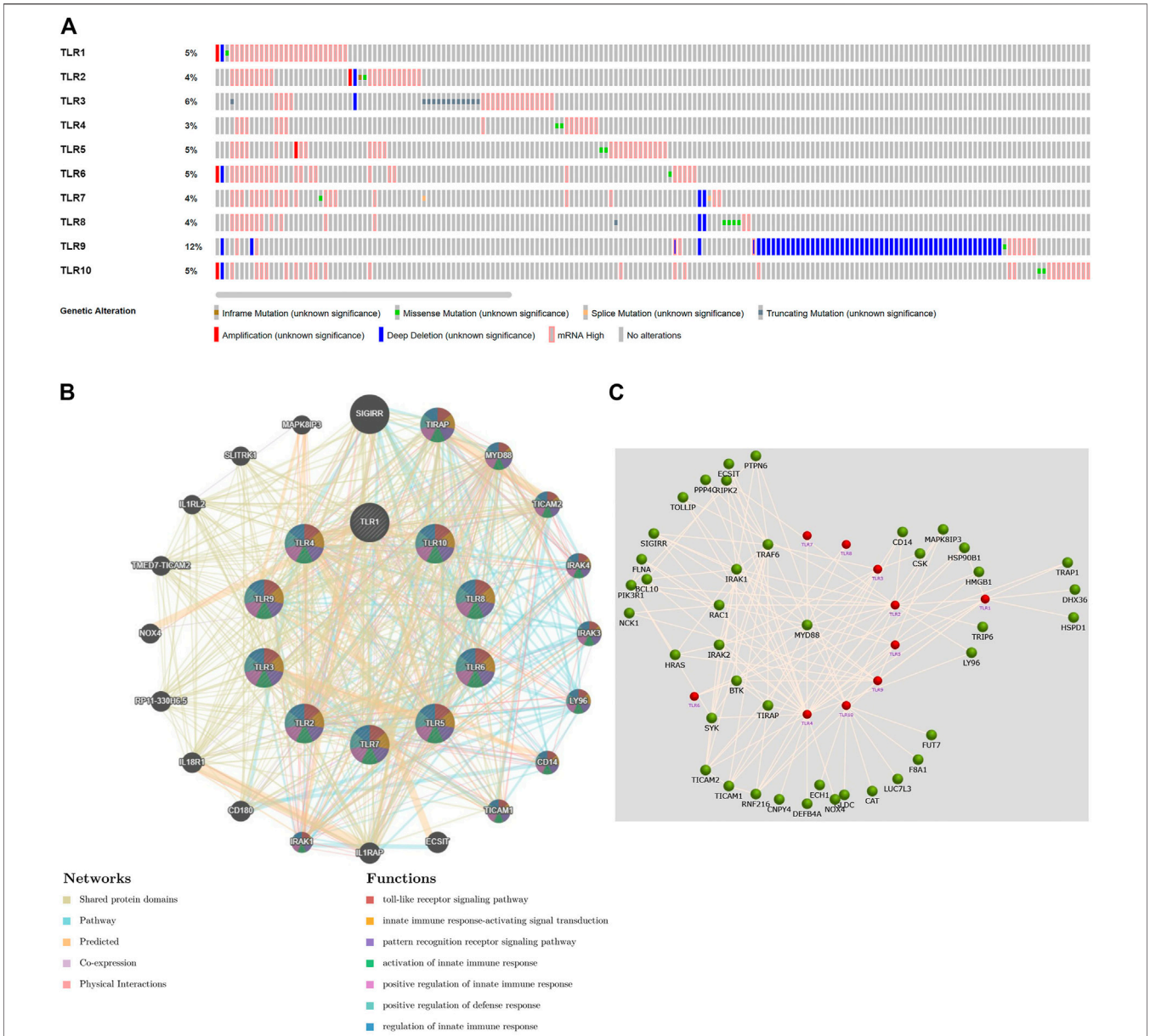


FIGURE 7 | Genetic alteration, function enrichment of PPI network, genetic interaction analyses of TLRs in KIRC patients. **(A)** Genetic alteration of different TLRs in KIRC. **(B)** Functional enrichment of TLRs-related PPI network in KIRC. **(C)** The key gene connecting the TLRs interaction was MYD88.

TABLE 1 | Correlation of co-expression between TLRs.

	TLR1	TLR2	TLR3	TLR4	TLR5	TLR6	TLR7	TLR8	TLR9	TLR10
TLR1	1.000	0.669	0.502	0.613	0.623	0.750	0.870	0.851	0.061	0.678
TLR2	0.669	1.000	0.157	0.352	0.605	0.774	0.679	0.751	0.297	0.661
TLR3	0.502	0.157	1.000	0.467	0.134	0.261	0.441	0.430	–0.268	0.212
TLR4	0.613	0.352	0.467	1.000	0.359	0.497	0.609	0.607	–0.079	0.365
TLR5	0.623	0.605	0.134	0.359	1.000	0.613	0.631	0.605	0.213	0.531
TLR6	0.750	0.774	0.261	0.497	0.613	1.000	0.746	0.792	0.364	0.696
TLR7	0.870	0.679	0.441	0.609	0.631	0.746	1.000	0.901	0.139	0.732
TLR8	0.851	0.751	0.430	0.607	0.605	0.792	0.901	1.000	0.182	0.694
TLR9	0.061	0.297	–0.268	–0.079	0.213	0.364	0.139	0.182	1.000	0.376
TLR10	0.678	0.661	0.212	0.365	0.531	0.696	0.732	0.694	0.376	1.000

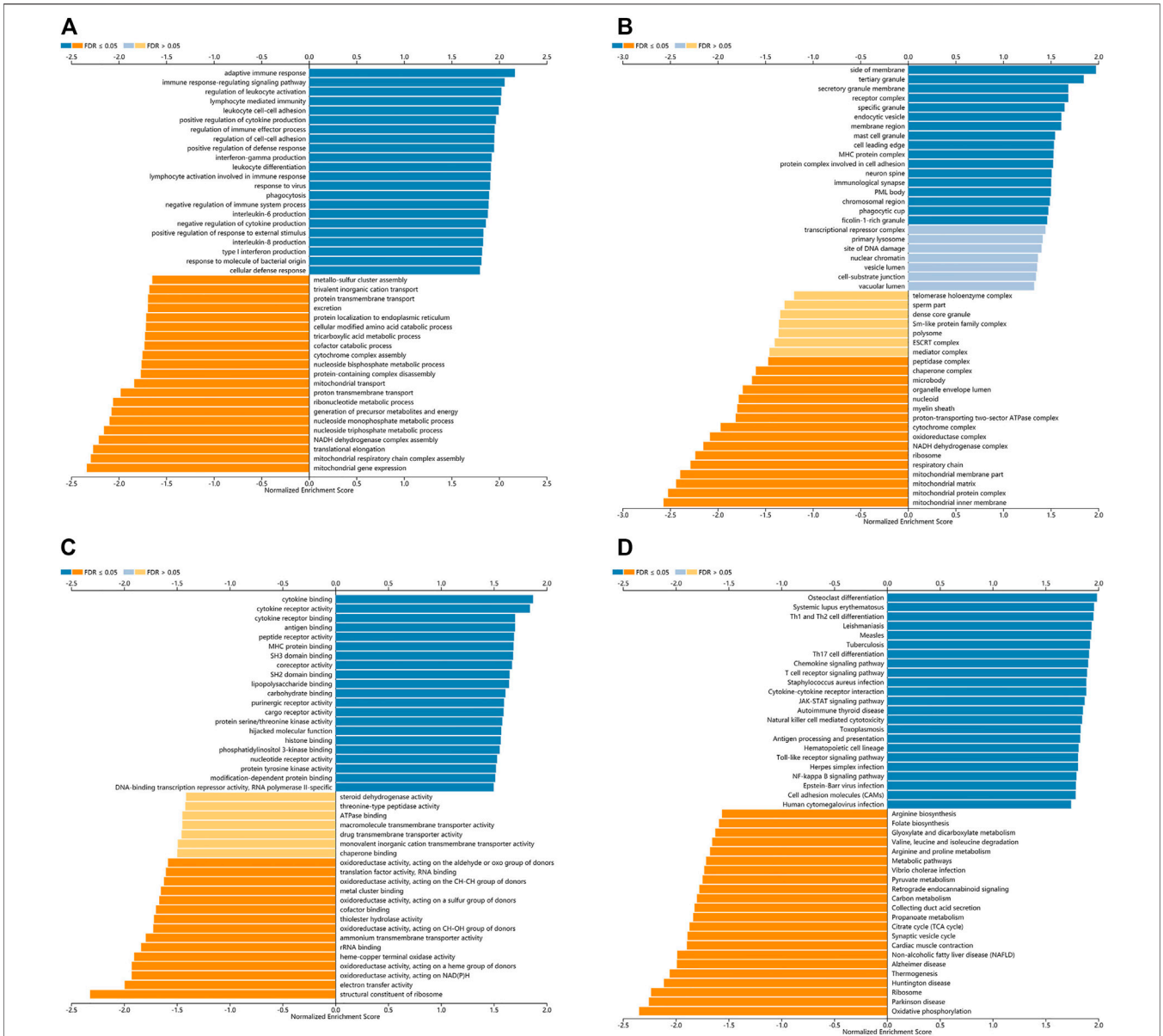


FIGURE 8 | The GO and KEGG pathway analyses of TLRs in KIRC. **(A)** Biological processes involved in TLRs in KIRC. **(B)** Cellular components involved in TLRs in KIRC. **(C)** Molecular functions involved in TLRs in KIRC. **(D)** KEGG pathways involved in TLRs in KIRC.

TABLE 2 | The key transcription factors regulating the expressions of TLRs in KIRC patients.

Key TF	Description	Target genes	*p* Value	FDR
RELA	v-rel reticuloendotheliosis viral oncogene homolog A (avian)	4 (TLR2, TLR3, TLR7, TLR9)	1.23E-05	2.31E-05
NFKB1	nuclear factor of kappa light polypeptide gene enhancer in B-cells 1	4 (TLR2, TLR3, TLR7, TLR9)	1.27E-05	2.31E-05
IRF8	interferon regulatory factor 8	2 (TLR3, TLR4)	1.39E-05	2.31E-05
IRF3	interferon regulatory factor 3	2 (TLR3, TLR4)	2.64E-05	3.31E-05
HIF1A	hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)	2 (TLR2, TLR6)	0.00084	0.00084

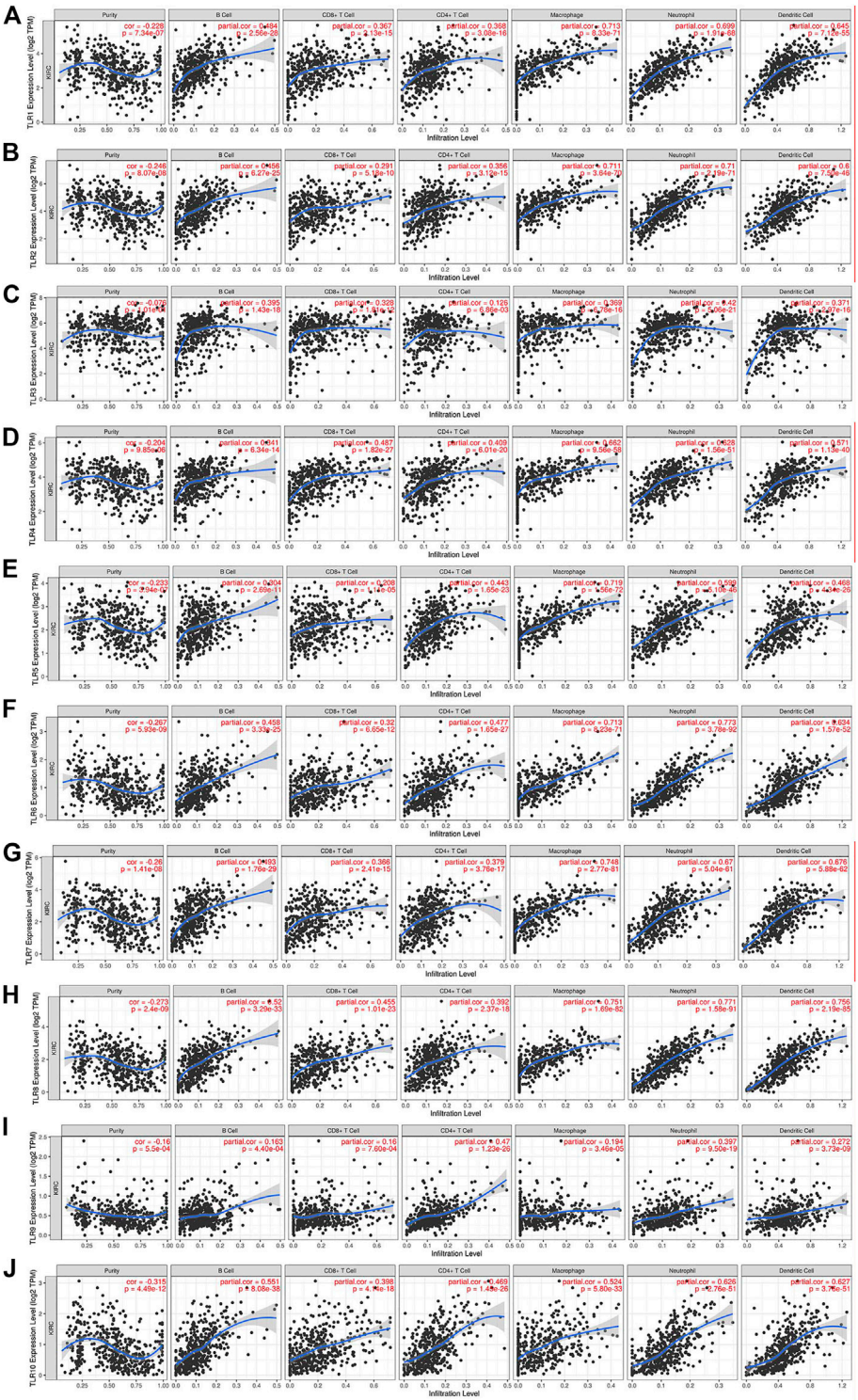


FIGURE 9 | The relationships between the expression levels of TLRs and immune cells infiltration in KIRC. The associations of immune cells infiltration levels with the expression levels of TLR1 **(A)**, TLR2 **(B)**, TLR3 **(C)**, TLR4 **(D)**, TLR5 **(E)**, TLR6 **(F)**, TLR7 **(G)**, TLR8 **(H)**, TLR9 **(I)** and TLR10 **(J)** in KIRC.

TABLE 3 | Cox proportional hazard model of TLRs and 6 kinds of immune cells infiltration in KIRC patients.

	Coef	HR	95%CI_l	5%CI_u	p.value	Sig
B_cell	−3.497	0.030	0.001	0.721	0.031	*
CD8_Tcell	−1.967	0.140	0.025	0.788	0.026	*
CD4_Tcell	−4.418	0.012	0.000	0.324	0.009	*
Macrophage	−0.565	0.568	0.029	11.176	0.710	
Neutrophil	2.300	9.977	0.067	1,480.962	0.367	
Dendritic	4.130	62.203	7.655	505.431	<0.001	***
TLR1	0.401	1.494	0.932	2.395	0.096	
TLR2	0.102	1.108	0.848	1.447	0.452	
TLR3	−0.216	0.806	0.679	0.957	0.014	*
TLR4	−0.394	0.674	0.479	0.949	0.024	*
TLR5	−0.082	0.921	0.676	1.254	0.600	
TLR6	0.570	1.768	1.012	3.089	0.045	*
TLR7	0.045	1.046	0.621	1.762	0.867	
TLR8	−0.926	0.396	0.221	0.711	0.002	**
TLR9	1.311	3.709	1.924	7.148	<0.001	***
TLR10	0.077	1.080	0.610	1.913	0.793	

*p < *0.05*, **p < *0.01*, ***p < *0.001*.

Key Transcription Factors Associated With Toll-Like Receptors in Kidney Renal Clear Cell Carcinoma Patients

Through TRRUST, we explored the transcription factors that regulated the expression of TLRs in KIRC patients. The results showed that the key transcription factors of TLRs were RELA, NFKB1, IRF8, IRF3 and HIF1A (**Table 2**). NFKB1 and RELA were key transcription factors that regulated the expressions of TLR2, TLR3, TLR7 and TLR9. IRF3 and IRF8 were key transcription factors that regulated the expressions of TLR3 and TLR4. HIF1A was the key transcription factor that regulated the expressions of TLR2 and TLR6.

Correlations Between the Expressions of Toll-Like Receptors and Immune Cell Infiltration Levels in Kidney Renal Clear Cell Carcinoma Patients

We comprehensively studied the correlations between the expressions of TLRs and the levels of immune cell infiltration in patients with KIRC by TIMER. To our surprise, the expressions of all TLRs in KIRC were positively correlated with the infiltration levels of dendritic cells, neutrophils, B cells, macrophages, $CD8^+$ T cells and $CD4^+$ T cells (**Figure 9**). Not only that, we also established a cox proportional hazard model of the effects of TLRs and six kinds of immune cells infiltration on patients with KIRC. The results indicated that B cells (coef = −3.497, p = 0.031), CD8+T cells (coef = −1.967, p = 0.026), CD4+T cells (coef = −4.418, p = 0.009), TLR3 (coef = −0.216, p = 0.014), TLR4 (coef = −0.394, p = 0.024) and TLR8 (coef = −0.926, p = 0.002) were negatively associated with the risk of survival in patients with KIRC, while dendritic cells (coef = 4.130, $p < 0.001$) and TLR9 (coef = 1.311, $p < 0.001$) were positively associated with the risk of survival in patients with KIRC (**Table 3**).

Verification the Expressions of TLR3 and TLR4 in Early Stage of Kidney Renal Clear Cell Carcinoma and Adjacent Nontumor Tissues

Finally, we used clinical samples to compare the differences of mRNA and protein expressions of TLR3 and TLR4 in the early stage of KIRC tissues and adjacent nontumor tissues. The characteristics of patients for verifying mRNA and protein expression levels were shown in **Tables 4**, **5**, respectively. The results of RT-qPCR showed that the expression of TLR3 mRNA in KIRC was significantly elevated than that in adjacent nontumor tissues, and TLR4 mRNA also showed the same trend (**Figure 10A**). What's more, the results of FC showed that the relative expression of TLR3 at the protein level in KIRC was significantly elevated than that in adjacent nontumor tissues, and the expression of TLR4 at the protein level showed the same trend (**Figures 10B,C**). In addition, the results of IHC were consistent with the results of RT-qPCR and FC (**Figure 10D**). Our results confirmed that TLR3 and TLR4 were significantly elevated in the early stage of KIRC compared with adjacent nontumor tissues.

DISCUSSION

KIRC can be cured in early diagnosis, but when the disease is metastatic, it is the cancer with the worst prognosis in the urinary system (Angulo et al., 2021). Therefore, early diagnosis of KIRC is important for its therapeutic efficacy and prognosis. However, there is currently a lack of clear clinical biomarkers that can be used to diagnose the early stage of KIRC (Siegel et al., 20182018), and prognosis of patients is mainly determined by TNM stage (Pichler et al., 2013; Qin et al., 2013; Zheng et al., 2017). In addition, molecular biomarkers can provide the possibility of accurate prediction of cancer prognosis and early diagnosis (Tamayo et al., 2011). Therefore, the research on molecular biomarkers for early diagnosis and prognosis of KIRC patients could bring great benefits to the majority of KIRC patients and provide a refined management strategy for KIRC patients.

Toll-like receptor (TLRs) are important initiators of innate and acquired immune responses (Zhang et al., 2021). Ten kinds of TLRs have been identified in humans, and they are expressed in varieties of cells, including B cells, T cells and many other non-immune cells (Nouri et al., 2021; Rameshbabu et al., 2021). There is growing evidence that TLRs play a significance role in a variety of pathological processes, including inflammation, tumor, autoimmune diseases, immunotherapy and vaccination (Vidya et al., 2018). Although the studies of the associations between TLRs and cancer have increased in recent years, there are few studies on the effects of TLRs on KIRC.

Therefore, we used multiple databases to study the relationships between TLRs and KIRC. First of all, we preliminarily studied the expressions of TLRs in kidney cancer through ONCOMINE, and found that the expressions of TLR1, TLR2, TLR3, TLR4, TLR7 and TLR8 in kidney cancer were significantly higher than those in the corresponding normal

TABLE 4 | The characteristics of patients for RT-qPCR.

Patients	Sex	Years of age	Tumor location	Tumor size (cm)	TNM stage	Histological type
Sample1	Male	40	Right	4.8 × 4 × 3.5	$T_1N_0M_0$	KIRC
Sample2	Male	50	Right	2.5 × 2 × 2	$T_1N_0M_0$	KIRC
Sample3	Female	41	Left	4.5 × 4 × 4	$T_1N_0M_0$	KIRC

TABLE 5 | The characteristics of patients for FC.

Patients	Sex	Years of age	Tumor location	Tumor size (cm)	TNM stage	Histological type
Sample4	Male	38	Right	2.5 × 2 × 1.5	$T_1N_0M_0$	KIRC
Sample5	Male	54	Right	2.5 × 2 × 2	$T_1N_0M_0$	KIRC
Sample6	Female	64	Left	9 × 9 × 7.5	$T_2N_0M_0$	KIRC

tissues, while the expression of TLR5 was significantly lower in kidney cancer. Moreover, we further studied the expressions of TLRs in KIRC by UALCAN. The results showed that the expressions of TLRs transcripts in KIRC were elevated than that in adjacent nontumor tissues, except for TLR5. We then evaluated the expressions of all TLRs in KIRC tumor tissues by GEPIA and found that TLR3 was particularly highly expressed in KIRC, followed by TLR4 and TLR2, which was consistent with the results of UALCAN database. Then we further evaluated the expression levels of TLRs at different stages of KIRC, and discovered that the expressions of TLR3 and TLR4 were significantly different at different stages, and TLR3 and TLR4 were significantly increased in the early stage of KIRC. Finally, we also used clinical samples to verify that the expressions of TLR3 and TLR4 were significantly elevated in the early stage of KIRC by RT-qPCR, FC and IHC. These results suggested that TLRs were likely to be important biomarkers for early diagnosis of KIRC, especially TLR3 and TLR4.

Next, we continued to study the effects of TLRs on the survival outcome of KIRC. Overexpressions of TLR1, TLR3, TLR4 and TLR8 significantly prolonged the OS in patients with KIRC, while downregulation of TLR9 significantly prolonged the OS. Moreover, Overexpressions of TLR1, TLR3 and TLR4 significantly prolonged the DFS in patients with KIRC. All of these results suggested that TLRs had the potential to become important biomarkers for predicting prognosis in patients with KIRC, especially TLR1, TLR3, and TLR4.

In order to learn more about TLRs and to understand the possible mechanisms of the effects of TLRs on KIRC patients, we investigated the genetic alteration of TLRs and co-expression of TLRs in KIRC using cBioPortal. There were frequent genetic alterations of TLRs in KIRC. Elevated mRNA expression and deep deletion were the most common changes. Some studies have shown that many factors contribute to the occurrence and development of tumors, and genetic alterations play an indispensable role in this process (Yap et al., 2015; Zeng et al., 2019). What's more, co-expression of TLRs was found a clear association, suggesting that all of these TLRs play a momentous synergistic role in the occurrence and development of KIRC.

Next, we concentrated on the PPI network, GO analysis and KEGG pathway analysis of TLRs. Not surprisingly, the functions of these TLRs were mainly related to toll-like receptor signaling pathway, activation of innate immune response, pattern recognition receptor signaling pathway, positive regulation of innate immune response, innate immune response-activating signal transduction, positive regulation of defense response and regulation of innate immune response. The GO and KEGG pathway analyses of TLRs indicated that TLRs were mainly associated with regulation of leukocyte activation, immune response-regulating signaling pathway, adaptive immune response, lymphocyte mediated immunity, leukocyte cell-cell adhesion, positive regulation of cytokine production, interferon-gamma production, regulation of immune effector process, regulation of cell-cell adhesion, cytokine binding, positive regulation of defense response, MHC protein complex, protein complex involved in cell adhesion, cytokine receptor activity, cytokine receptor binding, antigen binding, TNF signaling pathway, Th1 and Th2 cell differentiation, Toll-like receptor signaling pathway, Th17 cell differentiation, cell adhesion molecules and JAK-STAT signaling pathway. Some studies have demonstrated that tumorigenesis is closely related to immune dysfunction (Raja et al., 2018; Nakamura et al., 2020). Moreover, our study indicated that TLRs were closely related to adaptive immunity, innate immunity and other immune-related processes, and that genetic alterations in TLRs were very common in KIRC, so we have every reason to believe that the occurrence of KIRC is closely related to TLRs.

Next, through FunRich, we found that the most critical gene that affected the interaction between TLRs was MYD88. Most TLRs depend on MYD88 for the regulation of multiple signal pathways and immune responses (Kim et al., 2019). MyD88 is involved in the development of various cancers by acting downstream of TLRs (Skorka et al., 2021). The results of our study and previous conclusions suggested that MYD88 played a bridging role in human immune homeostasis mediated by TLRs.

In order to learn more about TLRs-related information, we have also explored TLRs-related transcription factors. Our study found that the key transcription factors of TLRs were RELA,

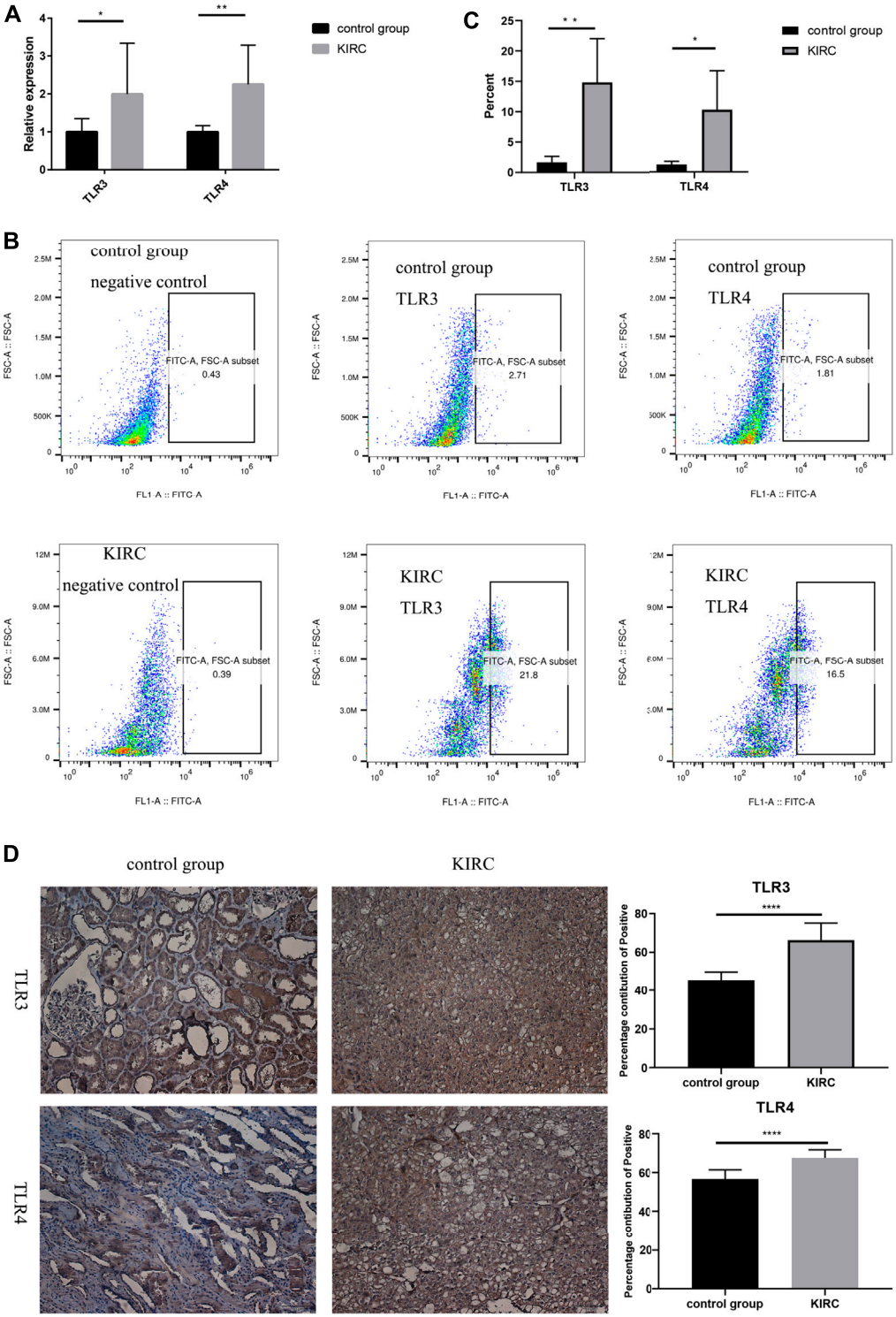


FIGURE 10 | Differential expressions of TLR3 and TLR4 in KIRC and adjacent nontumor tissues. **(A)** Expression levels of TLR3 mRNA and TLR4 mRNA in KIRC relative to adjacent nontumor tissues. **(B, C)** The proportion of $TLR3^+$ or $TLR4^+$ cells in KIRC and adjacent nontumor tissues was detected by flow cytometry. **(D)** The percentage contribution of TLR3+ or TLR4+ cells in KIRC and adjacent nontumor tissues was detected by IHC. Control group: adjacent nontumor tissues. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$.

NFKB1, IRF8, IRF3 and HIF1A. Previous studies demonstrated that RELA phosphorylation involved in the progression of various diseases including inflammatory disease and cancer by regulating NF-κB signaling (Lu and Yarbrough, 2015) and RELA also played a key role in mediating oncogene-induced aging (Lesina et al., 2016). NFKB1 is a cancer and inflammation inhibitor that plays an inhibitory role in the occurrence and development of a number of cancers by inhibiting the NF-κB signaling pathway (Cartwright et al., 2016; Concetti and Wilson, 2018). In addition, studies have shown that the loss of NFKB1 can lead to inflammation and the progression of cancer by increasing the expression of TNF (Low et al., 2020). IRF8, a tumor suppressor, is also a potential therapeutic option to overcome tumor drug resistance (Wu et al., 2020). IRF3, interferon regulatory factor 3, a tumor suppressor, plays an important role in inhibiting infection and cancer (King et al., 2017; Tian et al., 2020). HIF1A is a hypoxia inducible factor, and its absence increases tumor aggressiveness and metastatic activity (Tiwari et al., 2020). All these results provide insight into the complicated relationship among KIRC, TLRs and transcription factors. In addition, it also provides a further basis for TLRs to become early diagnostic biomarkers and judge the prognosis of patients with KIRC.

The tumor promoting or anticancer effects of TLRs may be related to the tumor microenvironment of immune cell infiltration and the types of cancers (Patra et al., 2020), therefore, we investigated the relationship between the expressions of TLRs in KIRC and the levels of immune cells infiltration by TIMER. Surprisingly, our results showed that the expressions of all TLRs in KIRC were significant positively correlated with the infiltration levels of dendritic cells, macrophages, neutrophils, B cells, $CD8^+$ T cells and $CD4^+$ T cells. There is growing evidence that immune cells infiltration is an important determinant of tumor therapeutic response and clinical outcome (Bindea et al., 2013; Liu et al., 2017). These results, combined with the differential expression of TLRs in KIRC and the significant effects of TLRs expressions on the prognosis of patients with KIRC, indicate that TLRs have the potential to be early diagnostic biomarkers of KIRC and biomarkers for judging the prognosis and immune status of KIRC patients. Further studies are needed to verify our results and explore how TLRs affect the immune microenvironment of KIRC.

CONCLUSION

In summary, the expression levels of TLRs in KIRC were generally different compared with adjacent normal tissues. Moreover, the expressions of TLR3 and TLR4 elevated significantly in the early stage of KIRC. Different TLRs had different effects on the prognosis of KIRC patients. TLRs can be used as important biomarkers for early diagnosis and prognosis assessment in patients with KIRC, especially TLR3 and TLR4. There were general genetic alterations and obvious co-expression correlation of TLRs in KIRC. The PPI network between TLRs was rather complex, and the key gene connecting the TLRs interaction was MYD88. The GO analysis and KEGG pathway analysis indicated that TLRs were closely related to adaptive immunity, innate immunity and other immune-related processes. RELA, NFKB1, IRF8, IRF3 and HIF1A were key transcription factors regulating the expressions of TLRs. What's more, the expressions of all TLRs in KIRC were significantly positively correlated with the infiltration levels of dendritic cells, macrophages, neutrophils, B cells, $CD8^+$ T cells and $CD4^+$ T cells. Taken together, the occurrence and development of KIRC is closely related to TLRs, and TLRs have the potential to be early diagnostic biomarkers of KIRC and biomarkers for judging the prognosis and immune status of KIRC patients. The results of our study may provide new insights into the selection of KIRC immunotherapy targets.

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.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of The First Affiliated Hospital of Guangxi Medical University (Approval Number: 2021KY-E-182). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

XZ, BG and ZM contributed to conception and design of the study. XZ and BG performed RT-qPCR, flow cytometry and IHC. XZ, BG QL and ZM performed the statistical analysis. QL collected the samples. XZ and BG wrote the manuscript. All authors have read and approved the manuscript.

FUNDING

This work was supported by the grants from the National Natural Science Foundation of China (81,770,759), the National Key Research and Development Program of China (2017YFC0908000), Major Project of Guangxi Innovation Driven (AA18118016), Guangxi key Laboratory for Genomic and Personalized Medicine (grant number 16–380-54, 17–259–45, 19–050–22, 19–185–33, 20–065–33), Guangxi Science and Technology base and Talent Project (2019AC17009) and Guangxi Clinical Research Center for Urology and Nephrology (2020AC03006).

REFERENCES

Angulo, J. C., Manini, C., Lopez, J. I., Pueyo, A., Colas, B., and Ropero, S. (2021). The Role of Epigenetics in the Progression of Clear Cell Renal Cell Carcinoma and the Basis for Future Epigenetic Treatments. *Cancers (Basel)* 13 (9). doi:10.3390/cancers13092071

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

Cartwright, T., Perkins, N. D., and Wilson, C. L. (2016). NFKB1: a Suppressor of Inflammation, Ageing and Cancer. *FEBS J.* 283 (10), 1812–1822. doi:10.1111/febs.13627

Chan, J. K., Roth, J., Oppenheim, J. J., Tracey, K. J., Vogl, T., Feldmann, M., et al. (2012). Alarmins: Awaiting a Clinical Response. *J. Clin. Invest.* 122 (8), 2711–2719. doi:10.1172/JCI62423

Chandrashekar, D. S., Bashel, B., Balasubramanya, S. A. H., Creighton, C. J., Ponce-Rodriguez, I., Chakravarthi, B., et al. (2017). UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia* 19 (8), 649–658. doi:10.1016/j.neo.2017.05.002

Chen, C. Y., Kao, C. L., and Liu, C. M. (2018). The Cancer Prevention, Anti-inflammatory and Anti-oxidation of Bioactive Phytochemicals Targeting the TLR4 Signaling Pathway. *Int. J. Mol. Sci.* 19 (9), 1. doi:10.3390/ijms19092729

Concetti, J., and Wilson, C. L. (2018). NFKB1 and Cancer: Friend or Foe. *Cells* 7 (9), 1. doi:10.3390/cells7090133

Dajon, M., Iribarren, K., and Cremer, I. (2017). Toll-like Receptor Stimulation in Cancer: A Pro- and Anti-tumor Double-Edged Sword. *Immunobiology* 222 (1), 89–100. doi:10.1016/j.imbio.2016.06.009

Fonseka, P., Pathan, M., Chitti, S. V., Kang, T., and Mathivanan, S. (2020). FunRich Enables Enrichment Analysis of OMICs Datasets. *J. Mol. Biol.* 433, 166747. doi:10.1016/j.jmb.2020.166747

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 (269), pl1. doi:10.1126/scisignal.2004088

Han, H., Cho, J. W., Lee, S., Yun, A., Kim, H., Bae, D., et al. (2018). TRRUST V2: an Expanded Reference Database of Human and Mouse Transcriptional Regulatory Interactions. *Nucleic Acids Res.* 46 (D1), D380–D6. doi:10.1093/nar/gkx1013

Hsieh, J. J., Purdue, M. P., Signoretti, S., Swanton, C., Albiges, L., Schmidinger, M., et al. (2017). Renal Cell Carcinoma. *Nat. Rev. Dis. Primers* 3, 17009. doi:10.1038/nrdp.2017.9

Kawasaki, T., and Kawai, T. (2014). Toll-like Receptor Signaling Pathways. *Front. Immunol.* 5, 461. doi:10.3389/fimmu.2014.00461

Kim, Y. C., Lee, S. E., Kim, S. K., Jang, H. D., Hwang, I., Jin, S., et al. (2019). Toll-like Receptor Mediated Inflammation Requires FASN-dependent MYD88 Palmitoylation. *Nat. Chem. Biol.* 15 (9), 907–916. doi:10.1038/s41589-019-0344-0

King, K. R., Aguirre, A. D., Ye, Y. X., Sun, Y., Roh, J. D., Ng, R. P., Jr., et al. (2017). IRF3 and Type I Interferons Fuel a Fatal Response to Myocardial Infarction. *Nat. Med.* 23 (12), 1481–1487. doi:10.1038/nm.4428

Lesina, M., Wormann, S. M., Morton, J., Diakopoulos, K. N., Korneeva, O., Wimmer, M., et al. (2016). RelA Regulates CXCL1/CXCR2-dependent Oncogene-Induced Senescence in Murine Kras-Driven Pancreatic Carcinogenesis. *J. Clin. Invest.* 126 (8), 2919–2932. doi:10.1172/JCI86477

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 (21), e108–e10. doi:10.1158/0008-5472.CAN-17-0307

Linehan, W. M., and Ricketts, C. J. (2019). The Cancer Genome Atlas of Renal Cell Carcinoma: Findings and Clinical Implications. *Nat. Rev. Urol.* 16 (9), 539–552. doi:10.1038/s41585-019-0211-5

Linehan, W. M., and Ricketts, C. J. (2013). The Metabolic Basis of Kidney Cancer. *Semin. Cancer Biol.* 23 (1), 46–55. doi:10.1016/j.semcancer.2012.06.002

Linehan, W. M., Srinivasan, R., and Schmidt, L. S. (2010). The Genetic Basis of Kidney Cancer: a Metabolic Disease. *Nat. Rev. Urol.* 7 (5), 277–285. doi:10.1038/nrurol.2010.47

Liu, X., Wu, S., Yang, Y., Zhao, M., Zhu, G., and Hou, Z. (2017). The Prognostic Landscape of Tumor-Infiltrating Immune Cell and Immunomodulators in Lung Cancer. *Biomed. Pharmacother.* 95, 55–61. doi:10.1016/j.biopha.2017.08.003

Low, J. T., Christie, M., Ernst, M., Dumoutier, L., Preaudet, A., Ni, Y., et al. (2020). Loss of NFKB1 Results in Expression of Tumor Necrosis Factor and Activation of Signal Transducer and Activator of Transcription 1 to Promote Gastric Tumorigenesis in Mice. *Gastroenterology* 159 (4), 1444–1458. e15. doi:10.1053/j.gastro.2020.06.039

Lu, X., and Yarbrough, W. G. (2015). Negative Regulation of RelA Phosphorylation: Emerging Players and Their Roles in Cancer. *Cytokine Growth Factor. Rev.* 26 (1), 7–13. doi:10.1016/j.cytogfr.2014.09.003

Marin-Acevedo, J. A., Dholaria, B., Soyano, A. E., Knutson, K. L., Chumsri, S., and Lou, Y. (2018). Next Generation of Immune Checkpoint Therapy in Cancer: New Developments and Challenges. *J. Hematol. Oncol.* 11 (1), 39. doi:10.1186/s13045-018-0582-8

Moch, H., Cubilla, A. L., Humphrey, P. A., Reuter, V. E., and Ulbright, T. M. (2016). The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part A: Renal, Penile, and Testicular Tumours. *Eur. Urol.* 70 (1), 93–105. doi:10.1016/j.eururo.2016.02.029

Nakamura, K., Smyth, M. J., and Martinet, L. (2020). Cancer Immunoediting and Immune Dysregulation in Multiple Myeloma. *Blood* 136 (24), 2731–2740. doi:10.1182/blood.2020006540

Nouri, Y., Weinkove, R., and Perret, R. (2021). T-cell Intrinsic Toll-like Receptor Signaling: Implications for Cancer Immunotherapy and CAR T-Cells. *J. Immunother. Cancer* 9 (11), e003065. doi:10.1136/jitc-2021-003065

Patra, M. C., Shah, M., and Choi, S. (2020). Toll-like Receptor-Induced Cytokines as Immunotherapeutic Targets in Cancers and Autoimmune Diseases. *Semin. Cancer Biol.* 64, 61–82. doi:10.1016/j.semcancer.2019.05.002

Petitprez, F., Ayadi, M., de Reyniès, A., Fridman, W. H., Sautès-Fridman, C., and Job, S. (2021). Review of Prognostic Expression Markers for Clear Cell Renal Cell Carcinoma. *Front. Oncol.* 11, 643065. doi:10.3389/fonc.2021.643065

Pichler, M., Hutterer, G. C., Chromecki, T. F., Jesche, J., Kampel-Kettner, K., Groselj-Strele, A., et al. (2013). Predictive Ability of the 2002 and 2010 Versions of the Tumour-Node-Metastasis Classification System Regarding Metastasis-free, Cancer-specific and Overall Survival in a European Renal Cell Carcinoma Single-centre Series. *BJU Int.* 111 (4 Pt B), E191–E195. doi:10.1111/j.1464-410X.2012.11584

Qin, C., Sun, L. J., Cui, L., Cao, Q., Zhu, J., Li, P., et al. (2013). Application of the Revised Tumour Node Metastasis (TNM) Staging System of clear Cell Renal Cell Carcinoma in Eastern China: Advantages and Limitations. *Asian J. Androl.* 15 (4), 550–557. doi:10.1038/aja.2012.161

Raja, J., Ludwig, J. M., Gettinger, S. N., Schalper, K. A., and Kim, H. S. (2018). Oncolytic Virus Immunotherapy: Future Prospects for Oncology. *J. Immunother. Cancer* 6 (1), 140. doi:10.1186/s40425-018-0458-z

Rameshbabu, S., Labadie, B. W., Argulian, A., and Patnaik, A. (2021). Targeting Innate Immunity in Cancer Therapy. *Vaccines* 9 (2), 138. doi:10.3390/vaccines9020138

Rhodes, D. R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., et al. (2004). ONCOMINE: a Cancer Microarray Database and Integrated Data-Mining Platform. *Neoplasia* 6 (1), 1–6. doi:10.1016/s1476-5586(04)80047-2

Sanchez-Gastaldo, A., Kempf, E., Gonzalez Del Alba, A., and Duran, I. (2017). Systemic Treatment of Renal Cell Cancer: A Comprehensive Review. *Cancer Treat. Rev.* 60, 77–89. doi:10.1016/j.ctrv.2017.08.010

Shetab Boushehri, M. A., and Lamprecht, A. (2018). TLR4-Based Immunotherapeutics in Cancer: A Review of the Achievements and Shortcomings. *Mol. Pharm.* 15 (11), 4777–4800. doi:10.1021/acs.molpharmaceut.8b00691

Siegel, R. L., Miller, K. D., and Jemal, A. (2018). Cancer Statistics. *CA Cancer J. Clin.* 68 (1), 7–30. doi:10.3322/caac.21442

Skorka, K., Wlasiuk, P., Karczmarczyk, A., and Giannopoulos, K. (2021). Aberrant Expression of TLR2, TLR7, TLR9, Splicing Variants of TLR4 and MYD88 in Chronic Lymphocytic Leukemia Patients. *J. Clin. Med.* 10 (4). doi:10.3390/jcm10040867

Tamayo, P., Cho, Y. J., Tsherniak, A., Greulich, H., Ambrogio, L., Schouten-van Meeteren, N., et al. (2011). Predicting Relapse in Patients with Medulloblastoma by Integrating Evidence from Clinical and Genomic Features. *J. Clin. Oncol.* 29 (11), 1415–1423. doi:10.1200/JCO.2010.28.1675

Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). GEPIA: a Web Server for Cancer and normal Gene Expression Profiling and Interactive Analyses. *Nucleic Acids Res.* 45 (W1), W98–W102. doi:10.1093/nar/gkx247

Tian, M., Wang, X., Sun, J., Lin, W., Chen, L., Liu, S., et al. (2020). IRF3 Prevents Colorectal Tumorigenesis via Inhibiting the Nuclear Translocation of Beta-Catenin. *Nat. Commun.* 11 (1), 5762. doi:10.1038/s41467-020-19627-7

Tiwari, A., Tashiro, K., Dixit, A., Soni, A., Vogel, K., Hall, B., et al. (2020). Loss of HIF1A from Pancreatic Cancer Cells Increases Expression of PPP1R1B and Degradation of P53 to Promote Invasion and Metastasis. *Gastroenterology* 159 (5), 1882–1897. e5. doi:10.1053/j.gastro.2020.07.046

Urban-Wojciuk, Z., Khan, M. M., Oyler, B. L., Fahraeus, R., Marek-Trzonkowska, N., Nita-Lazar, A., et al. (2019). The Role of TLRs in Anti-cancer Immunity and Tumor Rejection. *Front. Immunol.* 10, 2388. doi:10.3389/fimmu.2019.02388

Vasaikar, S. V., Straub, P., Wang, J., and Zhang, B. (2018). LinkedOmics: Analyzing Multi-Omics Data within and across 32 Cancer Types. *Nucleic Acids Res.* 46 (D1), D956–D63. doi:10.1093/nar/gkx1090

Vidya, M. K., Kumar, V. G., Sejian, V., Bagath, M., Krishnan, G., and Bhatta, R. (2018). Toll-like Receptors: Significance, Ligands, Signaling Pathways, and Functions in Mammals. *Int. Rev. Immunol.* 37 (1), 20–36. doi:10.1080/08830185.2017.1380200

Wang, L., Yu, K., Zhang, X., and Yu, S. (2018). Dual Functional Roles of the MyD88 Signaling in Colorectal Cancer Development. *Biomed. Pharmacother.* 107, 177–184. doi:10.1016/j.biopha.2018.07.139

Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., et al. (2010). The GeneMANIA Prediction Server: Biological Network Integration for Gene Prioritization and Predicting Gene Function. *Nucleic Acids Res.* 38 (Web Server issue), W214–W220. doi:10.1093/nar/gkq537

Wu, H., You, L., Li, Y., Zhao, Z., Shi, G., Chen, Z., et al. (2020). Loss of a Negative Feedback Loop between IRF8 and AR Promotes Prostate Cancer Growth and Enzalutamide Resistance. *Cancer Res.* 80 (13), 2927–2939. doi:10.1158/0008-5472.CAN-19-2549

Wu, K., Zhang, H., Fu, Y., Zhu, Y., Kong, L., Chen, L., et al. (2018). TLR4/MyD88 Signaling Determines the Metastatic Potential of Breast Cancer Cells. *Mol. Med. Rep.* 18 (3), 3411–3420. doi:10.3892/mmr.2018.9326

Xiao, W., Xiong, Z., Xiong, W., Yuan, C., Xiao, H., Ruan, H., et al. (2019). Melatonin/PGC1A/UCP1 Promotes Tumor Slimming and Represses Tumor Progression by Initiating Autophagy and Lipid browning. *J. Pineal Res.* 67 (4), e12607. doi:10.1111/jpi.12607

Yap, N. Y., Rajandram, R., Ng, K. L., Pailoor, J., Fadzli, A., and Gobe, G. C. (2015). Genetic and Chromosomal Aberrations and Their Clinical Significance in Renal Neoplasms. *Biomed. Res. Int.* 2015, 476508. doi:10.1155/2015/476508

Zeng, Q., Sun, S., Li, Y., Li, X., Li, Z., and Liang, H. (2019). Identification of Therapeutic Targets and Prognostic Biomarkers Among CXC Chemokines in the Renal Cell Carcinoma Microenvironment. *Front. Oncol.* 9, 1555. doi:10.3389/fonc.2019.01555

Zhang, Y., Liu, J., Wang, C., Liu, J., and Lu, W. (2021). Toll-Like Receptors Gene Polymorphisms in Autoimmune Disease. *Front. Immunol.* 12, 672346. doi:10.3389/fimmu.2021.672346

Zhao, S., Zhang, Y., Zhang, Q., Wang, F., and Zhang, D. (2014). Toll-like Receptors and Prostate Cancer. *Front. Immunol.* 5, 352. doi:10.3389/fimmu.2014.00352

Zheng, J., Wang, L., Peng, Z., Yang, Y., Feng, D., and He, J. (2017). Low Level of PDZ Domain Containing 1 (PDZK1) Predicts Poor Clinical Outcome in Patients with clear Cell Renal Cell Carcinoma. *EBioMedicine* 15, 62–72. doi:10.1016/j.ebiom.2016.12.003

Zou, X., and Mo, Z. (2021). CYP2J2 Is a Diagnostic and Prognostic Biomarker Associated with Immune Infiltration in Kidney Renal Clear Cell Carcinoma. *Biomed. Res. Int.* 2021, 3771866. doi:10.1155/2021/3771866

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 Molecular Biosciences

ORIGINAL RESEARCH
published: 28 January 2022
doi: 10.3389/fmolb.2021.789923

Validated Impacts of N6-Methyladenosine Methylated mRNAs on Apoptosis and Angiogenesis in Myocardial Infarction Based on MeRIP-Seq Analysis

Yingjie Zhang[1†], Wenjie Hua[1†], Yini Dang[2†], Yihui Cheng[1†], Jiayue Wang[1], Xiu Zhang[1], Meiling Teng[1], Shenrui Wang[1], Min Zhang[2], Zihao Kong[2], Xiao Lu[1] and Yu Zheng[1]**

[1]Department of Rehabilitation Medicine, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China, [2]Department of Gastroenterology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China

OPEN ACCESS

Edited by:
William C. Cho,
QEH, Hong Kong SAR, China

Reviewed by:
Geon-Woo Kim,
University of California, San Diego, United States
Ivana Josipovic,
Darmstadt University of Technology, Germany

***Correspondence:**
Yu Zheng
zhengyu8710@163.com
Xiao Lu
luxiao1972@163.com

†These authors have contributed equally to this work.

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 05 October 2021
Accepted: 26 November 2021
Published: 28 January 2022

Citation:
Zhang Y, Hua W, Dang Y, Cheng Y, Wang J, Zhang X, Teng M, Wang S, Zhang M, Kong Z, Lu X and Zheng Y (2022) Validated Impacts of N6-Methyladenosine Methylated mRNAs on Apoptosis and Angiogenesis in Myocardial Infarction Based on MeRIP-Seq Analysis. Front. Mol. Biosci. 8:789923. doi: 10.3389/fmolb.2021.789923

Objectives: N6-methyladenosine (m^6A) is hypothesized to play a role in the regulation of pathogenesis of myocardial infarction (MI). This study was designed to compare m^6A-tagged transcript profiles to identify mRNA-specific changes on pathophysiological variations after MI.

Methods: N6-methyladenosine methylated RNA immunoprecipitation sequencing (MeRIP-seq) and RNA sequencing (RNA-seq) were interacted to select m^6A-modified mRNAs with samples collected from sham operated and MI rat models. m^6A methylation regulated mRNAs were interacted with apoptosis/angiogenesis related genes in GeneCards. Afterwards, MeRIP-quantitative real-time PCR (MeRIP-qRT-PCR) was performed to measure m^6A methylation level of hub mRNAs. m^6A methylation variation was tested under different oxygen concentration or hypoxic duration in H9c2 cells and HUVECs. In addition, Western blot and qRT-PCR were employed to detect expression of hub mRNAs and relevant protein level. Flow cytometry and Tunel assay were conducted to assess apoptotic level. CCK-8, EdU, and tube formation assay were performed to measure cell proliferation and tube formation ability.

Results: Upregulation of Mettl3 was firstly observed *in vivo* and *in vitro*, followed by upregulation of m^6A methylation level. A total of 567 significantly changed m^6A methylation peaks were identified, including 276 upregulated and 291 downregulated peaks. A total of 576 mRNAs were upregulated and 78 were downregulated. According to combined analysis of MeRIP-seq and RNA-seq, we identified 26 significantly hypermethylated and downregulated mRNAs. Based on qRT-PCR and interactive analysis, Hadh, Kcnn1, and Tet1 were preliminarily identified as hub mRNAs associated with apoptosis/angiogenesis. MeRIP-qRT-PCR assay confirmed the results from MeRIP-seq. With the inhibition of Mettl3 in H9c2 cells and HUVECs, downregulated m^6A methylation level of total RNA and upregulated expression of hub mRNAs were observed. Increased m^6A level was verified in the gradient context in terms of prolonged hypoxic duration and decreased oxygen concentration. Under simulated hypoxia, roles of Kcnn1 and Tet1 in angiogenesis and Hadh, Tet1, and Kcnn1 in apoptosis were further confirmed with our validation experiments.

Conclusion: Roles of m^{6}A-modified mRNA transcripts in the context of MI were preliminarily verified. In the context of m^{6}A methylation, three hub mRNAs were validated to impact the process of apoptosis/angiogenesis. Our study provided theoretical basis and innovative targets for treatment of MI and paved the way for future investigations aiming at exploring upstream epigenetic mechanisms of pathogenesis after MI.

Keywords: myocardial infarction, m^{6}A methylation, mRNA, angiogenesis, apoptosis

1 INTRODUCTION

Myocardial infarction (MI), due to the reduction or interruption of the blood supply of the coronary artery, always results in ischemia of the corresponding myocardium leading to myocardial necrosis (Saleh and Ambrose, 2018). It is characterized by an elevated ST-segment in the electrocardiogram, and is one of the most common causes of death worldwide (Lu et al., 2015). At present, the reperfusion therapy to restore the blood circulation of the heart has become a common method for the treatment of MI, however the subsequent reperfusion injury would impair endothelial function and aggravate myocardial cell death (Thygesen et al., 2007; Puymirat et al., 2019). Upon these concerns, exploration of strategies on compensating myocardial cell regeneration and death after MI becomes the most warranted task in this research field. In addition, understanding the microscopic regulations in MI might be essential to reveal the pathophysiological mechanisms behind MI and might shed light on uncovering novel therapies for the treatment of MI.

Robust studies have explored the mechanisms of myocardial cell regeneration and death after MI. Integrin-linked kinase (ILK) has been reported to be an important factor regulating apoptosis and angiogenesis. In hypoxic condition, upregulation of ILK increased phosphorylation of protein kinase B and mammalian target of rapamycin, resulting in enhanced mesenchymal stem cells (MSCs) survival and vascular endothelial growth factors expression level. In addition, transplantation of MSCs rich in ILK could further improve angiogenesis at 3 weeks (Zeng et al., 2017). Exosomes derived from TIMP2-modified MSCs significantly increased the expression of antiapoptotic bcl-2, followed by decreased proapoptotic Bax and pro-caspase9 level, and finally attenuated apoptosis in MI injury via Akt/Sfrp2 pathway *in vivo* (Ni et al., 2019). However, the upstream regulation of these pathways has not been well documented. Recent studies on epigenetic regulation have revealed the relationships between epigenetic modifications and cardiovascular diseases. Epigenetics, including the reversible modification of DNA and protein, were proven to independently regulate gene expression of DNA and protein. It was not until recently that RNA modification was believed to be the third layer of epigenetics, regulating RNA processing and metabolism. There has been uncovered with more than 100 modifications in RNAs, including 5-methylcytosine (m^{5}C), N6-methyladenosine (m^{6}A), and N1-methyladenosine (m^{1}A). Among which, m^{6}A methylation was demonstrated to be the common and abundant internal modification of eukaryotic messenger RNA (mRNA) (Meyer et al., 2012). It is a dynamic reversible process regulated by methyltransferases (writers), demethylases (erasers), and binding proteins (readers). ALKBH5 was responsible for reducing m^{6}A methylation and ALKBH5 (one of the demethylases) knockout mice exhibited decreased cardiac regenerative ability and cardiac function after neonatal apex resection (Han et al., 2021). The expression of METTL3 was increased in cardiac fibrotic tissue with chronic myocardial infarction. It promoted proliferation of cardiac fibroblasts, fibroblast-to-myofibroblast transition, and collagens accumulation, while silence of METTL3 (one of the methyltransferases) alleviated cardiac fibrosis in MI mice (Li et al., 2021). These findings emphasized the importance of m^{6}A methylation in individual mRNAs and provided new insights into therapeutic strategies. Nonetheless, there is limited knowledge of the whole picture of m^{6}A modification on mRNAs after MI and how performance of m^{6}A methylated mRNAs on the downstream functional phenotypes.

Upon the above concerns, we aimed to systematically compare the m^{6}A-tagged transcript profiles of heart tissue from rat MI models with those from sham operated (SO) rats to identify gene-specific changes in mRNA methylation and contribute to the development of MI. Specifically, we first observed the change of methyltransferases *in vivo* and *vitro* followed by the verification of m^{6}A methylation in the gradient context of prolonged hypoxic duration and decreased oxygen concentration. We then identified m^{6}A methylation regulated mRNAs between the SO group and the MI group by combined analysis of N6-methyladenosine methylated RNA immunoprecipitation sequencing (MeRIP-seq) and RNA sequencing (RNA-seq), then top 10 hypermethylated and downregulated mRNAs were selected and validated in rats through quantitative real-time reverse transcription-polymerase chain reaction (qRT-PCR). Afterward, we interacted the hub mRNAs with genes related to apoptosis and angiogenesis in open-source datasets. We finally validated the role of methyltransferase in the regulation of m^{6}A methylation and selected hub mRNAs in angiogenesis and apoptosis. With the exploration of potential roles for the m^{6}A modified mRNA transcripts in the physiological and pathological mechanisms underlying MI, a theoretical basis and innovative targets could be identified for the treatment of MI.

2 MATERIALS AND METHODS

2.1 Models

2.1.1 Animal MI Model

Male Sprague-Dawley (SD) rats aged 10 weeks and weighing 250–300 g were collected from Beijing Vital River Laboratory Animal Technology Co. Ltd. The current study was carried out

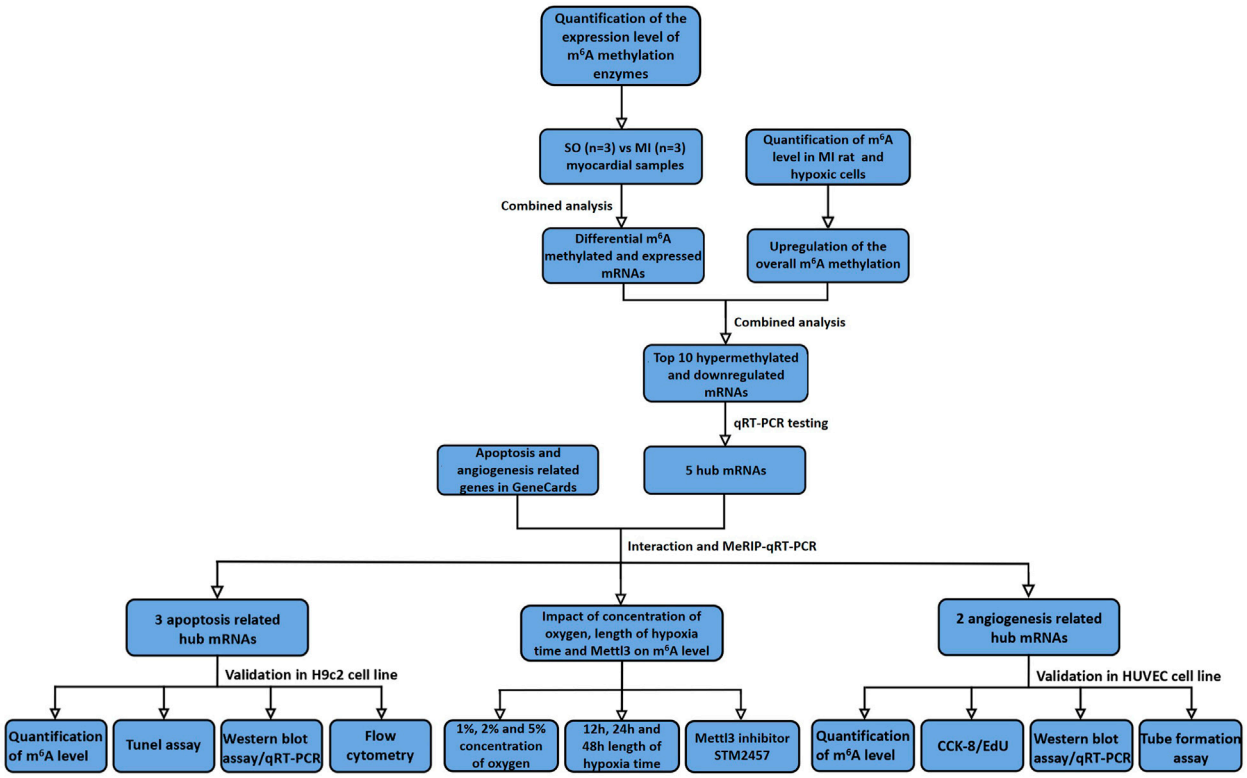


FIGURE 1 | Flowchart of study procedure. SO, sham operated; MI, myocardial infarction; qRT-PCR, quantitative reverse transcription polymerase chain reaction; MeRIP-qRT-PCR, MeRIP-quantitative real-time PCR; CCK-8, cell counting kit-8; EdU, 5-ethynyl-20-deoxyuridine.

in accordance with the guidelines of the Chinese Council on Animal Protection and approved by the Institutional Animal Care and Use Committee of Nanjing Medical University (reference number of 10091). SD rats were randomly divided into the sham operated (SO) group and the myocardial infarction (MI) group. Ligation of left anterior descending coronary artery was conducted in the MI group and myocardial ischemia was confirmed with elevated ST-segment in electrocardiogram. However, string went through the corresponding myocardial region without ligation in the SO group. Finally, heart samples were collected from the infarction area in the MI group and the corresponding region in the SO group. Heart samples were refrigerated at −80°C for further detection.

2.1.2 Cell Hypoxic Model

H9c2 cells were cultured in Dulbecco's modified eagle medium (Gibico, Waltham, CA) supplemented with 10% fetal bovine serum (Gibico, Waltham, CA) and 1% of penicillin/streptomycin at 37°C in an incubator with 95% humid air and 5% CO_2 (Wang M. et al., 2020). Human umbilical vein endothelial cells (HUVECs) were maintained in endothelial cell growth medium-2 bullet kit (Lonza, Basel, BS, CH) at 37°C in an incubator with 95% humid air and 5% CO_2 (Chen et al., 2020). To simulate myocardial ischemia, H9c2 cells and HUVECs were managed with hypoxia. Specifically, they were placed in a hypoxic incubator containing 94% N_2, 5% CO_2, and 1% O_2 for 24 h. Meanwhile, the control group was maintained in a normal atmosphere of 95% air and 5% CO_2 at 37°C (Zhu et al., 2021). Afterward, H9c2 cells and HUVECs were cultured in consistent 1% O_2, 2% O_2, and 5% O_2 hypoxic condition for 24 h to explore the impact of different concentrations of oxygen on m^{6}A methylation level. These cells were also cultured with 1% oxygen concentration for 12, 24, and 48 h to explore the impact of different hypoxic duration on m^{6}A methylation level. H9c2 cells and HUVECs were then treated with 25 μM Mettl3 inhibitor (STM2457) for 24 h to explore the role of Mettl3 on m^{6}A methylation level (Yankova et al., 2021).

2.2 MeRIP-Seq and Bioinformatic Analysis

The detailed procedure of MeRIP-Seq analysis and validation of newly discovered hub mRNAs on apoptosis and angiogenesis are demonstrated in **Figure 1**.

2.2.1 Methylated RNA Immunoprecipitation Sequencing

Total RNA was isolated and purified using TRIzol reagent (Invitrogen, Carlsbad, CA) following the manufacturer's procedure. The RNA amount and purity of each sample was quantified using NanoDrop ND-1000 (NanoDrop, Wilmington, DE). The RNA integrity was assessed by Bioanalyzer 2100 (Agilent, Santa Clara, CA) with RIN number >7.0 and confirmed by electrophoresis with denaturing agarose gel. Afterward, Epicentre Ribo-Zero Gold Kit (Illumina, San Diego, CA) was

used to deplete ribosomal RNA (rRNA) from total RNA. The ribosomal-depleted RNA was fragmented into small pieces using Magnesium RNA Fragmentation Module (New England Biolabs, Ipswich, MA). Then the cleaved RNA fragments were incubated with m^{6}A-specific antibody (Synaptic Systems, Walldorf, BW, DE) in immunoprecipitation (IP) buffer (50 mM Tris-HCl, 750 mM NaCl, and 0.5% Igepal CA-630). Subsequently, the IP RNA was reverse transcribed to create cDNA by SuperScript™ II Reverse Transcriptase (Invitrogen, Waltham, CA). Eluted fragments containing m^{6}A and untreated input control fragments were converted to construct the strand-specific cDNA library by dUTP method (Dominissini et al., 2013). The average insert size of the final cDNA library was 300 ± 50 bp. We finally performed 2 × 150 bp paired-end sequencing (PE150) with an illumina Novaseq™ 6000 (LC-Bio Technology CO., Ltd., Hangzhou, China) (Meng et al., 2014; Li M. et al., 2019).

2.2.2 RNA-Seq

Total RNA was isolated and purified using TRIzol reagent (Invitrogen, Carlsbad, CA). For RNA-seq analysis, rRNA was depleted, according to the protocol of the Epicentre Ribo-Zero Gold Kit (Illumina, San Diego, CA). Subsequently, the ribosomal-depleted RNA was fragmented into small pieces using Magnesium RNA Fragmentation Module (New England Biolabs, Ipswich, MA). The fragments were converted to construct the strand-specific cDNA library with dUTP method and were sequenced by illumina Novaseq™ 6000 (LC-Bio Technology CO., Ltd., Hangzhou, China) (Wang Q. et al., 2020).

2.2.3 Bioinformatic Analysis

Fastp software was used to remove the reads containing adaptor contamination, low quality bases, and undetermined bases with default parameters (Chen et al., 2018). Then sequence quality of IP and input samples were verified with fastp. HISAT2 were used to map reads to the reference genomeRattus norvegicus (Version 101) (Kim et al., 2015). Mapped reads of IP and input libraries were then provided with "exomePeak" package (Meng et al., 2014). m^{6}A peaks from the corresponding libraries were visualized with IGV software (Broad Institute, Cambridge, MA) (Thorvaldsdóttir et al., 2013). MEME and HOMER software were used for *de novo* and known motif findings followed by localization of the motif with respect to peak summit (Bailey et al., 2009). The information of m^{6}A peaks was obtained by intersection with gene architecture using "ChIPseeker" package (Yu et al., 2015). Then StringTie was used to obtain expression for all mRNAs from input libraries by calculating Fragments Per Kilobase of exon model per Million mapped fragments (FPKM) (Pertea et al., 2015). The differentially expressed mRNAs were selected with $|\log2FC|>1$ and p value < 0.05 by "edgeR" package (Robinson et al., 2010).

2.3 Validation

Expression of top 10 hypermethylated and downregulated mRNAs was first validated with qRT-PCR testing. Subsequently, the m^{6}A RNA methylation level was measured in histological and cellular level. qRT-PCR and Western blot were performed to verify the expression of genes and corresponding proteins. MeRIP-quantitative real-time PCR was performed to measure m^{6}A methylation level of hub genes. Flow cytometry and Tunel assay were carried out to assess the apoptosis level. Cell counting kit-8 (CCK-8), 5-ethynyl-20-deoxyuridine (EdU), and tube formation assay were performed to test cell proliferation and tube formation ability.

2.3.1 Quantification of Total m^{6}A Methylation Level

m^{6}A RNA methylation level was detected using the EpiQuik™ m^{6}A RNA Methylation Quantification kit (Epigentek, New York, NY) according to the manufacturer's protocol (Liu et al., 2020). Briefly, a negative control and a standard curve consisting of six different concentrations (ranged from 0.02 to 1 ng of m^{6}A) were prepared. There was 200 ng of total RNA used for each reaction. After RNA binding to the 96-well plates, diluted capture anti-m^{6}A antibodies were added, then 100 μl of developer solution to each well, and incubated at room temperature for 10 min without light. There was 100 μl of stop solution added afterward to each well to stop enzyme reaction. The optical density at 450 nm was measured using a microplate reader (Thermo Fisher Scientific, Waltham, CA). Percentage of m^{6}A within the total RNA was calculated (Dang et al., 2020; Liu et al., 2020).

2.3.2 Screening Strategy for Angiogenesis and Apoptosis Related Genes

We first obtained m^{6}A methylation regulated hub genes by qRT-PCR, then, the hub genes were interacted with genes related to angiogenesis and apoptosis in the GeneCards. Results were visualized by Venn diagram (Liang et al., 2019).

2.3.3 MeRIP-Quantitative Real-Time PCR

RNA sample from myocardial tissue and H9c2 cells was fragmented (300 nt) after incubation with fragmentation buffer at 94°C for 4 min. A total of 5% of fragmented RNA was saved as input control. The procedure of m^{6}A-IP sample preparation was similar to that of MeRIP-seq (Deng et al., 2021). Finally, both input control and m^{6}A-IP samples were subjected to qRT-PCR with gene-specific primers listed in **Supplementary Table S1**.

2.3.4 qRT-PCR

Total RNA was extracted from cells and myocardial tissues and complementary DNA (cDNA) was synthesized using total RNA with the PrimeScript™ RT reagent kit with gDNA Eraser (TaKaRa, Kusatsu, Japan) according to the manufacturer's instructions. qRT-PCR was performed with the Pro-17 Steponeplus system (Applied Biosystems, Carlsbad, CA) (Saeidi et al., 2018). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as internal control to normalize the expression of genes. There are 25 pairs of primer listed in **Supplementary Table S2**. Relative expression of differentially expressed genes were analyzed with the $2^{-\triangle\triangle CT}$ method (Jozefczuk and Adjaye, 2011).

2.3.5 Western Blot Analysis

H9c2 cells and HUVECs were lysed using RIPA lysis buffer (Beyotime, Shanghai, China). Protein concentration was determined using the BCA Protein Assay kit (New Cell and Molecular Biotech, Suzhou, China). Specifically, the corresponding protein was separated with 7.5 and 10% SDS-

PAGE and transferred to polyvinylidene difluoride membranes. Afterward, the membranes were blocked with 5% skim milk for 2 h in room temperature (Qiu et al., 2020). They were incubated with primary antibodies against GAPDH (Proteintech, Wuhan, China); Tet1 (Signalway Antibody, College Park, MD); Hadh (Signalway Antibody, College Park, MD); Kcnn1 (Signalway Antibody, College Park, MD); Caspase3 (Proteintech, Wuhan, China); and Cleavd-caspase3 (Cell Signaling Technology, Danvers, MA); bcl-2 (Abcam, Cambridge, MA); Bax (Abcam, Cambridge, MA); HIF-1α (Abcam, Cambridge, MA); and Mettl3 (Proteintech, Wuhan, China) overnight at 4°C. Then, the membranes were incubated with horseradish peroxidase-conjugated secondary antibodies (Abcam, Cambridge, MA) for 1 h. The bands were analyzed with chemiluminescence Western blotting detection system (Tanon, Shanghai, China) (Gu et al., 2018; Ren et al., 2018; Li X. et al., 2020).

2.3.6 Flow Cytometry

Annexin V-Fluorescein isothiocyanate (FITC) and propidium iodide (PI) staining kit (Vazyme, Nanjing, China) were used to identify apoptotic H9c2 cells. Flow cytometry analysis was performed through Cytoflex (Beckman Coulter, Brea, CA) and data was analyzed using FlowJo software (Tree Star, SanCarlos, CA). The second quadrant (FITC+/PI+) showed the late apoptotic cells and the fourth quadrant (FITC+/PI-) presented the early apoptotic cells (Luo et al., 2019; Zhang et al., 2019; Wei J. et al., 2020; Li Y. et al., 2020).

2.3.7 Tunel Analysis

TUNEL BrightGreen Apoptosis Detection kit (Vazyme, Nanjing, China) was used for Tunel staining based on the manufacturer's instructions. Briefly, the H9c2 was fixed in 4% PFA at 4°C for 25 min. Then the cells were incubated with Proteinase K (20 μg/ml) for 5 min. After incubation with 1×equilibration buffer for 30 min, the cells were treated with BrightGreen Labeling Mix and Recombinant TdT Enzyme for 1 h at 37°C. The nucleus was stained with DAPI (2 μg/ml) away from light for 10 min (Shi et al., 2021). The Tunel staining images were photographed with a fluorescence microscope (Bio-tek, Winusky, VT) (Zhang et al., 2016; Li T. et al., 2019; Luo et al., 2019; Sharma et al., 2021).

2.3.8 Cell Counting Kit-8

Cell suspension with a concentration of 5×10^3 each well was digested with trypsin and inoculated into a 96-well plate with 100 μL per well (Yu et al., 2021). After conventional culture for 24, 48, and 72 h, 10 μl CCK-8 solution (MedChemExpress, Shanghai, China) was added to each well and incubated for another 2 h. The absorbance value at 450 nm was finally measured by enzyme labeling instrument (Chen et al., 2019; Chen and Ling, 2019; Li X. et al., 2020; Zhou et al., 2020; Sun et al., 2021).

2.3.9 EdU Assay

EdU incorporation assay kit (Ribobio, Guangzhou, China) was used for the measurement of cell proliferation (Liu et al., 2018). The cell suspension with a concentration of 8×10^3 each well was digested with trypsin and inoculated into a 96-well plate with 100 μl per well, followed by the addition of 50 μM EdU diluent. After 2 h, cells were fixed in 4% paraformaldehyde, cultivated with 100 μl of 0.5% Triton X-100 and mixed with 100 μL of 1×Apollo® 567 fluorescent staining solution (Xu et al., 2019). The cell nucleus was finally subjected to DAPI staining in a dark environment. Images were finally obtained from Cytation1 (Bio-tek, Winusky, VT).

2.3.10 Tube Formation Assay

Matrigel (BD Biosciences, Franklin Lake, NJ) was used for tube formation assay to assess the tube-forming ability of HUVECs (Cheng et al., 2017). Briefly, it was dissolved at 4°C overnight, and each well of the 96-well plate was then coated with 50 μl of Matrigel. The plate was then left to polymerize at 37°C for 1 h incubation. The HUVECs (2×10^4 cells/100 μl) were then seeded into each well. After 6 h of incubation at 37°C, tube formation was photographed at ×50 magnification. Tube formations were calculated with the number of branches using ImageJ software (National Institutes of Health, Bethesda, MD) (Li et al., 2017; Hanlon et al., 2019).

2.4 Statistical Analysis

Data were presented as mean with standard deviation. Results of qRT-PCT, MeRIP-qRT-PCR, total m^{6}A methylation level, tube formation assay, and flow cytometry were analyzed with Student's t-test. CCK-8 assay data across 24, 48, and 72 h was analyzed with two-way analysis of variance (ANOVA). Statistical significance was considered with p value less than 0.05. All analyses were performed using SPSS 22.0 (International Business Machines Corporation, Armonk, NY) and GraphPad Prism 9 (GraphPad, San Diego, CA).

3 RESULTS

3.1 Overview of Methylated RNA Immunoprecipitation Sequencing

In the MeRIP-seq library, on average 86,616,150 and 90,985,883 valid reads were obtained in two groups of myocardial samples, while 81,704,807 and 90,279,353 valid reads were obtained in the RNA-seq library (**Supplementary Table S3**). In myocardial IP samples, the average mapping ratios of valid reads in the SO and the MI groups were 90.79 and 91.50%, respectively. The average mapping ratios of valid reads were 92.53 and 91.78% in the input samples (**Supplementary Table S4**). The valid data that mapped to the reference genome can be defined as the alignment to exon, intron, and intergenic according to the regional information. The average ratios of IP and input samples to exons were 67.91 and 55.20% in the SO group while 68.34 and 52.29% in the MI group, respectively (**Supplementary Figure S1**).

3.2 Mettl3 Induced Upregulation of m^{6}A Methylation Level in Myocardial Tissue, H9c2 Cells, and HUVECs

Upregulation of Mettl3 was first observed *in vivo* (rat MI model) and *in vitro* (hypoxic H9c2 cells and HUVECs), followed by

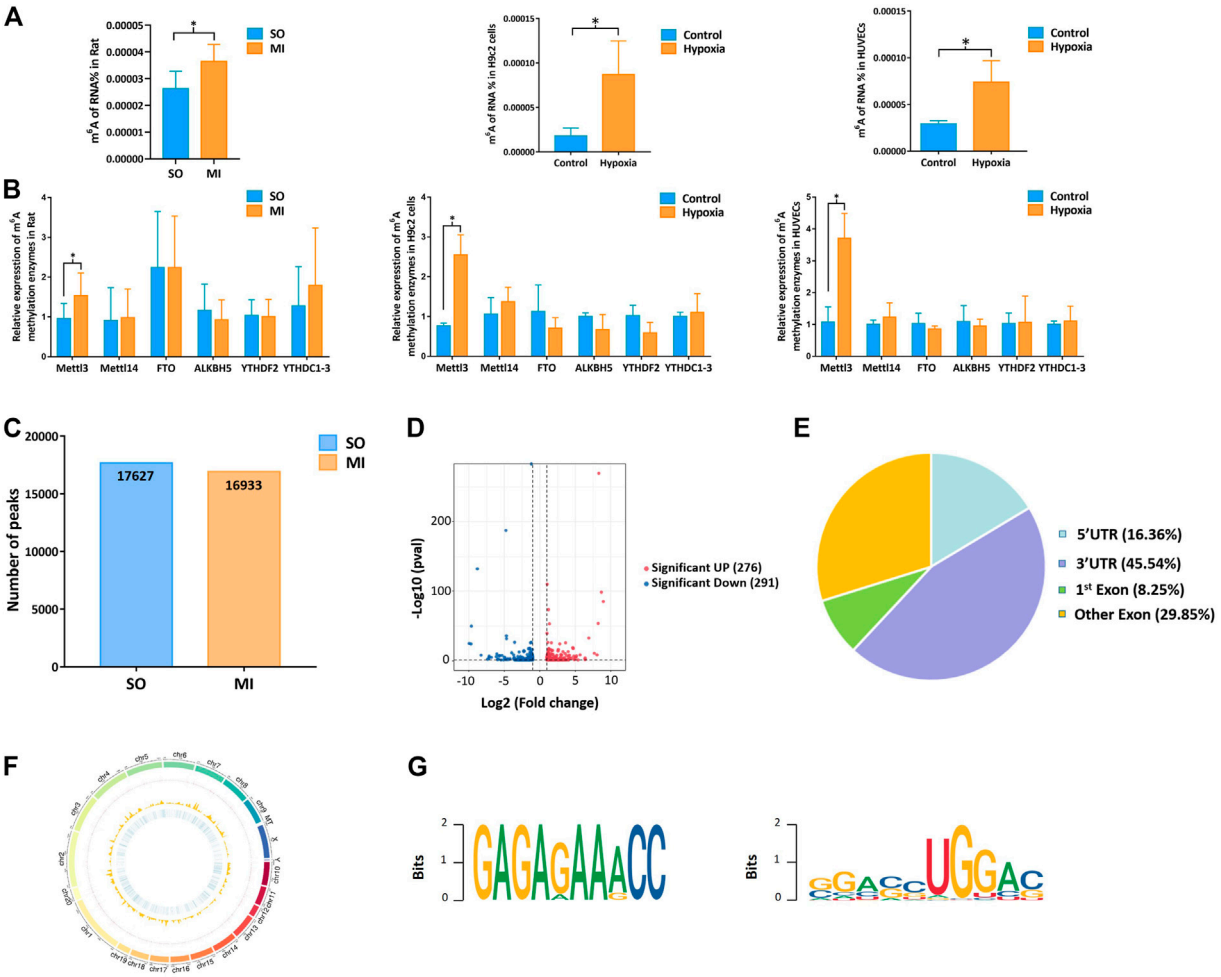


FIGURE 2 | Mettl3 induced upregulation of m^{6}A methylation level in myocardial tissue, H9c2 cells, and HUVECs. **(A)** m^{6}A methylation level of MI rat and hypoxic cells; **(B)** expression of m^{6}A methylation enzymes in myocardial tissue samples and cells by qRT-PCR; **(C)** number of peaks in the SO and MI group; **(D)** volcano plots for differentially methylated m^{6}A peaks; **(E)** distribution of differentially methylated m^{6}A peaks in the SO and MI group; **(F)** distribution patterns of differentially m^{6}A methylated peaks on chromosomes; **(G)** sequence motif for the m^{6}A peak regions. UTR, untranslated region; Exon, expressed region; SO, sham operated; MI, myocardial infarction. *$p < 0.05$.

upregulation of m^{6}A methylation level (**Figures 2A,B**). Through the analysis of MeRIP-seq, we identified 17,627 distinct m^{6}A peaks in 9889 mRNAs in the SO samples and 16,933 distinct m^{6}A peaks in 9764 mRNAs in the MI samples (**Figure 2C**). There were 567 significantly variated peaks identified, in which 276 peaks were upregulated and 291 were downregulated (**Figure 2D**). The top 20 differentially m^{6}A methylated peaks are shown in **Table 1**. We analyzed the distribution patterns of differentially m^{6}A methylated peaks. A total of 45.54% of the m^{6}A methylated peaks harbored in the 3′UTR and 16.36% enriched in the 5′UTR (**Figure 2E**). All differentially m^{6}A methylated peaks within mRNAs were mapped to chromosomes (**Figure 2F**). The top five chromosomes harboring the most m^{6}A peaks were listed as follows: chr1 (646), chr10 (436), chr11 (130), chr12 (176), and chr13 (170). Finally, we conducted motif prediction for samples and demonstrated the predicted motif in **Figure 2G**.

3.3 Hub mRNAs Associated with Angiogenesis and Apoptosis

Based on the results of RNA-seq, 576 mRNAs were significantly upregulated, and 78 mRNAs were downregulated. The top 20 differentially expressed mRNAs are listed in **Table 2**. Volcano plots and heatmap plots of differentially expressed mRNAs are shown in **Figures 3A,B**. Combined analysis of m^{6}A methylation and mRNAs expression levels was performed according to the following thresholds: $|\log2FC|>1$, $p < 0.05$ for m^{6}A methylation and $|\log2FC|>0.5$, $p < 0.05$ for mRNAs expression. As a result, we obtained 377 mRNAs where their m^{6}A peaks and mRNA expression both changed significantly. The relationships between m^{6}A methylation and mRNAs expression are shown in the four-quadrant graph and Venn diagram (**Figures 3C,D**). Accordingly, there were 124 significantly hypomethylated and upregulated mRNAs, 26 significantly hypermethylated and

TABLE 1 | Top 20 differentially methylated mRNAs

mRNA	Chromosome	Peak region	Peak start	Peak end	*p*-value	Log2FC	Up/Down
Bach1	chr11	3′ UTR	27,397,948	27,398,514	6.30E-86	8.96	Up
Rrn3	chr10	3′ UTR	3,213,487	3,213,996	9.99E-270	8.32	Up
Senp5	chr11	Exon	72,064,886	72,067,150	2.45E-09	8.06	Up
LOC100910130	X	Exon	156,395,518	156,397,956	1.00E-11	7.68	Up
Hnrnph3	chr20	Exon	27,176,410	27,177,224	1.26E-33	6.91	Up
Gspt1	chr10	Exon	4,412,071	4,426,636	8.50E-03	6.39	Up
Bdkrb1	chr6	3′ UTR	129,439,859	129,440,455	1.20E-04	6.38	Up
Scart1	chr1	Exon	212,701,635	212,704,224	7.20E-02	5.70	Up
Natd1	chr10	5′ UTR	45,065,415	45,071,643	2.90E-02	5.39	Up
Gata6	chr18	Exon	2,417,213	2,417,483	6.46E-08	5.18	Up
Tap2	chr20	5′ UTR	3,995,544	3,996,581	7.94E-26	–9.94	Down
Nfat5	chr19	5′ UTR	38,533,016	38,533,316	3.98E-25	–9.69	Down
Sp3	chr3	Exon	59,646,075	59,646,253	1.25E-50	–9.63	Down
Cd248	chr1	Exon	220,353,416	220,354,637	1.26E-50	–8.79	Down
AABR07048992	chr5	5′ UTR	99,032,868	99,033,107	1.00E-132	–8.25	Down
Nfs1	chr3	5′ UTR	151,687,398	151,688,149	6.76E-09	–7.34	Down
Rpusd2	chr3	Exon	110,836,072	110,836,252	2.18E-04	–7.17	Down
Ranbp2	chr20	Exon	28,047,206	28,047,481	9.77E-06	–6.81	Down
Capns1	chr1	Exon	91,063,753	91,063,928	6.30E-11	–6.04	Down
Raver2	chr5	Exon	119,938,132	119,938,341	1.50E-02	–5.95	Down

FC, fold change.

TABLE 2 | Top 20 differentially expressed mRNAs.

mRNA	Chromosome	Gene start	Gene end	*p*-value	Log2FC	Up/Down
Gm47305	chr6	91,680,317	91,680,386	8.33E-67	12.06	Up
Ccn4	chr7	107,695,215	107,723,772	1.10E-23	8.30	Up
Mmp12	chr8	5,606,592	5,616,493	1.27E-15	8.28	Up
Col11a1	chr2	216,863,428	217,056,523	1.12E-57	7.95	Up
Grem1	chr3	105,203,309	105,214,989	1.22E-16	7.65	Up
Comp	chr16	20,798,437	20,807,070	1.57E-68	7.59	Up
Card14	chr10	108,440,950	108,468,310	3.02E-19	7.48	Up
Angptl7	chr5	165,312,130	165,316,652	2.84E-49	7.37	Up
Fam180a	chr4	62,844,971	62,860,446	4.73E-33	7.36	Up
Ccn5	chr3	160,207,913	160,219,331	1.24E-134	7.18	Up
Ptprh	chr1	72,810,545	72,859,161	1.54E-05	–6.83	Down
Fam111a	chr1	229,003,961	229,019,527	3.80E-07	–4.96	Down
Gp6	chr1	73,040,901	73,064,641	2.80E-03	–4.22	Down
7SK	chr17	90,799,609	90,799,914	2.00E-06	–3.90	Down
AABR07069186	chr8	10,985,553	10,988,077	4.10E-04	–3.47	Down
Gzmc	chr15	35,392,162	35,394,792	3.81E-04	–3.41	Down
Sec14l5	chr10	10,591,251	10,629,735	2.36E-13	–3.34	Down
Pdzk1	chr2	198,965,685	198,999,323	5.50E-03	–3.31	Down
Klk12	chr1	99,706,780	99,711,004	3.10E-02	–3.25	Down
Alas2	X	23,167,696	23,187,341	2.90E-04	–3.18	Down

FC, fold change.

downregulated mRNAs, 173 significantly hypermethylated and upregulated, and 54 significantly hypomethylated and downregulated mRNAs. The top 10 hypermethylated and downregulated mRNAs are list in **Table 3**. Meanwhile, we provided the detailed information of the top 10 hypermethylated and upregulated mRNAs, the top 10 hypomethylated and downregulated mRNAs, and the top 10 hypomethylated and upregulated mRNAs in **Supplementary Tables S5–S7**. Expression of hub mRNAs, including Hadh, Arfgef3, Sez6, Psmg3, Kcnn1, Tet1, Myo1b, Ptprz1, Ank2, and Pwwp3b were compared between the SO and the MI groups by qRT-PCR testing. Among them, the expression of Hadh, Arfgef3, Sez6, Kcnn1, and Tet1 were significantly downregulated in the MI group (**Figure 3E**). After interaction of the above three hub mRNAs with apoptosis and angiogenesis-related genes in GeneCards, Hadh, Tet1, and Kcnn1 were finally identified to be significantly associated with apoptosis, and Tet1 and Kcnn1 were significantly correlated to angiogenesis (**Figure 3F**).

3.4 m^{6}A Methylation Influenced by Hypoxia in a Time- and Dose-Dependent Pattern

Upon our previous observation, m^{6}A methylation level was upregulated by Mettl3 *in vivo* and *in vitro* (**Figures 2A,B**). We

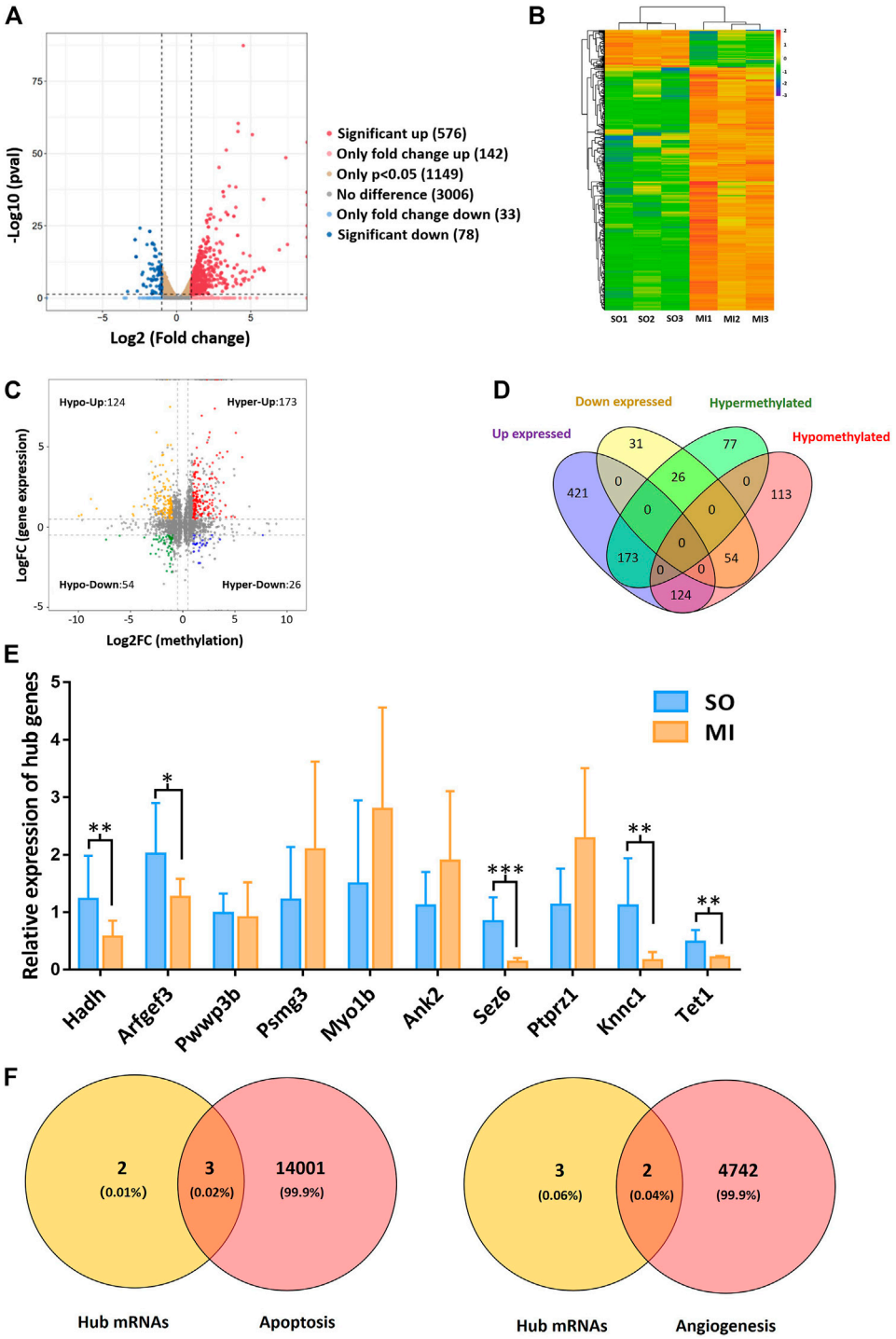


FIGURE 3 | Hub mRNAs associated with angiogenesis and apoptosis. **(A)** Volcano plots for differentially expressed mRNAs; **(B)** heatmap plots for differentially expressed mRNAs; **(C)** four-quadrant graph for m^{6}A methylation regulated mRNAs; **(D)** Venn diagram showing the relationships between m^{6}A methylation and mRNAs expression; **(E)** validation of the top 10 hypermethylated and downregulated mRNAs in MI rats; **(F)** Venn diagram showing the relationship between hub mRNAs and apoptosis/angiogenesis. Hypo-Up, hypomethylated and upregulated; hyper-Up, hypermethylated and upregulated; Hypo-Down, hypomethylated and downregulated; hyper-Down, hypermethylated and downregulated. SO, sham operated; MI, myocardial infarction. $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

afterward detected upregulated m^{6}A methylation level of hub mRNAs with MeRIP-qRT-PCR, which further confirmed our results from transcriptome-wide MeRIP-seq analysis (**Figures 4A,B**). With the inhibition of Mettl3 in H9c2 cells and HUVECs, downregulated m^{6}A methylation level of total RNA and upregulated expression of hub mRNAs were detected (**Figures**

TABLE 3 | Top 10 hypermethylated and downregulated mRNAs.

mRNA	Chromosome	m^{6}A regulation	Regulation	FPKM of MI input	FPKM of SO input
Hadh	chr2	Up	Down	87.12	249.88
Arfgef3	chr1	Up	Down	0.40	1.10
Pwwp3b	X	Up	Down	0.53	1.11
Sez6	chr10	Up	Down	0.35	0.71
Psmg3	chr12	Up	Down	1.28	2.62
Kcnn1	chr16	Up	Down	0.85	1.70
Tet1	X	Up	Down	0.29	0.58
Myo1b	chr9	Up	Down	12.53	21.41
Ptprz1	chr4	Up	Down	0.25	0.41
Ank2	chr2	Up	Down	11.74	18.99

FPKM, fragments per kilobase of exon model per million mapped fragments; MI, myocardial infarction; SO, sham operated.

4C–F). After treating with different oxygen concentration (1% O_2, 2% O_2, and 5% O_2) for 24 h, m^{6}A methylation level in H9c2 cells and HUVECs progressively decreased with the increase of oxygen concentration (**Figure 4G**). Afterward, these cells were treated with different hypoxic duration (12, 24, and 48 h) at 1% of oxygen. m^{6}A methylation level in H9c2 cells and HUVECs gradually increased with prolonged hypoxic duration (**Figure 4H**). Variation of m^{6}A methylation in the simulated hypoxic context demonstrated a time- and dose-dependent pattern.

3.5 m^{6}A Methylation Regulated Hub mRNAs Play a Role on Apoptosis *in vitro*

As presented in **Figures 5A,B**, Hadh, Tet1, and Kcnn1 were significantly downregulated after hypoxia. The apoptosis rate of cardiomyocytes exposed to hypoxia increased significantly (**Figures 5C,D**). As compared to the control group, the expressions of cleaved-caspase3 and Bax increased significantly, while the expressions of bcl-2 decreased in the hypoxia group based on Western blot (**Figure 5E**).

3.6 m^{6}A Methylation Regulated Hub mRNAs Play a Role on Angiogenesis *in vitro*

Kcnn1 and Tet1 determined by qRT-PCR and Western blot were significantly decreased in the hypoxia group (**Figures 6A,B**). The effects of hypoxia on cell proliferation and angiogenesis were revealed by CCK-8, EdU, and tube formation assay. Cell viability and EdU-positive cells were significantly decreased due to hypoxia (**Figures 6C,D**). All indicators of tube formation ability, including the total branching points, total tube length, and total loops, were markedly decreased in the hypoxia group (**Figure 6E**).

4 DISCUSSION

DNA methylation and histone modification have been extensively investigated in genetic and cellular biology (Panneerdoss et al., 2018). However, roles of reversible RNA methylation in the cardiovascular field have been under development in recent years (Mongelli et al., 2020). In the hypoxia/reoxygenation-treated cardiomyocytes, low-expression of METTL3 and over-expression of ALKBH5 were observed to inhibit autophagy and apoptosis (Song et al., 2019). In addition, over-expression of ALKBH5 was verified to reduce infarct size, restore cardiac function, and facilitate cardiomyocyte proliferation after MI (Han et al., 2021). These promising results preliminarily revealed the roles of m^{6}A methylation on mRNAs in ischemic heart diseases. However, information regarding systematic and comprehensive analysis is limited. In the current study, we observed the overall m^{6}A methylation level was significantly upregulated with the upregulation of Mettl3 in MI rat and hypoxic cells. We demonstrated the landscape changes of m^{6}A methylation on mRNA regulated apoptosis and angiogenesis in MI. Meanwhile, we explored the impact of oxygen concentration and hypoxic duration on m^{6}A methylation. Upon which, several hypermethylated and downregulated mRNAs were successfully identified after MI by combined analysis of MeRIP-seq and RNA-seq. It further provided potential targets for treatment and scientific purposes. However, how and what performance of these m^{6}A methylated mRNAs may have on MI need to be further elaborated.

Followed by the newly proposed mechanisms of upstream methylation on mRNA in this research field, the final destination is to compensate or improve the downstream functional impairment launched by MI. According to the literature review, apoptosis plays a role in the process of tissue damage after MI, which appears to have pathological and therapeutic implications (Krijnen et al., 2002). Therefore, the upcoming question in front of us was whether the m^{6}A methylated mRNAs would play a role in the regulation of cardiomyocyte apoptosis. In this regard, we interacted our newly identified m^{6}A methylated mRNAs and apoptosis related genes in the GeneCards and three of them were correlated to apoptosis (e.g., Tet1, Hadh, and Kcnn1). Over-expression of Tet1 increased cell apoptosis and inhibited cell growth in osteosarcoma cells (Teng et al., 2019). Over-expression of Hadh linked to cell apoptosis in acute myeloid leukemia (Wei L. et al., 2020). However, the relationship between Kcnn1 and apoptosis has not been validated according to previous studies. We afterward validated the m^{6}A methylation regulated mRNAs on apoptosis in hypoxic H9c2 cells and found the m^{6}A methylation level was upregulated in hypoxic H9c2 cells. Our results of MeRIP-qRT-PCR demonstrated that m^{6}A methylation of three hub mRNAs was significantly upregulated, while their expression levels were downregulated in

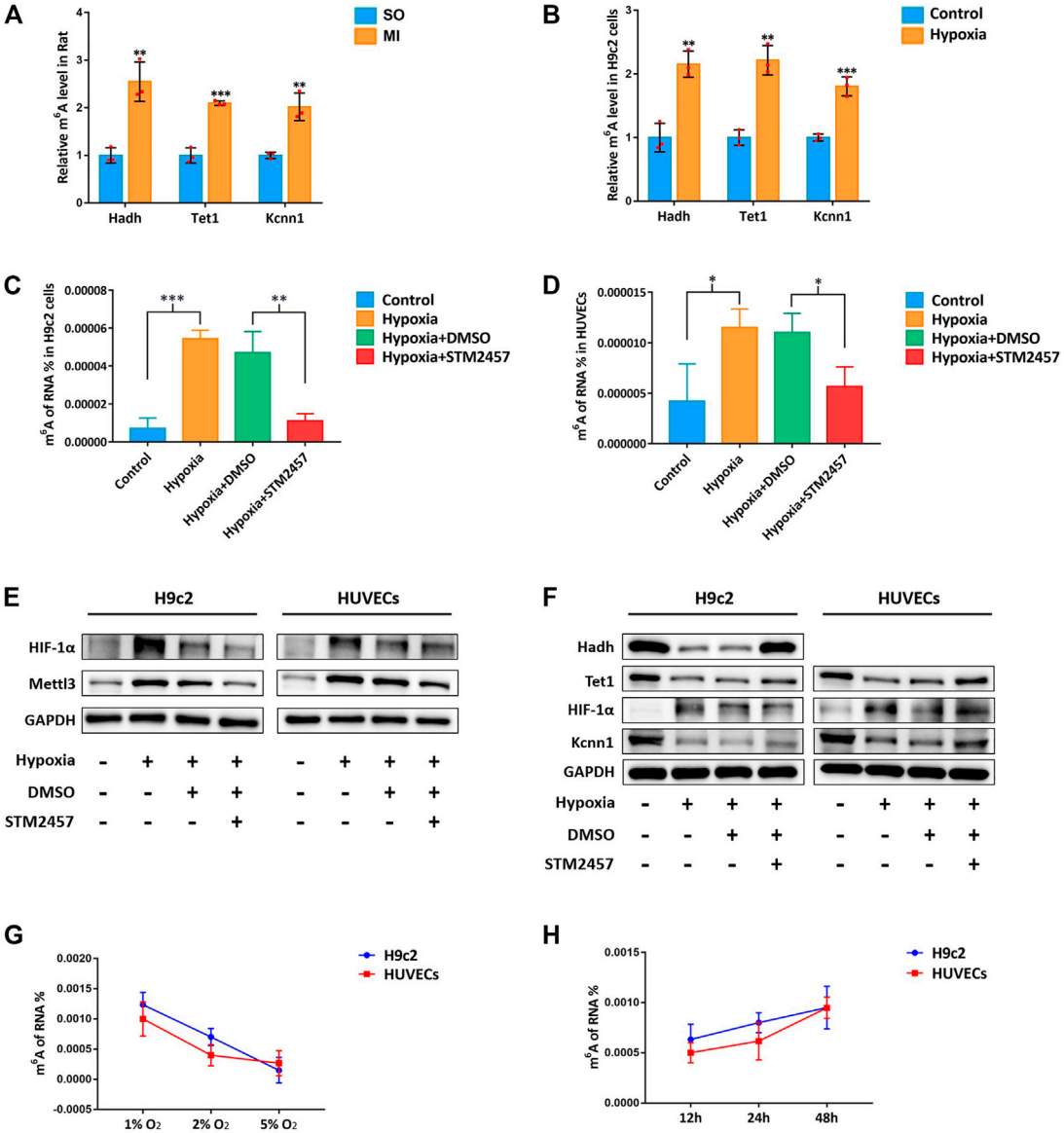


FIGURE 4 | m^{6}A methylation influenced by hypoxia in a time- and dose-dependent pattern. **(A)** MeRIP-qRT-PCR validation of hub mRNAs in MI rat; **(B)** MeRIP-qRT-PCR validation of hub mRNAs in H9c2 cells; **(C)** m^{6}A methylation levels of H9c2 cells with Mettl3 inhibitor; **(D)** m^{6}A methylation levels of HUVECs with Mettl3 inhibitor; **(E)** Mettl3 protein level of H9c2 cells and HUVECs with Mettl3 inhibitor. HIF-1α was used as positive control. GAPDH was used as loading control; **(F)** protein level of hub mRNAs in H9c2 cells and HUVECs with Mettl3 inhibitor; **(G)** m^{6}A methylation levels of H9c2 cells and HUVECs with different concentrations of oxygen; **(H)** m^{6}A methylation levels of H9c2 cells and HUVECs with different hypoxic duration. SO, sham operated; MI, myocardial infarction; DMSO, dimethyl sulfoxide. *p < 0.05, **p < 0.01, ***p < 0.001. *STM2457 is the inhibitor of Mettl3.

MI rat and hypoxic cells. In addition, inhibition of Mettl3 downregulated the overall m^{6}A level and upregulated the expression of Hadh, Tet1, and Kcnn1 in hypoxic H9c2 cells. Apoptosis was accompanied with variations of hub proteins, for instance, decrease of bcl-2 and increase of Bax, Caspase3, and cleaved-caspase3. Taken together, these three hub mRNAs regulated by m^{6}A methylation were preliminarily verified to impact the process of cardiomyocyte apoptosis after MI.

As the exacerbation of myocardial injury, it is likely leading to heart failure and sudden death (Zouggari et al., 2013). Early revascularization in the infarcted region is essential for promoting the survival of myocardial cells, reducing infarct size, and improving the prognosis of patients with MI (Keeley et al., 2003; Guo et al., 2017). Upon this perspective, angiogenesis has been reported to be an integral and indispensable part of the myocardial healing process following ischemic events (Cochain et al., 2013). The inhibition of angiogenesis accelerated heart failure in a murine model and various treatments were proposed to ameliorate infarction size, left ventricular remodeling, and cardiac function after MI via promoting angiogenesis in animal

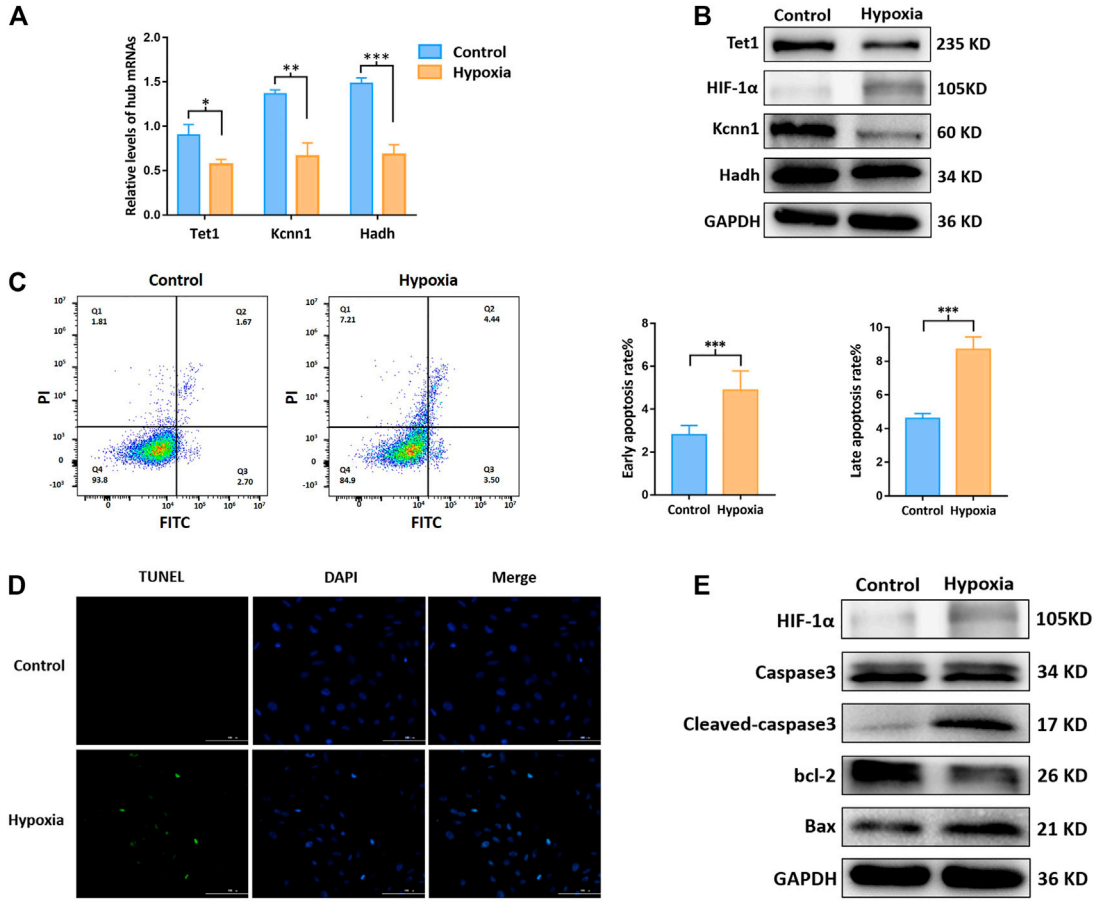


FIGURE 5 | m^{6}A methylation regulated hub mRNAs play a role on apoptosis *in vitro*. **(A)** Expression of apoptosis related mRNAs by qRT-PCR; **(B)** protein level of Tet1, Kcnn1, and Hadh by Western blot; **(C)** apoptosis rate by flow cytometry; **(D)** apoptotic status by Tunel staining; **(E)** protein level (Caspase3, cleaved-caspase3, bcl-2, Bax, and HIF-1α) by Western blot. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

models of MI (Shiojima et al., 2005; Zeng et al., 2010; Shindo et al., 2016; Gou et al., 2020). In the past several decades, attentions have been shifted to investigate the mechanical associations between angiogenesis and epigenetics. mRNAs are responsible for protein translation and can directly regulate the synthesis of proteins such as vascular endothelial growth factor in the context of MI. However, the upstream regulation of mRNAs has not been well clarified. In the current study, two m^{6}A methylation regulated mRNAs (e.g., Kcnn1 and Tet1) were detected by interaction analysis of five hub mRNAs and angiogenesis-related genes. Kcnn1 is one of the members in the calcium activated potassium channel subfamily. It was previously reported to be associated with the development of atrial fibrillation while its role in angiogenesis in the context of MI has not been reported (Park et al., 2014; Rahm et al., 2021). On the other hand, Tet1, as one of translocation enzymes mediating 5-methylcytosine (5mC) hydroxylation, was reported to participate in the facilitation of DNA demethylation (Gomes et al., 2020). In addition, inhibition of Tet1 expression may contribute to tumor growth and angiogenesis *in vivo* (Si et al., 2019). Upon this condition, we explored the impact of Kcnn1 and Tet1 on angiogenesis in hypoxia-induced HUVECs to further verify their roles after MI. Not surprisingly, in the context of hypermethylated and downregulated Kcnn1 and Tet1, CCK-8, EdU, and tube formation assay potentially indicated that hypoxia inhibits endothelial cell proliferation and angiogenesis. Meanwhile, we found the overall level of m^{6}A methylation of total RNA was upregulated in hypoxic endothelial cells. Inhibition of Mettl3 downregulated the overall m^{6}A level and upregulated the expression of Tet1 and Kcnn1 in hypoxic HUVECs. These results suggested that m^{6}A methylated Kcnn1 and Tet1 may play an essential role in angiogenesis. Further studies to be conducted in knockdown and over-expression models would be helpful to validate the role of m^{6}A methylation on mRNAs in the process of angiogenesis.

Similar to other MeRIP-seq and RNA-seq studies, the reproducibility of MeRIP-seq was considered poor. Fortunately, heart tissues of our animal model were consistently collected from similar regions in a standard environment, which guaranteed the homogeneity at the histological level to some extent. In addition, we performed a rescue experiment to verify the reproducibility of our sequencing results. However, it needs to be further verified with well-designed *in vitro* or *in vivo* studies. Finally, our results supplemented information of m^{6}A methylated mRNAs

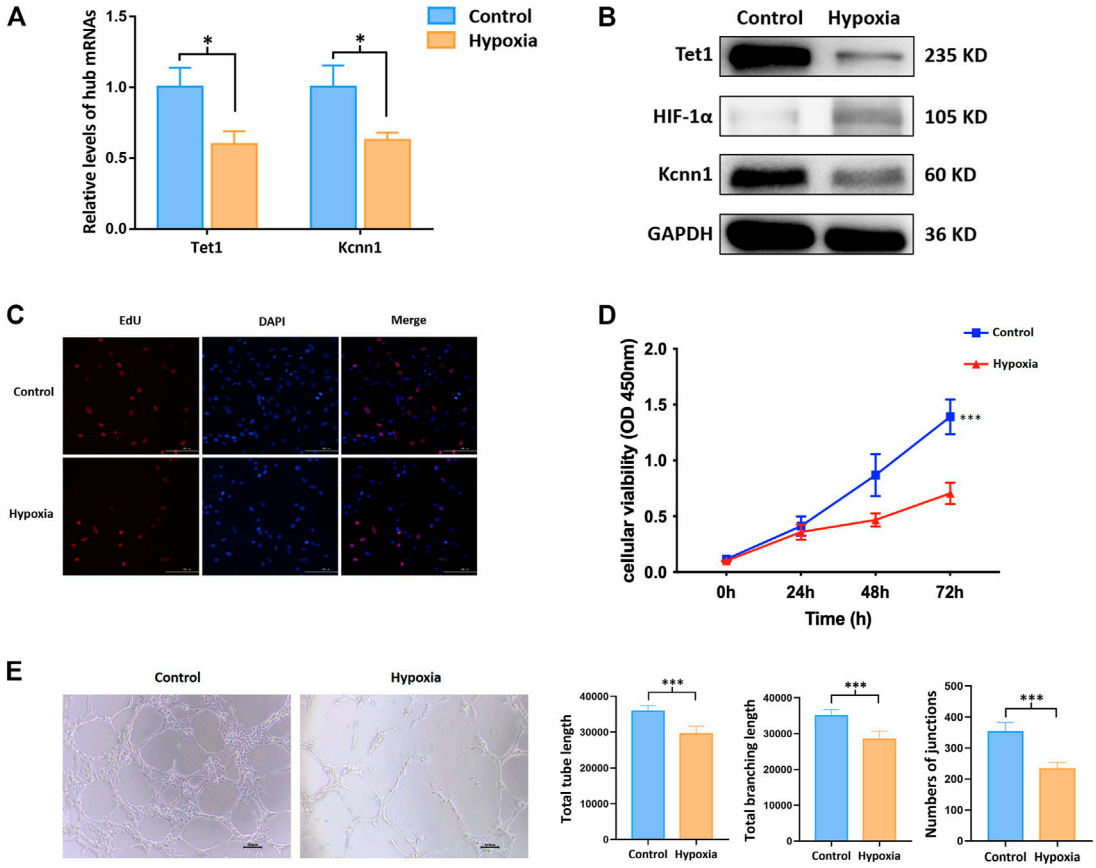


FIGURE 6 | m^{6}A methylation regulated hub mRNAs play a role on angiogenesis *in vitro*. **(A)** Expression of angiogenesis related mRNAs by qRT-PCR; **(B)** protein level of Tet1, Kcnn1, and HIF-1α by Western blot; **(C)** cell proliferation by EdU assay; **(D)** cell viability by CCK-8 assay; **(E)** tube formation. *$p < 0.05$; ***$p < 0.001$.

in MI, which was conducive to the broadening of research ideas in this direction and provided reference for further studies.

5 CONCLUSION

The overview of upstream epigenetic changes after MI were demonstrated with emphasizing the essential role of m^{6}A methylated mRNAs after MI. Five m^{6}A methylation regulated mRNAs were newly identified by interaction analysis of MeRIP-seq and RNA-seq. In the context of m^{6}A methylation, three of hub mRNAs were validated to impact the process of cardiomyocyte apoptosis and angiogenesis. Our study paved the way for future investigations aiming at exploring the upstream epigenetic mechanisms in the pathogenesis of MI.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the GEO repository, accession number GSE189593.

ETHICS STATEMENT

The animal study was reviewed and approved by Institutional Animal Care and Use Committee of Nanjing Medical University.

AUTHOR CONTRIBUTIONS

X.L., Y.Z., and Y.D. conceived and designed the current study. Y.Z., W.H., Y.C., J.W., X.Z., M.T., and S.W. established MI rat model and collected RNA samples from myocardial tissues. Y.C. and Y.Z. performed data processing, normalization, and bioinformatic analyses. Y.Z. and W.H. completed experimental verification. Y.Z., W.H., and Y.C. drafted the manuscript. Y.Z. and Y.D. contributed to the modification of the manuscript. All authors have read and approved the manuscript.

FUNDING

This study was funded by the National Natural Science Foundation of China (Grant number: 81772441, 81902288, 82072546, and 82000529). The funders had no role in the

study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

The authors acknowledge the Jiangsu province hospital core facility center for providing experimental instruments, research assistants here for their valuable help, the medical staff from the Department of Rehabilitation Medicine in the First Affiliated Hospital of Nanjing Medical University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.789923/full#supplementary-material

REFERENCES

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: Tools for Motif Discovery and Searching. *Nucleic Acids Res.* 37, W202–W208. doi:10.1093/nar/gkp335

Chen, B., and Ling, C. H. (2019). Long Noncoding RNA AK027294 Acts as an Oncogene in Non-small Cell Lung Cancer by Up-Regulating STAT3. *Eur. Rev. Med. Pharmacol. Sci.* 23, 1102–1107. doi:10.26355/eurrev_201902_17000

Chen, J., Chen, T., Zhu, Y., Li, Y., Zhang, Y., Wang, Y., et al. (2019). circPTN Sponges miR-145-5p/miR-330-5p to Promote Proliferation and Stemness in Glioma. *J. Exp. Clin. Cancer Res.* 38, 398. doi:10.1186/s13046-019-1376-8

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: an Ultra-fast All-In-One FASTQ Preprocessor. *Bioinformatics* 34, i884–i890. doi:10.1093/bioinformatics/bty560

Chen, Z. L., Chen, Y. X., Zhou, J., Li, Y., Gong, C. Y., and Wang, X. B. (2020). LncRNA HULC Alleviates HUVEC Inflammation and Improves Angiogenesis after Myocardial Infarction through Down-Regulating miR-29b. *Eur. Rev. Med. Pharmacol. Sci.* 24, 6288–6298. doi:10.26355/eurrev_202006_21527

Cheng, H.-W., Chen, Y.-F., Wong, J.-M., Weng, C.-W., Chen, H.-Y., Yu, S.-L., et al. (2017). Cancer Cells Increase Endothelial Cell Tube Formation and Survival by Activating the PI3K/Akt Signalling Pathway. *J. Exp. Clin. Cancer Res.* 36, 27. doi:10.1186/s13046-017-0495-3

Cochain, C., Channon, K. M., and Silvestre, J.-S. (2013). Angiogenesis in the Infarcted Myocardium. *Antioxid. Redox Signaling* 18, 1100–1113. doi:10.1089/ars.2012.4849

Dang, Y., Xu, J., Yang, Y., Li, C., Zhang, Q., Zhou, W., et al. (2020). Ling-gui-zhu-gan Decoction Alleviates Hepatic Steatosis through SOCS2 Modification by N6-Methyladenosine. *Biomed. Pharmacother.* 127, 109976. doi:10.1016/j.biopha.2020.109976

Deng, K., Ning, X., Ren, X., Yang, B., Li, J., Cao, J., et al. (2021). Transcriptome-wide N6-Methyladenosine Methylation Landscape of Coronary Artery Disease. *Epigenomics* 13, 793–808. doi:10.2217/epi-2020-0372

Dominissini, D., Moshitch-Moshkovitz, S., Salmon-Divon, M., Amariglio, N., and Rechavi, G. (2013). Transcriptome-wide Mapping of N6-Methyladenosine by m6A-Seq Based on Immunocapturing and Massively Parallel Sequencing. *Nat. Protoc.* 8, 176–189. doi:10.1038/nprot.2012.148

Gomes, A. M., Pinto, T. S., Costa Fernandes, C. J., Silva, R. A., and Zambuzzi, W. F. (2020). Wortmannin Targeting Phosphatidylinositol 3-kinase Suppresses Angiogenic Factors in Shear-stressed Endothelial Cells. *J. Cell Physiol* 235, 5256–5269. doi:10.1002/jcp.29412

Gou, L., Xue, C., Tang, X., and Fang, Z. (2020). Inhibition of Exo-miR-19a-3p Derived from Cardiomyocytes Promotes Angiogenesis and Improves Heart Function in Mice with Myocardial Infarction via Targeting HIF-1α. *Aging* 12, 23609–23618. doi:10.18632/aging.103563

Gu, M., Wang, J., Wang, Y., Xu, Y., Zhang, Y., Wu, W., et al. (2018). MiR-147b Inhibits Cell Viability and Promotes Apoptosis of Rat H9c2 Cardiomyocytes via Down-Regulating KLF13 Expression. *Acta Biochim. Biophys. Sin (Shanghai).* 50, 288–297. doi:10.1093/abbs/gmx144

Guo, Y., Luo, F., Liu, Q., and Xu, D. (2017). Regulatory Non-coding RNAs in Acute Myocardial Infarction. *J. Cell. Mol. Med.* 21, 1013–1023. doi:10.1111/jcmm.13032

Han, Z., Wang, X., Xu, Z., Cao, Y., Gong, R., Yu, Y., et al. (2021). ALKBH5 Regulates Cardiomyocyte Proliferation and Heart Regeneration by Demethylating the mRNA of YTHDF1. *Theranostics* 11, 3000–3016. doi:10.7150/thno.47354

Hanlon, M. M., Rakovich, T., Cunningham, C. C., Ansboro, S., Veale, D. J., Fearon, U., et al. (2019). STAT3 Mediates the Differential Effects of Oncostatin M and TNFα on RA Synovial Fibroblast and Endothelial Cell Function. *Front. Immunol.* 10, 2056. doi:10.3389/fimmu.2019.02056

Jozefczuk, J., and Adjaye, J. (2011). Quantitative Real-Time PCR-Based Analysis of Gene Expression. *Methods Enzymol.* 500, 99–109. doi:10.1016/b978-0-12-385118-5.00006-2

Keeley, E. C., Boura, J. A., and Grines, C. L. (2003). Primary Angioplasty versus Intravenous Thrombolytic Therapy for Acute Myocardial Infarction: a Quantitative Review of 23 Randomised Trials. *The Lancet* 361, 13–20. doi:10.1016/s0140-6736(03)12113-7

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a Fast Spliced Aligner with Low Memory Requirements. *Nat. Methods* 12, 357–360. doi:10.1038/nmeth.3317

Krijnen, P. A., Nijmeijer, R., Meijer, C. J., Visser, C. A., Hack, C. E., and Niessen, H. W. (2002). Apoptosis in Myocardial Ischaemia and Infarction. *J. Clin. Pathol.* 55, 801–811. doi:10.1136/jcp.55.11.801

Li, M., Ye, J., Zhao, G., Hong, G., Hu, X., Cao, K., et al. (2019a). Gas6 Attenuates Lipopolysaccharide-induced TNF-α E-xpression and A-poptosis in H9C2 C-ells through NF-κB and MAPK I-nhibition via the Axl/PI3K/Akt P-athway. *Int. J. Mol. Med.* 44, 982–994. doi:10.3892/ijmm.2019.4275

Li, T., Hu, P.-S., Zuo, Z., Lin, J.-F., Li, X., Wu, Q.-N., et al. (2019b). METTL3 Facilitates Tumor Progression via an m6A-igf2bp2-dependent Mechanism in Colorectal Carcinoma. *Mol. Cancer* 18, 112. doi:10.1186/s12943-019-1038-7

Li, T., Zhuang, Y., Yang, W., Xie, Y., Shang, W., Su, S., et al. (2021). Silencing of METTL3 Attenuates Cardiac Fibrosis Induced by Myocardial Infarction via Inhibiting the Activation of Cardiac Fibroblasts. *FASEB j.* 35, e21162. doi:10.1096/fj.201903169R

Li, X., Wang, X., Liu, Y.-s., Wang, X.-d., Zhou, J., and Zhou, H. (2020a). Downregulation of miR-3568 Protects against Ischemia/Reperfusion-Induced Cardiac Dysfunction in Rats and Apoptosis in H9C2 Cardiomyocytes through Targeting TRIM62. *Front. Pharmacol.* 11, 17. doi:10.3389/fphar.2020.00017

Li, Y., Zhu, H., Wei, X., Li, H., Yu, Z., Zhang, H., et al. (2017). LPS Induces HUVEC Angiogenesis *In Vitro* through miR-146a-Mediated TGF-B1 Inhibition. *Am. J. Transl Res.* 9, 591–600.

Li, Y., Ren, S., Xia, J., Wei, Y., and Xi, Y. (2020b). EIF4A3-Induced Circ-BNIP3 Aggravated Hypoxia-Induced Injury of H9c2 Cells by Targeting miR-27a-3p/BNIP3. *Mol. Ther. - Nucleic Acids* 19, 533–545. doi:10.1016/j.omtn.2019.11.017

Liang, Y., Zhang, X., Zou, J., Shi, Y., Wang, Y., Tai, J., et al. (2019). Pharmacology Mechanism of Flos Magnoliae and Centipeda Minima for Treating Allergic Rhinitis Based on Pharmacology Network. *Drug Dev. Ind. Pharm.* 45, 1547–1555. doi:10.1080/03639045.2019.1635150

Liu, T., Yang, S., Sui, J., Xu, S. Y., Cheng, Y. P., Shen, B., et al. (2020). Dysregulated N6-methyladenosine Methylation Writer METTL3 Contributes to the Proliferation and Migration of Gastric Cancer. *J. Cel Physiol* 235, 548–562. doi:10.1002/jcp.28994

Liu, Y., Chen, X., Cheng, R., Yang, F., Yu, M., Wang, C., et al. (2018). The Jun/miR-22/HuR Regulatory axis Contributes to Tumourigenesis in Colorectal Cancer. *Mol. Cancer* 17, 11. doi:10.1186/s12943-017-0751-3

Lu, L., Liu, M., Sun, R., Zheng, Y., and Zhang, P. (2015). Myocardial Infarction: Symptoms and Treatments. *Cell Biochem Biophys* 72, 865–867. doi:10.1007/s12013-015-0553-4

Luo, G., Jian, Z., Zhu, Y., Zhu, Y., Chen, B., Ma, R., et al. (2019). Sirt1 Promotes Autophagy and Inhibits Apoptosis to Protect Cardiomyocytes from Hypoxic Stress. *Int. J. Mol. Med.* 43, 2033–2043. doi:10.3892/ijmm.2019.4125

Meng, J., Lu, Z., Liu, H., Zhang, L., Zhang, S., Chen, Y., et al. (2014). A Protocol for RNA Methylation Differential Analysis with MeRIP-Seq Data and exome Peak R/Bioconductor Package. *Methods* 69, 274–281. doi:10.1016/j.ymeth.2014.06.008

Meyer, K. D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C. E., and Jaffrey, S. R. (2012). Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3′ UTRs and Near Stop Codons. *Cell* 149, 1635–1646. doi:10.1016/j.cell.2012.05.003

Mongelli, A., Atlante, S., Bachetti, T., Martelli, F., Farsetti, A., and Gaetano, C. (2020). Epigenetic Signaling and RNA Regulation in Cardiovascular Diseases. *Ijms* 21, 509. doi:10.3390/ijms21020509

Ni, J., Liu, X., Yin, Y., Zhang, P., Xu, Y.-W., and Liu, Z. (2019). Exosomes Derived from TIMP2-Modified Human Umbilical Cord Mesenchymal Stem Cells Enhance the Repair Effect in Rat Model with Myocardial Infarction Possibly by the Akt/Sfrp2 Pathway. *Oxidative Med. Cell Longevity* 2019, 1–19. doi:10.1155/2019/1958941

Panneerdoss, S., Eedunuri, V. K., Yadav, P., Timilsina, S., Rajamanickam, S., Viswanadhapalli, S., et al. (2018). Cross-talk Among Writers, Readers, and Erasers of M 6 A Regulates Cancer Growth and Progression. *Sci. Adv.* 4, eaar8263. doi:10.1126/sciadv.aar8263

Park, J. H., Lee, J. S., Ko, Y.-G., Lee, S. H., Lee, B. S., Kang, S.-M., et al. (2014). Histological and Biochemical Comparisons between Right Atrium and Left Atrium in Patients with Mitral Valvular Atrial Fibrillation. *Korean Circ. J.* 44, 233–242. doi:10.4070/kcj.2014.44.4.233

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie Enables Improved Reconstruction of a Transcriptome from RNA-Seq Reads. *Nat. Biotechnol.* 33, 290–295. doi:10.1038/nbt.3122

Puymirat, E., Cayla, G., Cottin, Y., Elbaz, M., Henry, P., Gerbaud, E., et al. (2019). Twenty-year Trends in Profile, Management and Outcomes of Patients with ST-Segment Elevation Myocardial Infarction According to Use of Reperfusion Therapy: Data from the FAST-MI Program 1995-2015. *Am. Heart J.* 214, 97–106. doi:10.1016/j.ahj.2019.05.007

Qiu, L., Zhao, Q., Dai, L., Zhu, A., Xu, X., Zhao, S., et al. (2020). Long Non-coding RNA DANCR Alleviates Hypoxia-Caused H9c2 Cells Damage through up Regulation of HIF-1α. *Artif. Cell Nanomedicine, Biotechnol.* 48, 533–541. doi:10.1080/21691401.2020.1725026

Rahm, A. K., Wieder, T., Gramlich, D., Müller, M. E., Wunsch, M. N., El Tahry, F. A., et al. (2021). Differential Regulation of K Ca 2.1 (KCNN1) K + Channel Expression by Histone Deacetylases in Atrial Fibrillation with Concomitant Heart Failure. *Physiol. Rep.* 9, e14835. doi:10.14814/phy2.14835

Ren, Q., Zhao, S., Ren, C., and Ma, Z. (2018). RETRACTED: Astragalus Polysaccharide Alleviates LPS-Induced Inflammation Injury by Regulating miR-127 in H9c2 Cardiomyoblasts. *Int. J. Immunopathol Pharmacol.* 31, 205873841875918. doi:10.1177/2058738418759180

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, 139–140. doi:10.1093/bioinformatics/btp616

Saeidi, L., Ghaedi, H., Sadatamini, M., Vahabpour, R., Rahimipour, A., Shanaki, M., et al. (2018). Long Non-coding RNA LY86-AS1 and HCG27_201 Expression in Type 2 Diabetes Mellitus. *Mol. Biol. Rep.* 45, 2601–2608. doi:10.1007/s11033-018-4429-8

Saleh, M., and Ambrose, J. A. (2018). Understanding Myocardial Infarction. *F1000Res* 7, 1378. doi:10.12688/f1000research.15096.1

Sharma, R., Iovine, C., Agarwal, A., and Henkel, R. (2021). TUNEL Assay-Standardized Method for Testing Sperm DNA Fragmentation. *Andrologia* 53, e13738. doi:10.1111/and.13738

Shi, W., Sun, Y., Wang, J., Tang, Y., Zhou, S., Xu, Z., et al. (2021). Trem1 Mediates Neuronal Apoptosis via Interaction with SYK after Spinal Cord Ischemia-Reperfusion Injury. *Am. J. Transl Res.* 13, 6117–6125.

Shindo, T., Ito, K., Ogata, T., Hatanaka, K., Kurosawa, R., Eguchi, K., et al. (2016). Low-Intensity Pulsed Ultrasound Enhances Angiogenesis and Ameliorates Left Ventricular Dysfunction in a Mouse Model of Acute Myocardial Infarction. *Arterioscler Thromb. Vasc. Biol.* 36, 1220–1229. doi:10.1161/atvbaha.115.306477

Shiojima, I., Sato, K., Izumiya, Y., Schiekofer, S., Ito, M., Liao, R., et al. (2005). Disruption of Coordinated Cardiac Hypertrophy and Angiogenesis Contributes to the Transition to Heart Failure. *J. Clin. Invest.* 115, 2108–2118. doi:10.1172/jci24682

Si, Y., Liu, J., Shen, H., Zhang, C., Wu, Y., Huang, Y., et al. (2019). Fisetin Decreases TET1 Activity and CCNY/CDK16 Promoter 5hmC Levels to Inhibit the Proliferation and Invasion of Renal Cancer Stem Cell. *J. Cel Mol Med* 23, 1095–1105. doi:10.1111/jcmm.14010

Song, H., Feng, X., Zhang, H., Luo, Y., Huang, J., Lin, M., et al. (2019). METTL3 and ALKBH5 Oppositely Regulate m6A Modification of TFEB mRNA, Which Dictates the Fate of Hypoxia/reoxygenation-Treated Cardiomyocytes. *Autophagy* 15, 1419–1437. doi:10.1080/15548627.2019.1586246

Sun, X., Chen, P., Chen, X., Yang, W., Chen, X., Zhou, W., et al. (2021). KIF4A Enhanced Cell Proliferation and Migration via Hippo Signaling and Predicted a Poor Prognosis in Esophageal Squamous Cell Carcinoma. *Thorac. Cancer* 12, 512–524. doi:10.1111/1759-7714.13787

Teng, S., Ma, C., Yu, Y., and Yi, C. (2019). Hydroxyurea Promotes TET1 Expression and Induces Apoptosis in Osteosarcoma Cells. *Biosci. Rep.* 39, 1. doi:10.1042/bsr20190456

Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-Performance Genomics Data Visualization and Exploration. *Brief. Bioinform.* 14, 178–192. doi:10.1093/bib/bbs017

Thygesen, K., Alpert, J. S., and White, H. D. (2007). Universal Definition of Myocardial Infarction. *J. Am. Coll. Cardiol.* 50, 2173–2195. doi:10.1016/j.jacc.2007.09.011

Wang, M., Liu, J., Zhao, Y., He, R., Xu, X., Guo, X., et al. (2020a). Upregulation of METTL14 Mediates the Elevation of PERP mRNA N6 Adenosine Methylation Promoting the Growth and Metastasis of Pancreatic Cancer. *Mol. Cancer* 19, 130. doi:10.1186/s12943-020-01249-8

Wang, Q., Liu, B., Wang, Y., Bai, B., Yu, T., and Chu, X. m. (2020b). The Biomarkers of Key miRNAs and Target Genes Associated with Acute Myocardial Infarction. *PeerJ* 8, e9129. doi:10.7717/peerj.9129

Wei, J., Xie, Q., Liu, X., Wan, C., Wu, W., Fang, K., et al. (2020a). Identification the Prognostic Value of Glutathione Peroxidases Expression Levels in Acute Myeloid Leukemia. *Ann. Transl Med.* 8, 678. doi:10.21037/atm-20-3296

Wei, L., Zhou, Q., Tian, H., Su, Y., Fu, G.-h., and Sun, T. (2020b). Integrin β3 Promotes Cardiomyocyte Proliferation and Attenuates Hypoxia-Induced Apoptosis via Regulating the PTEN/Akt/mTOR and ERK1/2 Pathways. *Int. J. Biol. Sci.* 16, 644–654. doi:10.7150/ijbs.39414

Xu, M., Chen, X., Lin, K., Zeng, K., Liu, X., Xu, X., et al. (2019). lncRNA SNHG6 Regulates EZH2 Expression by Sponging miR-26a/b and miR-214 in Colorectal Cancer. *J. Hematol. Oncol.* 12, 3. doi:10.1186/s13045-018-0690-5

Yankova, E., Blackaby, W., Albertella, M., Rak, J., De Braekeleer, E., Tsagkogeorga, G., et al. (2021). Small-molecule Inhibition of METTL3 as a Strategy against Myeloid Leukaemia. *Nature* 593, 597–601. doi:10.1038/s41586-021-03536-w

Yu, F., Lin, Y., Ai, M.-M., Tan, G.-J., Huang, J.-L., and Zou, Z.-R. (2021). Knockdown of Circular RNA hsa_circ_PVT1 Inhibited Laryngeal Cancer Progression via Preventing Wnt4/β-Catenin Signaling Pathway Activation. *Front. Cel Dev. Biol.* 9, 658115. doi:10.3389/fcell.2021.658115

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

Zeng, B., Lin, G., Ren, X., Zhang, Y., and Chen, H. (2010). Over-expression of HO-1 on Mesenchymal Stem Cells Promotes Angiogenesis and Improves Myocardial Function in Infarcted Myocardium. *J. Biomed. Sci.* 17, 80. doi:10.1186/1423-0127-17-80

Zeng, B., Liu, L., Wang, S., and Dai, Z. (2017). ILK Regulates MSCs Survival and Angiogenesis Partially through AKT and mTOR Signaling Pathways. *Acta Histochem.* 119, 400–406. doi:10.1016/j.acthis.2017.04.003

Zhang, D., Wang, B., Ma, M., Yu, K., Zhang, Q., and Zhang, X. (2019). lncRNA HOTAIR Protects Myocardial Infarction Rat by Sponging miR-

519d-3p. *J. Cardiovasc. Trans. Res.* 12, 171–183. doi:10.1007/s12265-018-9839-4

Zhang, D. X., Ma, D. Y., Yao, Z. Q., Fu, C. Y., Shi, Y. X., Wang, Q. L., et al. (2016). ERK1/2/p53 and NF-Kb Dependent-PUMA Activation Involves in Doxorubicin-Induced Cardiomyocyte Apoptosis. *Eur. Rev. Med. Pharmacol. Sci.* 20, 2435–2442.

Zhou, J., Jiang, Y. Y., Chen, H., Wu, Y. C., and Zhang, L. (2020). Tanshinone I Attenuates the Malignant Biological Properties of Ovarian Cancer by Inducing Apoptosis and Autophagy via the Inactivation of PI3K/AKT/mTOR Pathway. *Cell Prolif* 53, e12739. doi:10.1111/cpr.12739

Zhu, W., Sun, L., Zhao, P., Liu, Y., Zhang, J., Zhang, Y., et al. (2021). Macrophage Migration Inhibitory Factor Facilitates the Therapeutic Efficacy of Mesenchymal Stem Cells Derived Exosomes in Acute Myocardial Infarction through Upregulating miR-133a-3p. *J. Nanobiotechnol* 19, 61. doi:10.1186/s12951-021-00808-5

Zouggari, Y., Ait-Oufella, H., Bonnin, P., Simon, T., Sage, A. P., Guérin, C., et al. (2013). B Lymphocytes Trigger Monocyte Mobilization and Impair Heart Function after Acute Myocardial Infarction. *Nat. Med.* 19, 1273–1280. doi:10.1038/nm.3284

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 Molecular Biosciences

ORIGINAL RESEARCH
published: 31 January 2022
doi: 10.3389/fmolb.2021.800747

Complex Involvement of the Extracellular Matrix, Immune Effect, and Lipid Metabolism in the Development of Idiopathic Pulmonary Fibrosis

Weiping Qian[1,2†], Shu Xia[1,2†], Xiaoyun Yang[3,2†], Jiaying Yu[3,2], Bingpeng Guo[1,2], Zhengfang Lin[3,2], Rui Wei[1,2], Mengmeng Mao[1,2], Ziyi Zhang[1,2], Gui Zhao[1,2], Junye Bai[1,2], Qian Han[1,2†], Zhongfang Wang[3,2*†] and Qun Luo[1,2*†]*

[1]Department of Respiratory Medicine, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, [2]National Clinical Center for Respiratory Disease, Guangzhou Institute of Respiratory Health, Guangzhou, China, [3]State Key Laboratory of Respiratory Disease, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

Background and objective: Idiopathic pulmonary fibrosis (IPF) is an aggressive fibrotic pulmonary disease with spatially and temporally heterogeneous alveolar lesions. There are no early diagnostic biomarkers, limiting our understanding of IPF pathogenesis.

Methods: Lung tissue from surgical lung biopsy of patients with early-stage IPF ($n = 7$), transplant-stage IPF ($n = 2$), and healthy controls ($n = 6$) were subjected to mRNA sequencing and verified by real-time quantitative PCR (RT-qPCR), immunohistochemistry, Western blot, and single-cell RNA sequencing (scRNA-Seq).

Results: Three hundred eighty differentially expressed transcripts (DETs) were identified in IPF that were principally involved in extracellular matrix (ECM) remodeling, lipid metabolism, and immune effect. Of these DETs, 21 (DMD, MMP7, POSTN, ECM2, MMP13, FASN, FADS1, SDR16C5, ACAT2, ACSL1, CYP1A1, UGT1A6, CXCL13, CXCL5, CXCL14, IL5RA, TNFRSF19, CSF3R, S100A9, S100A8, and S100A12) were selected and verified by RT-qPCR. Differences in DMD, FASN, and MMP7 were also confirmed at a protein level. Analysis of scRNA-Seq was used to trace their cellular origin to determine which lung cells regulated them. The principal cell sources of DMD were ciliated cells, alveolar type I/II epithelial cells (AT cells), club cells, and alveolar macrophages (AMs); MMP7 derives from AT cells, club cells, and AMs, while FASN originates from AT cells, ciliated cells, and AMs.

Abbreviations: ACAT2, acetyl-CoA acetyltransferase 2; ACSL1, acyl-CoA synthetase long chain family member 1; CSF3R, colony stimulating factor 3 receptor; CXCL13, C-X-C motif chemokine ligand 13; CXCL14, C-X-C motif chemokine ligand 14; CXCL5, C-X-C motif chemokine ligand 5; CYP1A1, cytochrome P450 family 1 subfamily A member 1; DMD, dystrophin; ECM2, extracellular matrix protein 2; FADS1, fatty acid desaturase 1; FASN, fatty acid synthase; IL5RA, interleukin 5 receptor subunit alpha; MMP13, matrix metallopeptidase 13; MMP7, matrix metallopeptidase 7; POSTN, periostin; S100A12, S100 calcium binding protein A12; S100A8, S100 calcium binding protein A8; S100A9, S100 calcium binding protein A9; SDR16C5, short chain dehydrogenase/reductase family 16C member 5; TNFRSF19, tumor necrosis factor receptor superfamily member 19; UGT1A6, UDP glucuronosyltransferase family 1 member A6; SLB, surgical lung biopsy; IPF, idiopathic pulmonary fibrosis; HC, healthy controls.

OPEN ACCESS

Edited by:
William C. Cho, QEH, Hong Kong SAR, China

Reviewed by:
Ivette Buendia-Roldan, Instituto Nacional de Enfermedades Respiratorias-México (INER), Mexico
Thomas John Brett, Washington University in St. Louis, United States

***Correspondence:**
Qian Han
hanqian1020@yahoo.com
Zhongfang Wang
wangzhongfang@gird.cn
Qun Luo
luoqunx@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *24 October 2021*
Accepted: *06 December 2021*
Published: *31 January 2022*

Citation:
Qian W, Xia S, Yang X, Yu J, Guo B, Lin Z, Wei R, Mao M, Zhang Z, Zhao G, Bai J, Han Q, Wang Z and Luo Q (2022) Complex Involvement of the Extracellular Matrix, Immune Effect, and Lipid Metabolism in the Development of Idiopathic Pulmonary Fibrosis. Front. Mol. Biosci. 8:800747. doi: 10.3389/fmolb.2021.800747

Conclusion: Our data revealed a comprehensive transcriptional mRNA profile of IPF and demonstrated that ECM remodeling, lipid metabolism, and immune effect were collaboratively involved in the early development of IPF.

Keywords: idiopathic pulmonary fibrosis, mRNA sequencing, extracellular matrix remodeling, lipid metabolism, immune effect

INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive, fibrosing interstitial lung disease (ILD) of unknown cause, characterized by a histopathologic and/or radiologic pattern of usual interstitial pneumonia (UIP) with spatial heterogeneity, temporal heterogeneity, and honeycomb lesions (Lederer and Martinez, 2018; Lynch et al., 2018; Plantier et al., 2018; Raghu et al., 2018). IPF is difficult to differentiate from other ILDs (Schoenheit et al., 2011). Effective treatment is largely confined to lung transplantation, but this is limited by donor availability and transplantation complications (Mushiroda et al., 2008; Lederer and Martinez, 2018; Raghu et al., 2018). IPF treatments in the clinic have consistently failed, in part due to the limited understanding of IPF and lack of predictive diagnostic/prognostic biomarkers. Over the past decades, many studies (Martinez et al., 2017; Desai et al., 2018; Morse et al., 2019; Sivakumar et al., 2019) have revealed that IPF is associated with extracellular matrix (ECM) remodeling, lipid metabolism, and immune effect. For example, repeat exposure to a damaging environment can cause alveolar epithelial damage, while alveolar type II epithelial cells (AT2) are abnormally activated to initiate pulmonary fibrosis and promote proliferation, apoptosis, aging, and partial epithelial–mesenchymal transition (Martinez et al., 2017). In addition, Morse et al. reported that macrophages are highly plastic and functionally heterogeneous in IPF (Morse et al., 2019). Patients with IPF also show downregulation of cholesterol homeostasis pathway-related genes (Sivakumar et al., 2019) and low peripheral blood HDL-C levels (Aihara et al., 2013). Although these studies have shed light on pathways that could lead to IPF, our understanding of their involvement remains very limited. Several studies have utilized mRNA-Seq (Konishi et al., 2009; Steele et al., 2015; Kusko et al., 2016; Paplińska-Goryca et al., 2019; Wang et al., 2019) to identify genes and/or pathways that are differentially regulated, providing clues to specific mechanisms that underlie the occurrence and development of IPF. Nonetheless, these have been based on transplant-stage IPF samples. Because of the heterogeneity of UIP in IPF, a broader array of IPF lung tissue (including early and transplant stage) would provide more complete biological information about the IPF transcriptome and more clues to the common pathogenesis of IPF.

In the present study, we hypothesized that mRNA-Seq analysis of IPF lung tissue obtained from surgical lung biopsy (SLB) and lung transplantation would provide more comprehensive biological information and facilitate early diagnosis and a better understanding of the comprehensive pathogenesis of heterogeneous IPF. We performed mRNA-Seq on lung tissue from a cohort of patients with early-stage IPF ($n = 7$) who underwent SLB and those with transplant-stage IPF ($n = 2$) who underwent lung transplantation and compared the results with tissue obtained from healthy controls. Of the 380 differentially expressed transcripts (DETs), 21 target DETs that were mainly related to ECM remodeling, immune effect, and lung lipid metabolism-related pathways were selected for further verification by real-time quantitative PCR (RT-qPCR) and single-cell RNA sequencing (scRNA-Seq) analysis. Our findings highlighted the central and coordinating roles of different cells, pathways, and genes in the development of IPF.

MATERIALS AND METHODS

Study Design and Inclusion Criteria for Participants

This was a retrospective, single-center study that involved patients with IPF and healthy control (HC) treated at the ILD Center of the First Affiliated Hospital of Guangzhou Medical University between 2014 and 2018. The study protocol was approved by the local independent ethics committee (2020-71). Informed consent was obtained from participants.

mRNA Expression Analysis

Total RNAs of tissue were extracted using a Trizol reagent kit, and sequencing and analysis performed using the Illumina novaseq 6000 platform and Bowtie2, HISAT2 (v2.1.0), StringTie (v1.3.1), and R package gmodels, respectively. Patient inclusion and exclusion criteria, methods of mRNA expression analysis, 21 target gene selection and RT-qPCR, protein expression, and cell sources of DMD, MMP7, and FASN are detailed in the **Supplementary Materials**.

Statistical Analysis

Total transcript expression data were analyzed using Fisher's exact test or Wald test at a false discovery rate (FDR) of 5% and $|\log2FC| > 1$. Statistical analysis for RT-qPCR and Western blot (WB) was performed with a two-tailed Student's t-test using GraphPad Prism software (v 8.0.2.263), or SPSS (v25.0). Results are shown as mean ± SD in the table or mean ± SEM in figures. All expression experiments conducted *in vitro* were repeated at least three times with three samples. Statistical significance was set at $p < 0.05$.

RESULTS

Clinical Characteristics of Participants

Ten patients with suspected IPF and six healthy controls were initially recruited to the study. IPF was confirmed in nine out of the ten patients (seven with early-stage disease and two with transplant-stage disease, all with a pathological UIP pattern).

TABLE 1 | Characteristics of participants in mRNA-Seq study.

	IPF (*n* = 9)	HC (*n* = 6)	*t* value	*p*-value
Age, year	57.67 ± 8.34	33.33 ± 10.50	5.00	0.0044
Male gender, No. (%)	9 (100)	6 (100)	NA	NA
Race (No.)	Asian (9)	Asian (6)	NA	NA
Lung sample source	9	6	NA	NA
SLB	7	0	NA	NA
Transplantation	2	6	NA	NA
FVC pred, %	79.69 ± 26.38	NA	NA	NA
DLco pred, %	58.75 ± 22.57	NA	NA	NA

Data are presented as mean ± SD. The p-value was obtained using a two-tailed Student's t-test. Abbreviations: SLB, surgical lung biopsy; NA, not available; FVC, forced vital capacity; DLco, diffusing capacity of lungs for carbon monoxide.

Compared to healthy lung tissue sections (**Figures 1A, D**), patients with early-stage IPF had fewer fibrotic lesions [shown in high-resolution computed tomography (HRCT) (**Figure 1B**)], retained more alveolar units (shown by hematoxylin and eosin staining) (**Figure 1E**) and had better lung function than those with transplant-stage IPF which is characterised by typical UIP along with significant honeycomb change (**Figures 1C, F**). Characteristics and comparison of both groups are shown in **Table 1**, **Supplementary Table S1**, **Figure 1**, and **Supplementary Figure S2**.

IPF mRNA-Seq Data

The mRNA-Seq experiments were performed using nine IPF and six HC lung tissue samples, and 101,590 transcripts were identified. Principal component analysis of all transcripts revealed that the transcriptional mRNAs of IPF tissue differed to those of the HC samples (**Figure 2A**). A total of 380 DETs in the IPF group were identified at an FDR of 5% and $|\log2FC| > 1$, of which 195 were upregulated and 185 were downregulated. Interestingly, there were some significantly different DETs, including DMD, RBM6, S100A8, S100A9, and CYP1A1 (**Figure 2B** and **Supplementary Table S2**). **Figure 2C** shows the heat map of 380 DETs cluster analysis.

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Pathway Enrichment Analyses

Following Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses for downregulated and upregulated DETs, significant BP (biological process) terms of GO and KEGG pathways were identified and are shown in **Figure 3**. Processes enriched by upregulated DETs involved principally ECM organization, extracellular structure organization, and cell adhesion. Those enriched BPs were largely related to response to stress, small molecule metabolic processes, and lipid metabolic processes. The

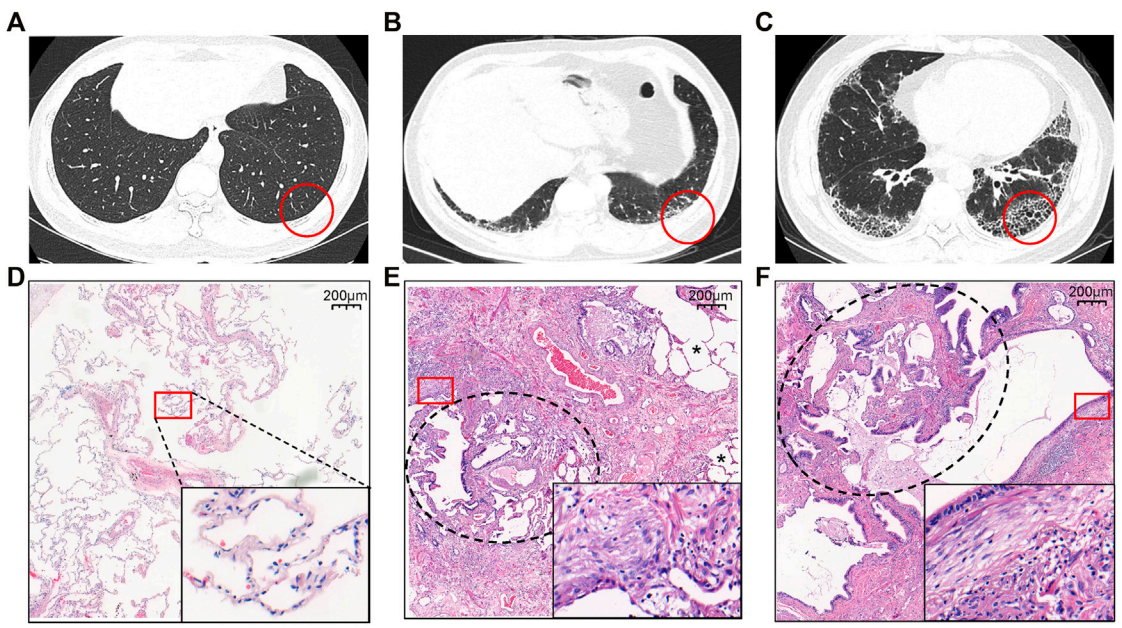


FIGURE 1 | Representative HRCT **(A–C)** and histology **(D–F)** findings in patients with IPF **(B, C, E, F)** and healthy controls **(A, D)**. **(A and D)** The normal appearance of HRCT and histopathology from HC. **(B)** The HRCT topographies in the IPF patients at early stage of SLB demonstrating subpleural, basal-predominant (often heterogeneous) subtle reticulation, mild ground glass opacity (GGO), or distortion (indeterminate for UIP). **(C)** The HRCT topographies in the IPF patients at the transplant stage from lung transplantation demonstrating a typical UIP with characteristics of subpleural, basal-predominant (heterogeneous) honeycombing, and bronchiolectasis. **(E and F)** Histopathology patterns of UIP, characterized by dense fibrosis with a predilection for subpleural and paraseptal parenchyma with associated architectural distortion in the form of microscopic honeycomb change black circle pointed juxtaposed with relatively unaffected lung parenchyma (*) and fibroblast foci (square circle). Among them, E comes from the IPF lung tissue of SLB, with more relatively unaffected lung parenchyma (*), while F is from transplanted IPF lung tissue, with honeycomb change (black circle). The red circle is the area of the biopsy shown in HRCT, while the square circle is used for magnification to present very dense fibrosis (collagen fibers and fibroblasts). Abbreviation: SLB, surgical lung biopsy.

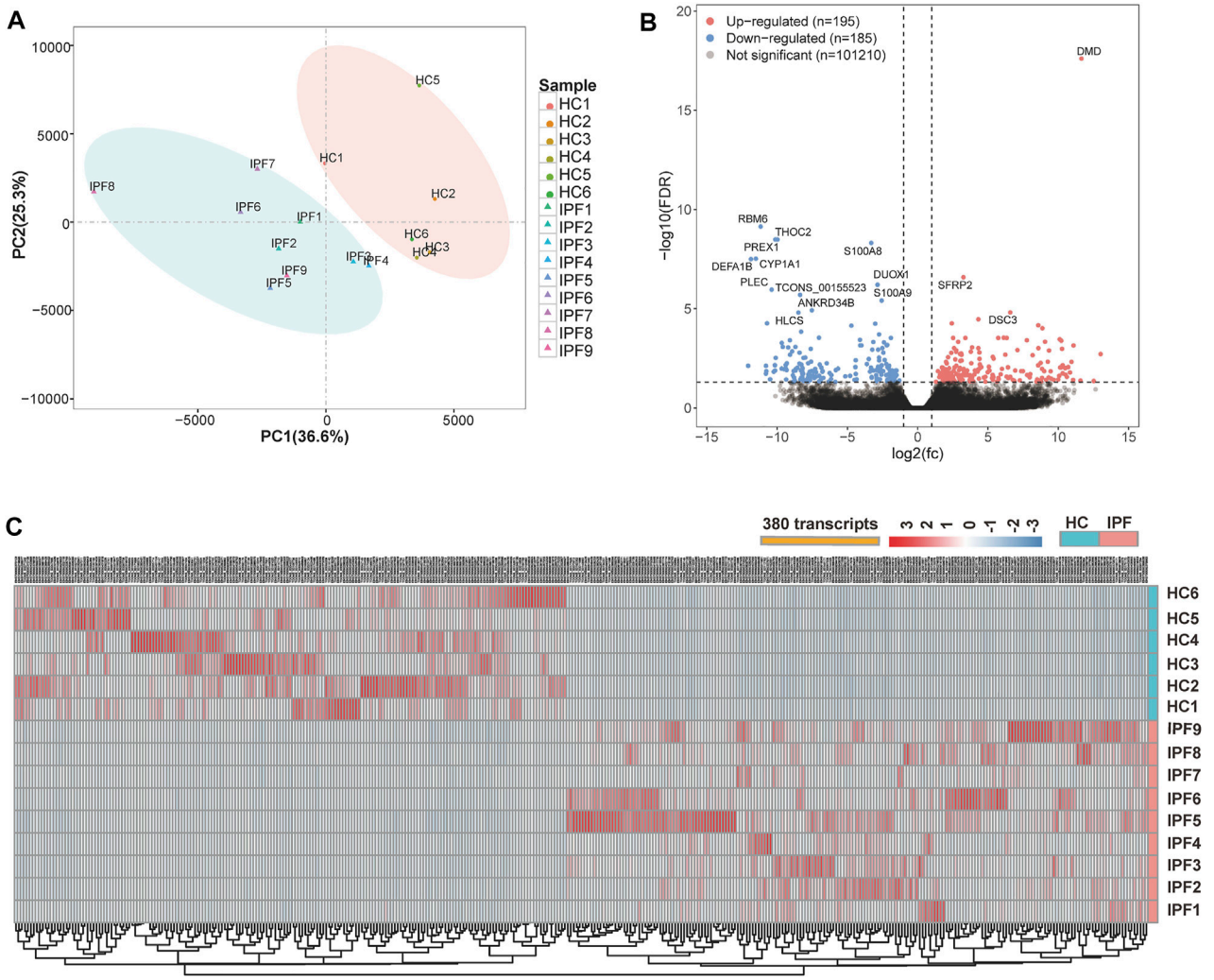


FIGURE 2 | Overview of the differential expression of transcripts between lung tissue samples from patients with IPF and healthy controls. **(A)** Principal component analysis of the mRNA-Seq for lung tissue samples shows a clear separation of the samples from HC and samples from patients with IPF. **(B)** Volcano plots depict significantly dysregulated gene changes of the IPF versus control at a false discovery rate (FDR) < 0.05 and |log2FC| > 1. Downregulated genes are shown in blue, and upregulated genes shown in red. DMD is a transcript with the lowest *p*-value. **(C)** Heat map shows the overall distribution of each differentially expressed transcript in all samples from IPF and HC. Abbreviations: FC, fold change; IPF1–IPF7, the lung samples of IPF with early stage from SLB; IPF8–IPF9, the lung samples of IPF with transplant stage from lung transplantation.

KEGG pathways enriched by upregulated DETs were mainly the Toll-like receptor signaling pathway and ECM–receptor interaction, while KEGG pathways enriched by downregulated DETs involved fatty acid metabolism and fatty acid biosynthesis. Interestingly, the cytokine–cytokine receptor interaction pathway was enriched by both upregulated and downregulated DETs. Taken together, these enriched pathways were classified into three main biological processes: ECM remodeling, lipid metabolism, and immune effect.

Selection and Verification of Target Genes

Target genes were selected based on the criteria of |log2FC| > 2 and genes expressed in at least two-thirds of IPF or HC samples. In total, 21 target genes were selected and classified into the following groups: 1) DMD, MMP7, POSTN, ECM2, and MMP13 related to ECM remodeling, 2) FASN, FADS1, SDR16C5, ACAT2, ACSL1, CYP1A1, and UGT1A6 associated with lipid metabolism, and 3) CXCL13, CXCL5, CXCL14, IL5RA, TNFRSF19, CSF3R, S100A9, S100A8, and S100A12 allied with immune effect. Following mRNA-Seq analysis, we determined that each of the 21 target genes may have multiple DETs: two transcripts in DMD, two in ECM, four in CYP1A1, three in SDR16C5, two in FASN, and three in CSF3R (**Figure 4A**). In addition, analysis of IPF scRNA-Seq data (unpublished) with deeper sequencing revealed that, except for CYPA1A1, the trend of the relative expression in IPF over HC was similar to that of mRNA-Seq (**Figure 4B**). RT-qPCR results also revealed that 11 genes (DMD, MMP7, POSTN, ECM2, MMP13, UGT1A6, CXCL13, CXCL5, CXCL14, IL5RA, and TNFRSF19) were

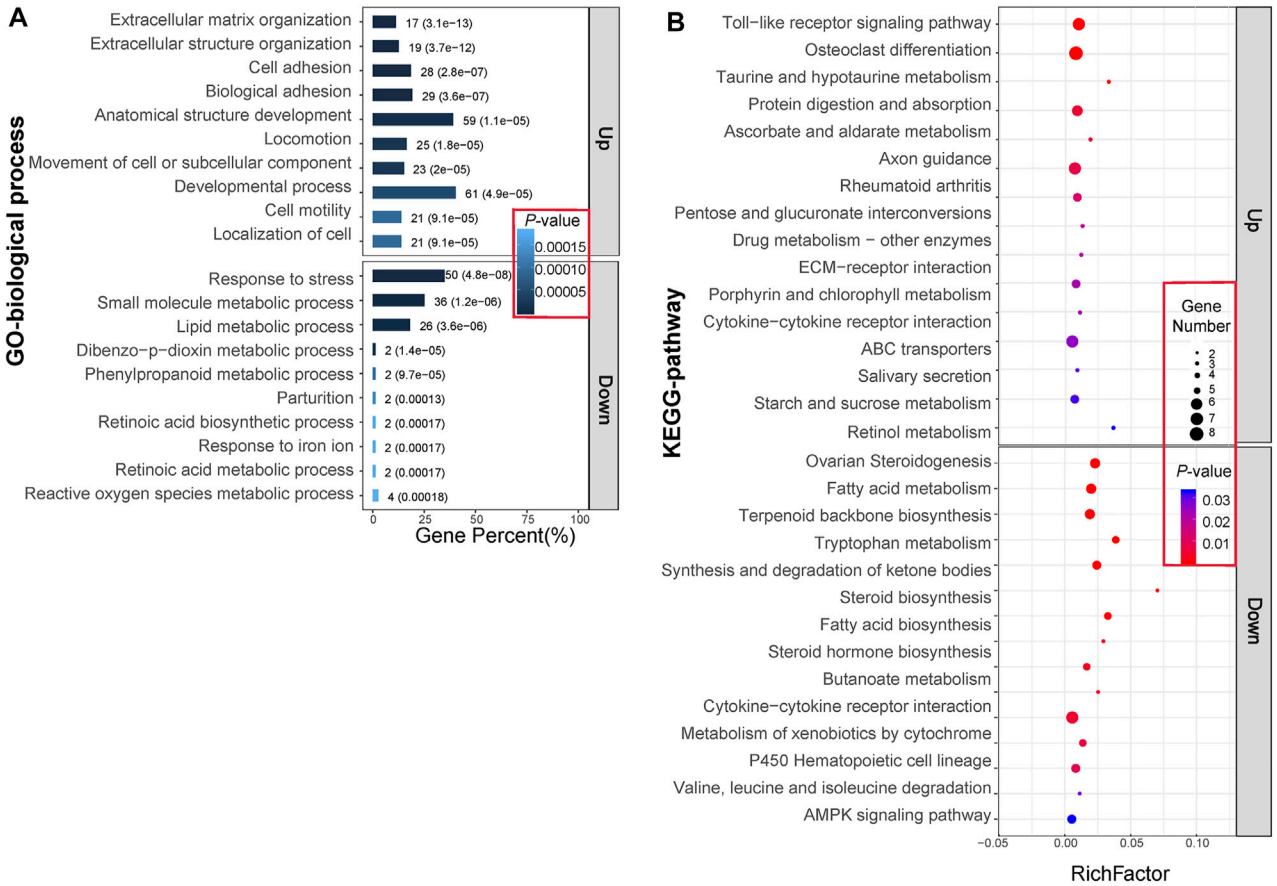


FIGURE 3 | GO **(A)** and KEGG **(B)** enrichment analysis of 380 differentially expressed transcripts (DETs). **(A)** The bar graphs represent the top 10 BP terms with the lowest *p*-value for upregulated and downregulated gene expression patterns. The numbers and parenthesis value on each bar graph separately represent the numbers of DETs enriched in it and the *p*-value. **(B)** The solid dots represent the top 15 KEGG pathways with the lowest *p*-values for upregulated and downregulated gene expression patterns. Different colored dots indicate the different *p*-value, while dots of different diameters represent the numbers of DETs enriched. Abbreviations: Up, upregulation in IPF group; Down, downregulation in IPF group.

upregulated, while 10 genes (FASN, FADS1, SDR16C5, ACAT2, ACSL1, CYP1A1, CSF3R, S100A9, S100A8, and S100A12) were downregulated when compared with HC (**Figure 4C**).

Further Analyses of DMD, MMP7, and FASN

To further verify the regulated genes, WB analysis was performed using IPF and HC tissue for DMD (the most differentially expressed transcript in mRNA-Seq analysis), MMP7 and FASN (the most significant fold change and significant statistical change in RT-qPCR). The immunohistology chemistry (IHC) results in **Figure 5A** show that: 1) DMD protein was expressed in the majority of the lung tissues and lung cells, mainly in the nucleus around the alveoli and around the bronchial lumen but also in the cytoplasm and in the muscle cell membrane in the IPF and HC groups; 2) MMP7 protein was clearly evident in the cytoplasm around the alveoli and bronchial lumen in the IPF group, but expression was low in the HC group; and 3) protein distribution of FASN was similar to that of DMD, mainly in the cytoplasm around the alveoli and bronchial lumen, and the staining of HC was darker than that in IPF. Next, we performed WB and further conformed that the protein expression of DMD and MMP7 in the IPF group was also significantly higher than that in the HC group, while the protein expression of FASN was clearly lower (**Figure 5B**). Finally, using the IPF scRNA-Seq data, we compared the cell sources of differentially expressed DMD, MMP7, and FASN genes in IPF and HC (**Figures 5C, D**). We demonstrated that DMD was mainly expressed in ciliated cells, AT cells, and alveolar macrophages (AMs) in HC while mostly expressed in ciliated cells, AT cells, and club cells in IPF lung tissue and that DMD expression was upregulated in AT (1.5-fold) and club cells (2.7-fold) in IPF lung tissue (**Figure 5E**). Moreover, MMP7 was mainly expressed in AT cells, club cells, and AMs in both IPF and HC lung tissue, and it was 30.7-, 4.4-, and 11.1-fold upregulated in AT cells, club cells, and AMs of IPF. Finally, FASN was mainly expressed in AT cells, ciliated cells, and AMs in both groups and was about 2-fold downregulated in AT and ciliated cells in IPF. In addition, it is worthy to note that DMD, MMP7, and FASN were higher in early-stage than in the transplant-stage IPF tissue (**Supplementary Figure S1**).

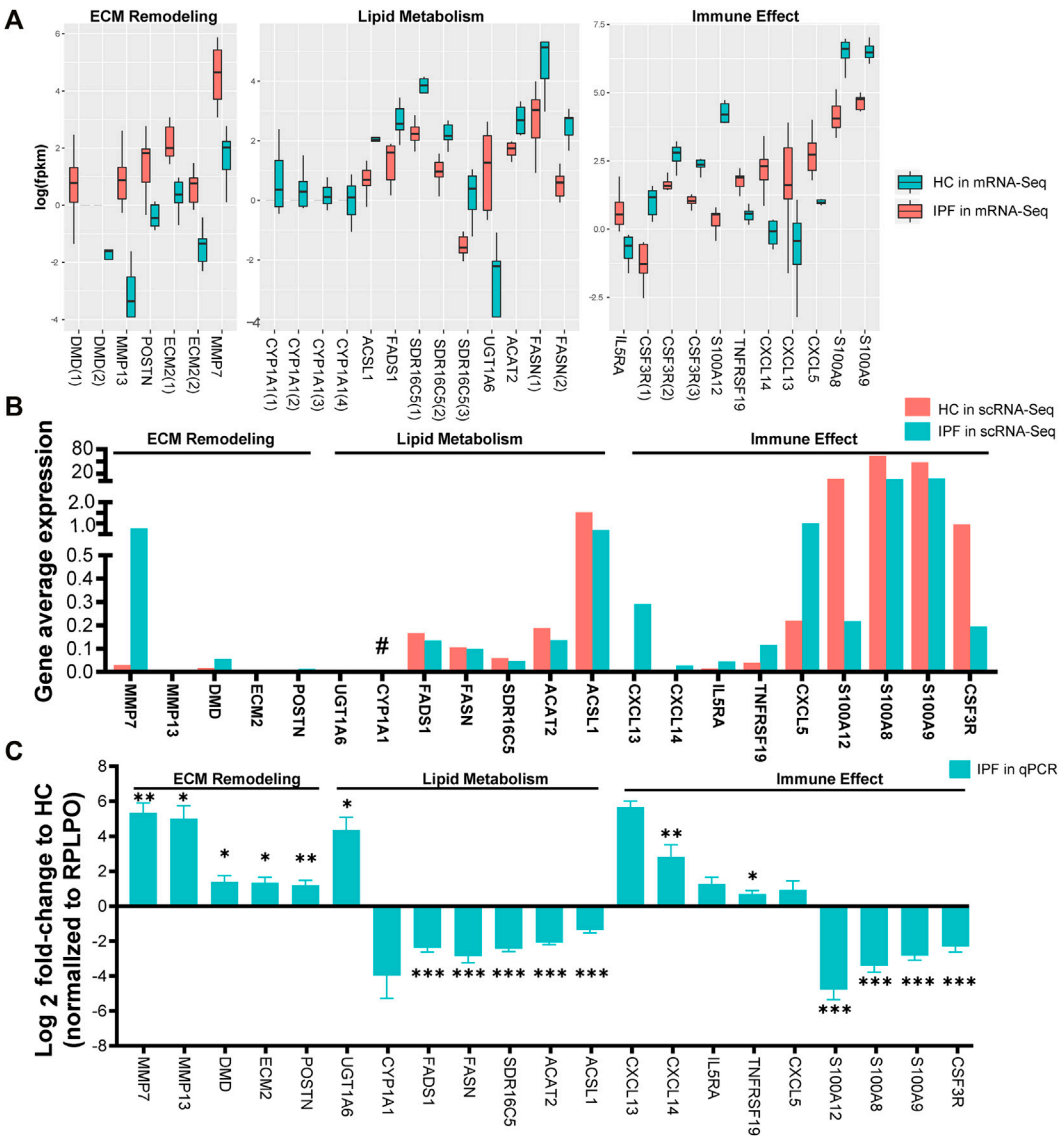


FIGURE 4 | Selection and verification of 21 target genes related to ECM remodeling, lipid metabolism, and immune effect. **(A)** The fpkm expression of 21 target genes related to ECM remodeling, lipid metabolism, and immune effect via mRNA-Seq analysis from lung tissues of IPF and HC. fpkm, fragments per kilobase million; Gene(NO.) represents having one or more differential transcripts of the gene in IPF compared with HC. **(B)** The fpkm expression of the 21 target genes via scRNA-Seq data from lung tissue of 9 (6 from SLB, 3 from lung transplantation) IPF and 4 HC. Compared with HC, the trend of IPF gene fpkm expression detected by scRNA-Seq was opposite to that of mRNA-Seq (#). **(C)** Relative expression of the 21 target genes in lung tissue of 11 (5 from SLB, 6 from lung transplantation) IPF and 7 HC via RT-qPCR experience. RPLPO was the internal reference. The bar graphs show mean ± SEM. Statistical significance was set at $p < 0.05$, and $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$ compared with HC group using a two-tailed Student's t-test.

DISCUSSION

IPF is usually characterized by a histopathologic and/or radiologic pattern of UIP with spatial heterogeneity, temporal heterogeneity, and the presence of honeycomb lesions (Lederer and Martinez, 2018; Lynch et al., 2018; Plantier et al., 2018; Raghu et al., 2018). Investigation of more comprehensive IPF specimens will help better understand the pathogenesis of IPF and unravel an in-depth mechanism underneath. Recently, an increasing number of studies have focused on the pathogenesis of IPF using high-throughput mRNA-Seq and scRNA-Seq. Nonetheless, the sources of tissue of the most previous IPF transcriptome studies are mainly from patients undergoing lung transplantation (Vukmirovic et al., 2017; Luzina et al., 2018; McDonough et al., 2019), which is generally considered as "late" IPF tissue. In this study, in addition to lung tissue specimens from transplant-stage IPF, more early-stage IPF specimens obtained from SLB were included for mRNA-Seq analysis. This is helpful to identify the complete transcriptional profile of IPF and help better the

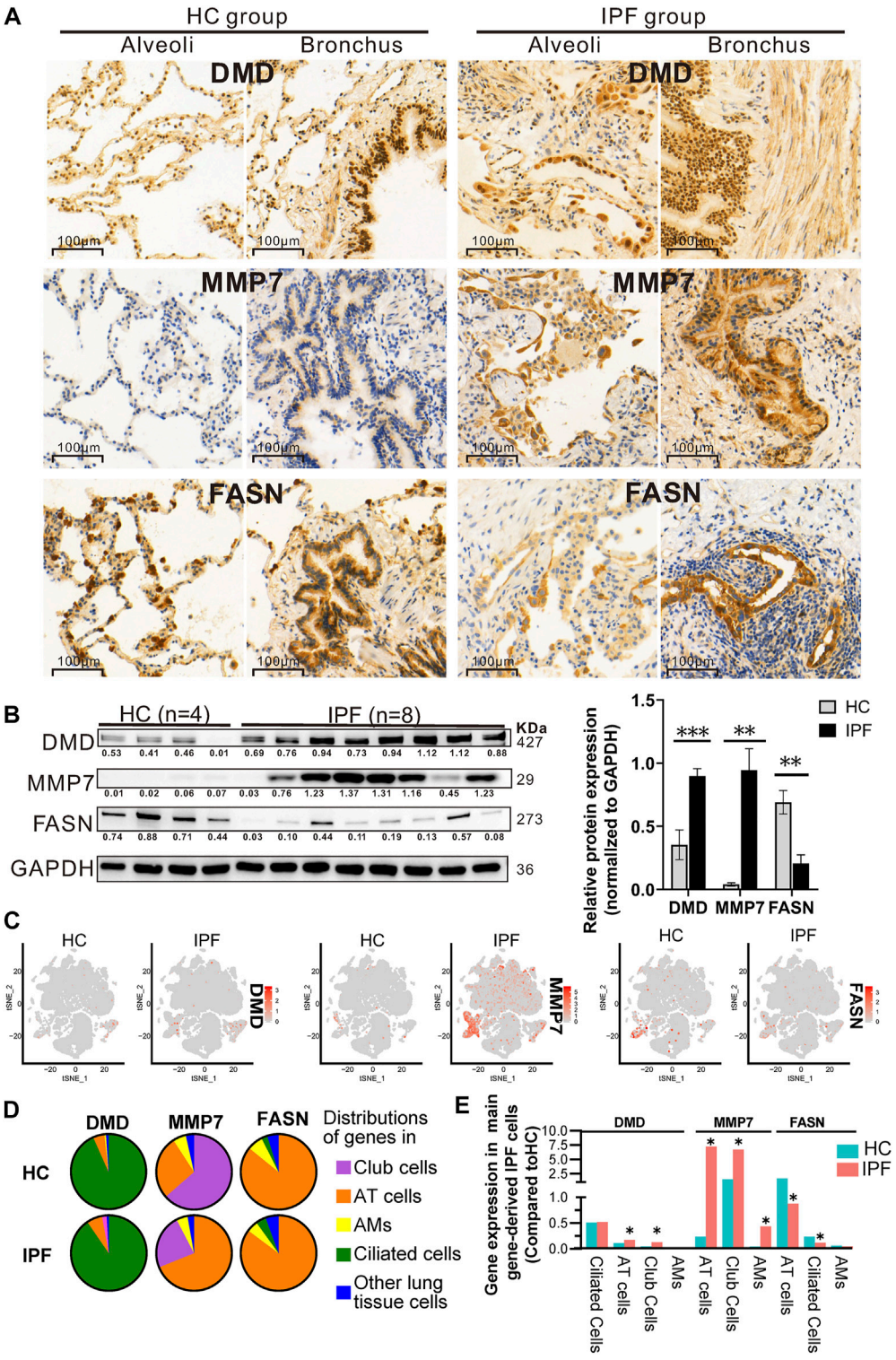


FIGURE 5 | Protein expression of and cell origins of DMD, MMP7, and FASN based on multiple analyses of IHC **(A)**, WB **(B)**, and scRNA-Seq **(C–E)**. **(A)** The protein distribution of DMD, MMP7, and FASN in lung tissue of IPF and HC via IHC experience. Each gene displays protein expression in the two fields of alveoli and bronchi. In addition, brown indicates that the gene has protein expression, and the darker the color, the greater the expression. **(B)** The protein expression of DMD, MMP7, and FASN in lung tissue of IPF and HC by WB experience, showing WB exposure bands (on the left) and bar charts of statistical analysis (on the right). Lung tissue derived from 8 IPF (4 from SLB and 4 from transplantation) and 4 HC. And GAPDH served as the internal reference. Error bars indicate mean ± SEM, and their three results were consistent with that of mRNA-Seq according to a two-tailed Student's t-test. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. **(C–E)** tSNE plots **(C)** displaying the

(*Continued*)

FIGURE 5 | expression level of DMD, MMP7, and FASN in all cell clusters from IPF and HC, pie chart **(D)** showing three lung cell sources expressing DMD, MMP7, and FASN genes, and histogram **(E)** showing the differential expression of the first three cells of DMD, MMP7, and FASN in IPF and HC groups. Compared with HC, the gene expression of DMD, MMP7, and FASN in IPF was upregulated or downregulated by at least 1.5-fold (*).

understanding of earlier development of IPF. In the present study, we identified several important pathways and categorized them into three biological processes, including ECM remodeling, lipid metabolism, and immune effect, and genes (DMD, MMP7, POSTN, ECM2, MMP13, FASN, FADS1, SDR16C5, ACAT2, ACSL1, CYP1A1, UGT1A6, CXCL13, CXCL5, CXCL14, IL5RA, TNFRSF19, CSF3R, S100A9, S100A8, and S100A12) involved in the processes. Additionally, we demonstrated the differential expression of DMD, MMP7, and FASN in AT cells, AMs, club cells, and ciliated cells of IPF.

Several genes such as MMP7, CSF3R, and S100A12 identified and characterized in our study have previously been reported (Deng et al., 2018; Chung et al., 2019; McDonough et al., 2019). For instance, we demonstrated that MMP7 was mainly enriched in ECM organization and ECM–receptor interaction, similar to a previous study, which described that MMP7 in blood levels has been consistently associated with IPF disease progression and survival (McDonough et al., 2019). We found that MMP7 mainly upregulated in the bronchi and alveoli of IPF; in addition, the main cell sources of MMP7 were AT cells, club cells, and AMs based on scRNA-Seq analysis. FASN was mainly enriched in fatty acid synthase and mainly downregulated in AT cells, ciliated cells, and AMs in the lung tissues of IPF. In fact, the phenomenon of lipid metabolism disorder in IPF patients has also been verified in animal models of fibrosis; for example, Chung et al. (2019) found the loss of AEC2 cell-specific FASN in bleomycin-induced pulmonary fibrosis in mice and the loss of lipid synthesis in AEC2 cells during mitochondrial injury could aggravate pulmonary fibrosis.

Despite the confirmation of inflammatory cytokines and innate and adaptive immune cell infiltration in IPF, several clinical experiments of immunosuppressive therapy for IPF have failed. This has downplayed the role of chronic inflammation and immune effect in the pathogenesis of IPF (Thannickal et al., 2014). Nonetheless, according to our current research results, failure of therapies based on immune intervention may be due to limited knowledge about the involvement of immune disorders in IPF. In our study, a great number of immune factors were altered in IPF patients, including upregulation of CXCL5, CXCL13, CXCL14, IL5RA, and TNFRSF19 and downregulation of CSF3R, S100A8, S100A9, and S100A12. Leurs and Lindholm (2013) reported that S100A12 was involved in the activation of signal transduction pathways in endothelial cells, vascular smooth muscle cells, and inflammatory cells, further leading to the transcription and secretion of proinflammatory cytokines and cell adhesion molecules. It has been reported that a high concentration of S100A12 in peripheral blood indicated a low overall survival rate of IPF, suggesting that the excessive inflammatory-immune response led to a worse prognosis (Richards et al., 2012). In addition to immune-related genes, there are also reports about immune cells in relation to IPF. Duffield et al. (2005) and Wynn (2011) revealed AMs were depleted in the early inflammation/maintenance phase of the fibrotic reaction in CD11b-DTR mice. This may have led to a relative reduction in scar formation and myofibroblasts, although fibrosis continued if AMs were consumed in the late remodeling/recovery stage of fibrosis.

DMD is a novel gene closely related to ECM remodeling in IPF. We found that DMD was mainly expressed by ciliated cells and secondarily expressed in AT and club cells in the lung tissue. Interestingly, there was an about 1.5-fold upregulation of DMD in AT and club cells of IPF tissue, while there was no significant difference between the ciliated cells of IPF and HC groups, suggesting that the role of DMD in IPF development may be cell type-dependent. It is worthy of note that the DMD gene is one of the largest protein-coding genes in the human genome, existing on the X chromosome, and encodes and synthesizes 427 KDa dystrophin protein that is distributed principally in the myocardium, smooth muscle, and skeletal muscle (Tuffery-Giraud et al., 2017). Edematous cells in Duchenne and Becker muscular dystrophy, X-linked dilated cardiomyopathy, and other muscle diseases result in cell fragmentation, leading to increased intracellular calcium ions that serve as a second messenger to activate the inflammatory cascade in organs (Sironi et al., 2002; Nardes et al., 2012; Kamdar and Garry, 2016). At present, the mainstream theory about the muscle degeneration and fibrosis causes of Duchenne muscular dystrophy (Mann et al., 2011) is that the integrity of the dystrophin protein in muscle cell membrane is lost and destroyed. When mechanically stimulated, sodium and calcium influx can be amplified and promote inflammatory cells to release various fibrogenic factors with consequent proliferation and differentiation of fibroblasts into myofibroblasts. Both can secrete ECM and ECM remodeling factors, with a subsequent imbalance of ECM synthesis and degradation and with consequent fibrosis. The histological feature of IPF is the gradual deposition of ECM and the gradual decrease in parenchymal cells. We hypothesize that dystrophin protein lesions occur in the lung tissue of IPF, especially in AT and club cells, and may lead to an imbalance between ECM synthesis and degradation in the lungs, promoting the occurrence and development of IPF.

The results of our study, from changes in the overall molecular level of IPF to alterations in the lung cells, showed that ECM remodeling, lipid metabolism, and immune effect are complex during the early development of IPF, suggesting that no single cell, gene, or pathway can contribute to the complex and heterogeneous nature of the disease. In addition, our findings supplement and reinforce the present transcriptome profile of human IPF disease.

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: NGDC Genome Sequence Archive, accession no: HRA001639.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the First Affiliated Hospital of Guangzhou Medical University ethics committee (ethical number: 2020-71). The study procedures were carried out in accordance with the tenets of the Declaration of Helsinki. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

WQ, SX, XY, QH, ZW, and QL designed the study. WQ, SX, XY, QH, ZW, and QL coordinated the study. WQ, JY, BG, and QL included patients. WQ, XY, JY, and ZL performed the experiments. WQ, SX, MM, QH, and QL participated in multidisciplinary discussion. RW, ZZ, GZ, JB, ZL, and JY participated in data collection. WQ, SX, XY, BG, and JY performed statistical analyses. WQ, SX, XY, QH, ZW, and QL wrote the first draft of the manuscript. All authors critically revised the manuscript for intellectually important content. All authors gave final approval for the publication of the work, and all accepted responsibility for the integrity of the work.

FUNDING

This work was supported by the Guangzhou Institute of Respiratory Health Independent research project (No. 2019GIRHZ05) and the National Key Technologies R&D Programme Precision Medicine Research (No. 2016YFC0901100). Neither the funders nor the study sponsor had a role in study design, in the collection, analysis, and interpretation of data, in the writing of the report, or in the decision to submit the report for publication.

ACKNOWLEDGMENTS

We thank Guangzhou Genedenovo Biotechnology Co., Ltd., for assisting in sequencing and bioinformatics analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.800747/full#supplementary-material

REFERENCES

Aihara, K., Handa, T., Nagai, S., Tanizawa, K., Ikezoe, K., Watanabe, K., et al. (2013). Impaired Endothelium-dependent Vasodilator Response in Patients with Pulmonary Fibrosis. *Respir. Med.* 107 (2), 269–275. doi:10.1016/j.rmed.2012.10.005

Chung, K.-P., Hsu, C.-L., Fan, L.-C., Huang, Z., Bhatia, D., Chen, Y.-J., et al. (2019). Mitofusins Regulate Lipid Metabolism to Mediate the Development of Lung Fibrosis. *Nat. Commun.* 10 (1), 3390. doi:10.1038/s41467-019-11327-1

Deng, Y., Li, Z., Liu, J., Wang, Z., Cao, Y., Mou, Y., et al. (2018). Targeted Resequencing Reveals Genetic Risks in Patients with Sporadic Idiopathic Pulmonary Fibrosis. *Hum. Mutat.* 39 (9), 1238–1245. doi:10.1002/humu.23566

Desai, O., Winkler, J., Minasyan, M., and Herzog, E. L. (2018). The Role of Immune and Inflammatory Cells in Idiopathic Pulmonary Fibrosis. *Front. Med.* 5, 43. doi:10.3389/fmed.2018.00043

Duffield, J. S., Forbes, S. J., Constandinou, C. M., Clay, S., Partolina, M., Vuthoori, S., et al. (2005). Selective Depletion of Macrophages Reveals Distinct, Opposing Roles during Liver Injury and Repair. *J. Clin. Invest.* 115 (1), 56–65. doi:10.1172/jci200522675

Kamdar, F., and Garry, D. J. (2016). Dystrophin-deficient Cardiomyopathy. *J. Am. Coll. Cardiol.* 67 (21), 2533–2546. doi:10.1016/j.jacc.2016.02.081

Konishi, K., Gibson, K. F., Lindell, K. O., Richards, T. J., Zhang, Y., Dhir, R., et al. (2009). Gene Expression Profiles of Acute Exacerbations of Idiopathic Pulmonary Fibrosis. *Am. J. Respir. Crit. Care Med.* 180 (2), 167–175. doi:10.1164/rccm.200810-1596oc

Kusko, R. L., Brothers, J. F., 2nd, Tedrow, J., Pandit, K., Huleihel, L., Perdomo, C., et al. (2016). Integrated Genomics Reveals Convergent Transcriptomic Networks Underlying Chronic Obstructive Pulmonary Disease and Idiopathic Pulmonary Fibrosis. *Am. J. Respir. Crit. Care Med.* 194 (8), 948–960. doi:10.1164/rccm.201510-2026oc

Lederer, D. J., and Martinez, F. J. (2018). Idiopathic Pulmonary Fibrosis. *N. Engl. J. Med.* 378 (19), 1811–1823. doi:10.1056/nejmra1705751

Leurs, P., and Lindholm, B. (2013). The AGE-RAGE Pathway and its Relation to Cardiovascular Disease in Patients with Chronic Kidney Disease. *Arch. Med. Res.* 44 (8), 601–610. doi:10.1016/j.arcmed.2013.11.002

Luzina, I. G., Salcedo, M. V., Rojas-Peña, M. L., Wyman, A. E., Galvin, J. R., Sachdeva, A., et al. (2018). Transcriptomic Evidence of Immune Activation in Macroscopically normal-appearing and Scarred Lung Tissues in Idiopathic Pulmonary Fibrosis. *Cel. Immunol.* 325, 1–13. doi:10.1016/j.cellimm.2018.01.002

Lynch, D. A., Sverzellati, N., Travis, W. D., Brown, K. K., Colby, T. V., Galvin, J. R., et al. (2018). Diagnostic Criteria for Idiopathic Pulmonary Fibrosis: a Fleischner Society White Paper. *Lancet Respir. Med.* 6 (2), 138–153. doi:10.1016/s2213-2600(17)30433-2

Mann, C. J., Perdiguero, E., Kharraz, Y., Aguilar, S., Pessina, P., Serrano, A. L., et al. (2011). Aberrant Repair and Fibrosis Development in Skeletal Muscle. *Skeletal Muscle* 1 (1), 21. doi:10.1186/2044-5040-1-21

Martinez, F. J., Collard, H. R., Pardo, A., Raghu, G., Richeldi, L., Selman, M., et al. (2017). Idiopathic Pulmonary Fibrosis. *Nat. Rev. Dis. Primers* 3, 17074. doi:10.1038/nrdp.2017.74

McDonough, J. E., Ahangari, F., Li, Q., Jain, S., Verleden, S. E., Herazo-Maya, J., et al. (2019). Transcriptional Regulatory Model of Fibrosis Progression in the Human Lung. *JCI Insight* 4 (22), e131597. doi:10.1172/jci.insight.131597

Morse, C., Tabib, T., Sembrat, J., Buschur, K. L., Bittar, H. T., Valenzi, E., et al. (2019). Proliferating SPP1/MERTK-Expressing Macrophages in Idiopathic Pulmonary Fibrosis. *Eur. Respir. J.* 54 (2), 1802441. doi:10.1183/13993003.02441-2018

Mushiroda, T., Wattanapokayakit, S., Takahashi, A., Nukiwa, T., Kudoh, S., Ogura, T., et al. (2008). A Genome-wide Association Study Identifies an Association of a Common Variant in TERT with Susceptibility to Idiopathic Pulmonary Fibrosis. *J. Med. Genet.* 45 (10), 654–656. doi:10.1136/jmg.2008.057356

Nardes, F., Araújo, A. P. Q. C., and Ribeiro, M. G. (2012). Mental Retardation in Duchenne Muscular Dystrophy. *J. Pediatr. (Rio J.* 88 (1), 6–16. doi:10.2223/jped.2148

Paplińska-Goryca, M., Goryca, K., Misiukiewicz-Stępień, P., Nejman-Gryz, P., Proboszcz, M., Górska, K., et al. (2019). mRNA Expression Profile of Bronchoalveolar Lavage Fluid Cells from Patients with Idiopathic Pulmonary Fibrosis and Sarcoidosis. *Eur. J. Clin. Invest.* 49 (9), e13153. doi:10.1111/eci.13153

Plantier, L., Cazes, A., Dinh-Xuan, A. T., Bancal, C., Marchand-Adam, S., and Crestani, B. (2018). Physiology of the Lung in Idiopathic Pulmonary Fibrosis. *Eur. Respir. Rev.* 27 (147), 170062. doi:10.1183/16000617.0062-2017

Raghu, G., Remy-Jardin, M., Myers, J. L., Richeldi, L., Ryerson, C. J., Lederer, D. J., et al. (2018). Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *Am. J. Respir. Crit. Care Med.* 198 (5), e44–e68. doi:10.1164/rccm.201807-1255st

Richards, T. J., Kaminski, N., Baribaud, F., Flavin, S., Brodmerkel, C., Horowitz, D., et al. (2012). Peripheral Blood Proteins Predict Mortality in Idiopathic Pulmonary Fibrosis. *Am. J. Respir. Crit. Care Med.* 185 (1), 67–76. doi:10.1164/rccm.201101-0058oc

Schoenheit, G., Becattelli, I., and Cohen, A. H. (2011). Living with Idiopathic Pulmonary Fibrosis. *Chron. Respir. Dis.* 8 (4), 225–231. doi:10.1177/1479972311416382

Sironi, M., Cagliani, R., Pozzoli, U., Bardoni, A., Comi, G. P., Giorda, R., et al. (2002). The Dystrophin Gene Is Alternatively Spliced throughout its Coding Sequence. *FEBS Lett.* 517 (1-3), 163–166. doi:10.1016/s0014-5793(02)02613-3

Sivakumar, P., Thompson, J. R., Ammar, R., Porteous, M., McCoubrey, C., Cantu, E., 3rd, et al. (2019). RNA Sequencing of Transplant-Stage Idiopathic Pulmonary Fibrosis Lung Reveals Unique Pathway Regulation. *ERJ Open Res.* 5 (3), 00117–2019. doi:10.1183/23120541.00117-2019

Steele, M. P., Luna, L. G., Coldren, C. D., Murphy, E., Hennessy, C. E., Heinz, D., et al. (2015). Relationship between Gene Expression and Lung Function in Idiopathic Interstitial Pneumonias. *BMC Genomics* 16, 869. doi:10.1186/s12864-015-2102-3

Thannickal, V. J., Zhou, Y., Gaggar, A., and Duncan, S. R. (2014). Fibrosis: Ultimate and Proximate Causes. *J. Clin. Invest.* 124 (11), 4673–4677. doi:10.1172/jci74368

Tuffery-Giraud, S., Miro, J., Koenig, M., and Claustres, M. (2017). Normal and Altered Pre-mRNA Processing in the DMD Gene. *Hum. Genet.* 136 (9), 1155–1172. doi:10.1007/s00439-017-1820-9

Vukmirovic, M., Herazo-Maya, J. D., Blackmon, J., Skodric-Trifunovic, V., Jovanovic, D., Pavlovic, S., et al. (2017). Identification and Validation of Differentially Expressed Transcripts by RNA-Sequencing of Formalin-Fixed, Paraffin-Embedded (FFPE) Lung Tissue from Patients with Idiopathic Pulmonary Fibrosis. *BMC Pulm. Med.* 17 (1), 15. doi:10.1186/s12890-016-0356-4

Wang, H., Wang, M., Xiao, K., Zhang, X., Wang, P., Xiao, S., et al. (2019). Bioinformatics Analysis on Differentially Expressed Genes of Alveolar Macrophage in IPF. *Exp. Lung Res.* 45 (9-10), 288–296. doi:10.1080/01902148.2019.1680765

Wynn, T. A. (2011). Integrating Mechanisms of Pulmonary Fibrosis. *J. Exp. Med.* 208 (7), 1339–1350. doi:10.1084/jem.20110551

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 Molecular Biosciences

ORIGINAL RESEARCH
published: 03 February 2022
doi: 10.3389/fmolb.2021.799497

Comprehensive Analysis of the Prognostic and Immunological Role of PAFAH1B in Pan-Cancer

*Yixiao Yuan[1†], Xiulin Jiang[2†], Lin Tang[1†], Juan Wang[1] and Lincan Duan[1]**

[1]Department of Thoracic Surgery, The Third Affiliated Hospital of Kunming Medical University, Kunming, China, [2]Key Laboratory of Animal Models and Human Disease Mechanisms of Chinese Academy of Sciences and Yunnan Province, Kunming Institute of Zoology, Kunming, China

Platelet-activating factor acetylhydrolase 1B3 (PAFAH1B3) plays a critical role in cancer initiation, metastasis, and progression; however, it remains unknown how PAFAH1B3 impacts cancer diagnosis and prognosis or regulates the immune response to different types of cancer. In this study, PAFAH1B3 was elevated in human pan-cancer, and this correlated with greater pathology and poor prognosis, in particular for non-small cell lung cancer (NSCLC) and liver hepatocellular carcinoma (LIHC). In addition, PAFAH1B3 expression was positively associated with tumor mutational burden (TMB), microsatellite instability (MSI), immune cell infiltration, immune-modulatory related gene expression, and diverse cancer drug sensitivity in human cancer. Increased PAFAH1B3 expression correlated with poor overall survival (OS), disease-specific survival (DSS), and progression-free interval (PFI) of NSCLC and LIHC, and has potential as an independent risk factor for overall survival (OS), disease-specific survival (DSS), and progression-free interval (PFI) during LIHC. Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis confirmed that PAFAH1B3 is primarily involved in immune regulation. More importantly, results demonstrated that PAFAH1B3 was upregulated in liver cancer cells lines and that knockdown of this gene significantly inhibited cell proliferation, migration, and invasion in liver hepatocellular carcinoma (LIHC). In summary, this study elucidates the clinical significance and biological function of PAFAH1B3 during liver hepatocellular carcinoma (LIHC) and may serve as a potential biomarker for the diagnosis and prognosis of various cancer types.

Keywords: PAFAH1B3, pan-cancer, immune cell infiltration, drug sensitivity, prognostic biomarker, multi-omics integrative analysis, cell proliferation, cell migration

OPEN ACCESS

Edited by:
Matteo Becatti,
University of Firenze, Italy

Reviewed by:
Giuseppe Jurman,
Bruno Kessler Foundation (FBK), Italy
Longhai Li,
Bozhou People's Hospital, China

****Correspondence:***
Lincan Duan
duanmumuhuosan@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *21 October 2021*
Accepted: *20 December 2021*
Published: *03 February 2022*

Citation:
Yuan Y, Jiang X, Tang L, Wang J and Duan L (2022) Comprehensive Analysis of the Prognostic and Immunological Role of PAFAH1B in Pan-Cancer. Front. Mol. Biosci. 8:799497. doi: 10.3389/fmolb.2021.799497

INTRODUCTION

Cancer affects millions of people each year and poses a substantial societal and economic burden worldwide. Despite available surgical and chemotherapeutic treatment modalities, cancer prognosis often remains poor. Cancer hallmark genes (e.g., BRCA1, CDK1, E2F1, and EGFR) are responsible for the most essential phenotypic characteristics of malignant transformation and progression (Nagy et al., 2021). Thus, there is an urgent need to identify specific molecular targets to improve cancer diagnosis and treatment.

Platelet-activating factor acetylhydrolase 1B3 (PAFAH1B3), one of the catalytic subunits of PAFAH, plays an important role in apoptosis (Monillas et al., 2015), cancer metastasis (Stafforini,

2015), and angiogenesis during cancer (Sun et al., 2015). PAFAH1B3 is involved in diverse cancer-related signaling pathways, including PAF and WNT, and facilitates cancer progression (Manya et al., 1999; Livnat et al., 2010). By modulating tumor-suppressing lipids, PAFAH1B3 promotes cancer cell aggressiveness (Kume and Shimizu, 1997). In addition, Michael et al. showed that PAFAH1B3 may be a potential target for tyrosine kinase inhibitors (TKIs) in breast cancer (BRCA) (Fiedler et al., 2018). PAFAH1B3 was found to play a crucial role in the brain development process. lissencephaly associated mutations will destroy the interaction between PAFAH1B3 and PAFAH1B2, leading to inhibitions in the neuronal migration (Xing et al., 2011). These studies indicate that PAFAH1B3 regulates diverse biological functions in cancer initiation, metastasis, and progression, and may be a promising prognostic and therapeutic biomarker for pan-cancer. However, the specific role of PAFAH1B3 in diagnosis, prognosis, and immune regulation in various types of cancer remains unexplored.

In this study, public databases were used for the first time to show that PAFAH1B3 is highly expressed in diverse cancer types. PAFAH1B3 expression correlates significantly with pathology and poor prognosis and is highly accurate at predicting cancer progression. In addition, PAFAH1B3 expression was positively associated with tumor mutational burden (TMB), microsatellite instability (MSI), immune cell infiltration, immune-modulatory related gene expression, and diverse cancer drug sensitivity in human cancer. Finally, high expression of PAFAH1B3 correlated with poor overall survival (OS), disease-specific survival (DSS), and progression-free interval (PFI) in non-small cell lung cancer (NSCLC) and liver hepatocellular carcinoma (LIHC), and has the potential as an independent risk factor for overall survival (OS), disease-specific survival (DSS), and progression-free interval (PFI) in liver hepatocellular carcinoma (LIHC). Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was used to confirm that PAFAH1B3 is primarily involved in immune regulation. More importantly, results demonstrated that PAFAH1B3 was up-regulated in liver hepatocellular carcinoma (LIHC) cells lines. Knockdown of PAFAH1B3 significantly inhibited cell proliferation, migration, and invasion in liver hepatocellular carcinoma (LIHC). In summary, PAFAH1B3 is a potential biomarker for diagnosis and prognosis in different cancer types and a promising molecular target for LIHC.

MATERIALS AND METHODS

Analysis of the Expression of PAFAH1B3 in Pan-Cancer

We utilized TIMER (https://cistrome.shinyapps.io/timer/) (Li et al., 2017), TCGA (https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga), Genotype-Tissue Expression (GTEx) database, ualcan database (http://ualcan.path.uab.edu/) (Chandrashekar et al., 2017), and CCLE database (https://portals.broadinstitute.org/ccle/) (Ghandi et al., 2019) to examine the expression of PAFAH1B3 in pan-cancer tissue and cancer cells lines (ns, $p \geq 0.05$; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

Analysis the Prognosis and Clinical Information of PAFAH1B3 in Pan-Cancer

We employed the GEPIA databases (http://gepia.cancer-pku.cn/) and prognoscan databases (http://dna00.bio.kyutech.ac.jp/PrognoScan/index.html) (Mizuno et al., 2009; Tang et al., 2017) to examine the OS and RFS of PAFAH1B3 in pan-cancer; additionally, the correlation between the pathology stage and PAFAH1B3 expression was analysis by GEPIA, the correlation between the tumor stage and PAFAH1B3 expression was analysis by GEPIA (ns, $p \geq 0.05$; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

Analysis of the Gene Mutation of PAFAH1B3 in Pan-Cancer

The gene mutation information of PAFAH1B3 in pan-cancer was analyzed by Cerami et al. (2012).

Analysis of the Function of PAFAH1B3 in Pan-Cancer

We utilized the cbioportal database (https://www.cbioportal.org/) to analyze the co-expression genes in pan-cancer. KEGG enrichment analyses was analysed by the cluster Profiler package and using ggplot2 package for visualization (Yu et al., 2012; Ito and Murphy, 2013).

Analysis of the Immunological Functions of PAFAH1B3 in Pan-Cancer

We employed the TIMER (https://cistrome.shinyapps.io/timer/) and XCELL tools (https://xcell.ucsf.edu/) to analyze the immunological roles of PAFAH1B3 (Li et al., 2017; Aran et al., 2017), including the correlation between the diverse immune cells and immune regulator. The TISIDB (http://cis.hku.hk/TISIDB/) was utilized to analysis the expression of PAFAH1B3 in molecular subtypes and immune subtypes of diverse cancers (Ru et al., 2019). The TMB and MSI scores were obtained from TCGA. Correlation analysis between the PAFAH1B3 expression and TMB or MSI was performed using spearman's methods (ns, $p \geq 0.05$; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

Correlation Between PAFAH1B3 Expression and Cancer Drug Sensitivity

We utilized GDSC and CTRP databases to analyze the correlation between PAFAH1B3 expression and drug sensitivity (Basu et al., 2013; Yang et al., 2013).

Cell Culture

The LO2 cell line was purchased from the cell bank of Kunming Institute of Zoology, and cultured in DMEM media (Lonza, CC-3170). Liver cancer cell lines, including HepG2, Hu7, and SMCC-

7721, were purchased from Cobioer, China, with STR document, HepG2, Hu7, and SMCC-7721 cells were all cultured in DMEM medium (Corning) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. The siRNA for PAFAH1B3 were synthesized by RIBOBIO, and a scrambled siRNA was synthesized as a negative control. Transfection was performed using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Total RNA was collected 48 h after transfection.

Quantitative Real-Time PCR

The qRT-PCR assay was performed as documented (Jiang et al., 2018). The primer sequences are as follows: PAFAH1B3-F: ACA TCCGGCCCAAGATTGTG, PAFAH1B3-R: GGGCTGTCG CTCATTCACC, PAFAH1B1-F: TCTTGGTCAGAAACGAGA CCC, PAFAH1B1-R: GTGGTCGAATGAAATGTCCTGTA. PAFAH1B2-F: CAAACCCAGCAGCTATTCCG, PAFAH1B2-R: GAACAGTACATCAGGCTCTTTGT. BRCA1-F: GCTCGT GGAAGATTTCGGTGT, BRCA1-R: TCATCAATCACGGAC GTATCATC. CDK1-F: ACCGAAGGGAGAACGACGAA, CDK1-R: GAACGCTTTGAACTTCCCGAT. E2F1-F: ACG CTATGAGACCTCACTGAA, E2F1-R: TCCTGGGTCAAC CCCTCAAG. β-actin-F: CTTCGCGGGCGACGAT, β-actin-R: CCATAGGAATCCTTCTGACC. The expression quantification was obtained with the $2^{-\Delta\Delta CT}$ method (ns, $p \geq 0.05$; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

Cell Proliferation and Colony Formation Assays

Cell proliferation, colony formation, and tumor sphere formation assay were performed as previously documented (Xiong et al., 2021). Briefly, for cell proliferation assay, indicated cells were plated into 12-well plates at a density of 1.5×10^4, and the cell numbers were subsequently counted each day using an automatic cell analyzer countstar (Shanghai Ruiyu Biotech Co., China, IC 1000). For the colony formation assay, indicated cells were seeded in a 6-well plate (China, NEST, Cat. 703001) with 600 cells per well supplemented with 2 ml cell culture medium, and the cell culture medium was changed every 3 days for 2–3 weeks, and then indicated cells were fixed with 4% PFA and stained with 0.5% crystal violet (ns, $p \geq 0.05$; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

Cell Migration and Invasion Assays

Cell migration assays was performed as previously documented (Xiong et al., 2021). Briefly, to produce a wound, the monolayer cells in a 6-well plate were scraped in a straight line with pipette tips. The plate was then washed with PBS to remove detached cells. Photographs of the scratch were taken at indicated time points using Nikon inverted microscope (Ti-S) (ns, $p \geq 0.05$; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

Statistical Analysis

Analysis the PAFAH1B3 expression pan-cancer was estimated using t-tests. The correlations between clinicopathological characteristics and PAFAH1B3 expression were evaluated using the Chi-squared test, Fisher exact test, Kruskal–Wallis (KW) test, Wilcoxon signed-rank test, Wilcoxon rank sum test, and logistic regression. Through univariate and multivariate analysis combined with Cox logistic regression models, other clinical factors impacting the survival and the PAFAH1B3 expression level were found. Kaplan-Meier analysis was employed to examine the survival time of patients stratified according to high or low level of the PAFAH1B3 expression. For all figures, *, **, *** indicate $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

RESULTS

Analysis of the Expression and Prognosis Values of PAFAH1B3 in Pan-Cancer

Tumor Immune Estimation Resource (TIMER) database analysis was used to assess PAFAH1B3 in diverse cancers, and was shown to be significantly elevated in adenoid cystic carcinoma (ACC), bladder urothelial carcinoma (BLCA), cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), head and neck squamous cell carcinoma (HNSC), kidney renal papillary cell carcinoma (KIRP), LIHC, lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), prostrate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid cancer (THCA), and uterine corpus endometrial carcinoma (UCEC). Interestingly, lower expression of PAFAH1B3 was observed in KICH and KIRC (**Figure 1A**). Given that some cancers lack normal tissue data in TCGA databases, PAFAH1B3 expression was also assessed in pan-cancers using the TCGA/GTEx databases. PAFAH1B3 expression was higher in ACC, BLCA, BRCA, cervical squamous cell carcinoma (CESC), COAD, CHOL, diffuse large B-cell lymphoma (DLBCL), ESCA, glioblastoma (GBM), HNSC, KIRP, low-grade glioma (LGG), LIHC, LUAD, LUSC, ovarian cancer (OV), pancreatic adenocarcinoma (PAAD), PRAD, READ, skin cutaneous melanoma (SKCM), STAD, tenosynovial giant cell tumor (TGCT), THCA, thymus cancer (THYM), UCEC, and uterine carcinosarcoma (UCS) (**Figure 1B**). In addition, Cancer Cell Line Encyclopedia (CCLE) databases analysis showed that PAFAH1B3 was overexpressed in many cancer cell lines (**Figure 1C**). To verify these results, UALCAN database analysis was used to assess PAFAH1B3 protein expression in human cancers. PAFAH1B3 was significantly elevated in breast cancer, colon cancer, ovarian cancer, clear cell renal cell carcinoma (RCC), and UCEC (**Figure 1D**). Overall, these results showed that PAFAH1B3 was upregulated in many human cancer types.

Since PAFAH1B3 expression was associated with the pathology of many cancer types, the ability of PAFAH1B3 to prognose pan-cancer was explored. OS, DSS, and PFI analysis of various cancer types showed that increased PAFAH1B3 expression correlated with poor overall survival for ACC, LIHC, LUAD, mesothelioma (MESO), sarcoma (SARC), and SKCM (**Figure 2A**), and poor DSS in BLCA, SARC, LIHC, LUAD, MESO, and SKCM (**Figure 2B**).

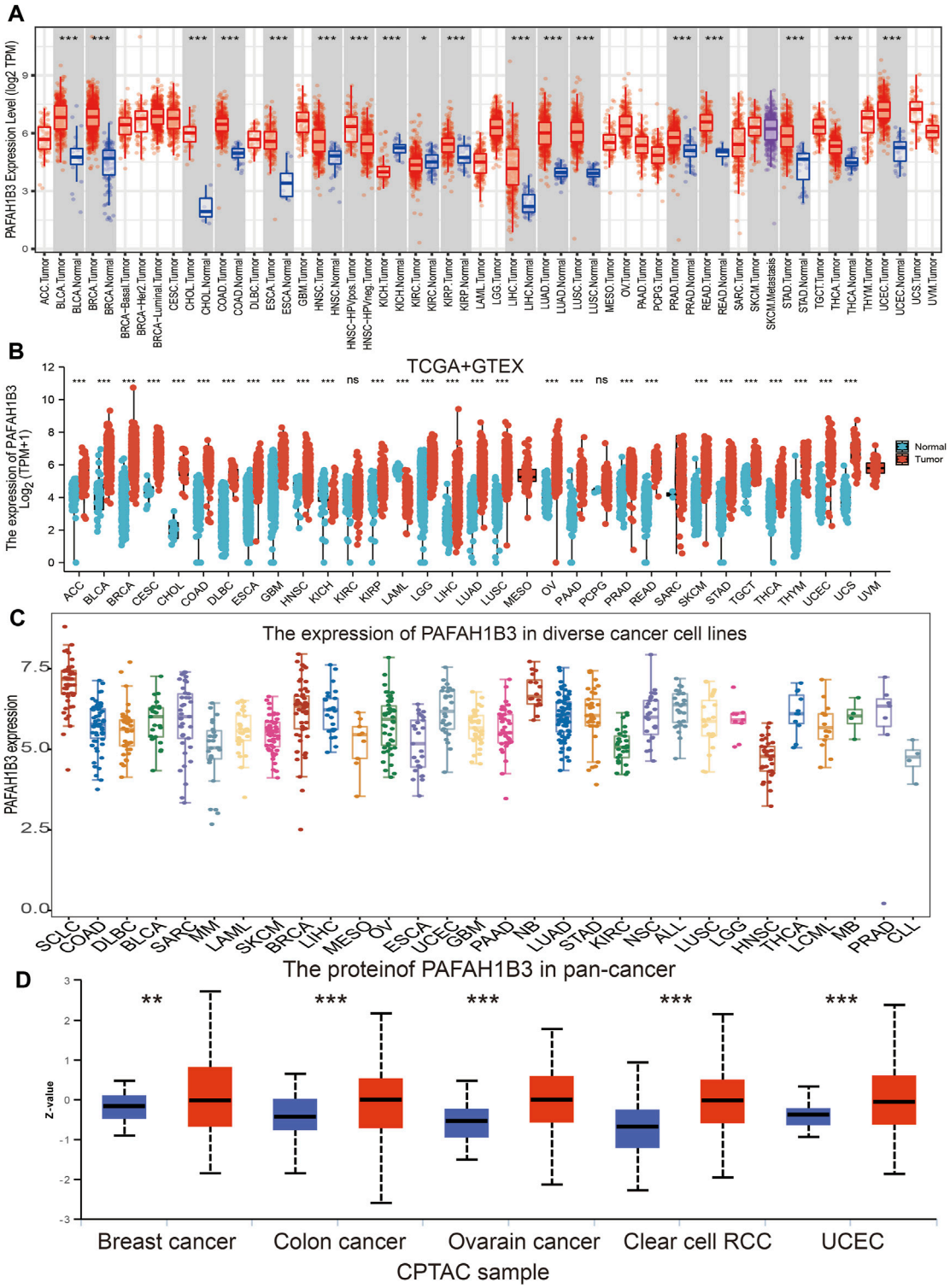


FIGURE 1 | Expression analysis for PAFAH1B3 in human cancers. **(A)** The expression of PAFAH1B3 in pan-cancer analysis by TIMER database. **(B)** The expression of PAFAH1B3 in pan-cancer analysis by TCGA/GTEx database. **(C)** The expression of PAFAH1B3 in diverse cancer cells lines examine by the CCLE database. **(D)** The protein expression of PAFAH1B3 in various cancers analysis by Ualcan database.

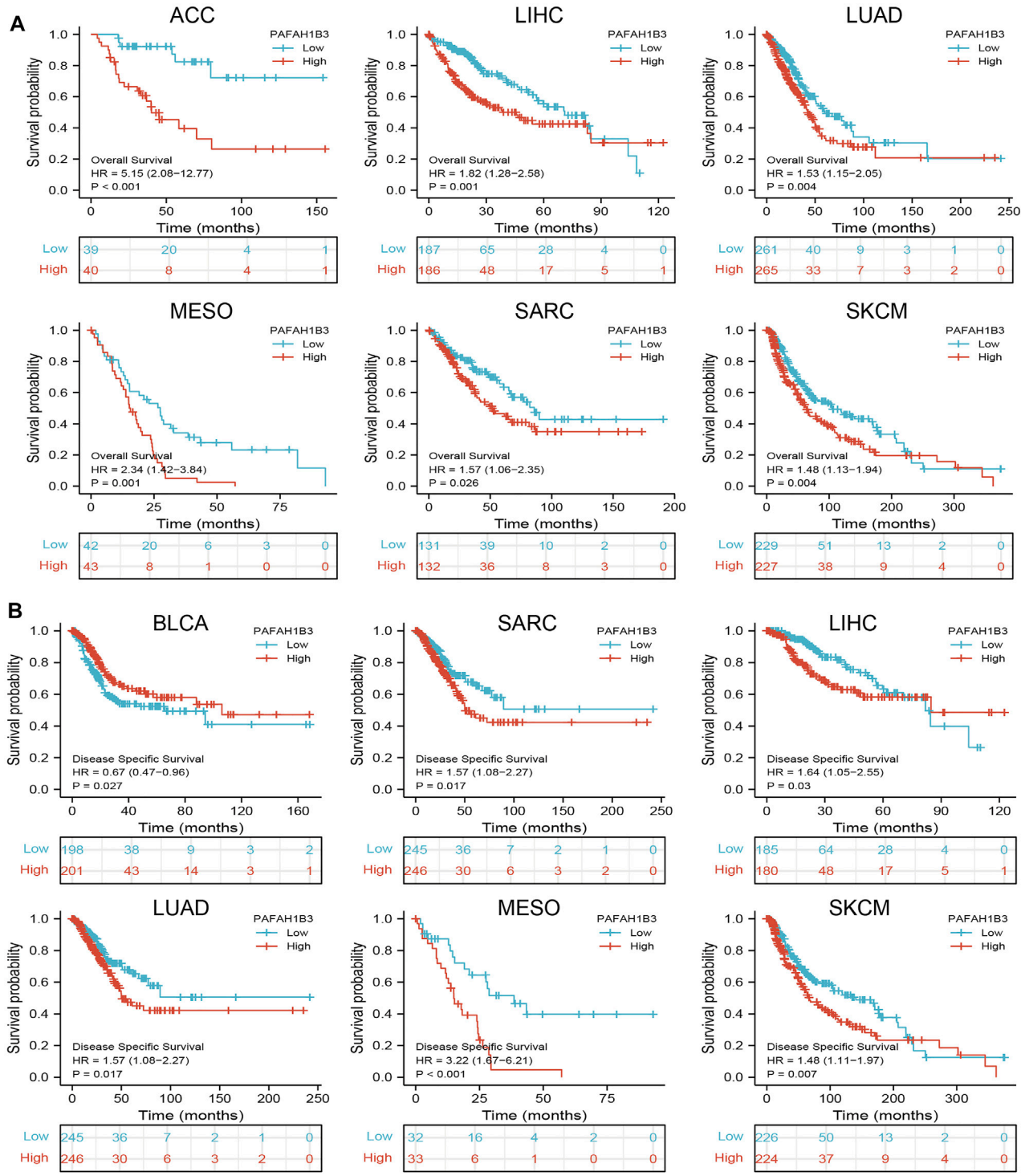


FIGURE 2 | Analysis of the prognosis values for PAFAH1B3 in human cancers. **(A)** The OS of PAFAH1B3 in diverse cancers was examined by the TCGA database. **(B)** The DSS for PAFAH1B3 in pan-cancer examine by TCGA database.

Since previous results have shown that PAFAH1B3 expression correlates with prognosis for a wide range of cancers, this study assessed whether PAFAH1B3 may act as a detection index for cancer diagnosis. Receiver operating characteristic (ROC) curve analysis was used to assess the diagnostic value of PAFAH1B3 in various human cancers and found that it had moderate accuracy (AUC > 0.75) in predicting BLCA, BRCA, CHOL, COAD, ESCA, GBM, HNSC, KICH, KIRP, LAML, LGG, LIHC, LUAD, LUSC,

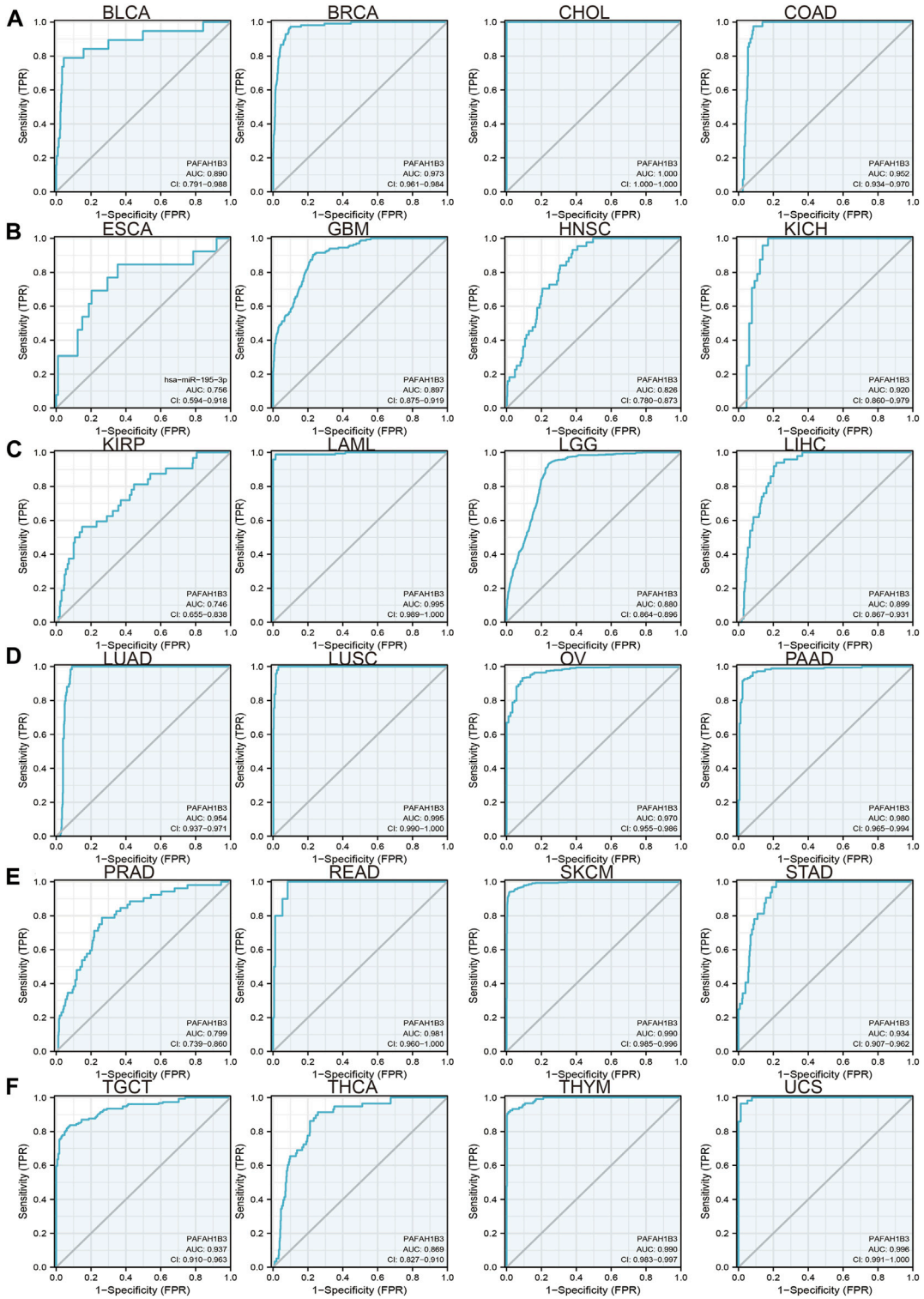


FIGURE 3 | ROC curve analyses and AUC values for PAFAH1B3 in diverse cancer. **(A–D)** ROC curve analysis and AUC values for PAFAH1B3 in BLCA, BRCA, CHOL, and COAD. **(B)** ROC curve analysis and AUC values for PAFAH1B3 in ESCA, GBM, HNSC, and KICH. **(C)** ROC curve analysis and AUC values for PAFAH1B3 in KIRP, LAML, LGG, and LIHC. **(D)** ROC curve analysis and AUC values for PAFAH1B3 in LUAD, LUSC, OV, and PAAD. **(E)** ROC curve analysis and AUC values for PAFAH1B3 in PRAD, READ, STAD, and SKCM. **(F)** ROC curve analysis and AUC values for PAFAH1B3 in TGCT, THCA, THYM, and THYM.

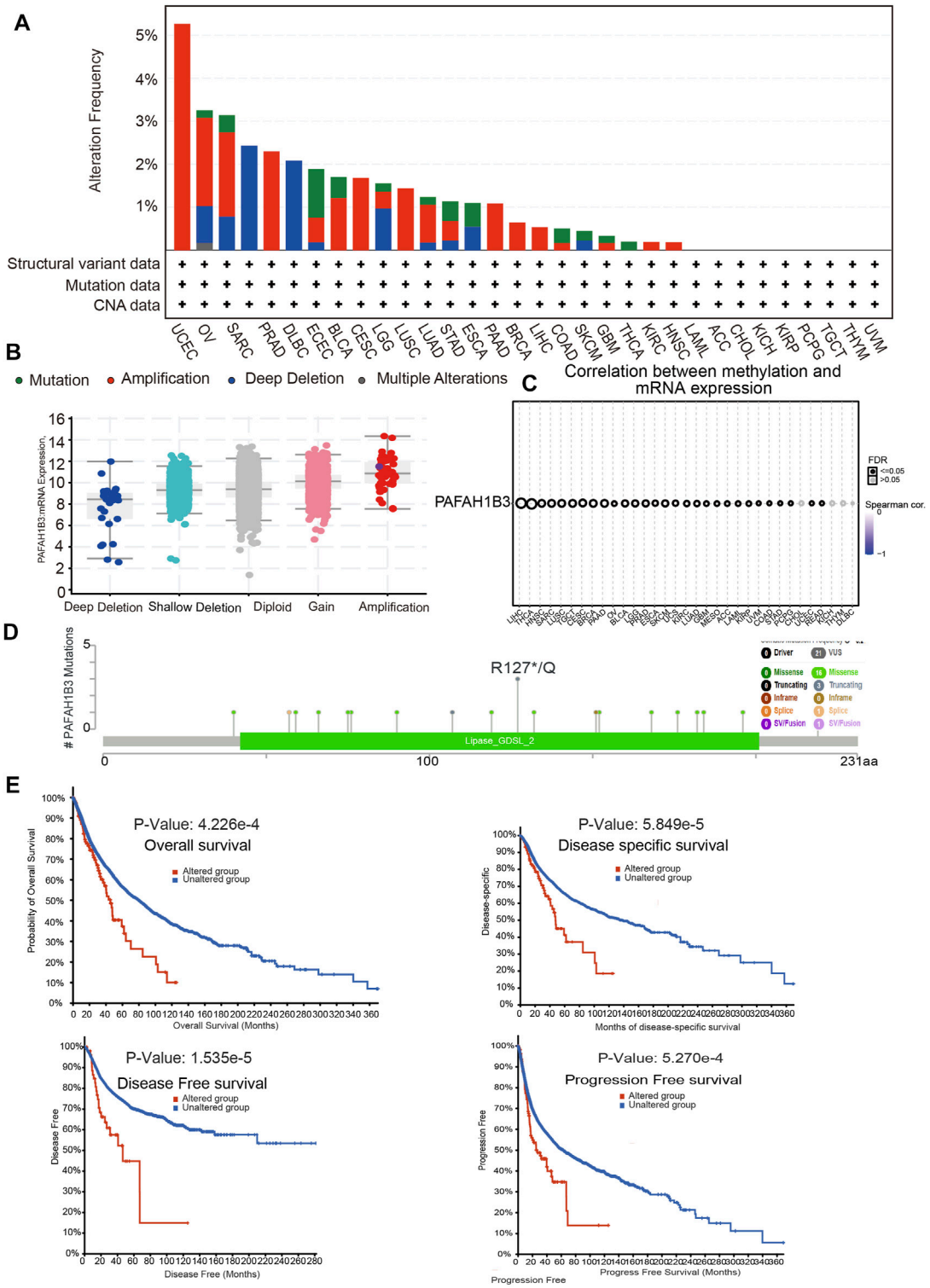


FIGURE 4 | PAFAH1B3 mutation landscape in various cancer types. **(A)** PAFAH1B3 mutation level in pan-cancer examined by the cBioPortal database. **(B)** PAFAH1B3 mutation frequency in pan-cancer examined by the cBioPortal database. **(C)** The DNA methylation of PAFAH1B3 in pan-cancer **(D)** Mutation diagram of PAFAH1B3 in pan-cancer was examined by the cBioPortal database. **(E)** The correlation between PAFAH1B3 mutation and poor prognosis in patients.

OV, PAAD, PRAD, THCA, THYM, and UCS (**Figures 3A–F**) and high accuracy (AUC > 0.90) in predicting BRCA, CHOL, COAD, KICH, LAML, LUAD, LUSC, OV, PAAD, READ, SKCM, STAD, TGCT, THCA, THYM, and UCS. These results confirm that PAFAH1B3 has the potential to act as a detection index for the diagnosis of many cancer types with high sensitivity and specificity.

Analysis of the Mutation Landscape of PAFAH1B3 in Various Cancer Types

Alterations in PAFAH1B3 copy numbers were assessed using the cBioPortal database. In various human cancers, the mutation frequency was higher in UCEC, OV, SARC, and PRAD than in other cancers (**Figure 4A**). Amplification was the most common type of alteration, followed by shallow depletion and diploid (**Figure 4B**). DNA methylation analysis showed that PAFAH1B3 expression was negatively associated with DNA methylation in LIHC, THCA, HNSC, and SARC (**Figure 4C**). To examine the PAFAH1B3 mutation landscape in various cancer types, 21 VUS, 16 missense sites, three truncation sites, one splice, and one fusion situated between amino acids 0 and 231 were identified in PAFAH1B3 using the cBioPortal database (**Figure 4D**). PAFAH1B3 genetic alterations were associated with overall survival, disease-specific survival, and progression-free interval in cancer patients (**Figure 4E**). These results confirm that PAFAH1B3 genetic alterations affect PAFAH1B3 expression and prognostic ability.

Correlations Between PAFAH1B3 Expression and Immune and Molecular Subtypes, Tumor Mutational Burden, and Microsatellite Instability in Pan-Cancer

The correlation between PAFAH1B3 expression and development of different immune and molecular subtypes in pan-cancer was assessed using the TISIDB database. Results showed that PAFAH1B3 has different expression patterns in pan-cancer (**Supplementary Figures S1A–D**). For example, while PAFAH1B3 was highly expressed in C2 of LIHC, it was poorly expressed in C3. PAFAH1B3 was uniquely expressed in different molecular subtypes of cancer (**Supplementary Figures S2A–D**). For example, PAFAH1B3 expression was higher in the C2C-CIMP subtype of LIHC. These results confirm that PAFAH1B3 expression is associated with different immune and molecular subtypes of cancer.

Correlations Between PAFAH1B3 Expression and Tumor Mutational Burden and Microsatellite Instability in Pan-Cancer

Tumor Mutational Burden (TMB), the number of DNA mutations in cancer, has emerged as a sensitive and specific biomarker in response to immune checkpoint inhibitors (Addeo et al., 2021). PAFAH1B3 expression was positively associated with the TMB in MESO, LUAD, STAD, PAAD, ACC, LGG, DLBC, UVM, and PRAD ($r > 0.2$, $p < 0.01$), and negatively associated with the TMB in THYM (**Supplementary Figure S3A**). Microsatellite instability (MSI) represents a hypermutable state of DNA sequences caused by the lack of DNA mismatch repair activity (Boland and Goel, 2010). PAFAH1B3 expression was positively correlated with the MSI in DLBC, STAD, PAAD, MESO, ESCA, UCEC, and LIHC ($r > 0.15$, $p < 0.01$), and negatively correlated with the MSI in TGCT, LAML, COAD, UCS, and READ (**Supplementary Figure S3B**).

Correlation Between PAFAH1B3 Expression and Immune Infiltration, Drug Sensitivity

Immune cell infiltration is important to cancer progression. TIMER results showed that PAFAH1B3 expression correlated with $CD8^+$ T cell abundance in 27 cancers, $CD4^+$ T cell abundance in 28 cancers, neutrophil abundance in 30 cancers, dendritic cell (DC) abundance in 30 cancers, macrophage abundance in 27 cancers, and B cell abundance in 29 cancers (**Figure 5A**). To verify these results, xCell was used to assess the correlation between PAFAH1B3 expression and immune cell infiltration in many cancer types. Expression correlated positively with 38 immune cell types in 25 cancers and correlated negatively with 32 immune cell types in two cancers (**Figure 5B**). Findings indicate that PAFAH1B3 expression is significantly correlated with immune cell infiltration during human cancer.

To further determine the relationship between PAFAH1B3 and the tumor microenvironment, TCGA analysis was used to measure the correlation between PAFAH1B3 and immune checkpoint-related genes. PAFAH1B3 expression was positively associated with immune checkpoint-related genes such as CD274, CTLA4, HAVCR2, LAG3, PDCD1, PDCD1LG2, SIGLEC15, and TIGIT, in 31 cancers (**Supplementary Figure S4**). The TISIDB tool showed that PAFAH1B3 expression was positively associated with genes for 28 tumor-infiltrating lymphocytes, 45 immune-stimulators, 24 immune-inhibitors, 41 chemokines, 18 receptors, and 21 MHCs in pan-cancer (**Supplementary Figures S5A,B**). These findings indicate that PAFAH1B3 plays an important role in regulating the immune response to human cancer.

The correlation between PAFAH1B3 expression and sensitivity to different drugs was assessed using cancer cell lines from the Genomics of Drug Sensitivity in Cancer (GDSC) database and the Cancer Therapeutics Response Portal (CTRP) database. In the GDSC database, PAFAH1B3 expression was positively correlated with Z-LLNle-CHO, XMD8-85, CGP-60474, A-770041, dasatinib, bortezomib, AZ628 and JW-7-52-1 ($r > 0.17$, $p < 0.0001$) sensitivity, and negatively correlated with Navitoclax, GSK1070916, WZ3105, and SB52334 ($r < -0.10$, $p < 0.0001$) sensitivity (**Supplementary Figure S6A**). In the CTRP database, PAFAH1B3 expression was positively correlated with trametinib, dasatinib, BRD-K44224150, and BRD-A05715709 ($r > 0.10$, $p < 0.0001$) sensitivity, and negatively associated with GSK-J4, SR8278, BRD-K48334597, MI-2, HBX-41108, pifithrin-mu, cerulenin, SB-525334, valdecoxib, belinostat, PX-12, and skepinone ($r < -0.20$, $p <$

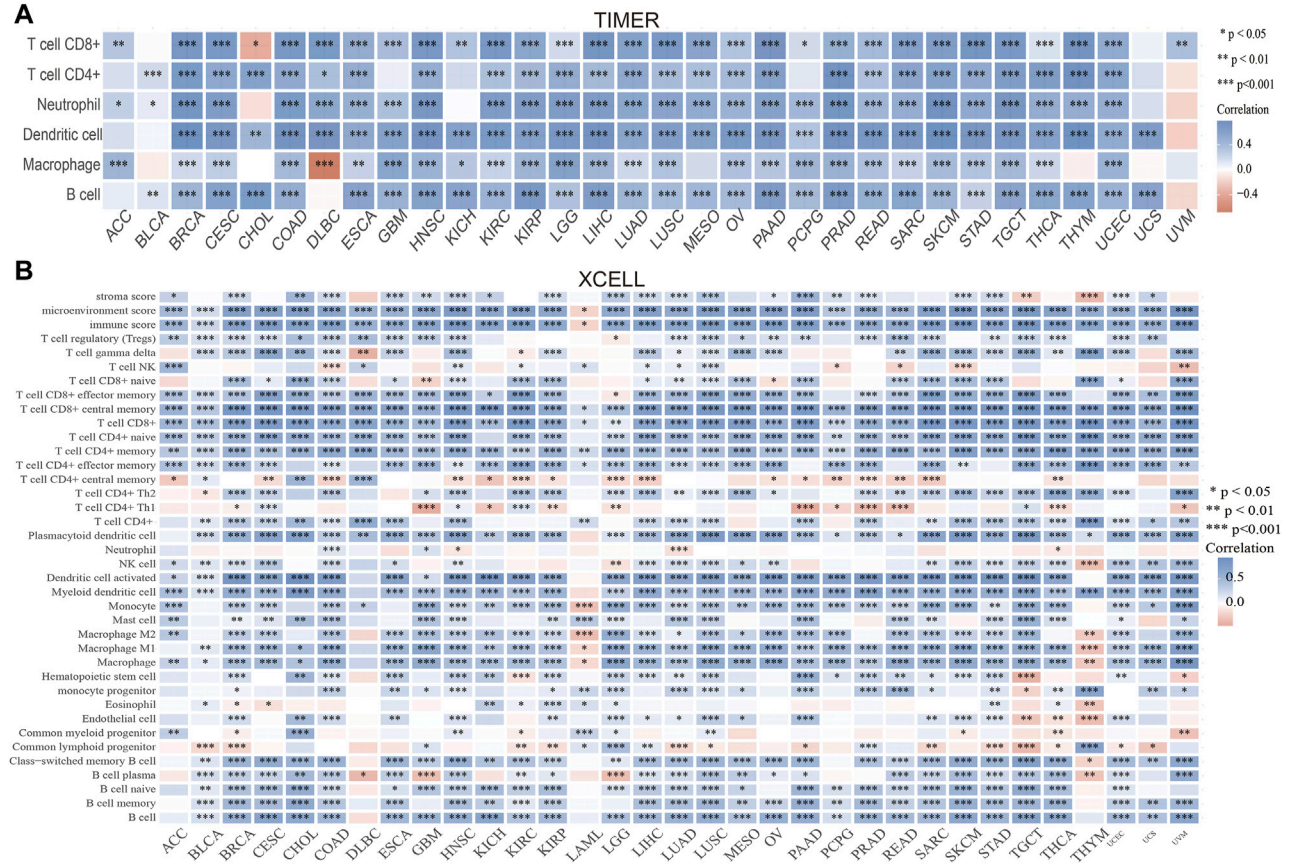


FIGURE 5 | Analysis of the correlation between the PAFAH1B3 expression and immune cells infiltration. **(A)** The correlation between PAFAH1B3 expression and immune cells infiltration in pan-cancer was examined by the TIMER database. **(B)** The correlation between PAFAH1B3 expression and diverse immune cells infiltration in pan-cancer examine by Xcell database. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

0.0001) sensitivity (**Supplementary Figure S6B**). These results suggest that PAFAH1B3 is significantly associated with diverse drug sensitivity in different cancer cell lines and may be a promising therapeutic target for cancer.

Correlation Between PAFAH1B3 and Clinical Characteristics in NSCLC and LIHC

Comprehensive bioinformatics was performed to assess the correlation between PAFAH1B3 overexpression and NSCLC pathology. Overexpressed PAFAH1B3 in NSCLC (**Figure 6A**) was strongly associated with pathologic stage, TNM stage, residual tumor, and outcome of primary therapy (**Figure 6B**). High expression of PAFAH1B3 had a worse OS for most clinical and demographic NSCLC subgroups including pathologic stage, TN stage, residual tumor, gender, age, smoking status, and race (**Figure 6C**). These results confirm that PAFAH1B3 plays a critical role in the progression of NSCLC.

The correlation between PAFAH1B3 expression and pathology of LIHC was also assessed. High PAFAH1B3 expression was significantly associated with histologic stage, tumor status, pathologic stage, TNM stage, residual tumor, vascular invasion, race, BMI, gender, age, weight, and height (**Figures 7A–D**). In addition, overexpression of PAFAH1B3 had a worse OS in most clinical and demographic subgroups of LIHC, including pathologic stage, histologic stage, TNM stage, residual tumor, gender, age, adjacent hepatic tissue inflammation, height, and weight (**Figures 8A–D** and **Table 1**). Univariate and multivariate Cox regression analyses showed that TM stage, pathologic stage, tumor status, and PAFAH1B3 expression were significantly associated with OS (**Table 2**). These results confirm that PAFAH1B3 plays a critical role in LIHC progression.

A nomogram was created to integrate PAFAH1B3 as a LIHC biomarker, including the TNM stages, tumor status, histologic stage to predict OS, DSS, and PFI. The C-indices of OS, DSS, and PFI were 0.680, 0.871, and 0.808, respectively. The calibration curves all presented desirable predictions for the three nomograms for 1-, 3-, and 5-years clinical outcomes (**Figures 9A–F**). Thus, this nomogram may be a model for predicting LIHC survival with PAFAH1B3 than an individual prognostic factor.

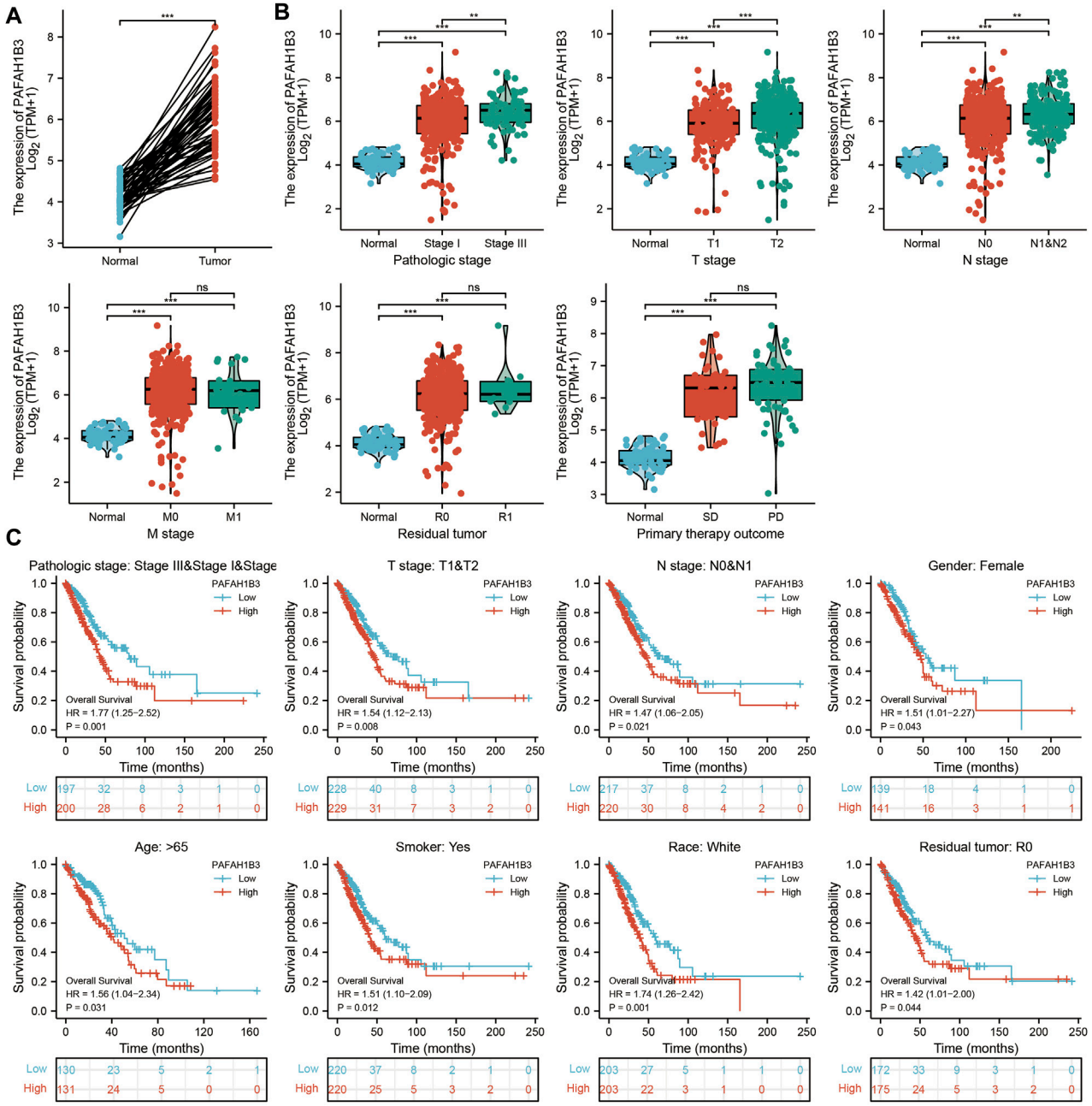


FIGURE 6 | PAFAH1B3 Is Correlated With Clinical Characteristics and Prognosis in NSCLC. **(A)** The expression of PAFAH1B3 in NSCLC. **(B)** The correlation between PAFAH1B3 expression and Clinical Characteristics in NSCLC, including the pathologic stage, TNM stage, residual tumor, and primary therapy outcome. **(C)** The prognosis of PAFAH1B3 in NSCLC.

Analysis of the Biological Functions of PAFAH1B3 in LIHC

To perform Gene Ontology (GO) and KEGG enrichment analysis, 1,013 DEGs (lncRNA and mRNA) were acquired with threshold values of |log2 fold-change (FC)| > 2.0 and adjusted *p*-value < 0.01, including 813 upregulated genes and 200 downregulated genes (**Figures 10A–D**). For the cellular components of GO term, the DEGs mainly involved those on the external side of the plasma membrane, including MHC class II protein complex, immunological synapse, MHC protein complex, secretory granule membrane, clathrin-coated endocytic vesicle membrane, endocytic vesicle, tertiary granule membrane, tertiary granule, clathrin-coated endocytic vesicle, endocytic vesicle membrane, ficolin-1-rich granule, an integral component of the luminal side of endoplasmic reticulum

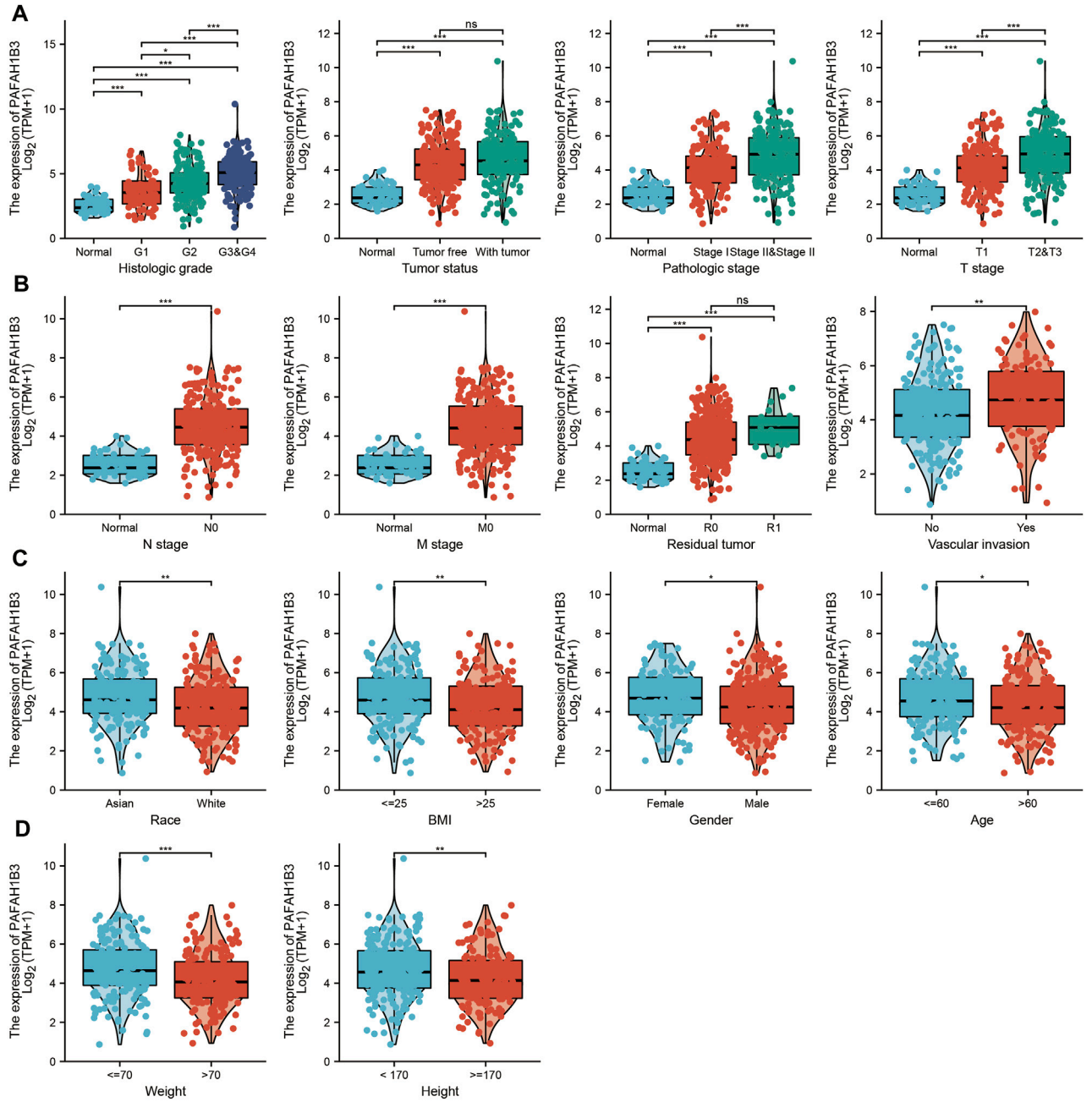


FIGURE 7 | PAFAH1B3 Is Correlated With Clinical Characteristics in LIHC. **(A–D)** The correlation between PAFAH1B3 expression and Clinical Characteristics in LIHC, including the histologic stage, Tumor status, pathologic stage, TNM stage, residual tumor, vascular invasion, race, BMI, Gender, age, weight, and height.

membrane, luminal side of the endoplasmic reticulum membrane, colin-1-rich granule membrane, clathrin-coated vesicle membrane, and lysosomal membrane (**Figure 10E**). For the biology process of GO term, the DEGs mainly involved T cell activation, lymphocyte differentiation, regulation of lymphocyte activation, leukocyte cell-cell adhesion, regulation of T cell activation, regulation of leukocyte cell-cell adhesion, positive regulation of cell activation, T cell differentiation, positive regulation of leukocyte cell-cell adhesion, positive regulation of leukocyte activation, regulation of lymphocyte proliferation, regulation of mononuclear cell proliferation, regulation of leukocyte proliferation, and positive regulation of T cell activation (**Figure 10F**). For the molecular function of GO term, the DEGs mainly involved cytokine receptor activity, carbohydrate-binding, cytokine binding, MHC protein binding, MHC protein complex binding, MHC class II receptor activity, G protein-coupled chemoattractant receptor activity, chemokine receptor activity, C-C chemokine receptor

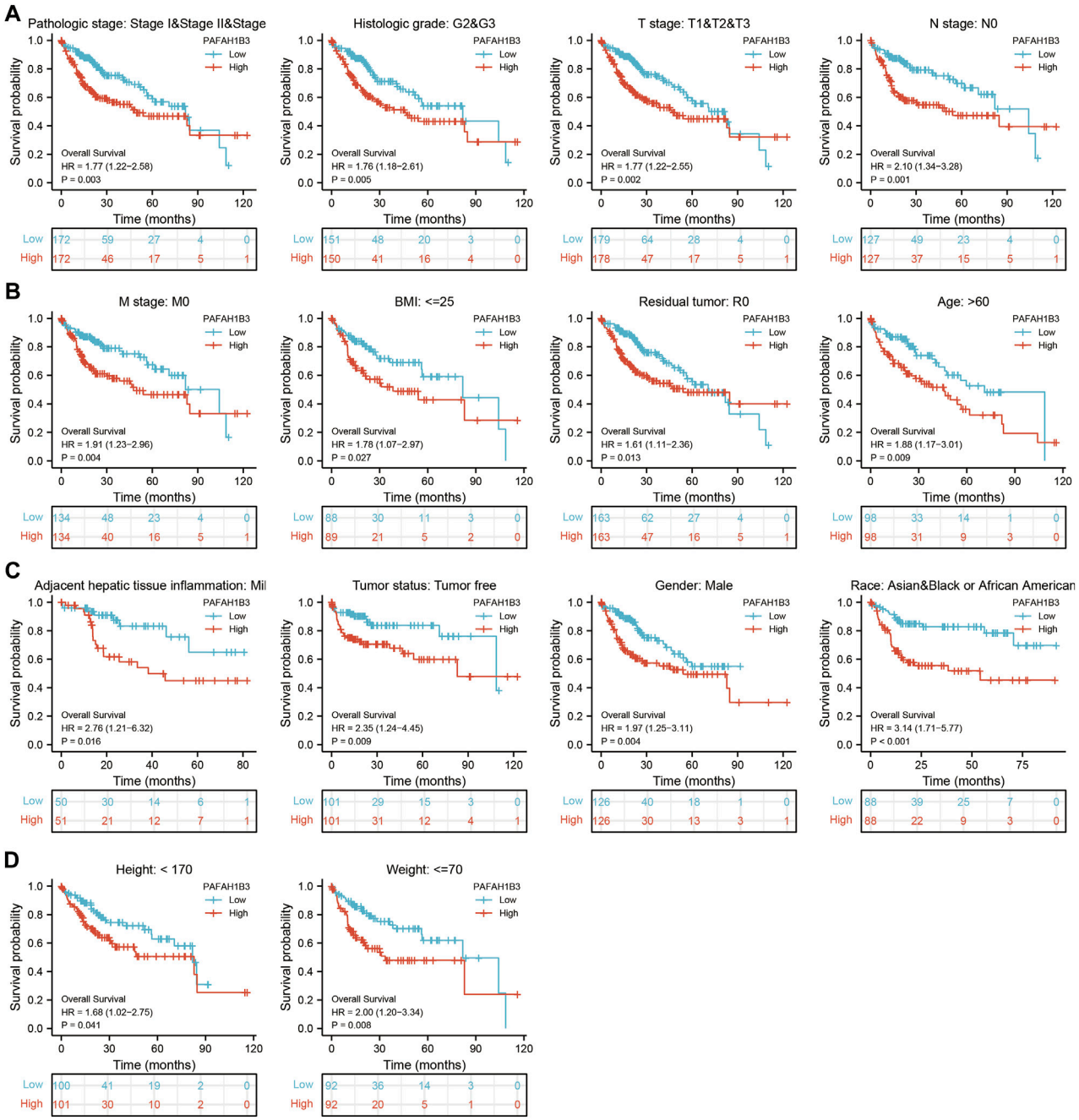


FIGURE 8 | PAFAH1B3 Is Correlated With Prognosis in LIHC. **(A–D)** The correlation between PAFAH1B3 and OS in different clinical subgroups of LIHC, including pathologic stage, histologic stage, TNM stage, residual tumor, gender, age, adjacent hepatic tissue inflammation, height, and weight.

activity, chemokine binding, C-C chemokine binding, MHC class I protein binding, pattern recognition receptor activity, coreceptor activity, and immunoglobulin binding (**Figure 10G**). KEGG enrichment analysis showed that the DEGs were mainly involved in hematopoietic cell lineage, Th1 and Th2 cell differentiation, Th17 cell differentiation, cell adhesion, the intestinal immune network for IgA production, allograft rejection, *Staphylococcus aureus* infection, graft-versus-host disease, type I diabetes mellitus, *Leishmaniasis* infection, autoimmune thyroid disease, B cell receptor signaling, primary immunodeficiency, T cell receptor signaling, tuberculosis, inflammatory bowel disease, natural killer cell-mediated cytotoxicity, and chemokine signaling (**Figure 10H**).

GSEA enrichment of the DEGs was also assessed and results indicated that these genes mainly participated in JAK-STAT3 signaling, cell adhesion, chemokine signaling, T cell receptor

TABLE 1 | PAFAH1B3 expression associated with clinical pathological characteristics in LIHC (logistic regression).

Characteristics	Total(N)	Odds ratio (OR)	*p* Value
T stage (T2 and T3 vs. T1)	358	2.426 (1.591–3.723)	<0.001
Pathologic stage (Stage III & Stage IV vs. Stage I and Stage II)	350	1.922 (1.182–3.162)	0.009
Histologic grade (G3 and G4 vs. G1 & G2)	369	3.315 (2.132–5.217)	<0.001
Age (>60 vs. ≤60)	373	0.657 (0.436–0.987)	0.044
Weight (>70 vs. ≤70)	346	0.504 (0.327–0.772)	0.002
Height (≥170 vs. <170)	341	0.494 (0.317–0.764)	0.002
Race (Black or African American vs. Asian)	177	0.810 (0.295–2.262)	0.681

TABLE 2 | Univariate and multivariate Cox regression analyses of clinical characteristics associated with OS of LIHC.

Characteristics	Total(N)	Univariate analysis		Multivariate analysis	
		Hazard ratio (95% CI)	*p* Value	Hazard ratio (95% CI)	*p* Value
T stage	370				
T1	183	Reference			
T2	94	1.428 (0.901–2.264)	0.129	1.390 (0.755–2.561)	0.291
T3 and T4	93	2.949 (1.982–4.386)	<0.001	1.776 (0.238–13.258)	0.575
M stage	272				
M0	268	Reference			
M1	4	4.077 (1.281–12.973)	0.017	1.110 (0.263–4.688)	0.887
Pathologic stage	349				
Stage I and Stage II	259	Reference			
Stage III and Stage IV	90	2.504 (1.727–3.631)	<0.001	1.483 (0.203–10.856)	0.698
Tumor status	354				
Tumor free	202	Reference			
With tumor	152	2.317 (1.590–3.376)	<0.001	1.915 (1.202–3.052)	0.006
PAFAH1B3	373	1.235 (1.095–1.392)	<0.001	1.198 (1.036–1.385)	0.015

signaling, Toll-like receptor signaling, neuro-active ligand-receptor interaction, cytokine receptor interaction, MAPK signaling, vascular smooth muscle contraction, apoptosis, focal adhesion, and WNT signaling (**Figures 11A–C**). Findings show that PAFAH1B3 plays an important role in regulating immune responses.

PAFAH1B3 Knockdown Inhibits LIHC Cell Proliferation and Migration

A loss of function assay was performed to examine the functional role of PAFAH1B3 in LIHC cancer cells. PAFAH1B3 was upregulated in LIHC cell lines (**Figure 12A**) and inhibited by siRNA in SMC7721 and Hu7 cells. Knockdown efficacy was verified using real-time RT-PCR, using cell lines expressing a negative control as the control (**Figure 12B**). As expected, PAFAH1B3 knockdown inhibited SMC7721 and Hu7 cell proliferation (**Figure 12C**) and colony formation ability (**Figures 12D,E**). Trans-well and invasion assays showed that cancer cell migration and invasion were dramatically repressed in PAFAH1B3 knockdown cells compared with the control group (**Figures 12F,G**). These results support an oncogenic role for PAFAH1B3 in LIHC.

DISCUSSION

Emerging evidence shows that PAFAH1B3 plays an important role in apoptosis (Nagy et al., 2021), cancer metastasis (Monillas et al., 2015), and angiogenesis during cancer (Stafforini, 2015). Previous studies show that PAFAH1B3 is highly expressed in HSCC and is positively correlated with cervical lymph node metastasis. Depletion of PAFAH1B3 suppresses cell proliferation, migration, and induces apoptosis, thereby disrupting cell cycle processes (Xu et al., 2019). PAFAH1B3 may serve as a potential therapeutic target for HSCC patients. However, no studies have analyzed the clinical significance of PAFAH1B3 in pan-cancer.

Pan et al. found that PAFAH1B3 was more highly expressed in the HSCC tumor tissues than adjacent non-tumor samples. Moreover, increased PAFAH1B3 expression was positively correlated with cervical lymph node metastasis and adverse clinical outcome in HSCC (Xu et al., 2019). Furthermore, depletion of PAFAH1B3 inhibited the cell proliferation via modulating cell apoptosis and disrupting cell cycle process, and the migratory and invasive capacities were also attenuated in the absence of PAFAH1B3 (Xu et al., 2019). Yuan et al. found that PAFAH1B3 was overexpressed in osteosarcoma tissues and cell lines. Moreover, depletion of PAFAH1B3 inhibits the osteosarcoma cell proliferation and induced cell apoptosis *in vitro*, and also reduced osteosarcoma growth *in vivo*. This research confirmed that PAFAH1B3 could be a novel therapeutic target for osteosarcoma patients (Fan et al., 2021).

In this study, PAFAH1B3 was highly expressed in ACC, BLCA, BRCA, CESC, COAD, CHOL, DLBC, ESCA, GBM, HNSC, KIRP, LGG, LIHC, LUAD, LUSC, OV, PAAD, PRAD,

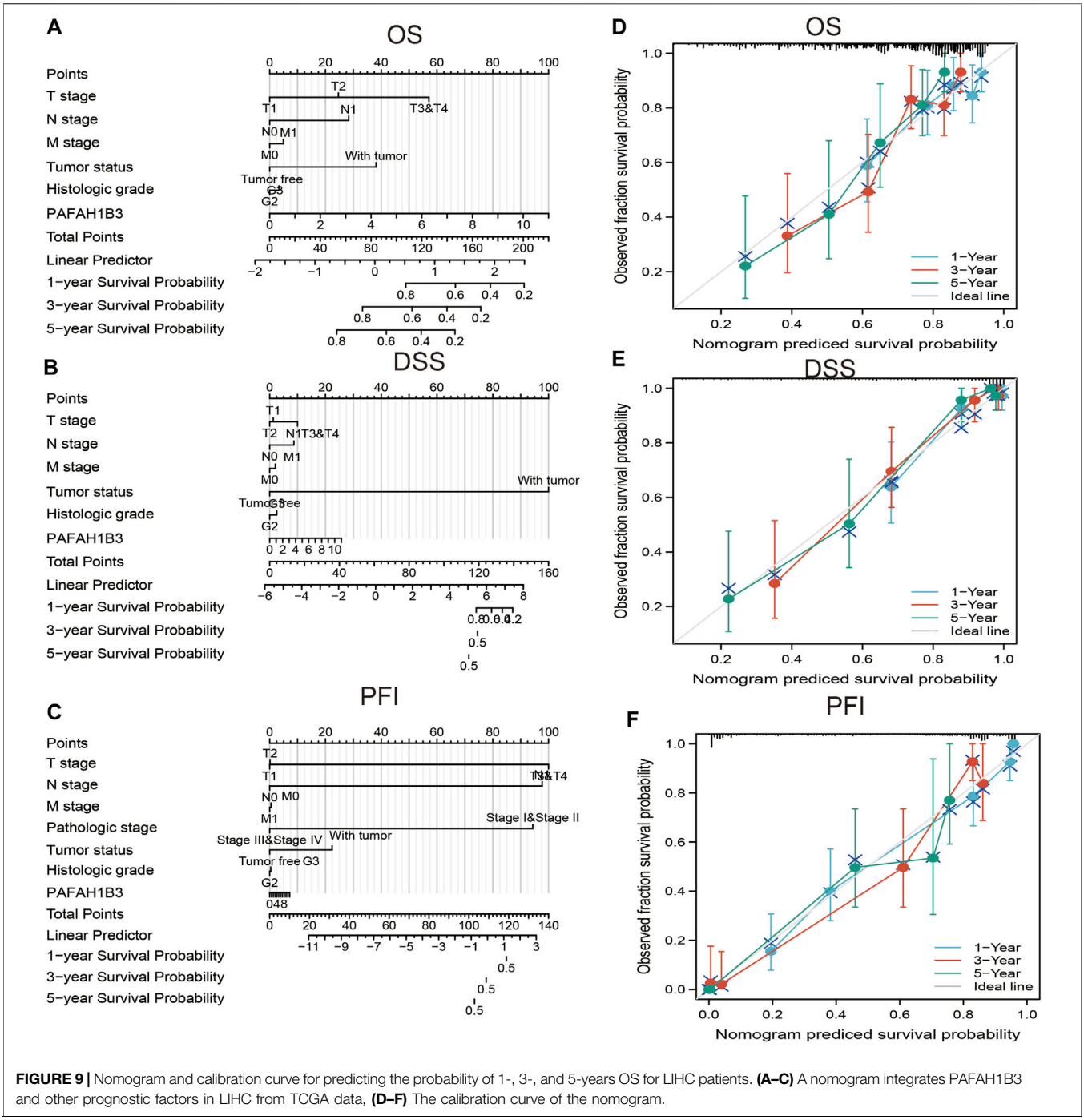


FIGURE 9 | Nomogram and calibration curve for predicting the probability of 1-, 3-, and 5-years OS for LIHC patients. **(A–C)** A nomogram integrates PAFAH1B3 and other prognostic factors in LIHC from TCGA data, **(D–F)** The calibration curve of the nomogram.

READ, SKCM, STAD, TGCT, THCA, THYM, UCEC, and UCS, and high expression was associated with pathology in BRCA, HNSC, KIRC, LIHC, LUAD, TGCT, and THCA. Increased PAFAH1B3 expression also correlated with poor prognosis in ACC, LIHC, LUAD, MESO, SARC, and SKCM. Findings confirmed that PAFAH1B3 had a moderate accuracy (AUC > 0.75) in predicting BLCA, BRCA, CHOL, COAD, ESCA, GBM, HNSC, KICH, KIRP, LAML, LGG, LIHC, LUAD, LUSC, OV, PAAD, PRAD, THCA, THYM, and UCS, confirming that PAFAH1B3 has the potential to act as a detection index for the diagnosis of diverse cancers.

PAFAH1B3 expression was also associated with immune and molecular subtypes of cancer, including UCEC, BLCA, BRCA, STAD, SKCM, MESO, LUSC, LUAD, LIHC, LGG, CESC, and KIRC. TMB and MSI play a significant role in tumor immunotherapy. PAFAH1B3 expression was positively correlated with the TMB in MESO, LUAD, STAD, PAAD, ACC, LGG, DLBC, UVM, and PRAD, negatively correlated with the TMB in THYM, positively correlated

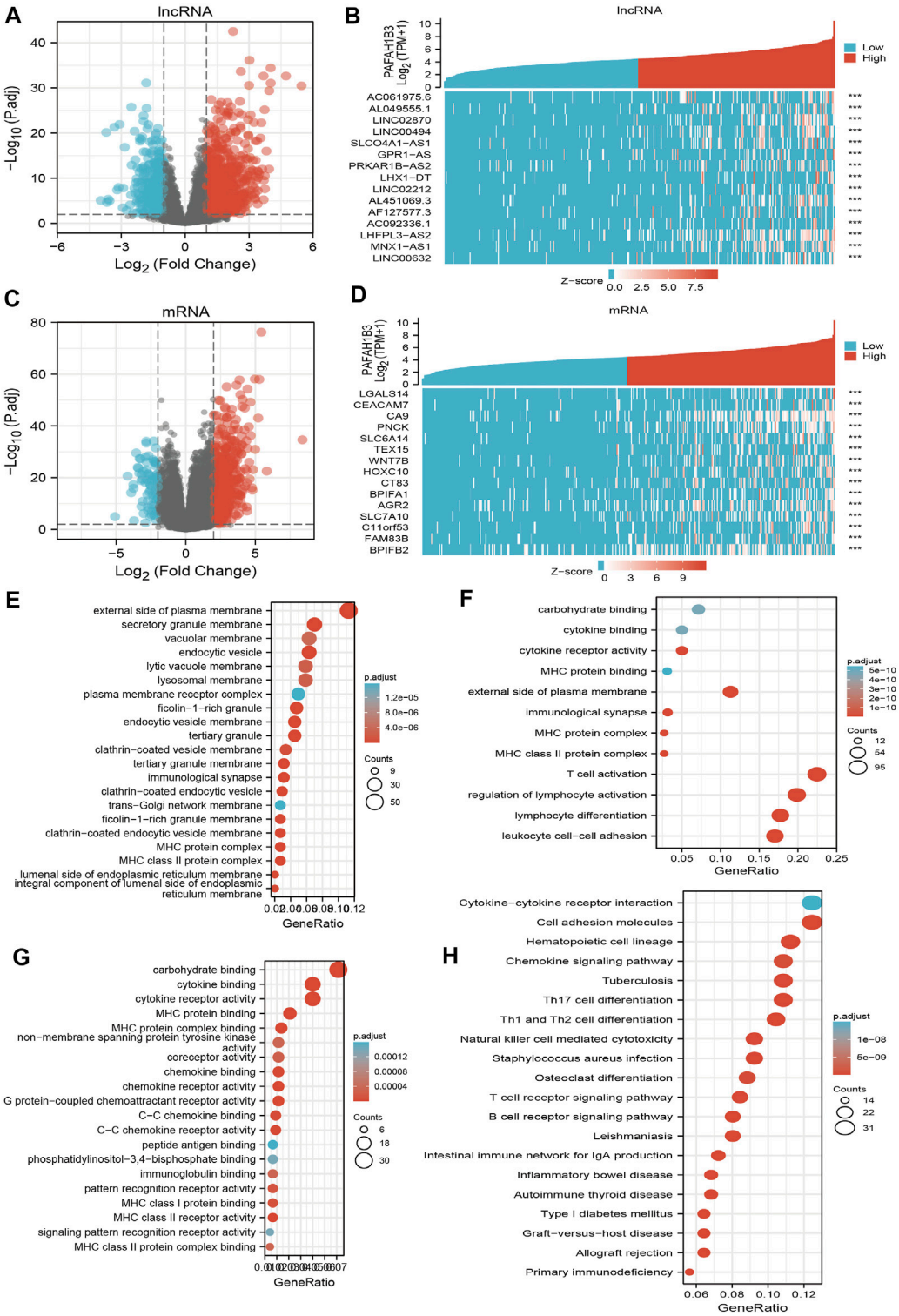


FIGURE 10 | Identifying differentially expressed genes between high and low expression of PAFAH1B3 groups. **(A)** Volcano plot of differential lncRNA profiles between PAFAH1B3 high expression and PAFAH1B3 low expression. **(B)** Hot map of the top 15 DEGs (lncRNA) between PAFAH1B3 high expression and PAFAH1B3 low expression. **(C)** Volcano plot of differential mRNA profiles between PAFAH1B3 high expression and PAFAH1B3 low expression. **(D)** Hot map of the top 15 DEGs (mRNA) between PAFAH1B3 high expression and PAFAH1B3 low expression. **(E–G)** The GO term of PAFAH1B3 analysis by using differentially expressed genes. **(H)** The KEGG term of PAFAH1B3 analysis by using differentially expressed genes.

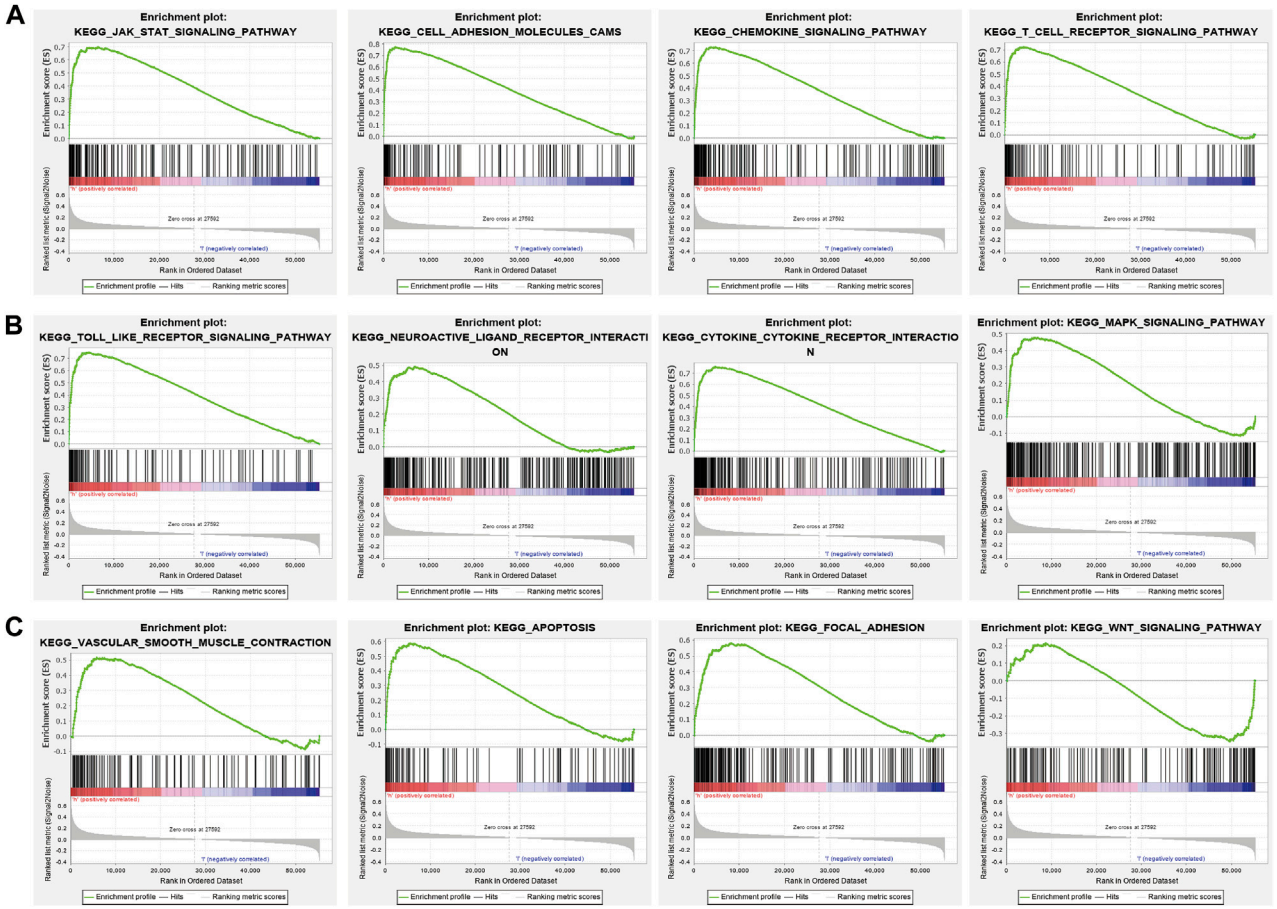


FIGURE 11 | GSEA enrichment PAFAH1B3-related signaling pathway in LIHC **(A–C)** GSEA enrichment PAFAH1B3-related signaling pathway in LIHC.

with the MSI in DLBC, STAD, PAAD, MESO, ESCA, UCEC, and LIHC, and negatively correlated with the MSI in TGCT, LAML, COAD, UCS, and READ. These results confirm that PAFAH1B3 may serve as a tumor immunotherapy-related biomarker. PAFAH1B3 was primarily involved in cell proliferation and oxidative phosphorylation signaling pathways in various cancers.

Chen et al. showed that PAFAH1B3 was up-regulated in gastric cancer. High PAFAH1B3 expression was significantly correlated with high M1 macrophage and CD8-positive T cell infiltration scores. PAFAH1B3 knockdown inhibited the proliferation, migration, and the activation of oncogenic signaling in gastric cancer cells (Xie et al., 2021). Immune cell infiltration plays an indispensable role in cancer progression. In this study, PAFAH1B3 expression was significantly correlated with $CD8^+$ T cell abundance in 27 cancers, $CD4^+$ T cell abundance in 28 cancers, neutrophil abundance in 30 cancers, DC abundance in 30 cancers, macrophage abundance in 27 cancers, and B cell abundance in 29 cancers. PAFAH1B3 expression also correlated positively with immune checkpoint-related genes such as CD274, CTLA4, HAVCR2, LAG3, PDCD1, PDCD1LG2, SIGLEC15, and TIGIT in 31 cancers. TISIDB analysis showed that PAFAH1B3 expression was positively associated with genes for 28 tumor-infiltrating lymphocytes, 45 immune-stimulators, 24 immune-inhibitors, 41 chemokines, 18 receptors, and 21 MHCs in pan-cancer. These findings indicate that PAFAH1B3 plays an important role in regulating the immune response during human cancer. PAFAH1B3 was also significantly associated with diverse drug sensitivity in many cancer cell lines and maybe a promising therapeutic target for cancer.

This study further assessed the correlation between PAFAH1B3 and clinical characteristics and prognosis of NSCLC. High PAFAH1B3 expression was significantly associated with pathologic stage, TNM stage, residual tumor, and primary therapy outcome. Higher expression of PAFAH1B3 had a worse OS in most clinical and demographic subgroups of NSCLC, including pathologic stage, TN stage, residual tumor, gender, age, smoking status, and race.

As the most common subtype of NSCLC, accumulating evidence has confirmed that LUAD and LUSC differ from each other in their bio-pathology, molecular, clinical characteristics, and therapeutic effect (Faruki et al., 2017). In this study, we found that high PAFAH1B3 expression was significantly associated with histologic stage, tumor status, pathologic stage, TNM stage, residual tumor, vascular invasion, race, BMI, gender, age, weight, and height in LIHC. Univariate and multivariate Cox regression analyses showed that TM stage, pathologic stage, tumor status, and PAFAH1B3

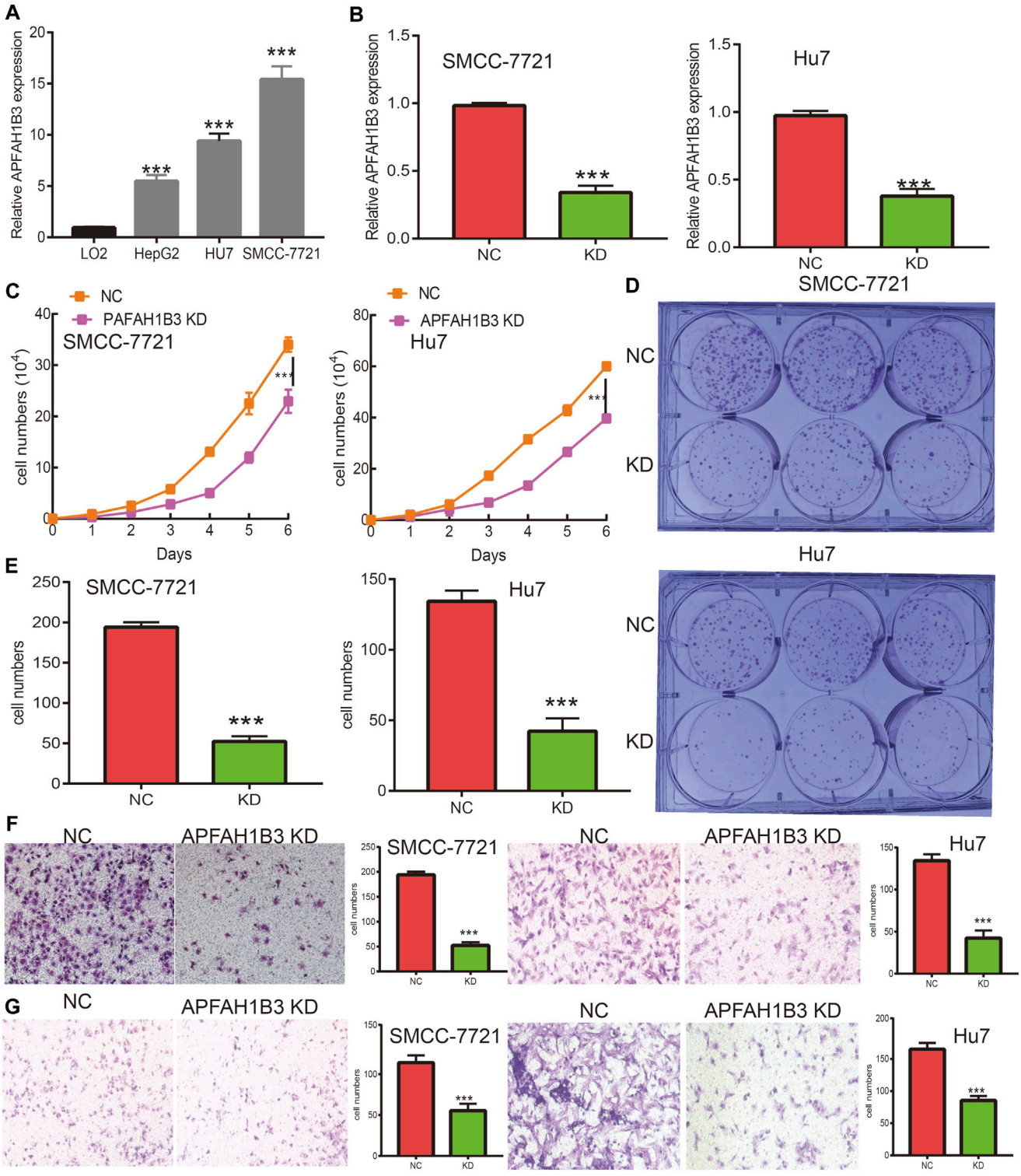


FIGURE 12 | Knockdown of PAFAH1B3 inhibits LIHC progression. **(A)** The expression PAFAH1B3 in LIHC cells lines was examined by qRT-PCR assay. **(B)** Establishment of PAFAH1B3 knockdown cell lines in SMC-7721 and Hu7 verified by qRT-PCR assay. **(C)** knockdown of PAFAH1B3 dramatically inhibits SMC-7721 and Hu7 cells proliferation examined by growth curve assay. **(D–E)** knockdown of PAFAH1B3 dramatically inhibits SMC-7721 and Hu7 cells colony formation ability. **(F–G)** knockdown of PAFAH1B3 dramatically inhibits SMC-7721 and Hu7 cells migration and invasion abilities.

expression were significantly associated with the OS. TNM stages, tumor status, histologic stage, and PAFAH1B3 expression were also included in a nomogram to predict OS, DSS, and PFI during LIHC. The C-indices of OS, DSS, and PFI were 0.680, 0.871, and 0.808, respectively.

Previous study showed that PAFAH1B3 plays a functional role in spindle formation and meiotic progression during bovine oocyte maturation (Vandenberghe et al., 2018). Aberrant higher expression of PAFAH1B3 promotes the cell proliferation and inhibits cell apoptosis of osteosarcoma (Xu et al., 2019). Recent study confirmed that high PAFAH1B3 expression was associated with high M1 macrophage and CD8-positive T cell infiltration scores (Xie et al., 2021).

To better understand the role of PAFAH1B3 in LIHC, KEGG enrichment analysis showed that these DEGs were primarily involved in the hematopoietic cell lineage, Th1 and Th2 cell differentiation, Th17 cell differentiation, cell adhesion, the intestinal immune network for IgA production, allograft rejection, *Staphylococcus aureus* infection, graft-versus-host disease, type I diabetes mellitus, *Leishmaniasis* infection, autoimmune thyroid disease, B cell receptor signaling, primary immunodeficiency, T cell receptor signaling, tuberculosis, inflammatory bowel disease, natural killer cell-mediated cytotoxicity, and chemokine signaling.

Bastian et al. found that platelet-activating factor acetylhydrolase expression in BRCA1 Mutant Ovarian cancer as a protective factor and potential negative regulator of the Wnt Signaling pathway. In this study, we showed that high PAFAH1B3 expression was associated with the JAK-STAT3 signaling, cell adhesion, chemokine signaling, T cell receptor signaling, Toll-like receptor signaling, neuro-active ligand-receptor interaction, cytokine receptor interaction, MAPK signaling, vascular smooth muscle contraction, apoptosis, focal adhesion, and Wnt signaling.

PAFAH1B3 is overexpressed in gastric cancer and knockdown of PAFAH1B3 inhibits proliferation, migration, and activation of oncogenic signaling in gastric cancer cells (Xie et al., 2021). Findings from this study showed that PAFAH1B3 was upregulated in LIHC cancer cell lines and knockdown of PAFAH1B3 inhibited the proliferation, migration, and invasion ability of LIHC cancer cells. These results demonstrate that PAFAH1B3 expression is correlated with LIHC progression.

CONCLUSION

In summary, this study showed that PAFAH1B3 was elevated in multiple types of human cancer, and high expression correlated with poor prognosis. High expression of PAFAH1B3 was also associated with TMB, MSI, immune cell infiltration, and sensitivity to multiple cancer drugs. Finally, PAFAH1B3 was shown to play a critical role in the progression of LIHC, in part by promoting cell proliferation, migration, and invasion. Results indicate that PAFAH1B3 may serve as a biomarker for the clinical detection of cancer. This study provides the first evidence that PAFAH1B3 impacts cancer progression and immune responses to human pan-cancer.

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

YY, XJ, and LT designed this work, performed related assay, and analyzed data. JW contributed to study materials. LD supervised and wrote the manuscript. All authors have read and approved the final version of the manuscript.

FUNDING

This work was supported by National Nature Science Foundation of China (82160508) and Yunnan Applied Basic Research Projects (YNWRMY-2019-067, 2019FE001) and Yunnan Province Specialized Training Grant for High-Level Healthcare Professionals (D-201614).

ACKNOWLEDGMENTS

We would like to thank the Core Technology Facility, Kunming Institute of Zoology, Chinese Academy of Sciences for providing us with bioinformatics analysis. And we are grateful to Guolan Ma for her technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.799497/full#supplementary-material

REFERENCES

Addeo, A., Friedlaender, A., Banna, G. L., and Weiss, G. J. (2021). TMB or Not TMB as a Biomarker: That Is the Question. *Crit. Rev. Oncology/Hematology* 163, 103374. doi:10.1016/j.critrevonc.2021.103374

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

Basu, A., Bodycombe, N. E., Cheah, J. H., Price, E. V., Liu, K., Schaefer, G. I., et al. (2013). An Interactive Resource to Identify Cancer Genetic and Lineage Dependencies Targeted by Small Molecules. *Cell* 154 (5), 1151–1161. doi:10.1016/j.cell.2013.08.003

Boland, C. R., and Goel, A. (2010). Microsatellite Instability in Colorectal Cancer. *Gastroenterology* 138 (6), 2073–2087. doi:10.1053/j.gastro.2009.12.064

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., et al. (2012). The cBio Cancer Genomics portal: an Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov.* 2 (5), 401–404. doi:10.1158/2159-8290.cd-12-0095

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 (8), 649–658. doi:10.1016/j.neo.2017.05.002

Fan, J., Yang, Y., Qian, J.-K., Zhang, X., Ji, J.-Q., Zhang, L., et al. (2021). Aberrant Expression of PAFAH1B3 Affects Proliferation and Apoptosis in Osteosarcoma. *Front. Oncol.* 11, 664478. doi:10.3389/fonc.2021.664478

Faruki, H., Mayhew, G. M., Serody, J. S., Hayes, D. N., Perou, C. M., and Lai-Goldman, M. (2017). Lung Adenocarcinoma and Squamous Cell Carcinoma Gene Expression Subtypes Demonstrate Significant Differences in Tumor Immune Landscape. *J. Thorac. Oncol.* 12 (6), 943–953. doi:10.1016/j.jtho.2017.03.010

Fiedler, E. R. C., Bhutkar, A., Lawler, E., Besada, R., and Hemann, M. T. (2018). *In Vivo* RNAi Screening Identifies Pafah1b3 as a Target for Combination Therapy with TKIs in BCR-Abl1+ BCP-ALL. *Blood Adv.* 2 (11), 1229–1242. doi:10.1182/bloodadvances.2017015610

Ghandi, M., Huang, F. W., Jané-Valbuena, J., Kryukov, G. V., Lo, C. C., McDonald, E. R., et al. (2019). Next-generation Characterization of the Cancer Cell Line Encyclopedia. *Nature* 569 (7757), 503–508. doi:10.1038/s41586-019-1186-3

Ito, K., and Murphy, D. (2013). Application of Ggplot2 to Pharmacometric Graphics. *CPT Pharmacometrics Syst. Pharmacol.* 2 (10), e79. doi:10.1038/psp.2013.56

Jiang, L.-P., Fan, S.-Q., Xiong, Q.-X., Zhou, Y.-C., Yang, Z.-Z., Li, G.-F., et al. (2018). GRK5 Functions as an Oncogenic Factor in Non-small-cell Lung Cancer. *Cell Death Dis* 9 (3), 295. doi:10.1038/s41419-018-0299-1

Kume, K., and Shimizu, T. (1997). Platelet-activating Factor (PAF) Induces Growth Stimulation, Inhibition, and Suppression of Oncogenic Transformation in NRK Cells Overexpressing the PAF Receptor. *J. Biol. Chem.* 272 (36), 22898–22904. doi:10.1074/jbc.272.36.22898

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 (21), e108–e110. doi:10.1158/0008-5472.can-17-0307

Livnat, I., Finkelshtein, D., Ghosh, I., Arai, H., and Reiner, O. (2010). PAF-AH Catalytic Subunits Modulate the Wnt Pathway in Developing GABAergic Neurons. *Front Cell Neurosci* 4, 19. doi:10.3389/fncel.2010.00019

Manya, H., Aoki, J., Kato, H., Ishii, J., Hino, S., Arai, H., et al. (1999). Biochemical Characterization of Various Catalytic Complexes of the Brain Platelet-Activating Factor Acetylhydrolase. *J. Biol. Chem.* 274 (45), 31827–31832. doi:10.1074/jbc.274.45.31827

Mizuno, H., Kitada, K., Nakai, K., and Sarai, A. (2009). PrognoScan: a New Database for Meta-Analysis of the Prognostic Value of Genes. *BMC Med. Genomics* 2, 18. doi:10.1186/1755-8794-2-18

Monillas, E. S., Caplan, J. L., Thévenin, A. F., and Bahnson, B. J. (2015). Oligomeric State Regulated Trafficking of Human Platelet-Activating Factor Acetylhydrolase Type-II. *Biochim. Biophys. Acta (Bba) - Proteins Proteomics* 1854 (5), 469–475. doi:10.1016/j.bbapap.2015.02.007

Nagy, Á., Munkácsy, G., and Győrffy, B. (2021). Pancancer Survival Analysis of Cancer Hallmark Genes. *Sci. Rep.* 11 (1), 6047. doi:10.1038/s41598-021-84787-5

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 (20), 4200–4202. doi:10.1093/bioinformatics/btz210

Stafforini, D. M. (2015). Diverse Functions of Plasma PAF-AH in Tumorigenesis. *Enzymes* 38, 157–179. doi:10.1016/bs.enz.2015.09.005

Sun, L., He, Z., Ke, J., Li, S., Wu, X., Lian, L., et al. (2015). PAF Receptor Antagonist Ginkgolide B Inhibits Tumourigenesis and Angiogenesis in Colitis-Associated Cancer. *Int. J. Clin. Exp. Pathol.* 8 (1), 432–440.

Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). GEPIA: a Web Server for Cancer and normal Gene Expression Profiling and Interactive Analyses. *Nucleic Acids Res.* 45 (W1), W98–w102. doi:10.1093/nar/gkx247

Vandenberghe, L. T. M., Heindryckx, B., Smits, K., Szymanska, K., Ortiz-Escribano, N., Ferrer-Buitrago, M., et al. (2018). Platelet-activating Factor Acetylhydrolase 1B3 (PAFAH1B3) Is Required for the Formation of the Meiotic Spindle during *In Vitro* Oocyte Maturation. *Reprod. Fertil. Dev.* 30 (12), 1739–1750. doi:10.1071/rd18019

Xie, T., Guo, X., Wu, D., Li, S., Lu, Y., Wang, X., et al. (2021). PAFAH1B3 Expression Is Correlated with Gastric Cancer Cell Proliferation and Immune Infiltration. *Front. Oncol.* 11, 591545. doi:10.3389/fonc.2021.591545

Xing, Z., Tang, X., Gao, Y., Da, L., Song, H., Wang, S., et al. (2011). The Human LIS1 Is Downregulated in Hepatocellular Carcinoma and Plays a Tumor Suppressor Function. *Biochem. Biophysical Res. Commun.* 409 (2), 193–199. doi:10.1016/j.bbrc.2011.04.117

Xiong, Q., Jiang, L., Liu, K., Jiang, X., Liu, B., Shi, Y., et al. (2021). miR-133b Targets NCAPH to Promote β-catenin Degradation and Reduce Cancer Stem Cell Maintenance in Non-small Cell Lung Cancer. *Sig Transduct Target. Ther.* 6 (1), 252. doi:10.1038/s41392-021-00555-x

Xu, J., Zang, Y., Cao, S., Lei, D., and Pan, X. (2019). Aberrant Expression of PAFAH1B3 Associates with Poor Prognosis and Affects Proliferation and Aggressiveness in Hypopharyngeal Squamous Cell Carcinoma. *Ott* 12, 2799–2808. doi:10.2147/ott.s196324

Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a Resource for Therapeutic Biomarker Discovery in Cancer Cells. *Nucleic Acids Res.* 41, D955–D961. doi:10.1093/nar/gks1111

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

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
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 in Molecular Biosciences

ORIGINAL RESEARCH
published: 07 March 2022
doi: 10.3389/fmolb.2021.793732

Systematic Analysis and Validation of the Prognosis, Immunological Role and Biology Function of the Ferroptosis-Related lncRNA GSEC/miRNA-101-3p/CISD1 Axis in Lung Adenocarcinoma

Xiulin Jiang[1,2†], *Yixiao Yuan*[1†], *Lin Tang*[1], *Juan Wang*[1], *Dahang Zhang*[1] *and Lincan Duan*[1*]

[1]The Department of Thoracic Surgery, The Third Affiliated Hospital of Kunming Medical University (Yunnan Tumor Hospital), Kunming, China, [2]Key Laboratory of Animal Models and Human Disease Mechanisms of Chinese Academy of Sciences, Kunming Institute of Zoology, Kunming, China

OPEN ACCESS

Edited by:
Matteo Becatti,
University of Firenze, Italy

Reviewed by:
Anirban Chakraborty,
University of Texas Medical Branch at Galveston, United States
Qiang Guo,
Hubei University of Medicine, China
Rui-Qing Zhai,
Harbin Medical University, China

***Correspondence:**
Lincan Duan
duanmumuhuosan@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 12 October 2021
Accepted: 22 December 2021
Published: 07 March 2022

Citation:
Jiang X, Yuan Y, Tang L, Wang J, Zhang D and Duan L (2022) Systematic Analysis and Validation of the Prognosis, Immunological Role and Biology Function of the Ferroptosis-Related lncRNA GSEC/miRNA-101-3p/CISD1 Axis in Lung Adenocarcinoma. Front. Mol. Biosci. 8:793732. doi: 10.3389/fmolb.2021.793732

Lung adenocarcinoma (LUAD) is the most common type of lung cancer, accounting for approximately 85% of pulmonary malignancies. Emerging evidence has demonstrated that ferroptosis plays a central role in both immunities as well as tumor proliferation. However, the clinical significance, immunological function, and upstream modulatory mechanism of ferroptosis-related genes in LUAD remain unclear. Here, we utilized various bioinformatics data to identify differentially expressed (DEGs) and prognosis-related ferroptosis (FRGs) genes in LUAD. Based upon identified DEGs, FRG, and ceRNA modulatory networks were constructed. Pearson's correlation analysis was used to evaluate the correlation between FRGs and the tumor mutational burden, microsatellite instability, tumor-infiltrating immunity, cellular checkpoint control, and drug sensitivity in LUAD. A loss-of-function analysis was performed to verify the function of CISD1 in LUAD progression. Our findings revealed that certain FRGs (CISD1, ATP5MC3, PGD, SLC7A11, ACSL3, and FANCD2) are significantly upregulated in LUAD and that their elevated expression is associated with both advanced tumor stage and unfavorable prognosis. Furthermore, Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment results revealed these FRGs to be primarily involved in ferroptosis and glutathione metabolism in LUAD. We constructed a prognostic FRG-based model capable of accurately predicting LUAD patient overall survival with high specificity. The upstream lncRNA GSEC/miRNA-101-3p regulatory axis involving CISD1, ATP5MC3, and PGD was identified to be relevant in tumor progression. We also found GSEC, CISD1, ATP5MC3, and PGD to be upregulated, with miRNA-101-3p downregulated, in the setting of LUAD. Immunohistochemical analysis revealed CISD1, ATP5MC3, and PGD overexpression in LUAD tissue samples; CISD1 knockdown was noted to significantly inhibit LUAD proliferation and migration. In summary, this study characterizes relevant functional roles of the lncRNA GSEC/miR-101-3p axis in the setting

of LUAD and suggests diagnostic and therapeutic biomarkers potentially useful in the clinical management of this illness.

Keywords: NSCLC, ferroptosis, ceRNA, immune cell infiltration, cell proliferation, cell migration

INTRODUCTION

Lung cancer remains the commonest fatal condition globally, with an estimated 2.09 million new cases and 1.76 million deaths annually (Rivera and Wakelee, 2016). Non-small cell lung cancer, the most frequently encountered disease subtype, typically results in adverse clinical outcomes and carries a 5-year survival rate of only 18% (Schwartz and Cote, 2016). Despite advances in early detection and treatment standardization, management strategies remain varied and range from chemo-radiotherapy to immunotherapy. Urgent identification of novel, specific biomarkers with clinicopathological and prognostic significance is thus necessary for developing successful lung adenocarcinoma (LUAD) management strategies.

Ferroptosis describes a recently discovered type of programmed cell death distinct from apoptosis that is iron-dependent and characterized by both lipid peroxidation and reactive oxygen species production (Wang et al., 2019; Luo et al., 2021; Xu et al., 2021). Ferroptosis has been reported to play an essential role in lung cancer progression; the interaction of LINC00336 with ELAVL1 was found to result in the inhibition of ferroptosis in LUAD (Wang et al., 2019). Similarly, the interaction of G3BP1 with lncRNA-P53RRA was reported to facilitate both ferroptosis and apoptosis in lung cancer *via* regulation of p53 expression (Mao et al., 2018). Ferroptosis was additionally reported to be regulated *via* cellular redox and cell cycle signaling pathways (Yang and Stockwell, 2016). Prior studies have suggested several oncogenic signaling pathways related to the ferroptosis process to inhibit tumor progression *via* modulation of ferroptosis (Kang et al., 2019; Mou et al., 2019). For instance, P53, a well-studied tumor-suppressor gene, was reported to suppress cystine/glutamate antiporter expression and thus inhibit the ferroptosis process (Mou et al., 2019). However, a comprehensive analysis has yet to establish the prognostic value of ferroptosis-related genes (FRGs) and their relevant upstream regulatory axis in LUAD.

Analysis of gene expression, mutation, DNA methylation, and prognostic value in LUAD revealed CISD1, ATP5MC3, PGD, SLC7A11, ACSL3, and FANCD2 to be significantly upregulated in LUAD and elevated expression of these FRGs to be associated with advanced tumor stage and poor prognosis. Furthermore, KEGG enrichment findings revealed these FRGs to primarily be involved in both ferroptosis and glutathione metabolism in the setting of LUAD. Six prognosis-related genes were also used to construct an FRG-based prognostic model capable of accurately predicting LUAD patient overall survival with high specificity. Finally, CISD1, ATP5MC3, PGD, SLC7A11, ACSL3, and FANCD2 expression was confirmed to significantly correlate with tumor mutational burden (TMB), microsatellite instability (MSI), immune cell infiltration, cellular checkpoint dysfunction, and cancer sensitivity to drugs. Additionally, we identified the upstream regulatory (namely, the lncRNA GSEC/miRNA-101-3p) axis of relevant FRGs (CISD1, ATP5MC3, PGD) and found GSEC, CISD1, ATP5MC3, and PGD to be upregulated, with miRNA-101-3p downregulated, in the setting of LUAD. Knockdown of CISD1 was noted to significantly inhibit the proliferative and migratory capabilities of LUAD cells. In summary, by characterizing the functional roles of the lncRNA GSEC/miR-101-3p axis in LUAD, we suggest diagnostic and therapeutic biomarkers potentially useful in the future clinical management of this condition.

MATERIALS AND METHODS

Data Collection

The Cancer Genome Atlas (TCGA)-LUAD cohort data and corresponding clinical information of 535 LUAD patients were downloaded from the TCGA website up to November 14, 2020 (https://portal.gdc.cancer.gov/repository). The gene expression profiles were normalized using the scale method provided in the "limma" R package. Data analysis was performed with the R (version 3.6.3) and ggplot2 (3.3.3) packages. The expression data were normalized to transcripts per kilobase million (TPM) values before further analysis.

Identification of Differentially Expressed 34 FRGs

A total of 34 FRGs were obtained from previous reviews (Su et al., 2020; Feng et al., 2020), which are shown in **Supplementary Table S1**. The difference in FRG expression in LUAD and normal tissues was identified using the "limma" packages. We then constructed a gene–gene interaction network for 34 FRGs using the GeneMANIA (http://www.genemania.org) database (Warde-Farley et al., 2010).

Gene Mutation Analysis of FRGs

The mutation frequency and oncoplot waterfall plot of 34 FRGs in LUAD patients were analyzed by the Gene Set Cancer Analysis (http://bioinfo.life.hust.edu.cn/web/GSCALite/) database (Liu et al., 2018).

Functional Enrichment Analysis

Gene Ontology (GO), consisting of the biological process (BP), cellular component (CC), and molecular function (MF) categories, was conducted with the "ggplot2" package in R software. Similarly, this package was also utilized to perform Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis (Ito and Murphy, 2013).

Development of the Ferroptosis-Related Gene Prognostic Model

Development of the FRG prognostic model was performed as documented (Lin et al., 2021). Briefly, we performed the Cox

regression analysis to examine the prognostic significance of the FRGs. For Kaplan–Meier curves, *p*-values and hazard ratios (HRs) with 95% confidence intervals (CIs) were generated by log-rank tests and univariate Cox proportional hazard regression. FRGs with a significant prognostic value was selected for further analysis. Based on these prognostic FRGs, LASSO Cox regression analysis was then used to construct the prognostic model. The TCGA LUAD patients were divided into low- and high-risk subgroups according to the median risk score, and the overall survival time was compared between the two subgroups *via* Kaplan–Meier analysis. The predictive accuracy of each gene and the risk score was evaluated by performing time receiver-operating characteristic (ROC) analysis. Considering the clinical characteristics, a predicted nomogram was developed to predict the 1-, 3-, and 5-year overall survival. A forest was used to show the *p*-value, HR, and 95% CI of each variable through the "forest plot" R package.

Prediction of lncRNA and ceRNA Network Construction

We used starbase (http://starbase.sysu.edu.cn/) to predict the potential upstream miRNAs of FRGs and examine the expression, prognosis, and correlation between miRNA-101-3p and lncRNA, as well as to predict the binding sites among miRNA, mRNA, and lncRNA (Li et al., 2014). We utilized lncLocator (www.csbio.sjtu.edu.cn/bioinf/lncLocator.) and CPC2 (http://cpc2.cbi.pku.edu.cn) to explore the subcellular localization and the protein-coding ability of lncRNAs (Kang et al., 2017; Cao et al., 2018).

Analysis of the Immunological Roles of FRGs in Lung Cancer

We utilize TIMER (https://cistrome.shinyapps.io/timer/) and XCELL tools (https://xcell.ucsf.edu/) to examine the immunological roles of FRGs (Aran et al., 2017; Li et al., 2017), including the correlation between diverse immune cells and immune regulators. The TMB and MSI scores were obtained from TCGA. A correlation analysis between the FRG expression and TMB or MSI was performed using Spearman's method.

Analysis of the Correlation Between the FRGs Expression and Drug Sensitivity

We utilize the Genomics of Drug Sensitivity in Cancer (GDSC) (www.cancerRxgene.org) and the Cancer Therapeutics Response Portal (CTRP) databases to analyze the correlation between FRG expression and drug sensitivity (Basu et al., 2013; Yang et al., 2013).

Cells and Cell Culture Conditions

The BEAS-2B cell line was purchased from the cell bank of Kunming Institute of Zoology and cultured in BEGM media (Lonza, Morrisville, NC, USA, CC-3170). Lung cancer cell lines, including A549 and H1975, were purchased from Cobioer (Nanjing, China), with the STR document; A549 and H1975 cells were all cultured in RPMI 1640 medium (Corning, Tewksbury, MA, USA) supplemented with 10% fetal bovine serum (Cat# 10099141C, Gibco, Grand Island, NY, USA) and 1% penicillin/streptomycin.

Quantitative Real-Time PCR

The qRT-PCR assay was performed as documented (Jiang et al., 2018). Total RNA was extracted according to the manufacturer's protocol, and then reverse-transcribed using RT Reagent Kit (Takara Bio, Beijing, China, Cat# RR047A; Tiangen Biotech, Beijing, China, Cat# KR211-02). Real-time PCR was performed by FastStart Universal SYBR Green Master Mix (Roche, Cat# 04194194001; Tiangen Biotech, Beijing, China, Cat# FP411-02) using an Applied Biosystems 7500 machine. The primer sequences are listed as follows: GSEC-F: TTCCAA TTAACCTGGCCGGAG, GSEC-R: GTCAGCCAACCCATT GCAAC, PGD-F: ATGGCCCAAGCTGACATCG, PGD-R: AAAGCCGTGGTCATTCATGTT, CISD1-F: CTGACTTCC AGTTCCAGCGT, CISD1-R:TGATCAGAGGGCCCACATTG, ATP5G3-F: ATGTTCGCCTGCGCCAAG, ATP5G3-R: GGC AAACAAGATCAAGAACGCA, miRNA-101-3p: TACAGT ACTGTGATAACTGAA. β-Actin was used to normalize expression levels: β-actin-F: CTTCGCGGGCGACGAT, β-actin-R: CCATAGGAATCCTTCTGACC. The expression quantification was obtained with the 2−ΔΔCt method.

Cell Proliferation and Cell Migration Assays

Cell proliferation assay was performed as previously described (Xu et al., 2020). Briefly, the indicated cells were plated onto 12-well plates, and the cell numbers were subsequently counted each day using an automated cell analyzer Countstar (Shanghai Ruiyu Biotech Co., Shanghai, China, IC1000). Cell migration assay was performed as previously described (Xiong et al., 2021). For transwell assay, 1-2×104 cells in 100 l serum-free medium were plated in an 8.0-μm, 24-well plate chamber insert (Corning Life Sciences, catalog no. 3422), with a medium containing 10% FBS at the bottom of the insert. Cells were incubated for 24 h and then fixed with 4% paraformaldehyde for 20 min. After washing, cells were stained with 0.5% crystal violet-blue. The positively stained cells were examined under the microscope.

Immunohistochemistry Staining

The immunohistochemistry staining assay was performed as documented (Dixon et al., 2012). Briefly, cancer tissues were collected; the primary antibody was incubated overnight and the second antibody incubated. Finally, the instrument was developed. Detailed information of the primers used in this study is as follows: PGD, Catalog number: #13389, dilution, 1: 100, CST, Shanghai, China; CISD1, Catalog number: #83775, dilution, 1:200, CST, Shanghai, China; and ATP5MC3, Catalog number: ab129742, dilution, 1:100, Abcam, Shanghai, China.

Statistical Analysis

Analysis of the FRG expression lung cancer was estimated using t-tests. For survival analysis, the HR and *p*-value were calculated employing univariate Cox regression analysis. Kaplan–Meier analysis was used to examine the survival time of patients

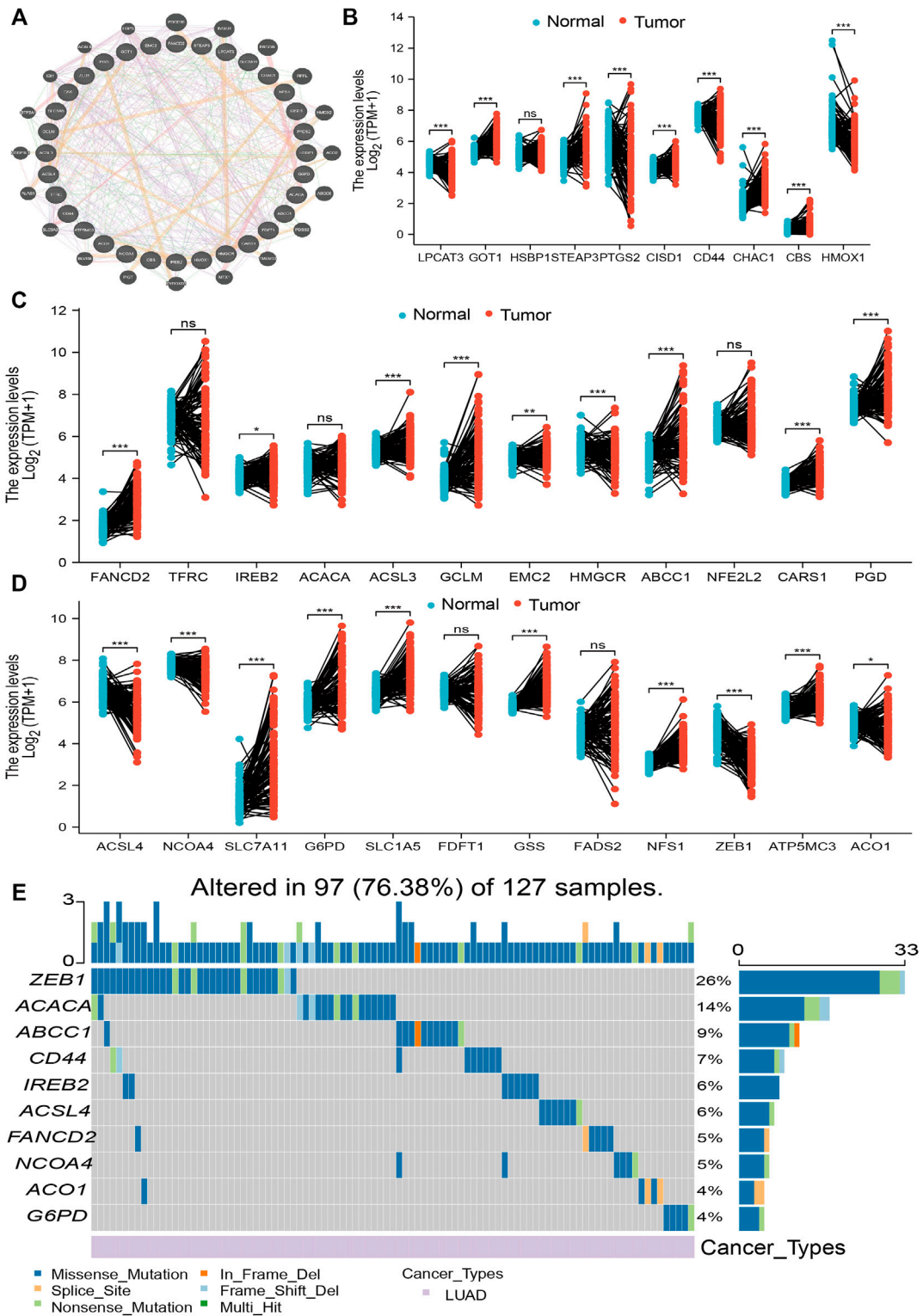


FIGURE 1 | Analysis of the expression and gene mutation of FRGs in LUAD. **(A)** The gene–gene interaction networks of FRG analysis by the GeneMANIA database. **(B–D)** The expression of FRGs in LUAD is examined by the TCGA database. **(E)** The mutation frequency and classification of FRGs in LUAD were examined by GSCA tools. For all figures, *, **, and *** indicate $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

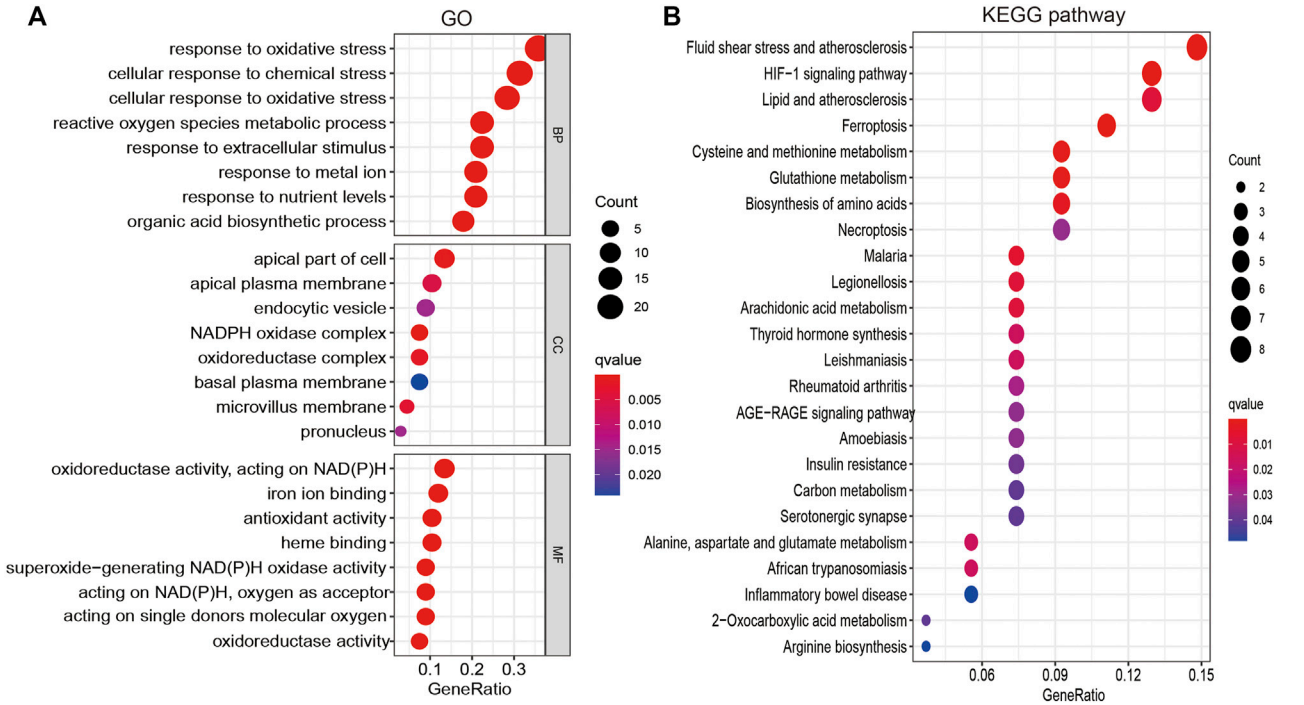


FIGURE 2 | Analysis of the functions of FRGs in LUAD. **(A)** The biological process involved by FRGs in LUAD. **(B)** The KEGG pathway is involved by FRGs in LUAD.

stratified according to high or low levels of the FRG expression. *p*-values less than 0.05 were considered statistically significant. For all figures, *, **, and *** indicate $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

RESULTS

Analysis of FRG Expression and Mutation in LUAD

The expression of 34 FRGs and their function in the setting of LUAD was analyzed using data obtained from the TCGA database (**Supplementary Table S1**). Findings confirmed 26 genes to be upregulated and seven genes to be downregulated in lung cancer; no other significant differences were noted (**Figures 1A–D**). Furthermore, analysis of FRG mutations in LUAD revealed 97 of 127 (76.38%) samples to exhibit genetic mutations; *ZEB1* exhibited the highest mutation frequency and was followed by *ACACA* and *ABCC1* (**Figure 1E**). Finally, analysis of correlations among copy number variation (CNV), DNA methylation, and FRG expression in LUAD revealed FDFT1, IREB2, NFS1, TFRC, HSBP1, ACO1, LPCAT3, CARS, GSS, ACACA, ACSL3, CISD1, SLC1A5, and NCOA4 expression to positively correlate with their respective CNV (**Supplementary Figure S1A**). Meanwhile, DNA methylation was found to negatively associate with CBS, FDFT1, STEAP3, FADS2, SCL1A5, GCLM, GOT1, SLC7A11, CD44, ACSL3, and G6PD expression in LUAD (**Supplementary Figure S1B**). Collectively, these data indicate that CNV and DNA methylation exert a critical influence on FRG expression in LUAD.

Analysis of FRG Function in LUAD

To explore the function of FRGs in LUAD, GO and KEGG enrichment analyses were performed. Results revealed 34 FRGs to be primarily involved in cellular responses to oxidative and other chemical stress, reactive oxygen species metabolism, extracellular stimuli, nutrient levels, biosynthetic processes, and iron ion binding (**Figure 2A**). Furthermore, KEGG pathway enrichment analysis revealed these FRGs to be primarily involved in vascular fluid dynamics and pathogenesis of atherosclerosis, HIF-1 signaling, lipid metabolism, ferroptosis, amino acid metabolism, necroptosis, thyroid hormone synthesis, serotonergic communication, and inflammatory bowel disease pathogenesis, as well as the metabolism of arachidonic acid, aspartate, alanine, glutamate, and 2-monocarboxylic acid (**Figure 2B**). Collectively, these data confirm that FRGs influence cancer cell metabolism, with ferroptosis playing an essential role in LUAD progression.

The Prognostic Value of FRGs in LUAD

An investigation of the correlation between FRG expression and tumor stage revealed only ATP5G3, CISD1, and PGD expression to positively correlate with the LUAD stage (**Figure 3A**). Evaluation of the prognostic value of FRGs revealed elevated SLC7A11, PGD, FANCD2, CISD1, ATP5G3, and ASCL3 expression to closely correlate with adverse clinical outcomes in LUAD (**Figure 3B**). The diagnostic value of FRGs was evaluated using ROC analysis. Findings confirmed SLC7A11, FANCD2, CISD1, and ATP5G3 expression to accurately predict the LUAD prognosis area under the curve (AUC > 0.8) (**Figure 3C**). Higher levels of SLC7A11, FANCD2, SIDS1, ATP5G3, and ASCL3 expression were similarly found to correlate with disease-specific survival in LUAD

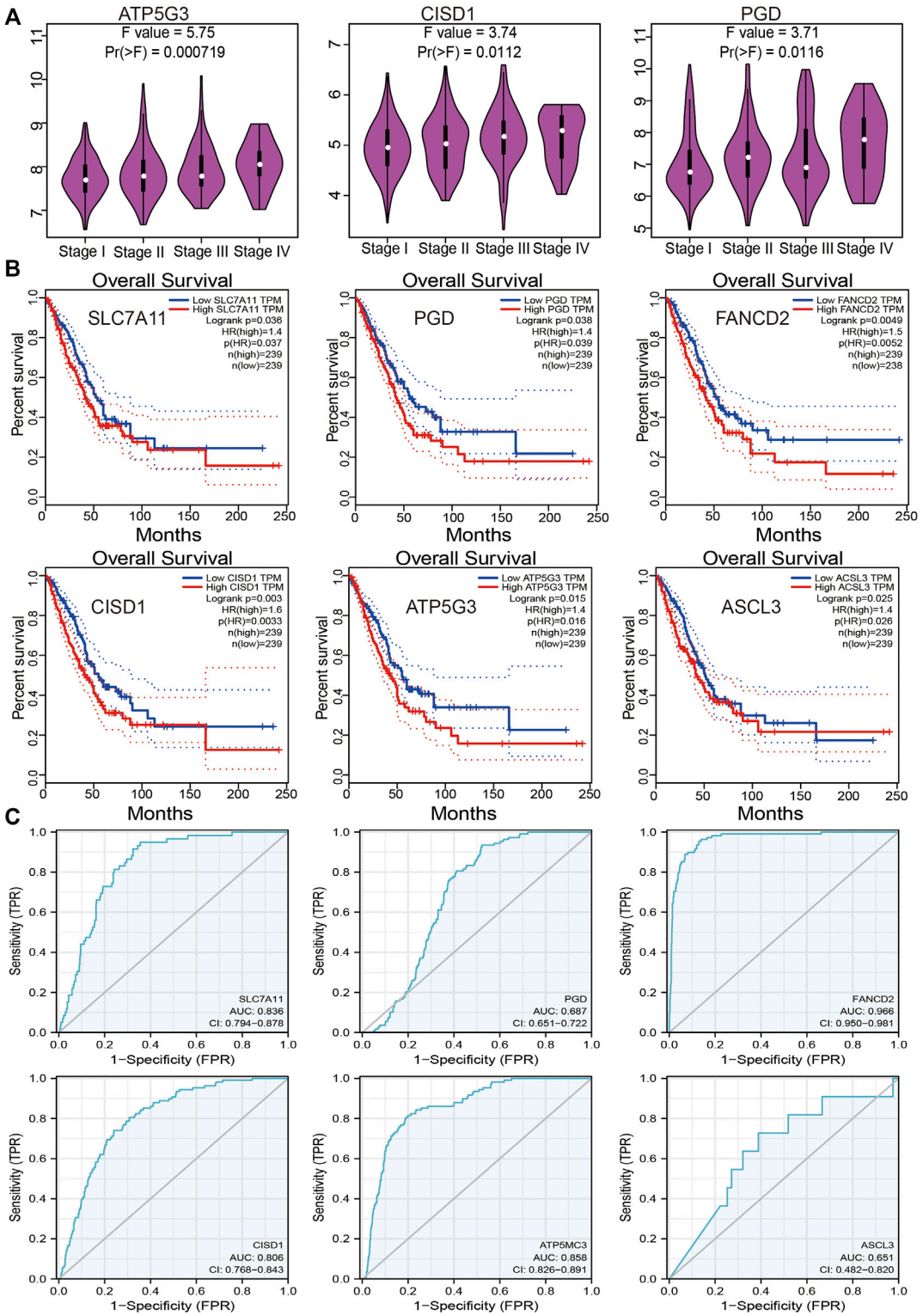


FIGURE 3 | Analysis of the prognosis of FRGs in LUAD. **(A)** The pathologic stage of FRGs in LUAD is examined by the GEPIA database. **(B)** The overall survival of FRGs in LUAD was examined by the GEPIA database. **(C)** The ROC curve of FRGs in LUAD is examined by the TCGA database.

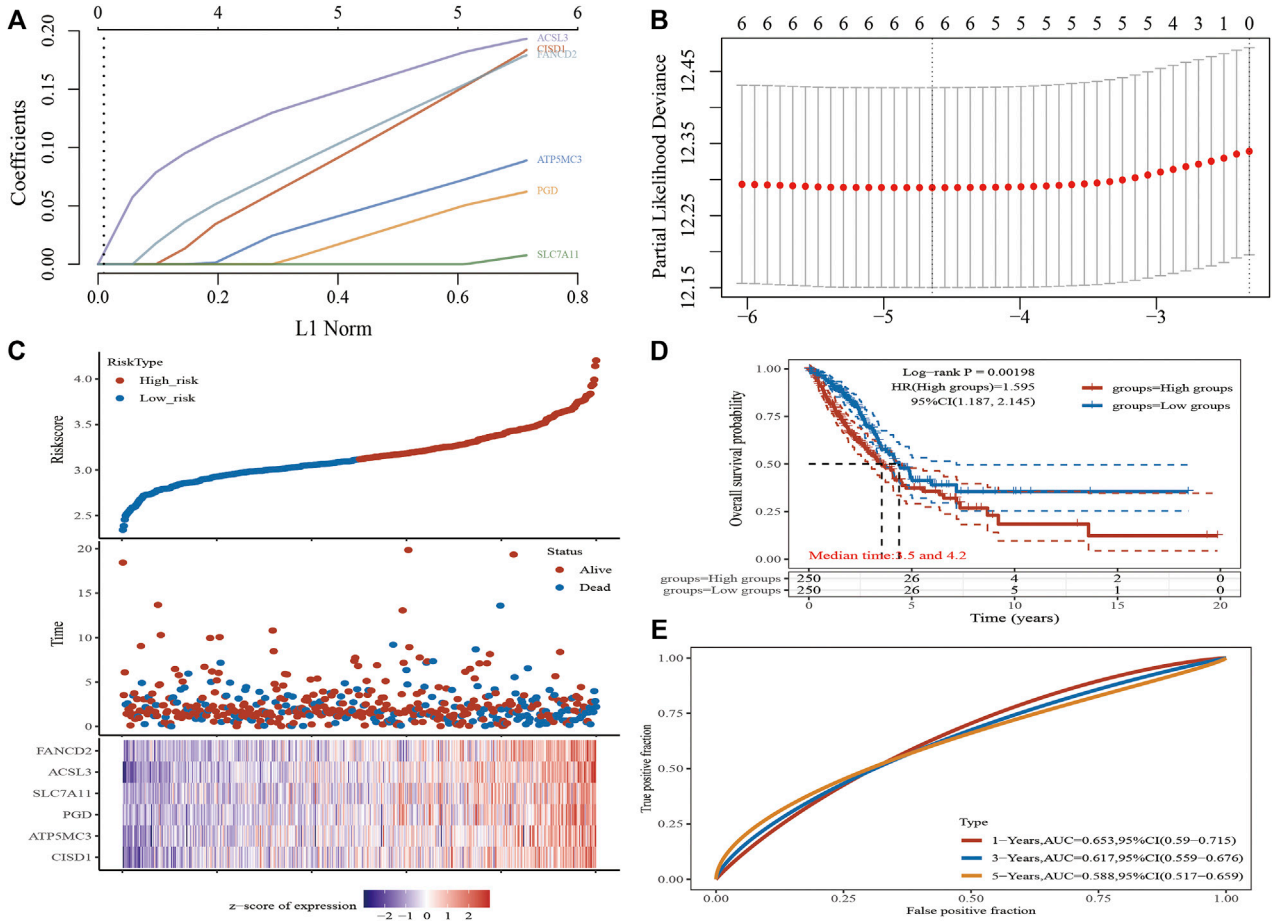


FIGURE 4 | Construction of a prognostic FRG model in LUAD. **(A)** LASSO coefficient profiles of six FRGs. **(B)** Plots of the ten-fold cross-validation error rates. **(C)** Distribution of risk score, survival status, and the expression of six prognostics FRGs in LUAD. **(D,E)** Overall survival curves for LUAD patients in the high-/low-risk group and the ROC curve of measuring the predictive value. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

(**Supplementary Figure S2A**). Elevated SLC7A11, ATP5G3, and ASCL3 expressions were found to correlate with progression-free survival in LUAD (**Supplementary Figure S2B**). Our findings thus confirm that FRG plays an essential role in LUAD progression.

Construction of an FRG-Based Prognostic Model

A Lasso Cox regression prognostic model based on the six aforementioned FRGs was constructed (**Figures 4A,B**) with the risk score=(0.0387)*PGD+(0.082)*ATP5MC3+(0.2118)*CISD1+(0.0029)*SLC7A11+(0.2104)*ACSL3+(0.1528)*FANCD2. Patients suffering LUAD were divided into two groups based on the risk score. Risk score distribution, survival status, and gene expression data are presented in **Figure 4C**. Patient risk of death increased and survival time decreased as the risk score increased (**Figure 4C**). Kaplan–Meier survival analysis suggested that LUAD patients with high-risk scores possess a worse overall survival probability as compared to those with low-risk scores (median time = 3.3 vs. 4.2 years, p = 0.00198; **Figure 4D**), with AUCs of 0.653, 0.617, and 0.588 noted on 1-, 3-, and 5-year receiver operating characteristic (ROC) curves, respectively (**Figure 4E**). The above results were verified by the GEO dataset; Kaplan–Meier survival analysis suggested that LUAD patients with high-risk scores possess a worse overall survival probability as compared to those with low-risk scores (median time = 2.9 vs. 4.9 years, p = 6.52e-08), with AUCs of 0.66, 0.693, and 0.664 noted on 1-, 3-, and 5-year ROC curves, respectively (**Supplementary Figures S6A–D**). A nomogram predictive of patient survival state was constructed. Univariate and multivariate analyses revealed ATP5MC3 expression and TNM stage to be independent factors affecting LUAD prognosis (**Figures 5A,B**). The nomogram could accurately predict 1-, 3-, and 5-year overall survival rates as compared with an ideal model of the entire cohort (**Figures 5C,D**).

Analysis of the Correlation Between FRG Expression and TMB, MSI, and Drug Sensitivity

Emerging evidence has suggested that TMB and MSI may serve as potential predictive biomarkers for immunotherapy efficacy in

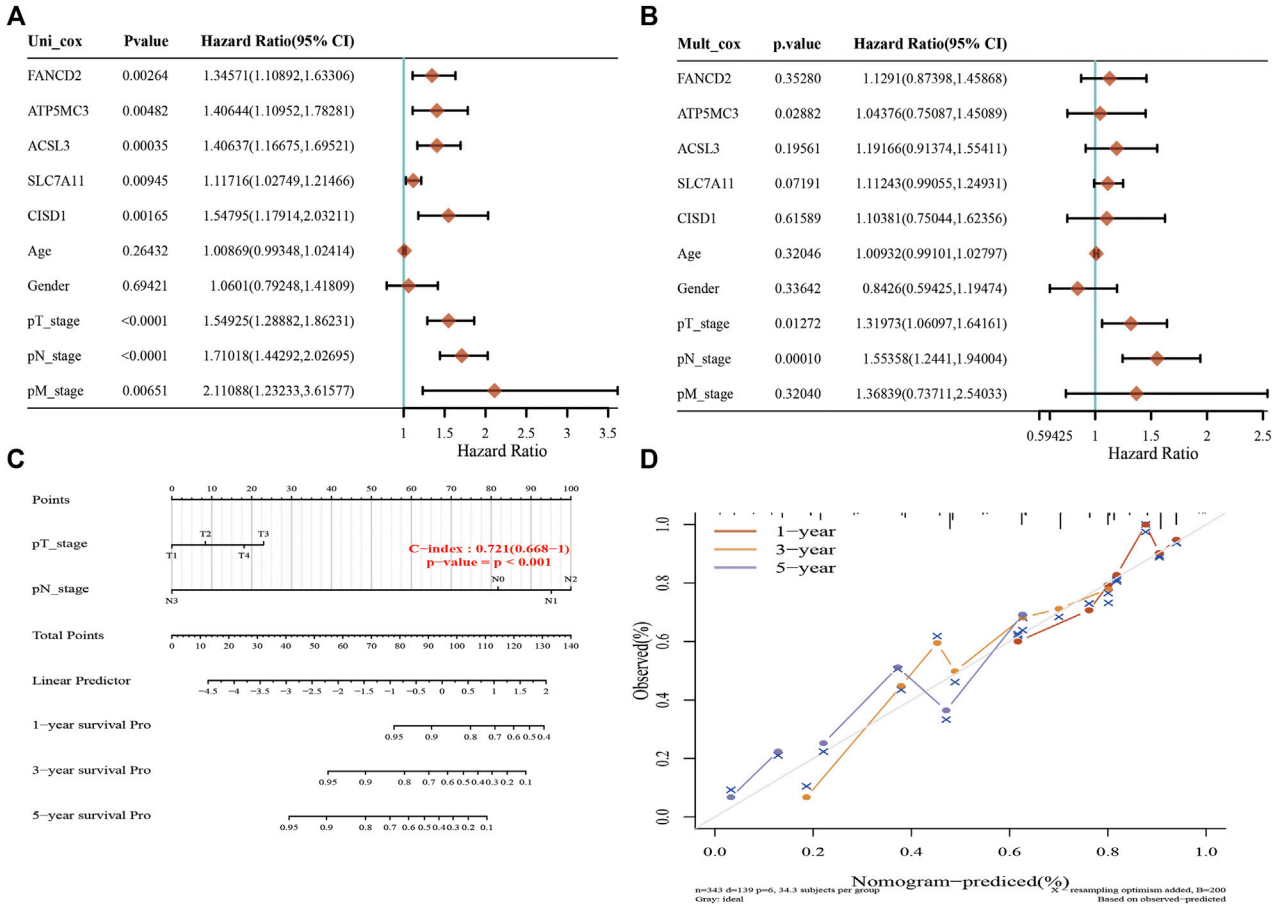


FIGURE 5 | Construction of a predictive nomogram in LUAD. **(A,B)** Hazard ratio and *p*-value of the constituents involved in univariate and multivariate Cox regression considering clinical parameters and six prognostics FRGs in LUAD. **(C,D)** Nomogram predicts the 1-, 3-, and 5-year overall survival rate of LUAD patients and calibration curve for the overall survival nomogram model in the discovery group. A dashed diagonal line represents the ideal nomogram. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

the setting of lung cancer (Goodman et al., 2019). As such, several studies have reported that the expression of FRGs significantly correlates with tumor immune infiltration. Correlation analysis revealed FANCD2, ATP5MC3, and PGD expression to positively correlate with TMB (**Supplementary Figure S3A**); FANCD2 expression was also noted to positively correlate with MSI. Meanwhile, CISD1 expression was found to negatively correlate with MSI in LUAD (**Supplementary Figure S3B**). To explore potential therapeutic targets, gene-set co-expression analysis was utilized to explore the relationship between FRG expression and drug sensitivity. Results revealed SLC7A11 expression to positively correlate with sensitivity to PRIMA-1, PX-12, necrosulfonamide, methylstat, piperlongumine, SMER-3, NSC95397, manumycin A, ML162, PL-DI1S, 3R-RSL-3, pifithrin-mu, and cerulenin ($r > 0.36$). Expression of ACSL3 was found to positively correlate with sensitivity to tozasertib, PRIMA-1, PX-12, manumycin A, BRD-K30748066, ML210, BRD-A94377914, 1S, 3R-RSL-3, and docetaxel ($r > 0.30$). On the contrary, PGD expression was found to negatively correlate with sensitivity to tivantinib, SCH-79797, BI-2536, GW-405833, GSK461364, nakiterpiosin, docetaxel, SB-743921, bafilomycin A1, linifanib, ceranib-2, BRD-K70511574, BRD-K01737880, FQI-2, and BRD-K30748066 ($r < -0.22$). Expression of ATP5MC3 was found to negatively correlate with sensitivity to BRD-K30748066, marinopyrrole A, COL-3, dinaciclib, BI-2536, alvocidib, methotrexate, and bafilomycin A1 ($r < -0.22$). The expression of FANCD2 was found to negatively correlate with sensitivity to GSK-J4, BRD-K30748066, COL-3, docetaxel, GSK461364, BRD-K66453893, tivantinib, BI-2536, narciclasine, BRD-K70511574, and triazolothiadiazine ($r < -0.34$; **Supplementary Figure S3D**). These findings thus underscore that FRG expression either positively or negatively correlates with LUAD sensitivity to drugs.

FRG Expression Associated With LUAD Immune Infiltration

Analysis of CISD1, FANCD2, PGD, ASCL1, ATP5MC3, and SLC7A11 expression in the setting of C1 and C2 LUAD subtypes revealed high CISD1, FANCD2, and ATP5MC3

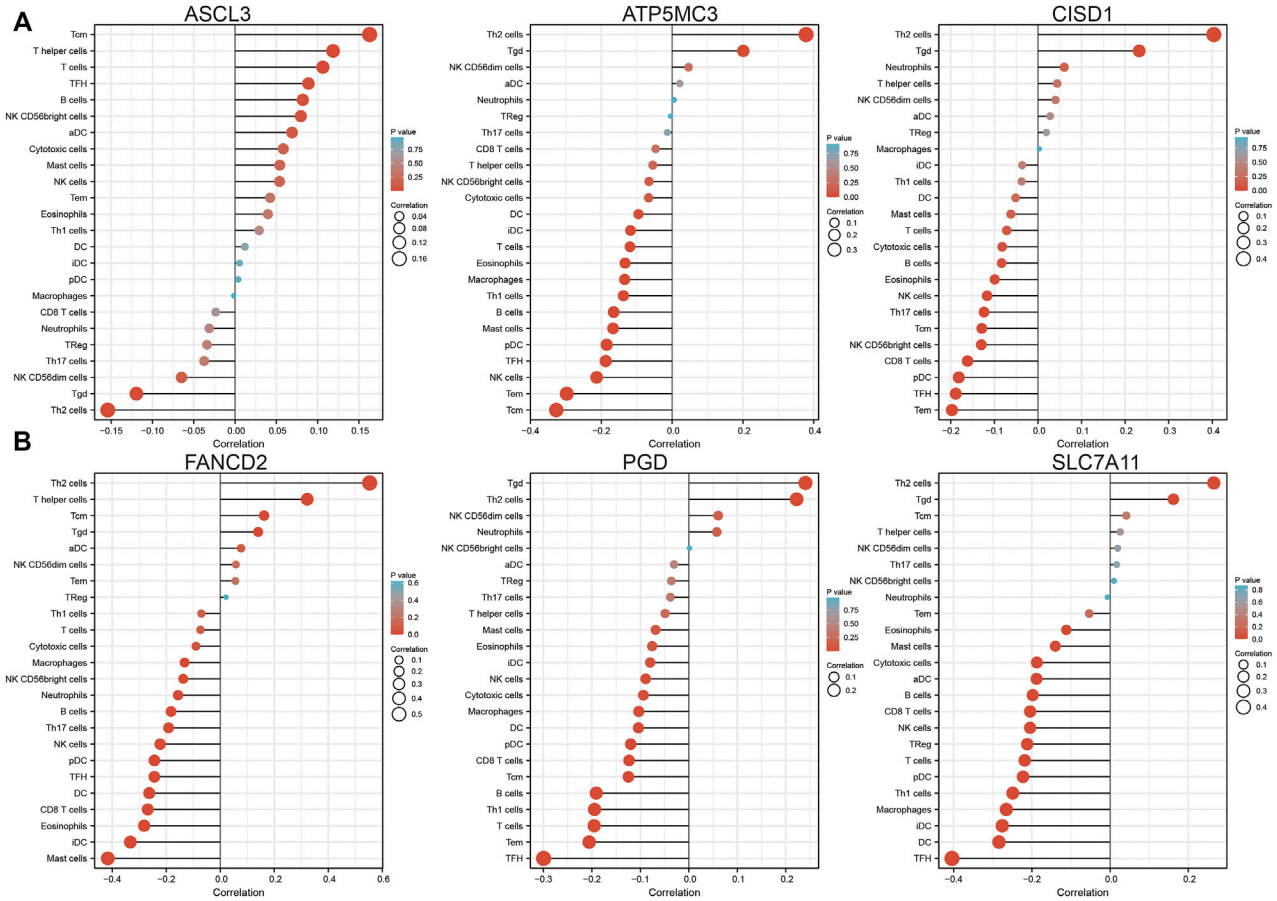


FIGURE 6 | Analysis of the association between FRG expression and immune infiltration level in LUAD. **(A,B)** The association between FRG expression and immune infiltration level in LUAD.

expression in the C2 subtype and high PGD and SLC7A11 expression in the C1 subtype (**Supplementary Figures S4A,B**). Analysis of data obtained from the TIMER database revealed FRG somatic copy number alterations to significantly correlate with levels of immune cell infiltration in LUAD (**Supplementary Figures S5A–F**). As ferroptosis plays crucial roles in both immunity and pulmonary carcinogenesis (Huang et al., 2021), the correlation between CISD1, FANCD2, PGD, ASCL3, ATP5MC3, and SLC7A11 expression and tumor immune infiltration in LUAD was evaluated utilizing data obtained from the TIMER database. Expression of ASCL3 was found to positively correlate with the level of infiltration by central memory T cells, T helper cells, T follicular helper cells, B cells, natural killer cells, CD56 bright natural killer cells, activated dendritic cells, cytotoxic cells, and mast cells. On the contrary, ATP5MC3 expression was found to negatively correlate with the level of infiltration by cytotoxic cells, dendritic cells, interstitial dendritic cells, T cells, eosinophils, macrophages, T helper one cells, B cells, mast cells, plasmacytoid dendritic cells, T follicular helper cells, natural killer cells, and central and effector memory T cells. The expression of CISD1 was found to negatively correlate with the level of infiltration by mast cells, T cells, cytotoxic cells, B cells, eosinophils, natural killer cells, T helper 17 cells, CD56 bright natural killer cells, CD8 T cells, plasmacytoid dendritic cells, T follicular helper cells, and central and effector memory T cells (**Figure 6A**). The expression of FANCD2 was found to negatively correlate with the level of infiltration by cytotoxic cells, macrophages, CD56 bright natural killer cells, neutrophils, B cells, T helper 17 cells, natural killer cells, plasmacytoid dendritic cells, T follicular helper cells, dendritic cells, CD8 T cells, eosinophils, interstitial dendritic cells, and mast cells. Similarly, the PGD expression was found to negatively correlate with the level of infiltration by mast cells, eosinophils, interstitial dendritic cells, natural killer cells, cytotoxic cells, macrophages, dendritic cells, plasmacytoid dendritic cells, CD8 T cells, B cells, T helper one cell, T cells, T follicular helper cells, and effector and memory T cells. The expression of SLC7A11 was found to negatively correlate with the level of infiltration by eosinophils, mast cells, cytotoxic cells, activated dendritic cells, B cells, CD8 T cells, natural killer cells, regulatory T cells, T cells, plasmacytoid dendritic cells, T helper one cell, macrophages, interstitial dendritic cells, dendritic cells, and T follicular helper cells (**Figure 6B**).

As immune checkpoints play a crucial role in tumor immunosuppression, we analyzed the distribution of a relevant

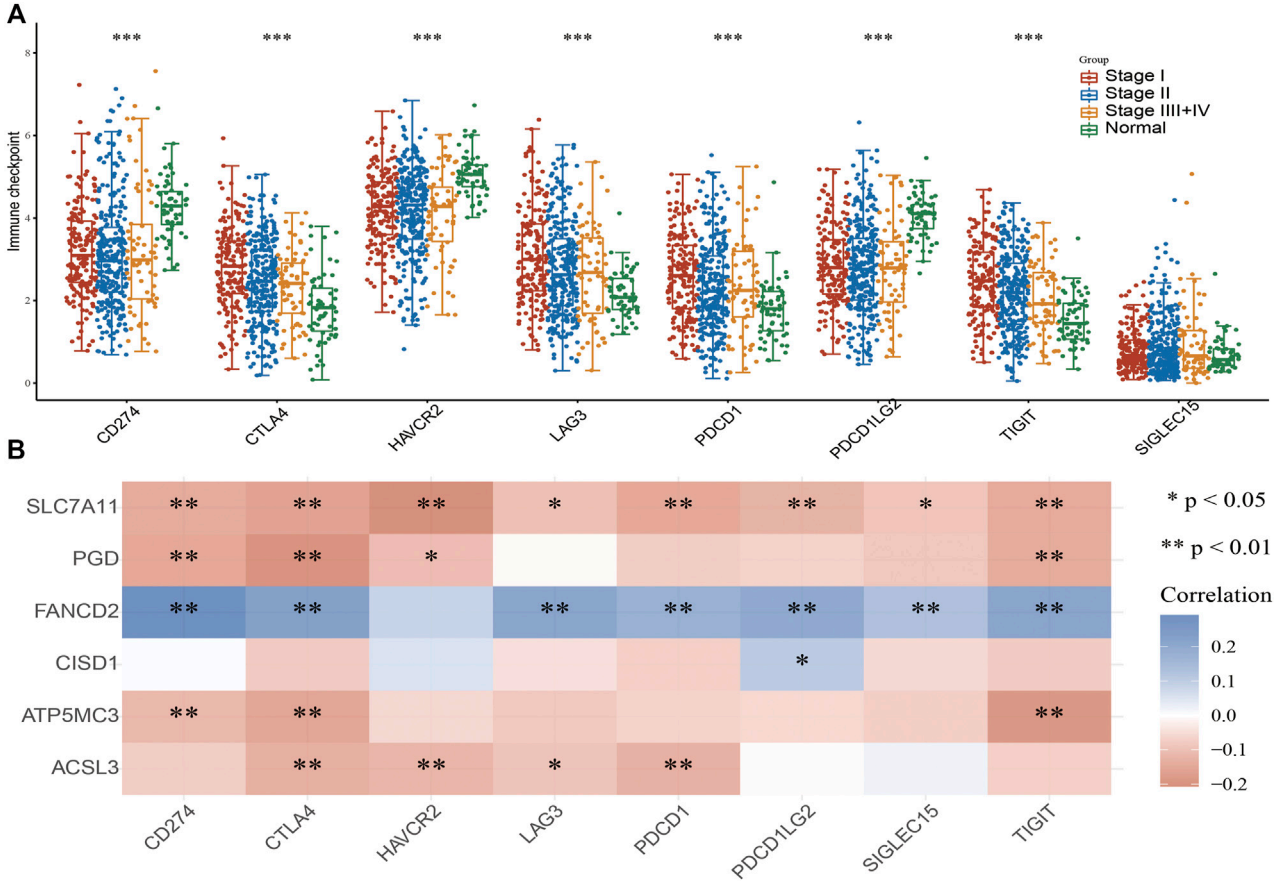


FIGURE 7 | Analysis of the correlation between FRG expression and diverse immune modulators. **(A)** The expression of immune checkpoint-related genes in diverse pathologic stages and normal samples. **(B)** The correlation between FRGs expression and diverse immune checkpoint-related genes in LUAD. For all figures, *, **, and *** indicate $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

gene expression in stage I–IV LUAD tissues (**Figure 7A**). We additionally evaluated the relationship between FRG and immune checkpoint-related gene (including CD274, CTLA4, HAVCR2, LAG3, PDCD1, PDCD1LG12, TIGIT, SIGLEC15) expression in LUAD *via* Pearson correlation analysis; results revealed the expression of SLC7A11, PGD, CISD1, ATP5MC3, and ACSL3 to negatively correlate with that of CD274, CTLA4, HAVCR2, LAG3, PDCD1, PDCD1LG12, TIGIT, and SIGLEC15. Meanwhile, FANCD2 expression was found to positively correlate with checkpoint-related gene expression (**Figure 7B**). These findings confirm that FRG expression significantly correlates with that of immune checkpoint-related genes in LUAD.

Analysis of the Upstream FRG Molecular Regulatory Axis

The aforementioned evidence suggests that ATP5MC3, PGD, and CISD1 expression correlates with both stage and progression of LUAD. To further explore the relevant upstream FRG regulatory axis, data obtained from various public databases were used to construct a network of mRNA–miRNA–lncRNA interactions. We found miRNA-101-3p to most significantly bind the 3′UTR of ATP5MC3, PGD, and CISD1 (**Figure 8A**). Further analysis revealed miRNA-101-3p to be decreased in lung cancer based on TCGA-LUAD and GSE74190 datasets (**Figures 8B,C**). Moreover, downregulated miRNA-101-3p expression correlated with the manifestation of poorer clinical features and prognosis among lung cancer patients (**Figures 8D,E**). Analysis of miRNA-101-3p ROC curve data revealed an AUC value of 0.864 among lung cancer patients (**Figure 8F**). We also found miRNA-101-3p expression to significantly negatively correlate with ATP5MC3, PGD, and CISD1 expression in LUAD (**Figure 8G**). To predict and obtain the PVT1 and GSEC lncRNAs, starbase and lncbase were utilized (**Figures 8H–J**). According to the ceRNA theory, lncRNA negatively and positively correlates with miRNA and mRNA expression, respectively. We found only GSEC to negatively correlate with miRNA-101-3p expression (**Figure 8K**) and positively correlate with ATP5MC3, PGD, and CISD1 expression in LUAD (**Figure 9A**). Importantly, GSEC was found to be highly expressed in the setting of this malignancy based on the GSE81089 dataset (**Figure 9B**) and

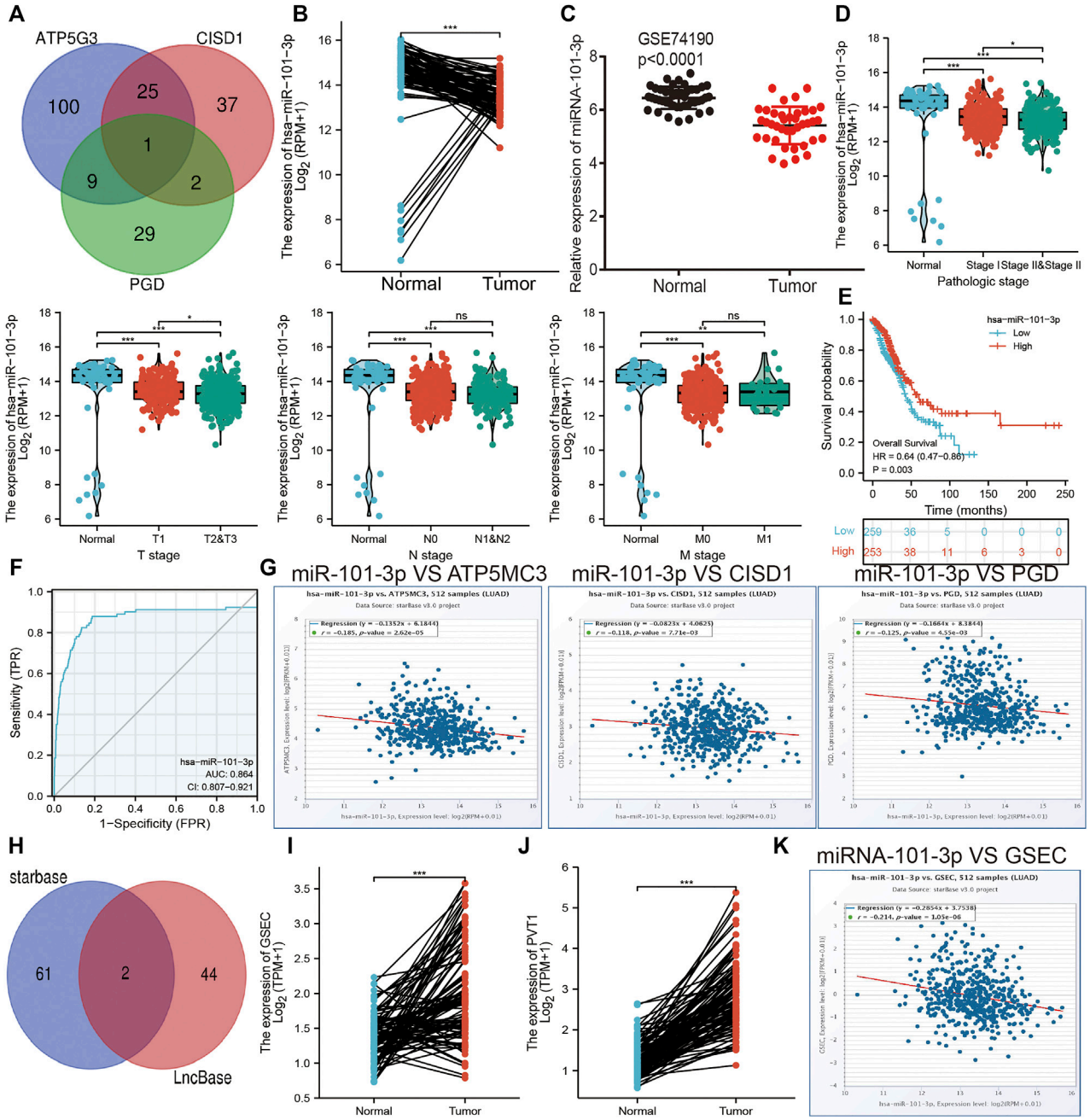


FIGURE 8 | Prediction and analysis of the upstream miRNAs of FRGs in LUAD. **(A)** Prediction of the potential miRNAs of FRGs in LUAD examined by starbase. **(B,C)** The expression of miRNA-101-3p in LUAD was examined by TCGA/GEO databases. **(D)** The correlation between FRGs and clinical features in LUAD. **(E)** The prognosis value of miRNA-101-3p in LUAD. **(F)** ROC curve analyses and AUC values for miRNA-101-3p in LUAD. **(G)** Pearson's correlation analysis determined the correlation between miRNA-101-3p and ATP5MC3, CISD1, and PGD expression in LUAD examined by starbase. **(H)** Explored the potential lncRNAs of miRNA-101-3p in LUAD by starbase. **(I,J)** The expression of GESC and PVT1 in LUAD was examined by starbase. **(K)** Pearson's correlation analysis determined the correlation between miRNA-101-3p and GSEC in LUAD. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

primarily localized to the cytoplasm as determined using the lncLocator and AnnoLnc RNA tools (**Figure 9C**). However, GSEC was not found to possess coding potential (**Figure 9D**). High levels of GSEC expression were found to correlate not only with tumor stage but also with poor prognosis in LUAD. Analysis of the GSEC ROC curve data revealed an AUC value of 0.813 in LUAD patients (**Figures 9E–G**). In summary, our findings confirm that the lncRNA GSEC/

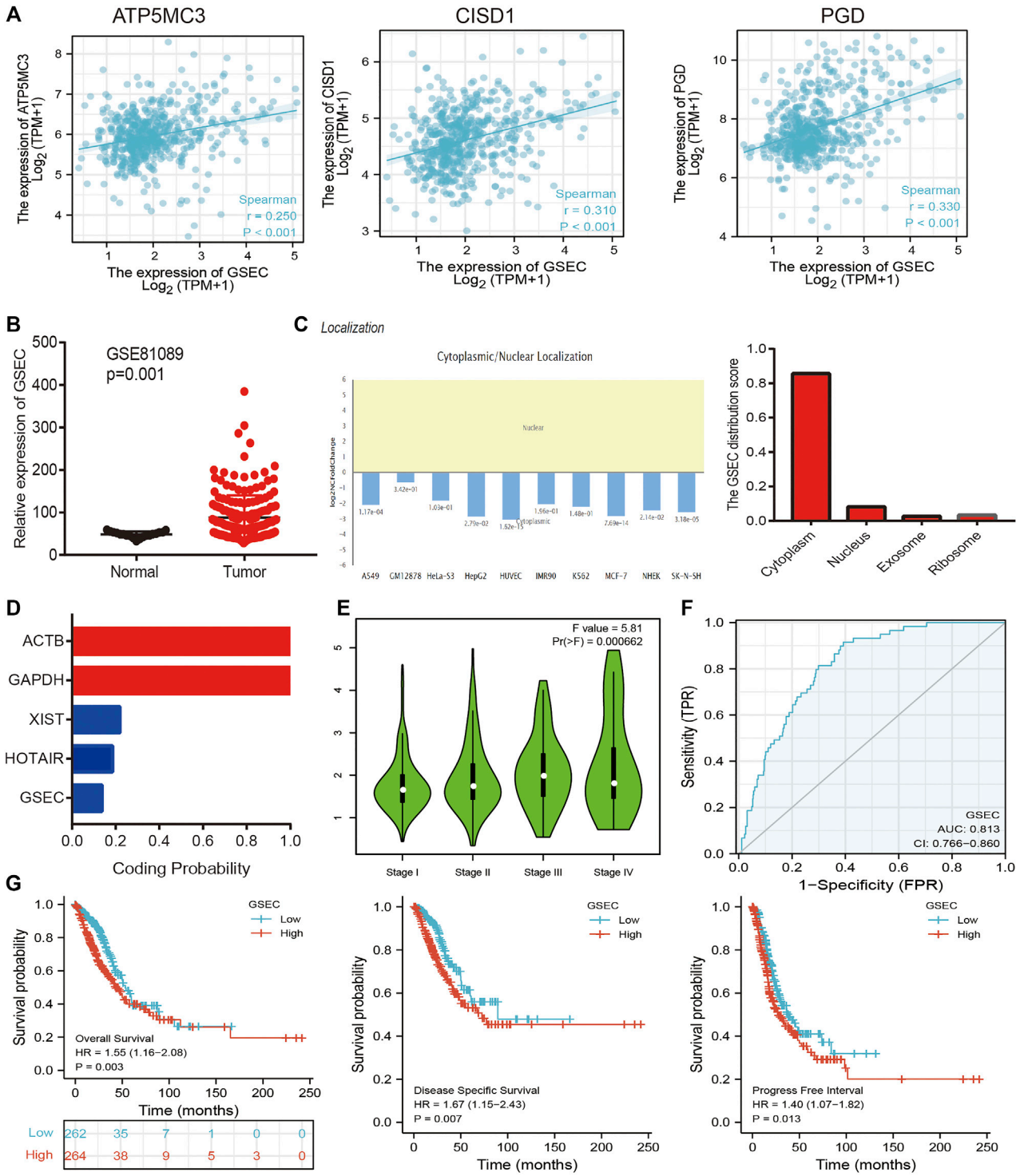


FIGURE 9 | Construction of lncRNA/miRNA/FRG interaction network. **(A)** Pearson's correlation analysis determined the correlation between GSEC and ATP5MC3, CISD1, and PGD in LUAD. **(B)** The expression of GESC in LUAD is examined by the GEO dataset. **(C)** The subcellular location of GESC is examined by lncLocator and Annolnc2 databases. **(D)** The coding ability of GESC is examined by coding potential calculator databases. **(E)** The pathologic stage of GESC in LUAD is examined by GEPIA databases. **(F)** ROC curve analyses and AUC values for GSEC in LUAD. **(G)** The prognosis of GESC in LUAD is examined by GEPIA databases. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

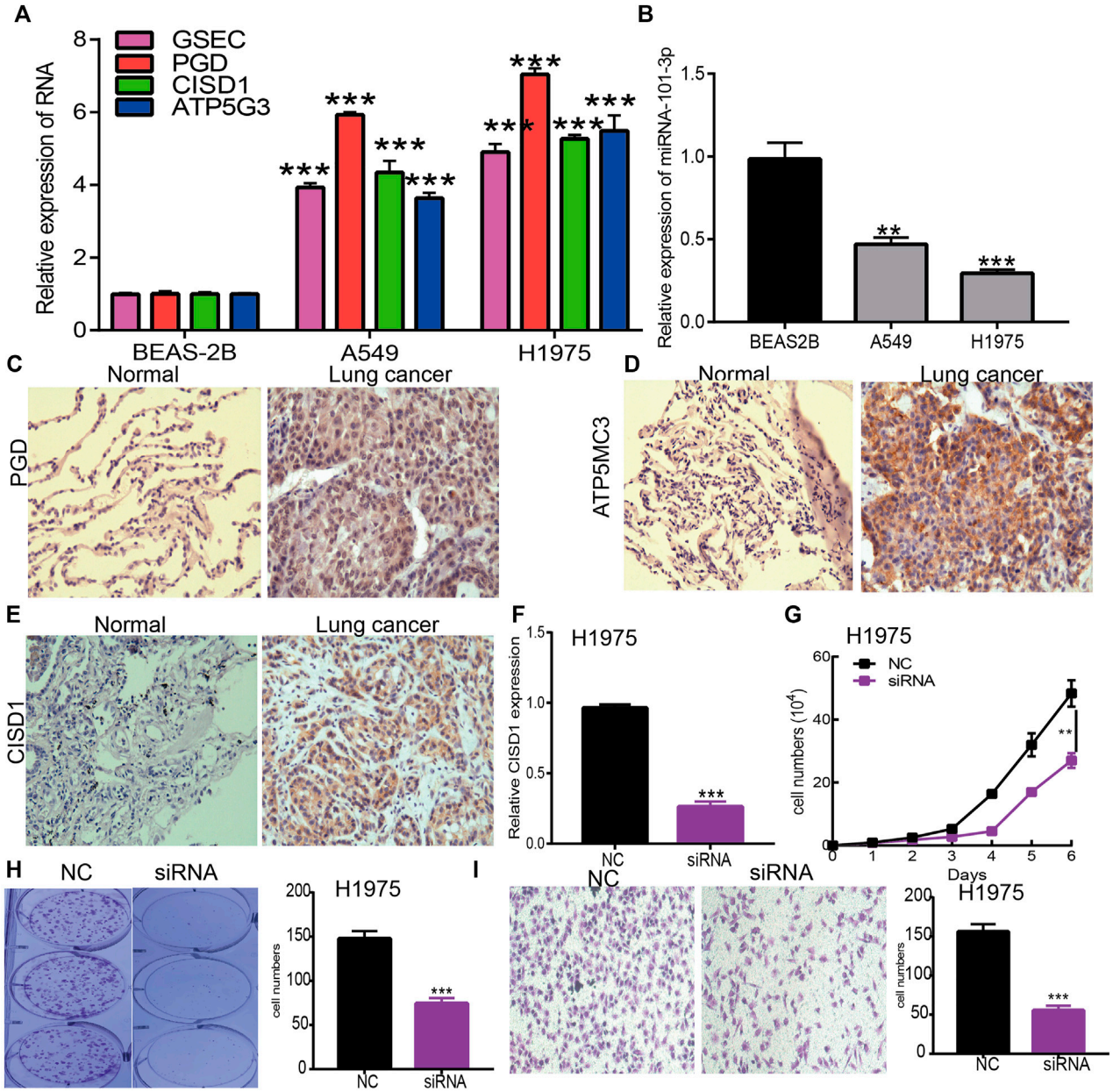


FIGURE 10 | Depletion of CISD1 inhibits LUAD cell proliferation and migration. **(A,B)** The expression of GSEC, FRGs, and miRNA-101-3p in LUAD cell lines was examined by qRT-PCR assay. **(C–E)** The expression of PGD, ATP5MC3, and CISD1 in LUAD tissue was examined by IHC assay. **(F)** The expression of CISD1 in H1975 cells after knockdown of CISD1 was examined by qRT-PCR assay. **(G,H)** The growth curve and colony formation assays were utilized to detect the silencing of CISD1 on the growth of H1975 cells. **(I)** The transwell assay was utilized to detect the silencing of CISD1 on the migration of H1975 cells. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

miRNA-101-3p axis modulates ATP5MC3, PGD, and CISD1 expression in LUAD.

Knockdown of CISD1 Inhibits LUAD Proliferation and Migration

To verify the aforementioned findings, a qRT-PCR assay was used to detect GSEC, ATP5MC3, PGD, CISD1, and miRNA-101-3p expression in LUAD cell lines. While we found GSEC, ATP5MC3, PGD, and CISD1 expression to be upregulated in LUAD cells (**Figure 10A**), miRNA-101-3p expression was noted to be decreased (**Figure 10B**). The immunohistochemical evaluation further revealed PGD, ATP5MC3, and CISD1 overexpression in LUAD tissues (**Figures 10C–E**). As prior literature has detailed the roles of PGD and ATP5MC3 in LUAD, we focused on further studying the influence of CISD1 on LUAD cell proliferation and migration *via* the utilization of target siRNA transfected into H1975 cells. We found that CISD1

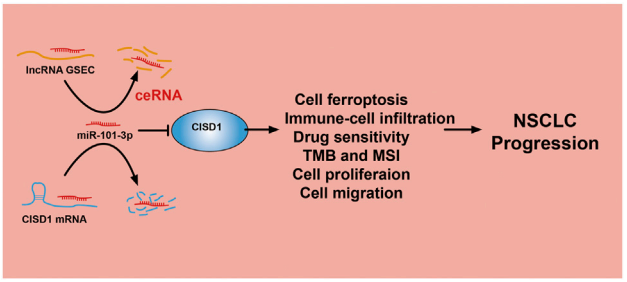


FIGURE 11 | A working model for lncRNA GSEC/miR-101-3p/CISD1 axis in LUAD.

expression was decreased after CISD1 knockdown in this cell line (**Figure 10F**). Growth curve and colony formation findings revealed that CISD1 knockdown significantly inhibits cell proliferation (**Figures 10G,H**). Transwell assays indicated that the migratory capability of H1975 cells was significantly suppressed *via* CISD1 downregulation (**Figure 10I**). These findings confirm that CISD1 functions as an oncogene, promoting cell growth and migration among LUAD cells.

DISCUSSION

An iron-dependent type of programmed cell death, ferroptosis differs from other forms of cell death such as apoptosis, necrosis, and autophagy. The ferroptosis process generally results in excessive lipid peroxidation and is induced by abnormalities in cellular redox processes. Importantly, ferroptosis is understood to play an indispensable role in the elimination of carcinogenic cells (Hirschhorn and Stockwell, 2019). Activation of Ras/mitogen-activated protein kinase signaling increases the sensitivity of cancer cells to ferroptosis and thus leads to the maintenance of relative iron abundance in cancer *via* control of transferrin receptor and ferritin expression (Dixon et al., 2012). Knockdown of FMS-like tyrosine kinase three was reported to reduce lipid peroxidation and inhibit p22phox activity, resulting in the inhibition of ferroptosis among malignant cells (Song et al., 2018). Recently, 5-monophosphate was reported to increase Beclin1 phosphorylation and inhibit system Xc activity, thus facilitating ferroptosis (Kang et al., 2018). However, the modulatory mechanism of the ferroptosis process and associated FRG expression correlated with LUAD progression remains to be elucidated.

Here, we first investigated the expression and prognostic value of FRGs in LUAD. We found a total of 26 genes to be upregulated; seven genes were found to be downregulated in LUAD as compared with normal tissue. Meanwhile, the expression of FDFT1, IREB2, NFS1, TFRC, HSBP1, ACO1, LPCAT3, CARS, GSS, ACACA, ACSL3, CISD1, SLC1A5, and NCOA4 was found to positively correlate with respective CNV in LUAD. Levels of DNA methylation were found to negatively associate with CBS, FDFT1, STEAP3, FADS2, SCL1A5, GCLM, GOT1, SLC7A11, CD44, ACSL3, and G6PD expression in LUAD. Collectively, these data indicate that CNV and DNA methylation exert critical effects on FRG expression in LUAD.

Functional enrichment analysis revealed 33 FRGs to be primarily involved in vascular fluid dynamics and atherosclerosis, HIF-1a signaling, lipid metabolism, ferroptosis, cysteine and methionine metabolism, glutathione metabolism, amino acid biosynthesis, necroptosis, arachidonic acid metabolism, thyroid hormone synthesis, serotonergic communication, alanine metabolism, aspartate metabolism, glutamate metabolism, inflammatory bowel disease, 2-monocarboxylic acid metabolism, and arginine biosynthesis. Our findings demonstrate that FRGs play an important role in both LUAD pathogenesis and illness progression.

Prognosis analysis revealed increased SLC7A11, PGD, FANCD2, CISD1, ATP5G3, and ASCL3 expression to closely correlate with poor LUAD clinical prognosis. Importantly, levels of ATP5G3, CISD1, and PGD expression were found to positively correlate with tumor stage. Analysis ROC curve data revealed SLC7A11, FANCD2, CISD1, and ATP5G3 to be highly accurate markers in LUAD (AUC > 0.8). These findings suggest that FRG expression correlates with LUAD prognosis. A highly accurate five-gene prognostic model capable of predicting overall LUAD patient survival was constructed based on SLC7A11, PGD, FANCD2, SIDS1, ATP5G3, and ASCL3 using Lasso Cox regression analysis. A predictive nomogram useful in predicting 3- and 5-year overall survival rates as compared to an ideal model in the studied cohort was similarly developed. Here, we identified a pyroptosis-related prognostic gene signature for LUAD, further advancing potential prognostic methods relevant to LUAD.

Here, SLC7A11, PGD, FANCD2, CIDS1, ATP5G3, and ASCL3 were found to serve as effective gene signatures for predicting LUAD prognosis. An important cystine/glutamate antiporter regulated by p53, SLC7A11 plays suppressive roles in the ferroptosis process. The SLC7A11-mediated importation of cystine enhances glutathione biosynthesis and increases GPX4-mediated detoxification of lipid peroxides, thereby inhibiting ferroptosis (Stockwell and Jiang, 2020). A prior study reported PGD capable of predicting overall survival in the setting of papillary thyroid carcinoma (Yang et al., 2021). In addition, FANCD2 was reported to inhibit ferroptosis and protect from bone marrow injury (Song et al., 2016).

Importantly, we found FRG expression to significantly correlate with tumor immune infiltration. Furthermore, we found that SLC7A11, PGD, CISD1, ATP5MC3, and ACSL3 expression significantly negatively correlates with CD274, CTLA4, HAVCR2, LAG3, PDCD1, PDCD1LG12, TIGIT, and SIGLEC15 expression. However, FANCD2 expression was found to positively correlate with checkpoint-related gene expression in LUAD. As such, these data confirm that FRG expression is associated with immune checkpoint-related gene expression in LUAD.

The lncRNA GSEC was previously reported to upregulate EIF5A2 expression *via* miR-588 sponging and subsequently facilitate cell proliferation, migration, and invasion capabilities in osteosarcoma (Liu et al., 2020). High levels of GSEC expression were reported in triple-negative breast cancer tissue and cell lines,

with GSEC knockdown significantly decreasing these proliferative, migratory, and invasive capabilities. The sponging of miR-202-5p by GSEC, along with resultant increased AXL expression, was reported to enhance LUAD progression (Zhang et al., 2021). Similarly, miRNA-101-3p was reported to inhibit both proliferative and migratory capabilities of gastric carcinoma cells and enhance apoptosis *via* PIM-1 expression modulation (Wu et al., 2019). Indeed, in an LUAD setting of miRNA-101-3p downregulation, overexpression of miR-101-3p mimics was reported to reduce rates of tumor growth and progression *in vitro* (Liu et al., 2021). In this study, data obtained from various public databases were used to illustrate novel mRNA–miRNA–lncRNA interactions (namely, the lncRNA GSEC/miR-101-3p/ATP5MC3/PGD/CISD1 axis) detailing modulation of FRG expression following the ceRNA hypothesis (**Figure 11**). Although we found GSEC, ATP5MC3, PGD, and CISD1 to be upregulated in LUAD cell lines, miRNA-101-3p expression was found to be decreased. The immunohistochemical evaluation also confirmed increased expression of ATP5MC3, PGD, and CISD1 in LUAD tissue. Knockdown of CISD1 was found to significantly inhibit the proliferative and migratory capabilities of LUAD cells.

CONCLUSION

Here, we found that GSEC, CISD1, ATP5MC3, and PGD are upregulated, and miRNA-101-3p is downregulated, in LUAD. Knockdown of CISD1 was found to significantly inhibit the proliferative and migratory capabilities of LUAD cells. In summary, this study characterizes the functional roles of the lncRNA GSEC/miR-101-3p/CISD1 axis in LUAD and suggests potential diagnostic and therapeutic biomarkers for future clinical application in the management of this illness.

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.

AUTHOR CONTRIBUTIONS

XJ and YY designed this work, performed the related assay, and analyzed the data. LT, JW, and DZ contributed to the study materials. LD supervised and wrote the manuscript. All authors have read and approved the final version of the manuscript.

FUNDING

This work was supported by the National Nature Science Foundation of China (82160508) and Yunnan Applied Basic Research Projects (YNWRMY-2019-067, 2019FE001) and Yunnan Province Specialized Training Grant for High-Level Healthcare Professionals (D-201614).

ACKNOWLEDGMENTS

The authors would like to thank support from the Department of Thoracic Surgery, The Third Affiliated Hospital of Kunming Medical University (Yunnan Tumor Hospital), Kunming, China.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.793732/full#supplementary-material

REFERENCES

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

Basu, A., Bodycombe, N. E., Cheah, J. H., Price, E. V., Liu, K., Schaefer, G. I., et al. (2013). An Interactive Resource to Identify Cancer Genetic and Lineage Dependencies Targeted by Small Molecules. *Cell* 154 (5), 1151–1161. doi:10.1016/j.cell.2013.08.003

Cao, Z., Pan, X., Yang, Y., Huang, Y., and Shen, H.-B. (2018). The lncLocator: a Subcellular Localization Predictor for Long Non-coding RNAs Based on a Stacked Ensemble Classifier. *Bioinformatics* 34 (13), 2185–2194. doi:10.1093/bioinformatics/bty085

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 (5), 1060–1072. doi:10.1016/j.cell.2012.03.042

Feng, H., Schorpp, K., Jin, J., Yozwiak, C. E., Hoffstrom, B. G., Decker, A. M., et al. (2020). Transferrin Receptor Is a Specific Ferroptosis Marker. *Cel Rep.* 30 (10), 3411–3423. doi:10.1016/j.celrep.2020.02.049

Goodman, A. M., Sokol, E. S., Frampton, G. M., Lippman, S. M., and Kurzrock, R. (2019). Microsatellite-Stable Tumors with High Mutational Burden Benefit from Immunotherapy. *Cancer Immunol. Res.* 7 (10), 1570–1573. doi:10.1158/2326-6066.cir-19-0149

Hirschhorn, T., and Stockwell, B. R. (2019). The Development of the Concept of Ferroptosis. *Free Radic. Biol. Med.* 133, 130–143. doi:10.1016/j.freeradbiomed.2018.09.043

Huang, J., Liu, D., Wang, Y., Liu, L., Li, J., Yuan, J., et al. (2021). Ginseng Polysaccharides Alter the Gut Microbiota and Kynurenine/tryptophan Ratio, Potentiating the Antitumour Effect of Antiprogrammed Cell Death 1/programmed Cell Death Ligand 1 (Anti-PD-1/pd-L1) Immunotherapy. *Gut* 2021 [Epub ahead of print], 321031. doi:10.1136/gutjnl-2020-321031

Ito, K., and Murphy, D. (2013). Application of Ggplot2 to Pharmacometric Graphics. *CPT Pharmacometrics Syst. Pharmacol.* 2 (10), e79. doi:10.1038/psp.2013.56

Jiang, L.-P., Fan, S.-Q., Xiong, Q.-X., Zhou, Y.-C., Yang, Z.-Z., Li, G.-F., et al. (2018). GRK5 Functions as an Oncogenic Factor in Non-small-cell Lung Cancer. *Cell Death Dis* 9 (3), 295. doi:10.1038/s41419-018-0299-1

Kang, R., Kroemer, G., and Tang, D. (2019). The Tumor Suppressor Protein P53 and the Ferroptosis Network. *Free Radic. Biol. Med.* 133, 162–168. doi:10.1016/j.freeradbiomed.2018.05.074

Kang, R., Zhu, S., Zeh, H. J., Klionsky, D. J., and Tang, D. (2018). BECN1 Is a New Driver of Ferroptosis. *Autophagy* 14 (12), 2173–2175. doi:10.1080/15548627.2018.1513758

Kang, Y.-J., Yang, D.-C., Kong, L., Hou, M., Meng, Y.-Q., Wei, L., et al. (2017). CPC2: a Fast and Accurate Coding Potential Calculator Based on Sequence Intrinsic Features. *Nucleic Acids Res.* 45 (W1), W12–w16. doi:10.1093/nar/gkx428

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

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 (21), e108–e110. doi:10.1158/0008-5472.can-17-0307

Lin, W., Chen, Y., Wu, B., chen, Y., and Li, Z. (2021). Identification of the Pyroptosis-related P-rognostic G-ene S-ignature and the A-ssociated R-egulation axis in L-ung A-denocarcinoma. *Cell Death Discov.* 7 (1), 161. doi:10.1038/s41420-021-00557-2

Liu, C.-J., Hu, F.-F., Xia, M.-X., Han, L., Zhang, Q., and Guo, A.-Y. (2018). GSCALite: a Web Server for Gene Set Cancer Analysis. *Bioinformatics* 34 (21), 3771–3772. doi:10.1093/bioinformatics/bty411

Liu, R., Ju, C., Zhang, F., Tang, X., Yan, J., Sun, J., et al. (2020). LncRNA GSEC Promotes the Proliferation, Migration and Invasion by Sponging miR-588/EIF5A2 axis in Osteosarcoma. *Biochem. Biophysical Res. Commun.* 532 (2), 300–307. doi:10.1016/j.bbrc.2020.08.056

Liu, S.-H., Hsu, K.-W., Lai, Y.-L., Lin, Y.-F., Chen, F.-H., Peng, P.-H., et al. (2021). Systematic Identification of Clinically Relevant miRNAs for Potential miRNA-Based Therapy in Lung Adenocarcinoma. *Mol. Ther. - Nucleic Acids* 25, 1–10. doi:10.1016/j.omtn.2021.04.020

Luo, L., Zheng, Y., Lin, Z., Li, X., Li, X., Li, M., et al. (2021). Identification of SHMT2 as a Potential Prognostic Biomarker and Correlating with Immune Infiltrates in Lung Adenocarcinoma. *J. Immunol. Res.* 2021, 6647122. doi:10.1155/2021/6647122

Mao, C., Wang, X., Liu, Y., Wang, M., Yan, B., Jiang, Y., et al. (2018). A G3BP1-Interacting lncRNA Promotes Ferroptosis and Apoptosis in Cancer via Nuclear Sequestration of P53. *Cancer Res.* 78 (13), 3484–3496. doi:10.1158/0008-5472.CAN-17-3454

Mou, Y., Wang, J., Wu, J., He, D., Zhang, C., Duan, C., et al. (2019). Ferroptosis, a New Form of Cell Death: Opportunities and Challenges in Cancer. *J. Hematol. Oncol.* 12 (1), 34. doi:10.1186/s13045-019-0720-y

Rivera, G. A., and Wakelee, H. (2016). Lung Cancer in Never Smokers. *Adv. Exp. Med. Biol.* 893, 43–57. doi:10.1007/978-3-319-24223-1_3

Schwartz, A. G., and Cote, M. L. (2016). Epidemiology of Lung Cancer. *Adv. Exp. Med. Biol.* 893, 21–41. doi:10.1007/978-3-319-24223-1_2

Song, X., Xie, Y., Kang, R., Hou, W., Sun, X., Epperly, M. W., et al. (2016). FANCD2 Protects against Bone Marrow Injury from Ferroptosis. *Biochem. Biophysical Res. Commun.* 480 (3), 443–449. doi:10.1016/j.bbrc.2016.10.068

Song, X., Zhu, S., Chen, P., Hou, W., Wen, Q., Liu, J., et al. (2018). AMPK-mediated BECN1 Phosphorylation Promotes Ferroptosis by Directly Blocking System Xc- Activity. *Curr. Biol.* 28 (15), 2388–2399. doi:10.1016/j.cub.2018.05.094

Stockwell, B. R., and Jiang, X. (2020). The Chemistry and Biology of Ferroptosis. *Cel Chem. Biol.* 27 (4), 365–375. doi:10.1016/j.chembiol.2020.03.013

Su, Y., Zhao, B., Zhou, L., Zhang, Z., Shen, Y., Lv, H., et al. (2020). Ferroptosis, a Novel Pharmacological Mechanism of Anti-cancer Drugs. *Cancer Lett.* 483, 127–136. doi:10.1016/j.canlet.2020.02.015

Wang, M., Mao, C., Ouyang, L., Liu, Y., Lai, W., Liu, N., et al. (2019). Long Noncoding RNA LINC00336 Inhibits Ferroptosis in Lung Cancer by Functioning as a Competing Endogenous RNA. *Cell Death Differ* 26 (11), 2329–2343. doi:10.1038/s41418-019-0304-y

Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., et al. (2010). The GeneMANIA Prediction Server: Biological Network Integration for Gene Prioritization and Predicting Gene Function. *Nucleic Acids Res.* 38, W214–W220. doi:10.1093/nar/gkq537

Wu, F., Huang, W., Yang, L., and Xu, F. (2019). MicroRNA-101-3p Regulates Gastric Cancer Cell Proliferation, Invasion and Apoptosis by Targeting PIM 1 Expression. *Cel Mol Biol (Noisy-le-grand)* 65 (7), 118–122. doi:10.14715/cmb/2019.65.7.20

Xiong, Q., Jiang, L., Liu, K., Jiang, X., Liu, B., Shi, Y., et al. (2021). miR-133b Targets NCAPH to Promote β-catenin Degradation and Reduce Cancer Stem Cell Maintenance in Non-small Cell Lung Cancer. *Sig Transduct Target. Ther.* 6 (1), 252. doi:10.1038/s41392-021-00555-x

Xu, G., Wang, H., Li, X., Huang, R., and Luo, L. (2021). Recent Progress on Targeting Ferroptosis for Cancer Therapy. *Biochem. Pharmacol.* 190, 114584. doi:10.1016/j.bcp.2021.114584

Xu, P., Jiang, L., Yang, Y., Wu, M., Liu, B., Shi, Y., et al. (2020). PAQR4 Promotes Chemoresistance in Non-small Cell Lung Cancer through Inhibiting Nrf2 Protein Degradation. *Theranostics* 10 (8), 3767–3778. doi:10.7150/thno.43142

Yang, D., Wang, J., Li, C., Shi, L., and Zhang, M. (2021). Ferroptosis-related Gene Model to Predict Overall Survival of Papillary Thyroid Carcinoma. *Am. J. Otolaryngol.* 42 (6), 103163. doi:10.1016/j.amjoto.2021.103163

Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a Resource for Therapeutic Biomarker Discovery in Cancer Cells. *Nucleic Acids Res.* 41, D955–D961. doi:10.1093/nar/gks1111

Yang, W. S., and Stockwell, B. R. (2016). Ferroptosis: Death by Lipid Peroxidation. *Trends Cel Biol.* 26 (3), 165–176. doi:10.1016/j.tcb.2015.10.014

Zhang, J., Du, C., Zhang, L., Wang, Y., Zhang, Y., and Li, J. (2021). lncRNA GSEC Promotes the Progression of Triple Negative Breast Cancer (TNBC) by Targeting the miR-202-5p/AXL Axis. *Ott* Vol. 14, 2747–2759. doi:10.2147/ott.s293832

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 Molecular Biosciences

ORIGINAL RESEARCH
published: 09 March 2022
doi: 10.3389/fmolb.2022.840412

The Prognostic Value of Lysine Acetylation Regulators in Hepatocellular Carcinoma

Liying Sun[1,2†], *Jian Zhang*[3†], *Kai Wen*[4], *Shenglan Huang*[1,2], *Dan Li*[1,2], *Yongkang Xu*[1,2] *and Jianbing Wu*[1,2*]

[1]Department of Oncology, The Second Affiliated Hospital of Nanchang University, Nanchang, China, [2]Jiangxi Key Laboratory of Clinical and Translational Cancer Research, Second Affiliated Hospital of Nanchang University, Nanchang, China, [3]Department of Orthopedics, The Second Affiliated Hospital of Nanchang University, Nanchang, China, [4]Department of Hepatobiliary Surgery, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, Guangzhou, China

Background: Hepatocellular carcinoma (HCC) is a tumor with high morbidity and mortality worldwide. lysine acetylation regulators (LARs) dynamically regulate Lysine acetylation modification which plays an important regulatory role in cancer. Therefore, we aimed to explore the potential clinical prognostic value of LARs in HCC.

Methods: Differentially expressed LARs in normal liver and HCC tissues were obtained from The Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC) datasets. To identify genes with prognostic value and establish the risk characteristics of LARs, consensus clustering was employed. We used univariate Cox regression survival analysis and LASSO Cox regression based on LARs to determine the independent prognostic signature of HCC. CIBERSORT and Gene Set Enrichment Analysis (GSEA) were used to estimate immune infiltration and functional enrichment analysis respectively. The expression of LAR was detected by Real-time quantitative polymerase chain reaction (RT-qPCR). statistical analyses were conducted using SPSS and R software.

Results: In this study, the 33 LARs expression data and corresponding clinical information of HCC were obtained using TCGA and ICGC datasets. We found majority of the LARs were differentially expressed. Consensus cluster analysis was carried out based on the TCGA cohort, and three HCC subtypes (cluster 1, 2, and 3) were obtained. The LA3 subgroup had the worst clinical outcomes. Nine key LARs were identified to affect prognosis. The results showed that LARs signature has a strong independent prognostic value in HCC patients, whether in the training datasets or in the testing datasets. GSEA results showed that various tumor-related processes and pathways

OPEN ACCESS

Edited by:
William C. Cho,
QEH, Hong Kong SAR, China

Reviewed by:
Adrian Drazic,
University of Bergen, Norway
Wenzhi Guo,
First Affiliated Hospital of Zhengzhou University, China

****Correspondence:***
Jianbing Wu
hhgwjb@163.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *21 December 2021*
Accepted: *21 February 2022*
Published: *09 March 2022*

Citation:
Sun L, Zhang J, Wen K, Huang S, Li D, Xu Y and Wu J (2022) The Prognostic Value of Lysine Acetylation Regulators in Hepatocellular Carcinoma. Front. Mol. Biosci. 9:840412. doi: 10.3389/fmolb.2022.840412

Abbreviations: AUC, area under the curve; ATCC, American Type Culture Collection; CDF, cumulative distribution function; DMEM, Dulbecco's modified Eagle medium; FBS, fetal bovine serum; GSEA, gene set enrichment analyses; HCC, hepatocellular carcinoma; HDACs, histone deacetylases; HR, hazard ratio; ICGC, International Cancer Genome Consortium; KATs, lysine acetyltransferases; LARs, lysine acetylation regulators; LASSO, least absolute shrinkage and selection operator; OS, overall survival; PTMs, post-translational modification; PPI, protein-protein interaction; PCA, principal component analysis; ROC, receiver operating characteristic; TCGA, The Cancer Genome Atlas; TMB, tumor mutation load; t-SNE, t-distributed stochastic neighbor embedding.

were abundant in the high-risk groups. RT-qPCR results showed that HAT1, HDAC1, HDAC2, HDAC4, and HDAC11 were highly expressed in HCC cells.

Conclusion: Our results suggest that LARs play critical roles in HCC and are helpful for individual prognosis monitoring and clinical decision-making of HCC.

Keywords: lysine acetylation regulators, hepatocellular carcinoma, prognostic signature, overall survival, nomogram

INTRODUCTION

Hepatocellular carcinoma (HCC) is one of the most common malignant tumors worldwide, ranking sixth in cancer incidence and fourth in mortality worldwide (Kulik and El-Serag, 2019; Huang et al., 2021). HCC risk factors include viral infections, obesity, nonalcoholic fatty liver disease, and high intake of aflatoxin (Caruso et al., 2021). Progress has been made in recent years regarding HCC, and the clinical prognosis of HCC has improved to a certain extent.(Fan, 2012). However, due to the high heterogeneity and high incidence of recurrence and metastasis, the overall prognosis with HCC remains unsatisfactory (Nault et al., 2020). Therefore, it is urgent to need to develop novel therapeutic targets for HCC patients, and act as effective predictive markers to evaluate HCC patient prognosis.

Cancer is a multifactorial disease that results from the interaction between genetic abnormalities and epigenetic changes. Epigenetic changes are an important sign of cancer progression (Baylin and Jones, 2016; Feinberg et al., 2016; Cavalli and Heard, 2019). Post-translational modification (PTM) is an important way to regulate protein function and affect cell behavior, and act as a signal marker in cancer cells (Chen et al., 2020; Figlia et al., 2020). Lysine acetylation is an important reversible and dynamic protein PTM that is exceedingly important for gene expression. It plays an important role in transcription factor activity, chromatin remodeling, and metabolic enzyme activity, and is related to tumorigenesis, tumor progression, and metastasis (Choudhary et al., 2014; Kaypee et al., 2016; Narita et al., 2019). Lysine acetylation is a reversible epigenome modification regulated by two clusters of opposing enzymes: lysine acetyltransferases (KATs) and histone deacetylases (HDACs) (Falkenberg and Johnstone, 2014; Sheikh and Akhtar, 2019). KATs are primarily divided into two types according to their cellular localization and acetylated chromatin histone ability: type a KATs and type b KATs. The main KAT families are the GCN5-related N-acetyltransferase (GNAT) (Bienvenut et al., 2020), p300/CBP (KAT3) (Lipinski et al., 2020)and MYST (MOZ, Ybf2, Sas2, and TIP60) families (Baell et al., 2018). Moreover, two other KAT families belong to the nuclear receptor family of transcription factor-related KATs and KATs. HDACs are roughly divided into classical HDACs, including type I (homologues of yeast Rpd3, including HDAC 1, 2, 3, and 8), II (homologues of yeast Hda1, including HDAC 4, 5, 6, 7, 9, 10), IV (HDAC 11), and NAD + -dependent III HDAC or sirtuins, similar to yeast Sir2 (Li et al., 2020). Although the molecular regulation mechanism of LARs in HCC has been discussed, there is a lack of comprehensive research (Chai et al., 2021). Therefore, we systematically studied the role of LARs in HCC in order to provide potential prognostic markers and therapeutic targets.

In this study, we grouped 368 HCC patients based on LARs and identified three subgroups with differences in prognosis. The risk score be calculated by the LASSO-Cox regression to establish a LARs signature. We also discuss the relationship between the LAR risk model and immunity. Finally, we performed experiments to verify the LARs mRNA expression. Our findings reveal the possible role of LARs in HCC, which is of substantial significance for accurate HCC treatment.

MATERIALS AND METHODS

Data Acquisition

RNA-seq data for HCC samples were selected from the TCGA dataset (https://portal.gdc.cancer.gov/), and clinical data related to the patient were downloaded simultaneously. The study included 368 HCC and 50 normal liver tissue samples. The RNA-seq data of the verification set and the corresponding clinicopathological information were downloaded from ICGC dataset (https://dcc.icgc.org/), including 232 HCC samples and 202 normal samples adjacent to cancer. We obtained 33 LARs were obtained from previously published literature (Narita et al., 2019), of which 13 belonged to the acetyltransferase family, and 20 belonged to the deacetylase class. The 33 LARs information is provided in **Supplementary Table S1**.

Analysis of the LARs Regulatory Factors in HCC

The differential expression of LARs in HCC and normal liver tissue samples was analyzed with the "limma" R package and visualized by heatmap. The protein-protein interactions (PPI) network between differentially expressed LARs were analyzed with the STRING website (http://string-db.org/). In addition, the correlation between these LARs was calculated using Pearson correlation network analysis.

Consensus Clustering Analysis

In the TCGA cohort, the expression of 33 LARs were analyzed to determine the HCC subtype by the "ConsensusClusterPlus" R package (Wilkerson and Hayes, 2010). The method was based on the classical K-Means algorithm, with the Euclidean distance,

iterated 50 times, and 80% of tumor samples were taken in each iteration. The number of clusters was set to 2–9, and we determine the best cluster number through the clustering score for the cumulative distribution function (CDF). According to different subgroups, the survival curve was drawn with survival R packet. Combined with clinical characteristics, heat maps were used to analyze the expression and distribution of LARs in different subgroups.

Establishment and Verification of a Prognostic Risk Model

In the TCGA dataset, the survival information of HCC patients and the expression data of LARs were combined, and Cox regression analysis was performed to obtain the hazard ratio (HR) with its 95% confidence interval (CI) for each LARs. According to HR to identify protective genes (HR < 1) and dangerous genes (HR > 1), LARs with $p < 0.05$, were selected for the next analysis. According to the LARs related to prognosis, nine risk genes were identified using the LASSO algorithm, and the risk coefficient of each gene was obtained. We constructed a risk-scoring equation based on LAR expression:

$$Risk\ score = \sum_{i=1}^{n} (Coef_i * x_i)$$

where $Coef_i$ refers to the risk coefficient of the gene and x_i is the expression value of the gene.

Taking the median risk score as the cutoff value, HCC patients were classified into high-risk and low-risk group. Using the "survival" R package to perform Kaplan–Meier survival curve analysis, we drew a risk curve from low to high patient risk values, and the package "timeROC" was used to depict the ROC curve, calculate the area under the curve (AUC), and evaluate the accuracy, the sensitivity and the specificity of the model. Principal component analysis (PCA) and t-distribution stochastic neighbor embedding (t-SNE) methods were used to study the distribution of different populations by using R packages stats and Rtsne.

Construction and Verification of Predictive Nomogram

Clinicopathological characteristics and risk scores were integrated with survival data to obtain independent prognostic analysis input files, and The R package "survival" was used to perform univariate Cox and multivariate Cox regression analyses on the input files to evaluate the role of predictive model risk scores in prognostic prediction. The results were a forest map for visualization. Then, we used the "rms" R package to construct these independent prognostic factors, and constructed 1-, 3-, and 5-years forecast nomograms and corresponding calibration charts. The concordance index (C-index) was applied to evaluate the performance of the nomogram, and the calibration curve of the nomogram was used to evaluate the consistency between actual and predicted value.

Immune Infiltration Analysis

Based on the TCGA database, the "CIBERSORT" R package was used to calculate the infiltration proportions of 22 immune cells in each HCC patient and displayed with a barplot, heatmap, and violin chart. The correlation between the expression and distribution of 22 infiltrating immune cells in the two groups was analyzed, and the related heat map was drawn. Red denotes a positive correlation, blue indicates a negative correlation.

Gene Set Enrichment Analysis

Gene expression data were analyzed by Gene Set Enrichment Analysis (GSEA) software (http://www.gsea-msigdb.org/gsea/index.jsp). The related pathways and molecular mechanisms of HCC patients were evaluated using the "Hallmark" and the "KEGG" gene sets. The minimum and maximum gene set sizes were 10 and 500, respectively. The nominal (NOM) p-value < 0.05 and false discovery ratio (FDR) q value < 0.25 were considered statistically significant.

Cell Culture

Human hepatic cell line, THLE-2 was purchased from the American Type Culture Collection (MA, USA), and HCC cell line (MHCC-97H, HepG2, LM3) was purchased from the National Collection of Authenticated Cell Cultures (Shanghai, China). The cell lines were cultured in Dulbecco's modified Eagle medium (Gibco, TX, USA), supplemented with 10% fetal bovine serum (Gibco) in a humidified incubator (37°C, 5% CO_2).

Reverse-Transcription Quantitative PCR

Total RNA was extracted from cells using TRIzol reagent (Invitrogen, MA, USA). cDNA was synthesized using the PrimeScript RT reagent kit (Takara, Shiga, Japan) following the manufacturer's instructions. The samples were amplified by qPCR using SYBR Green qPCR Master Mix (Thermo Fisher Scientific, MA, USA). The primer sequence information is provided in **Supplementary Table S2**.

Statistical Analysis

All statistical analyses were conducted using R. 4.0.4 (https://www.r- project. org/) and SPSS Statistics software (version 25, https://www.ibm.com/products/software). The student's t-test was used to compare the differences between the two groups. A value of $p < 0.05$ indicated that a statistically significant difference.

RESULTS

Expression Characteristics and Interactions of LARs

The expression characteristics of 33 LARs were analyzed using TCGA and ICGC datasets. Compared with normal liver tissues, 23 same LARs were abnormally expressed in HCC in the ICGC and TCGA datasets (**Figures 1A,B**) Genes marked in red indicate

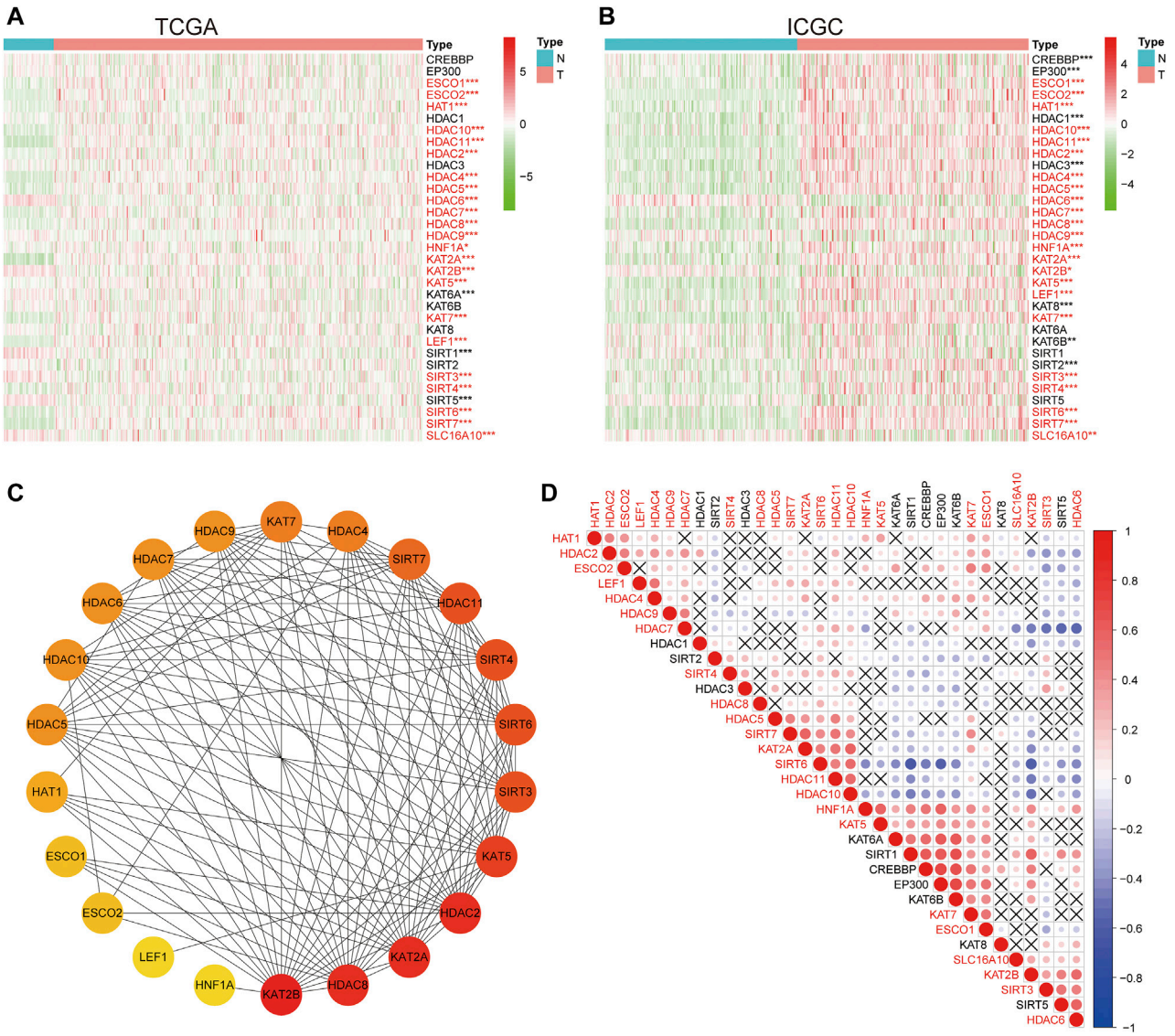


FIGURE 1 | The expression characteristics and correlation of LARs in hepatocellular carcinoma (HCC). **(A)** Heatmaps present the overall expression of LARs in HCC tissues and normal liver tissues from The Cancer Genome Atlas (TCGA). The red mark indicates that the gene is differentially expressed in the two data sets. **(B)** Heatmaps presented the overall expression of LARs in HCC tissues and normal liver tissues from the International Cancer Genome Consortium (ICGC). **(C)** STRING online tool was used to analyze the interaction of LARs regulatory factors. **(D)** Spearman correlation analysis of 33 regulatory genes in the HCC cohort. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

that they were differentially expressed in the two databases. Considering the similarity of biological functions of LARs, we analyzed the interaction and correlation between 33 LARs differentially expressed in the TCGA dataset. A PPI network analysis showed that these LARs frequently interact with each other, and HDACs and KATs have a particularly high interaction with other LARs (**Figure 1C**). **Figure 1D** shows the correlation of 33 LARs in HCC. We found that there was a significant correlation between the gene expression patterns of LARs in the same functional category, and there was a high correlation between acetyltransferase and deacetylase. In correlation analysis, there was a high correlation among KAT2A, HDAC10, SIRT6 and SIRT7, while SIRT6 was negatively correlated with the expression of KAT2B, HDAC6 and HDAC7 in HCC.

Identification of Hepatoma Subsets by Consensus Clustering

In order to establish the prognostic characteristics based on LARs, Pearson correlation coefficient analysis was carried out. Consensus clustering analysis was used to divide 368 samples into k groups (k = 2–9) (**Figures 2A,B**). According to the cumulative distribution function (CDF) value, k = 3 is the best number of clusters for dividing HCC queues, that is, cluster 1,

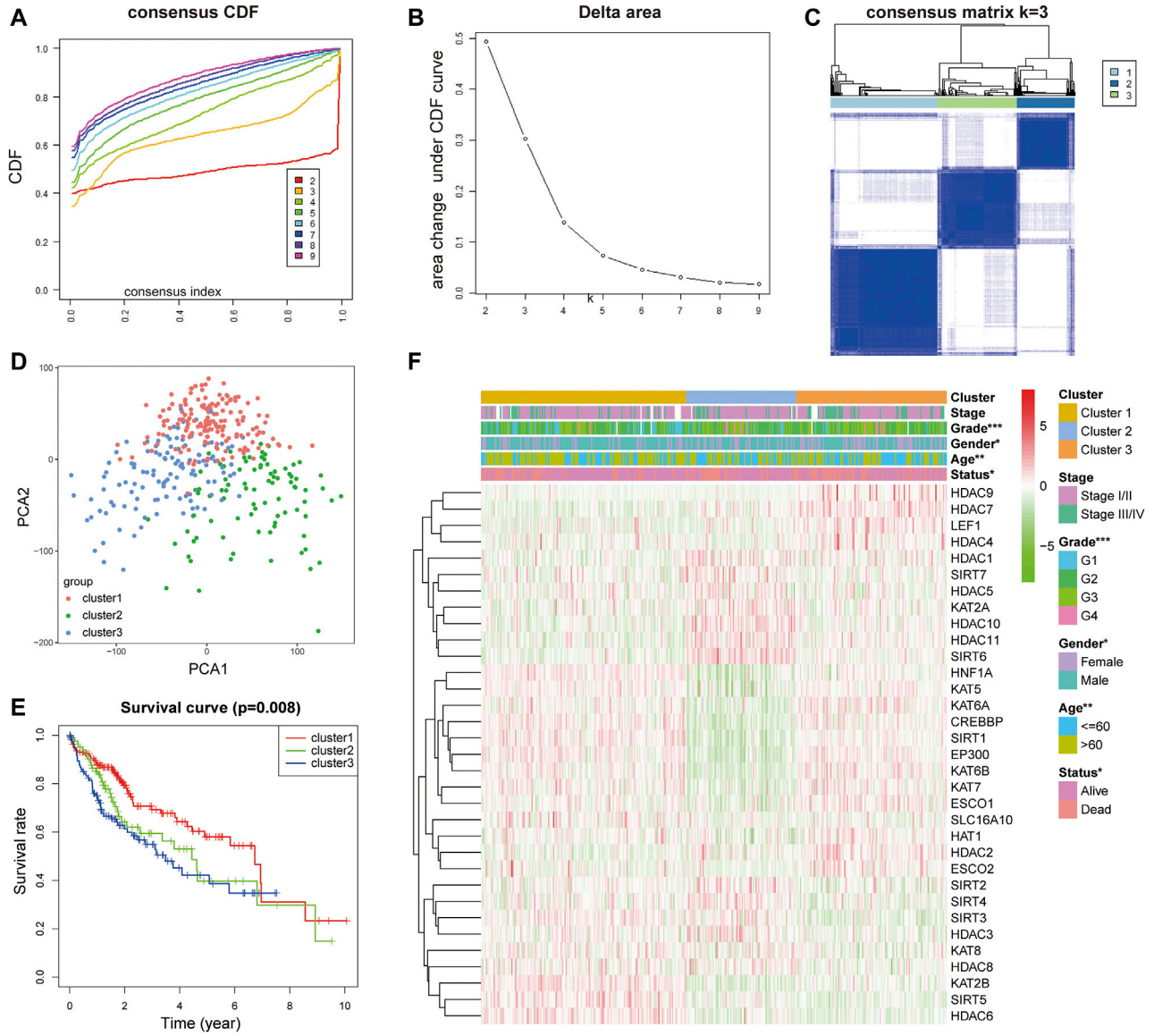


FIGURE 2 | (A) Consensus clustering cumulative distribution function (CDF) for k = 2 to 9. **(B)** Relative change in area under the CDF curve for k = 2 to 9. **(C)** Consensus clustering matrix for k = 3. **(D)** PCA of the total RNA expression profile in TCGA dataset. **(E)** Kaplan–Meier survival curves for HCC in the three subgroups defined by the consensus expression of 33 LARs. **(F)** The different expression levels of LARs and clinicopathological feature contributions.

cluster 2 and cluster 3 (**Figure 2C**). In order to understand the differences among the three subgroups in more detail, we used principal component analysis (PCA) to evaluate the classification of mRNA expression profiles, and the analysis showed that there were significant differences among the three subgroups (**Figure 2D**). In addition, the results of survival analysis showed that there was a great difference in OS time among the three clusters, and the OS of patients with cluster 3 was significantly lower than that of cluster 1 (**Figure 2E**). Subsequently, we evaluated the relationship between clusters and clinicopathological features (**Figure 2F**). There are differences in pathological characteristics such as grade, age, sex, and status among the three groups, and there are also differences in expression levels of LARs between the three clusters.

Building a LARs Signature Using LASSO Cox Regression

In the TCGA dataset, we performed univariate Cox regression analysis to explore the prognostic values of LARs. We found that 10 LARs were significantly associated with OS in HCC patients (**Figure 3A**). Among the 10 genes, ESCO2, HAT1, HDAC1, HDAC2, HDAC4, and HDAC11 are risk factors for HCC, with a risk ratio >1, while KAT8, HDAC6, SIRT3, and SIRT5 are protective factors, with a risk ratio <1. LASSO cox

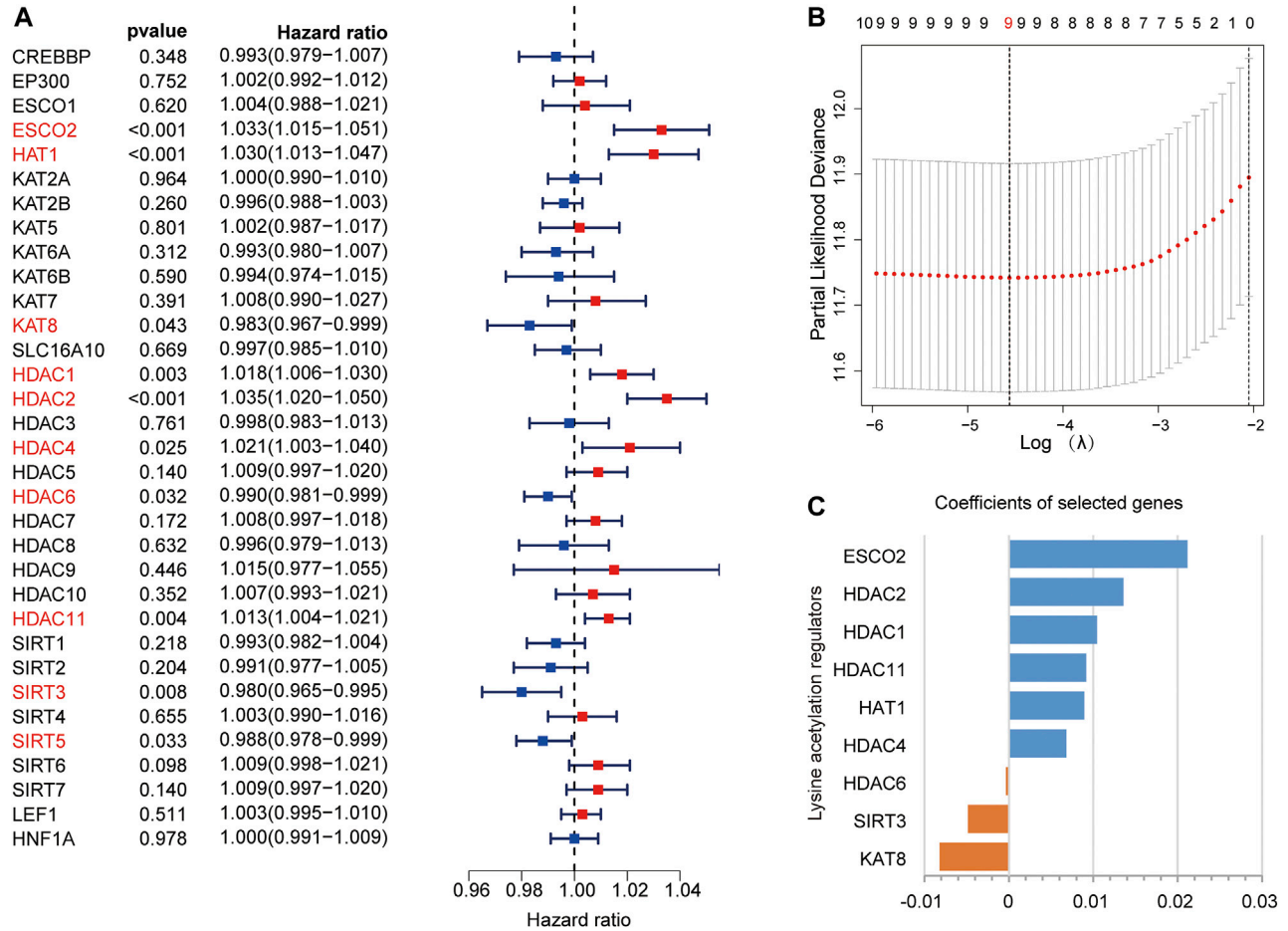


FIGURE 3 | (A) Univariate Cox regression analysis was used to calculate the hazard ratios (HRs), 95% confidence intervals, and *p*-values for screening the prognostic LARs. **(B)**, **(C)** LASSO regression was performed to calculate the minimum criteria **(B)** and coefficients **(C)**.

regression analysis was then used to build a prognostic model from the 10 LARs for HCC patients in the TCGA dataset. Based on the minimum standard, the prognostic features based on 9 LARs are successfully developed (**Figure 3B**, **Supplementary Figure S1A**). Nine LARs coefficients were visualized from high to low using the LASSO algorithm, as shown in **Figure 3C**. Detailed information is provided in **Supplementary Table S3**.

Evaluation and Validating of the Prognostic Model in TCGA Dataset

According to the risk score, HCC patients in the training group were divided into high-risk group and low-risk group. Survival analysis showed that patients in the high-risk group had worse survival compared with those in the low-risk group (**Figure 4A**). The increase in LARs-related gene expression and mortality in HCC patients were parallel to the increase in risk score (**Figure 4B**). The AUC values of 1-year, 2-years and 3-years calculated by TCGA dataset are 0.732, 0.701 and 0.695 respectively (**Figure 4C**), indicating that our LARs prognostic characteristics have a good predictive ability. PCA and t-SNE analysis revealed that distribution patterns of patients were significantly different in the different risk groups (**Figures 4D,E**). Then, Kaplan-Meier survival analysis was performed to investigate the prognostic value of a single LAR in HCC patients. The results showed that the expression of ESCO2, HDAC2, HDAC11, HDAC1, HAT1, HDAC6, and SIRT3 were found to be linked to the prognosis of HCC (**Figures 4F–N**). To evaluate the predictive value of LARs signature obtained from the training set, we used the ICGC database as external verification dataset. Using the same formula, the risk scores of the patients in the validation set were calculated and grouped. As shown in **Supplementary Figures S2A,B**, the survival time of patients in the high-risk group was significantly shorter than that in the low-risk group ($p < 0.001$). Moreover, the verification results showed that the AUC expression of LAR-related mRNA was 0.752, 0.656, and 0.685 at 1, 2, and 3 years, respectively (**Supplementary Figure S2C**). The PCA and t-SNE analysis successfully separated and confirmed two patient subgroups (**Supplementary Figures S2D–E**). Kaplan-Meier survival curve revealed that patients with high expression of ESCO2, HDAC2,

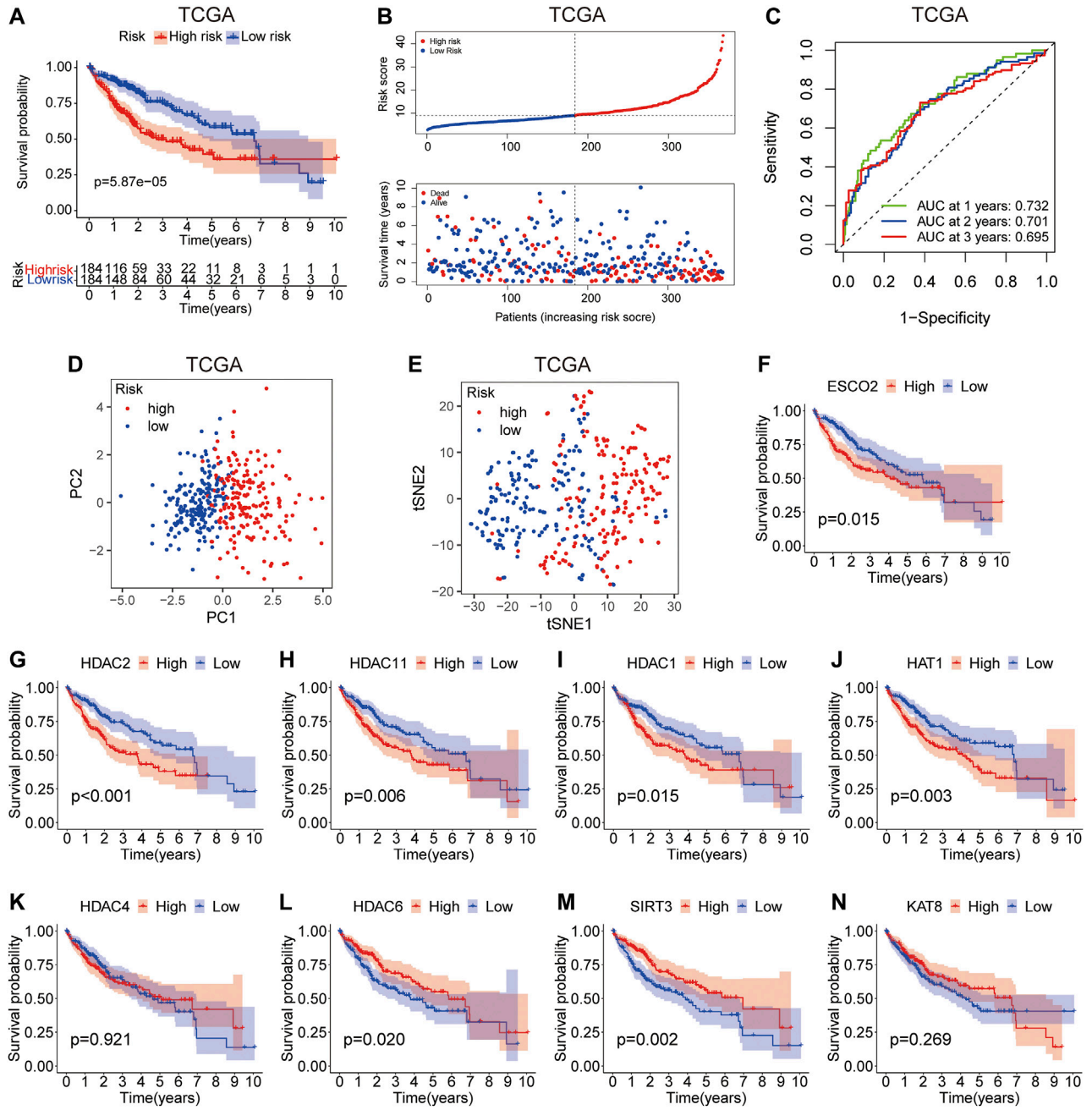


FIGURE 4 | Prognostic value of the risk model in TCGA cohort. **(A)** Survival analysis of patients in the high-risk and low-risk groups based on the prediction risk score formula. **(B)** The median value of risk scores with survival and statuses of HCC patients depends on the risk model in the TCGA cohort and the distribution of risk scores with survival and statuses of HCC patients depends on the risk model in the TCGA cohort. **(C)** 1-, 2-, and 3-years receiver operating characteristics (ROC) curves of the risk model for assessing the prognostic performance of the gene signature in the TCGA cohort. **(D)** Principal component analysis of HCC patients in TCGA cohort. **(E)** t-distributed stochastic neighbor embedding (t-SNE) analysis of HCC patients in TCGA cohort. **(F–N)** Kaplan–Meier survival analysis of the association between mRNA expression of LARs and OS in HCC patients. The HCC sample information was derived from TCGA databases.

HDAC11, HAT1 and HDAC6 had poorer OS expression than patients with low expression. (**Supplementary Figures S2F–N**), which is consistent with results from training set. These results revealed that the LARs signature is excellent in predicting the prognosis of HCC in the validation set.

Independent Prognostic Value of LARs Signature

According to the heat map of risk and clinical correlation, the high-risk score was positively correlated with tumor classification, stage, grade, age, and status in the TCGA dataset

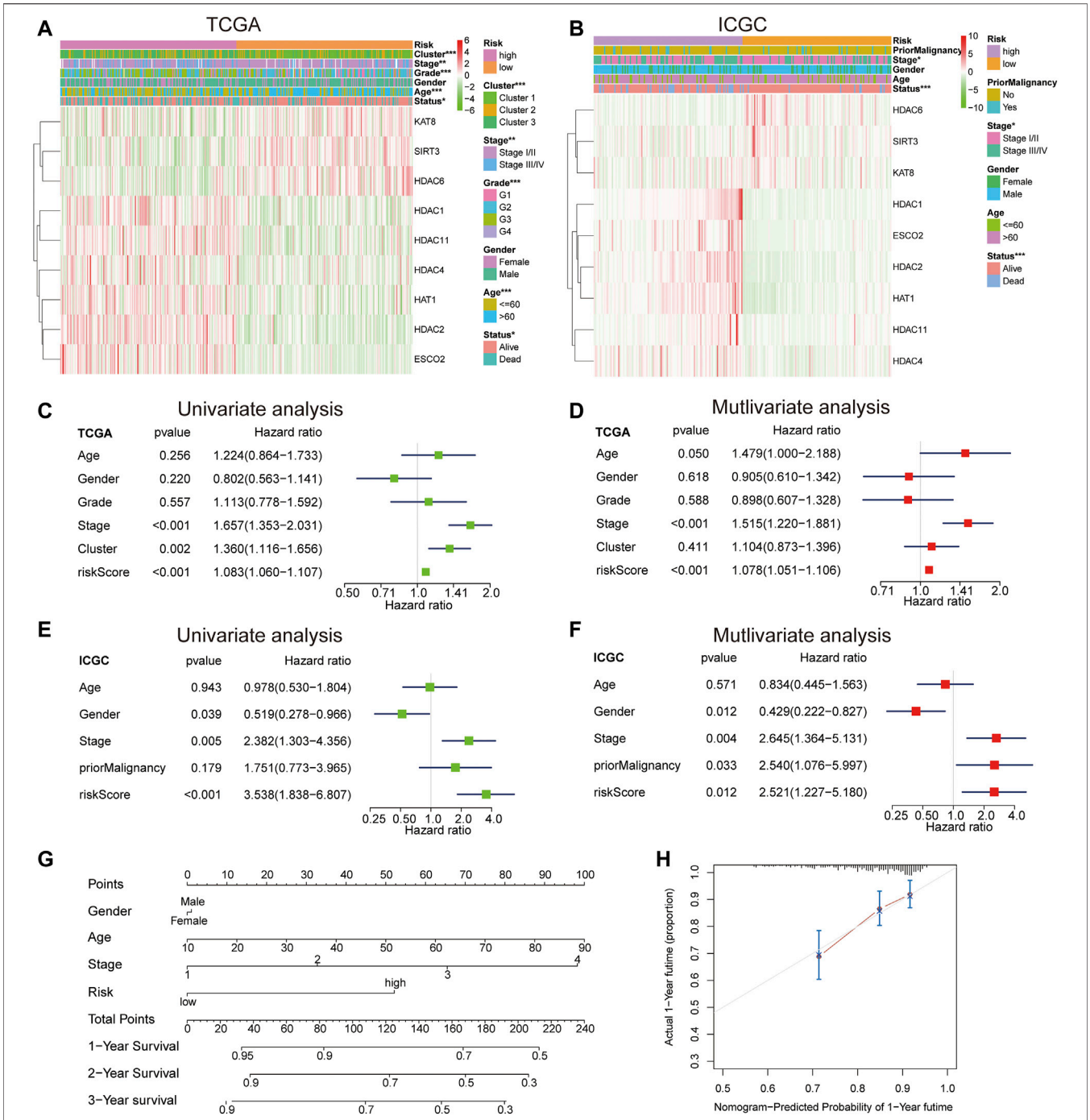


FIGURE 5 | (A) The differential expression levels of the included 9 LARs and the distributions of clinicopathological characteristics were compared between low- and high-risk subgroups in the TCGA cohort. **(B)** The differential expression levels of the included 9 LARs and the distributions of clinicopathological characteristics were compared between low- and high-risk subgroups in the ICGC cohort. **(C)**, **(D)** Univariate **(C)** and multivariate **(D)** Cox regression analyses of the OS and clinicopathological features of patients from TCGA datasets. **(E)**, **(F)** Univariate **(E)** and multivariate **(F)** Cox regression analyses of the OS and clinicopathological features of patients from ICGC datasets. **(G)** A line chart was established to predict the 1-, 3-, and 5-years survival rates of HCC patients. **(H)** The 1-year alignment diagram calibration curve of the entire TCGA queue.

(**Figure 5A**). In the ICGC dataset, the high-risk group was positively correlated with stage and status (**Figure 5B**). According to the risk and clinical correlation heat map of TCGA and ICGC datasets, HDAC1, HDAC2, HDAC4, HDAC11, HAT1, and ESCO2 were suggested to be high-risk LARs. In TCGA cohort, according to univariate Cox regression analysis, staging (HR = 1.657, 95% CI, 1.353–2.031, $p < 0.001$) and risk score (HR = 1.083, 95%

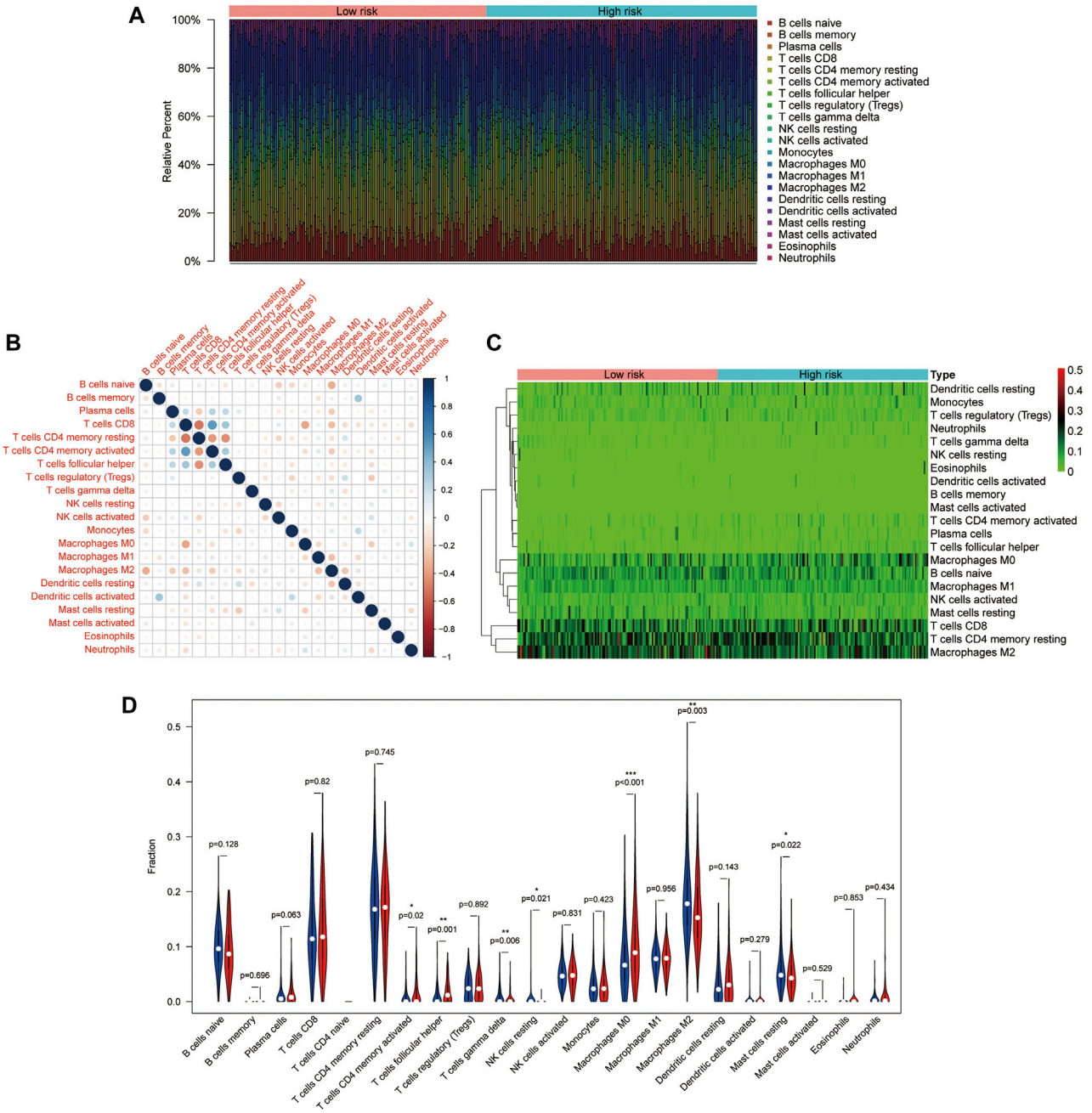


FIGURE 6 | Distribution and visualization of immune cell infiltration in HCC patients. **(A)** The histogram shows 21 specific immune components represented by different colors in each sample. **(B)** The correlation matrix of the proportion of 21 types of immune cells. **(C)** DEIRGs heat map of patients with high and low tumor mutational load (TMB). **(D)** Violin pictures with different degrees of immune cell infiltration in patients with high and low TMB. Blue and red represent low- and high-risk samples, respectively.

CI, 1.060–1.107, $p < 0.001$) were significantly correlated with OS (**Figure 5C**). Multivariate analysis by Cox regression showed that, after correcting for other confounding factors, staging (HR = 1.515, 95% CI, 1.220–1.881, $p < 0.001$) and risk score (HR = 1.078, 95% CI, 1.051–1.106, $p < 0.001$) were still statistically significant (**Figure 5D**). The results were validated using the ICGC queue (**Figures 5E,F**). Based on multivariate analysis by Cox regression, sex, age, stage, and risk scores were introduced into the line chart to quantitatively predict the OS of HCC patients (**Figure 5G**). The calibration curves of the 1-, 3-, and 5-years line diagrams were very close to the best prediction curve, and the predicted OS rate and the actual observation value were highly consistent. (**Figure 5H**, **Supplementary Figures S1B–F**).

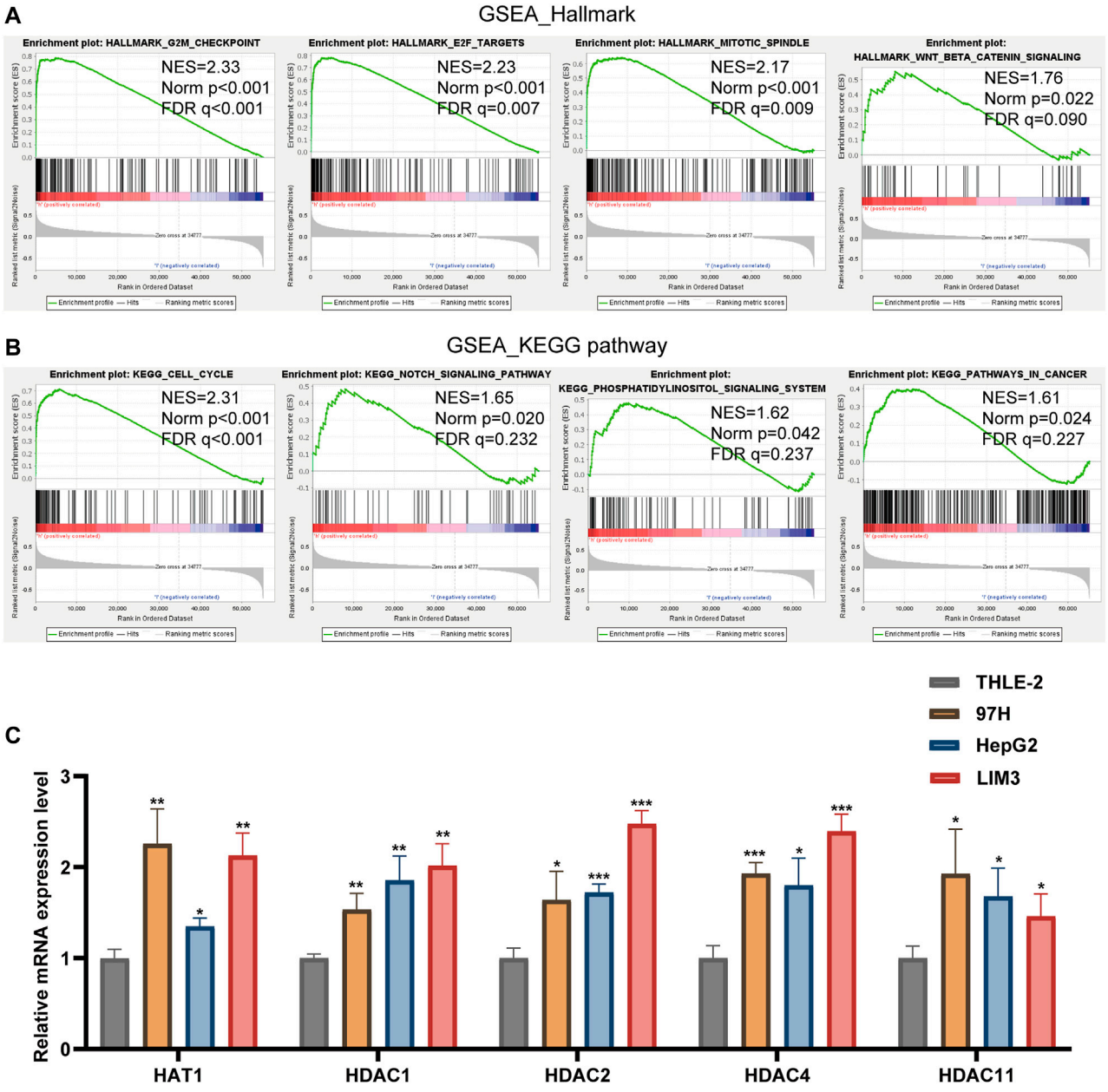


FIGURE 7 | (A,B) Gene Set Enrichment Analysis was conducted to predict the potential functions and pathways regulated by LARs based on TCGA datasets. **(C)** The RT-qPCR results of the 5 LARs genes was evaluated using the $2^{-\Delta\Delta CT}$ method. $^{*}p < 0.05$, $^{**}p < 0.01$ and $^{***}p < 0.001$.

Correlation Between LARs-Related Signature and Immune Infiltration

To further immune infiltration analysis, we employed CIBERSORT to calculate the infiltration abundance of 22 immune cell types between high-and low-risk HCC samples in the TCGA dataset. The infiltration ratio of immune cells was shown in **Figure 6A**. We also assessed the association between 22 types of immune cells, and the heat map revealed that the proportions of different subpopulations of tumor-infiltrating immune cells were weakly to moderately correlated (**Figure 6B**). Next, we used the heat map (**Figure 6C**) and the violin map (**Figure 6D**) to reveal possible differences in immune cell expression between groups. The results showed that memory activated $CD4^{+}$ T cells and macrophages M0 were higher expressed in the high-risk group, while T cells gamma delta, NK cells regulatory, Macrophages M2, T cells follicular helper, and Resting mast cells were more infiltrated in the low-risk group.

Gene Set Enrichment Analysis

GSEA was used to study tumor markers and signal pathways in patients with low-risk and high-risk HCC in TCGA dataset. We

determined that four tumor characteristics, namely E2F targets, G2/M checkpoint, mitotic spindle and Wnt/β-catenin signaling were significantly enriched in high-risk HCC patients (**Figure 7A**). GSEA-based KEGG pathway analysis confirmed that cell cycle, notch signaling pathway, pathways in cancer and phosphatidylinositol signaling system were enriched in the high-risk subgroup (**Figure 7B**).

Validation of LARs Expression Levels

We used RT-qPCR to verify the expression level of these prognosis-related LARs genes.

Our results showed that compared with the HCC cell line (THLE-2), the expression of HAT1, HDAC1, HDAC2, HDAC4 and HDAC11 were up-regulated in the HCC cell line (97H, HepG2, LIM3) (**Figure 7C**).

DISCUSSION

HCC is a serious malignant tumor worldwide. Furthermore, the incidence of HCC has continued to increase (Yang J. D. et al., 2019). Considering the huge heterogeneity of HCC, there is an urgent need to identify new prognostic biomarkers and establish a more accurate prognostic model. Lysine acetylation is a common cellular protein modification that regulates several cellular processes and participates in tumorigenesis and metastasis (Sabari et al., 2017). Related HCC studies have reported that LARs play an important role in HCC (Chai et al., 2021). Therefore, LARs may have considerable potential as biomarkers to predict HCC patient prognosis.

In this study, We found that the mRNA expression levels of majority of the evaluated LARS were closely related to the clinicopathological characteristics of HCC. Through the differential expression analysis of 852 genes, we screened 33 genes as potential prognostic factors to construct a prognostic model. The risk score model was constructed using LASSO Cox regression analysis, and nine differentially expressed genes (ESCO2, HAT1, KAT8, HDAC1, HDAC2, HDAC4, HDAC6, SIRT3, and HDAC11) were screened. The risk scoring model had good prediction effectiveness on both TCGA and ICGC datasets. Independent prognostic models suggest that LARs might serve as a potential prognostic prediction in HCC patients, and the high-risk groups exhibits remarkably lower OS rate of HCC patients than the low-risk group. The AUC value of the risk score model in the ICGC and TCGA cohorts performed well in predicting short-term survival (1–3 years). Univariate and multivariate Cox regression analysis of the two cohorts indicated that the prognostic characteristics of LAR-related genes were accurate, reliable, and explainable. GSEA results showed that various tumor-related signal transduction pathways were abundant in the high-risk groups. RT-qPCR results showed that HAT1, HDAC1, HDAC2, HDAC4, and HDAC11 were highly expressed in HCC cells.

We identified nine risk prognostic genes, namely HAT1, KAT8, ESCO2, HDAC1, HDAC2, HDAC4, HDAC6, SIRT3, and HDAC11. The risk score of this feature is related to invasive clinicopathological features and can also function as an independent prognostic factor for overall survival. HAT1 was identified as the first histone acetyltransferase. As a carcinogenic protein, HAT1 may promote cell proliferation and induce cisplatin resistance in HCC. Therefore, targeted HAT1 inhibitors are feasible strategies for the effective treatment of advanced HCC (Jin et al., 2017). KAT8 is mainly involved in the acetylation of histone H4 lysine 16 (H4K16) and certain non-histones. KAT8 consumption significantly promotes HCC cell migration and invasion. (Wei et al., 2021). The role of ESCO2 in HCC has not yet been reported. However, ESCO2 promotes proliferation and metastatic metabolic reprogramming of lung adenocarcinoma cells *in vitro* and *in vivo* (Zhu et al., 2021). Moreover, ESCO2 knockdown inhibits cell proliferation and induces apoptosis in human gastric cancer cells (Chen et al., 2018). HDAC1 and HDAC2 are enzymes that regulate gene transcription and participate in cell cycle progression, differentiation, apoptosis, and tumorigenesis. Pharmacological or transcriptional inhibition of HDAC1 and HDAC2 can lead to cell cycle arrest and apoptosis in HCC (Zhou et al., 2018). Targeted inhibition of HDAC4 has been shown to inhibit the growth and metastasis of HCC, and HDAC4 can promote the proliferation, migration and invasion of nasopharyngeal carcinoma cells *in vitro*, as well as tumor growth and lung metastasis *in vivo* (Freese et al., 2019; Cheng et al., 2021). HDAC6 is a tumor suppressor that inhibits Let-7i-5p to induce TSP1/CD47-mediated anti-tumorigenesis and phagocytosis of HCC. In our study, we found that HDAC6 expression was down-regulated in HCC patients, which is in line with the results of the current study (Yang H. D. et al., 2019). SIRT3 inhibits the growth and cell proliferation and promotes apoptosis in HCC cells. The low expression or lack of SIRT3 in HCC tissues suggests that SIRT3 expression may affect the occurrence and development of HCC (Liu et al., 2021). HDAC11 is the only IV histone deacetylase. It forms a complex with early growth response (Egr-1) of p53 transcription factor, which induces deacetylation of Egr-1, inhibits p53 transcription, and promotes the occurrence of HCC (Gong, 2019).

Tumor microenvironment plays an important role in the occurrence and development of tumor. In many solid organ malignant tumors, tumor infiltrating immune cells have high prognostic value for tumor progression and patient survival (Gajewski et al., 2013; Lamplugh and Fan, 2021). In this study, the proportion of memory-activated CD4 T cells and M0 macrophages was higher in patients with high risk score, which confirmed the role of LARs in the regulation of tumor immune invasion. Functional enrichment analysis showed that LARs was mainly involved in the immune pathway. Therefore, the poor prognosis of high-risk HCC patients may be due to this tumor immunosuppressive microenvironment.

As far as we know, this is the first study to identify the LAR genes related to the prognosis of HCC and to construct the related risk model for prognostication. The model is verified in two external independent HCC queues. LARs features have strong and stable prognostic value and have a broad prospect in clinical application. However, this study still has its limitations. First of all, although our model has been well verified in two large

databases, TCGA and ICGC, our study is still a retrospective study, and some prospective studies are needed to verify its clinical application. In addition, it is very important to verify the functional characteristics and molecular mechanism of LARs gene through biological experiments and clinical studies. All in all, further investigations with some wet lab evidence are needed to validate our findings and improve the statistical power to achieve more meaningful results.

CONCLUSION

In summary, our results suggest that LARs' risk score model can be used as a potential prognostic factor of HCC, which may be helpful for personalized management of cancer in the clinical environment.

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

JW designed the overall study and revised the paper. JZ performed public data interpretation. LS and JZ drafted the manuscript. KW supervised the experiments. SH, DL, and YX contributed to the data analysis. LS and JZ contributed equally to this study. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (No. 82060435).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.840412/full#supplementary-material

Supplementary Figure 1 | (A) Plots for LASSO expression coefficients of 10 LARs. **(B)** The 2-year alignment diagram calibration curve of the entire TCGA queue. **(C)** The 3-year alignment diagram calibration curve of the entire TCGA queue. **(D)** The 1-year alignment diagram calibration curve of the entire ICGC queue. **(E)** The 2-year alignment diagram calibration curve of the entire ICGC queue. **(F)** The 3-year alignment diagram calibration curve of the entire ICGC queue.

Supplementary Figure 2 | Validation of the risk model in the ICGC cohort. **(A)** Survival analysis of patients in the high-risk and low-risk groups based on the prediction risk score formula. **(B)** The median value of risk scores with survival and statuses of HCC patients depends on the risk model in the ICGC cohort and the distribution of risk scores with survival and statuses of HCC patients depends on the risk model in the ICGC cohort. **(C)** One-, two-, and 3-year ROC curves of the risk model for assessing the prognostic performance of the gene signature in the ICGC cohort. **(D)** Principal component analysis of HCC patients in the ICGC cohort. **(E)** t-SNE analysis of HCC patients in the ICGC cohort. **(F–N)** Kaplan–Meier survival analysis of the association between mRNA expression of LARs and OS in HCC patients. The HCC sample information was derived from the ICGC databases.

Supplementary Table S1 | Primers used in this study.

Supplementary Table S2 | Information of the nine LARs.

REFERENCES

Baell, J. B., Leaver, D. J., Hermans, S. J., Kelly, G. L., Brennan, M. S., Downer, N. L., et al. (2018). Inhibitors of Histone Acetyltransferases KAT6A/B Induce Senescence and Arrest Tumour Growth. *Nature.* 560 (7717), 253–257. doi:10.1038/s41586-018-0387-5

Baylin, S. B., and Jones, P. A. (2016). Epigenetic Determinants of Cancer. *Cold Spring Harb Perspect. Biol.* 8 (9), a019505. doi:10.1101/cshperspect.a019505

Bienvenut, W. V., Brünje, A., Boyer, J. B., Mühlenbeck, J. S., Bernal, G., Lassowskat, I., et al. (2020). Dual Lysine and N-terminal Acetyltransferases Reveal the Complexity Underpinning Protein Acetylation. *Mol. Syst. Biol.* 16 (7), e9464. doi:10.15252/msb.20209464

Caruso, S., O'Brien, D. R., Cleary, S. P., Roberts, L. R., and Zucman-Rossi, J. (2021). Genetics of Hepatocellular Carcinoma: Approaches to Explore Molecular Diversity. *Hepatology.* 73 (Suppl. 1), 14–26. doi:10.1002/hep.31394

Cavalli, G., and Heard, E. (2019). Advances in Epigenetics Link Genetics to the Environment and Disease. *Nature.* 571 (7766), 489–499. doi:10.1038/s41586-019-1411-0

Chai, X., Guo, J., Dong, R., Yang, X., Deng, C., Wei, C., et al. (2021). Quantitative Acetylome Analysis Reveals Histone Modifications that May Predict Prognosis in Hepatitis B-related Hepatocellular Carcinoma. *Clin. Translational Med.* 11 (3), e313. doi:10.1002/ctm2.313

Chen, H., Zhang, L., He, W., Liu, T., Zhao, Y., Chen, H., et al. (2018). ESCO2 Knockdown Inhibits Cell Proliferation and Induces Apoptosis in Human Gastric Cancer Cells. *Biochem. Biophysical Res. Commun.* 496 (2), 475–481. doi:10.1016/j.bbrc.2018.01.048

Chen, L., Liu, S., and Tao, Y. (2020). Regulating Tumor Suppressor Genes: post-translational Modifications. *Sig Transduct Target. Ther.* 5 (1), 90. doi:10.1038/s41392-020-0196-9

Cheng, C., Yang, J., Li, S.-W., Huang, G., Li, C., Min, W.-P., et al. (2021). HDAC4 Promotes Nasopharyngeal Carcinoma Progression and Serves as a Therapeutic Target. *Cell Death Dis.* 12 (2), 137. doi:10.1038/s41419-021-03417-0

Choudhary, C., Weinert, B. T., Nishida, Y., Verdin, E., and Mann, M. (2014). The Growing Landscape of Lysine Acetylation Links Metabolism and Cell Signalling. *Nat. Rev. Mol. Cell Biol.* 15 (8), 536–550. doi:10.1038/nrm3841

Falkenberg, K. J., and Johnstone, R. W. (2014). Histone Deacetylases and Their Inhibitors in Cancer, Neurological Diseases and Immune Disorders. *Nat. Rev. Drug Discov.* 13 (9), 673–691. doi:10.1038/nrd4360

Fan, S. T. (2012). Hepatocellular Carcinoma-Resection or Transplant? *Nat. Rev. Gastroenterol. Hepatol.* 9 (12), 732–737. doi:10.1038/nrgastro.2012.158

Feinberg, A. P., Koldobskiy, M. A., and Göndör, A. (2016). Epigenetic Modulators, Modifiers and Mediators in Cancer Aetiology and Progression. *Nat. Rev. Genet.* 17 (5), 284–299. doi:10.1038/nrg.2016.13

Figlia, G., Willnow, P., and Teleman, A. A. (2020). Metabolites Regulate Cell Signaling and Growth via Covalent Modification of Proteins. *Developmental Cell.* 54 (2), 156–170. doi:10.1016/j.devcel.2020.06.036

Freese, K., Seitz, T., Dietrich, P., Lee, S. M. L., Thasler, W. E., Bosserhoff, A., et al. (2019). Histone Deacetylase Expressions in Hepatocellular Carcinoma and Functional Effects of Histone Deacetylase Inhibitors on Liver Cancer Cells *In Vitro*. *Cancers.* 11 (10), 1587. doi:10.3390/cancers11101587

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

Gong, D., Zeng, F., and Wu, J. (2019). Inhibition of Histone Deacetylase 11 Promotes Human Liver Cancer Cell Apoptosis. *Am. J. Transl Res.* 11 (2), 983–990.

Huang, X., Zhang, G., Tang, T., and Liang, T. (2021). Identification of Tumor Antigens and Immune Subtypes of Pancreatic Adenocarcinoma for mRNA Vaccine Development. *Mol. Cancer.* 20 (1), 44. doi:10.1186/s12943-021-01310-0

Jin, X., Tian, S., and Li, P. (2017). Histone Acetyltransferase 1 Promotes Cell Proliferation and Induces Cisplatin Resistance in Hepatocellular Carcinoma. *Oncol. Res.* 25 (6), 939–946. doi:10.3727/096504016X14809827856524

Kaypee, S., Sudarshan, D., Shanmugam, M. K., Mukherjee, D., Sethi, G., and Kundu, T. K. (2016). Aberrant Lysine Acetylation in Tumorigenesis: Implications in the Development of Therapeutics. *Pharmacol. Ther.* 162, 98–119. doi:10.1016/j.pharmthera.2016.01.011

Kulik, L., and El-Serag, H. B. (2019). Epidemiology and Management of Hepatocellular Carcinoma. *Gastroenterology.* 156 (2), 477–491. doi:10.1053/j.gastro.2018.08.065

Lamplugh, Z., and Fan, Y. (2021). Vascular Microenvironment, Tumor Immunity and Immunotherapy. *Front. Immunol.* 12, 811485. doi:10.3389/fimmu.2021.811485

Li, G., Tian, Y., and Zhu, W.-G. (2020). The Roles of Histone Deacetylases and Their Inhibitors in Cancer Therapy. *Front. Cell Dev. Biol.* 8, 576946. doi:10.3389/fcell.2020.576946

Lipinski, M., Muñoz-Viana, R., Del Blanco, B., Marquez-Galera, A., Medrano-Relinque, J., Caramés, J. M., et al. (2020). KAT3-dependent Acetylation of Cell Type-specific Genes Maintains Neuronal Identity in the Adult Mouse Brain. *Nat. Commun.* 11 (1), 2588. doi:10.1038/s41467-020-16246-0

Liu, M., Yu, J., Jin, H., Wang, S., Ding, J., Xing, H., et al. (2021). Bioinformatics Analysis of the SIRT Family Members and Assessment of Their Potential Clinical Value. *Onco Targets Ther.* 14, 2635–2649. doi:10.2147/OTT.S298616

Narita, T., Weinert, B. T., and Choudhary, C. (2019). Functions and Mechanisms of Non-histone Protein Acetylation. *Nat. Rev. Mol. Cell Biol.* 20 (3), 156–174. doi:10.1038/s41580-018-0081-3

Nault, J. C., Martin, Y., Caruso, S., Hirsch, T. Z., Bayard, Q., Calderaro, J., et al. (2020). Clinical Impact of Genomic Diversity From Early to Advanced Hepatocellular Carcinoma. *Hepatology.* 71 (1), 164–182. doi:10.1002/hep.30811

Sabari, B. R., Zhang, D., Allis, C. D., and Zhao, Y. (2017). Metabolic Regulation of Gene Expression through Histone Acylations. *Nat. Rev. Mol. Cell Biol.* 18 (2), 90–101. doi:10.1038/nrm.2016.140

Sheikh, B. N., and Akhtar, A. (2019). The many Lives of KATs - Detectors, Integrators and Modulators of the Cellular Environment. *Nat. Rev. Genet.* 20 (1), 7–23. doi:10.1038/s41576-018-0072-4

Wei, S., Liu, W., Sun, N., Wu, Y., Song, H., Wang, C., et al. (2021). MOF Upregulates the Estrogen Receptor α Signaling Pathway by its Acetylase Activity in Hepatocellular Carcinoma. *Cancer Sci.* 112 (5), 1865–1877. doi:10.1111/cas.14836

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: a Class Discovery Tool with Confidence Assessments and Item Tracking. *Bioinformatics (Oxford, England).* 26 (12), 1572–1573. doi:10.1093/bioinformatics/btq170

Yang, H. D., Kim, H. S., Kim, S. Y., Na, M. J., Yang, G., Eun, J. W., et al. (2019). HDAC6 Suppresses Let-7i-5p to Elicit TSP1/CD47-Mediated Anti-Tumorigenesis and Phagocytosis of Hepatocellular Carcinoma. *Hepatology.* 70 (4), 1262–1279. doi:10.1002/hep.30657

Yang, J. D., Hainaut, P., Gores, G. J., Amadou, A., Plymoth, A., and Roberts, L. R. (2019). A Global View of Hepatocellular Carcinoma: Trends, Risk, Prevention and Management. *Nat. Rev. Gastroenterol. Hepatol.* 16 (10), 589–604. doi:10.1038/s41575-019-0186-y

Zhou, H., Cai, Y., Liu, D., Li, M., Sha, Y., Zhang, W., et al. (2018). Pharmacological or Transcriptional Inhibition of Both HDAC1 and 2 Leads to Cell Cycle Blockage and Apoptosis via p21Waf1/Cip1 and p19INK4d Upregulation in Hepatocellular Carcinoma. *Cell Prolif.* 51 (3), e12447. doi:10.1111/cpr.12447

Zhu, H.-e., Li, T., Shi, S., Chen, D.-x., Chen, W., and Chen, H. (2021). ESCO2 Promotes Lung Adenocarcinoma Progression by Regulating hnRNPA1 Acetylation. *J. Exp. Clin. Cancer Res.* 40 (1), 64. doi:10.1186/s13046-021-01858-1

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

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

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 14 March 2022
doi: 10.3389/fmolb.2022.753318

Gene Expression-Based Signature Can Predict Sorafenib Response in Kidney Cancer

*Alexander Gudkov[1], Valery Shirokorad[2], Kirill Kashintsev[2], Dmitriy Sokov[3], Daniil Nikitin[4], Andrey Anisenko[4], Nicolas Borisov[5], Marina Sekacheva[6], Nurshat Gaifullin[7], Andrew Garazha[8], Maria Suntsova[6], Elena Koroleva[5], Anton Buzdin[5,8,6,9] and Maksim Sorokin[1,5,8,10]**

[1]I. M. Sechenov First Moscow State Medical University, Moscow, Russia, [2]Moscow City Oncological Hospital №. 62, Moscow, Russia, [3]Moscow City Clinical Oncological Dispensary №. 1, Moscow, Russia, [4]Oncobox Ltd., Moscow, Russia, [5]Moscow Institute of Physics and Technology, Moscow, Russia, [6]World-Class Research Center "Digital Biodesign and Personalized Healthcare", Sechenov First Moscow State Medical University, Moscow, Russia, [7]Department of Pathology, Faculty of Medicine, Lomonosov Moscow State University, Moscow, Russia, [8]OmicsWay Corp, Walnut, CA, United States, [9]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia, [10]European Organization for Research and Treatment of Cancer (EORTC), Biostatistics and Bioinformatics Subgroup, Brussels, Belgium

OPEN ACCESS

Edited by:
William C. Cho,
QEH, Hong Kong SAR, China

Reviewed by:
Byeong Hwa Yun,
United States Food and Drug Administration, United States
Mahendra K. Singh,
University of Miami, United States
Yuriy L. Orlov,
I. M. Sechenov First Moscow State Medical University, Russia

***Correspondence:**
Maksim Sorokin
sorokin@oncobox.com

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 04 August 2021
Accepted: 28 January 2022
Published: 14 March 2022

Citation:
Gudkov A, Shirokorad V, Kashintsev K, Sokov D, Nikitin D, Anisenko A, Borisov N, Sekacheva M, Gaifullin N, Garazha A, Suntsova M, Koroleva E, Buzdin A and Sorokin M (2022) Gene Expression-Based Signature Can Predict Sorafenib Response in Kidney Cancer. Front. Mol. Biosci. 9:753318. doi: 10.3389/fmolb.2022.753318

Sorafenib is a tyrosine kinase inhibitory drug with multiple molecular specificities that is approved for clinical use in second-line treatments of metastatic and advanced renal cell carcinomas (RCCs). However, only 10–40% of RCC patients respond on sorafenib-containing therapies, and personalization of its prescription may help in finding an adequate balance of clinical efficiency, cost-effectiveness, and side effects. We investigated whether expression levels of known molecular targets of sorafenib in RCC can serve as prognostic biomarker of treatment response. We used Illumina microarrays to profile RNA expression in pre-treatment formalin-fixed paraffin-embedded (FFPE) samples of 22 metastatic or advanced RCC cases with known responses on next-line sorafenib monotherapy. Among them, nine patients showed partial response (PR), three patients—stable disease (SD), and 10 patients—progressive disease (PD) according to Response Evaluation Criteria In Solid Tumors (RECIST) criteria. We then classified PR + SD patients as "responders" and PD patients as "poor responders". We found that gene signature including eight sorafenib target genes was congruent with the drug response characteristics and enabled high-quality separation of the responders and poor responders [area under a receiver operating characteristic curve (AUC) 0.89]. We validated these findings on another set of 13 experimental annotated FFPE RCC samples (for 2 PR, 1 SD, and 10 PD patients) that were profiled by RNA sequencing and observed AUC 0.97 for 8-gene signature as the response classifier. We further validated these results in a series of qRT-PCR experiments on the third experimental set of 12 annotated RCC biosamples (for 4 PR, 3 SD, and 5 PD patients), where 8-gene signature showed AUC 0.83.

Keywords: renal cell carcinoma, kidney cancer, gene signature, mRNA expression, RNA sequencing, microarray profiling, sorafenib response, tyrosine kinase inhibitor

INTRODUCTION

According to the estimates for 2020, globally there were ~431,000 new cases of kidney cancer and ~179,000 associated deaths (Sung et al., 2021). Renal cell carcinoma (RCC) is the most common subtype of kidney cancer in adults, responsible for nearly 90% of all cases and prone to distant metastasis (He et al., 2021). RCC arises from the renal parenchyma, and the incidence of RCC is still increasing in most countries (Bhatt and Finelli, 2014; Du et al., 2020). Approximately 25–30% of RCC patients are diagnosed at a metastatic or locally advanced disease stage, and another third of RCC patients will recur after receiving a successful first-line treatment (Sánchez-Gastaldo et al., 2017). RCC includes several different histological subtypes with distinct biological behaviors and prognoses.

RCCs are frequently characterized by inactivation of the *VHL* tumor suppressor gene. This leads to deficiency of its encoded protein, which is part of an E3 ubiquitin ligase complex that degrades alpha subunit of hypoxia inducible factor 2 (HIF-2α). The resulting excessive accumulation of HIF-2α can transcriptionally upregulate oncogenic hypoxia-responsive genes, including platelet-derived growth factor (PDGF) and vascular endothelial growth factor (VEGF) (Choueiri and Kaelin, 2020). In turn, VEGF and PDGF promote angiogenesis, cell growth and survival, and RCC progression by activating the respective tyrosine kinase receptors PDGFR and VEGFR. This leads to high vascularization of RCC and to its high metastatic potential (He et al., 2021). Patients with metastatic RCC are insensitive to chemotherapy and radiotherapy, and have a poor survival prognosis (Choueiri and Motzer, 2017).

Sorafenib is a tyrosine kinase inhibitor with multiple specificities that targets at least ten tyrosine kinase molecules: RAF1, BRAF, RET, FLT1, FGFR1, KIT, PDGFRB, FLT3, FLT4, and KDR (Adnane et al., 2006). Sorafenib is thought to have a dual suppressive effect on tumors by blocking both angiogenesis, and cell proliferation and survival through the inhibition of VEGFR/PDGFR and BRAF/RET/FLT/FGFR/KIT/KDR signaling axis, respectively (Wilhelm et al., 2004, 2006). It is the first targeted drug approved for treatment of metastatic or locally advanced RCC by US FDA, which revolutionized treatment of kidney cancer and accelerated development and registration of other targeted therapeutics (Escudier et al., 2019). Since then, several other specific agents against VEGF, PDGF, and their receptors have been approved for the treatment of RCC, including sunitinib, axitinib, cabozanitinib, lenvatinib, pazopanib, and bevacizumab (Dizman et al., 2020). In addition, mTOR-specific inhibitors temsirolimus and everolimus were also approved for the treatment of this disease (Dizman et al., 2020). According to the National Comprehensive Cancer Network (NCCN) guidelines, sorafenib and sunitinib are recommended as drugs for metastatic RCC, where sorafenib has a lower toxicity than sunitinib (Deng et al., 2019). Moreover, sorafenib is also approved as the first-line treatment of metastatic RCC according to the latest guidelines of Chinese Society of Clinical Oncology (He et al., 2021). In clinical practice, it is also effective in hepatocellular and thyroid cancers (Escudier et al., 2019), myeloid leukemia, mesothelioma, and prostate cancer (Méndez-Blanco et al., 2018).

However, only 10–40% of RCC patients respond on sorafenib-containing therapeutic schemes (Larkin and Eisen, 2006; Escudier et al., 2007; Guevremont et al., 2009), and personalization of its prescriptions may help in finding an adequate balance of clinical efficiency, cost-effectiveness, and side effects. Nowadays, there are no clinical biomarkers of response on sorafenib treatment in RCC, and the molecular mechanisms of sorafenib resistance in RCC are not sufficiently understood (He et al., 2021). Currently, several RCC sorafenib resistance biomarkers were identified by RNA expression assays in drug responder and non-responder tumors: long non-coding RNAs *GAS5* (Liu et al., 2019) and *SRLR* (Xu et al., 2017), microRNA *miR-21*, and genes *ANGPTL3* (Bao et al., 2018), *CXCR4*, *CD34* (Aziz et al., 2014), *FGFR1* (Ho et al., 2015), *FRS2A*, *GLUT1*, *HO-1* (Zheng et al., 2015), *SOX5*, and *SOX9* (Li et al., 2015; He et al., 2021). In addition, expression of AKT protein was reported to be a biomarker of enhanced resistance against sorafenib in RCC patients (Jonasch et al., 2010). However, despite those important observations, no diagnostic settings were constructed to predict effectiveness of sorafenib for RCC patients.

Drug target expression levels, e.g., determined by immunohistochemistry, are already used as the biomarkers of treatment response in some solid cancers (Hechtman et al., 2017) including breast cancer (Nicolini et al., 2018) and gastric cancer (Abrahao-Machado and Scapulatempo-Neto, 2016). In addition, profiling of gene expression at RNA level is a powerful tool for discovery of drug efficiency biomarkers and for cancer therapy personalization (Buzdin et al., 2019). Previously, we showed that gene expression levels established from standardized RNA sequencing data can be used as robust estimators of the corresponding protein levels for several cancer biomarkers in tumor biosamples, including formalin-fixed paraffin-embedded (FFPE) specimens (Sorokin et al., 2020b).

In this study, we investigated FFPE biosamples of pre-treatment RCC tissues from 47 patients with known response status on next-line monotherapy with sorafenib. Illumina microarrays were used to profile RNA expression in FFPE samples of 22 metastatic or advanced RCC cases. Among them, nine patients showed partial response (PR), three patients—stable disease (SD), and 10 patients—progressive disease (PD) according to RECIST criteria. We then classified PR + SD patients as "responders" and PD patients as "poor responders". We found that gene signature including eight sorafenib target genes was congruent with the drug response characteristics, and enabled high-quality separation of the responders and poor responders [area under a receiver operating characteristic curve (AUC) 0.89]. We validated these findings on another set of 13 experimental annotated FFPE RCC samples (for 2 PR, 1 SD, and 10 PD patients) that were profiled by RNA sequencing and observed AUC 0.97 for 8-gene signature as the response classifier. We further validated these results in a series of qRT-PCR experiments on the third experimental set of 12 annotated RCC biosamples (for 4 PR, 3 SD, and 5 PD patients), where 8-gene signature showed AUC 0.83.

TABLE 1 | Clinical information for RCC patients profiled using Illumina HumanHT-12 WG-DASL V4.0 R2 gene expression arrays

Patient ID	Response status	Age	Gender	T	N	M	Grade
18	Partial response	66	Male	2	0	1	4
26	Stable disease	64	Female	3	2	1	4
27	Progressive disease	53	Male	3	0	1	4
31	Partial response	62	Male	2	2	0	3
36	Progressive disease	60	Male	2	0	0	2
37	Partial response	49	Female	3	1	1	4
46	Progressive disease	45	Male	3	0	0	3
49	Progressive disease	66	Female	3	0	0	3
54	Partial response	55	Female	2	0	0	2
58	Progressive disease	65	Female	3	1	0	3
60	Progressive disease	59	Male	2	0	0	2
62	Progressive disease	58	Male	1	0	0	1
72	Partial response	56	Female	3	1	0	3
73	Progressive disease	48	Male	3c	0	0	3
74	Partial response	53	Male	4	2	1	4
88	Stable disease	59	Female	3	0	0	3
91	Stable disease	67	Female	3	2	1	4
94	Progressive disease	74	Female	3	1	1	4
97	Partial response	70	Female	3	0	0	3
122	Partial response	61	Male	3a	0	0	3
128	Partial response	68	Female	3	0	1	4
135	Progressive disease	50	Male	3	0	0	3

TABLE 2 | Clinical information for RCC patients profiled using Illumina HiSeq3000 next-generation sequencing platform in this study

Patient ID	Response	Age	Gender	T	N	M	Grade
KC11	Progressive disease	62	Female	3	1	1	4
KC14	Progressive disease	68	Female	3	0	1	4
KC19	Partial response	46	Female	3	0	0	3
KC21	Progressive disease	41	Male	3	0	1	4
KC23	Progressive disease	53	Male	3	0	1	4
KC26	Stable disease	55	Male	3	0	1	4
KC36	Progressive disease	64	Female	3a	0	1	4
KC37	Partial response	54	Male	3a	0	1	4
KC46	Progressive disease	55	Male	3b	2	1	4
KC57	Progressive disease	58	Male	3b	0	1	4
KC92	Progressive disease	55	Male	3	0	0	3
KC93	Progressive disease	65	Female	2	0	0	2
KC96	Progressive disease	47	Male	3	0	0	3

TABLE 3 | Outline of clinical information of patients whose samples were profiled using RT-PCR platform in this study

Patient ID	Response	Age	Gender	T	N	M	Grade
III-1	Partial response	67	Male	3	0	1	4
III-2	Stable disease	45	Female	2	0	0	2
III-3	Partial response	48	Female	3	0	0	3
III-4	Progressive disease	65	Female	1	0	1	4
III-5	Progressive disease	59	Male	3	1	1	4
III-6	Progressive disease	53	Female	4	0	0	3
III-7	Progressive disease	58	Male	1	0	0	1
III-8	Progressive disease	51	Male	3b	2	1	4
III-9	Partial response	71	Female	4	1	0	3
III-10	Stable disease	59	Male	1	0	0	1
III-11	Stable disease	70	Male	3	1	1	4
III-12	Partial response	47	Male	3	1	1	4

MATERIALS AND METHODS

Patients and Samples

All patients enrolled in this study have previously signed written informed consents to participate in the observational clinical investigation, and for publication of depersonalized molecular profiles and study results in the form of gene activity profiles associated with age, sex, and results of sorafenib monotherapy treatment estimated according to RECIST criteria (Eisenhauer et al., 2009). The patients provided written informed consent that their tumor samples will be subjected to gene expression profiling using either microarray technology or next-generation sequencing or qRT-PCR. Twenty-two patients signed agreement that their biosamples are profiled with the Illumina HT-12 bead arrays (**Table 1**). Thirteen patients signed agreement that their biosamples are profiled by RNA sequencing using Illumina HiSeq3000 next generation sequencing platform (**Table 2**). Twelve patients signed agreement that their biosamples are profiled by qRT-PCR, but not using expression microarrays or RNA sequencing (**Table 3**).

The study was conducted in accordance with the Declaration of Helsinki ethical principles. The patient groups, design of this study, and its public presentation in the form of a research paper were approved by the local ethical committees at I.M. Sechenov First Moscow State Medical University, Moscow City Oncological Hospital №. 62, and Moscow City Clinical Oncological Dispensary №. 1.

Biosamples were collected prospectively in the period from May 2015 until July 2020. All biosamples obtained from all the patients in this study were FFPE RCC tumor blocks obtained from primary tumor sites and evaluated by a pathologist, with cancer cell content of at least 60%. All patients were treated with sorafenib in first-line therapy and their responses were assessed according to RECIST criteria (Eisenhauer et al., 2009).

Twenty-two samples from kidney cancer patients were analyzed using Illumina HumanHT-12 WG-DASL V4.0 R2 gene expression array (**Table 1**). Among them, patients with progressive disease were considered as poor responders (n = 10), whereas patients with partial response and stable disease were classified as the responders (n = 12).

Gene expression for 13 other RCC samples was profiled by RNA sequencing using Illumina HiSeq3000 next generation sequencing platform (**Table 2**). Similarly, patients with progressive disease were considered as poor responders (n = 10), whereas patients with partial response and stable disease were classified as the responders (n = 3).

Finally, patients in the third set of 12 RCC patients were profiled by quantitative reverse transcription PCR (qRT-PCR) assay (**Table 3**). According to the aforementioned criteria, five patients were considered poor responders, and seven patients—treatment responders.

Gene Expression Assays

RNA Extraction

To isolate RNA, 10-μm-thick paraffin slices were trimmed from each FFPE RCC tissue block using microtome. RNA was

TABLE 4 | Sequences of qRT-PCR primers used in this study

Target gene	Oligonucleotide sequence[a] (5′–3′)
RAF1	F, *CTGGCTCCCTCAGGTTTAAGAA*
	R, *AAGCTCCCTGTATGTGCTCC*
FLT3	F, *CTCAAGGAAACGGCCATCCT*
	R, *AACACGGCCATCCACATTCT*
FLT1	F, *TGTCGTGTAAGGAGTGGACC*
	R, *GCACCTGCTGTTTTCGATGT*
FGFR1	F, *GAGTGACTTCCACAGCCAGA*
	R, *GGATGCACTGGAGTCAGCAG*
BRAF	F, *CAGAGGACAGTGGTACCTGC*
	R, *CAGCACAGCACTCTGGGATT*
PDGFRB	F, *GCAAAACCACCATTGGGGAC*
	R, *TGCGTTCACAGAGACGTTGA*
KDR	F, *GAAACTGACTTGGCCTCGGT*
	R, *CACGACTCCATGTTGGTCACT*
KIT	F, *GCACAATGGCACGGTTGAAT*
	R, *GGTGTGGGGATGGATTTGCT*
ACTB	F, *ACAGAGCCTCGCCTTTGC*
	R, *CGCGGCGATATCATCATCCA*
VCP	F, *TGGAAGCGTATCGACCCATC*
	R, *CTTTGAACTCCACAGCACGC*
DIABLO	F, *AATGGCGGCTCTGAAGAGTT*
	R, *AAACTCGAGCCAAGCAGGAA*
EIF3B	F, *GGCGAACACCATCTTCTGGA*
	R, *TGTCCACAAACGCTAAGGCA*
PSMB2	F, *GCAGCAGCTAACTTCACACG*
	R, *AGCCAGGAGGAGGTTCACAT*
POLR2C	F, *TCTTCATCGCTGAGGTTCCC*
	R, *ATCCAAGCCTGTGAGCAATGA*

[a]*F—forward, R—reverse.*

extracted from FFPE slices using QIAGEN RNeasy FFPE Kit following the manufacturer's protocol. RNA 6000 Nano or Qubit RNA Assay kits were used to measure RNA concentration. RNA Integrity Number (RIN) was measured using Agilent 2100 bio-Analyzer.

Microarray Gene Expression Profiling

Gene expression profiling was done according to Lezhnina et al. (2014) at Dr. Olga Kovalchuk's laboratory, University of Lethbridge, Alberta, Canada. The profiling was done using Illumina HumanHT-12 WG-DASL V4.0 R2 gene expression bead arrays. BeadChips were scanned using Illumina BeadArray Reader and the Bead Scan Software (Illumina).

RNA Sequencing

RNA sequencing was done according to Suntsova et al. (2019) and Sorokin et al. (2020d) at the Department of Pathology and Laboratory Medicine, University of California Los Angeles. For depletion of ribosomal RNA and library construction, KAPA RNA Hyper with rRNA erase kit (HMR only) was used. Different adaptors were used for multiplexing samples in one sequencing run. Library concentrations and quality were measured using Qubit ds DNA HS Assay kit (Life Technologies) and Agilent Tapestation (Agilent). RNA sequencing was done using Illumina HiSeq 3000 equipment for single-end sequencing, 50 bp read length, for approximately 30 million (mln) raw reads per sample. Data quality check was done on Illumina SAV. De-multiplexing was performed with Illumina Bcl2fastq2 v 2.17 program.

Quantitative Reverse Transcription PCR

Quantitative reverse transcription PCR (qRT-PCR) panel was developed to measure the expression level of eight target and six housekeeping genes in kidney cancer samples using Evrogen Reverse transcription polymerase and Evrogen Taq polymerase kit with SYBR Green for the PCR product detection. PCR mix composition included (25 μl total volume)

Buffer (HS-qPCRmix-HS SYBR; Evrogen, Moscow, Russia)—5 μl;
Primers 1 μl (0.4 μM each);
RNA solution—1–3 μl (2–6 ng total RNA per mix);
MMLV-RT (Evrogen)—2 μl;
Water—13–15 μl.

The oligonucleotide sequences for PCR primers are listed in **Table 4**. Following reverse transcription reaction, the PCR mix was melted at 95°C for 5 min, and then the following cycling conditions were applied using CFX Touch Real-Time PCR Detection System (BioRad):

95°C—30 s.
60°C—30 s.
72°C—30 s.

Each experiment was carried out in quadruplicate.

Processing of Gene Expression Data

Illumina HumanHT-12 WG-DASL V4.0 R2 Gene Expression Array

Probe IDs were mapped to HGNC gene symbols (Yates et al., 2017) using the manufacturer's annotation table. Gene expression values were normalized using quantile normalization protocol (Bolstad et al., 2003) prior to further processing. R package preprocessCore was used to perform quantile normalization.

Illumina HiSeq3000 RNAseq Profiles

RNA sequencing FASTQ files were processed with STAR aligner (Dobin et al., 2013) in "GeneCounts" mode with the Ensembl human transcriptome annotation (Build version GRCh38 and transcript annotation GRCh38.89). Ensembl gene IDs were converted to HGNC gene symbols using Complete HGNC dataset (https://www.genenames.org/, accessed on 2017 July 13). In total, expression levels were established for 36,596 annotated genes with the corresponding HGNC identifiers. Raw gene counts were normalized using R DESeq2 package (Love et al., 2014).

Quantitative Reverse Transcription PCR

For each sorafenib target gene from the RCC drug sensitivity gene signature (*RAF1*, *BRAF*, *FLT1*, *FGFR1*, *KIT*, *PDGFRB*, *FLT3*, *FLT4*, *KDR*), we performed normalization using expression of six housekeeping genes selected according to Chang et al. (2011) (*ACTB*, *GAPDH*, *POLR2C*, *PSMB2*, *DIABLO*, *VCP*). For each

gene, we calculated ΔCt by subtracting the value of the threshold cycle of cDNA amplification of a target gene from the geometric mean value of the threshold cycle of cDNA amplification of the housekeeping genes. The gene signature score was calculated as sum of ΔCt values for all genes included in the signature.

Gene Expression Analysis and Visualization

Differential gene expression analysis was performed using Student *t*-test. The observed clinical responses were used for investigation of molecular signature using ROC-AUC analysis (Fawcett, 2006). Area under a receiver operating characteristic curve (ROC-AUC) values were calculated using ROCR package in R environment (Sing et al., 2005). Patient survival analysis and visualization were performed using R packages *survival*, *survminer*, and *ggplot2*.

Sorafenib *In Vitro* Efficiency Data

From Genomics of Drug Sensitivity in Cancer (GDSC) database (https://www.cancerrxgene.org/downloads/bulk_download, accessed on 2021 March 30), we downloaded log10-transformed IC_{50} values for sorafenib in 732 cancer cell lines corresponding to 13 different tumor types and 50 subtypes (**Supplementary Table S1**). For each cell line, we downloaded raw gene expression data from ArrayExpress database (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-3610/) in CEL format, experimentally profiled using Affymetrix Human Genome U219 Array. CEL files were normalized and background correction was applied using *rma* function of *affy* R package.

Mutation Analysis

For mutation analysis, we used gene expression and genetic features data from GDSC database (https://www.cancerrxgene.org). We used data for 802 solid and 167 blood cancer cell lines with available genetic mutational profiles. For each mutation, we compared IC_{50} values for sorafenib between mutant and wild-type cell lines using non-parametric Mann–Whitney *U* test. Then we applied false discovery rate (FDR) correction to adjust for comparing multiple genetic features. Genetic features with FDR-corrected *p*-values <0.1 and more than 2-fold IC_{50} differences were considered as significant. Mann–Whitney *U* tests and FDR correction were performed using *scipy* and *statsmodels* Python libraries implemented in GDSC web interface.

RESULTS

Study Population

In total, 47 RCC patients were enrolled in this study (21 female and 26 male patients, age range 41–74, mean 58 y.o.). The biosamples were FFPE RCC tumor tissue blocks collected in the period from May 2015 until July 2020. Gene expression was profiled using three different methods: microarray hybridization using Illumina HT-12 bead array, Illumina RNA sequencing, and qRT-PCR. Each patient provided a written informed consent and agreed that his/her biosample is profiled with one of the aforementioned methods. In the microarray group, there were 22 patients including 11 women and 11 men, age range 45–74, mean 59 y.o.; in RNAseq group—13 patients including 5 women and 8 men, age range 41–68, mean 56 y.o.; in qRT-PCR group, there were 12 patients including 7 men and 5 women, age range 45–71, mean 58 y.o. (**Tables 1**–**3**). The patients whose response status on next-line sorafenib monotherapy treatment according to RECIST criteria was "Progressive disease" were considered as poor responders, and the patients with statuses "Partial response" and "Stable disease" were considered as the responders. No "Complete response" outcomes according to RECIST (disappearance of all target lesions) were detected. This is in line with a previous study by Escudier et al., where only 1 out of 451 RCC patients treated with sorafenib had a complete response (Escudier et al., 2007). In total, 25 patients were classified as the poor responders, and 22—as the responders (**Tables 1**–**3**).

Specifically, there were 10 non-responders and 12 responders in the microarray group, 10 non-responders and three responders in the RNAseq group, and five non-responders and seven responders in the qRT-PCR group.

Differential Gene Expression Analysis and Generation of Sorafenib Response Signature

In the samples profiled by Illumina microarrays, we screened differential gene expression between the responder and poor responder biosamples. We aimed to generate sorafenib response gene signature and focused on expression levels of sorafenib target genes to avoid over-training. Using prior knowledge such as biological function of the genes is a well-established technique for feature selection as reviewed in Hira and Gillies (2015). At the single gene level, we observed a significant difference between the responders and poor responders only for *FLT1* and *PDGFRB* genes, which were both upregulated in the responders group (**Table 5**; **Figure 1**). Multiple logistic regression analysis did not provide significant coefficients for any of the sorafenib target genes.

In the previous studies, drug response statuses could correlate with the drug target gene expression levels (Tkachev et al., 2020), and for generating sorafenib drug response signature, we selected sorafenib target genes whose expression levels were greater in the responders than in the poor responders. Except two genes that were downregulated in the responders (*RET* and *FLT4*), the remaining eight sorafenib target genes that were upregulated were used to construct the molecular signature. Complex models with relatively small number of samples are often overfitted; therefore, we calculated the signature score as sum of log10-transformed normalized gene expression values, thus reducing data dimensionality.

This signature was tested to predict sorafenib response status in the microarray-profiled dataset. To assess the signature biomarker quality, we used AUC value as the measure. AUC reflects biomarker robustness and depends on its sensitivity and specificity (Borisov et al., 2020). It varies between 0.5 and 1, and the typical discrimination threshold is 0.7, where greater values denote high-quality biomarkers, and *vice versa* (Boyd, 1997). AUC is often used for scoring different types of molecular

TABLE 5 | Differential expression analysis of sorafenib responders (n = 12) and poor responders (n = 10) in microarray-profiled RCC samples

HGNC gene ID	*T*-test *p*-value	Log_2(fold change responders *vs.* poor responders)
RAF1	0.56	0.072
BRAF	0.37	0.071
RET	0.54	−0.037
FLT1	0.0032*	1.155
FGFR1	0.1	0.149
KIT	0.72	0.029
PDGFRB	0.013	0.273
FLT3	0.67	0.011
FLT4	0.54	−0.010
KDR	0.2	0.119

*p < *0.05*.

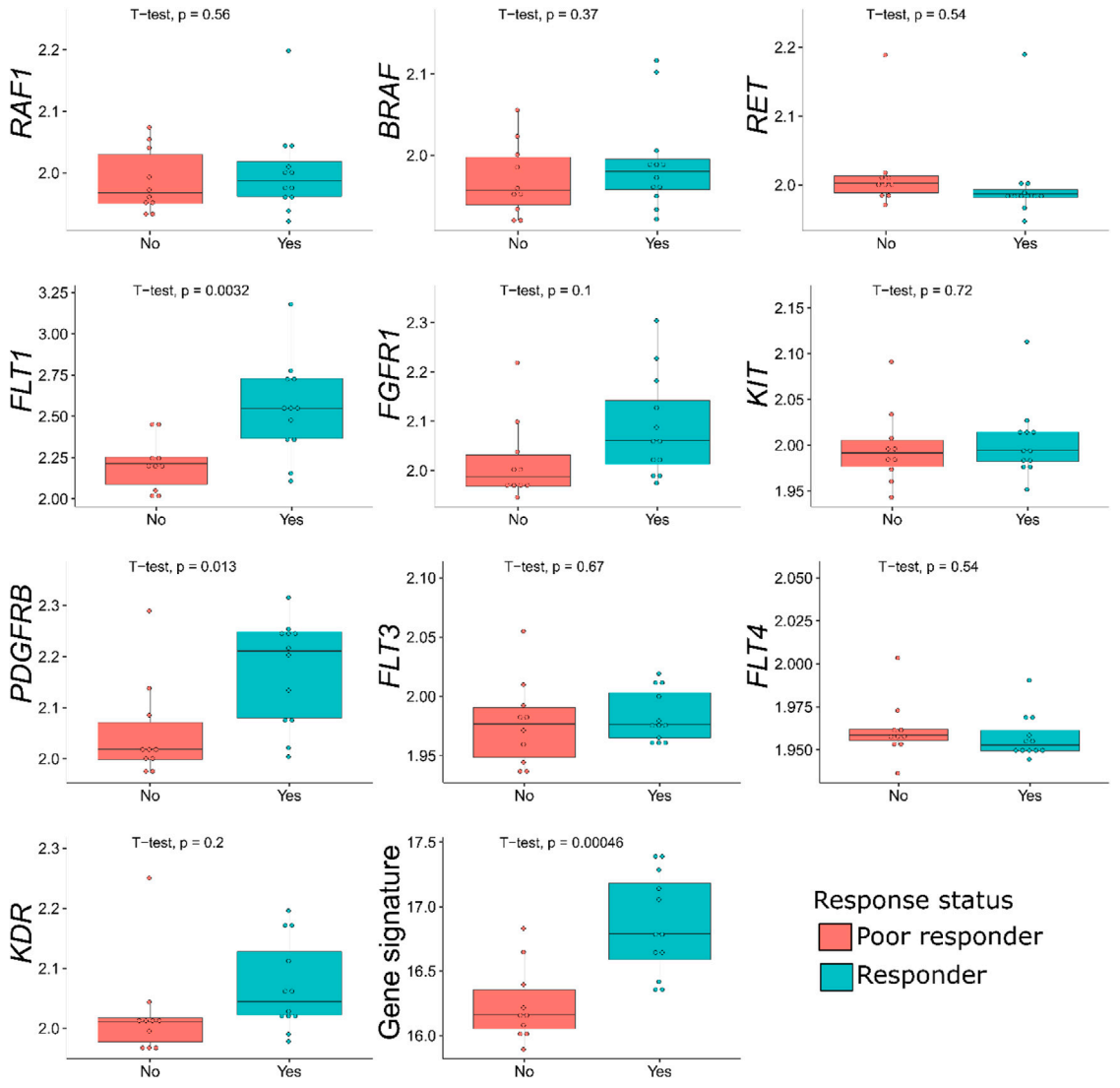


FIGURE 1 | Distribution of sorafenib target gene expressions, and of the gene signature generated, among the sorafenib responder and poor responder groups of 22 RCC samples profiled by microarrays. For every gene, log10-transformed normalized expression is shown.

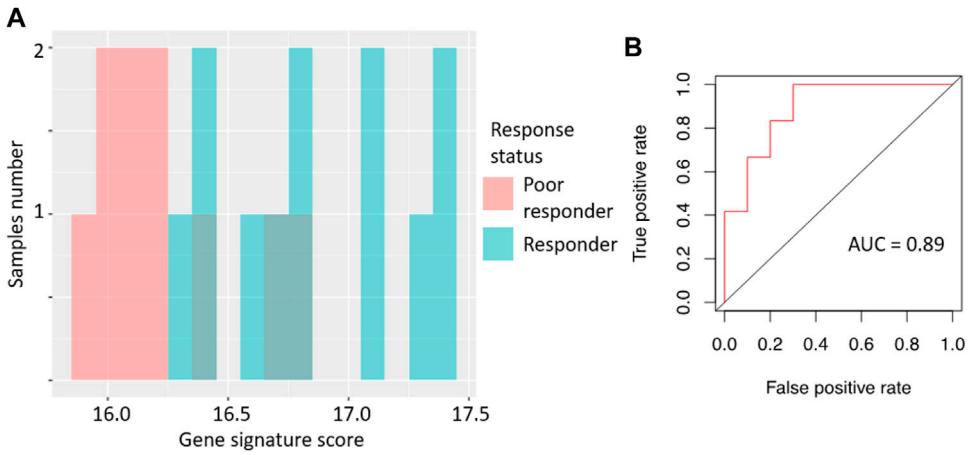


FIGURE 2 | Performance of sorafenib response gene signature in microarray-profiled RCC set. **(A)** Distribution of gene signature score in 22 RCC samples profiled by expression microarrays. **(B)** ROC (receiver operating characteristic) curve for prediction of response status by gene signature score in 22 RCC samples profiled by expression microarrays. Validation of sorafenib response gene signature.

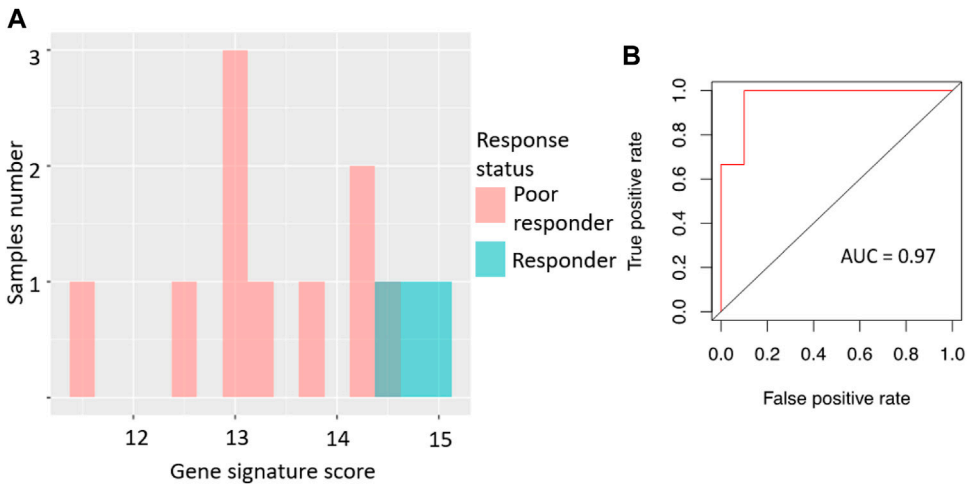


FIGURE 3 | Performance of sorafenib response gene signature in RNAseq-profiled RCC set. **(A)** Distribution of gene signature score in 13 RCC samples profiled by RNA sequencing. **(B)** ROC (receiver operating characteristic) curve for prediction of response status by gene signature score in 13 RCC samples profiled by RNA sequencing.

biomarkers in oncology (Liu et al., 2018; Tanioka et al., 2018; Chen et al., 2019; Sorokin et al., 2020a).

For the gene signature biomarker capacity, we obtained AUC value 0.89 (**Figure 2B**), which evidences its high prediction robustness. Using an assumption of equal frequency of type I and type II errors, we obtained threshold gene signature score of 16.41. This threshold corresponded to sensitivity 0.83, specificity 0.8, and Matthew's correlation coefficient (MCC) 0.63; error matrix is shown on **Supplementary Table S2**. Interestingly, *t*-test *p*-value of the gene signature for comparison between the good and poor sorafenib responders (p = 0.00046) was lower than the respective *p*-value for any of the single sorafenib target genes (**Figure 1**).

We then tested the ability of the sorafenib response gene signature to predict good/poor response status using two alternative experimental platforms (RNAseq and qRT-PCR) and different sets of annotated RCC biosamples (n = 13 and n = 12, respectively).

Specifically, the signature score for RNAseq data was calculated in the same way as for the microarray dataset: sum of the log-transformed expression values for the same eight sorafenib target genes. For the qRT-PCR dataset, we totalized ΔCt values for the selected sorafenib targets. For those two platforms, we obtained AUC scores of 0.97 and 0.83, respectively (**Figures 3**, **4**; error matrices are shown on **Supplementary Tables S3, S4**, respectively).

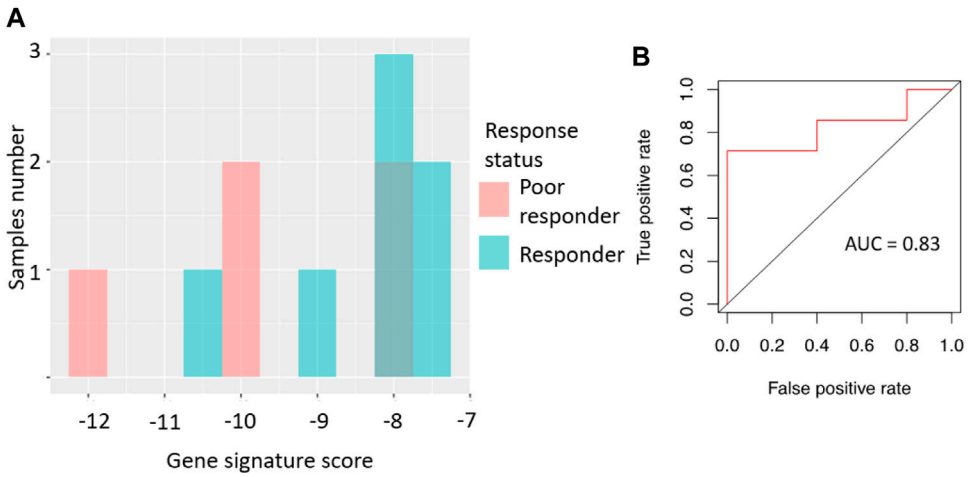


FIGURE 4 | Performance of sorafenib response gene signature in microarray-profiled RCC set. **(A)** Distribution of gene signature score in 22 RCC samples profiled by expression microarrays. **(B)** ROC (receiver-operator characteristic) curve for prediction of response status by gene signature score in 22 RCC samples profiled by expression microarrays. *In vitro* validation of sorafenib response gene signature.

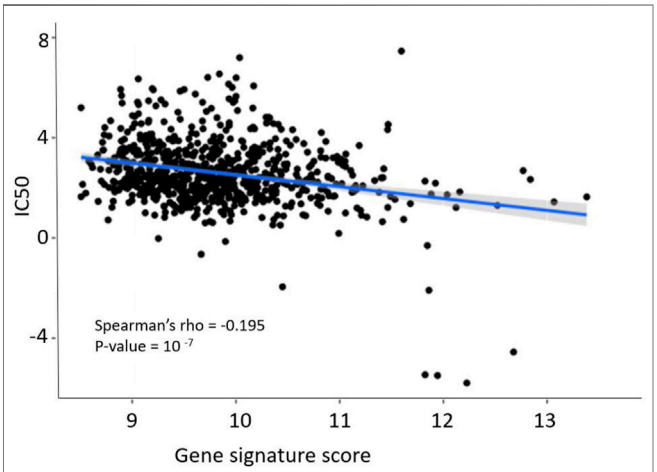


FIGURE 5 | Dependence of sorafenib resistance gene signature score and sorafenib IC_{50} in GDSC pan-cancer dataset. Blue line and shadow around it show linear approximation and 5% confidence interval. Figure built using *ggplot* function in R.

For the RNAseq dataset, we also used an assumption of equal frequency of type I and type II errors and obtained threshold gene signature score of 14.35, sensitivity 1.00, specificity 0.9, and MCC 0.82. Similarly, in the case of qRT-PCR dataset, a threshold of −8.03 was obtained, which corresponded to sensitivity 0.71, specificity 1.0, and MCC 0.71.

The high scores of AUC, MCC, sensitivity, and specificity values observed for all three cohorts suggest in favor of the proposed sorafenib response gene signature usefulness as the new combinatorial expression biomarker.

We further validated the sorafenib response gene signature using bioinformatics analysis of publicly available cell line gene expression data annotated with sorafenib sensitivity information. We calculated molecular signature scores for 735 samples of different cancer cell lines extracted through the GDSC database (Yang et al., 2013).

We then compared gene signature scores with the log10-transformed IC_{50} micromolar values of sorafenib. IC_{50} shows sorafenib concentration that reduces cell viability by 50%, and therefore IC_{50} is an inverse measure of drug efficiency (high IC_{50} suggests strong drug resistance, and low IC_{50} means high sensitivity to a drug). We observed a statistically significant negative correlation between sorafenib IC_{50} and gene signature score (**Figure 5**), Spearman correlation −0.195, $p = 10^{-7}$.

We then modeled ability of the gene signature to predict sorafenib IC_{50} in the tumor cell lines. Using GDSC data, we selected top and bottom 5% cell lines by sorafenib IC_{50}, and associated them with the sorafenib poor and good responders, respectively. In this test, AUC value for prediction of high or low sorafenib IC_{50} by gene signature score was 0.77 (**Figure 6**).

Based on the assumption of equal importance of type I and type II errors, in this setting we obtained gene signature score threshold of 9.8, MCC coefficient 0.32, sensitivity 0.63, and specificity 0.66; error matrix for this analysis is shown on **Supplementary Table S5**.

Mutations Associated With Sorafenib Activity *In Vitro*

Using the GDSC dataset, we further investigated the connection between sorafenib IC_{50} and annotated mutations in the GDSC cell lines. *P*-value cut-off was set according to GDSC default parameters (threshold FDR corrected $p < 0.1$ and fold change >2). With the internal GDCS analytic interface, we identified mutations in two genes that were statistically significantly linked with IC_{50} of sorafenib: *FLT3* and *SMARCA4* (**Figure 7**; **Table 6**). The observed genetic features contained different driver mutations in both genes. Noteworthy, *FLT3* gene product is one of the molecular targets of sorafenib. Thus, strong linkage of driver mutations in this gene with the sensitivity to sorafenib directly confirms its implication

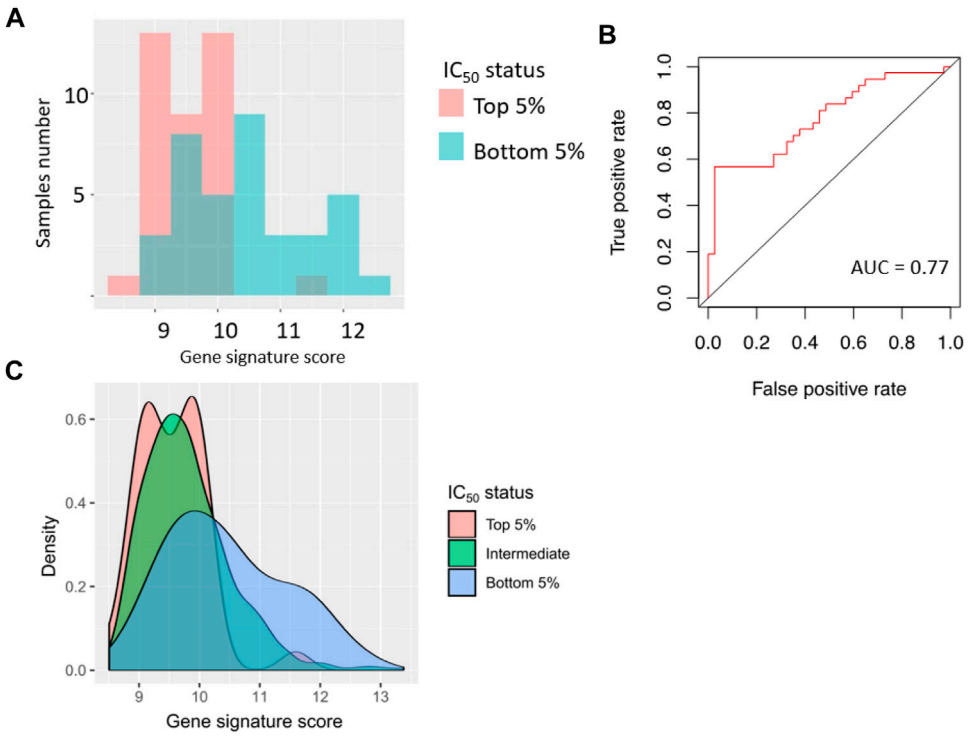


FIGURE 6 | Performance of sorafenib response gene signature in 76 cancer cell lines (top 5% and bottom 5% cell lines from GDSC dataset, sorted by sorafenib IC_{50}). **(A)** Distribution of gene signature score in 76 cancer cell lines. **(B)** ROC (receiver operating characteristic) curve for prediction of response status by gene signature score in 76 cancer cell lines. **(C)** Distribution of gene signature score in all cancer cell lines.

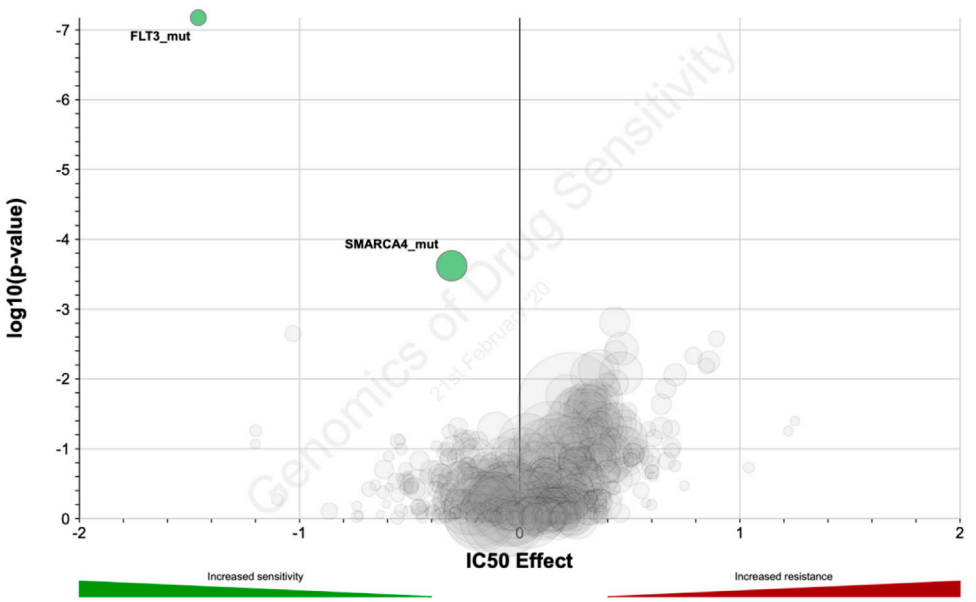


FIGURE 7 | Distribution of log10-transformed *p*-value and IC_{50} difference between groups with and without gene-specific mutations and copy number alternations in GDSC database.

TABLE 6 | Genes which mutations are statistically significantly associated with sorafenib IC_{50} in GDSC data collection

Gene mutation	IC_{50} effect size (fold change)	*p*-value	FDR	Number of altered cell lines
*FLT3*_mut	−1.46	6.67×10^{-08}	0.000042	11 (1.6%)
*SMARCA4*_mut	−0.309	0.000239	0.0754	41 (6%)

in the mechanisms of cancer cells' resistance to sorafenib. For the second gene (*SMARCA4*), we found no previous associations with sorafenib efficacy in the literature. However, molecular function of this gene product is ATP-dependent chromatin remodeling and overall transcriptional activation, and *SMARCA4* mutations are linked with many cancers (Fountzilas et al., 2021; Nambirajan and Jain, 2021; Pastorczak et al., 2021).

DISCUSSION

Sorafenib is a targeted tyrosine kinase inhibitor (TKI) with multiple molecular specificities, which is widely used to treat kidney cancer due to relative clinical efficacy and affordability (Sheng et al., 2016; Cai et al., 2017). However, sorafenib response rate in RCC varies between 10 and 40% (Larkin and Eisen, 2006; Escudier et al., 2007; Guevremont et al., 2009), thus personalized approach is needed to select the patients who would more likely benefit from the treatment with this drug.

High-throughput gene expression profiling is becoming a powerful tool for finding new cancer biomarkers (Buzdin et al., 2019; Tsimberidou et al., 2020). Moreover, aggregating gene expression levels into functional groups like molecular pathways or gene signatures can increase efficiency of the biomarkers and even enhance stability of experimental data (Borisov et al., 2017; Buzdin et al., 2018). Previously we used this approach to establish biomarkers of trastuzumab response in metastatic/recurrent HER2-positive breast cancers (Sorokin et al., 2020a), ramucirumab response in gastric cancer (Sorokin et al., 2020d), and for building gene signature for ganglioside GD2 expression in cancer cells (Sorokin et al., 2020c).

In this study, we identified and validated an 8-gene expression signature that predicts sorafenib response in RCC patients. The signature was validated on the independent patient groups using three different methods of gene expression profiling: by Illumina HT-12 microarrays, by RNA sequencing, and by qRT-PCR. The sorafenib response signature includes eight sorafenib target genes: *RAF1*, *BRAF*, *FLT1*, *FGFR1*, *KIT*, *PDGFRB*, *FLT3*, and *KDR*. Among them, increased expressions of single genes *FLT1* and *PDGFRB* were positively associated with the sorafenib response, whereas other genes showed similar trends, which were however not statistically significant. At the same time, the gene signature could show better efficacy than any of the separately taken enclosing genes, thus evidencing better efficacy of a cumulative complex expression biomarker. On the other hand, significant association of sorafenib target gene *FLT3* was confirmed at the level of driver mutations in GDSC data, thus implying a peculiar role for this gene in the sorafenib activity mechanism.

Sorafenib has a strong overlap in the molecular specificities with regorafenib (Granito et al., 2021) and with several other TKI drugs (Shah et al., 2020; Das et al., 2021), and theoretically the same drug target–based gene signature approach can be translated on finding new response biomarkers for other TKIs as well, and for different cancer types. However, such an approach would require accumulating enough tumor gene expression data connected with the specific drug response statuses, which is frequently a difficult task to implement. For example, to the best of our knowledge, the high-throughput experimental expression profiles that were associated here with the sorafenib response are the first such RCC dataset published in the literature. Accumulation and publishing of more molecular profiles connected with the TKI response statuses in RCC and other tumors would clearly enhance development of next-generation drug response prediction biomarkers.

For the current sorafenib 8-gene expression signature, we developed a qRT-PCR–based diagnostic panel that enables cost-effective molecular profiling. The panel was validated on an independent cohort of RCC patients with AUC = 0.83, which opens an avenue for further molecular testing on bigger patient cohorts and, if successful, for the development of diagnostic tools supporting personalized sorafenib prescriptions. Such a study would also be needed to validate the exact threshold developed for qRT-PCR signature established herein.

Interestingly, the same 8-gene signature was also validated using GDSC project cell line gene expression data connected with the tested drug sensitivities (Yang et al., 2013): a modest (−0.195) yet highly statistically significant ($p = 10^{-7}$) correlation was observed for the gene signature score and sorafenib IC_{50}. The GDSC collection accumulated data for various cancer cell lines. Cell lines are heterogeneous and derived from tumors of various origin, not only kidney cancer. In addition, *in vitro* culturing may have an impact on gene expression. Despite all these factors, we still obtained statistically significant performance of the gene signature. Potentially, this may indicate that this gene signature is not specific to RCC but may be also predictive for the other cancer types. Thus, further clinical investigations are needed to assess its performance in cancers other than RCC.

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/bioproject/PRJNA749745/, https://www.ncbi.nlm.nih.gov/geo/, GSE180925.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Vitamed clinic ethical committee. 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

AGu, VS, KK, DS, MSe, and NG collected and interpreted patient data; NB, DN, and MSo performed bioinformatical analysis; AA and MSu performed molecular analyses; AGa, EK, and AB supervised the study; AB, MSo, and DN wrote the paper.

FUNDING

This study was financed by the Ministry of Science and Higher Education of the Russian Federation within the framework of state support for the creation and development of World-Class Research Centers "Digital biodesign and personalized healthcare" No. 075-15-2020-926.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.753318/full#supplementary-material

REFERENCES

Abrahao-Machado, L. F., and Scapulatempo-Neto, C. (2016). HER2 Testing in Gastric Cancer: An Update. *Wjg* 22, 4619–4625. doi:10.3748/WJG.V22.I19.4619

Adnane, L., Trail, P. A., Taylor, I., and Wilhelm, S. M. (2006). Sorafenib (BAY 43-9006, Nexavar), a Dual-Action Inhibitor that Targets RAF/MEK/ERK Pathway in Tumor Cells and Tyrosine Kinases VEGFR/PDGFR in Tumor Vasculature. *Methods Enzymol.* 407, 597–612. doi:10.1016/S0076-6879(05)07047-3

Aziz, S. A., Sznol, J. A., Albiges, L., Zito, C., Jilaveanu, L. B., Camp, R. L., et al. (2014). Microvessel Area as a Predictor of Sorafenib Response in Metastatic Renal Cell Carcinoma. *Cancer Cel Int* 14. doi:10.1186/1475-2867-14-4

Bao, Y., Yang, F., Liu, B., Zhao, T., Xu, Z., Xiong, Y., et al. (2018). Angiopoietin-like Protein 3 Blocks Nuclear Import of Fak and Contributes to Sorafenib Response. *Br. J. Cancer* 119, 450–461. doi:10.1038/s41416-018-0189-4

Bhatt, J. R., and Finelli, A. (2014). Landmarks in the Diagnosis and Treatment of Renal Cell Carcinoma. *Nat. Rev. Urol.* 11, 517–525. doi:10.1038/nrurol.2014.194

Bolstad, B. M., Irizarry, R. A., Åstrand, M., and Speed, T. P. (2003). A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Variance and Bias. *Bioinformatics* 19, 185–193. doi:10.1093/bioinformatics/19.2.185

Borisov, N., Sorokin, M., Tkachev, V., Garazha, A., and Buzdin, A. (2020). Cancer Gene Expression Profiles Associated with Clinical Outcomes to Chemotherapy Treatments. *BMC Med. Genomics* 13. doi:10.1186/s12920-020-00759-0

Borisov, N., Suntsova, M., Sorokin, M., Garazha, A., Kovalchuk, O., Aliper, A., et al. (2017). Data Aggregation at the Level of Molecular Pathways Improves Stability of Experimental Transcriptomic and Proteomic Data. *Cell Cycle* 16, 1810–1823. doi:10.1080/15384101.2017.1361068

Boyd, J. (1997). Mathematical Tools for Demonstrating the Clinical Usefulness of Biochemical Markers. *Scli* 57, 46–63. doi:10.1080/00365519709168308

Buzdin, A., Sorokin, M., Garazha, A., Glusker, A., Aleshin, A., Poddubskaya, E., et al. (2020). RNA Sequencing for Research and Diagnostics in Clinical Oncology. *Semin. Cancer Biol.* 60, 311–323. doi:10.1016/j.semcancer.2019.07.010

Buzdin, A., Sorokin, M., Garazha, A., Sekacheva, M., Kim, E., Zhukov, N., et al. (2018). Molecular Pathway Activation - New Type of Biomarkers for Tumor Morphology and Personalized Selection of Target Drugs. *Semin. Cancer Biol.* 53, 110–124. doi:10.1016/j.semcancer.2018.06.003

Cai, W., Kong, W., Dong, B., Zhang, J., Chen, Y., Xue, W., et al. (2017). Comparison of Efficacy, Safety, and Quality of Life between Sorafenib and Sunitinib as First-Line Therapy for Chinese Patients with Metastatic Renal Cell Carcinoma. *Chin. J. Cancer* 36. doi:10.1186/S40880-017-0230-7

Chang, C.-W., Cheng, W.-C., Chen, C.-R., Shu, W.-Y., Tsai, M.-L., Huang, C.-L., et al. (2011). Identification of Human Housekeeping Genes and Tissue-Selective Genes by Microarray Meta-Analysis. *PLoS One* 6, e22859. doi:10.1371/journal.pone.0022859

Chen, L., Zhou, Y., Tang, X., Yang, C., Tian, Y., Xie, R., et al. (2019). EGFR Mutation Decreases FDG Uptake in Non-small Cell Lung Cancer via the NOX4/ROS/GLUT1 axis. *Int. J. Oncol.* 54, 370–380. doi:10.3892/ijo.2018.4626

Choueiri, T. K., and Kaelin, W. G. (2020). Targeting the HIF2-VEGF axis in Renal Cell Carcinoma. *Nat. Med.* 26, 1519–1530. doi:10.1038/s41591-020-1093-z

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

Das, A., Mahapatra, S., Bandyopadhyay, D., Samanta, S., Chakraborty, S., Philpotts, L. L., et al. (2021). Bleeding with Vascular Endothelial Growth Factor Tyrosine Kinase Inhibitor: A Network Meta-Analysis. *Crit. Rev. Oncology/Hematology* 157, 103186. doi:10.1016/j.critrevonc.2020.103186

Deng, H., Liu, W., He, T., Hong, Z., Yi, F., Wei, Y., et al. (2019). Comparative Efficacy, Safety, and Costs of Sorafenib vs. Sunitinib as First-Line Therapy for Metastatic Renal Cell Carcinoma: A Systematic Review and Meta-Analysis. *Front. Oncol.* 9. doi:10.3389/fonc.2019.00479

Dizman, N., Philip, E. J., and Pal, S. K. (2020). Genomic Profiling in Renal Cell Carcinoma. *Nat. Rev. Nephrol.* 16, 435–451. doi:10.1038/s41581-020-0301-x

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

Du, Z., Chen, W., Xia, Q., Shi, O., and Chen, Q. (2020). Trends and Projections of Kidney Cancer Incidence at the Global and National Levels, 1990-2030: a Bayesian Age-Period-Cohort Modeling Study. *Biomark. Res.* 8, 16. doi:10.1186/s40364-020-00195-3

Eisenhauer, E. A., Therasse, P., Bogaerts, J., Schwartz, L. H., Sargent, D., Ford, R., et al. (2009). New Response Evaluation Criteria in Solid Tumours: Revised RECIST Guideline (Version 1.1). *Eur. J. Cancer* 45, 228–247. doi:10.1016/J.EJCA.2008.10.026

Escudier, B., Eisen, T., Stadler, W. M., Szczylik, C., Oudard, S., Siebels, M., et al. (2007). Sorafenib in Advanced Clear-Cell Renal-Cell Carcinoma. *N. Engl. J. Med.* 356, 125–134. doi:10.1056/NEJMoa060655

Escudier, B., Worden, F., and Kudo, M. (2019). Sorafenib: Key Lessons from over 10 Years of Experience. *Expert Rev. Anticancer Ther.* 19, 177–189. doi:10.1080/14737140.2019.1559058

Fawcett, T. (2006). An Introduction to ROC Analysis. *Pattern Recognition Lett.* 27, 861–874. doi:10.1016/J.PATREC.2005.10.010

Fountzilas, E., Kurzrock, R., Vo, H. H., and Tsimberidou, A.-M. (2021). Wedding of Molecular Alterations and Immune Checkpoint Blockade: Genomics as a Matchmaker. *J. Natl. Cancer Inst.* 113, 1634–1647. doi:10.1093/jnci/djab067

Granito, A., Forgione, A., Marinelli, S., Renzulli, M., Ielasi, L., Sansone, V., et al. (2021). Experience with Regorafenib in the Treatment of Hepatocellular Carcinoma. *Therap. Adv. Gastroenterol.* 14, 175628482110169. doi:10.1177/17562848211016959

Guevremont, C., Jeldres, C., Perrotte, P., and Karakiewicz, P. I. (2009). Sorafenib in the Management of Metastatic Renal Cell Carcinoma. *Curr. Oncol.* 16 (Suppl. 1), 27–32. doi:10.3747/co.v16i0.430

He, Y., Luo, Y., Huang, L., Zhang, D., Wang, X., Ji, J., et al. (2021). New Frontiers against Sorafenib Resistance in Renal Cell Carcinoma: From Molecular Mechanisms to Predictive Biomarkers. *Pharmacol. Res.* 170, 105732. doi:10.1016/j.phrs.2021.105732

Hechtman, J. F., Benayed, R., Hyman, D. M., Drilon, A., Zehir, A., Frosina, D., et al. (2017). Pan-Trk Immunohistochemistry Is an Efficient and Reliable Screen for the Detection of NTRK Fusions. *Am. J. Surg. Pathol.* 41, 1547–1551. doi:10.1097/PAS.0000000000000911

Hira, Z. M., and Gillies, D. F. (2015). A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data. *Adv. Bioinformatics* 2015, 1–13. doi:10.1155/2015/198363

Ho, T. H., Liu, X.-D., Huang, Y., Warneke, C. L., Johnson, M. M., Hoang, A., et al. (2015). The Impact of FGFR1 and FRS2α Expression on Sorafenib Treatment in Metastatic Renal Cell Carcinoma. *BMC Cancer* 15. doi:10.1186/s12885-015-1302-1

Jonasch, E., Corn, P., Pagliaro, L. C., Warneke, C. L., Johnson, M. M., Tamboli, P., et al. (2009). Upfront, Randomized, Phase 2 Trial of Sorafenib versus Sorafenib and Low-Dose Interferon Alfa in Patients with Advanced Renal Cell Carcinoma. *Cancer* 116, NA. doi:10.1002/cncr.24685

Larkin, J. M., and Eisen, T. (2006). Renal Cell Carcinoma and the Use of Sorafenib. *Ther. Clin. Risk Manag.* 2, 87–98.

Lezhnina, K., Kovalchuk, O., Zhavoronkov, A. A., Korzinkin, M. B., Zabolotneva, A. A., Shegay, P. V., et al. (2014). Novel Robust Biomarkers for Human Bladder Cancer Based on Activation of Intracellular Signaling Pathways. *Oncotarget* 5, 9022–9032. doi:10.18632/oncotarget.2493

Li, X. L., Chen, X. Q., Zhang, M. N., Chen, N., Nie, L., Xu, M., et al. (2015). SOX9 Was Involved in TKIs Resistance in Renal Cell Carcinoma via Raf/MEK/ERK Signaling Pathway. *Int. J. Clin. Exp. Pathol.* 8, 3871–3881. Available at: https://pubmed.ncbi.nlm.nih.gov/26097571.

Liu, L., Pang, X., Shang, W., Xie, H., Feng, Y., and Feng, G. (2019). Long Noncoding RNA GAS5 Sensitizes Renal Cell Carcinoma to Sorafenib via miR-21/SOX5 Pathway. *Cell Cycle* 18, 257–263. doi:10.1080/15384101.2018.1475826

Liu, T., Cheng, G., Kang, X., Xi, Y., Zhu, Y., Wang, K., et al. (2018). Noninvasively Evaluating the Grading and IDH1 Mutation Status of Diffuse Gliomas by Three-Dimensional Pseudo-continuous Arterial Spin Labeling and Diffusion-Weighted Imaging. *Neuroradiology* 60, 693–702. doi:10.1007/s00234-018-2021-5

Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8

Méndez-Blanco, C., Fondevila, F., García-Palomo, A., González-Gallego, J., and Mauriz, J. L. (2018). Sorafenib Resistance in Hepatocarcinoma: Role of Hypoxia-Inducible Factors. *Exp. Mol. Med.* 50, 1–9. doi:10.1038/s12276-018-0159-1

Nambirajan, A., and Jain, D. (2021). Recent Updates in Thoracic SMARCA4-Deficient Undifferentiated Tumor. *Semin. Diagn. Pathol.* 38, 83–89. doi:10.1053/j.semdp.2021.06.001

Nicolini, A., Ferrari, P., and Duffy, M. J. (2018). Prognostic and Predictive Biomarkers in Breast Cancer: Past, Present and Future. *Semin. Cancer Biol.* 52, 56–73. doi:10.1016/j.semcancer.2017.08.010

Pastorczak, A., Krajewska, K., Urbanska, Z., Szmyd, B., Salacinska-Los, E., Kobos, J., et al. (2021). Ovarian Carcinoma in Children with Constitutional Mutation of SMARCA4: Single-Family Report and Literature Review. *Fam. Cancer* 20, 355–362. doi:10.1007/s10689-021-00258-w

Sánchez-Gastaldo, A., Kempf, E., González del Alba, A., and Duran, I. (2017). Systemic Treatment of Renal Cell Cancer: A Comprehensive Review. *Cancer Treat. Rev.* 60, 77–89. doi:10.1016/j.ctrv.2017.08.010

Shah, A. A., Kamal, M. A., and Akhtar, S. (2021). Tumor Angiogenesis and VEGFR-2: Mechanism, Pathways and Current Biological Therapeutic Interventions. *Cdm* 22, 50–59. doi:10.2174/1389200221666201019143252

Sheng, X., Chi, Z., Cui, C., Si, L., Li, S., Tang, B., et al. (2016). Efficacy and Safety of Sorafenib versus Sunitinib as First-Line Treatment in Patients with Metastatic Renal Cell Carcinoma: Largest Single-center Retrospective Analysis. *Oncotarget* 7, 27044–27054. doi:10.18632/ONCOTARGET.7395

Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T. (2005). ROCR: Visualizing Classifier Performance in R. *Bioinformatics* 21, 3940–3941. doi:10.1093/bioinformatics/bti623

Sorokin, M., Ignatev, K., Barbara, V., Vladimirova, U., Muraveva, A., Suntsova, M., et al. (2020a). Molecular Pathway Activation Markers Are Associated with Efficacy of Trastuzumab Therapy in Metastatic HER2-Positive Breast Cancer Better Than Individual Gene Expression Levels. *Biochem. Mosc.* 85, 758–772. doi:10.1134/S0006297920070044

Sorokin, M., Ignatev, K., Poddubskaya, E., Vladimirova, U., Gaifullin, N., Lantsov, D., et al. (2020b). RNA Sequencing in Comparison to Immunohistochemistry for Measuring Cancer Biomarkers in Breast Cancer and Lung Cancer Specimens. *Biomedicines* 8, 114. doi:10.3390/BIOMEDICINES8050114

Sorokin, M., Kholodenko, I., Kalinovsky, D., Shamanskaya, T., Doronin, I., Konovalov, D., et al. (2020c). RNA Sequencing-Based Identification of Ganglioside GD2-Positive Cancer Phenotype. *Biomedicines* 8, 142. doi:10.3390/BIOMEDICINES8060142

Sorokin, M., Poddubskaya, E., Baranova, M., Glusker, A., Kogoniya, L., Markarova, E., et al. (2020d). RNA Sequencing Profiles and Diagnostic Signatures Linked with Response to Ramucirumab in Gastric Cancer. *Cold Spring Harb. Mol. Case Stud.* 6, a004945. doi:10.1101/mcs.a004945

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

Suntsova, M., Gaifullin, N., Allina, D., Reshetun, A., Li, X., Mendeleeva, L., et al. (2019). Atlas of RNA Sequencing Profiles for normal Human Tissues. *Sci. Data* 6, 36. doi:10.1038/s41597-019-0043-4

Tanioka, M., Fan, C., Parker, J. S., Hoadley, K. A., Hu, Z., Li, Y., et al. (2018). Integrated Analysis of RNA and DNA from the Phase III Trial CALGB 40601 Identifies Predictors of Response to Trastuzumab-Based Neoadjuvant Chemotherapy in HER2-Positive Breast Cancer. *Clin. Cancer Res.* 24, 5292–5304. doi:10.1158/1078-0432.CCR-17-3431

Tkachev, V., Sorokin, M., Garazha, A., Borisov, N., Buzdin, A., Tkachev, V., et al. (2020). Oncobox Method for Scoring Efficiencies of Anticancer Drugs Based on Gene Expression Data. *Methods Mol. Biol.* 2063, 235–255. doi:10.1007/978-1-0716-0138-9_17

Tsimberidou, A. M., Fountzilas, E., Bleris, L., and Kurzrock, R. (2020). Transcriptomics and Solid Tumors: The Next Frontier in Precision Cancer Medicine. *Semin. Cancer Biol.* doi:10.1016/j.semcancer.2020.09.007

Wilhelm, S., Carter, C., Lynch, M., Lowinger, T., Dumas, J., Smith, R. A., et al. (2006). Discovery and Development of Sorafenib: A Multikinase Inhibitor for Treating Cancer. *Nat. Rev. Drug Discov.* 5, 835–844. doi:10.1038/nrd2130

Wilhelm, S. M., Carter, C., Tang, L., Wilkie, D., McNabola, A., Rong, H., et al. (2004). BAY 43-9006 Exhibits Broad Spectrum Oral Antitumor Activity and Targets the RAF/MEK/ERK Pathway and Receptor Tyrosine Kinases Involved in Tumor Progression and Angiogenesis. *Cancer Res.* 64, 7099–7109. doi:10.1158/0008-5472.CAN-04-1443

Xu, Z., Yang, F., Wei, D., Liu, B., Chen, C., Bao, Y., et al. (2017). Long Noncoding RNA-SRLR Elicits Intrinsic Sorafenib Resistance via Evoking IL-6/STAT3 axis in Renal Cell Carcinoma. *Oncogene* 36, 1965–1977. doi:10.1038/onc.2016.356

Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a Resource for Therapeutic Biomarker Discovery in Cancer Cells. *Nucleic Acids Res.* 41, D955–D961. doi:10.1093/nar/gks1111

Yates, B., Braschi, B., Gray, K. A., Seal, R. L., Tweedie, S., and Bruford, E. A. (2017). Genenames.org: The HGNC and VGNC Resources in 2017. *Nucleic Acids Res.* 45, D619–D625. doi:10.1093/nar/gkw1033

Yuan, J.-L., Zheng, W.-X., Yan, F., Xue, Q., Wu, G.-J., Qin, W.-J., et al. (2015). Heme Oxygenase-1 Is a Predictive Biomarker for Therapeutic Targeting of Advanced clear Cell Renal Cell Carcinoma Treated with Sorafenib or Sunitinib. *Ott* 8, 2081–2088. doi:10.2147/OTT.S86222

Conflict of Interest: DN and AA were employed by the company Oncobox Ltd. AG, AB, and MS (14th author) were employed by the company OmicsWay Corp.

The reviewer (YLO) declared a shared affiliation, with no collaboration, with several of the authors (AGu, MS, MS, AAB, MS) to the handling editor at the time of the review.

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 Molecular Biosciences

ORIGINAL RESEARCH
published: 24 May 2022
doi: 10.3389/fmolb.2022.895927

Long Non-Coding RNA AP000695.2 Acts as a Novel Prognostic Biomarker and Regulates the Cell Growth and Migration of Lung Adenocarcinoma

Chunyan Wang[1†], Jishu Guo[2†], Rongyan Jiang[3], Chenyang Wang[1], Chenglong Pan[1], Zhi Nie[4,5] and Xiulin Jiang[6*]*

[1]Department of Pathology, First Affiliated Hospital of Kunming Medical University, Kunming, China, [2]School of Ecology and Environmental Science, Yunnan University, Kunming, China, [3]Department of Cardiovascular Medicine, the Bozhou Hospital Affiliated to Anhui Medical University, Bozhou Anhui, China, [4]Department of Neurology, First Affiliated Hospital of Kunming Medical University, Kunming, China, [5]Yunnan Province Clinical Research Center for Neurological Diseases, Kunming, China, [6]Kunming College of Life Science, University of Chinese Academy of Sciences, Beijing, China

Long non-coding RNAs (lncRNAs) are tumor-associated biological molecules and have been found to be implicated in the progression of lung adenocarcinoma (LUAD). LncRNA-AP000695.2 (ENSG00000248538) is a long non-coding RNA (lncRNA) that is widely increased in many tumor types including lung adenocarcinoma (LUAD). However, the aberrant expression profile, clinical significance, and biological function of AP000695.2 in human lung adenocarcinoma (LUAD) need to be further investigated. This study mines key prognostic AP000695.2 and elucidates its potential role and molecular mechanism in regulating the proliferation and metastasis of LUAD. Here, we discovered that AP000695.2 was significantly upregulated in lung adenocarcinoma tissues compared with healthy adjacent lung tissue and higher in LUAD cell lines than in normal human bronchial epithelial cell lines. A higher expression of AP000695.2 was positively correlated with aggressive clinicopathological characteristics, and AP000695.2 served as an independent prognostic indicator for the overall survival, disease-free survival, and progression-free survival in patients with LUAD. Receiver operating curve (ROC) analysis revealed the significant diagnostic ability of AP000695.2 (AUC = 0.838). Our *in vivo* data confirmed that AP000695.2 promotes the proliferation, migration, and invasion of LUAD cells. GSEA results suggested that AP000695.2 co-expressed genes were mainly enriched in immune-related biological processes such as JAK-STAT signaling pathway and toll-like receptor signaling pathway. Single-sample GSEA analysis showed that AP000695.2 is correlated with tumor-infiltrating immune cells in lung adenocarcinoma. Our findings confirmed that AP000695.2 was involved in the progression of lung adenocarcinoma, providing a novel prognostic indicator and promising diagnostic biomarker in the future.

Keywords: lncRNA, lung adenocarcinoma, prognosis biomarker, cell proliferation, cell apoptosis, cell growth, cell migration

OPEN ACCESS

Edited by:
William C. Cho,
QEH, Hong Kong SAR, China

Reviewed by:
Longhai Li,
Bozhou People's Hospital, China
Yu Peng,
Peking Union Medical College Hospital (CAMS), China
Chen Li,
Free University of Berlin, Germany

****Correspondence:***
Zhi Nie
niezhi@mail.kiz.ac.cn
Xiulin Jiang
1149155183@qq.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *14 March 2022*
Accepted: *07 April 2022*
Published: *24 May 2022*

Citation:
Wang C, Guo J, Jiang R, Wang C, Pan C, Nie Z and Jiang X (2022) Long Non-Coding RNA AP000695.2 Acts as a Novel Prognostic Biomarker and Regulates the Cell Growth and Migration of Lung Adenocarcinoma. Front. Mol. Biosci. 9:895927. doi: 10.3389/fmolb.2022.895927

INTRODUCTION

Lung cancer is the leading cause of cancer-related death worldwide, according to cancer statistics 2020. The incidence rate of lung cancer ranks second, while the death rate of lung cancer ranks first (Siegel et al., 2020). Lung cancer includes small cell lung carcinoma (SCLC) and non-small cell lung carcinoma (NSCLC). NSCLC includes lung adenocarcinoma (ADC), lung squamous cell carcinoma (SCC), and large-cell lung carcinoma. NSCLC accounts for approximately 85% of all cases (Molina et al., 2008). Despite various treatments applied to the diagnostic and therapeutic for lung cancer, the five-year survival rate of lung cancer remains poor (Reck and Rabe 2017). As an important treatment of lung cancer, chemotherapy is widely used in the clinical treatment of patients. However, the drug resistance of patients to chemotherapeutic drugs has become a major challenge to patient prognosis, leading to lung cancer recurrence and metastasis (Zhang et al., 2020). Cisplatin (DDP) was reported to be a common chemotherapeutic drug which plays an important role in the treatment of lung cancer (Dasari and Bernard Tchounwou 2014). Nevertheless, the drug resistance to chemotherapeutic drugs usually leads to poor treatment effects, leading to malignant progression and recurrence of LUAD (Fennell et al., 2016). Elucidating the complex molecular mechanism underlying drug resistance and identifying the key molecules regulating drug resistance are crucial for the treatment of lung cancer.

LncRNAs are a kind of ncRNAs whose transcripts with a length of more than 200 nucleotides do not have protein-coding potential. Mounting evidence has demonstrated that lncRNA abnormal expression and over-activation are usually involved in cancer initiation and progression (Zhang et al., 2020). LncRNAs can modulate the gene expression via influencing the structure of chromatin (Wang et al., 2011), histone modification (Luco et al., 2010), and to sponging microRNA (Yang et al., 2014). Aberrantly expressed lncRNAs have been reported to correlate with the development and progression of lung cancer (Chen et al., 2020a). It has been confirmed that lncRNA plays an important role in regulating cancer chemoresistance (Zhang et al., 2020). For example, Chen et al. found that lncRNA SNHG14 was upregulated in the A549/DDP cell line compared to that in the A549 cell line, and SNHG14 promotes the DDP-resistance of non-small cell lung cancer cells by modulating the miR-133a/HOXB13 signaling pathway (Xu et al., 2020). Furthermore, it has been confirmed that lincAK126698 improves the expression of β-catenin and promotes DDP-induced apoptosis (Yang et al., 2013). Moreover, it has been confirmed that lncRNA bladder cancer-associated transcript 1 was highly expressed in DDP-resistant NSCLC cells and promotes chemoresistance via modulating the miR-17/ATG7 axis (Huang et al., 2019). LncRNA-NNT-AS1 is elevated in DDP-resistant NSCLC tissues and cells, and the overexpression of lncRNA NNT-AS1 boosts cell proliferation and inhibits cell apoptosis by regulating the MAPK signaling pathway (Cai et al., 2018). Wang J. et al (2020) showed that LINC01116 was highly expressed in cisplatin-resistant LUAD specimens and A549/DDP cells. Depletion of LINC01116 reduced cell viability and cell migration, elevated cell apoptosis, and improved the sensitivity to DDP in A549/DDP cells. In our previous study, we developed a new method called CVAA (Cross-Value Association Analysis), which functions without a normalization and distribution assumption. We applied it to large-scale pan-cancer transcriptome data generated by The Cancer Genome Atlas (TCGA) project and successfully discovered numerous new differentially expressed genes (DEGs) (Jiang et al., 2021). AP000695.2 is one of these DEGs. Based on our analysis, AP000695.2 is a long non-coding RNA that is highly expressed in various human cancers, including lung cancer. However, the clinical significance and function of AP000695.2 associated with DDP resistance in LUAD remain to be elucidated.

In this study, we explored the diagnostic and prognostic significance of AP000695.2 in lung adenocarcinoma by data mining in The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO) datasets. Subsequently, gene set enrichment analysis (GSEA) was used to examine the possible biological functions and signaling pathways of AP000695.2 in lung adenocarcinoma. Moreover, we also examined the relationship between AP000695.2 expression and immune cell infiltration levels to explore the possible mechanisms by which AP000695.2 affected lung cancer occurrence and progression. Finally, cell viability assay, flow cytometry, colony formation, transwell, and wound healing assays were used to determine the biological function of AP000695.2 in lung adenocarcinoma.

MATERIALS AND METHODS

Data Collection

TCGA-LUAD dataset and clinical information of LUAD patients were downloaded from TCGA website (https://portal.gdc.cancer.gov/repository). AP000695.2 expression data from GSE81089 datasets were downloaded from the GEO database and validated AP000695.2 expression.

Nomogram Construction and Evaluation

Based on the multivariate Cox analysis results, we established a nomogram to predict the prognosis of LUAD patients. According to the prognosis model, we calculated each patient's risk score as the total score of each parameter, which could predict the prognosis of LUAD patients. The accuracy estimation of nomogram prediction was obtained from a calibration plot. The nomogram discrimination was determined using a concordance index (C-index), and 1,000 resamples were used in the calculation by the bootstrap approach. In this study, all statistical tests were two-tailed, with a statistical significance level of 0.05.

Gene Set Enrichment Analysis

Using the clusterProfiler package, the subtype-specific gene expression patterns and potential cellular pathways were elucidated on GSEA software (Subramanian et al., 2005). Based on the AP000695.2 expression level, we divided gene expression data into high-AP000695.2 and low-AP000695.2

groups, and each analysis included 1,000 times of gene set permutations. A p value of less than 0.05 was considered statistically significant.

Immune Infiltration Analysis by ssGSEA

We used a GSVA R package to examine the lung adenocarcinoma immune infiltration of 24 tumor-infiltrating immune cells in tumor samples through ssGSEA (Bindea et al., 2013; Hänzelmann et al., 2013). The correlation between AP000695.2 and infiltration levels of immune cells was analyzed by the Spearman correlation, and these immune cells with the different expression groups of AP000695.2 were analyzed by the Wilcoxon rank-sum test.

Cell Culture

The BEAS-2B cell line was purchased from the Cell Bank of Kunming Institute of Zoology and cultured in BEGM (Lonza, CC-3170). Lung cancer cell lines, including A549, H1299, and SPC-A1, were purchased from Cobioer, China, with STR documents, and were cultured in an RPMI-1640 medium (Corning) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin.

Constructs, Lentiviral Preparation, and Establishment of Different Cell Lines

For shRNA knockdown experiments, independent shRNAs targeting a different region of AP000695.2 RNA were constructed using a pLKO.1 vector (Addgene), and the oligo sequences were provided as follows: lentiviruses were generated according to the manufacturer's protocol, and indicated cells were infected by viruses twice with 48 and 72 h viral supernatants containing 4 μg/ml polybrene, and stable cell lines were established by appropriate puromycin selection. The two independent AP000695.2 targeting sequences are: shRNA#1, 5′-GTGTTGTATGACCCCGTTTTC-3' and shRNA#2, 5′-ATACGCACTGACAAACAACAC-3'.

Cell Proliferation Assays

For the cell proliferation assay, a total of 2×10^4 indicated cells were plated into 12-well plates in triplicate, and the exact cell numbers for each day were determined using an automatic cell analyzer countstar. Cell migration assay was performed as previously described (Jiang et al., 2022). To produce a wound, the monolayer cells in a 6-well plate were scraped in a straight line with pipette tips. The plate was then washed with warm PBS to remove detached cells. Photographs of the scratch were taken at indicated time points using a Nikon inverted microscope (Ti-S). The gap width was calculated with GraphPad Prism software. For transwell assay, $1–2\times10^4$ cells in 100 μL serum-free medium were plated in an 8.0-μm, 24-well plate chamber insert, with a medium containing 10% FBS at the bottom of the insert. The cells were incubated for 24 h and then fixed with 4% paraformaldehyde for 20 min. After washing, the cells were stained with 0.5% crystal violet-blue. The positively stained cells were examined under the microscope.

Real-Time RT-PCR Assay

A real-time RT-PCR assay was performed as previously described (Jiang et al., 2022). The primer used in this study is as follows: β-actin-F: AAGTGTGACGTGGACATCCGC, β-actin-R: CCGGACTCGTCATACTCCTGCT, AP000695.2-F: GATGAAAGACCGCGTTGTTT, and AP000695.2-R: CTCTTCGGAGGAAATAC.

Statistical Analysis

Data are represented as mean ± SEM, and error bars indicate SEM. p values were calculated by either unpaired or paired two-tailed Student's t-test, $^*p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$. All analyses were performed using GraphPad Prism software (GraphPad Software, Inc.).

RESULTS

The Expression and Prognostic Value of AP000695.2 in Pan-Cancer

To explore the expression of AP000695.2 in diverse human cancer, the AP000695.2 RNA expression data of multiple human cancers and normal tissues were examined. Based on the best cutoff score, we found that AP000695.2 expression was higher in various tumor tissues. As shown in **Figure 1A**, AP000695.2 expression was extremely significant in bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), cervical squamous cell carcinoma (CESC), cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), glioblastoma multiforme (GBM), 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), ovarian serous cystadenocarcinoma (OV), pancreatic adenocarcinoma (PAAD), skin cutaneous melanoma (SKCM), stomach adenocarcinoma (STAD), uterine corpus endometrial carcinoma (UCEC), and uterine carcinosarcoma (UCS) (**Figure 1A**). Furthermore, we found that AP000695.2 with high expression was associated with adverse clinical outcomes in ESAD, GBM, KICH, KIRC, LGG, LIHC, LUAD, MESO, PAAD, SARC, STAD, and UCEC (**Figures 1B–D**).

The Correlation Between AP000695.2 Expression Levels and Clinical Parameters in Lung Adenocarcinoma

We simultaneously analyzed the expression profiles of AP000695.2 (ENSG00000248538) based on TCGA database. Results confirmed that AP000695.2 was highly expressed in lung adenocarcinoma compared to the normal samples (**Figure 2A**). There were 59 pairs of lung adenocarcinoma cancer samples and matched adjacent normal samples from TCGA data. We found that AP000695.2 was elevated in lung adenocarcinoma samples than in matched adjacent normal samples (**Figure 2B**). Moreover, we found that AP000695.2

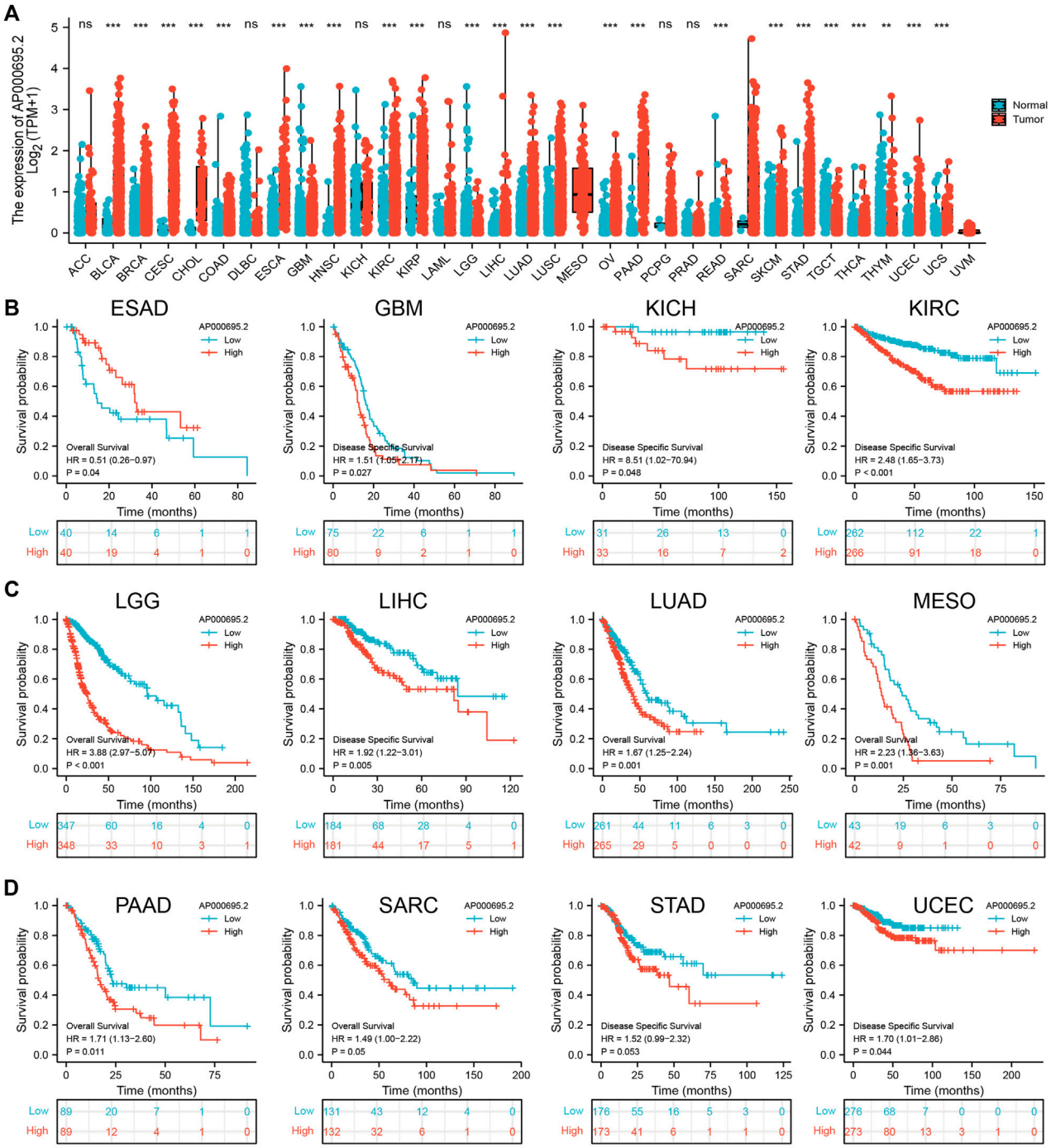


FIGURE 1 | Expression level and prognosis of lncRNA-AP000695.2 in human cancer. **(A)** Expression of lncRNA-AP000695.2 in normal and tumor tissues in TCGA and GTEx data, based on the best cutoff score to distinguish high- and low-expression groups. **(B–D)** Correlation between AP000695.2 expression and cancer survival prognosis. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

was increased in lung cancer tissues by analyzing the GEO dataset (**Figure 2C**). To examine the clinical relevance of AP000695.2 in LUAD, 535 LUAD patients with clinical characteristics were divided into two subgroups based on the mean value of relative AP000695.2 expression. We then explored the correlations between AP000695.2 expression and clinical features, including pathologic stage, TNM stage, and residual tumor. We found that AP000695.2 expression was significantly correlated with pathologic stage, TNM stage, and residual tumor (**Figures 2D–G**). Furthermore, ROC analysis

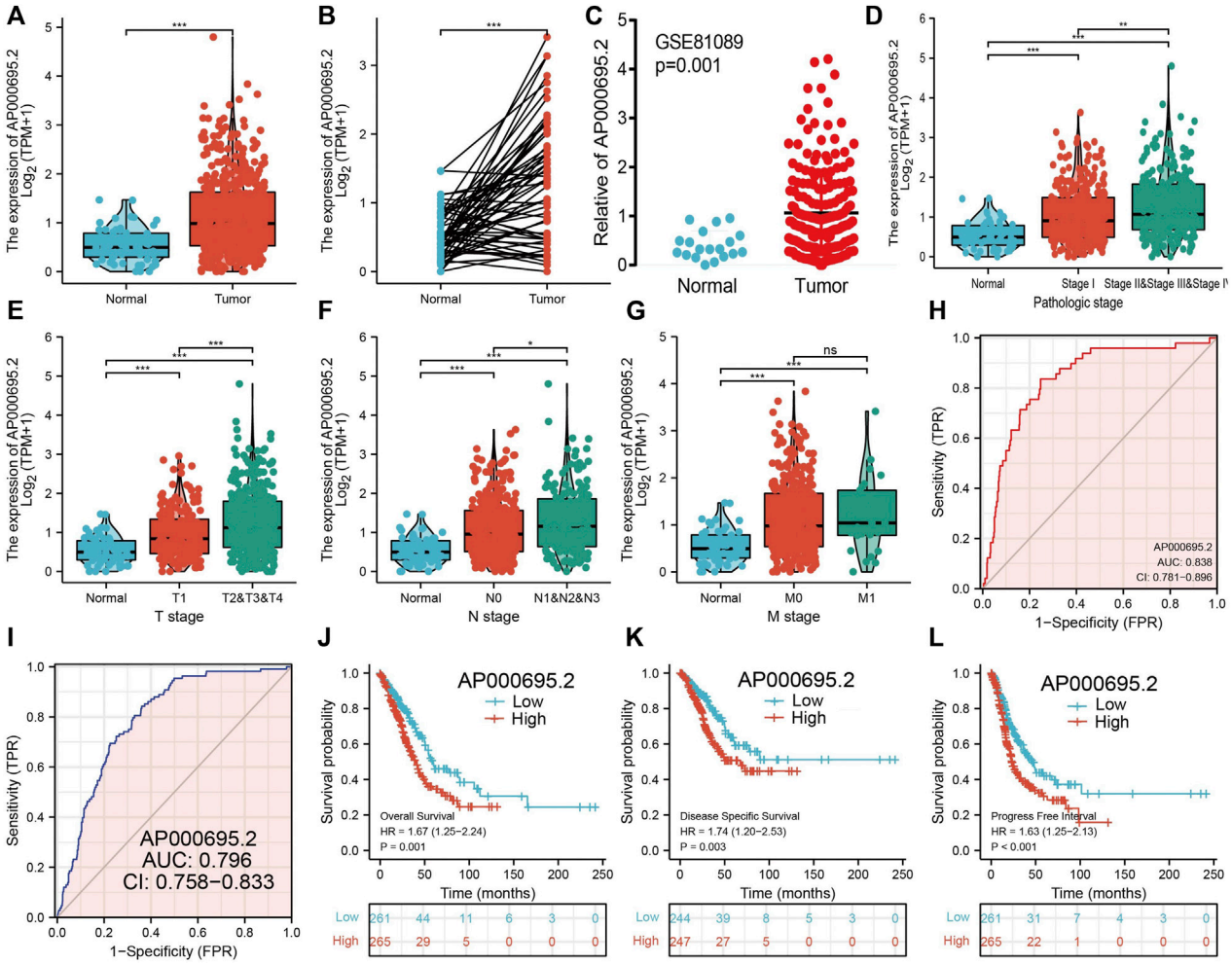


FIGURE 2 | Clinical significance of lncRNA-AP000695.2 in lung adenocarcinoma. **(A)** Expression of lncRNA-AP000695.2 in lung cancer based on TCGA dataset. **(B)** Expression levels of lncRNA-AP000695.2 in 59 paired adjacent normal tissues and paired samples. **(C)** Expression levels of lncRNA-AP000695.2 in lung cancer based on the GEO dataset. **(D–G)** Correlation between lncRNA-AP000695.2 expression and clinical parameters includes pathological and TNM stages. **(H–I)** ROC curves were used to determine the diagnostic value of lncRNA-AP000695.2 in lung adenocarcinoma based on TCGA-LUAD and GEO datasets. **(J–L)** Kaplan–Meier survival curves showed that lung adenocarcinoma patients with high-lncRNA-AP000695.2 expression exhibited poor overall survival, disease-specific survival, and progression-free survival based on TCGA-LUAD dataset. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

showed that AP000695.2 could be used to differentiate LUAD patients from normal controls with a specificity (AUC = 0.838) (**Figure 2H**). This result is also validated by the GEO dataset (**Figure 2I**). Moreover, Kaplan–Meier analysis showed that LUAD patients with higher AP000695.2 expression were associated with adverse overall survival (OS), disease-free survival, and progression-free survival (PFS) (**Figures 2J-L**).

Univariate and Multivariate Cox Regression Analyses

Univariate Cox regression analysis was conducted to determine whether the AP000695.2 expression level and pathologic stage might be valuable prognostic biomarkers in TCGA-LUAD cohort. In univariate Cox regression analysis, the high expression of AP000695.2, pathologic stage, and TNM stage was associated with overall survival in LUAD patients. To assess whether AP000695.2 could be an independent prognostic factor for LUAD patients, multivariate Cox survival analysis showed that AP000695.2 expression could be an independent prognostic factor for LUAD (**Tables 1**).

To examine whether the prognostic values of AP000695.2 was applicable to other clinical features, all patients were divided into various subgroups based on the clinical features. Survival analyses performed in the subgroups indicated that AP000695.2 performed well in subgroups such as stages I-II ($p = 0.003$), T1-T2 ($p = 0.004$), N0-N1 ($p = 0.009$), M0 ($p = 0.003$), R0 ($p = 0.01$), female ($p = 0.005$), male ($p = 0.032$), race, white ($p = 0.027$), >65years ($p = 0.015$), <65years ($p = 0.002$), CR ($p = 0.002$), and smoker ($p = 0.001$) in TCGA-LUAD cohort (**Figures 3A–C**).

TABLE 1 | Univariate and multivariate Cox regression analyses of different parameters on overall survival in lung adenocarcinoma.

Characteristic	Total (N)	Univariate analysis		Multivariate analysis	
		Hazard ratio (95% CI)	*p*-value	Hazard ratio (95% CI)	*p*-value
T stage	523				
T1 and T2	457				
T3 and T4	66	2.317 (1.591–3.375)	<0.001	1.638 (1.018–2.635)	0.042
N stage	510				
N0 and N1	437				
N3 and N2	73	2.321 (1.631–3.303)	<0.001	1.293 (0.626–2.674)	0.488
Pathologic stage	518				
Stage II and Stage I	411				
Stage IV and Stage III	107	2.664 (1.960–3.621)	<0.001	1.802 (0.839–3.871)	0.131
M stage	377				
M0	352				
M1	25	2.136 (1.248–3.653)	0.006	1.192 (0.541–2.626)	0.664
AC022784 1	526	1.251 (1.141–1.372)	<0.001	1.168 (1.053–1.296)	0.003

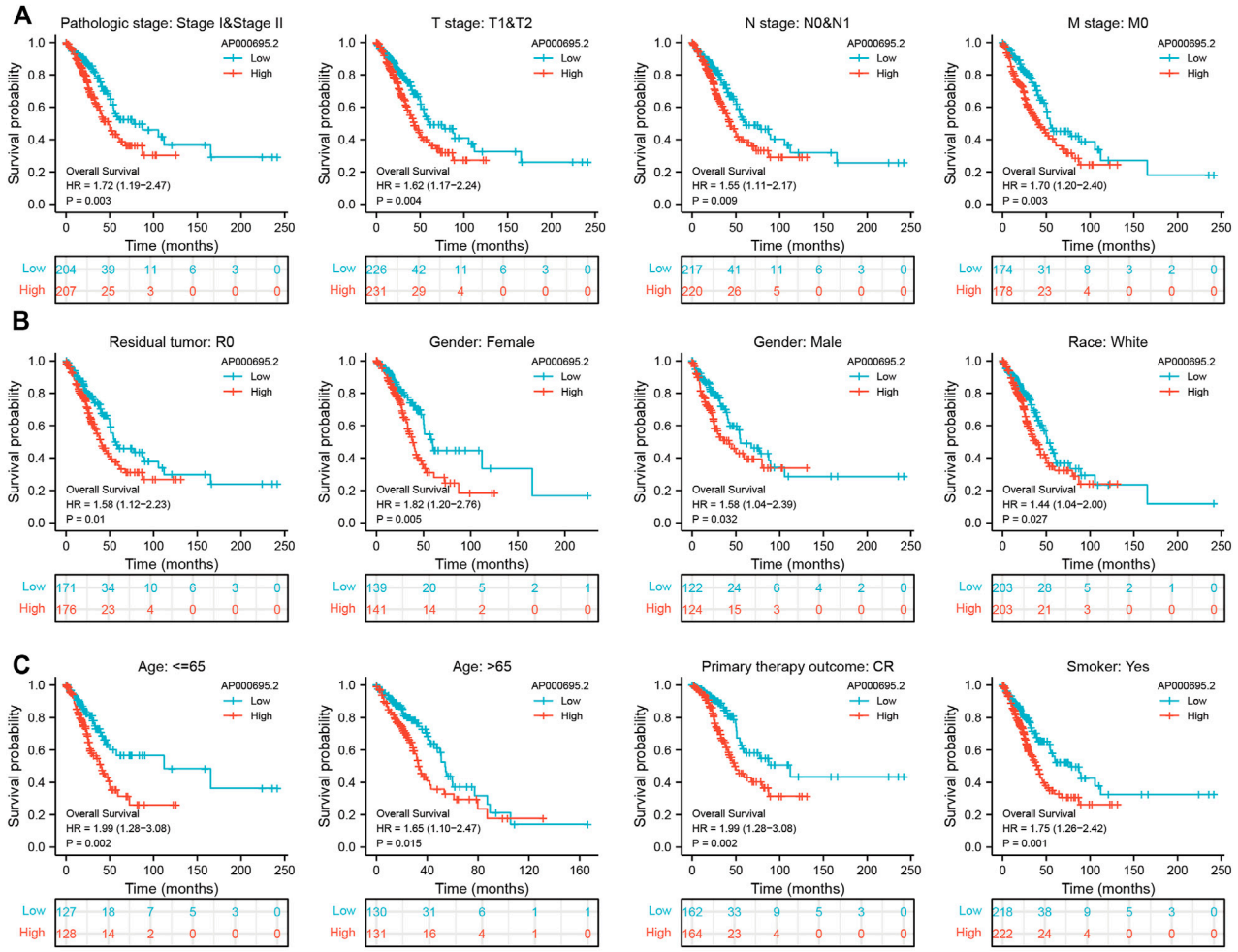


FIGURE 3 | Kaplan–Meier survival analyses for prognostic values of AP000695.2 in different subgroups stratified by clinical features. **(A–C)** Kaplan–Meier curves for overall survival of AP000695.2 in subgroups including stages I-II, T1-T2, N0-N1, M0, R0, female, male, race, white, age>65years, age <65years, CR, and smoker in TCGA-LUAD cohort.

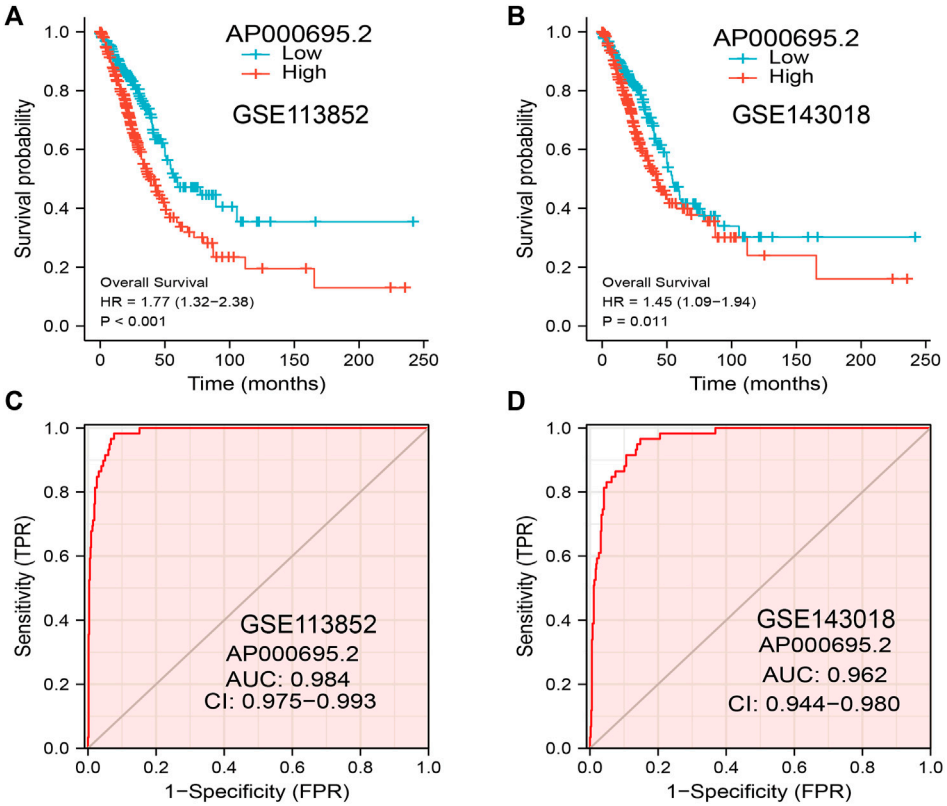


FIGURE 4 | Predictive powers for the prognosis of AP000695.2 in LUAD patients. **(A–B)** Kaplan–Meier survival validation of the prognostic values of AP000695.2 in LUAD patients based on the GEO cohorts. **(C–D)** ROC curves show the AUC of AP000695.2 for predicting the overall survival of LUAD patients.

Validation of the Prognostic Significance of AP000695.2 in Lung Adenocarcinoma Cohorts

The GEO dataset was used to validate the prognostic significance of AP000695.2 in LUAD, and we showed that the LUAD patients in the high-expression group had shorter OS in the GEO cohorts (**Figures 4A,B**). Moreover, the ROC curve was also used to assess the predictive power of AP000695.2 in OS, and the AUC for the overall survival rate of LUAD patients was 0.984 and 0.962 in the GEO cohort (**Figures 4C,D**). Taken together, these results suggest that AP000695.2 is a moderately sensitive index for predicting the prognosis of LUAD patients and can act as an effective prognostic biomarker in LUAD.

Construction and Validation of an AP000695.2-Based Nomogram

The multivariate analysis result confirmed that AP000695.2 is an independent prognostic factor in LUAD. We then constructed a prediction model for overall survival, disease-free survival, and progression-free survival by integrating AP000695.2 expression and pathologic stage. We established a nomogram to integrate AP000695.2 as an LUAD biomarker, and higher total points on the nomogram for OS, PFS, and DFS indicated a worse prognosis (**Figures 5A–F**). These findings indicated that the nomogram could well predict clinical outcomes of lung adenocarcinoma patients.

AP000695.2-Related Signaling Pathways Based on Gene Set Enrichment Analysis

To explore the influence of potential signaling pathways of AP000695.2 in LUAD, gene set enrichment analysis (GSEA) was performed on the datasets with high and low expressions of AP000695.2. We only selected the top nine datasets with a high normalized enrichment score (NES) and significant *p*-value. The results showed that cell apoptosis, focal adhesion, cell cycle, cell adhesion molecules cams, JAK-STAT signaling pathway, MAPK signaling pathway, natural killer cell-mediated cytotoxic, tight junction, T-cell receptor signaling pathway, toll-like receptor signaling pathway, chemokine signaling pathway, and cytokine–cytokine receptor interaction were significantly enriched in the high-AP000695.2 expression group in TCGA datasets (**Figures 6A–C**). These results suggested that AP000695.2 might participate in the regulation of cell apoptosis and immune response in LUAD.

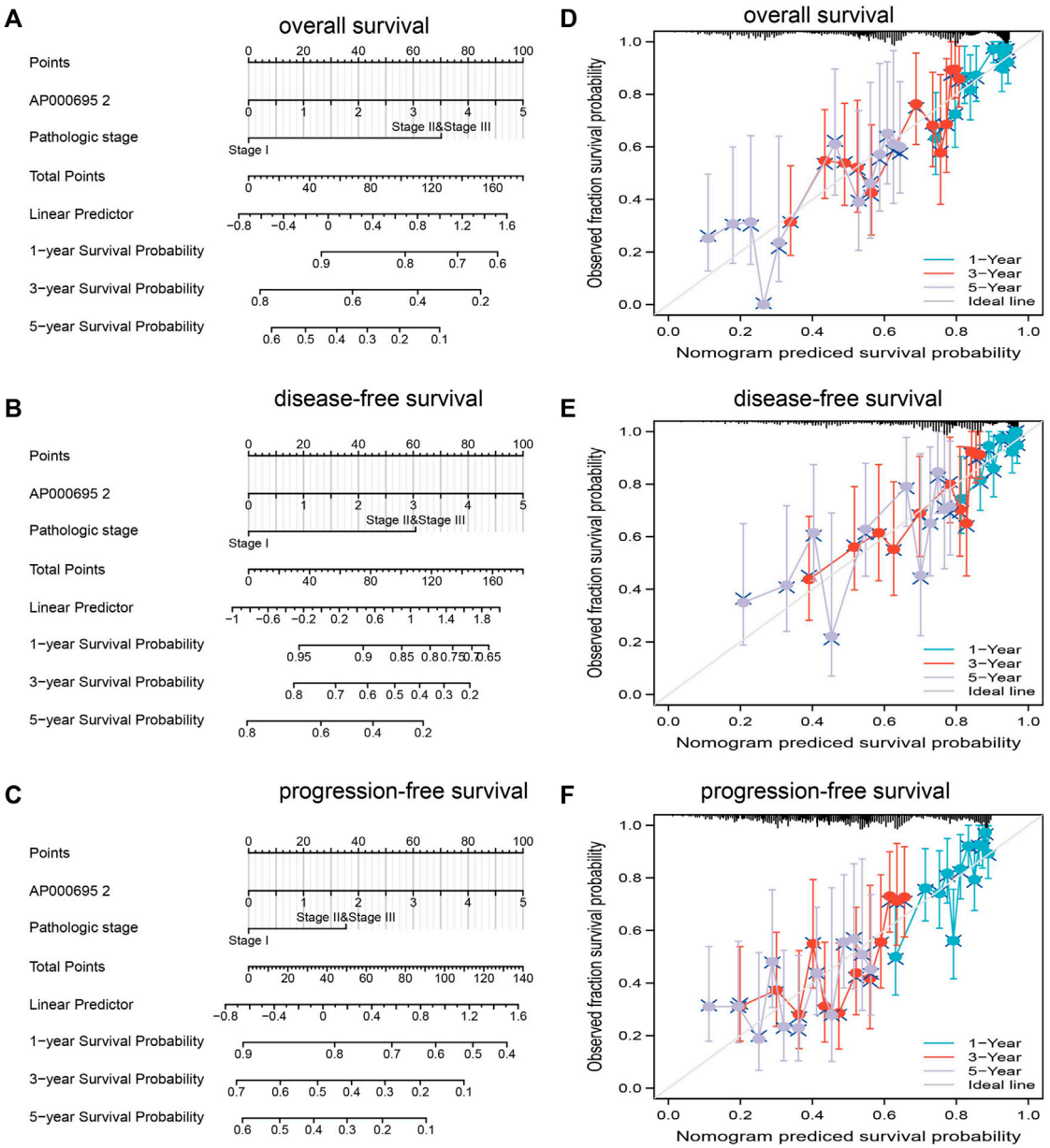


FIGURE 5 | Construction and performance validation of the lncRNA-AP000695.2-based nomogram for lung adenocarcinoma patients. **(A–D)** Nomogram to predict the overall survival, disease-free survival, and progression-free survival of lung cancer patients, and the calibration curve and Hosmer–Lemeshow test of nomograms in TCGA-lung adenocarcinoma cohort for overall survival, disease-specific survival, and progression-free survival.

Correlation Between AP000695.2 Expression and Immune Infiltration

ssGSEA with Spearman's rank correlation was utilized to determine the relationship between AP000695.2 expression and infiltrating immune cells in LUAD. The results confirmed that AP000695.2 was positively correlated with infiltration levels of Th2 cells, neutrophils, macrophages, NK CD56dim cells, Ted, Th1 cells, NK cells, TReg, pDC, and aDC ($p < 0.001$) and was negatively correlated with that of eosinophils, CD8 T cells, NK CD56bright cells, Th17 cells, B cells, mast cells, Tcm, and TFH (**Figure 7A**). Furthermore, we found that patients in the AP000695.2 high-expression group showed an increase in the numbers of infiltrating aDC, macrophages, neutrophils, NK CD56dim cells, NK cells, pDC, Tgd, Th1 cells, and Th2 cells. On the contrary, patients in the AP000695.2 low-expression group showed a reduction in the numbers of infiltrating eosinophils, CD8 T cells, NK CD56bright cells, Th17 cells, B cells, mast cells, Tcm, and TFH (**Figure 7B**).

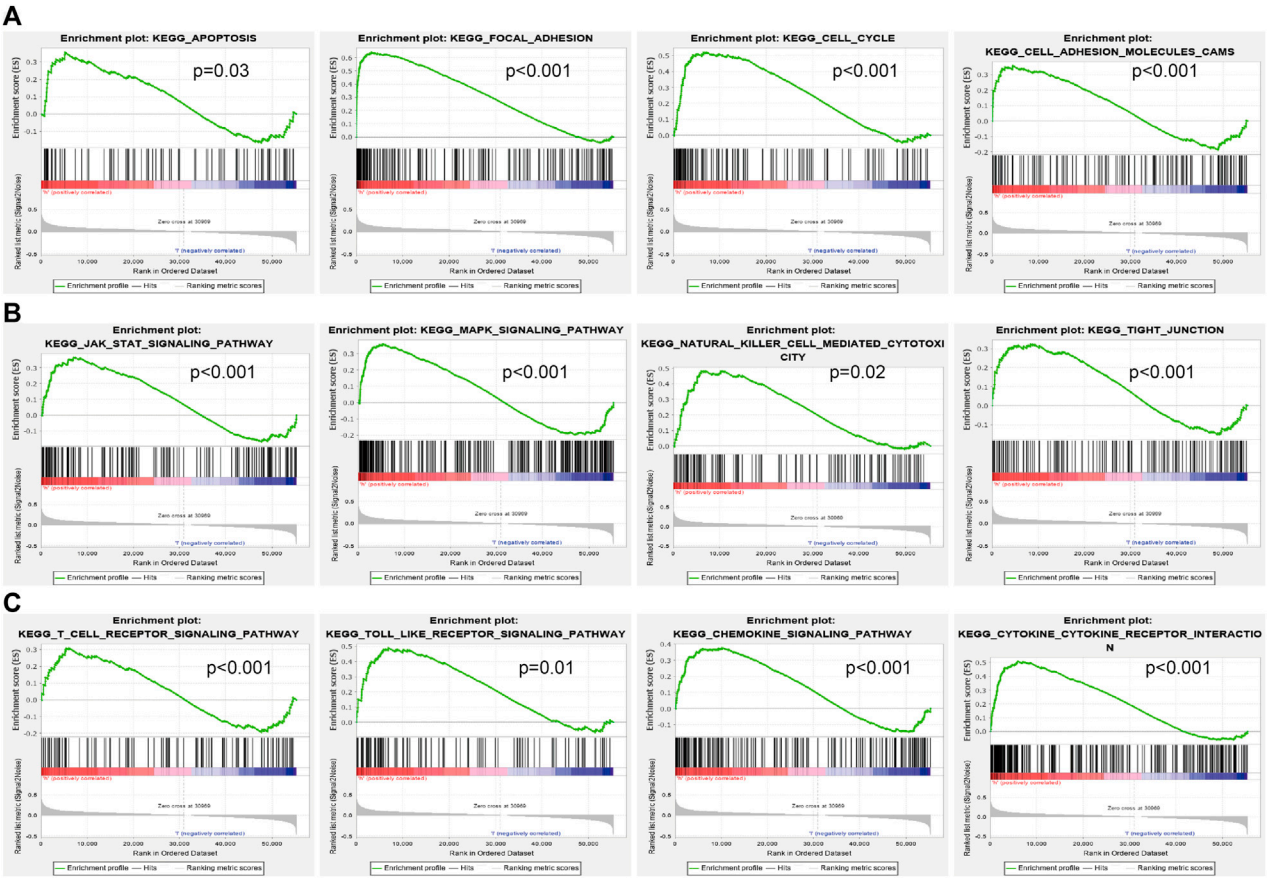


FIGURE 6 | Identification of lncRNA-AP000695.2-related signaling pathways in lung adenocarcinoma. **(A–B)** Top 12 significant KEGG pathways associated with AP000695.2 were examined by GSEA software.

AP000695.2 Promotes Proliferation and Migration of Lung Adenocarcinoma Cells *In Vitro*

The aforementioned studies suggested that AP000695.2 expression was elevated in LUAD tissues, and AP000695.2 may involve in the progression of LUAD. To further determine the function of AP000695.2 in LUAD, we found that AP000695.2 was elevated in H1650, H1975, and H1299 lung cancer cell lines compared to the human bronchial epithelial cells (BEAS-2B) (**Figure 8A**). Furthermore, specific shRNA for AP000695.2 was utilized to construct H1299 and H1975 cells with stable knockdown of AP000695.2 expression. The knockdown efficiencies were examined by qRT-PCR assay (**Figures 8B,C**). The knockdown of AP000695.2 inhibited the cell proliferation capacity of H1299 and H1975 cells (**Figures 8D,E**) on BrdU and colony assays. Moreover, transwell assay and wound healing showed that depletion of the AP000695.2 expression level significantly reduced the cell migration and invasion abilities of H1299 and H1975 cells (**Figures 8F,G**). These results confirmed that AP000695.2 plays a role of an oncogene in lung cancer cells.

AP000695.2 Facilitates Cisplatin Resistance in Lung Adenocarcinoma *In Vitro*

We utilized the GDSC database to explore the potential drug that correlated with AP000695.2. The results confirmed that AP000695.2 was positively correlated with the diverse drug, including cisplatin (DDP) ($r = 0.43$, $p < 0.001$) which was reported to be a common chemotherapeutic drug for LUAD patients (**Figure 9A**). To further determine the function of DDP resistance in LUAD, we utilized DDP-sensitive, parental LUAD cell lines H1299 and their isogenic DDP-resistant counterparts H1299 as experimental cell lines. We found that AP000695.2 was elevated in H1299-DDP-resistant cells than in H1299 DDP-sensitive cells (**Figure 9B**). Then, we treated LUAD cells with DDP and calculated the IC_{50} value. The results confirmed that AP000695.2 knockdown significantly decreased the IC_{50} value of DDP (**Figure 9C**). Likewise, AP000695.2 knockdown significantly promotes the DDP-induced cell apoptosis in H1299-DDP-resistant cells (**Figures 9D,E**). Overall, an increased AP000695.2 expression could promote DDP resistance of LUAD cells.

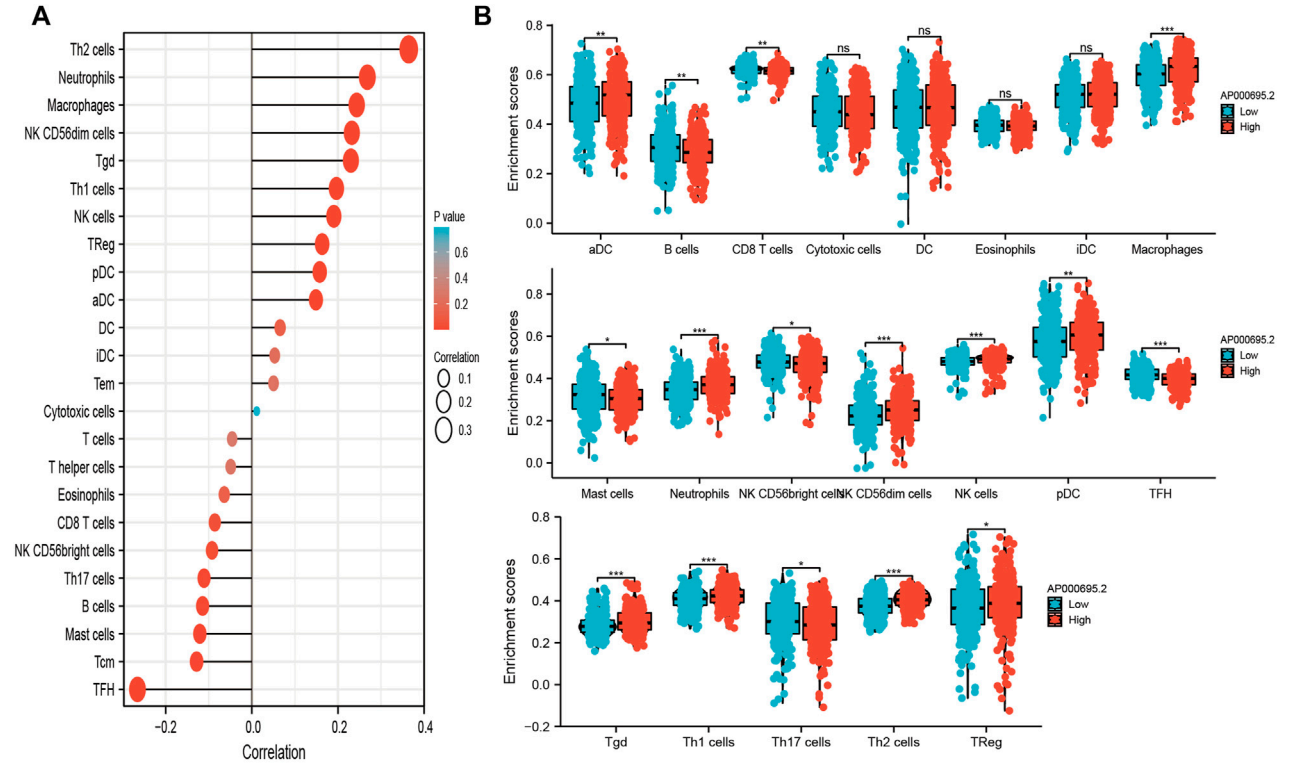


FIGURE 7 | Correlation analysis of lncRNA-AP000695.2 expression and infiltration levels of immune cells in LUAD tissues. **(A)** Correlation between the relative abundances of 24 immune cells and lncRNA AP000695.2 expression level. **(B)** Box plots of the correlations between lncRNA-AP000695.2 or molecular model expression and infiltration levels of immune cells. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

DISCUSSION

Despite there being many different treatments, the prognosis of lung cancer is an advanced stage and the survival rate of patients is very low (Hirsch et al., 2017). The initiation and progression of lung cancer is an exceedingly complicated process that involves genetic mutations, tumor microenvironment, and the dysregulation of epigenetic pathways (Govindan et al., 2012; Duruisseaux and Esteller 2018). It has been shown that lncRNAs maybe serve as effective and specific molecular markers for lung cancer diagnosis. For example, NEAT1 was highly expressed in NSCLC tissues, and its higher expression was associated with the TNM stage and lymphatic metastasis, and the ROC curve of NEAT1 in NSCLC was 0.878, which showed good diagnostic potential in NSCLC (Jen et al., 2017). Li et al. reported that lncRNA UPLA1 was overexpressed in NSCLC tissues and correlated with tumor size and lymph node metastasis, and lncRNA UPLA1 had an ROC curve of 0.756 for discriminating NSCLC from normal controls, with a sensitivity and specificity of 76.2% and 72.1%, respectively (Han et al., 2020). In this study, we found that lncRNA-AP000695.2 was highly expressed in LUAD and correlated with adverse clinical features, including pathologic stage, TNM stage, and residual tumor. The survival analysis results showed that patients with high expression of AP000695.2 correlated with OS, DFS, and PFS in the LUAD patients of TCGA data.

Furthermore, univariate and multivariate analyses as standard and reliable statistical methods were utilized to confirm whether lncRNA can be regarded as an independent tumor marker for predicting the prognosis of lung cancer patients. By univariate and multivariate analyses, some lncRNAs have also been identified as independent prognostic markers in lung cancer. For instance, lncRNA CASC9 was shown to be increased in NSCLC and correlated with poor prognosis. Forced lncRNA CASC9 expression promotes NSCLC cell proliferation and chemoresistance via epigenetic repression of DUSP1, and univariate and multivariate analyses reported that lncRNA CASC9 can be regarded as independent prognostic markers in lung cancer (Chen et al., 2020b). In this study, we confirmed that AP000695.2 expression could be an independent prognostic factor for LUAD. We also established a nomogram to integrate AP000695.2 as an LUAD biomarker, and higher total points on the nomogram for overall survival, progression-free survival (PFS), and disease-specific survival (DSS) indicated a worse prognosis.

LncRNAs are a class of RNAs with more than 200 nucleotides in length. Although lncRNAs do not have protein-coding capabilities, they play essential roles in various biological processes and diseases (Wang et al., 2018). The aberrant regulation of lncRNAs is associated with tumorigenesis, metastasis, and drug resistance (Liang et al., 2021). It has been confirmed that lncRNA TUC338 via activating the MAPK pathway leads to lung cancer progression (Zhang et al., 2018). Additionally, You et al. found that lncRNA-RP11-468E2.5

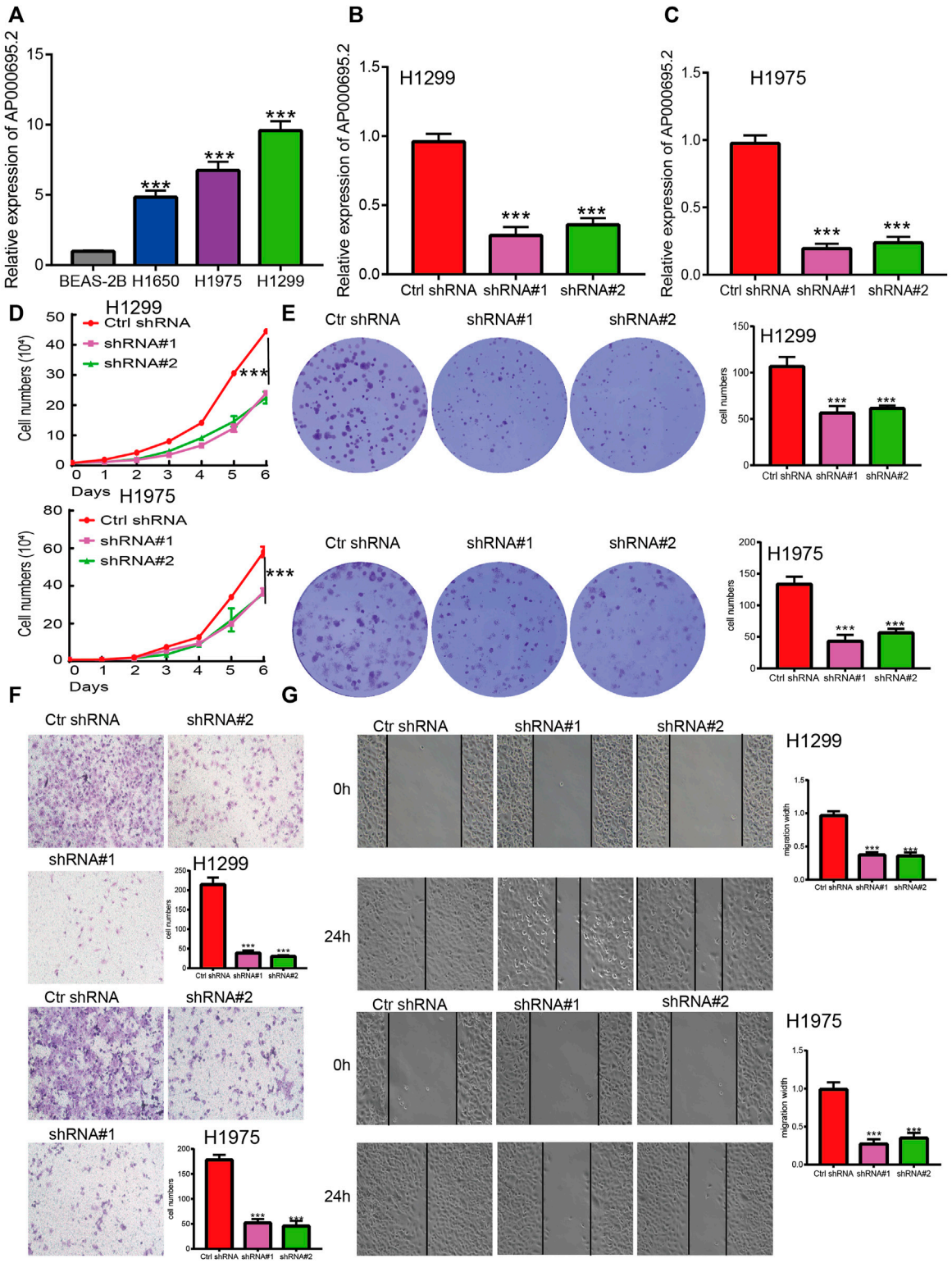


FIGURE 8 | lncRNA-AP000695.2 promotes LUAD cell proliferation, migration, and invasion *in vitro*. **(A)** Relative expression level of lncRNA AP000695.2 in lung adenocarcinoma cancerous cell lines, including H1650, H1975, and H1299 examined by real-time RT-PCR, compared with normal human bronchial epithelial cell lines: BEAS-2B. **(B–C)** Establishment of lncRNA-AP000695.2 knockdown cell lines in H1299 and H1975 cells verified by real-time RT-PCR. **(D–E)** Knockdown of lncRNA-AP000695.2 significantly inhibits cell proliferation in H1299 and H1975 cells as measured by BrdU and colony formation. **(F–G)** Knockdown of lncRNA AP000695.2 dramatically inhibits H1299 and H1975 cells' migration ability examined by transwell and wound healing assays. $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

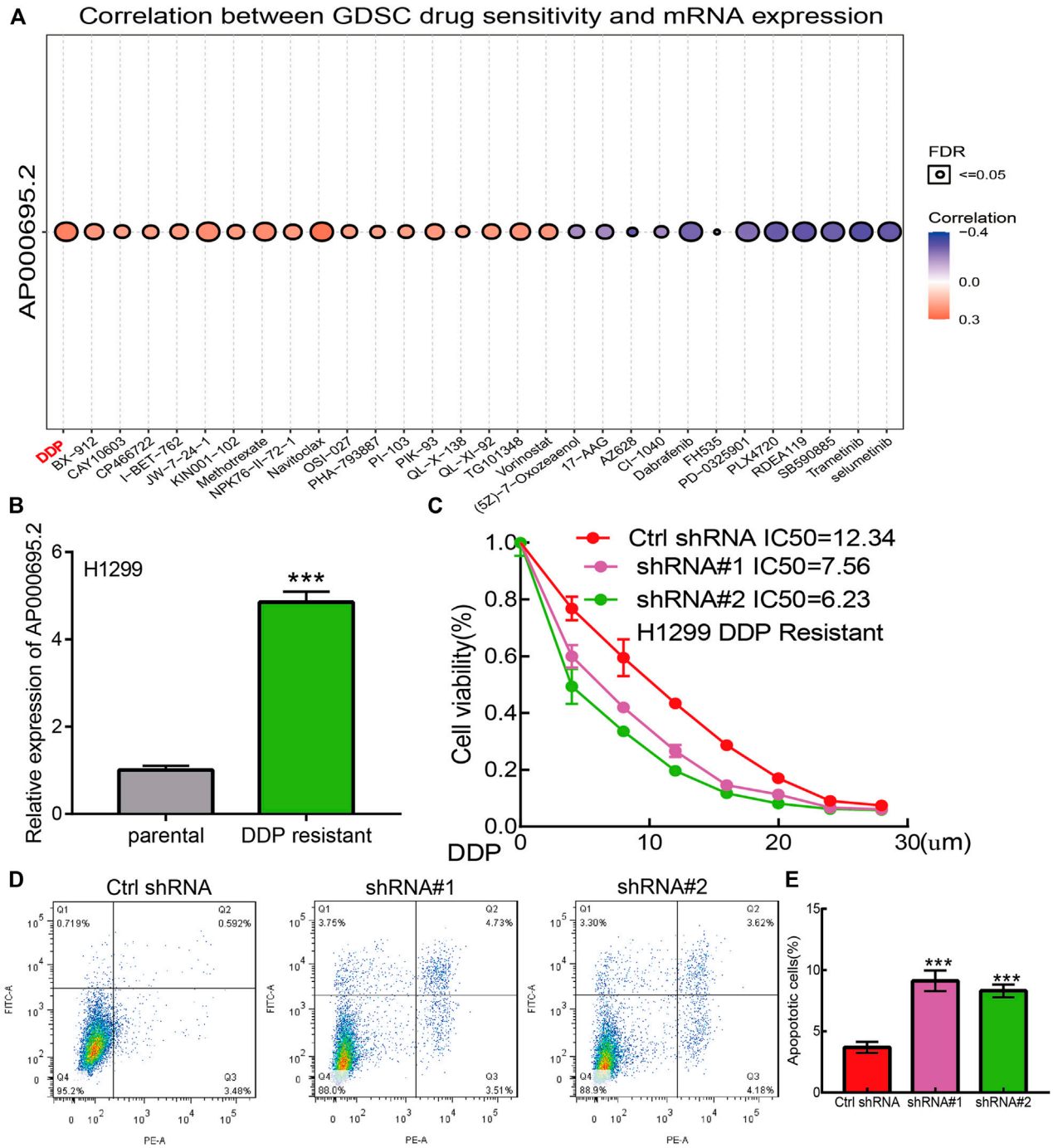


FIGURE 9 | lncRNA-AP000695.2 contributes to DDP resistance of LUAD. **(A)** Relationships between AP000695.2 expression and different drugs were examined by the GDSC database. **(B)** Expression of AP000695.2 in DDP-resistant and matched parental LUAD cells, normalized to 18sRNA expression. **(C)** AP000695.2 knockdown increases the sensitivity of H1299/CDDP-resistant cells to CDDP, detected by CCK-8 assay. **(D–E)** AP000695.2 knockdown increases the CDDP-induced apoptosis rate in H1299/DDP-resistant cells detected by FCM.

regulates colorectal cancer cell proliferation and promotes apoptosis by modulating the JAK/STAT signaling pathway by targeting STAT5 and STAT6 (Jiang et al., 2019). For example, Li et al. found that ZFPM2-AS1 was found to facilitate cell proliferation, migration, and invasion via involvement in the JAK-STAT and AKT pathways in NSCLC (Wang X et al., 2020). Moreover, Guo et al. found that lncRNA- XLOC_098,131, via sponging miR-548s and thereby upregulation of FOS expression, leads to the production of more immunoglobulins and the promotion of antigen presentation (Fan et al., 2019). In this study, we found that the high expression of

AP000695.2 was highly associated with cell apoptosis, focal adhesion, cell cycle, cell adhesion molecules cams, JAK-STAT signaling pathway, MAPK signaling pathway, natural killer cell-mediated cytotoxic, tight junction, T-cell receptor signaling pathway, toll-like receptor signaling pathway, chemokine signaling pathway, and cytokine–cytokine receptor interaction that were significantly enriched, which indicated that AP000695.2 might have a crucial role in immune response regulation and cell proliferation. The JAK-STAT signaling pathway was reported to play crucial roles in the regulation of intracellular signal transduction and cellular immune response. These results indicated that AP000695.2 may be correlated with the JAK-STAT signaling pathway, MAPK signaling pathway, and toll-like receptor signaling pathway in cancer development and progression.

The tumor microenvironment plays a crucial role in cancer progression and can be used as a biomarker for diagnosis and prognosis (Lee and Cheah 2019). For instance, tumor-associated macrophages are regarded as essential components of the tumor microenvironment and play critical roles in the modulation of cancer progression (Pan et al., 2020). In this finding, we found that AP000695.2 expression was positively associated with infiltration levels of Th2 cells, neutrophils, macrophages, NK CD56dim cells, Tgd, Th1 cells, NK cells, TReg, pDC, and aDC ($p < 0.001$) and was negatively correlated with that of eosinophils, CD8 T cells, NK CD56bright cells, Th17 cells, B cells, mast cells, Tcm, and TFH. Furthermore, we found that patients in the AP000695.2 high-expression group showed an increase in the numbers of infiltrating aDC, macrophages, neutrophils, NK CD56dim cells, NK cells, pDC, Tgd, Th1 cells, and Th2 cells. This correlation means that AP000695.2 expression was significantly positively correlated with the infiltrating levels of aDC, macrophages, neutrophils, NK CD56dim cells, NK cells, pDC, Tgd, Th1 cells, and Th2 cells in LUAD. On the contrary, patients in the AP000695.2 low-expression group showed a reduction in the numbers of infiltrating eosinophils, CD8 T cells, NK CD56bright cells, Th17 cells, B cells, mast cells, Tcm, and TFH. This correlation means that AP000695.2 expression was significantly negatively correlated with the infiltrating levels of eosinophils, CD8 T cells, NK CD56bright cells, Th17 cells, B cells, mast cells, Tcm, and TFH in LUAD. Based on the aforementioned findings, we proposed that AP000695.2 may be involved in the immune response by affecting diverse immune cells infiltrating in the tumor microenvironment of lung cancer. In *in vitro* assay, we show that knockdown of AP000695.2 in H1299 and H1975 inhibited cell proliferation and migration.

Cisplatin is a common drug in lung cancer chemotherapy. However, owing to primary or acquired drug resistance, DDP has not achieved a satisfactory therapeutic effect in most LUAD patients. Therefore, identifying and validating the key regulators involved in drug resistance, especially epigenetic modifications, can provide crucial information for overcoming DDP resistance in LUAD. In this finding, by the analysis of TCGA and GDSC datasets, we found the key lncRNAs involved in DDP resistance in LUAD. Among these lncRNAs, AP000695.2 was elevated in DDP-resistant LUAD cells than in DDP-sensitive cells, confirming that AP000695.2 was related to DDP resistance. Then, we found that the knockdown of AP000695.2 significantly improved the sensitivity of lung cancer cells to DDP and induced cell apoptosis. Through loss-of-function experiments performed *in vitro*, we uncover that AP000695.2 promoted DDP resistance in LUAD. To our knowledge, this finding is the first to confirm the crucial role of AP000695.2 in DDP resistance in LUAD.

This study improves our understanding of the correlation between AP000695.2 and LUAD, but some limitations still exist. First, although we explored the correlation between AP000695.2 and immune infiltration in LUAD patients, there is a lack of experiments to validate the function of AP000695.2 in the tumor microenvironment regulation of LUAD. Second, we uncover that depletion of AP000695.2 inhibits cell proliferation and cell migration of LUAD cells. However, the potential molecular mechanisms of AP000695.2 in cancer progression need to be explored in further studies. Third, we did not conduct the Western blot experiments to detect the expression of the related signaling pathway molecules (JAK/MAPK, etc.) after knocking down AP000695.2 in lung cancer cells. In the future, we will pay more attention to the function of AP000695.2 in tumor metastasis and tumor microenvironment regulation of LUAD. Furthermore, we will perform more *in vivo* and *in vitro* experiments to explore the function and the potential molecular mechanisms of AP000695.2 in tumor metastasis and tumor microenvironment regulation of LUAD.

CONCLUSION

In conclusion, this study uncovers the biological function of AP000695.2 in LUAD for the first time. AP000695.2 was highly expressed in LUAD and associated with adverse clinical outcomes in LUAD patients, and AP000695.2 expression in LUAD is associated with pathologic stage, TNM stage, and residual tumor. Furthermore, AP000695.2 plays a significant role in regulating cell proliferation, cell migration, and DDP resistance. Our findings showed that AP000695.2 has the clinical potential to reverse DDP resistance and achieve better clinical outcomes in LUAD patients and may serve as a promising diagnostic and prognostic biomarker for LUAD.

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.

AUTHOR CONTRIBUTIONS

CW and JG designed this work and performed related assay. RJ, CW, and CP analyzed the data. ZN and XJ supervised and wrote the manuscript. All authors have read and approved the final version of the manuscript.

FUNDING

This study was supported in part by grants from the Yunnan Province Clinical Research Center for Neurological Diseases (202002AA100204), National Natural Science Foundation of

China (82160461 to CW), Department of Science and Technology of Yunnan Province-Kunming Medical University (2019FE001 (-063) to CW), and Yunnan Health Training Project of High Level Talents (D2018058 to CW).

REFERENCES

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

Cai, Y., Dong, Z. Y., and Wang, J. Y. (2018). LncRNA NNT-AS1 is a Major Mediator of Cisplatin Chemoresistance in Non-small Cell Lung Cancer through MAPK/Slug Pathway. *Eur. Rev. Med. Pharmacol. Sci.* 22 (15), 4879–4887. doi:10.26355/eurrev_201808_15624

Chen, Z., Lei, T., Chen, X., Gu, J., Huang, J., Lu, B., et al. (2020). Long Non-coding RNA in Lung Cancer. *Clinica Chim. Acta* 504, 190–200. doi:10.1016/j.cca.2019.11.031

Chen, Z., Chen, Q., Cheng, Z., Gu, J., Feng, W., Lei, T., et al. (2020). Long Non-coding RNA CASC9 Promotes Gefitinib Resistance in NSCLC by Epigenetic Repression of DUSP1. *Cell Death Dis* 11 (10), 858. doi:10.1038/s41419-020-03047-y

Dasari, S., and Bernard Tchounwou, P. (2014). Cisplatin in Cancer Therapy: Molecular Mechanisms of Action. *Eur. J. Pharmacol.* 740, 364–378. doi:10.1016/j.ejphar.2014.07.025

Duruisseaux, M., and Esteller, M. (2018). Lung Cancer Epigenetics: From Knowledge to Applications. *Semin. Cancer Biol.* 51, 116–128. doi:10.1016/j.semcancer.2017.09.005

Fan, H., Lv, Z., Gan, L., Ning, C., Li, Z., Yang, M., et al. (2019). A Novel lncRNA Regulates the Toll-like Receptor Signaling Pathway and Related Immune Function by Stabilizing FOS mRNA as a Competitive Endogenous RNA. *Front. Immunol.* 10, 838. doi:10.3389/fimmu.2019.00838

Fennell, D. A., Summers, Y., Cadranel, J., Benepal, T., Christoph, D. C., Lal, R., et al. (2016). Cisplatin in the Modern Era: The Backbone of First-Line Chemotherapy for Non-small Cell Lung Cancer. *Cancer Treat. Rev.* 44, 42–50. doi:10.1016/j.ctrv.2016.01.003

Govindan, R., Ding, L., Griffith, M., Subramanian, J., Dees, N. D., Kanchi, K. L., et al. (2012). Genomic Landscape of Non-small Cell Lung Cancer in Smokers and Never-Smokers. *Cell* 150 (6), 1121–1134. doi:10.1016/j.cell.2012.08.024

Han, X., Jiang, H., Qi, J., Li, J., Yang, J., Tian, Y., et al. (2020). Novel lncRNA UPLA1 Mediates Tumorigenesis and Prognosis in Lung Adenocarcinoma. *Cel Death Dis.* 11 (11), 999. doi:10.1038/s41419-020-03198-y

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

Hirsch, F. R., Scagliotti, G. V., Mulshine, J. L., Kwon, R., Curran, W. J., Wu, Y.-L., et al. (2017). Lung Cancer: Current Therapies and New Targeted Treatments. *Lancet* 389 (10066), 299–311. doi:10.1016/s0140-6736(16)30958-8

Huang, F. X., Chen, H. J., Zheng, F. X., Gao, Z. Y., Sun, P. F., Peng, Q., et al. (2019). LncRNA BLACAT1 Is Involved in Chemoresistance of Non-small Cell Lung Cancer Cells by Regulating Autophagy. *Int. J. Oncol.* 54 (1), 339–347. doi:10.3892/ijo.2018.4614

Jen, J., Tang, Y.-A., Lu, Y.-H., Lin, C.-C., Lai, W.-W., and Wang, Y.-C. (2017). Oct4 Transcriptionally Regulates the Expression of Long Non-coding RNAs NEAT1 and MALAT1 to Promote Lung Cancer Progression. *Mol. Cancer* 16 (1), 104. doi:10.1186/s12943-017-0674-z

Jiang, L., Zhao, X.-H., Mao, Y.-L., Wang, J.-F., Zheng, H.-J., and You, Q.-S. (2019). Long Non-coding RNA RP11-468E2.5 Curtails Colorectal Cancer Cell Proliferation and Stimulates Apoptosis via the JAK/STAT Signaling Pathway by Targeting STAT5 and STAT6. *J. Exp. Clin. Cancer Res.* 38 (1), 465. doi:10.1186/s13046-019-1428-0

Jiang, X., Yuan, Y., Tang, L., Wang, J., Liu, Q., Zou, X., et al. (2021). Comprehensive Pan-Cancer Analysis of the Prognostic and Immunological Roles of the METTL3/lncRNA-SNHG1/miRNA-140-3p/UBE2C Axis. *Front. Cel Dev. Biol.* 9, 765772. doi:10.3389/fcell.2021.765772

Jiang, X., Yuan, Y., Tang, L., Wang, J., Zhang, D., Cho, W. C., et al. (2022). Identification and Validation Prognostic Impact of MiRNA-30a-5p in Lung Adenocarcinoma. *Front. Oncol.* 12, 831997. doi:10.3389/fonc.2022.831997

Lee, S. S., and Cheah, Y. K. (2019). The Interplay between MicroRNAs and Cellular Components of Tumour Microenvironment (TME) on Non-small-cell Lung Cancer (NSCLC) Progression. *J. Immunol. Res.* 2019, 3046379. doi:10.1155/2019/3046379

Liang, R., Li, X., Li, W., Zhu, X., and Li, C. (2021). DNA Methylation in Lung Cancer Patients: Opening a "window of Life" under Precision Medicine. *Biomed. Pharmacother.* 144, 112202. doi:10.1016/j.biopha.2021.112202

Luco, R. F., Pan, Q., Tominaga, K., Blencowe, B. J., Pereira-Smith, O. M., and Misteli, T. (2010). Regulation of Alternative Splicing by Histone Modifications. *Science* 327 (5968), 996–1000. doi:10.1126/science.1184208

Molina, J. R., Yang, P., Cassivi, S. D., Schild, S. E., and Adjei, A. A. (2008). Non-small Cell Lung Cancer: Epidemiology, Risk Factors, Treatment, and Survivorship. *Mayo Clinic Proc.* 83 (5), 584–594. doi:10.4065/83.5.584

Pan, Y., Yu, Y., Wang, X., and Zhang, T. (2020). Tumor-Associated Macrophages in Tumor Immunity. *Front. Immunol.* 11, 583084. doi:10.3389/fimmu.2020.583084

Reck, M., and Rabe, K. F. (2017). Precision Diagnosis and Treatment for Advanced Non-small-cell Lung Cancer. *N. Engl. J. Med.* 377 (9), 849–861. doi:10.1056/nejmra1703413

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

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. U.S.A.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102

Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., et al. (2011). A Long Noncoding RNA Maintains Active Chromatin to Coordinate Homeotic Gene Expression. *Nature* 472 (7341), 120–124. doi:10.1038/nature09819

Wang, J., Su, Z., Lu, S., Fu, W., Liu, Z., Jiang, X., et al. (2018). LncRNA HOXA-AS2 and its Molecular Mechanisms in Human Cancer. *Clin. Chim. Acta* 485, 229–233. doi:10.1016/j.cca.2018.07.004

Wang J, J., Gao, J., Chen, Q., Zou, W., Yang, F., Wei, C., et al. (2020). LncRNA LINC01116 Contributes to Cisplatin Resistance in Lung Adenocarcinoma. *Onco Targets Ther.* 13, 9333–9347. doi:10.2147/ott.s244879

Wang X, X., Tang, J., Zhao, J., Lou, B., and Li, L. (2020). ZFPM2-AS1 Promotes the Proliferation, Migration, and Invasion of Human Non-small Cell Lung Cancer Cells Involving the JAK-STAT and AKT Pathways. *PeerJ* 8, e10225. doi:10.7717/peerj.10225

Xu, L., Xu, Y., Yang, M., Li, J., Xu, F., and Chen, B.-L. (2020). LncRNA SNHG14 Regulates the DDP-Resistance of Non-small Cell Lung Cancer Cell through miR-133a/HOXB13 Pathway. *BMC Pulm. Med.* 20 (1), 266. doi:10.1186/s12890-020-01276-7

Yang, Y., Li, H., Hou, S., Hu, B., Liu, J., and Wang, J. (2013). The Noncoding RNA Expression Profile and the Effect of lncRNA AK126698 on Cisplatin Resistance in Non-small-cell Lung Cancer Cell. *PLoS One* 8 (5), e65309. doi:10.1371/journal.pone.0065309

Yang, F., Xue, X., Zheng, L., Bi, J., Zhou, Y., Zhi, K., et al. (2014). Long Non-coding RNA GHET1 Promotes Gastric Carcinoma Cell Proliferation by Increasing C-Myc mRNA Stability. *Febs J* 281 (3), 802–813. doi:10.1111/febs.12625

Zhang, Y. X., Yuan, J., Gao, Z. M., and Zhang, Z. G. (2018). LncRNA TUC338 Promotes Invasion of Lung Cancer by Activating MAPK Pathway. *Eur. Rev. Med. Pharmacol. Sci.* 22 (2), 443–449. doi:10.26355/eurrev_201801_14193

Zhang, X., Xie, K., Zhou, H., Wu, Y., Li, C., Liu, Y., et al. (2020). Role of Non-coding RNAs and RNA Modifiers in Cancer Therapy Resistance. *Mol. Cancer* 19 (1), 47. doi:10.1186/s12943-020-01171-z

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 Molecular Biosciences

ORIGINAL RESEARCH
published: 23 June 2022
doi: 10.3389/fmolb.2022.862570

Pulmonary Microbial Composition in Sepsis-Induced Acute Respiratory Distress Syndrome

Peng Zhang[1†], Baoyi Liu[2†], Weihao Zheng[1†], Yantang Chen[1], Zhentao Wu[1], Yuting Lu[1], Jie Ma[1], Wenjie Lu[2], Mingzhu Zheng[2], Wanting Wu[2], Zijie Meng[2], Jinhua Wu[3], Yan Zheng[4], Xin Zhang[2,5,6], Shuang Zhang[1]* and Yanming Huang[2,7]**

[1]Department of Critical Care Medicine, Jiangmen Central Hospital, Jiangmen, China, [2]Clinical Experimental Center, Jiangmen Key Laboratory of Clinical Biobanks and Translational Research, Jiangmen Central Hospital, Jiangmen, China, [3]Department of Clinical Laboratory, Jiangmen Central Hospital, Jiangmen, China, [4]Department of Research and Development, Guangdong Research Institute of Genetic Diagnostic and Engineering Technologies for Thalassemia, Hybribio Limited, Guangzhou, China, [5]Dongguan Key Laboratory of Medical Bioactive Molecular Developmental and Translational Research, Guangdong Provincial Key Laboratory of Medical Molecular Diagnostics, Guangdong Medical University, Dongguan, China, [6]Collaborative Innovation Center for Antitumor Active Substance Research and Development, Guangdong Medical University, Zhanjiang, China, [7]Department of Respiratory and Critical Care Medicine, Jiangmen Central Hospital, Jiangmen, China

Background: Acute respiratory distress syndrome (ARDS) is an unresolved challenge in the field of respiratory and critical care, and the changes in the lung microbiome during the development of ARDS and their clinical diagnostic value remain unclear. This study aimed to explore the role of the lung microbiome in disease progression in patients with sepsis-induced ARDS and potential therapeutic targets.

Methods: Patients with ARDS were divided into two groups according to the initial site of infection, intrapulmonary infection (ARDSp, 111 cases) and extrapulmonary infection (ARDSexp, 45 cases), and a total of 28 patients with mild pulmonary infections were enrolled as the control group. In this study, we sequenced the DNA in the bronchoalveolar lavage fluid collected from patients using metagenomic next-generation sequencing (mNGS) to analyze the changes in the lung microbiome in patients with different infectious site and prognosis and before and after antibiotic treatment.

Results: The Shannon–Wiener index indicated a statistically significant reduction in microbial diversity in the ARDSp group compared with the ARDSexp and control groups. The ARDSp group was characterized by a reduction in microbiome diversity, mainly in the normal microbes of the lung, whereas the ARDSexp group was characterized by an increase in microbiome diversity, mainly in conditionally pathogenic bacteria and intestinal microbes. Further analysis showed that an increase in *Bilophila* is a potential risk factor for death in ARDSexp. An increase in *Escherichia coli*, *Staphylococcus aureus*, *Candida albicans*, enteric microbes, or conditional pathogens may be risk factors for death in ARDSp. In contrast, *Hydrobacter* may be a protective factor in ARDSp.

Conclusion: Different initial sites of infection and prognoses are likely to affect the composition and diversity of the pulmonary microbiome in patients with septic ARDS. This study provides insights into disease development and exploration of potential therapeutic targets.

OPEN ACCESS

Edited by:
William C. Cho, QEH, Hong Kong SAR, China

Reviewed by:
Alejandra Hernández-Teran, Instituto Nacional de Enfermedades Respiratorias-México (INER), Mexico
Dinesh Yadav, University of Alabama at Birmingham, United States

***Correspondence:**
Xin Zhang
zhangx45@mail3.sysu.edu.cn
Shuang Zhang
13828075559@163.com
Yanming Huang
huangyanming_jxy@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *26 January 2022*
Accepted: *03 May 2022*
Published: *23 June 2022*

Citation:
Zhang P, Liu B, Zheng W, Chen Y, Wu Z, Lu Y, Ma J, Lu W, Zheng M, Wu W, Meng Z, Wu J, Zheng Y, Zhang X, Zhang S and Huang Y (2022) Pulmonary Microbial Composition in Sepsis-Induced Acute Respiratory Distress Syndrome. Front. Mol. Biosci. 9:862570. doi: 10.3389/fmolb.2022.862570

Keywords: acute respiratory distress syndrome(ARDS), sepsis, metagenomic next-generation sequencing(mNGS), pulmonary microbiome, acute respiratory distress syndrome of pulmonary, acute respiratory distress syndrome of extrapulmonary

1 INTRODUCTION

Acute respiratory distress syndrome (ARDS) is a clinical syndrome that occurs during severe infections, shock, trauma, and burns and is mostly caused by sepsis (Force et al., 2012; Matthay et al., 2019). It is characterized by hypoxemia and respiratory distress. Owing to its rapid progression, high mortality rate, and lack of effective treatments, ARDS is one of the leading causes of death in patients with acute and critical illnesses. In ARDS, damage to alveolar epithelial cells and pulmonary capillary endothelial cells causes increased alveolar-capillary permeability and the collection of protein-laden edema fluid in the alveolar lumen, eventually leading to diffuse interstitial lung edema (Matthay et al., 2019). In recent years, there has been increasing interest in research on ARDS, and the exploration of timely and effective treatments and prognostic markers for ARDS remains a hot topic of research.

In a previous study, we found that the use of metagenomic next-generation sequencing (mNGS) for ARDS caused by severe pneumonia improved clinical diagnosis and guided the clinical use of drugs, thereby improving patient prognosis (Zhang et al., 2020). In addition, changes in the lung microbiome in patients can be explored using mNGS. Research on the relationship between the lung microenvironment and the etiology of ARDS is still in its early stages. Dickson et al. (2018) suggested the presence of interactions between alterations in the pulmonary microbiome and ARDS. ARDS infections or other lung injuries can directly alter the lung microbiome, including ventilator-induced injury and aspiration. Alterations in the pulmonary microbiome may, in turn, contribute to lung injury by promoting increased pulmonary vascular permeability, the establishment of stark oxygen gradients, a surge in inflammatory molecules for bacterial growth (Freestone et al., 2012; Dickson et al., 2015), and damage to host defenses, ultimately altering the alveolar microenvironment. Once both lung microbial imbalance and lung injury occur, they interact via a positive feedback loop. Therefore, regulation of the microbiome is likely to be a potential therapeutic or prophylactic target for ARDS (Dickson, 2016). However, the relationship between alterations in the microbiome and ARDS and their influence on disease regression and prognosis remain to be further evaluated.

This study aimed to explore the role of the lung microbiome in disease progression in patients with sepsis-induced ARDS. Moreover, potential therapeutic targets were screened based on changes in the microbiome in the lung microenvironment.

2 MATERIALS AND METHODS

2.1 Ethics and Informed Consent

The study protocol was reviewed and approved by the Ethics Review Committee of Jiangmen Central Hospital (No. 2019-15). Written informed consent was obtained from the patients or their legal representatives before the collection of bronchoalveolar lavage fluid (BALF) samples by bronchoalveolar lavage using fiberoptic bronchoscopy.

2.2 Patients

A retrospective analysis was conducted on patients with ARDS caused by sepsis who were admitted to the intensive care unit (ICU) of Jiangmen Central Hospital from January 2018 to June 2021. The inclusion criteria were as follows: 1) the diagnosis of ARDS met the Berlin 2012 definition (Force et al., 2012), 2) the etiology of ARDS was sepsis, 3) the age was greater than 18 years, and 4) the clinical profile was complete. The exclusion criteria were as follows: 1) ARDS caused by non-infectious factors, 2) age <18 years, and 3) an incomplete clinical profile. The patients were divided into two groups according to their initial infection status: intrapulmonary infection-induced ARDS (ARDSp group) and extrapulmonary infection-induced ARDS (ARDSexp group). The ARDSp group included 111 patients, the ARDSexp group included 45 patients, and the control group included 28 patients, including patients with mild pulmonary infection and non-ARDS, all of whom had a good prognosis and did not return to the ICU within 90 days of being transferred out of the ICU.

2.3 General Treatment Plan

All patients with sepsis were treated according to the sepsis guidelines (Rhodes et al., 2017) and empirical anti-infective therapy in conjunction with clinical indicators of infection and imaging information. Patients with ARDS were mechanically ventilated according to the ARDS ventilation guidelines (Bein et al., 2016; Griffiths et al., 2019), and the anti-infective regimen was adjusted according to the patient's inflammatory indicators, imaging data, and microbiological tests.

2.4 BALF Collection Process

All patients were intubated and mechanically ventilated in the ICU, and BALF specimens were obtained using a fiberoptic bronchoscope (Chinese Thoracic Society, 2017). Baseline specimens were collected within 24 h of ARDS diagnosis in the ICU before antibiotic administration. Some patients were treated for 7 days, and post-treatment specimens were retained. The baseline and post-treatment specimens were sent to the laboratory for pathogenic culture. The remaining specimens were also sent to the clinical experimental center for DNA extraction and stored at -20°C for research purposes. All laboratory consumables were purchased from Guangzhou Jet Bio-Filtration Co., Ltd. (China). The final specimens were sent to Guangdong Longsee Biomedical Co., Ltd. for metagenomic sequencing, including Bacterial nucleic acid purification, DNA library preparation, high-throughput sequencing, bioinformatics analysis, and pathogenic data interpretation (Miao et al., 2018).

2.5 Experimental Groups

As shown in **Figure 1**, patients in all groups were stratified according to their prognosis. Those who improved in the ICU

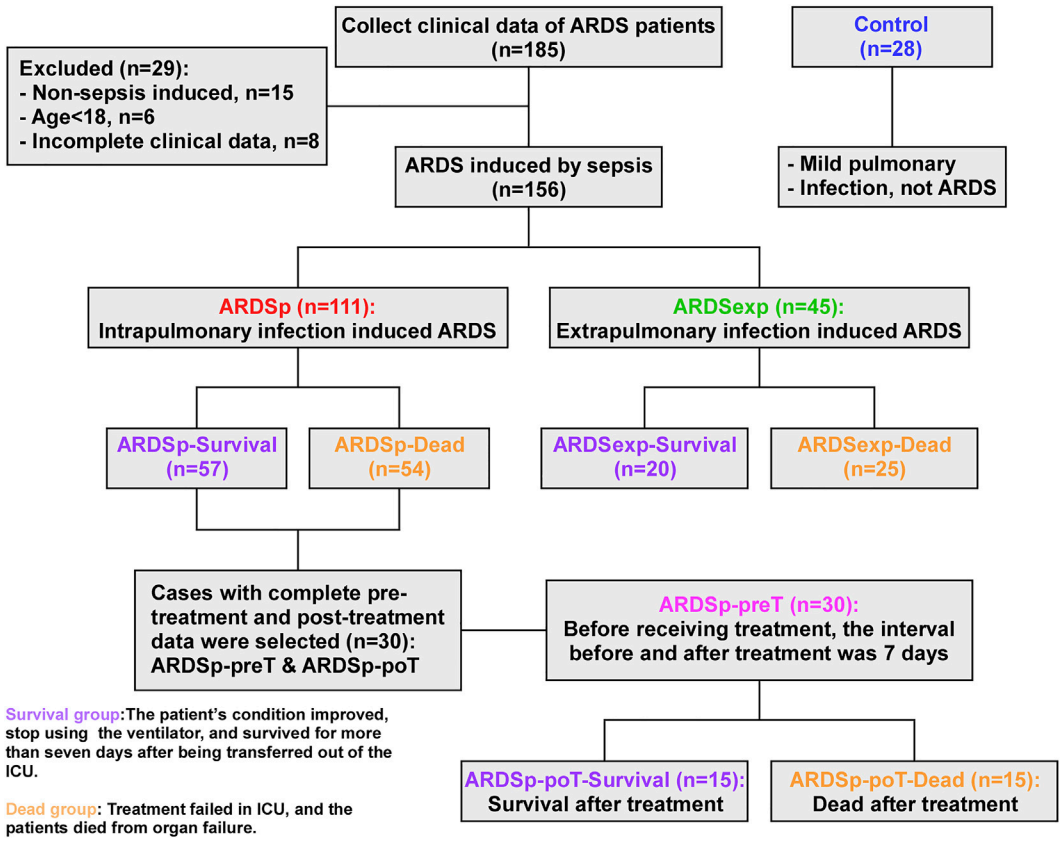


FIGURE 1 | Schematic diagram of the experimental groupings.

with treatment and ventilation and could be successfully transferred out of the ICU within 7 days were defined as the "survival group," while those who did not benefit from ICU care and died of organ failure were defined as the "dead group." Based on different etiologies with different prognoses, 57 cases were classified as the ARDSp-survival group, 54 cases as the ARDSp-dead group, 20 cases as the ARDSexp-survival group, and 25 cases as the ARDSexp-dead group. We analyzed the similarities and differences in the microbiome between the different groups and searched for microbial markers associated with prognosis using metagenomic DNA sequencing of the BALF collected from these patients.

Thirty patients with complete data in the ARDSp group, including pre-treatment and post-treatment data, were selected and divided into the pre-treatment group (ARDSp-preT group) and post-treatment group (ARDSp-poT group). The ARDSp-poT-survival (post-treatment) group was defined as the post-treatment-survival group, with 15 cases, based on the improvement after ICU care allowing the cessation of ventilator use and survival within 7 days of transfer out of ICU transfer. Conversely, the ARDSp-poT-dead group was defined as a failure of treatment in the ICU and death due to organ failure (15 cases). The control group included 28 cases. The similarities and differences in the microbiome between the different groups were also analyzed.

2.6 Pathogenic and Background Microorganisms

In total, 2,728 microorganisms were sequenced in this study. The MetaPhlAn database was used for the taxonomic assignment (Truong et al., 2015). As the RPM values of different microorganisms varied considerably, the microorganisms were divided into pathogenic and background microorganisms and were analyzed separately. Pathogenic microorganisms had significantly higher RPM values than the background microorganisms. We extracted common nosocomial infection pathogenic microorganisms, including common bacteria, fungi, viruses, and specific pathogens, based on 2019 CHINET surveillance data (Hu et al., 2020) and common pathogenic microorganisms in sequencing laboratory testing, and defined 57 microorganisms as pathogenic microorganisms (**Supplementary Table S1**). The remaining microorganisms were considered background microorganisms, with a total of 1,040 at the genus level.

2.7 Metagenomic Next-Generation Sequencing and Analysis

2.7.1 Nucleic Acid Extraction and Library Preparation

DNA was extracted using nucleic acid extraction kit (#20150013, Hybribio, China) and stored at -20°C to prevent degradation.

A DNA library was constructed using the VAHTS® Universal Plus DNA Library Prep Kit for Illumina

(Guangdong Longsee Biomedical Co., Ltd, China). All experiments were performed on ice. The reaction system was prepared in sterile PCR tubes according to the manufacturer's instructions, and DNA fragmentation and adapter ligation were performed by polymerase chain reaction (PCR). The products were purified using VAHTS DNA Clean Beads. Library amplification was then performed by PCR and the amplified products were purified again. PCR primer information is shown in **Supplementary Table S2**.

2.7.2 Bioinformatics Analysis

After library preparation, high-throughput sequencing was performed using the NextSeq CN500 sequencing platform, and sequencing data were formatted using the official Illumina software bcl2fastq to obtain FASTQ files for individual samples. Quality control was performed using the quality control software fastp (version 0.20.0) to remove bases with an average qc value <20 within 4 bp of the end and to remove reads <70 bp in length in the FASTQ files. After quality control, the host genome was removed by matching the quality-controlled reads to the human genome (hg39) using Bowtie (version 1.2.2) and retaining unmatched reads for subsequent analysis. After removing the human reads, the remaining reads were matched to a reference genome database of five pathogenic microorganisms (Longsee Clinical Pathogen Sequence Database) using bowtie2 (version 2.3.4.1) in an end-to-end matching mode. The results were further standardized in Python and R before analysis (RPM = number of reads on match/total number of reads ∗ 1000000).

2.8 Statistical Analysis

We divided the microorganisms into pathogenic and background microorganisms, and counted the number of positive and negative sequencing results for each group separately. Among pathogenic microorganisms, those with RPM values ≥1 were defined as positive and those with RPM values <1 as negative; among background microorganisms, those with RPM values >0 were defined as positive and those with RPM values = 0 as negative.

To analyze the positive rate of the microbiota, we used the chi-square test or Fisher's test and presented the data using a histogram. To analyze the abundance of the microbiota, we log2-processed the RPM values from the sequencing results, compared the median values using a paired t-test, and presented the data using a heatmap. Analyses were performed using GraphPad 9.3 or R4.1.2. Statistical significance was set as $p < 0.05$. Cox and cluster analyses were performed using the SPSS software (version 26.0) and multiple experiment viewer, respectively. Principal component analysis (PCA) was performed using GraphPad prism 9.3 and the Adonis analysis was performed on the Omicshare platform. The Shannon–Wiener index was calculated as follows: $H = \sum(pi)(\ln pi)$.

3 RESULTS

3.1 Patient Clinical Characteristics

A total of 156 patients with sepsis-induced ARDS were selected for this study and were divided into two groups according to their initial infection status: intrapulmonary infection-induced ARDS (ARDSp, n = 111), and extrapulmonary infection-induced ARDS groups (ARDSexp, n = 45), and a control group (n = 28, **Figure 1**). No significant statistical difference in age and sex was found between the ARDS and control groups (**Table 1**). The ARDS group was divided into the ARDSp group and the ARDSexp group for comparison, but there was no significant statistical difference in the basic characteristics of the two groups. (**Table 2**). Pre-treatment laboratory tests showed that the ARDSexp group had significantly different results compared with the ARDSp group, including PCT ($p < 0.001$), PLT ($p = 0.011$), Scr ($p = 0.041$), T. Bil ($p = 0.014$), Lac ($p < 0.001$), pre-treatment APACHE II ($p < 0.001$), and the SOFA score ($p < 0.001$; **Table 3**). In addition, a higher number of patients in the ARDSexp group required treatment with vasoactive drugs ($p < 0.001$) and CRRT ($p < 0.001$) than in the ARDSp group, with no statistical difference in the 90-day all-cause mortality (**Table 4**). These findings suggest that patients in the ARDSexp group were sicker than those in the ARDSp group; however, there was no significant difference in the prognosis between the two groups.

TABLE 1 | General information for the ARDS group versus the control group.

	ARDS (n = 156)	Control (*n* = 28)	*p* value
Age (years)			
≥60, n (%)	95 (60.9)	13 (46.4)	0.152
<60, n (%)	61 (39.1)	15 (53.6)	
Gender			
Male, n (%)	106 (67.9)	16 (57.1)	0.265
Female, n (%)	50 (32.1)	12 (42.9)	

3.2 Comparison of the Lung Microbiome Between the ARDSp and ARDSexp Groups

ARDS may be caused by several pathogenic microorganisms, including bacteria, fungi, viruses, and specific pathogens. We first compared the composition and abundance of pathogenic microorganisms among the control, ARDSp, and ARDSexp groups. Shannon's diversity index results suggested that the ARDSp group had less microbiome diversity compared with the ARDSexp and control groups (**Supplementary Figure 7A**). Further analysis revealed that the ARDSp group had a higher positive rate for *Escherichia coli*, *Staphylococcus haemolyticus*, and fungi than the control group (**Figure 2A**); the ARDSexp group had a higher positive rate for *Escherichia coli* than both groups (**Figure 2B**). No statistically significant difference in the pathogenic microorganisms was found between the ARDSp and ARDSexp groups (**Supplementary Figure 1A**).

Furthermore, we performed PCA on background microorganisms among the control, ARDSp, and ARDSexp groups. As shown in the graph (**Figure 2C**), the control group had the most concentrated community composition, whereas both the ARDSp and ARDSexp groups had a larger number of different microorganism species than the control group. This

TABLE 2 | Patient characteristics and baseline information for the ARDS group versus the ARDSexp group.

	ARDSp (*n* = 111)	ARDSexp (*n* = 45)	*p* value
Age (years)			
≥60, n (%)	68 (61.3)	27 (60.0)	0.884
<60, n (%)	43 (38.7)	18 (40.0)	
Gender			
Male, n (%)	77 (69.4)	29 (64.4)	0.550
Female, n (%)	34 (30.6)	16 (35.6)	
Underlying diseases			
Hypertension, n (%)	35 (31.5)	11 (24.4)	0.379
Coronary heart disease, n (%)	11 (9.9)	3 (6.7)	0.521
COPD, n (%)	19 (17.1)	4 (8.9)	0.189
Chronic nephrosis, n (%)	12 (10.8)	3 (6.7)	0.426
Diabetes, n (%)	19 (17.1)	11 (24.4)	0.293
Immunosuppression, n (%)	15 (13.5)	3 (6.7)	0.225
Tumor, n (%)	9 (8.1)	4 (8.9)	0.873
Smoking, n (%)	35 (31.5)	8 (17.8)	0.082
Drinking, n (%)	14 (12.6)	4 (8.9)	0.435
Primary site of infection			
Lung, n (%)	111 (100)	0 (0)	—
Blood, n (%)	0 (0)	7 (15.6)	—
Gastrointestinal tract, n (%)	0 (0)	18 (40.0)	—
Liver, gallbladder, and pancreas, n (%)	0 (0)	9 (20.0)	—
Skin, n (%)	0 (0)	10 (22.2)	—
Urinary system, n (%)	0 (0)	1 (2.2)	

The difference between the two groups was tested by the chi-square test. $p < 0.05$ *was considered statistically significant. COPD, chronic obstructive pulmonary disease.*

TABLE 3 | Comparison of laboratory examination, ventilator parameters, APACHE II score, and SOFA score before treatment between the two groups of patients.

	ARDSp (*n* = 111)	ARDSexp (*n* = 45)	*p* value
Laboratory examination before treatment			
PCT (ug/L)	1.9 (0.4, 10.0)	16.2 (6.0, 72.8)	< 0.001*
WBC (109/L)	13.0 (6.9, 17.2)	13.2 (8.2, 19.2)	0.197
PLT (109/L)	176 (112, 218)	115 (57, 204)	0.011*
Scr (μmol/L)	90 (70, 188)	142 (89, 260)	0.041*
T.Bil (mmol/L)	12.4 (7.4, 22.9)	21.0 (11.3, 50.9)	0.014*
ALT (IU/L)	24 (13, 51)	37 (25, 87)	0.259
Lac (mmol/L)	1.8 (1.2, 2.8)	3.6 (1.7, 7.6)	< 0.001*
OI	148 (106, 181)	164 (140, 210)	0.084
APACHE II score before treatment	21 (18, 23)	22 (19, 25)	< 0.001*
SOFA score before treatment	7 (5, 8)	8 (6, 10)	< 0.001*

Patient physiological index measurements are presented as median (interquartile). $p < 0.05$ *is considered statistically significant. PCT, procalcitonin; WBC, white blood cell; PLT, platelet; Scr, serum creatinine; T.Bil, total bilirubin; ALT, alanine aminotransferase; Lac, lactic acid; OI, oxygen index; APACHE, acute physiology and chronic health evaluation; SOFA, sequential organ failure assessment.*

TABLE 4 | Comparison of special ICU treatment and prognosis between two groups of patients.

	ARDSp (*n* = 111)	ARDSexp (*n* = 45)	*p* value
Use of vasoactive drugs, n (%)	51 (45.9)	39 (86.7)	< 0.001*
CRRT, n (%)	13 (11.7)	20 (44.4)	< 0.001*
ECMO, n (%)	8 (7.2)	0 (0)	0.106
Prone ventilation, n (%)	22 (19.8)	3 (6.7)	0.053
All-cause mortality at 90 days, n (%)	54 (48.6)	25 (55.6)	0.619

The difference between the two groups was tested by chi-square test. $p < 0.05$ *was considered statistically significant. CRRT, continuous renal replacement therapy; ECMO, extracorporeal membrane oxygenation.*

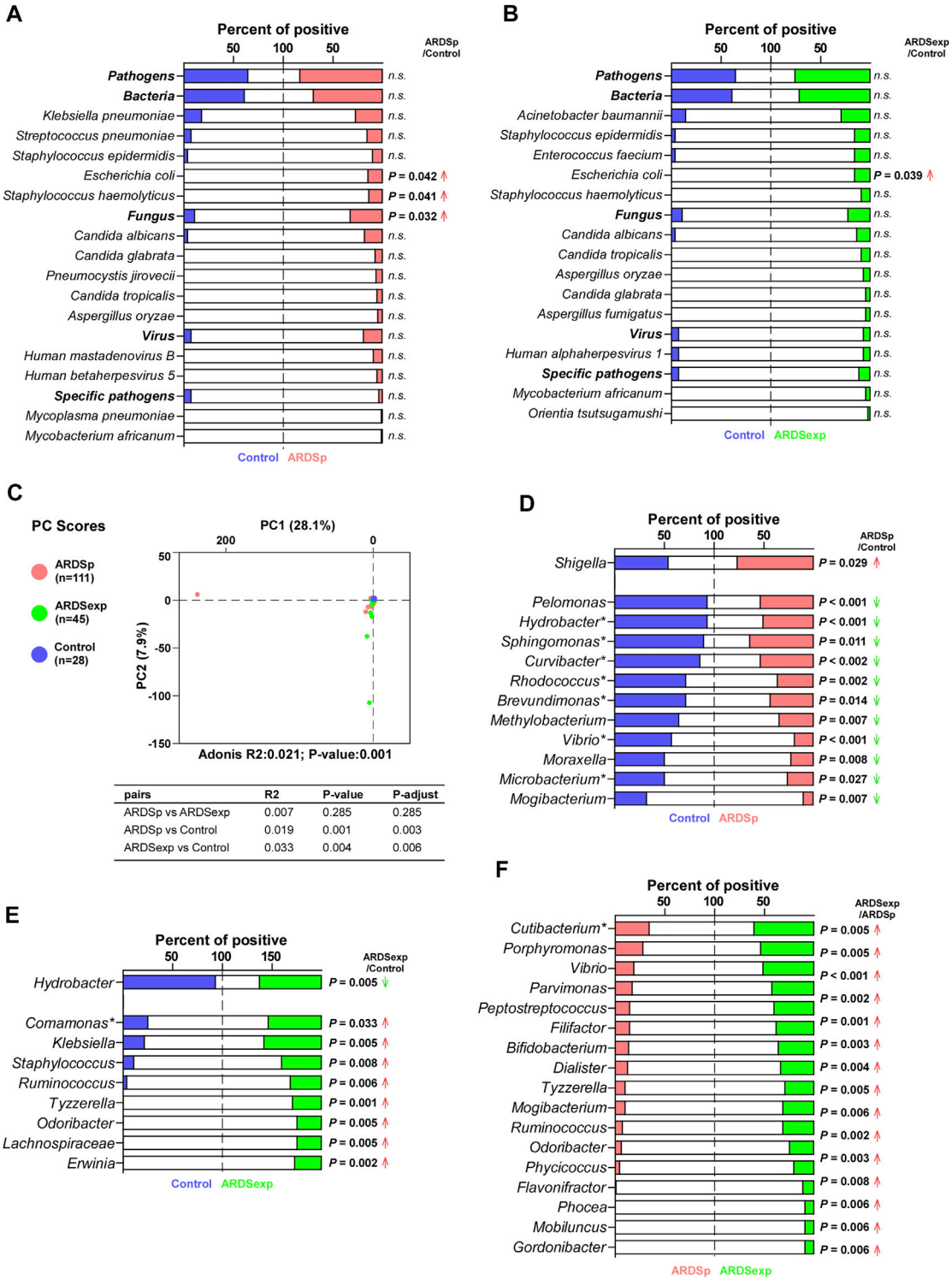


FIGURE 2 | Comparison of the pulmonary microbiome between the ARDSp and ARDSexp groups. Comparison of pathogenic microorganisms. **(A)** Compared with the control group, the ARDSp group had a higher positive rate for *Escherichia coli*, *Staphylococcus haemolyticus,* and *Fungus*. **(B)** Compared with the control group, the ARDSexp group had a higher positive rate for *Escherichia coli*. Comparison of background microorganisms: **(C)** PCA showed that the community composition of ARDSp and ARDSexp groups was different from that of the control group. **(D)** ARDSp group had a predominantly reduced positive rate for the ARDSp group compared with the control group, except for *Shigella*. **(E)** ARDSexp group showed a predominant increase in the positive rate in the ARDSexp group compared with the control group, except for *Hydrobacter*. **(F)** ARDSexp group has a higher positive rate of the microbiome compared with the ARDSp group. R2: variation; P-adjust: *p* value was adjusted by using the Benjamini–Hochberg (BH) method. Note: asterisk represents the microbiome with a simultaneous increase or decrease in positive rate and abundance comparisons.

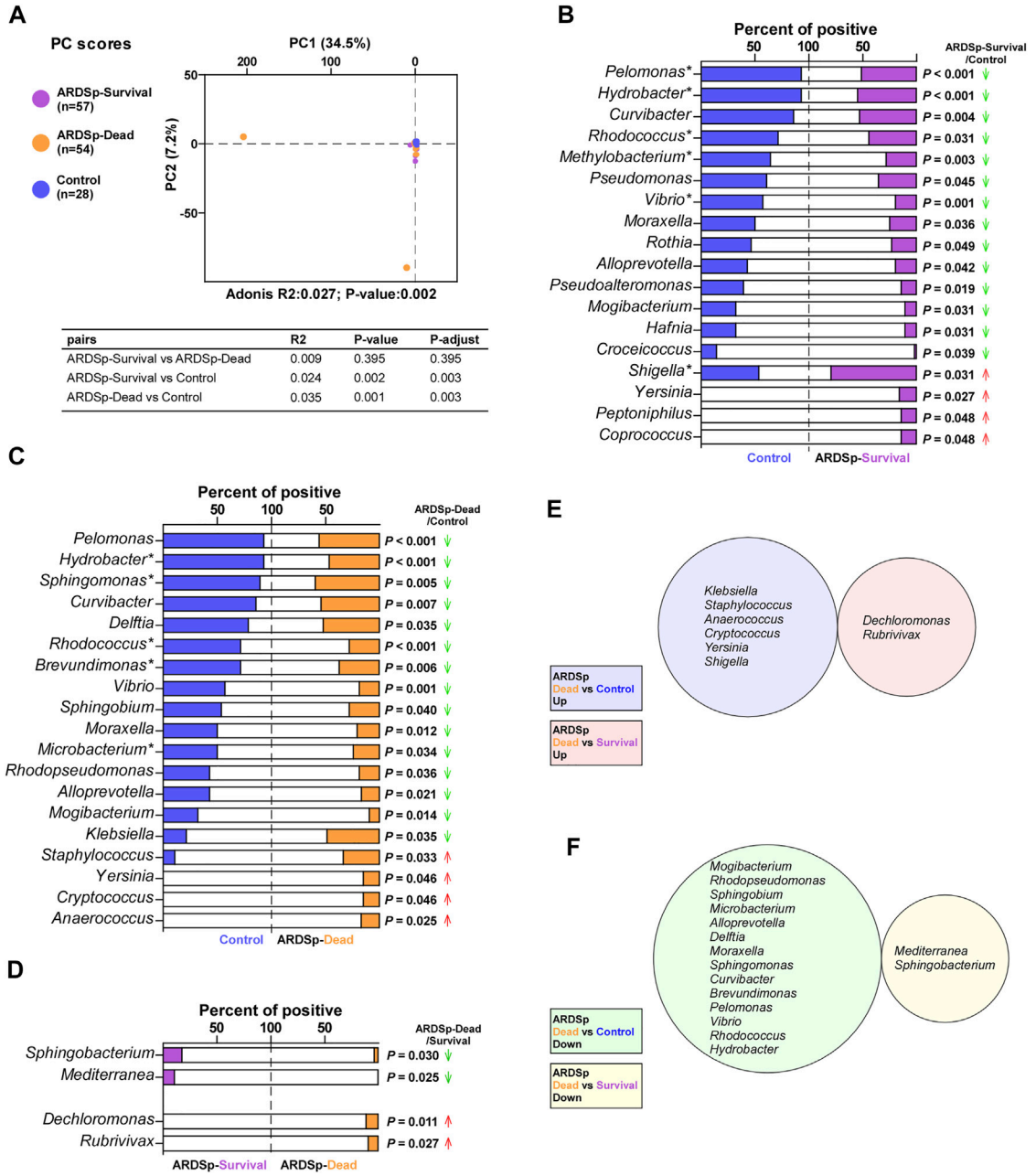


FIGURE 3 | Microbial analysis associated with the prognosis of the ARDSp group. Comparison of background microorganisms: **(A)** PCA showed the community composition of ARDSp-survival and ARDSp-dead groups was significantly different from that of the control group. **(B)** ARDSp-survival group had a predominantly reduced rate of positivity compared with the control group. **(C)** ARDSp-dead group also had a predominantly reduced positive rate when compared with the control group. **(D)** There were four pathogenic microorganisms with statistically significant differences in the positive rate in the ARDSp-dead group compared with the ARDSp-survival group. **(E)** No simultaneous increases in the background microbiome were found in the ARDSp-dead group compared with the control and ARDSp-survival groups. **(F)** No simultaneous decreases in the background microbiome were found in the ARDSp-dead group compared with the control and ARDSp-survival groups. R2: variation; P-adjust: *p* value was adjusted by using the Benjamini–Hochberg (BH) method.

indicates that the community composition of the background microorganisms was relatively consistent in the control group, and there were a few abnormally increased background microorganisms in the ARDSp and ARDSexp groups. In a subsequent comparison between the ARDSp, ARDSexp, and control groups, we found that the ARDSp group was characterized by a reduction in both positivity and abundance compared with the control group (**Figure 2D**; **Supplementary Figure 1B**), suggesting that the ARDSp microbiome was characterized by a reduction in diversity. Therefore, we speculate that an increase in pathogenic microorganisms inhibits the growth of the "normal" respiratory microbiome.

Interestingly, increased positivity and an abundance of intestinal microbes such as *Shigella* in ARDSp were observed. In contrast, the ARDSexp group presented a predominant increase in the positivity of microbes including *Comamonas*, *Klebsiella*, *Staphylococcus*, *Ruminococcus*, *Tyzzerella*, *Odoribacter*, Lachnospiraceae, and *Erwinia,* implying that the ARDSexp microbiome was characterized by an increased diversity of the microbiome (**Figure 2E**; **Supplementary Figure 1C**), most of which are of intestinal origin. In addition, the ARDSexp group had a higher positive rate and abundance of microbes than the ARDSp group such as *Cutibacterium* (**Figure 2F**; **Supplementary Figure 1D**), which were also mainly of intestinal origin.

3.3 Microbial Analysis Associated With the Prognosis of the ARDSp Group

We divided the ARDSp group into ARDSp-survival and ARDSp-dead groups according to different prognoses and performed PCA together with the control group (**Figure 3A**). PCA results showed that there was a relatively large amount of taxonomic overlap between the control and ARDSp-survival groups, with more obviously isolated specimens in the ARDSp-dead group. This suggests that the control and ARDSp-survival groups had a relatively similar microbial community composition, whereas the ARDSp-dead group had several abnormally increased background microorganisms.

The pathogenic microorganisms and background microbiome that increased or decreased simultaneously in the ARDSp-dead group compared with the control and ARDSp-survival groups are probably related to prognosis. Therefore, we compared the composition and abundance of microorganisms among these three groups. The Shannon index result suggested that the ARDSp group had less diversity than the control group; however, there was no statistically significant difference in the microbial diversity between the ARDSp-survival and ARDSp-dead groups (**Supplementary Figure 2B**). In addition, the ARDSp-dead group had a significantly higher positive rate for pathogenic microorganisms including total pathogens, total fungi, *Staphylococcus haemolyticus,* and *Escherichia coli* than the control group (**Supplementary Figure 2B**). Compared with the ARDSp-survival group, the ARDSp-dead group had a significantly higher positive rate for *Pseudomonas aeruginosa* (**Supplementary Figure 2C**).

Moreover, the ARDSp-survival and ARDSp-dead groups showed a predominant decrease in the positivity and abundance of background microorganisms compared with the control group (**Figures 3B–D**; **Supplementary Figure 2E,F**). Nevertheless, no prognosis-related microbes were identified among pathogenic or background microorganisms (**Figures 3E,F**).

3.4 Microbial Analysis Associated With the Prognosis of the ARDSexp Group

We divided the ARDSexp group into ARDSexp-survival and ARDSexp-dead groups according to different prognoses and performed PCA together with the control group (**Figure 4A**). The PCA results revealed that there was a relatively large taxonomic overlap between the control and ARDSexp-survival groups, whereas the ARDSexp-dead group had more distinct isolated specimens. This suggests that the control and ARDSexp-survival groups had a similar microbial community composition, whereas the ARDSexp-dead group had several abnormally increased background microorganisms.

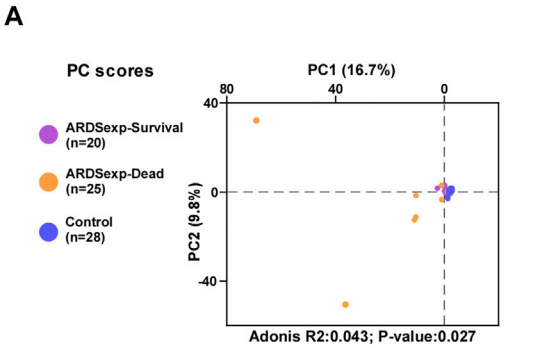
Therefore, we compared the differences in the composition and abundance of microorganisms among the three groups. The Shannon index result indicated a similar diversity between the ARDSexp-survival, ARDSexp-dead, and control groups (**Supplementary Figure 7C**). The ARDSexp-dead group had significantly higher positive rates for *Escherichia coli* and *Haemophilus influenzae* than the control and ARDSexp-survival groups, respectively (**Supplementary Figure 3A,B**). However, no pathogenic microbial markers that might be associated with prognosis were identified.

In addition, the ARDSexp-survival and ARDSexp-dead groups showed a predominant increase in the positivity and abundance of background microorganisms compared with the control group (**Figures 4B–D** and **Supplementary Figure 3E,F**). In contrast, the ARDSp group showed a decrease in these microorganisms. Notably, there was a simultaneous increase in the microorganisms of the ARDSexp-dead group including *Bifidobacterium*, *Bilophila*, *Mediterranea*, *Anaerostipes*, *Bacillus*, *Dorea*, and *Collinsella* (**Figure 4E**), when compared with the ARDSexp-survival and control groups. Further prognostic survival analysis was conducted and the results indicated that these seven microbes were strongly associated with poor prognosis (**Figures 5A–G**). In contrast, the Cox univariate analysis showed that increased pre-treatment APACHE II and SOFA scores and increased *Bifidobacterium*, *Bilophila*, *Mediterranea*, *Bacillus*, *Dorea*, and *Collinsella* may be risk factors for ARDS (**Figure 5H**). Furthermore, the Cox multivariate analysis also indicated that the increase in *Bilophila* was most likely associated with mortality in patients with ARDS (**Figure 5I**).

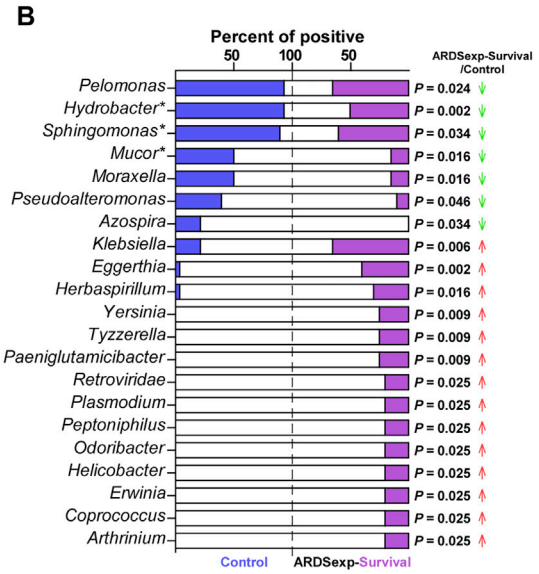
3.5 Changes in the Lung Microbiome Before and After Treatment in the ARDSp Group

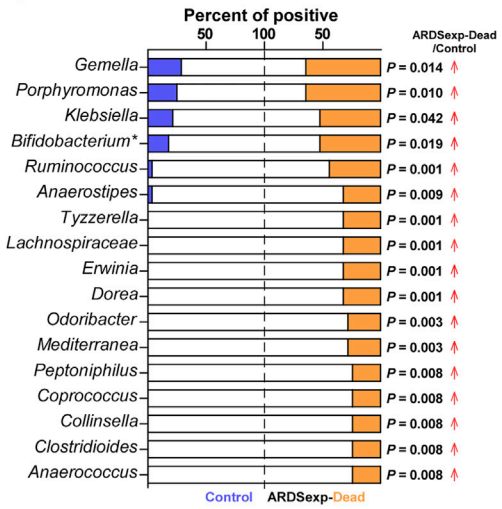
To explore the changes in the lung microbiome before and after treatment in the ARDSp group, 30 cases with complete pre-and post-treatment comparative data were selected for analysis and screened for pathogenic organisms associated with death (potential risk factors for death) and markers associated with survival (potential protective factors for survival). Alpha diversity analysis showed that the ARDSp-preT group had lesser microbial diversity than the control group. There was a minor but not statistically significant increase in the microbial diversity after treatment compared with pre-treatment (**Supplementary Figure 7D**).

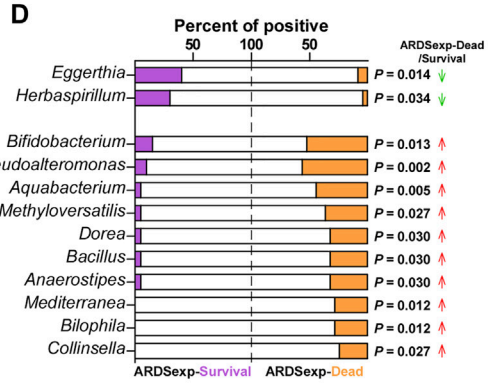
We conducted a cluster analysis to determine differences in pathogenic microorganisms and found that *Escherichia coli*, *Staphylococcus aureus*, and *Candida albicans* were significantly increased in the ARDSp-poT-dead group, whereas *Acinetobacter*



pairs	R2	P-value	P-adjust
ARDSexp-Survival vs ARDSexp-Dead	0.025	0.504	0.504
ARDSexp-Survival vs Control	0.040	0.014	0.042
ARDSexp-Dead vs Control	0.031	0.050	0.075







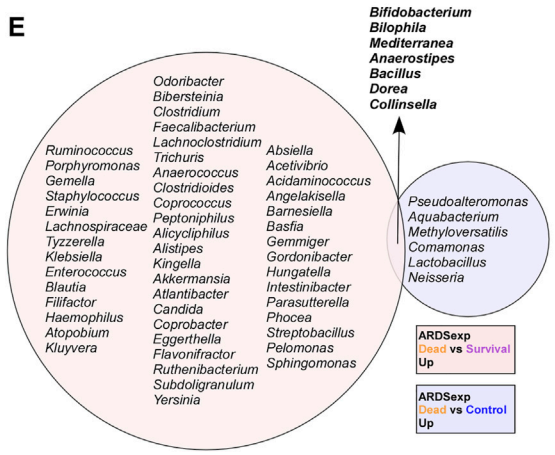


FIGURE 4 | Microbial analysis associated with the prognosis of the ARDSexp group. Comparison of background microorganisms: **(A)** PCA showed that the community composition of the ARDSexp-dead group was different from that of the control group. **(B)** ARDSexp-survival group had a predominantly elevated positive rate compared with the control group. **(C)** ARDSexp-dead group had an elevated positive rate compared with the control group. **(D)** ARDSexp-dead group had a predominantly elevated positive rate when compared with the ARDSexp-survival group. **(E)** ARDSexp-dead group had seven increased background microbiomes, *Bifidobacterium*, *Bilophila*, *Mediterranea*, *Anaerostipes*, *Bacillus*, *Dorea*, and *Collinsella* compared with the control and ARDSexp-survival groups. R2: variation; P-adjust: *p* value was adjusted by using the Benjamini–Hochberg (BH) method.

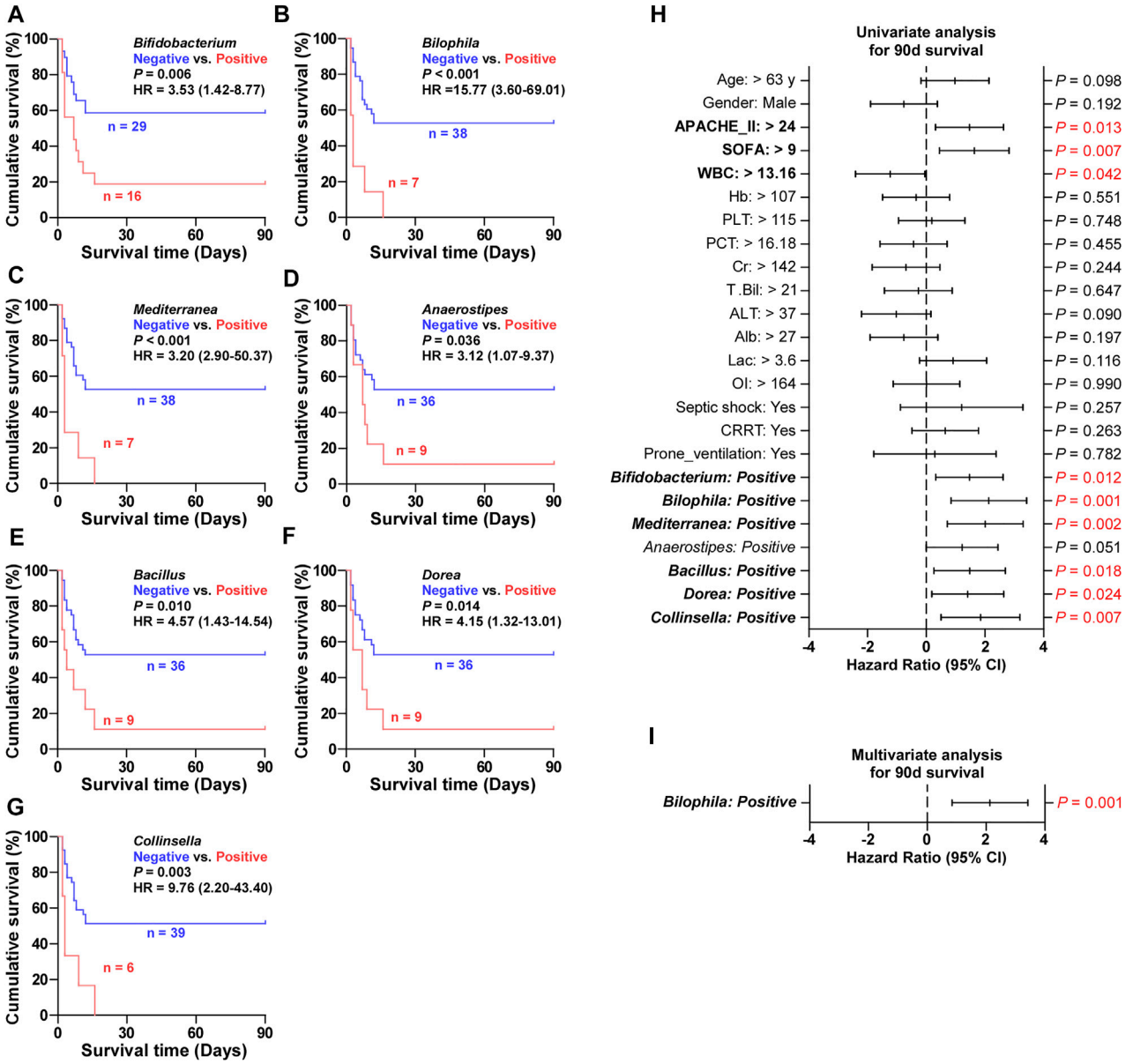


FIGURE 5 | Validation of screened microbial markers associated with ARDSexp prognosis. **(A–G)** Kaplan–Meier analysis revealed that patients with one of the increased background microbiomes including *Bifidobacterium*, *Bilophila*, *Mediterranea*, *Anaerostipes*, *Bacillus*, *Dorea*, and *Collinsella*, had a shorter survival time. **(H)** Cox univariate analysis revealed that elevated pre-treatment APACHE II and SOFA scores and increased *Bifidobacterium*, *Bilophila*, *Mediterranea*, *Anaerostipes*, *Bacillus*, *Dorea*, and *Collinsella* were likely to be contributing factors to the death of ARDS patients. **(I)** Cox multivariate analysis revealed that *Bilophila* was the most significant and underlying risk factor for mortality in ARDS patients.

baumannii and *Acinetobacter nosocomialis* were increased or unchanged, mainly in the ARDSp-poT-survival group (**Figure 6A**). An analysis of the microbial composition and abundance revealed increasing trends in positivity (p = 0.014, p = 0.080, and p = 0.002, respectively) and abundance (p = 0.524, p = 0.015, and p = 0.001, respectively) for *Escherichia coli*, *Staphylococcus aureus,* and *Candida albicans* in the ARDSp-poT-dead group (**Figures 6B,D**). The microorganisms that showed a significant increasing trend in positivity in the ARDSp-poT-dead group included *Klebsiella pneumoniae*, *Escherichia coli*, and *Candida albicans* (**Supplementary Figure 4A**). There was also a decreasing trend in the positivity and abundance of *Acinetobacter baumannii* and *Acinetobacter nosocomialis*; however, the result was not statistically significant (**Figures 6C,E**).

In addition, the PCA showed that the ARDSp-poT-survival and control groups were relatively concentrated in background microorganism composition, and the ARDSp-poT-dead and

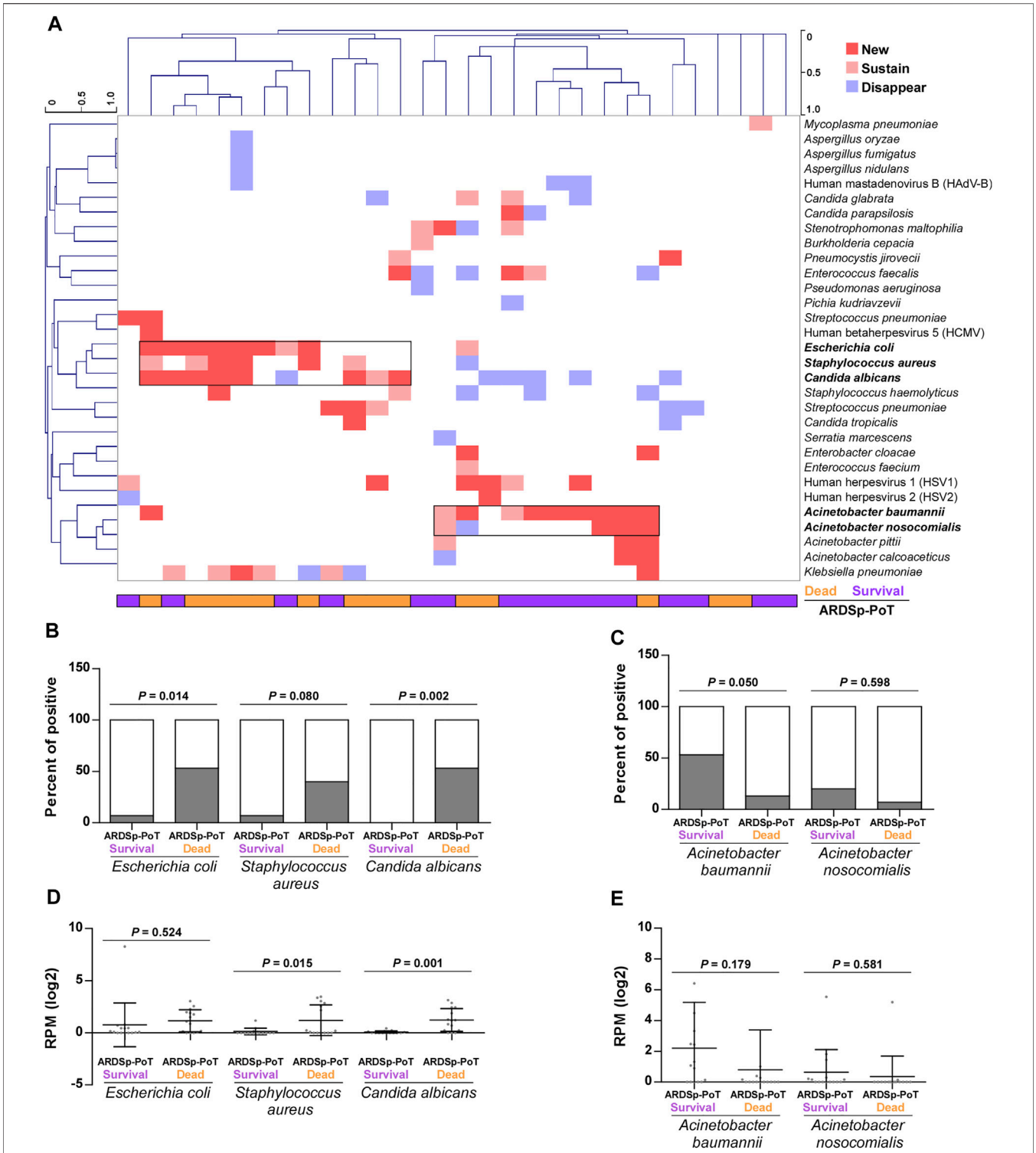
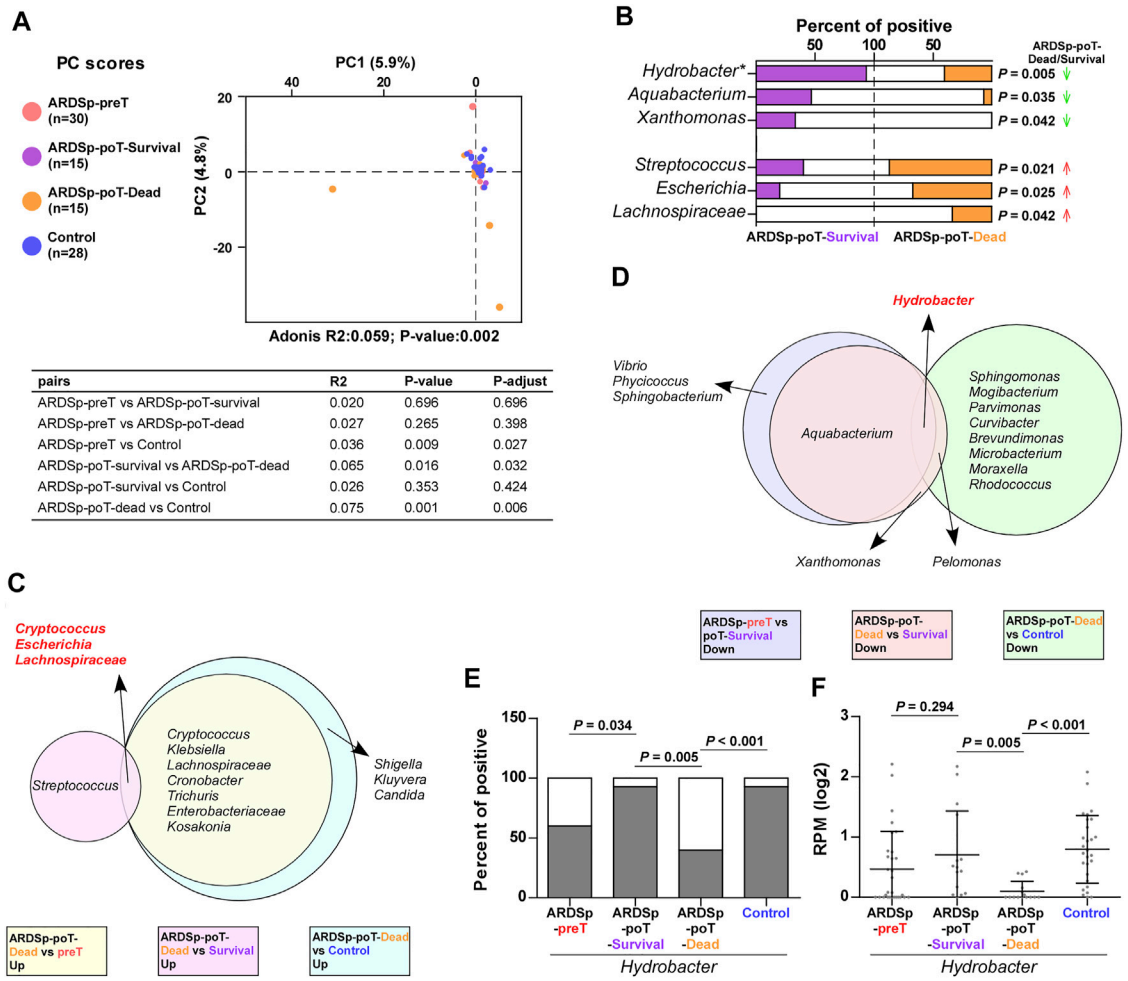


FIGURE 6 | Changes in pulmonary pathogenic microorganisms before and after treatment in the ARDSp group (30 pairs). **(A)** Cluster analysis revealed that *Escherichia coli*, *Staphylococcus aureus,* and *Candida albicans* increased significantly in the ARDSp-poT-dead group, while *Acinetobacter baumannii* and *Acinetobacter nosocomialis* increased or remained unchanged mainly in the ARDSp-poT-survival group. **(B–E)** Comparison of the positive rate and abundance of the aforementioned five pathogenic microorganisms in the ARDSp-poT-dead group versus the ARDSp-poT-survival group.



pairs	R2	P-value	P-adjust
ARDSp-preT vs ARDSp-poT-survival	0.020	0.696	0.696
ARDSp-preT vs ARDSp-poT-dead	0.027	0.265	0.398
ARDSp-preT vs Control	0.036	0.009	0.027
ARDSp-poT-survival vs ARDSp-poT-dead	0.065	0.016	0.032
ARDSp-poT-survival vs Control	0.026	0.353	0.424
ARDSp-poT-dead vs Control	0.075	0.001	0.006

FIGURE 7 | Changes in pulmonary background microorganisms before and after treatment in the ARDSp group (30 pairs). **(A)** PCA showed the community composition of the ARDSp-poT-dead group was different from that of the control group. **(B)** There were six background microorganisms with statistically significant positive rates in the ARDSp-poT-dead group compared with the ARDSp-poT-survival group. **(C)** Co-increased microorganisms in the ARDSp-poT-dead group compared with the ARDSp-preT group, the ARDSp-poT-dead group compared with the ARDSp-poT-survival group, and the ARDSp-poT-dead group compared with the control group included *Cryptococcus*, *Escherichia*, and Lachnospiraceae. **(D)** Commonly reduced microbiome in the ARDSp-preT group compared with the ARDSp-poT-survival group, the ARDSp-poT-dead group compared with the ARDSp-poT-survival group, and the ARDSp-poT-dead group compared with the control group were *Hydrobacter*. **(E,F)** Comparison of *Hydrobacter* in the four groups in terms of positive rate and abundance. R2: variation; P-adjust: *p* value was adjusted by using the Benjamini–Hochberg (BH) method.

ARDSp-preT groups had significantly distinct background microorganisms. These results indicated that the microbial community composition of the control and ARDSp-poT-survival groups were relatively similar, whereas the ARDSp-poT-dead and ARDSp-preT groups had several unusual background microorganisms (**Figure 7A**).

We speculated that the microorganisms that were increased in the ARDSp-poT-survival group compared with the ARDSp-preT group (**Supplementary Figure 6E**), especially those decreased in the ARDSp-poT-dead group compared with the ARDSp-poT-survival group and the control group (**Figure 7B**; **Supplementary Figure 6C**), were potential ARDSp protective factors. Notably, the most common microorganism was *Hydrobacter* (**Figure 7D**); *Hydrobacter* had the highest positivity or abundance values in the control group and decreased as prognosis deteriorated or increased as prognosis improved (**Figures 7E,F**). In contrast, we speculated that the microorganisms that were increased in the ARDSp-poT-dead group compared with the ARDSp-preT group (**Supplementary Figure 6A**), in the ARDSp-poT-dead group compared with the ARDSp-poT-survival group (**Figure 7B**), and an increase of in the ARDSp-poT-dead group compared with the control group (**Supplementary Figure 6C**) were potential risk factors for death in patients with ARDSp. The common microorganisms identified simultaneously were *Cryptococcus*, *Escherichia*, and Lachnospiraceae (**Figure 7C**), which showed the least difference in either positivity or abundance in the control group. These three microorganisms increased as the prognosis worsened or decreased as the prognosis improved (**Supplementary Figure 4D-I**).

4 DISCUSSION

In this study, we investigated the changes in the lung microbiome of patients with sepsis-induced ARDS by mNGS sequencing. We compared the basic clinical characteristics and lung microbial composition of patients in each group and found significant differences between the ARDSp and ARDSexp groups. In addition, we found that the microbial diversity in ARDS induced by intrapulmonary infection was significantly decreased; however, there was no major difference among groups with different prognoses. The microbial profile of the ARDSp group was characterized by a reduced microbiome diversity, dominated by a decline in normal microbes in the lungs. The microbial profile of the ARDSexp group was characterized by an increased diversity of the microbiome, mainly in conditionally pathogenic bacteria and intestinal microbes. A comparison of the lung microbiome between the two groups indicated an increase in the pathogenic microorganisms *Escherichia coli*, *Staphylococcus aureus*, and *Candida albicans* in the lung, or an increase in enteric microbes or conditionally pathogenic bacteria, is potential a risk factor for death in ARDSp. The background microbiome *Hydrobacter* may be a protective factor for survival in ARDSp and the increase in *Bilophila* may be a mortality indicator in ARDSexp.

In addition, the differences in the microbiome between the ARDSp and ARDSexp groups at the baseline level were mainly manifested as differences in background bacterial microorganisms. The PCA results revealed significant differences in the microbial composition between the ARDSp, ARDSexp, and control groups. For example, the detection rate and abundance of the ARDSp microbiome were lower than those of the control and ARDSexp groups which had the highest detection rate and abundance, suggesting a decrease in the diversity of the ARDSp microbiome and an increase in the diversity of the ARDSexp microbiome. In ARDSexp, the epithelium and endothelium remain intact because of sepsis-induced increased vascular permeability and interstitial edema of the lungs; thus, the environment for the growth of the respiratory microbiome is not damaged. Increased vascular permeability leads to protein extravasation, which provides the necessary and abundant nutrients for bacterial growth to ensure that the normal respiratory microbiome is not reduced. The diversity of the microbiome possibly and significantly increases with the migration of the microbiome and the growth of environmental pathogens. Kyo et al., 2019analyzed the pulmonary microbiome of the BALF of patients with ARDS and discovered an increasing trend in pulmonary bacterial burdens, such as 16S rRNA gene copy number, and a pronounced decrease in α diversity. However, the results do not account for the origin of infection (pulmonary or extrapulmonary), and pneumonia was the main disease etiology, accounting for 65% of the cases. Dickson et al., 2016found that the lung microbiome was enriched with enteric bacteria and increased bacterial diversity in a mouse model of lung injury with abdominal sepsis caused by appendiceal ligation and puncture (Dickson et al., 2016), which is in accordance with our observations of the ARDSexp group. In our study, further analysis of the variation in the species of the ARDSp group versus the ARDSp group revealed that the background microorganisms *Hydrobacter*, *Sphingomonas*, *Curvibacter*, *Rhodococcus*, *Brevundimonas*, *Vibrio*, and *Microbacterium* were decreased in the ARDSp group compared with the control group. Several reports have indicated that *Sphingomonas*, *Brevundimonas*, and *Methylobacterium* are pulmonary microbes and conditionally pathogenic bacteria (Hilty et al., 2010; Huffnagle and Dickson, 2015; O'Dwyer et al., 2019). Therefore, the ARDSp microbiome is characterized by an increased abundance of pathogenic microorganisms and a decrease in other pulmonary respiratory microbiomes. In contrast, the ARDSexp group was characterized by an increase in different background microorganisms compared with the control group, both in terms of positivity and abundance. These microbes have been reported to be widespread, including oral and respiratory sources such as *Staphylococcus* and *Klebsiella* (Hostacka, 2001; Dyke, 2003) and intestinal microbes such as *Erwinia*, Lachnospiraceae, *Shigella*, and *Lactobacillus* (Starr and Chatterjee, 1972; Baker and The, 2018; Vacca et al., 2020). Hence, the microbiome of ARDSexp is characterized by an increase in the bacterial load, microbiome diversity, conditionally pathogenic bacteria, and intestinal microbes.

In addition, we compared the similarities and differences in the baseline levels of the lung microbiome in the survival and dead groups to screen for microbes associated with prognosis. In the ARDSp group, no microbe was identified. This result may be because the main etiology of ARDSp was severe pneumonia, which has a wide variety of infectious pathogenic microorganisms, and the infection characteristics and microbial features of different pathogenic microorganisms vary. Therefore, screening for meaningful microbes at the baseline is challenging. Future studies with a larger number of patients to quantity and stratify the analysis based on different pathogenic microorganisms, such as bacteria, fungi, and viruses are required to identify potential biomarkers. In the ARDSexp group, the microbiome was characterized by an increased lung bacterial load and microbiome diversity, and an increase in both conditionally pathogenic and enteric microbes and background microorganisms. Although pathogenic microorganisms related to prognosis were not identified, an increase in *Bilophila* among the background microbiome is most likely a risk factor for death in ARDSexp. *Bilophila*, a genus of intestinal microbes, can be isolated and cultured in abdominal infections, pulmonary infections, or infections at other sites. Moreover, the presence of increased bacterial load and Enterobacteriaceae in the pulmonary microbiome is associated with poor prognosis in patients with ARDS, which is consistent with our results. (Finegold et al., 1992; Cheung et al., 2019; Dahl et al., 2020). Therefore, an increase in *Bilophila* is a potential indicator for ARDSexp mortality.

ARDSp cases with complete data before and after treatment were selected for analysis and comparison. A cluster analysis of pathogenic bacteria showed that *Escherichia coli*, *Staphylococcus aureus,* and *Candida albicans* were more abundant in the ARDSp-poT-dead group than in the ARDSp-preT group. The ARDSp-poT-dead group exhibited increased positivity or

abundance of *Escherichia coli*, *Candida albicans*, *Staphylococcus aureus*, and *Klebsiella pneumoniae* compared with the ARDSp-poT-survival group. These results suggest that the pathogenic bacteria *Escherichia coli*, *Staphylococcus aureus*, and *Candida albicans* are potential risk factors for ARDSp death. In a retrospective analysis of the anti-infective regimens for these patients, we found that all regiments included treatment for *Escherichia coli* and *Staphylococcus aureus* and that 53.3% of non-surviving patients had received treatment for *Candida albicans*. These findings suggest that the increase in these pathogens was not strongly correlated with the anti-infective regimen, indicating a poor prognosis of ARDSp. *Escherichia*, Lachnospiraceae, *Cryptococcus,* and other enteric microbes and conditionally pathogenic bacteria may be associated with the risk of death in ARDSp, which is consistent with previous studies showing that the lung microbiome was enriched in enteric bacteria in mouse models of sepsis and patients with ARDS. Intestinal-specific microbes (*Bacteroides*) are common and abundant in the BALF of patients with ARDS and correlate with the intensity of systemic inflammation (Dickson et al., 2016), suggesting an interaction between the lower respiratory tract and the gastrointestinal tract. Dickson et al., 2020reported that increased bacterial load and gut-associated bacterial enrichment help predict the prognosis of patients with ARDS. In addition, the presence of *Candida* in septum cultures is associated with increased mortality in immunosuppressed patients (Pendleton et al., 2018). This observation may support the hypothesis that the increase in the pathogenic bacteria *Escherichia coli*, *Staphylococcus aureus,* and *Candida albicans*, and the increase in the intestinal microbiome, are important contributors to mortality in ARDSp; therefore, treatment targeting the intestinal microbiome in patients with ARDS should be considered.

We identified *Hydrobacter* as a possible protective factor against ARDSp. *Hydrobacter* was the most abundant in the control group and was associated with a better prognosis (**Figures 7E,F**). Little attention has been paid to the role of microbes such as *Hydrobacter* in the lower respiratory tract; however, the presence of a "normal microbiome" in the respiratory tract may be closely associated with the pathogenesis and development of ARDSp. *Hydrobacter* can be found in pure water (Eder et al., 2015); however, it remains unknown whether it belongs to the normal microbiota of the respiratory tract and whether it acts alone or in concert with other microbiomes. Extensive studies have demonstrated the positive effects of commensal microorganisms on human health (Nembrini et al., 2011; Fyhrquist et al., 2014; Zhang et al., 2021); therefore, microbial agents have potential clinical applications in maintaining lung function. Nevertheless, further studies are required to determine whether microbes such as *Hydrobacter* can serve as therapeutic targets for ARDS.

This study was limited by the absence of RNA sequencing data, missing information on RNA viruses, and microbial transcriptome alterations. The lack of 16s rRNA sequencing for microbial analyses hinders comparisons of the total bacterial abundance (using quantitative PCR), relative abundance (taxonomic composition of the specimen community), and community characteristics (e.g., diversity) for the whole sample. Greater efforts are needed to combine 16S rRNA and metagenomic sequencing to conduct more precise analyses of the community composition, diversity, evolutionary relationships, and gene functions. In addition, the results were potentially biased owing to inadequate sample quantity, the influence of retrospective analysis, and numerous clinical factors. Moreover, the application of BALF must consider the risk of sample contamination. However, in this study, all patients received an aseptically operated endotracheal tube; therefore, the risk of bacterial contamination was considered minimal. Antibiotic treatments may also affect alterations in the respiratory microbiome; however, previous studies have shown that in patients with traumatic ARDS, the antibiotic application is not significantly linked to the composition of the respiratory microbiome (Panzer et al., 2018). In addition, the patients were treated according to the criteria of the sepsis guidelines; thus, we believe that the impact of antibiotics was minimal. Further multicenter and prospective controlled studies are needed to recruit more patients and subdivide infections at different sites and with different pathogens to better understand the microbial profile of ARDS with different etiologies.

In general, the interaction or synergy between the lung microbiome and gut microbiome plays a regulatory role in the inflammatory immune response of the body. The association between microorganisms and their hosts is intricate and poorly understood. Their interaction should be balanced and mutually constrained, implying that no single microbe can affect the microbial function completely, while changes in any one part can influence the development of health and disease. To date, it remains unclear whether alterations in the microbial community in one region affect other regions or whether systemic effects produce a specific microbial community in a specific tissue (Lyon, 2017). Moreover, no systematic or large-scale studies have been undertaken; therefore, further longitudinal studies should be performed to correlate microorganisms with the severity of lung disease in humans and animals, thereby facilitating their application in etiological determination and disease control.

DATA AVAILABILITY STATEMENT

The clinical information presented in the study is included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Review Committee of Jiangmen Central Hospital (No. 2019-15). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

XZ, SZ, and YH conceived the project and designed the study. PZ, BL, and WL contributed to patient enrollment, experimental

design, manuscript writing, and data analysis equally. YC, ZW, YL, and JM collected and organized clinical information from patients. WL, MZ, WW, and ZM were responsible for sample handling and storage. JW and YZ collated the sequencing results and conducted diversity analysis.

FUNDING

This study was supported by grants from the National Natural Science Foundation of China (81802918), the China Postdoctoral Science Foundation Grant (2019M660206), the Science and Technology Project of Guangdong Province (2019A1515011565, 2018A030310007 and 2021A1515220180), the Medical Science Foundation of Guangdong Province (A2022089), the Science and Technology Project of Jiangmen (2020020500150004120, 2020030103140008978, 2019030102430012905, 2019-252-2-1, and 2018090106380023859), and the Medical Science Foundation of Jiangmen Central Hospital (J202004).

ACKNOWLEDGMENTS

The authors would like to thank Editage (www.editage.cn) for English language editing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.862570/full#supplementary-material

REFERENCES

Baker, S., and The, H. C. (2018). Recent Insights into Shigella. *Curr. Opin. Infect. Dis.* 31 (5), 449–454. doi:10.1097/QCO.0000000000000475

Bein, T., Grasso, S., Moerer, O., Quintel, M., Guerin, C., Deja, M., et al. (2016). The Standard of Care of Patients with ARDS: Ventilatory Settings and Rescue Therapies for Refractory Hypoxemia. *Intensive Care Med.* 42 (5), 699–711. doi:10.1007/s00134-016-4325-4

Cheung, S. G., Goldenthal, A. R., Uhlemann, A.-C., Mann, J. J., Miller, J. M., and Sublette, M. E. (2019). Systematic Review of Gut Microbiota and Major Depression. *Front. Psychiatry* 10, 34. doi:10.3389/fpsyt.2019.00034

Chinese Thoracic Society (2017). Chinese Expert Consensus on Bronchoalveolar Lavage Pathogen Detection for Infectious Diseases of the Lung (2017 Edition). *Chin. J. Tuberc. Respir. Dis.* 40, 8.

Dahl, W. J., Rivero Mendoza, D., and Lambert, J. M. (2020). Diet, Nutrients and the Microbiome. *Prog. Mol. Biol. Transl. Sci.* 171, 237–263. doi:10.1016/bs.pmbts.2020.04.006

Dickson, R. P., Erb-Downward, J. R., and Huffnagle, G. B. (2015). Homeostasis and its Disruption in the Lung Microbiome. *Am. J. Physiology-Lung Cell. Mol. Physiology* 309 (10), L1047–L1055. doi:10.1152/ajplung.00279.2015

Dickson, R. P., Schultz, M. J., van der Poll, T., Schouten, L. R., Falkowski, N. R., Luth, J. E., et al. (2020). Lung Microbiota Predict Clinical Outcomes in Critically Ill Patients. *Am. J. Respir. Crit. Care Med.* 201 (5), 555–563. doi:10.1164/rccm.201907-1487OC

Dickson, R. P., Singer, B. H., Newstead, M. W., Falkowski, N. R., Erb-Downward, J. R., Standiford, T. J., et al. (2016). Enrichment of the Lung Microbiome with Gut Bacteria in Sepsis and the Acute Respiratory Distress Syndrome. *Nat. Microbiol.* 1 (10), 16113. doi:10.1038/nmicrobiol.2016.113

Dickson, R. P. (2018). The Lung Microbiome and ARDS. It Is Time to Broaden the Model. *Am. J. Respir. Crit. Care Med.* 197 (5), 549–551. doi:10.1164/rccm.201710-2096ED

Dickson, R. P. (2016). The Microbiome and Critical Illness. *Lancet Respir. Med.* 4 (1), 59–72. doi:10.1016/S2213-2600(15)00427-0

Dyke, K. G. H. (2003). Staphylococcus Research. *Microbiol. Read.* 149 (Pt 10), 2697–2699. doi:10.1099/mic.0.26676-0

Eder, W., Peplies, J., Wanner, G., Frühling, A., and Verbarg, S. (2015). Hydrobacter Penzbergensis Gen. nov., Sp. nov., Isolated from Purified Water. *Int. J. Syst. Evol. Microbiol.* 65 (Pt 3), 920–926. doi:10.1099/ijs.0.000040

Finegold, S., Summanen, P., Hunt Gerardo, S., and Baron, E. (1992). Clinical Importance ofBilophila Wadsworthia. *Eur. J. Clin. Microbiol. Infect. Dis.* 11 (11), 1058–1063. doi:10.1007/BF01967799

Force, A. D. T., Ranieri, V. M., Rubenfeld, G. D., Thompson, B. T., Ferguson, N. D., Caldwell, E., et al. (2012). Acute Respiratory Distress Syndrome. *JAMA* 307 (23), 2526–2533. doi:10.1001/jama.2012.5669

Freestone, P. P., Hirst, R. A., Sandrini, S. M., Sharaff, F., Fry, H., Hyman, S., et al. (2012). *Pseudomonas aeruginosa* -Catecholamine Inotrope Interactions. *Chest* 142 (5), 1200–1210. doi:10.1378/chest.11-2614

Fyhrquist, N., Ruokolainen, L., Suomalainen, A., Lehtimäki, S., Veckman, V., Vendelin, J., et al. (2014). Acinetobacter Species in the Skin Microbiota Protect against Allergic Sensitization and Inflammation. *J. Allergy Clin. Immunol.* 134 (6), 1301–1309. e1311. doi:10.1016/j.jaci.2014.07.059

Griffiths, M. J. D., McAuley, D. F., Perkins, G. D., Barrett, N., Blackwood, B., Boyle, A., et al. (2019). Guidelines on the Management of Acute Respiratory Distress Syndrome. *BMJ Open Resp. Res.* 6 (1), e000420. doi:10.1136/bmjresp-2019-000420

Hilty, M., Burke, C., Pedro, H., Cardenas, P., Bush, A., Bossley, C., et al. (2010). Disordered Microbial Communities in Asthmatic Airways. *PLoS One* 5 (1), e8578. doi:10.1371/journal.pone.0008578

Hostacká, A. (2001). Klebsiella Species from the Viewpoint of Nosocomial Infections and Virulence Factors. *Epidemiol. Mikrobiol. Imunol.* 50 (2), 92–96. doi:10.3760/cma.j.issn.1001-0939

Hu, F., Guo, Y., Zhu, D., Wang, F., and Jiang, X. (2020). CHINET Surveillance of Bacterial Resistance across Tertiary Hospitals in 2019. *Chin. J. Infect. Chemother.* 20 (3), 481–491. doi:10.16718/j.1009-7708.2017.05.001

Huffnagle, G. B., and Dickson, R. P. (2015). The Bacterial Microbiota in Inflammatory Lung Diseases. *Clin. Immunol.* 159 (2), 177–182. doi:10.1016/j.clim.2015.05.022

Kyo, M., Nishioka, K., Nakaya, T., Kida, Y., Tanabe, Y., Ohshimo, S., et al. (2019). Unique Patterns of Lower Respiratory Tract Microbiota Are Associated with Inflammation and Hospital Mortality in Acute Respiratory Distress Syndrome. *Respir. Res.* 20 (1), 246. doi:10.1186/s12931-019-1203-y

Lyon, J. (2017). The Lung Microbiome: Key to Respiratory Ills? *JAMA* 317 (17), 1713–1714. doi:10.1001/jama.2017.3023

Matthay, M. A., Zemans, R. L., Zimmerman, G. A., Arabi, Y. M., Beitler, J. R., Mercat, A., et al. (2019). Acute Respiratory Distress Syndrome. *Nat. Rev. Dis. Prim.* 5 (1), 18. doi:10.1038/s41572-019-0069-0

Miao, Q., Ma, Y., Wang, Q., Pan, J., Zhang, Y., Jin, W., et al. (2018). Microbiological Diagnostic Performance of Metagenomic Next-Generation Sequencing when Applied to Clinical Practice. *Clin. Infect. Dis.* 67 (Suppl. l_2), S231–S240. doi:10.1093/cid/ciy693

Nembrini, C., Sichelstiel, A., Kisielow, J., Kurrer, M., Kopf, M., and Marsland, B. J. (2011). Bacterial-induced Protection against Allergic Inflammation through a Multicomponent Immunoregulatory Mechanism. *Thorax* 66 (9), 755–763. doi:10.1136/thx.2010.152512

O'Dwyer, D. N., Ashley, S. L., Gurczynski, S. J., Xia, M., Wilke, C., Falkowski, N. R., et al. (2019). Lung Microbiota Contribute to Pulmonary Inflammation and Disease Progression in Pulmonary Fibrosis. *Am. J. Respir. Crit. Care Med.* 199 (9), 1127–1138. doi:10.1164/rccm.201809-1650OC

Panzer, A. R., Lynch, S. V., Langelier, C., Christie, J. D., McCauley, K., Nelson, M., et al. (2018). Lung Microbiota Is Related to Smoking Status and to Development

of Acute Respiratory Distress Syndrome in Critically Ill Trauma Patients. *Am. J. Respir. Crit. Care Med.* 197 (5), 621–631. doi:10.1164/rccm.201702-0441OC

Pendleton, K. M., Dickson, R. P., Newton, D. W., Hoffman, T. C., Yanik, G. A., and Huffnagle, G. B. (2018). Respiratory Tract Colonization by Candida Species Portends Worse Outcomes in Immunocompromised Patients. *Clin. Pulm. Med.* 25 (6), 197–201. doi:10.1097/CPM.0000000000000279

Rhodes, A., Evans, L. E., Alhazzani, W., Levy, M. M., Antonelli, M., Ferrer, R., et al. (2017). Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. *ntensive Care Med.* 43 (3), 304–377. doi:10.1007/s00134-017-4683-6

Starr, M. P., and Chatterjee, A. K. (1972). The Genus Erwinia: Enterobacteria Pathogenic to Plants and Animals. *Annu. Rev. Microbiol.* 26, 389–426. doi:10.1146/annurev.mi.26.100172.002133

Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., et al. (2015). MetaPhlAn2 for Enhanced Metagenomic Taxonomic Profiling. *Nat. Methods* 12 (10), 902–903. doi:10.1038/nmeth.3589

Vacca, M., Celano, G., Calabrese, F. M., Portincasa, P., Gobbetti, M., and De Angelis, M. (2020). The Controversial Role of Human Gut Lachnospiraceae. *Microorganisms* 8 (4), 573. doi:10.3390/microorganisms8040573

Zhang, P., Chen, Y., Zheng, W., Wu, M., Wu, Z., Lu, Y., et al. (2021). Changes of Lung Microbiome of Acute Respiratory Distress Syndrome before and after Treatment under Open Airway. *Zhonghua Wei Zhong Bing Ji Jiu Yi Xue* 33 (9), 1063–1068. doi:10.3760/cma.j.cn121430-20210414-00558

Zhang, P., Chen, Y., Li, S., Li, C., Zhang, S., Zheng, W., et al. (2020). Metagenomic Next-Generation Sequencing for the Clinical Diagnosis and Prognosis of Acute Respiratory Distress Syndrome Caused by Severe Pneumonia: a Retrospective Study. *PeerJ* 8, e9623. doi:10.7717/peerj.9623

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 Molecular Biosciences

ORIGINAL RESEARCH
published: 28 June 2022
doi: 10.3389/fmolb.2022.884839

Tumor Microenvironment Heterogeneity-Based Score System Predicts Clinical Prognosis and Response to Immune Checkpoint Blockade in Multiple Colorectal Cancer Cohorts

Hufei Wang[1†], Zhi Li[2†], Suwen Ou[1†], Yanni Song[3†], Kangjia Luo[1], Zilong Guan[1], Lei Zhao[4], Rui Huang[1]* and Shan Yu[5]**

[1]Department of Colorectal Cancer Surgery, The Second Affiliated Hospital of Harbin Medical University, Harbin, China, [2]Department of Obstetrics and Gynecology, The Second Affiliated Hospital of Harbin Medical University, Harbin, China, [3]Department of Breast Surgery, Harbin Medical University Cancer Hospital, Harbin, China, [4]Department of Gastroenterology, The Second Affiliated Hospital of Harbin Medical University, Harbin, China, [5]Department of Pathology, The Second Affiliated Hospital of Harbin Medical University, Harbin, China

Despite immune checkpoint blockade (ICB) therapy contributed to significant advances in cancer therapy, only a small percentage of patients with colorectal cancer (CRC) respond to it. Identification of these patients will facilitate ICB application in CRC. In this study, we integrated multiple CRC cohorts (2,078 samples) to construct tumor microenvironment (TME) subtypes using TME indices calculated by CIBERSORT and ESTIMATE algorithms. Furthermore, a surrogate quantitative indicator, a tumor microenvironment immune gene (TMEIG) score system, was established using the key immune genes between TME clusters 1 and 2. The subsequent analysis demonstrated that TME subtypes and the TMEIG score system correlated with clinical outcomes of patients in multiple CRC cohorts and exhibited distinct immune statuses. Furthermore, Tumor Immune Dysfunction and Exclusion (TIDE) analysis indicated that patients with low TMEIG scores were more likely to benefit from ICB therapy. A study on two ICB cohorts (GSE78220 and IMvigor210) also validated that patients with low TMEIG scores exhibited higher ICB response rates and better prognoses after ICB treatment. The biomarker evaluation module on the TIDE website revealed that the TMEIG score was a robust predictive biomarker. Moreover, differential expression analysis, immunohistochemistry, qPCR experiments, and gene set prioritization module on the TIDE website demonstrated that the five genes that constitute the TMEIG score system (*SERPINE1*, *FABP4*, *SCG2*, *CALB2*, and *HOXC6*) were closely associated with tumorigenesis, immune cells, and ICB response indices. Finally, TMEIG scores could accurately predict the prognosis and ICB response of patients with CRC. *SERPINE1*, *FABP4*, *SCG2*, *CALB2*, and *HOXC6* might be potential targets related to ICB treatment. Furthermore, our study provided new insights into precision ICB therapy in CRC.

Keywords: tumor microenvironment, immune checkpoint therapy, colorectal cancer, prognosis, ICB response biomarkers

OPEN ACCESS

Edited by:
William C. Cho, QEH, Hong Kong SAR, China

Reviewed by:
Colt Egelston, City of Hope National Medical Center, United States
Iman Mamdouh Talaat, University of Sharjah, United Arab Emirates

***Correspondence:**
Lei Zhao
zhaolei181220@163.com
Rui Huang
huangrui2019@163.com
Shan Yu
yushan@hrbmu.edu.cn

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

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *27 February 2022*
Accepted: *16 May 2022*
Published: *28 June 2022*

Citation:
Wang H, Li Z, Ou S, Song Y, Luo K, Guan Z, Zhao L, Huang R and Yu S (2022) Tumor Microenvironment Heterogeneity-Based Score System Predicts Clinical Prognosis and Response to Immune Checkpoint Blockade in Multiple Colorectal Cancer Cohorts. Front. Mol. Biosci. 9:884839. doi: 10.3389/fmolb.2022.884839

INTRODUCTION

Colorectal cancer (CRC) is one of the most common malignant tumors globally, with high morbidity and mortality. Tumor immunotherapies involving immune checkpoint blockade (ICB) have contributed to significant advancements in the treatment of many tumors (Topalian et al., 2012), such as melanoma (Luke et al., 2017), bladder cancer (Pettenati and Ingersoll, 2018), and non–small cell lung cancer (Huang et al., 2021). However, most patients with CRC exhibit poor responses to immune checkpoint blockade (ICB) therapy. The biomarkers that predict the efficacy of ICB therapy include the expression of programmed death-ligand 1 (PD-L1) (Nishino et al., 2017), tumor mutation load (Snyder et al., 2014), mismatch repair deficiency (Le et al., 2015), and gut microbiota (Daillère et al., 2016; Routy et al., 2018). However, there is still a lack of effective tools to predict the ICB response in CRC, which impedes the application of ICB therapy in CRC. Therefore, there is an urgent need to establish effective and reliable tools for predicting response to ICB therapy and achieving precision therapy in patients with CRC.

The tumor microenvironment (TME) mainly contains tumor, immune and stromal cells, and small molecules (Vitale et al., 2019). TME of CRC exhibits remarkable heterogeneity (Zhang et al., 2020), which can cause variation in tumor biology, thus affecting the efficacy of multiple therapies (Casey et al., 2015; Wu and Dai, 2017), including chemotherapy (Hanoteau et al., 2019), radiotherapy (Yin et al., 2019; Sheng et al., 2020), and immune checkpoint therapy (Lei et al., 2020; Sheng et al., 2020). In addition, the TME can predict the prognosis of patients with CRC (Pagès et al., 2005). For example, a high M1:M2 density ratio in tumor stroma was associated with better cancer-specific survival (Väyrynen et al., 2021). Immune cells in the TME play critical roles in the efficacy of immunotherapy (Arce Vargas et al., 2018; Väyrynen et al., 2021). Patients with higher CD8 cells in the TME exhibit more favorable responses to ICB (Pagès et al., 2005). T cell-dendritic cell crosstalk facilitates successful anti–PD-1 immune therapy (Zhao et al., 2019). Therefore, studying TME heterogeneity will help reveal the biological characteristics of CRC, assist the implementation of precision therapy, and guide the application of ICB. However, the TME is an extremely complex system. It is critical to establish a simple surrogate gene model of the TME to predict the prognosis of patients and the efficacy of ICB therapy.

In the present study, we integrated transcriptome data of 1,175 patients with colorectal cancer from the GPL570 platform and then employed CIBERSORT, ESTIMATE, and ssGSEA algorithms to assess the characteristics of the TME. Based on TME heterogeneity, two TME subtypes (clusters 1 and 2) with different survival statuses were identified using a consensus clustering algorithm. Weighted gene co-expression network analysis (WGCNA), linear models for microarray data (LIMMA), and other bioinformatics analyses were used to identify the hub TME immune genes between subtypes. Then, patients were also divided into two tumor microenvironment immune gene (TMEIG) subtypes (clusters A and B) with different survival statuses according to the hub TME immune genes. Moreover, a robust prognostic scoring system (TMEIG score) was developed using these TME immune genes, which could effectively predict overall survival (OS), progression-free survival (PFS), and disease-specific survival (DSS) of patients with CRC. The prognostic TMEIG score system was also verified in multiple cohorts, such as TCGA-COAD. Notably, we validated that the TMEIG score could predict ICB response in multiple immunotherapy cohorts and is expected to guide the application of ICB in CRC.

MATERIALS AND METHODS

Data Source and Process

Ten data sets were downloaded from the public database, including clinical data and transcriptome data of 2,078 patients with CRC. GSE39582, GSE14333, GSE17536, GSE17537, and GSE72968 were all microarray data of the GPL570 platform and integrated as a training set (the combined GEO cohort). The CEL format data of the microarray was downloaded using the "GEOquery" package (Davis and Meltzer, 2007). ReadAffy function in the "affy" package was used to read data in CEL format (Gautier et al., 2004), background correction and standardization were carried out with RMA, and then the "SVA" package was used to remove batch effect among the data sets (Chakraborty et al., 2012). Probes corresponding to multiple genes were deleted, and the average expression level was taken when multiple probes corresponded to one gene. Clinical data and FPKM value (fragments per kilobase million) of TCGA-COAD were obtained from the UCSC website as an external validation set (Navarro Gonzalez et al., 2021). Then, FPKM was converted to TPM (transcripts per kilobase million) for subsequent analysis. GSE39582, GSE17536, and GSE17537 were used as the internal validation sets. The paired samples of GSE44076, GSE32323, GSE89076, and GSE113513 datasets have been reserved for verifying the gene expression level in the TMEIG score system. Detailed information on the data sets is shown in **Supplementary Table S1**.

Characteristics of the Tumor Microenvironment

CIBERSORT (Newman et al., 2015) and ESTIMATE (Becht et al., 2016) algorithms can infer the composition of 22 types of immune cells, immune score, stroma score, and tumor purity in the TME based on transcriptome data. In this study, transcriptome data from the combined GEO dataset (1,175 samples) and TCGA COAD dataset (471 tumors) were used for CIBERSORT and ESTIMATE analyses. After filtering out low expression immune cells, 15 types of immune cells were retained. ssGSEA is another algorithm for estimating immune cell

composition in solid tumor TME and can also be used to calculate adaptive and innate immune components of samples. In the present study, the adaptive and innate immune scores of each sample were obtained with the "GSVA" package. ssGSEA parameters were set as follows: method = "ssgsea," KCDF = "Gaussian."

Unsupervised Clustering

"ConsensusClusterPlus" is a re-sampling unsupervised clustering method to verify the rationality of clustering (Wilkerson and Hayes, 2010). In the combined GEO cohort, TME subtypes and TMEIG subtypes were obtained using "ConsensusClusterPlus" package. The parameters were set as follows: MaxK = 9, REPS = 1,000, pItem = 0.8, pFeature = 1, clusterAlg = "PAM," distance = "Euclidean," and seed = 1.

Weighted Gene Co-Expression Network Analysis

WGCNA analysis was used to identify gene modules most associated with traits (Langfelder and Horvath, 2008). Stromal score, immune score, estimate score, tumor purity, adaptive immune, innate immune, TME cluster 1, and TME cluster 2 were inputted as traits. The key parameters of WGCNA were set as follows: soft threshold power $\beta = 4$, scale-free R2 = 0.89. The relationship between modules and traits was analyzed using the Pearson correlation method. Gene significance (GS) and module membership (MM) are two important indicators in WGCNA analysis. GS is the correlation between the gene and trait. MM is defined as the correlation of the module eigengene and the gene expression profile. Genes with GS > 0.2 and MM > 0.8 are usually considered hub genes.

Functional Enrichment Analyses

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000) analyses were employed to explore the biological functions of the modules in WGCNA using the R package "clusterprofiler" (Yu et al., 2012). An adjusted *p*-value of less than 0.05 was regarded as statistically significant. In addition, Gene Set Enrichment Analysis (GSEA) was conducted (Subramanian et al., 2005). The gene sets "c2.cp.kegg.v6.2.symbols.gmt," "c5.all.v7.0.symbols.gmt," and "h.all.v7.2.symbols.gmt" on MSigDB website were chosen as the reference gene sets (Liberzon et al., 2015). The log fold change (FC) of differentially expressed genes between two groups was used as the input list for GSEA analysis. When analyzing the biological functions related to one gene, the Pearson correlation coefficient was used as the input list.

Construction and Validation of TMEIG Score

First, univariate Cox proportional hazards regression was employed to identify the prognostic genes using the "survival" R package. Genes with a *p*-value less than 0.05 were regarded as the candidates, input to least absolute shrinkage and selection operator (LASSO) regression (Friedman et al., 2010). After ten cross-validations, five prognostic genes and the corresponding coefficient remained when lambda = 0.0713387182. Then, TMEIG score was established as follows:

$$\textit{TMEIG score} = \sum_{i} \textit{Coefficient of } (i) \times \textit{Expression of gene } (i)$$

The regression coefficient of the gene was designated (i) in the LASSO regression model. The combined GEO cohort was used as a training set, whereas GSE39582, GSE17536, and GSE17537 were used as the internal validation sets. In addition, the TCGA COAD cohort served as the external validation set.

Survival Analysis

Only GSE39582, GSE17536, GSE17537, and GSE72968 had overall survival data in the combined GEO cohort (**Supplementary Table S2**). The survival time was converted to months, and samples with a survival time of less than 1 month were removed during survival analysis. Finally, 864 samples in the combined GEO cohort and 435 samples in the TCGA COAD cohort were used for survival analysis (**Supplementary Table S2**). According to the best cutoff value determined using the "survminer" package, the patients were divided into high and low expression groups. Log-rank test was employed to evaluate statistical significance. Kaplan–Meier (KM) plots were visualized using the "survminer" R package. The risk factors diagrams were visualized using the "ggrisk" R package.

Analysis of Mutation Data

The mutation data of TCGA COAD were downloaded from the TCGA website and analyzed using the "maftools" package (Mayakonda et al., 2018). The tumor mutation burden (TMB) was calculated using the following formula: (total mutation/total covered bases) × 106. The driver genes in somatic alterations were also identified using the "maftool" package.

ICB Response Prediction

Tumor Immune Dysfunction and Exclusion (TIDE) algorithm was employed to predict ICB response based on the gene expression related to T cell dysfunction (Dysfunction) and T cell exclusion (Exclusion). The lower the TIDE score is reportedly associated with a better immunotherapy response (Jiang et al., 2018). Furthermore, the scores of cancer-associated fibroblasts (CAF), Dysfunction, Exclusion, M2 macrophages (M2), myeloid-derived suppressor cells (MDSC), and TIDE were obtained from the TIDE website. The IMvigor210 cohort is a large cohort of patients with metastatic urothelial cancer under anti–PD-L1 therapy (atezolizumab), which can be downloaded from the Creative Commons 3.0 license. GSE78220 is an anti–PD-1 therapy cohort containing mRNA expression data from pre-treatment melanomas. The two cohorts were used to validate the predictive power of the TMEIG score for ICB response.

Cell Culture

The human CRC cell lines SW620, RKO, HCT116, HT29, and NCM460 (ATCC) were cultured in RPMI-1640 medium (Gibco, United States) supplemented with 10% fetal bovine serum (FBS, Biological Industries, United States) at 37°C in a humidified 5% CO_2 atmosphere.

RNA Extraction and Quantitative Real-Time PCR

Total RNAs of cell lines were extracted by TRIzol reagent (Invitrogen, United States) and then was reversely transcribed as cDNA *via* PrimeScript™ RT Master Mix (Takara, Japan). Quantitive real-time PCR was performed using PowerUp™ SYBR™ Green Master Mix (Applied Biosystems, United States) in the StepOne™ Real-Time PCR System (Applied Biosystems). Each reaction was tested in triplicate. ACTB was used as the internal reference, and the $2(^{-\Delta\Delta CT})$ method was used for calculating the relative mRNA expression. The following primer sets were used:

Human *FABP4*: Forward: 5′-GGGCCAGGAATTTGACGAAG-3′, Reverse: 5′-TCGTGGAAGTGACGCCTTTC-3′; Human *SCG2*: Forward: 5′-GTGAAGCGAGTTCCTGGTCA-3′, Reverse: 5′-ATG CTCTTTGATGGCCTGCT-3′; Human *CALB2*: Forward: 5′-GAA GGCAAGGAAAGGCTCTGG-3′, Reverse: 5′-GCCATCTCGATT TTCCCATCTG-3′; Human *SERPINE1*: Forward: 5′-CCTGGTTCT GCCCAAGTTCT-3′, Reverse: 5′-CCATGCGGGCTGAGACTA TG-3′; Human *HOXC6*: Forward: 5′-CACTAACCCTTCCTT ATCCTGCC-3′, Reverse: 5′-TCATAGGCGGTGGAATTGAGG-3′; Human *ACTB*: Forward: 5′-GATTCCTATGTGGGCGACGA-3′, Reverse: 5′-AGGTCTCAAACATGATCTGGGT-3′.

Immunohistochemistry

For the IHC experiment, we collected 16 pairs of CRC tissue (cancer and adjacent normal tissue) from patients who received surgery at the Department of Colorectal Cancer Surgery, the Second Affiliated Hospital of Harbin Medical University (Harbin, China) between January 2014 and December 2020. Ethics approval was also granted by the Ethics Committee of Harbin Medical University (No. KY 2022-063). The primary antibodies used in IHC were as follows: anti-FABP4 (Proteintech, #12802-1-AP, 1:200 dilution), anti-SCG2 (Proteintech, #20357-1-AP, 1:200 dilution), anti-CALB2 (Proteintech, #12278-1-AP, 1:200 dilution), anti-HOXC6 (Affinity, #DF3078, 1:150 dilution), and anti-PAI1 (SERPINE1) (Affinity, #AF5176, 1:200 dilution). Paraffin sections were incubated with primary antibodies at 4°C overnight, followed by treatment with HRP-conjugated secondary antibodies at 37°C temperature for 60 min following PBS rinse. Then, tissues were counter-stained with hematoxylin and further treated with DAB for 2 min. The IHC results were independently analyzed by two experienced pathologists. A staining scoring system was evaluated by both staining intensity (negative = 0, weak = 1, and strong = 2) and staining area (<5% = 0, 5%–25% = 1, 25%–50% = 2, 50%–75% = 3, and >75% = 4). The staining intensity score was computed, and the score of the staining area was the final staining score. A total score of ≤3 was considered a weak expression. A total score of >3 was considered a strong expression. The details of IHC performance and scoring system are described in **Supplementary Table S3**.

Statistical Analysis

Heat maps were visualized with the "ComplexHeatmap" package (Gu et al., 2016). The "ggplot2" package was used to visualize boxplots, scatter plots, and Sankey plots. The log-rank test and Pearson method were used for KM survival and correlation analyses, respectively. The difference between the two groups was tested by the Wilcox test. It should be noted that * represented a *p*-value less than 0.05, ** represented a *p*-value less than 0.01, *** represented a *p*-value less than 0.001, and **** represented a *p*-value less than 0.0001. All analyses were performed in R 4.0.3.

RESULTS

Depicting the Heterogeneity of the Tumor Microenvironment in a Large CRC Cohort

The flow diagram describes the construction of TME subtypes and the TMEIG score in CRC (**Figure 1**). We integrated microarray data of 1,175 patients with CRC from the GPL570 platform and then used the combat function of the "SVA" package to remove batch effects. The principal component analysis (PCA) diagrams of five cohorts before and after batch effect removal are shown in **Figures 2A,B**. The results indicated that the batch effect was negated, and the combined cohort could be used for subsequent analysis. To fully dissect the heterogeneity of the TME in patients with CRC, the CIBERSORT algorithm was used to assess the proportion of immune cells in the TME. Macrophages and mast cells were the most abundant immune populations in the combined GEO cohort, followed by memory resting CD4 and CD8 T cells. **Figures 2C,D** shows the proportion of immune cells in each patient, which partly reflects the heterogeneity of immune cells in the TME. A total of 15 types of immune cells were retained after eliminating low expression cells (such as memory B cells, CD4 naive T cells, gamma delta T cells, activated NK cells, monocytes, resting mast cells, and eosinophils). The detailed results of the CIBERSORT analysis are shown in **Supplementary Table S4**. Then, the ESTIMATE algorithm was used to calculate patients' immune scores and stromal scores. Collectively, CIBERSORT and ESTIMATE algorithms were used to comprehensively describe the correlations among the immune cells, immune score, and stromal score in the tumor microenvironment of patients with CRC (**Figure 2E**). Resting NK cells were inversely correlated with M0/M1/M2 macrophages (correlation values = −0.09, correlation values = −0.28, correlation values = −0.26; p-value < 0.05; **Supplementary Table S4**). Furthermore, CD8 T cells were negatively related to M0 macrophages (correlation values = −0.31, p-value < 0.05) and positively related to M1 macrophages (correlation values = 0.16, p-value < 0.05).

Tumor Microenvironment Cluster 2 has Better Survival and Exhibits a Different Immune State

Based on these quantitative indicators describing the TME, we conducted unsupervised clustering in these 1,175 patients using the "ConsesusClusterPlus" package. As shown in **Supplementary Figure S1**, the clustering result was the most stable when $K = 2$. The PCA plot also demonstrated significant differences between the two clusters (**Figure 3A**). Then, survival analysis was employed to compare the prognosis between the two TME clusters (**Supplementary Table S5**). The OS in TME cluster 2 was significantly better than that in TME cluster 1 (**Figure 3B** log-rank test, $p = 0.047$). Furthermore, we explored 11 critical biological

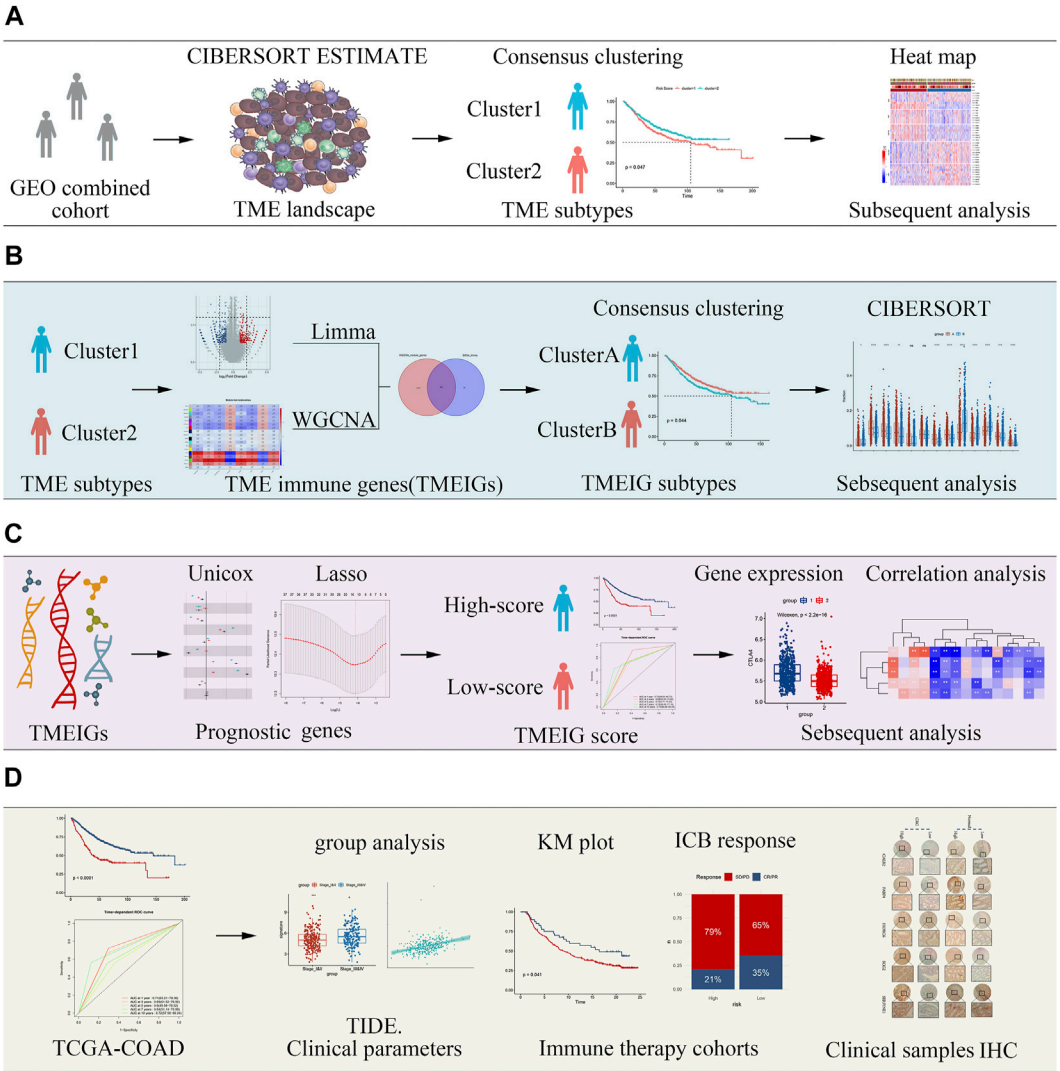


FIGURE 1 | Flow diagram of the study describing the process by which tumor microenvironment (TME) subtypes and the key tumor microenvironment immune gene (TMEIG) scoring system were identified. **(A)** Identification of TME subtypes. **(B)** Construction of TMEIG subtypes. **(C)** Establishment of TMEIG score. **(D)** Validation of TMEIG score in immune therapy cohorts.

gene signatures between the two TME subtypes using a heat map (Mariathasan et al., 2018). The results indicated that cell cycle genes and DNA damage repair (DDR) genes were markedly decreased, and angiogenesis (Angio) markers, TGFβ receptor and ligand (TGFβ), antigen-processing machinery (APM), and F-TBRS genes were significantly increased in TME cluster 1 as compared to TME cluster 2 (**Figure 3C**). In addition, CD8 Teff cells and immune checkpoint signatures (ICI) were highly expressed in TME cluster 1 (**Figures 3C–E**). The low expression of cell cycle-associated genes may indicate that tumor cells in TME cluster 1 were in a dormant phase and were not easily cleared by the immune system. A comparison of the immune score and stromal score revealed that the immune score and stromal score of TME cluster 1 were higher than those of TME cluster 2 (**Figure 3D**). CIBERSORT analysis demonstrated that immunosuppressive cells (M0, M1, and M2) were significantly reduced, and immunoreactive cells (CD8 T cell, CD4 memory resting T cells, resting dendritic cells, and activated dendritic cells) were significantly increased in the TME cluster 2 (**Figure 3F**). Furthermore, the GSEA results indicated that immune-related functions (activation of immune response, positive regulation of cytokine production, cytokine–cytokine receptor interaction, and IL6-JAK-STAT3 signaling) significantly varied between TME clusters 1 and 2 (**Figures 3G–I**; **Supplementary Table S6**).

Identification of Key Tumor Microenvironment Immune Genes Between Tumor Microenvironment Subtype

To identify key gene modules in TME clusters 1 and 2, WGCNA analysis was employed. The WGCNA analysis processes are shown in **Supplementary Figures S2A–E**. Adaptive immunity and innate immunity were derived from

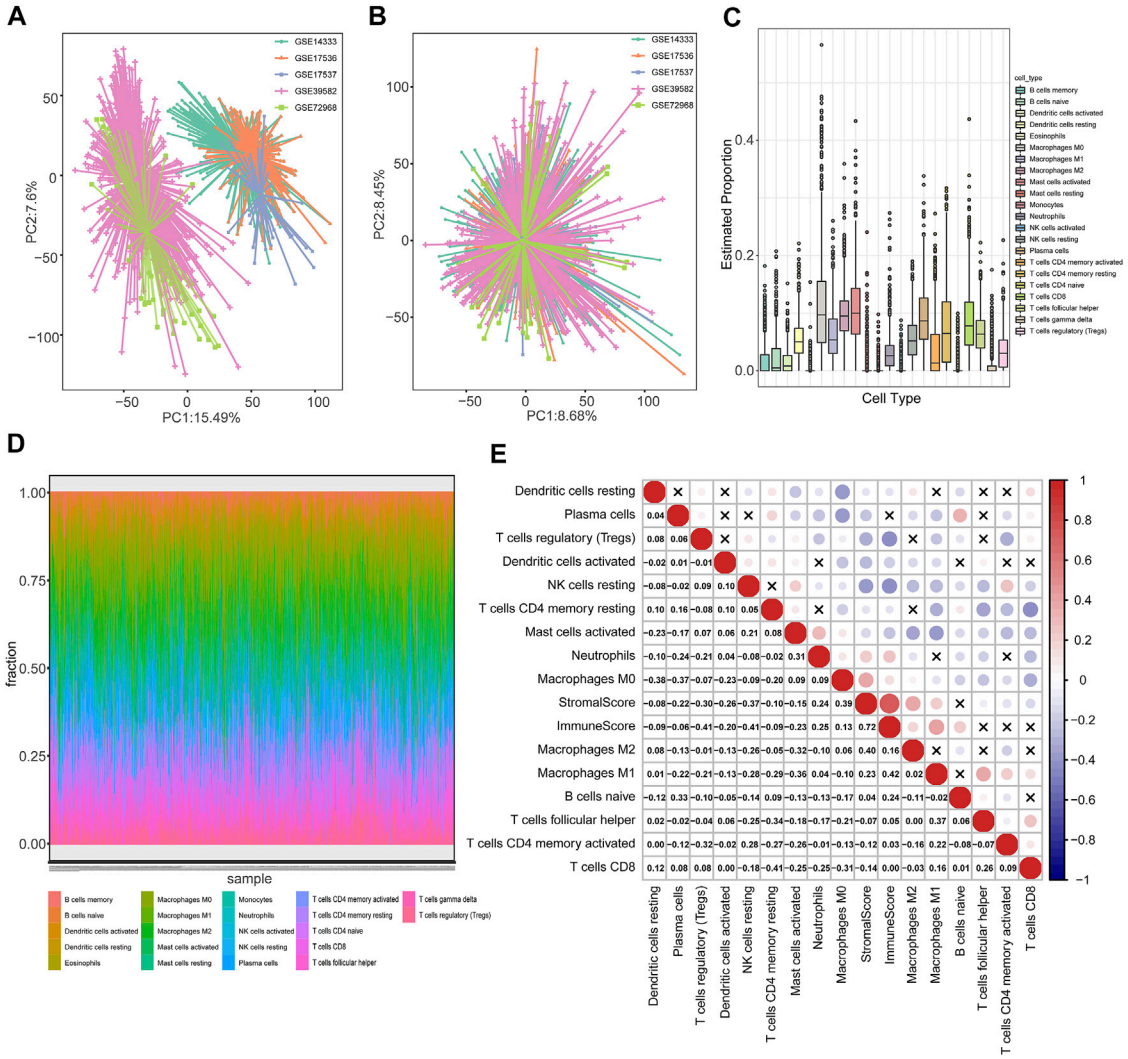


FIGURE 2 | Heterogeneity of the TME in patients with colorectal cancer. **(A,B)** PCA diagrams of five cohorts (GSE39582, GSE14333, GSE17536, GSE17537, and GSE72968) before and after batch effect removal. **(C)** Boxplot cells, immune score, and stromal score (calculated using the ESTIMATE algorithm) in TME of patients with CRC were analyzed using the "corrplot" R package. Red and blue colors represent positive and negative, depicting the proportion of 22 types of immune cells in TME estimated using the CIBERSORT algorithm. **(D)** The distribution of 22 types of immune cells in each patient. **(E)** Pearson correlation between immune correlations, respectively. The correlation *p*-values were less than 0.05 in all cases except those marked with "x" symbols.

ssGSEA analysis. Stromal score, immune score, estimate score, tumor purity, adaptive immunity, innate immunity, and TME clusters 1 and 2 were used as traits. The heat map of module–trait relationships is shown in **Figure 4A**. Results indicated that blue (cor = 0.79, p < 1e−200), brown (cor = 0.53, p = 2.2e−63), and green (cor = 0.98, p < 1e−200) modules displayed the high correlations with adaptive immunity (**Figures 4B–D**).

Thus, the blue, brown, and green modules were identified as the key modules. We performed GO and KEGG analyses to explore the biological functions of genes within the key modules. As shown in **Supplementary Figures S2F–H**, GO and KEGG terms were closely related to the immune function, such as regulation of immune system process, cytokine production, TNF signaling pathway, TNF superfamily cytokine production, and TNF superfamily cytokine production. In the three modules, 223 genes with GS > 0.2 and MM > 0.8 were identified as candidate genes. We used the "Limma" package to obtain the differentially expressed genes (DEGs) between TME clusters 1 and 2, and the results are shown in the volcano map (**Figure 4E**). p-value < 0.05 and logFC >0.5 were set as parameters, and 719 DEGs were obtained (698 upregulated and 21 downregulated). Since there were very few downregulated genes, we mainly used the upregulated genes to compare with the candidate genes of WGCNA. A total of 202 TMEIGs were eventually identified after comparing candidate genes with upregulated genes (**Figure 4F**).

TMEIG Cluster a Has a Better Prognosis

A total of 202 TMEIGs were again used for unsupervised clustering in the combined GEO cohort, and the clustering

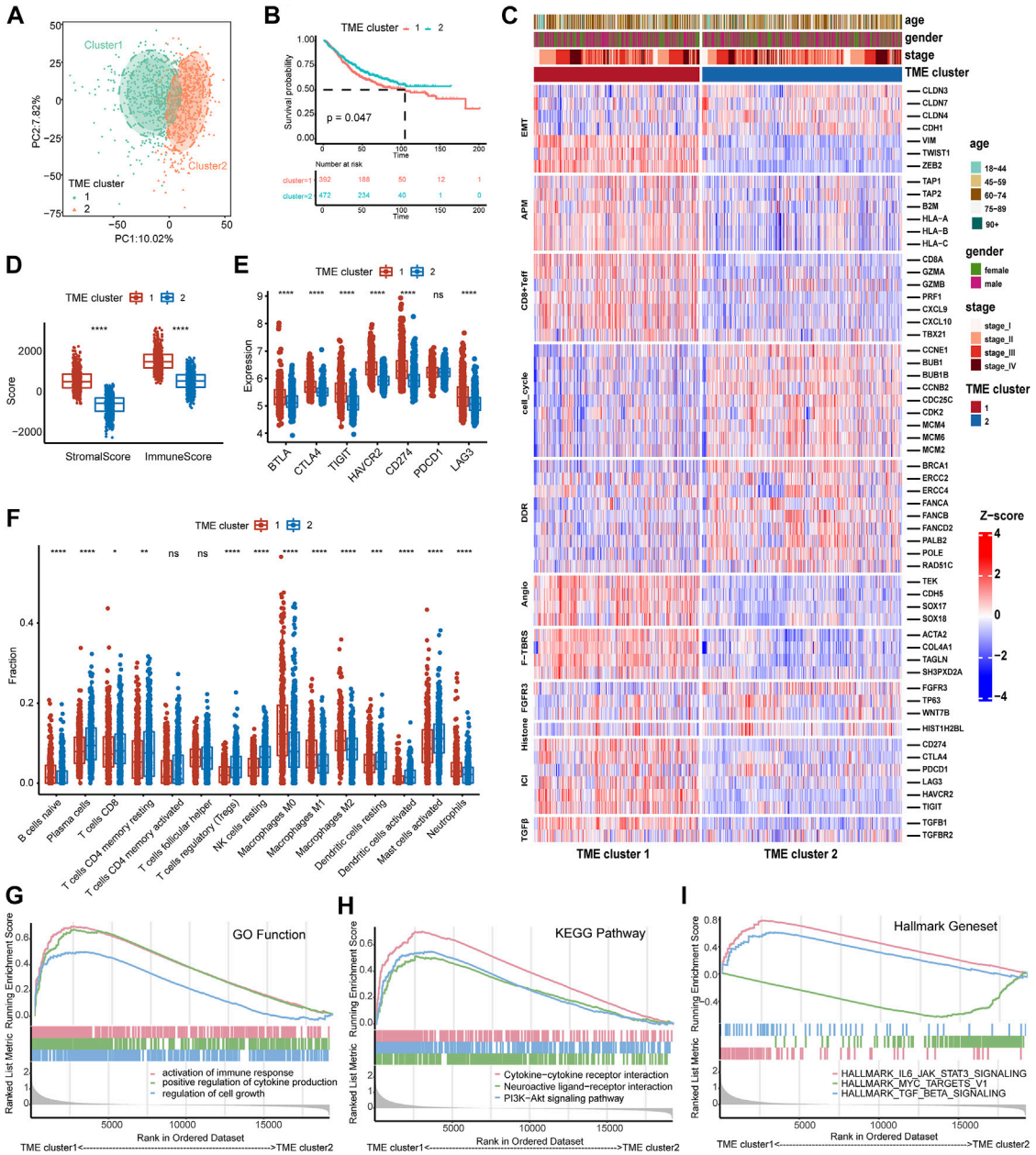


FIGURE 3 | Identification of the TME subtypes and analysis of biological functions. **(A)** PCA analysis demonstrates that the TME subtypes display distinct gene expression signatures. **(B)** The survival analysis of TME subtypes in the combined GEO cohort. **(C)** Relationship between TME subtypes and 11 critical biological pathways. Rows of the heat map represent gene expression grouped by pathway. Red and blue colors represent high and low expression, respectively. EMT (epithelial to mesenchymal transition), Angio (angiogenesis), ICI (immune checkpoint genes), DDR (DNA damage-repair), and APM (antigen-processing machinery). **(D)** The boxplot of the immune score and stromal score between TME subtypes calculated by ESTIMATE analysis. **(E)** The boxplot of seven immune checkpoint genes between TME subtypes. **(F)** The distribution of 15 types of immune cells between TME subtypes estimated by CIBERSORT analysis. **(G–I)** GSEA analysis of GO function, KEGG pathway, and Hallmark gene set of both TME subtypes. The difference between the two groups was assessed using the Wilcox test. The log-rank test was used for KM survival analysis.

process is shown in **Supplementary Figures S3A–E**. The clustering result was most stable when $k = 2$. PCA plot also revealed significant differences between the TMEIG subtypes (**Figure 4G**; **Supplementary Table S7**). The KM plots revealed that patients in TMEIG cluster A exhibited better OS (**Figure 4H**, log-rank test, $p = 0.047$). Similarly, the heat map of tumor-related pathways showed that cell cycle and DDR signatures were significantly decreased in TMEIG cluster B. Angio, transforming growth factor-beta (TGFβ), antigen processing machinery (APM), TGF-beta response signatures (TBRS) of fibroblasts (F-TBRS), and immune checkpoint signatures were increased in TMEIG cluster B as compared to that in cluster A

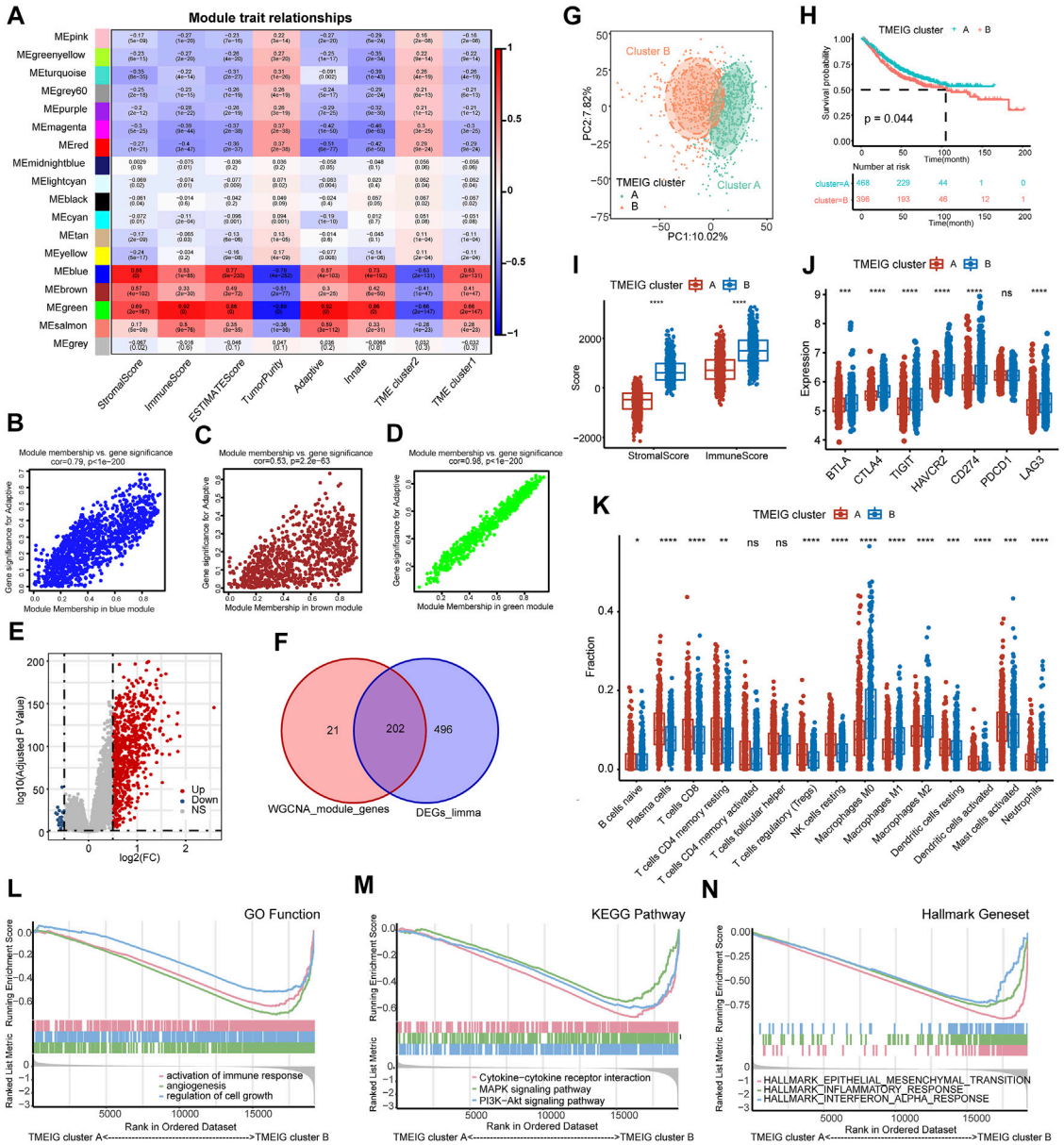


FIGURE 4 | Identification of TMEIG and TMEIG subtypes. **(A)** Heatmap of module trait relationships in the combined GEO cohort. Each row contains the corresponding correlation values and *p*-value. Red and blue colors represent the positive and negative correlations, respectively. **(B–D)** Scatter plots of the correlation between module eigengenes and adaptive immune in blue, brown, and green modules. **(E)** The volcano plot of the differentially expressed genes between TME clusters 1 and 2. **(F)** The intersection genes of WGCNA module genes and differentially expressed genes were considered the TMEIGs. **(G)** PCA plot demonstrates that the TMEIG subtypes display distinct gene expression patterns. **(H)** The survival analysis of TMEIG subtypes in the combined GEO cohort. **(I)** The boxplot of the immune score and stromal score between TMEIG subtypes. **(J)** The boxplot of seven immune checkpoint genes between TMEIG subtypes. **(K)** The distribution of 15 types of immune cells between TMEIG subtypes. **(L–N)** GSEA analysis of GO function, KEGG pathway, and Hallmark gene set between TMEIG subtypes. The difference between the two groups was assessed using the Wilcox test. The log-rank test was used for KM survival analysis.

(**Supplementary Figure S3J**). ESTIMATE analysis showed that TMEIG cluster B had higher immune and stromal scores than TMEIG cluster A (**Figure 4I**). CIBERSORT analysis showed that immunosuppressive cells (M0/M1/M2 macrophages) increased significantly in TMEIG cluster B, whereas immunoreactive cells (CD8 T cell, CD4 memory resting T cells, resting dendritic cells, and activated dendritic cells) decreased significantly compared to TMEIG cluster A (**Figure 4K**). GSEA results indicated that immune-related functions (activation of immune response, cytokine-cytokine receptor interaction, and inflammatory response) significantly varied between TMEIG cluster A and TMEIG cluster B (**Figures 4L–N**). These results demonstrated that TMEIG subtype clustering accurately reflected the differences between TME subtypes.

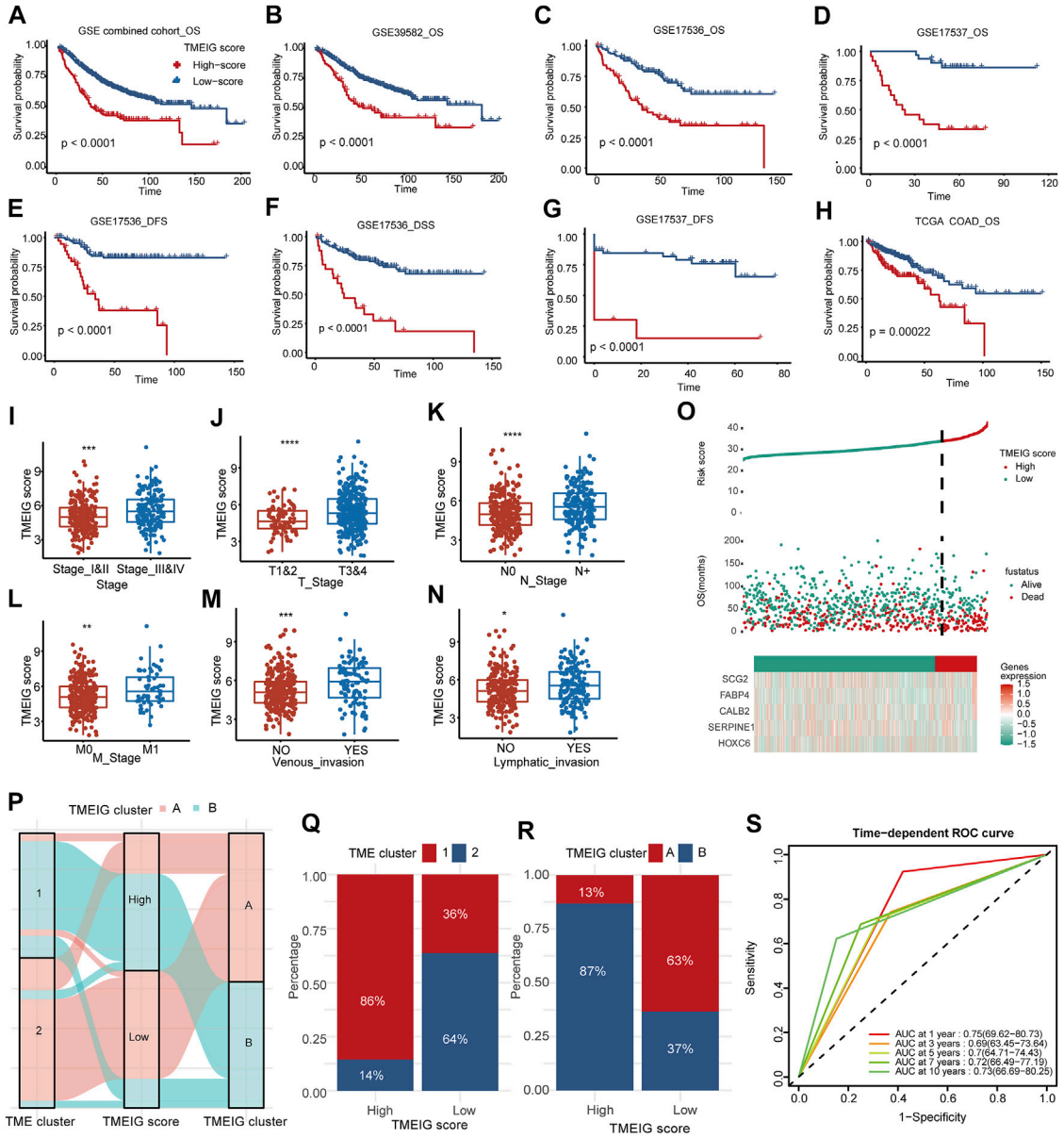


FIGURE 5 | Clinical significance of TMEIG score. **(A–H)** The survival analysis of TMEIG score in multiple colorectal cancer cohorts. OS represents overall survival, DFS represents disease-free survival, and DSS represents disease-specific survival. **(I–N)** The relationship between clinicopathological parameters and TMEIG score in TCGA COAD. The TMEIG score was transformed to log2 format for analysis. Clinicopathological parameters are collated from the UCSC website. **(O)** The risk factor diagrams of the combined GEO cohort. **(P)** The Sankey diagram revealed the correlation between the TME cluster, TMEIG score, and TMEIG cluster in the combined GEO cohort. **(Q,R)** The stacked histogram exhibits the distribution of the TME cluster and TMEIG cluster between high and low TMEIG score groups. **(S)** ROC plot shows the predictive value of the TMEIG score combined with age, sex, M stage, and TNM stage in the GSE39582 cohort using stepwise Cox regression. The difference between the two groups was assessed using the Wilcox test. The log-rank test was used for KM survival analysis.

Patients With High Tumor Microenvironment Immune Gene Scores Have a Poorer Prognosis in Multiple CRC Cohorts

Gene signature is a simple and effective model widely used in clinical practice (Paik et al., 2004; van 't Veer et al., 2002; Parker et al., 2009). To further facilitate the application of TME subtypes in CRC, we intended to construct a TMEIG score system. First, 662 genes were obtained by comparing 2,799 genes in blue, brown, and green modules with 698 DEGs (**Supplementary Figure S4A**). Univariate Cox regression analysis was performed in the combined GEO cohort and TCGA COAD cohort. With a *p*-value less than 0.05, 287 and 47 prognostic genes were obtained from the combined GEO and TCGA COAD cohorts, respectively. Then, Lasso regression was used to identify 27 common

genes in the combined GEO cohort (**Supplementary Figure S4B**). Details of the Lasso regression are shown in **Supplementary Figures S4C and D**. After cross-validating the results ten times, five genes (*SERPINE1*, *FABP4*, *SCG2*, *CALB2*, and *HOXC6*) and their corresponding lambda coefficients were obtained when lambda = 0.0713387182. The TMEIG score was constructed based on the expression and coefficient of the five genes as described in the methods. According to the optimal cutoff value, patients were divided based on whether their TMEIG scores were high or low (**Supplementary Table S8**). OS analysis suggested that the high TMEIG scores were associated with poor prognosis in patients with CRC (**Figure 5A**, log-rank test, $p < 0.0001$).

Internal validation cohorts indicated that the OS of the patients with high TMEIG scores was poorer than those with low scores (**Figures 5B–D**, GSE39582, GSE17536, GSE17537, log-rank test, $p < 0.0001$). In addition, the PFS and DSS in the low TMEIG score group were superior to those of the high TMEIG score group (**Figures 5E–G**, GSE17536 DFS, GSE17536 DSS, GSE17537 DFS, log-rank test, $p < 0.0001$). Then, the TCGA COAD cohort also revealed that the overall survival of the high TMEIG score group was poorer (**Figure 5H**; **Supplementary Table S8**). When analyzing the relationship between clinicopathological parameters and the TMEIG score, we observed that the scores of patients exhibiting stage III & IV, T 3 & 4, lymphatic invasion, and venous invasion were significantly higher (**Figures 5I–N**; **Supplementary Table S9**), suggesting the high TMEIG score was associated with poor clinical prognosis. Furthermore, the risk factor diagrams of the combined GEO and TCGA COAD cohorts indicated that the high TMEIG score group had significantly higher mortality than the low TMEIG score group (**Figure 5O**; **Supplementary Figure S4E**). All the results demonstrated poor prognoses in patients with high scores. The distribution of patients in the TME clusters, TMEIG clusters, and TMEIG score groups are shown in **Figure 5P**, which indicates that most patients in TME cluster 1 belonged to TMEIG cluster B and the high TMEIG score group. Consistent with the results, the high TMEIG score group had a higher proportion of patients in TME cluster 1 and TMEIG cluster B (**Figures 5Q,R**). Moreover, patients in TME cluster 1 and TMEIG cluster B exhibited higher TMEIG scores than that in TME cluster 2 and TMEIG cluster A (**Supplementary Figures S4F,G**). This evidence demonstrated that the TMEIG score could effectively surrogate TME and TMEIG subtypes. Furthermore, TMEIG score, age, sex, and T, N, M, and TNM stages were included in stepwise Cox regression in the GSE39582 cohort, which possessed comprehensive clinical information. Results indicated that TMEIG score combined with age, sex, M stage, and TNM stage exhibited the best predictive power (**Supplementary Figure 5S**, AUC: 0.75, 0.69, 0.7, 0.72, and 0.73 at 1, 3, 5, 7, and 10 years, respectively), and was validated in the TCGA COAD cohort (**Supplementary Figure S4H**).

Patients With High Tumor Microenvironment Immune Scores Are More Likely to Benefit From Immune Checkpoint Blockers

To evaluate whether the TMEIG score could predict the efficacy of ICB treatment, we analyzed the expression of crucial immune checkpoint molecules between high and low TMEIG score groups. The results showed that the expression of immune checkpoint molecules (*PD-1* [*PDCD1*], *PD-L1* [*CD274*], cytotoxic T-lymphocyte associated protein 4 [*CTLA4*], B- and T-lymphocyte attenuator [*BTLA*], T cell immunoglobulin and ITIM domain [*TIGIT*], hepatitis A virus cellular receptor 2 [*HAVCR2*], and lymphocyte-activation gene 3 [*LAG3*]) was significantly higher in the high score group (**Figure 6A**). Patients with CRC who have microsatellite instability-high (MSI-H) tumors are more likely to benefit from immune checkpoint inhibitors than patients with microsatellite stable (MSS)/MSI-low (MSI-L). To explore the relationship between TMEIG score and MSI status, the MSI status of TCGA COAD patients was downloaded from the supplements of previous studies focusing on MSI detection (**Supplementary Table S10**). There were 72 patients identified as MSI-H and 355 identified as MSI-L/MSS in TCGA COAD determined by MSI-PCR. We then analyzed whether the TMEIG score had prognostic value across MSI-H and MSI-L/MSS subgroups. KM plots demonstrated that patients with high TMEIG scores exhibited poor overall survival in MSI-H and MSI-L/MSS subgroups (**Supplementary Figures S5A,B**). Further analysis showed that patients with MSI-H possessed higher TMEIG scores (**Figure 6B**, $p = 6.9e{-}06$). Next, we explored the proportion of patients with MSI-H and MSI-L/MSS in high and low TMEIG score groups. We observed more MSI-H CRC patients in the high TMEIG score group (**Supplementary Figure S5C**, High: 26%, Low: 14%). Previous studies reported that TMB was a predictor of the efficacy of ICB therapy. When exploring the TMB of patients from the TCGA COAD cohort, results indicated no statistical difference between high and low TMEIG score groups (**Figure 6C**). In addition, the top six driver genes with the highest alteration frequency were further analyzed. The alteration frequency of *APC*, *TP53*, *TTN*, *KRAS*, *PIK3CA*, and *MUC16* in high and low TMEIG score groups are displayed in **Supplementary Figure S5D**. The ESTIMATE algorithm revealed that immune, stromal, and estimate scores were significantly higher in the high TMEIG score group (**Figure 6D**). Furthermore, pathway heat maps demonstrated that EMT, angiogenesis (vimentin [*VIM*], Twist-related protein 1 [*TWIST1*], zinc finger E-box binding homeobox 1 and 2 [*ZEB1* and *ZEB2*]), and T-cell factor-beta (TCFβ) signatures were significantly more activated in the high TMEIG score group. In contrast, cell cycle and DDR signatures were highly expressed in the low TMEIG score group (**Figure 6E**). These results indicated significant differences in the TME and biological pathways between high and low TMEIG score groups. To dissect the relationship between TMEIG score and ICB response, we used the TIDE algorithm to predict ICB response based on transcriptome signatures. The TIDE algorithm (**Figure 6F**) showed that the TMEIG score was positively correlated with CAF ($R = 0.68$, $p < 2.2e{-}16$), Dysfunction ($R = 0.37$, $p < 1.7e{-}15$), and Exclusion ($R = 0.14$, $p < 0.0036$), and negatively correlated with M2 macrophages ($R = 0.51$, $p < 2.2e{-}16$) and MDSC ($R = -0.3$, $p < 2.7e{-}10$). In addition, there was a strong

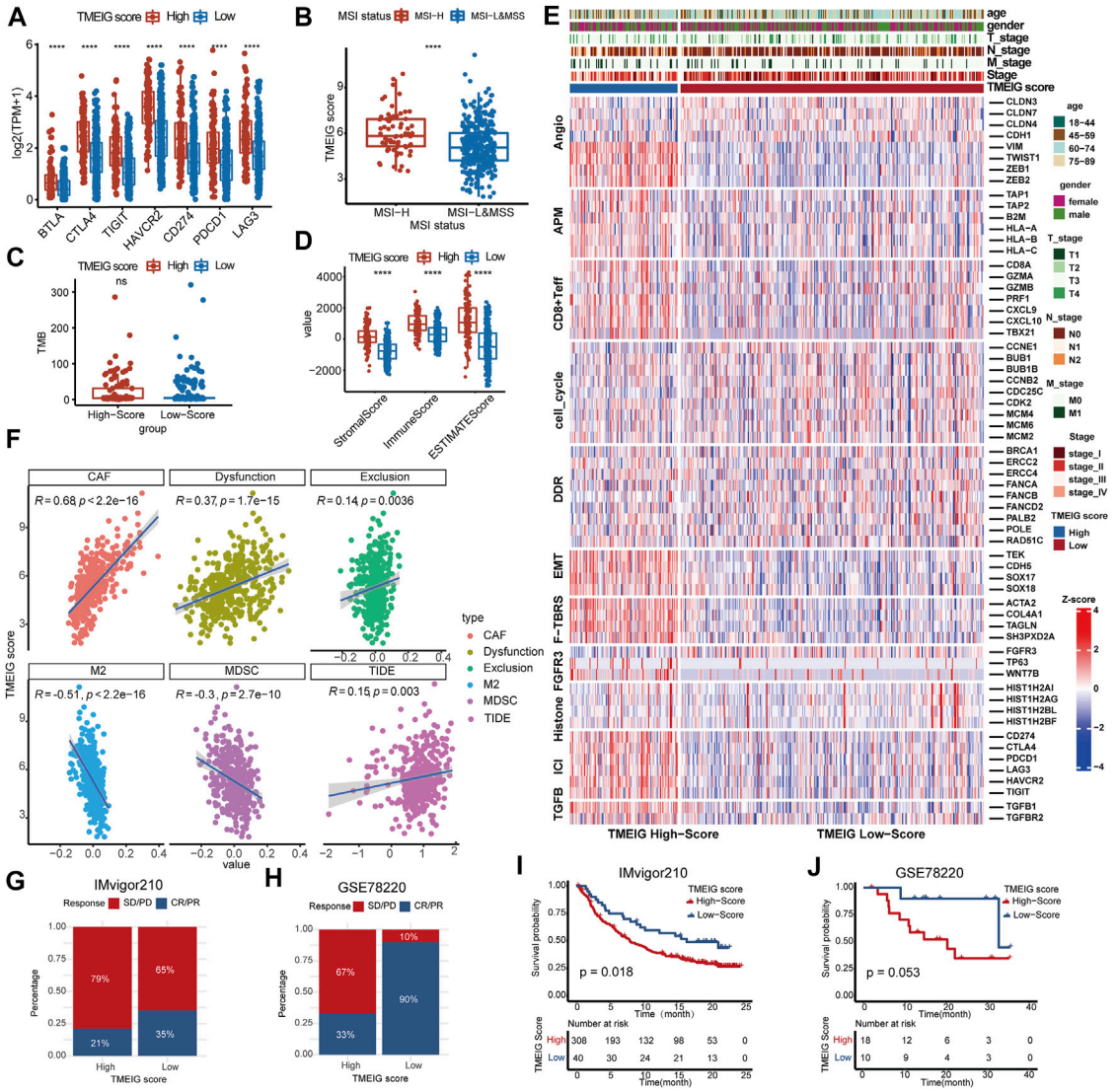


FIGURE 6 | The correlation between TMEIG score and ICB response. **(A)** The boxplot of seven immune checkpoint genes between the high and low TMEIG score groups in the TCGA COAD cohort. **(B)** The TMEIG score between MSI-H and MSI-L/MSS subgroups. The boxplot showed that patients with MSI-H possessed higher TMEIG scores than MSI-L/MSS (Wilcox test, p = 6.9e−06). The TMEIG score was transformed to log2 format for analysis. **(C)** TMB difference in the high and low TMEIG score groups in the TCGA COAD cohort. **(D)** Relationship between TME subtypes and 11 critical biological pathways in the TCGA COAD cohort. Rows of the heat map represent gene expression grouped by pathway. **(F)** The Pearson correlation analysis between TMEIG score and tumor-associated fibroblast (CAF), T cell dysfunction (Dysfunction), T cell exclusion (Exclusion), M2 macrophage (M2), myeloid-derived suppressor cell (MDSCs), and TIDE score. The TMEIG score was transformed to log2 format for analysis. **(G,H)** The stacked histogram exhibits the distribution of ICB response rates between high and low TMEIG score groups in IMvigor210 and GSE78220 cohorts. The blue (CR/PR) indicates patients who responded to ICB, whereas the red (SD/PD) represents patients who did not respond to ICB treatment. **(I** and **J)** Survival analysis in ICB treatment cohorts (IMvigor210 and GSE78220) using the log-rank test. The difference between the two groups was assessed using the Wilcox test.

positive correlation between the TMEIG and TIDE scores (**Figure 6F**, R = −0.15, $p < 0.003$).

Then, we explored the predictive power of the TMEIG score in two ICB therapy cohorts. In IMvigor210 and GSE78220 cohorts, the ICB response rates were significantly lower in the high TMEIG score group (**Figures 6G,H**). Notably, patients with high TMEIG scores exhibited worse overall survival (**Figures 6I,J**; IMvigor210: log-rank test, p = 0.018, GSE78220: log-rank test, p = 0.053). The insignificant result in GSE78220 (28 patients) can be attributed to the small sample size. Furthermore, the biomarker evaluation module on the TIDE website was used to assess the accuracy of the TMEIG score using multiple ICB cohorts as compared to other published biomarkers. The TMEIG score demonstrated an AUC of more than 0.5 in nine out of 16 ICB cohorts (**Supplementary Figure S6**), demonstrating its robustness as a predictive biomarker (Fu et al., 2020).

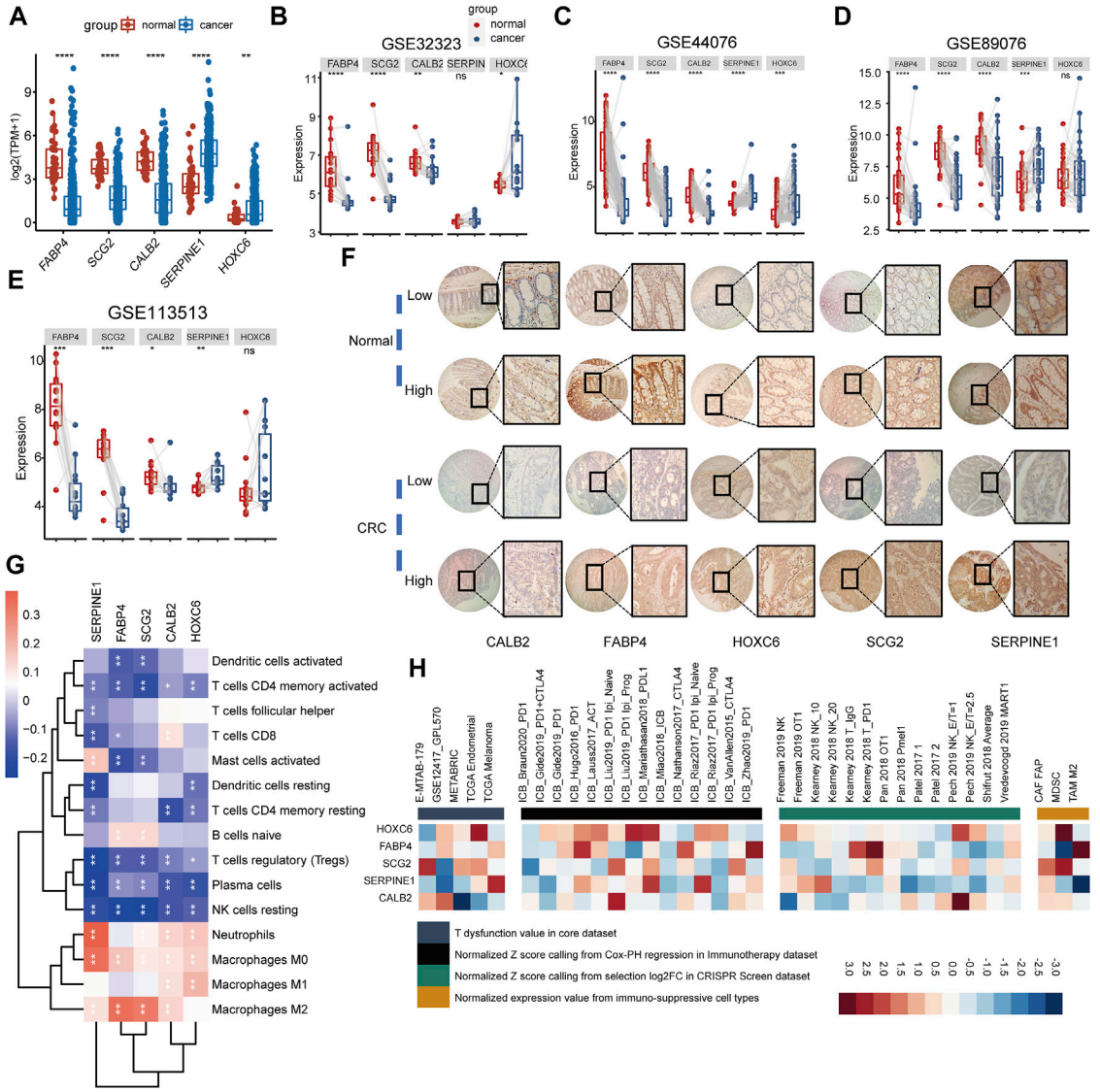


FIGURE 7 | Exploring the biological functions of biomarker genes. **(A)** Comparison of biomarker gene expression between normal tissue and cancer tissue in TCGA COAD. **(B–E)** Comparison of biomarker gene expression between cancer tissues and paired normal tissues, statistically assessed using Wilcox test. **(F)** The representative immunohistochemical images of FABP4, SCG2, CALB2, SERPINE1, and HOXC6. A total of 16 pairs of CRC tissue (cancer and adjacent normal tissue) were collected for IHC. **(G)** The heatmap shows the Pearson's correlation between five biomarker genes and immune cells in the combined GEO cohort. Red represents positive correlation, whereas blue represents negative correlation. **(H)** The correlation between five biomarker genes and four immunosuppressive indices (columns), including T cell dysfunction score (first column, T dysfunction value in core dataset), association with ICB survival outcome (second column, z-score in the Cox-PH regression in immunotherapy), log-fold change (logFC) in CRISPR screens (third column, helping identify regulators whose knockout can mediate the efficacy of lymphocyte-mediated tumor killing in cancer models), and T cell exclusion score (fourth column, assessing the gene expression levels in immunosuppressive cell types that drive T cell exclusion). Genes (rows) are ranked by average value across four immunosuppressive indices analyzed using the TIDE website.

The Biomarker Genes Are Differentially Expressed in CRC and Significantly Correlate With Immune Cells

To further understand the functions of the biomarker genes consisting of the TMEIG score, we analyzed the expression levels of *SERPINE1*, *FABP4*, *SCG2*, *CALB2*, and *HOXC6* in the TCGA-COAD cohort. The results demonstrated that the expression values of *SERPINE1* and *HOXC6* were significantly upregulated in tumors, whereas *FABP4*, *SCG2*, and *CALB2* were highly expressed in normal tissues (**Figure 7A**). Furthermore, the same results were found in paired differential expression in multiple CRC cohorts (**Figures 7B–E**, GSE32323, GSE44076, GSE89076, and GSE113513). Our immunohistochemistry (IHC) results revealed that *CALB2* exhibited relevant stronger staining in two cases of tumor (2/16), with the other 14 cases displaying low expression (14/16). For *FABP4*, seven and two cases in tumor and normal samples, respectively, exhibited

stronger staining, whereas nine and 14 cases in tumor and normal samples, respectively, were with low staining intensity. For *HOXC6*, seven and nine samples out of 16 exhibited strong staining in normal and tumor samples, respectively. The protein level of *SCG2* was high in 12 cases of CRC samples (12/16) and 11 cases of normal samples (11/16). *SEP1NG1* was found strongly stained in 11 cases of tumor samples (11/16) and nine cases of normal samples (9/16). Representative immunohistochemical images are shown in **Figure 7F**, and the high-resolution images are shown in **Supplementary Figure S7**. The qPCR experiments (**Supplementary Figure S8A**) revealed that the expression of *HOXC6* and *SERPINE1* was significantly upregulated in RKO and HT29 cell lines, and *FABP4* expression was downregulated in nearly all CRC cell lines analyzed. Although *SCG2* and *CALB2* were downregulated in patients with CRC in multiple cohorts, qPCR experiments showed that their expression was upregulated in several CRC cell lines (such as HCT116 and HT29). This discrepancy may be due to false positives in RNA sequencing or the heterogeneity between clinical tissues and tumor cells. Studies involving more clinical samples or cell lines may be needed to confirm the expression of the two genes at the RNA and protein levels in the future. KM plots of these genes are displayed in **Supplementary Figure S9A**. Results indicated that all genes were closely related to OS. GSEA analysis indicated that the five genes were involved in multiple cancer biological functions: cell motility, angiogenesis, cell migration, programmed cell death, MAPK signaling pathway, and PI3K-Akt signaling pathway. Notably, the immune-related pathway "cytokine-cytokine receptor interaction" was also significantly enriched in most of these genes (*SERPINE1*, *FABP4*, *SCG2*, and *HOXC6*) (**Supplementary Figure S9B**). Next, we summarized several immunological molecules from our previous studies, such as immune checkpoint genes and cytotoxic genes, and analyzed the correlation with five genes in TCGA COAD (**Supplementary Figures S8B,C**). Results demonstrated that most of the five genes were significantly correlated with immune checkpoint genes (*BTLA*, *CD274*, *CTLA4*, *HAVCR2*, *LAG3*, *PDCD1*, and *TIGIT*) and cytotoxic genes (granzyme A [*GZMA*], *GZMB*, *GZMK*, *GZMM*, interferon-gamma [*IFNG*], perforin 1 [*PRF1*], and tumor necrosis factor superfamily member 11 [*TNFSF11*]), which revealed that these genes might play an important role in tumor immunity. We then explored the correlation between the five genes and immune cells infiltrating the TME. As shown in the correlation heatmap, the five genes were positively related to macrophages (such as M0, M1, and M2), inversely correlated with resting NK cells and resting memory $CD4^+$ T cells, $CD8^+$ T cells, and activated memory $CD4^+$ T cells (**Figure 7G**), which might explain the poor ICB response in the high TMEIG score group. Moreover, the gene set prioritization module on the TIDE website indicated that *HOXC6* was the most appropriate target to treat TME resistance to ICB (**Figure 7H**). The expression of *HOXC6* was positively associated with T cell dysfunction in GSE12417, METABRIC, and TCGA Endometria datasets (**Figure 7H**, left panel). In addition, high *HOXC6* expression was also associated with poorer ICB outcomes in multiple cohorts treated with ICB (**Figure 7H**, second to left panel). Among the immune-suppressive cell types, *HOXC6* was highly expressed on the MDSC and CAF (**Figure 7H**, right panel).

DISCUSSION

Understanding the heterogeneity of the tumor microenvironment is required to elucidate the biological properties of CRC and guide the treatment strategies. Moreover, the TME heterogeneity is closely related to the efficacy of ICB therapy (Lee et al., 2014; Nishino et al., 2017; Cristescu et al., 2018; Mariathasan et al., 2018). Thus, understanding TME heterogeneity may provide new insights into CRC immunotherapy.

In this study, we constructed TME subtypes based on the TME landscape of patients with CRC, which can accurately distinguish the heterogeneity of the TME and predict the clinical prognosis. The patients were then re-clustered by TMEIGs identified by WGCNA and differential expression analysis. Two TMEIG subtypes were obtained, reflecting heterogeneity in TME and clinical prognosis. Gene signature is a simple and effective model widely used in clinical practice (Paik et al., 2004; van 't Veer et al., 2002; Parker et al., 2009). Therefore, we established a TMEIG score system to quantify the TME heterogeneity in patients with CRC. The Sankey plots revealed that the TME and TMEIG subtypes were consistent with the TMEIG score, suggesting that the TMEIG score could be utilized as a surrogate biomarker of TME heterogeneity.

Tumor mutation burden (TMB) (Chan et al., 2019), microsatellite instability (MSI) status (Ganesh et al., 2019), and immune checkpoint genes are important factors affecting ICB therapy. Patients with high levels of TMB and MSI-H exhibited better ICB therapy responses. In this study, there was no significant difference in TMB between high and low TMEIG score groups. However, patients with MSI-H possessed a higher TMEIG score, and there were more patients with MSI-high CRC in the high TMEIG score group. In addition, the expression of immune checkpoint molecules was higher in patients with high TMEIG scores. Patients with high expression of PD-L1 and PD-1 are more likely to benefit from ICB therapy (Topalian et al., 2012; Nishino et al., 2017). These results indicate that patients with high TMEIG scores tend to respond better to ICB therapy. However, ICB response is influenced by numerous factors, such as EMT (Jiang and Zhan, 2020), angiogenesis (Tian et al., 2017; Voron et al., 2014), and the TCF-β pathway (Mariathasan et al., 2018; Tauriello et al., 2018). EMT, angiogenesis, and TCFβ pathway activation inhibit the efficacy of immune checkpoint therapy. The pathway heatmap revealed that EMT, angiogenesis (*VIM, TWIST1, ZEB1*, and *ZEB2*), and TCFβ signatures were significantly activated in the high TMEIG score group (**Figure 7E**). Furthermore, the TIDE score predicted the efficacy of ICB therapy based on two mechanisms of tumor immune escape (T cell dysfunction and T cell exclusion) (Jiang et al., 2018). A higher TIDE score is related to poorer ICB response and survival in patients receiving anti–PD-1 and

anti-CTLA4 therapies (Jiang et al., 2018). In the present study, the TMEIG score was positively related to dysfunction and exclusion scores (**Figure 6F**). It indicated that patients with high TMEIG scores possessed fewer CTL cells, which were majorly dysfunctional in the TME. In line with the above results, the TMEIG score was positively correlated with the TIDE score ($R = 0.15$, $p = 0.003$), indicating that patients in the high TMEIG score group exhibited poorer ICB response. The prediction of ICB response by MSI, TMB, or PD-L1 is based on the presence of CTL cells in the TME. Hence, we speculated that patients with high TMEIG scores mainly tend to exhibit poor ICB therapy response due to fewer CTL cells that are primarily dysfunctional. Since there was no suitable public ICB treatment CRC cohorts at the time of publication, we only used transcriptome data from other tumor types to verify the predictive power of the TMEIG score. Nevertheless, validation in melanoma and urothelial cancer datasets may indirectly suggest that the TMEIG score predicts the efficacy of immune checkpoint therapy in CRC. In accordance with TIDE results, a higher TMEIG score was associated with poorer ICB response and prognosis in two ICB treatment cohorts. In conclusion, the evidence demonstrated that the TMEIG score might serve as a reliable ICB biomarker in CRC. We will further validate our results once transcriptome data of CRC patients undergoing immune checkpoint therapy becomes publicly available or establish our own cohort regarding this point.

In our study, the TMEIG score was determined by *SERPINE1, FABP4, SCG2, CALB2*, and *HOXC6* expression. *HOXC6* is a member of the homeobox family, which encode transcription factors that play a critical role in morphogenesis in all multicellular organisms. *HOXC6* expression was higher and negatively associated with prognosis in right-sided colon cancer than in left-sided colon cancer. This finding was further validated by tissue microarray analysis. *HOXC6* facilitated proliferation and metastasis through the dickkopf-1 (*DKK1*)/Wnt/β-catenin axis in right-sided colon cancer (Qi et al., 2021; Garris et al., 2018). The role of *FABP4*, which encodes the fatty acid–binding protein found in adipocytes, is unclear in CRC. A study demonstrated that *FABP4* was downregulated in CRC (Zhao et al., 2019). IHC and ELISA data from another study revealed that *FABP4* and plasma FABP4 concentrations were higher in CRC tissues than in normal tissues (Zhang et al., 2021). Thus, the role of *FABP4* in CRC must be investigated further. In addition, mRNA and protein levels of *SCG2*, a member of the chromogranin family of acidic secretory proteins, were significantly downregulated in CRC tissues (Wang et al., 2021; Fang et al., 2021). Mechanistically, *SCG2* inhibits tumor growth and angiogenesis by disrupting the activities of HIF-1α/VEGF in malignant CRC tissues (Fang et al., 2021). *In vitro* and *in vivo* studies have shown that *CALB2* promotes hepatocellular carcinoma metastasis *via* the TRPV2-Ca2+-ERK1/2 signaling pathway (Chu et al., 2022). Although fluorouracil (5-FU) treatment reduced the mRNA and protein expression of *CALB2* in CRC, their expression levels were not quantified and compared in tumor and normal tissues (Stevenson et al., 2011). *SERPINE1* expression is reportedly upregulated in CRC tissues and is associated with tumor invasiveness and aggressiveness (Mazzoccoli et al., 2012). Our study also reports the same trend (**Figures 7A–E**). Nevertheless, the roles of *HOXC6*, *SERPINE1*, *FABP4*, *SCG2*, and *CALB2* in tumorigenesis, cancer immunity, and ICB treatment are poorly understood. In the present study, IHC and qPCR results preliminarily elucidated the expression levels of these five molecules in CRC and normal tissues. Larger clinical sample sizes are required to verify mRNA and protein expression levels reported in this study and whether protein levels can be used to predict the prognosis of patients suffering from CRC and their response to ICB therapy. In addition, we observed that the five genes were significantly associated with immune cells of TME, immune checkpoint genes, and cytotoxic genes. Immune checkpoint genes and cytotoxic genes were collected from our previous study (Wang et al., 2022). Moreover, heatmaps also demonstrated that these genes, especially *HOXC6*, were closely associated with four immunosuppressive indices, including T cell dysfunction score, T cell exclusion score, association with ICB survival outcome, and logFC in CRISPR screens. Collectively, the roles of the five genes in tumor immunity are worthy of investigation, which will be the focus of our future research.

Our study has numerous advantages. First, datasets of the combined GEO cohort were downloaded from the GPL570 platform, which reduced the batch effect caused by different platform processes. Second, a large cohort with more than 1,000 samples was used for clustering, guaranteeing stable clustering results. Third, the prognostic power and predictive ICB response of the TMEIG score have been validated in multiple cohorts. However, the study design does have a few drawbacks. First, the predictive ICB response power of the TMEIG score was assessed in melanomas and metastatic urothelial cancer. These data must be verified using patients with CRC. Second, the relationship between the five molecules of the TMEIG score system and tumorigenesis, immune system, and ICB response were not investigated in this study. Future *in vivo* and *in vitro* studies from our group will focus on these aspects.

In conclusion, we identified the TME subtypes that comprehensively depicted the TME, revealed multiple aspects of CRC biology, and assessed variation in the prognosis of patients with CRC. TMEIG score is a robust marker to predict patients' prognosis and may serve as a predictor of ICB response in CRC. Moreover, we identified several potential targets that may play a critical role in ICB treatment, of which *HOXC6* may be the most significant.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Harbin Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

HW and YS analyzed the bioinformatic data, wrote the manuscript draft, and generated the figures and tables. ZL assisted in bioinformatics analysis and article structure design. SO generated the flow diagram and performed qPCR experiments. KL and ZG revised the manuscript and figures. LZ, RH, and SY conceived, designed, and guided the study and provided financial support.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant No. 81872034); Natural Science Foundation of Heilongjiang Province (Grant No. H2017016); Wu Jieping Medical Foundation (No. 320.6750.19092-41); Chen Xiao-Ping Foundation for the Development of Science and Technology of Hubei Province (Grant No. CXPJJH12000002-2020025); Ministry of Education Chunhui Project Cooperative Research Project (Grant No. HLJ2019010); Heilongjiang Natural Science Foundation of China (No. LH2020H120); and Haiyan Research Fund of Harbin Medical University Cancer Hospital (No. JJZD2020-04).

ACKNOWLEDGMENTS

The authors would like to thank Prof. Jun Xiang, Department of Colorectal Cancer Surgery, the Second Affiliated Hospital of Harbin Medical University; Prof. Zhiming Zhang, Department of Thoracic Surgery, the Second Affiliated Hospital of Harbin Medical University; and Prof. Yu Wang, Department of Ophthalmology, the Second Affiliated Hospital of Harbin Medical University, who helped complete this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.884839/full#supplementary-material

Supplementary Figure S1 | Processes of constructing TME subtypes. **(A–D)** Consensus matrixes of the combined GEO cohort for each *k* (*k* = 2–5), displaying the clustering stability using 1000 iterations of hierarchical clustering. **(E)** Empirical cumulative distribution function plots display consensus distributions for each *k*. When *k* = 2, the distribution reaches an approximate maximum, indicating maximum stability.

Supplementary Figure S2 | Details of the WGCNA analysis. **(A** and **B)** Analysis of the scale-free fit index and the mean connectivity for various soft-thresholding power values. **(C)** Hierarchical clustering dendrograms of co-expressed genes in modules. **(D** and **E)** The correlation between modules. **(F–H)** The GO and KEGG enrichment terms are in blue, brown, and green modules.

Supplementary Figure S3 | Details of constructing TMEIG subtypes. **(A–D)** Consensus matrixes of the combined GEO cohort for each *k* (*k* = 2–5), displaying the clustering stability using 1000 iterations of hierarchical clustering. **(E)** Empirical cumulative distribution function plots display consensus distributions for each *k*. When *k* = 2, the distribution reaches an approximate maximum, indicating maximum stability. **(F)** Relationship between TMEIG subtypes and 11 critical biological pathways.

Supplementary Figure S4 | Details of constructing the TMEIG score system. **(A)** The Venn diagrams show the intersection between genes in blue, brown, and green modules and DEGs. **(B)** The intersection of prognostic genes in the TCGA COAD cohort and the combined GEO cohort. **(C,D)** Details of the Lasso regression. **(E)** The risk factor diagrams of the TCGA COAD cohort. **(F** and **G)** The TMEIG score between TME Clusters as well as TMEIG Cluster. **(H)** ROC plot shows the predictive value of the TMEIG score combined with age, sex, M stage, and TNM stage in the TCGA COAD cohort using stepwise Cox regression.

Supplementary Figure S5 | Exploring the TMEIG score groups. **(A)** The KM plot of high and low TMEIG groups in MSI-H subgroups (log-rank *p*-value = 0.037). **(B)** The KM plot of high and low TMEIG groups in MSI-L/MSS subgroups (log-rank *p*-value = 0.0039). **(C)** The stacking histogram shows the proportion of patients with MSI-H and MSI-L/MSS in high and TMEIG score groups. Red represents patients with MSI-H, and blue represents those with MSI-L/MSS. **(D)** The OncoPrint shows the top six mutated genes between high and low TMEIG score groups, including *APC*, *TP53*, *TTN*, *KRAS*, *PIK3CA*, and *MUC16*. There are 100 and 274 patients in the high and low TMEIG score groups, respectively. Individual patients are represented in each column.

Supplementary Figure S6 | Comparison of TMEIG score and other biomarkers. AUC is employed to evaluate the prediction performance of the TMEIG score (Custom) and other common biomarkers on ICB response in 16 ICB treatment cohorts using the TIDE website.

Supplementary Figure S7 | High-definition images of IHC.

Supplementary Figure S8 | Exploring the role of the five biomarker genes. **(A)** The qPCR data of *FABP4*, *SCG2*, *CALB2*, *SERPINE1*, and *HOXC6*. NCM460 is a normal human colonic epithelial cell line, whereas SW620, RKO, HCT116, and HT29 are human CRC cell lines. NS, not significant. The statistical significance was assessed using one-way ANOVA. **(B)** Pearson correlation between the five genes and immune checkpoint genes (*BTLA*, *CD274*, *CTLA4*, *HAVCR2*, *LAG3*, *PDCD1*, and *TIGIT*) in TCGA COAD. **(C)** Pearson correlation between the five genes and cytotoxic genes (*GZMA*, *GZMB*, *GZMK*, *GZMM*, *IFNG*, *PRF1*, and *TNFSF11*) in TCGA COAD.

Supplementary Figure S9 | Survival and GSEA analyses of the five biomarker genes. **(A)** Survival analysis of *SERPINE1*, *FABP4*, *SCG2*, *CALB2*, and *HOXC6* in the TCGA COAD cohort. **(B)** GSEA analysis of *SERPINE1*, *FABP4*, *SCG2*, *CALB2*, and *HOXC6* in the combined GEO cohort. Data from "c2.cp.kegg.v6.2.symbols.gmt" and "c5.all.v7.0.symbols.gmt" in the MSigDB website were chosen as the reference gene sets.

Supplementary Table S1 | Cohorts information used in this study.

Supplementary Table S2 | Survival data of the combined GEO and TCGA COAD cohorts.

Supplementary Table S3 | Details of immunohistochemistry experiment in this study.

Supplementary Table S4 | CIBERSORT results of the combined GEO cohort.

Supplementary Table S5 | TME clusters in the combined GEO cohort.

Supplementary Table S6 | Results of GSEA analysis between TME clusters in the combined GEO cohort.

Supplementary Table S7 | TMEIG clusters in the combined GEO cohort.

Supplementary Table S8 | TMEIG score in the combined GEO and TCGA COAD cohorts.

Supplementary Table S9 | Clinical data of the TCGA COAD cohort.

Supplementary Table S10 | MSI information of the TCGA COAD cohort.

REFERENCES

Arce Vargas, F., Furness, A. J. S., Litchfield, K., Joshi, K., Rosenthal, R., Ghorani, E., et al. (2018). Fc Effector Function Contributes to the Activity of Human Anti-CTLA-4 Antibodies. *Cancer Cell* 33 (4), 649–e4. doi:10.1016/j.ccell.2018.02.010

Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the Population Abundance of Tissue-Infiltrating Immune and Stromal Cell Populations Using Gene Expression. *Genome Biol.* 17 (1), 218. doi:10.1186/s13059-016-1070-5

Casey, S. C., Amedei, A., Aquilano, K., Azmi, A. S., Benencia, F., Bhakta, D., et al. (2015). Cancer Prevention and Therapy through the Modulation of the Tumor Microenvironment. *Seminars cancer Biol.* 35 (Suppl.), S199–s223. doi:10.1016/j.semcancer.2015.02.007

Chakraborty, S., Datta, S., and Datta, S. (2012). Surrogate Variable Analysis Using Partial Least Squares (SVA-PLS) in Gene Expression Studies. *Bioinforma. Oxf. Engl.* 28 (6), 799–806. doi:10.1093/bioinformatics/bts022

Chan, T. A., Yarchoan, M., Jaffee, E., Swanton, C., Quezada, S. A., Stenzinger, A., et al. (2019). Development of Tumor Mutation Burden as an Immunotherapy Biomarker: Utility for the Oncology Clinic. *Ann. Oncol.* 30 (1), 44–56. doi:10.1093/annonc/mdy495

Chu, H., Zhao, Q., Shan, Y., Zhang, S., Sui, Z., Li, X., et al. (2022). All-Ion Monitoring-Directed Low-Abundance Protein Quantification Reveals CALB2 as a Key Promoter in Hepatocellular Carcinoma Metastasis. *Anal. Chem.* 94, 6102–6111. doi:10.1021/acs.analchem.1c03562

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. 362. doi:10.1126/science.aar3593

Daillère, R., Vétizou, M., Waldschmitt, N., Yamazaki, T., Isnard, C., Poirier-Colame, V., et al. (2016). Enterococcus Hirae and Barnesiella Intestinihominis Facilitate Cyclophosphamide-Induced Therapeutic Immunomodulatory Effects. *Immunity* 45 (4), 931–943. doi:10.1016/j.immuni.2016.09.009

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

Fang, C., Dai, L., Wang, C., Fan, C., Yu, Y., Yang, L., et al. (2021). Secretogranin II Impairs Tumor Growth and Angiogenesis by Promoting Degradation of Hypoxia-inducible Factor-1α in Colorectal Cancer. *Mol. Oncol.* 15 (12), 3513–3526. doi:10.1002/1878-0261.13044

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

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 (1), 21. doi:10.1186/s13073-020-0721-z

Ganesh, K., Stadler, Z. K., Cercek, A., Mendelsohn, R. B., Shia, J., Segal, N. H., et al. (2019). Immunotherapy in Colorectal Cancer: Rationale, Challenges and Potential. *Nat. Rev. Gastroenterol. Hepatol.* 16 (6), 361–375. doi:10.1038/s41575-019-0126-x

Garris, C. S., Arlauckas, S. P., Kohler, R. H., Trefny, M. P., Garren, S., Piot, C., et al. (2018). Successful Anti-PD-1 Cancer Immunotherapy Requires T Cell-Dendritic Cell Crosstalk Involving the Cytokines IFN-γ and IL-12. *Immunity* 49 (6), 1148–1161. doi:10.1016/j.immuni.2018.09.024

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

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex Heatmaps Reveal Patterns and Correlations in Multidimensional Genomic Data. *Bioinformatics* 32 (18), 2847–2849. doi:10.1093/bioinformatics/btw313

Hanoteau, A., Newton, J. M., Krupar, R., Huang, C., Liu, H.-C., Gaspero, A., et al. (2019). Tumor Microenvironment Modulation Enhances Immunologic Benefit of Chemoradiotherapy. *J. Immunother. Cancer* 7 (1), 10. doi:10.1186/s40425-018-0485-9

Huang, M.-Y., Jiang, X.-M., Wang, B.-L., Sun, Y., and Lu, J.-J. (2021). Combination Therapy with PD-1/pd-L1 Blockade in Non-small Cell Lung Cancer: Strategies and Mechanisms. *Pharmacol. Ther.* 219, 107694. doi:10.1016/j.pharmthera.2020.107694

Jiang, Y., and Zhan, H. (2020). Communication between EMT and PD-L1 Signaling: New Insights into Tumor Immune Evasion. *Cancer Lett.* 468, 72–81. doi:10.1016/j.canlet.2019.10.013

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T Cell Dysfunction and Exclusion Predict Cancer Immunotherapy Response. *Nat. Med.* 24 (10), 1550–1558. doi:10.1038/s41591-018-0136-1

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28 (1), 27–30. doi:10.1093/nar/28.1.27

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

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

Lee, K., Hwang, H., and Nam, K. T. (2014). Immune Response and the Tumor Microenvironment: How They Communicate to Regulate Gastric Cancer. *Gut Liver* 8 (2), 131–139. doi:10.5009/gnl.2014.8.2.131

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

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004

Luke, J. J., Flaherty, K. T., Ribas, A., and Long, G. V. (2017). Targeted Agents and Immunotherapies: Optimizing Outcomes in Melanoma. *Nat. Rev. Clin. Oncol.* 14 (8), 463–482. doi:10.1038/nrclinonc.2017.43

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

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

Mazzoccoli, G., Pazienza, V., Panza, A., Valvano, M. R., Benegiamo, G., Vinciguerra, M., et al. (2012). ARNTL2 and SERPINE1: Potential Biomarkers for Tumor Aggressiveness in Colorectal Cancer. *J. Cancer Res. Clin. Oncol.* 138 (3), 501–511. doi:10.1007/s00432-011-1126-6

Navarro Gonzalez, J., Zweig, A. S., Speir, M. L., Schmelter, D., Rosenbloom, K. R., Raney, B. J., et al. (2021). The UCSC Genome Browser Database: 2021 Update. *Nucleic Acids Res.* 49 (D1), D1046–D1057. doi:10.1093/nar/gkaa1070

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 (5), 453–457. doi:10.1038/nmeth.3337

Nishino, M., Ramaiya, N. H., Hatabu, H., and Hodi, F. S. (2017). Monitoring Immune-Checkpoint Blockade: Response Evaluation and Biomarker Development. *Nat. Rev. Clin. Oncol.* 14 (11), 655–668. doi:10.1038/nrclinonc.2017.88

Pagès, F., Berger, A., Camus, M., Sanchez-Cabo, F., Costes, A., Molidor, R., et al. (2005). Effector Memory T Cells, Early Metastasis, and Survival in Colorectal Cancer. *N. Engl. J. Med.* 353 (25), 2654–2666. doi:10.1056/NEJMoa051424

Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., et al. (2004). A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. *N. Engl. J. Med.* 351 (27), 2817–2826. doi:10.1056/NEJMoa041588

Parker, J. S., Mullins, M., Cheang, M. C. U., Leung, S., Voduc, D., Vickery, T., et al. (2009). Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. *Jco* 27 (8), 1160–1167. doi:10.1200/jco.2008.18.1370

Pettenati, C., and Ingersoll, M. A. (2018). Mechanisms of BCG Immunotherapy and its Outlook for Bladder Cancer. *Nat. Rev. Urol.* 15 (10), 615–625. doi:10.1038/s41585-018-0055-4

Qi, L., Chen, J., Zhou, B., Xu, K., Wang, K., Fang, Z., et al. (2021). HomeoboxC6 Promotes Metastasis by Orchestrating the DKK1/Wnt/β-Catenin axis in Right-Sided Colon Cancer. *Cell Death Dis.* 12 (4), 337. doi:10.1038/s41419-021-03630-x

Routy, B., Le Chatelier, E., Derosa, L., Duong, C. P. M., Alou, M. T., Daillère, R., et al. (2018). Gut Microbiome Influences Efficacy of PD-1-Based

Immunotherapy against Epithelial Tumors. *Science* 359. New York, NY, 91–97. doi:10.1126/science.aan3706

Sheng, H., Huang, Y., Xiao, Y., Zhu, Z., Shen, M., Zhou, P., et al. (2020). ATR Inhibitor AZD6738 Enhances the Antitumor Activity of Radiotherapy and Immune Checkpoint Inhibitors by Potentiating the Tumor Immune Microenvironment in Hepatocellular Carcinoma. *J. Immunother. Cancer* 8 (1), e000340. doi:10.1136/jitc-2019-000340

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 (23), 2189–2199. doi:10.1056/NEJMoa1406498

Stevenson, L., Allen, W. L., Proutski, I., Stewart, G., Johnston, L., McCloskey, K., et al. (2011). Calbindin 2 (CALB2) Regulates 5-fluorouracil Sensitivity in Colorectal Cancer by Modulating the Intrinsic Apoptotic pathwayCentral PMCID: PMCPMC3101240 Following Conflicts: Professor Patrick Johnston Is the Founder and Director of Almac Diagnostics. *PloS one* 6 (5), e20276. doi:10.1371/journal.pone.0020276

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. U.S.A.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102

Tauriello, D. V. F., Palomo-Ponce, S., Stork, D., Berenguer-Llergo, A., Badia-Ramentol, J., Iglesias, M., et al. (2018). TGFβ Drives Immune Evasion in Genetically Reconstituted Colon Cancer Metastasis. *Nature* 554 (7693), 538–543. doi:10.1038/nature25492

Tian, L., Goldstein, A., Wang, H., Ching Lo, H., Sun Kim, I., Welte, T., et al. (2017). Mutual Regulation of Tumour Vessel Normalization and Immunostimulatory Reprogramming. *Nature* 544 (7649), 250–254. doi:10.1038/nature21724

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

van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A. M., Mao, M., et al. (2002). Gene Expression Profiling Predicts Clinical Outcome of Breast Cancer. *Nature* 415 (6871), 530–536. doi:10.1038/415530a

Väyrynen, J. P., Haruki, K., Lau, M. C., Väyrynen, S. A., Zhong, R., Dias Costa, A., et al. (2021). The Prognostic Role of Macrophage Polarization in the Colorectal Cancer Microenvironment. *Cancer Immunol. Res.* 9 (1), 8–19. doi:10.1158/2326-6066.Cir-20-0527

Vitale, I., Manic, G., Coussens, L. M., Kroemer, G., and Galluzzi, L. (2019). Macrophages and Metabolism in the Tumor Microenvironment. *Cell metab.* 30 (1), 36–50. doi:10.1016/j.cmet.2019.06.001

Voron, T., Marcheteau, E., Pernot, S., Colussi, O., Tartour, E., Taieb, J., et al. (2014). Control of the Immune Response by Pro-angiogenic Factors. *Front. Oncol.* 4, 70. doi:10.3389/fonc.2014.00070

Wang, H., Yin, J., Hong, Y., Ren, A., Wang, H., Li, M., et al. (2021). SCG2 Is a Prognostic Biomarker Associated with Immune Infiltration and Macrophage Polarization in Colorectal Cancer. *Front. Cell Dev. Biol.* 9, 795133. doi:10.3389/fcell.2021.795133

Wang, H., Luo, K., Guan, Z., Li, Z., Xiang, J., Ou, S., et al. (2022). Identification of the Crucial Role of CCL22 in F. Nucleatum-Related Colorectal Tumorigenesis that Correlates with Tumor Microenvironment and Immune Checkpoint Therapy. *Front. Genet.* 13, 811900. doi:10.3389/fgene.2022.811900

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: a Class Discovery Tool with Confidence Assessments and Item Tracking. *Bioinforma. Oxf. Engl.* 26 (12), 1572–1573. doi:10.1093/bioinformatics/btq170

Wu, T., and Dai, Y. (2017). Tumor Microenvironment and Therapeutic Response. *Cancer Lett.* 387, 61–68. doi:10.1016/j.canlet.2016.01.043

Yin, Z., Li, C., Wang, J., and Xue, L. (2019). Myeloid-derived Suppressor Cells: Roles in the Tumor Microenvironment and Tumor Radiotherapy. *Int. J. Cancer* 144 (5), 933–946. doi:10.1002/ijc.31744

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

Zhang, Y., Song, J., Zhao, Z., Yang, M., Chen, M., Liu, C., et al. (2020). Single-cell Transcriptome Analysis Reveals Tumor Immune Microenvironment Heterogenicity and Granulocytes Enrichment in Colorectal Cancer Liver Metastases. *Cancer Lett.* 470, 84–94. doi:10.1016/j.canlet.2019.10.016

Zhang, Y., Zhang, W., Xia, M., Xie, Z., An, F., Zhan, Q., et al. (2021). High Expression of FABP4 in Colorectal Cancer and its Clinical Significance. *J. Zhejiang Univ. Sci. B* 22 (2), 136–145. doi:10.1631/jzus.B2000366

Zhao, D., Ma, Y., Li, X., and Lu, X. (2019). microRNA-211 Promotes Invasion and Migration of Colorectal Cancer Cells by Targeting FABP4 via PPARγ. *J. Cell Physiol.* 234, 15429–15437. doi:10.1002/jcp.28190

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 Molecular Biosciences

TYPE Original Research
PUBLISHED 05 September 2022
DOI 10.3389/fmolb.2022.815260

OPEN ACCESS

EDITED BY
William C. Cho,
QEH, Hong Kong SAR, China

REVIEWED BY
Yuriy L. Orlov,
I.M. Sechenov First Moscow State Medical University, Russia
Nguyen Minh Duc,
Pham Ngoc Thach University of Medicine, Vietnam

*CORRESPONDENCE
Feng Jiang,
drjiangfeng@163.com
Jie Ma,
majie@xinhuamed.com.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 15 November 2021
ACCEPTED 15 July 2022
PUBLISHED 05 September 2022

CITATION
Cao L, Zhao Y, Liang Z, Yang J, Wang J, Tian S, Wang Q, Wang B, Zhao H, Jiang F and Ma J (2022), Systematic analysis of MCM3 in pediatric medulloblastoma via multi-omics analysis.
Front. Mol. Biosci. 9:815260.
doi: 10.3389/fmolb.2022.815260

Systematic analysis of MCM3 in pediatric medulloblastoma *via* multi-omics analysis

Liangliang Cao[1], Yang Zhao[1†], Zhuangzhuag Liang[1†], Jian Yang[1], Jiajia Wang[1], Shuangwei Tian[1], Qinhua Wang[1], Baocheng Wang[1], Heng Zhao[1], Feng Jiang[1]* and Jie Ma[1]*

[1]Department of Pediatric Neurosurgery, Xin Hua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China

Minichromosome maintenance proteins are DNA-dependent ATPases that bind to replication origins and allow a single round of DNA replication. One member of this family, MCM3, is reportedly active in most cancers. To systematically elucidate the mechanisms affected by aberrant MCM3 expression and evaluate its clinical significance, we analyzed multi-omics data from the GEO database and validated them in cell lines and tumor samples. First, we showed the upregulation of MCM3 in medulloblastoma (MB) at bulk and single-cell RNA sequence levels and revealed the potential role of MCM3 *via* DNA replication. Then we found the dysregulation of MCM3 might result from abnormal methylation of MCM3. Moreover, we discovered that MCM3 might affect varied biological processes such as apoptosis, autophagy, and ferroptosis and that MCM3 was correlated with immune components such as fibroblast and neutrophils, which were associated with overall survival in different medulloblastoma subtypes. Furthermore, we found that MCM3 expression was correlated with the IC_{50} values of cisplatin and etoposide. The nomogram of MCM3-related genes showed the reliable and better prediction of 1- and 5-year survival compared to current histological and molecular classifications. Overall, the results of our study demonstrated that MCM3 might serve as a potential biomarker with clinical significance and better guidance than current histological and molecular classifications for clinical decision-making.

KEYWORDS

MCM3, medulloblastoma, multi-omics, DNA methylation, single cell, RNA sequencing, prognosis, nomogram

Introduction

Medulloblastoma (MB), the leading cause of cancer-related death in children, is one of the most common pediatric brain tumors (Hovestadt et al., 2020). In recent years, individualized therapy models have emerged based on molecular subtypes and risk stratification. Surgical resection, cytotoxic chemotherapy, and craniospinal irradiation (for non-infants usually ≥3 years of age at diagnosis) constitute the standard therapy for

MB. The estimated 5-year overall survival has remained unchanged during the past two decades, ranging from 60% to 80% (Lannering et al., 2012; von Bueren et al., 2016). Despite these high estimates, the drawbacks of current treatment strategies include toxic effects on neurocognition and the neuroendocrine systems, sluggish identical therapies concerning radiotherapy and cytotoxic chemotherapy in developing children, lack of indicators for novel clinical medications, etc. Therefore, more therapeutic targets and less toxic strategies are required.

Non-invasive methods have made great advances with the identification of molecular subtypes based on DNA methylation. Similarly, as another clinical evaluation method, magnetic resonance imaging (MRI) also helps the differentiation of medulloblastoma from other pediatric brain tumors and risk stratification based on different features of T1 and T2-weighted MRI (Duc et al., 2020; Minh Thong and Minh Duc, 2020; Zhang et al., 2022). In addition, cellular proliferation plays an essential role in tumor content, especially in highly malignant cancers (Gu et al., 2021; McCarthy et al., 2021; Newman et al., 2021; Qiu et al., 2021). As the molecular mechanisms involved have been uncovered gradually, increasing numbers of informative biomarkers have been identified to evaluate the degree of malignancy of various cancers, including proliferating cell nuclear antigen (PCNA) and marker of proliferation Ki-67 (MKi67). Additionally, eukaryotic DNA replication guarantees genome stability. The minichromosome maintenance (MCM) proteins play a role as subunits of pre-replication complexes in the G1 phase and bind to replication origins and restrict DNA synthesis to a single round of DNA replication (Madine et al., 1995; Sedlackova et al., 2020). MCM proteins can reflect the cell cycle status due to their stable state during the cell cycle and proteolysis in quiescent cells (G0) (Musahl et al., 1998; Madine et al., 2000). Some studies have reported other functions of MCM proteins in different cancers, such as their relationships to the immune response in brain gliomas (Söling et al., 2005), execution of apoptosis (Schwab et al., 1998), regulation of autophagy (Puustinen et al., 2020), resistance to anti-tumor therapies (Shrestha et al., 2021), and stemness of cancer cells (Wang et al., 2020). The dysregulated expression of MCM3 has also been demonstrated in varied tumors and could serve as a target or prognostic biomarker (Stewart et al., 2017; Iglesias-Gato et al., 2018; Zhao et al., 2020). To date, there is only one study has reported the expression of MCM3 in various MB cell lines, and evidence of its systematic roles in MB remains deficient.

The present study systematically analyzed the functions of MCM3 in pediatric MB combined with clinical tumor specimens *via* multi-mics bioinformatic analysis. The results revealed its potential roles as a therapeutic target and a tool for better guidance compared to current histological and molecular classifications for clinical decision-making.

Materials and methods

Public data collection and construction of the validation cohort

The normalized pediatric MB datasets of DNA Methylation, mRNA array, and single-cell RNA sequencing (including GSE85212, GSE54880, GSE85217, GSE42656, GSE50161, and GSE155446) were obtained from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) database. In addition, 62 clinical tumor tissues were collected from children diagnosed with primary MB who received surgical treatment in our medical center (Xinhua Hospital Affiliated to Shanghai Jiao Tong University of Medicine) between July 2012 and October 2017. The study protocol was approved by the Ethics Committee of the Xinhua Hospital Affiliated with Shanghai Jiao Tong University School of Medicine (Approval No. XHEC-D-2021-076, Approval Date. 2021-10-21). Written informed consents were obtained from all patients.

RNA sequencing of clinical samples

RNA-seq service was provided by MAJORBIO (Shanghai, China), and completed on a HiSeq4000 instrument. The RNA-seq reads were mapped to the hg38 reference genome using STAR (v2.5.3a) (Dobin et al., 2013). Fragments per kilobase of transcript per million fragments mapped (FPKM) was calculated, and a mean FPKM ≥ 1 was set as the threshold to determine the active genes in all samples.

Expression features of MCM3 in different cancers

The Gene Expression Profiling Interactive Analysis (GEPIA.2) database was used to determine the differential expression of MCM3 in various cancers (Tang et al., 2019). The expression features in different brain tumors, biological functions, and gene effects in MB cell lines were analyzed from the CCLE database (Ghandi et al., 2019). The function of MCM3 was investigated in the Biological General Repository for Interaction Datasets-Open Repository for CRISPR Screens (BioGRID ORCS) database (Oughtred et al., 2019). The gene effects on cancer cells were studied using the DepMap database (Tsherniak et al., 2017). The relationships between MCM3 and clinical features were investigated with GEO data and validated in our own data by utilizing the R programing language.

Differential gene expression and functional enrichment analyses

Differential expression analysis was performed with two datasets (GSE42656 AND GSE50161) using R/limma (Ritchie et al., 2015). The differentially expressed genes (DEGs) were obtained by the intersecting DEGs from the two datasets. The DEGs from GSE50161 were analyzed by comparing the tumor group to fetal and adult normal brain tissues, respectively, to reduce the impact of developmental genes. The criteria for DEG analysis were $p < 0.05$ and fold change (FC) > 2. The interaction network between DEGs was predicted using the online STRING tool (Szklarczyk et al., 2021). The hub genes and related biological processes were identified in Cytoscape (ClueGO) (Shannon et al., 2003).

Data processing and analysis of MCM3 in single-cell RNA sequencing data

The count matrix obtained from GEO was processed using the Seurat package to get the Seurat object, filtered with a criterion of >500 and <7,800 genes, and normalized using the NormalizeData function (Butler et al., 2018). Highly variable genes between cells were then identified using the FindVariableFeatures function for the subsequent principal component analysis (PCA). Ten principal components were presented for uniform manifold approximation and projection (UMAP) dimension reduction to obtain a two-dimensional representation of the cell state. The FindClusters function was used for clustering with a selection of resolution of 0.3. The singleR package was applied for cell annotation, in which non-immune cells were treated as tumor cells for simplified analysis (Aran et al., 2019). The expression values of MCM3 in different clusters or groups were analyzed using the FeaturePlot function in the Seurat package. Cells expressing MCM3 were extracted and classified as showing high or low expression levels according to the mean expression value. Moreover, the DEGs between them were identified using the FindMarkers function and analyzed with enrichment analysis to investigate the biological process affected.

Effects of MCM3 dysregulation

Single-sample gene set variation analysis (ssGSVA) was performed in R/GSVA to analyze the biological functions between high and low-risk classifications of MCM3 (Sonja et al., 2013). Gene ontology (GO) and pathway enrichment analysis (Kyoto Encyclopedia of Genes and Genomes (KEGG)) were performed in R/clusterProfiler (Yu et al., 2012) to analyze the intersected processes shared between the MCM3-correlated genes and DEGs. *P* values <0.05 and FDR <25% were considered statistically significant. Immune infiltration was evaluated using the "xCell" (Aran et al., 2017), "ESTIMATE" (Yoshihara et al., 2013), and "CIBERSORT" (Newman et al., 2015) packages in R. The immune indexes related to survival rate were identified using the "survival" package in R. Their correlations with MCM3 expression were also analyzed. The genes related to apoptosis, autophagy, and ferroptosis were obtained from corresponding online databases, including Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005), AmiGo 2 (Park et al., 2015), HAMdb (Wang et al., 2018) and FerrDb (Zhou and Bao, 2020).

DNA methylation analysis of MCM3 in MB

Missing values in the beta value matrixes were processed using the "impute" package in R. The probes were filtered and normalized using the "ChAMP" package (Andrew et al., 2014). The data quality was then checked by principle component analysis (PCA) and heatmaps. The MCM3 probes were analyzed via differential, correlation, and survival analyses to identify methylation sites affecting MCM3 expression.

Prognostic model construction and analysis of drug susceptibility

Survival data of GSE85217 was downloaded and filtered (age <18 years). This discovery set was then randomly divided into training and test sets in a 7:3 ratio. The data set of 62 RNA sequencing data from our medical center was treated as the independent validation set for the prognostic model. The overall survival and MCM3-related genes were identified using univariate Cox regression and lasso regression analyses. A nomogram containing a multigene panel and clinical features was used to predict the survival probability. The pRRophetic package in R was used to predict the drug sensitivity of each sample according to the gene expression matrix and to evaluate the correlation between MCM3 expression and IC_{50} values (Geeleher et al., 2014).

shRNA plasmid

For the generation of shRNA plasmids, double-strand oligonucleotides were annealed and cloned into the CMV-EGFP-F2A-puro vector. The oligonucleotides of shRNA were synthesized by OBIO Technology (Shanghai, China). The target oligonucleotides were:

shMCM3-1: GGATGAATCAGAGACAGAA; shMCM3-2: GCAGTCAATCGGCATGAAT; shMCM3-3: GCCTCACAG AATCCATCAA.

Cell transfection

D425 and D458 cell lines were kind gifts from Shanghai Jiao Tong University of Medicine School, China. The D283 and D458 cell lines were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 1% penicillin–streptomycin and 10% FBS. At 48 h post-infection, the cells were harvested and subjected to protein extraction or other cellular experiments.

RNA extraction and real-time RT-PCR

Total RNA was extracted using TRIzol Reagent (Takara, 9108) according to the manufacturer's instructions. A High-capacity cDNA Reverse Transcription Kit (Takara, RRO47A) was used to perform the reverse transcription reactions. Quantitative PCR was performed on an ABI VERTI Real-Time PCR instrument. The relative mRNA levels were normalized to GAPDH. The qPCR primer sequences were:

MCM3, 5′-TCAGAGAGATTACCTGGACTTCC-3′ (forward); 5′-TCAGCCGGTATTGGTTGTCAC-3′ (reverse).

Western blot assay

The target protein was extracted, and its concentration was quantified using a BCA Protein Assay Kit (Pierce, 23227). Protein samples were separated by sodium dodecyl sulfate-polyacrylamide gel (SDS-PAGE) and transferred onto polyvinylidene difluoride (PVDF) membranes (Millipore). The membranes were blocked with 5% fat-free milk (BD Biosciences, 232100) and then incubated with primary antibodies against MCM3 (1:1,000; Cell Signaling Technology, 13421S), MCM2 (1:1,000; Cell Signaling Technology, 12191), MCM7 (1:2,000; Proteintech, 66905-1), and CDC45 (1:1,000; Cell Signaling Technology, 9405S), respectively. The secondary antibodies were HRP-linked goat anti-mouse IgG (1 ml; Cell Signaling Technology, 7076). The chemical fluorescence images of the proteins were visualized using a chemiluminescent substrate (Epizyme Biotech, Shanghai, China).

Cell proliferation assay

The effects of etoposide (XY91494, X-Y Biotechnology, China) and cisplatin (ST1164, Beyotime, China) on cell proliferation under different conditions were determined using Cell Counting Kit-8 (CKK-8) reagents (B34304, Bimake, China). Cells were plated in 96-well plates (4 replicates per condition), treated with serial drug concentrations, and incubated in normoxic conditions (37°C, 5% CO_2, 21% O_2) for 24, 48, and 72 h. The CKK-8 assays were performed to determine the IC_{50} values at each time point.

Results

Dysregulated MCM3 expression in pediatric MB and identification of MCM3-related signaling pathways and genes

The RNA-seq data of different brain tumor cell lines from the Cancer Cell Line Encyclopedia (CCLE) showed the highest MCM3 mRNA expression level in MB (Figure 1A). The GEPIA 2 database also revealed higher MCM3 mRNA expression levels in cancer tissues compared to those in normal tissues in the Oncomine database (Supplementary Figure S1A). The CRISPR function of MCM3 data from the BioGRID ORCS database revealed its function in the cell cycle and potential roles in immune response (Figure 1B). Differential analysis in GSE42656 and GSE50161 showed MCM3 upregulation in malignant samples compared to normal (Figure 1C), adult (Supplementary Figure S1B), and fetal brain tissues (Supplementary Figure S1C, Supplementary Tables S1–S3). Moreover, 255 genes were dysregulated in the two data sets based on the criteria of the absolute logFC >1 and the adjusted *p*-value <0.05 (Figure 1D). The heatmap also showed the differential expression of 255-DEGs (Figure 1E; Supplementary Figures S1D,E). Hub genes analysis revealed the core role of MCM3 in these DEGs (Supplementary Figure S1F) and top 3 ranking among hub genes (Supplementary Figure S1G). Enrichment analysis *via* STRING and ClueGO indicated that DNA replication was the main pathway involved (Figure 1F). The gene effect analysis of the DepMap database also indicated the dependence of cancer cells on MCM3 (Figure 1G). In addition, knockdown of the MCM3 protein level was performed by sh-MCM3 in both D283 and D458 cell lines, which was confirmed by RT-PCR (Figure 1H) and WB (Figure I). The cell viability decreased significantly at 72 h in the sh-MCM3 groups compared to that in the sh-Control groups in both D283 and D458 cell lines. These results revealed the dysregulation of MCM3 in cancers, especially MB, and that MCM3 was essential for the survival of malignant cells in MB.

MCM3 drives the malignant transformation of non-immune cells *via* DNA replication-related pathways

A total of 38,328 cells were divided into 30 clusters comprising six cell types including astrocytes, neurons, B-cells, T-cells, macrophages, and monocytes (Figure 2A) by SingleR. MCM3 expression in all clusters was heterogeneous (Figure 2B). Heterogeneity was also observed in malignant cells, namely non-immune cells, and non-tumor clusters (Figure 2C). Compared to non-tumor cells, MCM3 was significantly overexpressed in tumor cells (Figure 2D). We then investigated the effect of MCM3 on cell biology processes at

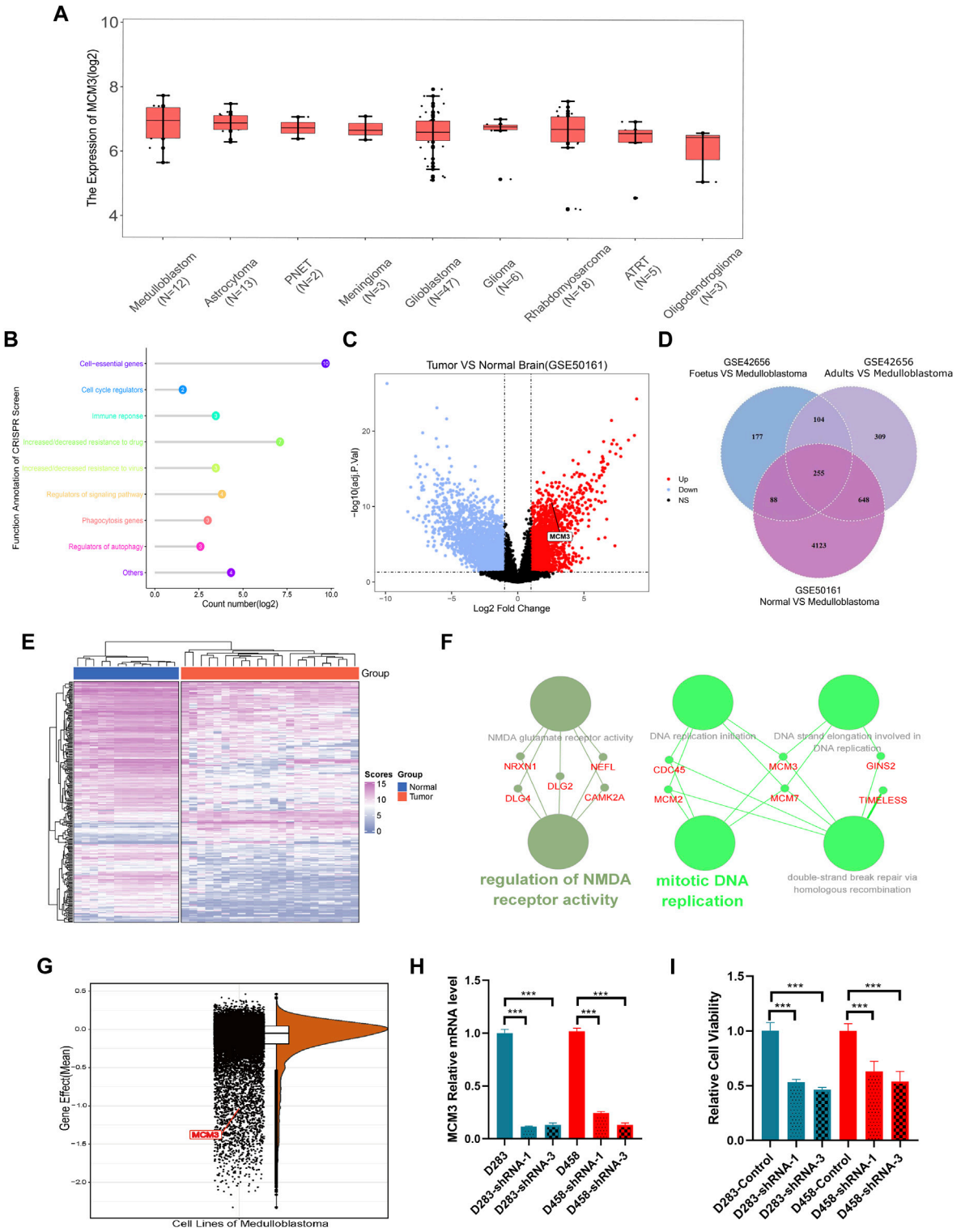
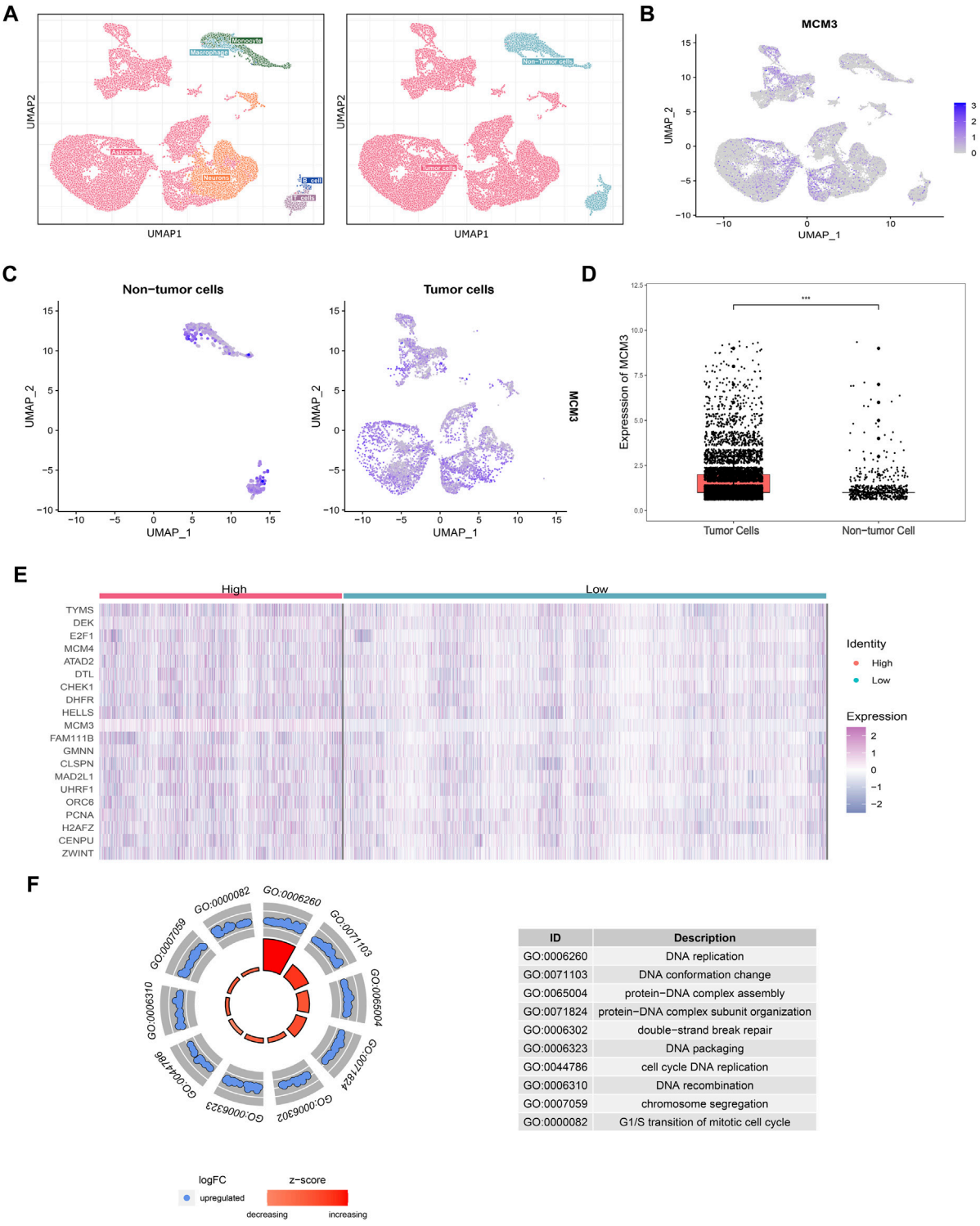


FIGURE 1
MCM3 expression levels among brain tumor cell lines **(A)**. Functional annotation of MCM3 from the BioGRID ORCS database by CRISPR **(B)**. MCM3 upregulation in the tumor group of GSE50161 **(C)**. 255 DEGs showed dysregulation in the three contrast sets **(D)**. The normal and medulloblastoma samples from GSE50161 clustered respectively according to the expression of DEGs **(E)**. Hub genes analysis *via* Cytoscape showing the main biological functions with which MCM3 is involved **(F)**. Gene effects (namely the dependency of the cell on genes) of MCM3 reflect its essentiality for cancer cell survival **(G)**. Confirmation of MCM3 knockdown by shRNA-1 and shRNA-3 on RNA level by QT-PCR **(H)** in both D283 and D458 cell lines **(H)**. Significantly decreased cell viability in sh-MCM3 cell lines **(I)**.



ID	Description
GO:0006260	DNA replication
GO:0071103	DNA conformation change
GO:0065004	protein−DNA complex assembly
GO:0071824	protein−DNA complex subunit organization
GO:0006302	double−strand break repair
GO:0006323	DNA packaging
GO:0044786	cell cycle DNA replication
GO:0006310	DNA recombination
GO:0007059	chromosome segregation
GO:0000082	G1/S transition of mitotic cell cycle

FIGURE 2
Cell annotations according to SingleR and cell classifications of malignant and non-malignant Cell annotations according to SingleR and cell classifications of malignant and non-malignant cells **(A)**. Heterogeneity of MCM3 expression in different clusters **(B)**. Heterogeneity of MCM3 express ion in tumor and non-tumor cells **(C)**. MCM3 over-expression in tumor cells compared to non-tumor cells **(D)**. Top 20 DEGs **(E)** and the top10 GO **(F)** annotations of 159 DEGs between the two groups of malignant cells (high and low MCM3 expression).

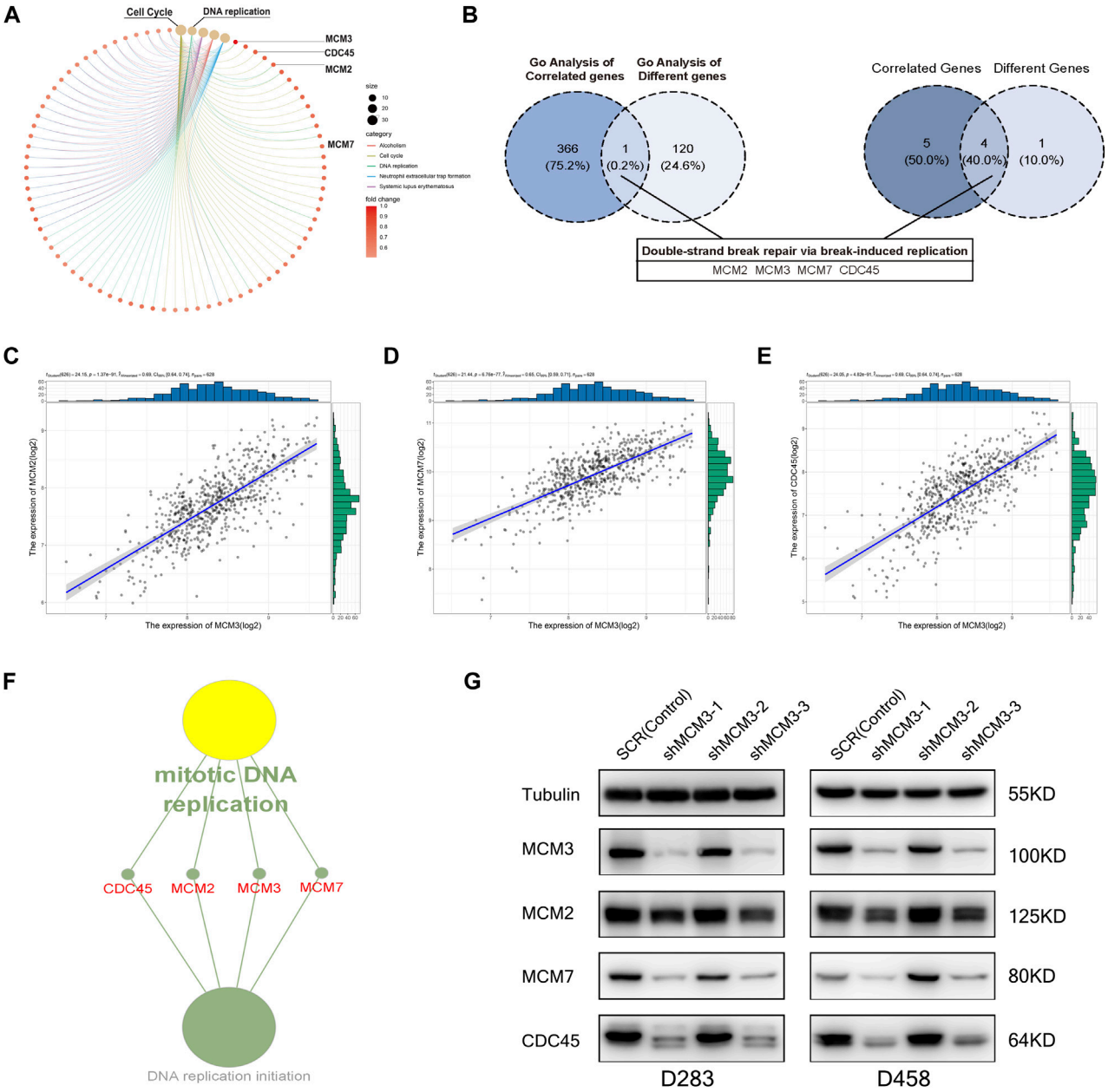


FIGURE 3
GO enrichment analysis of DEGs for different MCM3 expression levels showing the dysregulation of the cell cycle and DNA replication **(A)**. GO analysis of MCM3-correlated genes and DEGs revealed overlap only for "Double-strand break repair *via* break-induced replication", with only four genes in this biological process differentially expressed and significantly correlated with MCM3 **(B)**. Significant correlations of MCM3 with MCM2, MCM7, and CDC45 were **(C–E)**. MCM2, MCM3, MCM7, and CDC45 are the main components of the pre-replicative complex **(F)**. MCM7 and CDC45 are correlated with MCM3 at the protein level *via* WB analysis in D283 and D458 cell lines **(G)**.

the single-cell level. A total of 7,536 cells expressing MCM3 were classified as high and low levels according to the mean MCM3 expression level. The top 20 DEGs of the two groups were then identified (Figure 2E). GO analysis was performed of the 159 DEGs after filtering for log fold-change >0.25 and a minimum fraction of 0.25. Like the enrichment analysis in bulk RNA sequencing, the biological processes associated with DNA replication and cell cycle were significantly enriched, in addition to double-strand break repair (Figure 2F). Therefore, MCM3 overexpression might be associated with the malignant transformation of cells *via* the dysregulation of DNA replication and cell cycle.

TABLE 1 Data of RNA-seq included for Kaplan-Meier analysis.

	GSE85217	Our cohort
	(*n* = 628)	(*n* = 62)
Gender		
Female	212 (33.8%)	25 (40.3%)
Male	405 (64.5%)	37 (59.7%)
NA	11 (1.8%)	NA
Age		
Mean (SD)	7.30 (4.10)	5.20 (3.22)
Median (Min, Max)	7.00 (0.24,17.3)	4.45 (0.334,16.1)
Histology		
Classic	340 (54.1%)	41 (66.1%)
Desmoplastic	86 (13.7%)	7 (11.3%)
LCA	67 (10.7%)	9 (14.5%)
MBEN	17 (2.7%)	5 (8.1%)
NA	118 (18.8%)	NA
Metastasis		
Yes	165 (26.3%)	28 (45.2%)
No	333 (53.0%)	34 (54.8%)
NA	130 (20.7%)	NA
Dead		
Yes	147 (23.4%)	21 (33.9%)
No	398 (63.4%)	41 (66.1%)
NA	83 (13.2%)	NA
Overall Survival (years)		
Mean (SD)	4.88 (3.66)	3.93 (2.74)
Median (Min, Max)	3.92 (0.0192,19.0)	3.29 (0.186,10.0)
NA	90 (14.3)	NA

MCM3 affects tumorigenesis *via* CDC45-MCM2-7-GINS helicases

To investigate the potential mechanism by which MCM3 drives tumorigenesis in MB, we performed GSEA on samples with high and low MCM3expression levels. Dysregulation of MCM3 expression was mainly associated with cell cycle and DNA replication (Figure 3A). Moreover, enrichment analysis of the MCM3-correlated genes and 255 DEGs respectively showed an intersection only for "Double-strand break repair *via* break-induced replication" (Figure 3B). Only four genes—MCM2, MCM3, MCM7, and CDC45—met both the criterion of different expression, and were correlated with MCM3 expression and were included in further analysis (Figures 3C–E). The results of the enrichment analysis demonstrated their important roles in recruitment in the pre-replicative complex (pre-RC) during the initiation of DNA replication (Figure 3F). Furthermore, in both D458 and D283 cells, the protein levels of MCM7 and CDC45 significantly decreased as with MCM3 knockdown (Figure 3G). MCM2 is more stable than other MCMs and might maintain its protein level by reducing cytoplasmic proteolysis, or some other mechanism (Sedlackova et al., 2020). Therefore, MCM3 might play a role in MB via the dysfunction of pre-replicative complex and CDC45-MCM-GINS (CMG) helicases formed by MCM2, MCM3, MCM7, and CDC45.

MCM3 expression is related to clinical parameters

Table 1 includes data from 628 patients with prognosis data from GSE85217 and 62 patients with prognosis and RNA-seq data from our cohort. MCM3 expression did not change significantly with age (Figure 4A) or between sexes (Figure 4B); however, children with metastasis showed lower MCM3 expression levels (Figure 4C). Moreover, MCM3 was significantly correlated with histology and molecular subtypes (Figures 4D,E). Thus, more malignant histology and molecular subgroups, such as Large Cell and Anaplastic (LCA), Sonic Hedgehog (SHH)-MB, and Group3-MB, showed higher expression levels of MCM3. In addition, the patients from GSE85217 were classified into high and low-risk groups according to the cut-off value determined by ROC analysis after excluding adult samples. The low-MCM3 group showed better overall survival (Figure 4F), which was validated in our cohort (Figure 4G). Furthermore, our cohort showed that patients with high MCM3 expression had a high risk of metastasis (Figure 4H). In addition, the Sankey plots also revealed higher proportions of low-risk patients in WNT and Group4 MB, and better prognosis in the low-risk group (Figure 3I). Therefore, MCM3 expression was related to current major clinical parameters and had potential clinical significance.

MCM3 dysregulation remodels the immune microenvironment and affects multiple cell death-related processes

To systematically investigate the effect of MCM3 dysregulation in MB, we first analyzed tumor-infiltrating cells via three algorithms, including ESTIMATE, CIBERSORT, and XCELL. The results revealed the differential infiltration of immune cells, such as cytotoxic lymphocytes, fibroblasts, CD4 T cells, macrophages, etc. (Supplementary Figure S2). MCM3 was correlated with varied components of the immune microenvironment based on the ESTIMATE score (Figure 5A). Subgroup analysis of four molecular subgroups revealed that the varied immune components were associated with OS in different subtypes, such as fibroblasts in SHH (Figure 5B), neutrophils and stromal score in SHH (Supplementary Figures S3A,B), neutrophils in WNT MB (Supplementary Figure S3C), and stromal and estimate scores

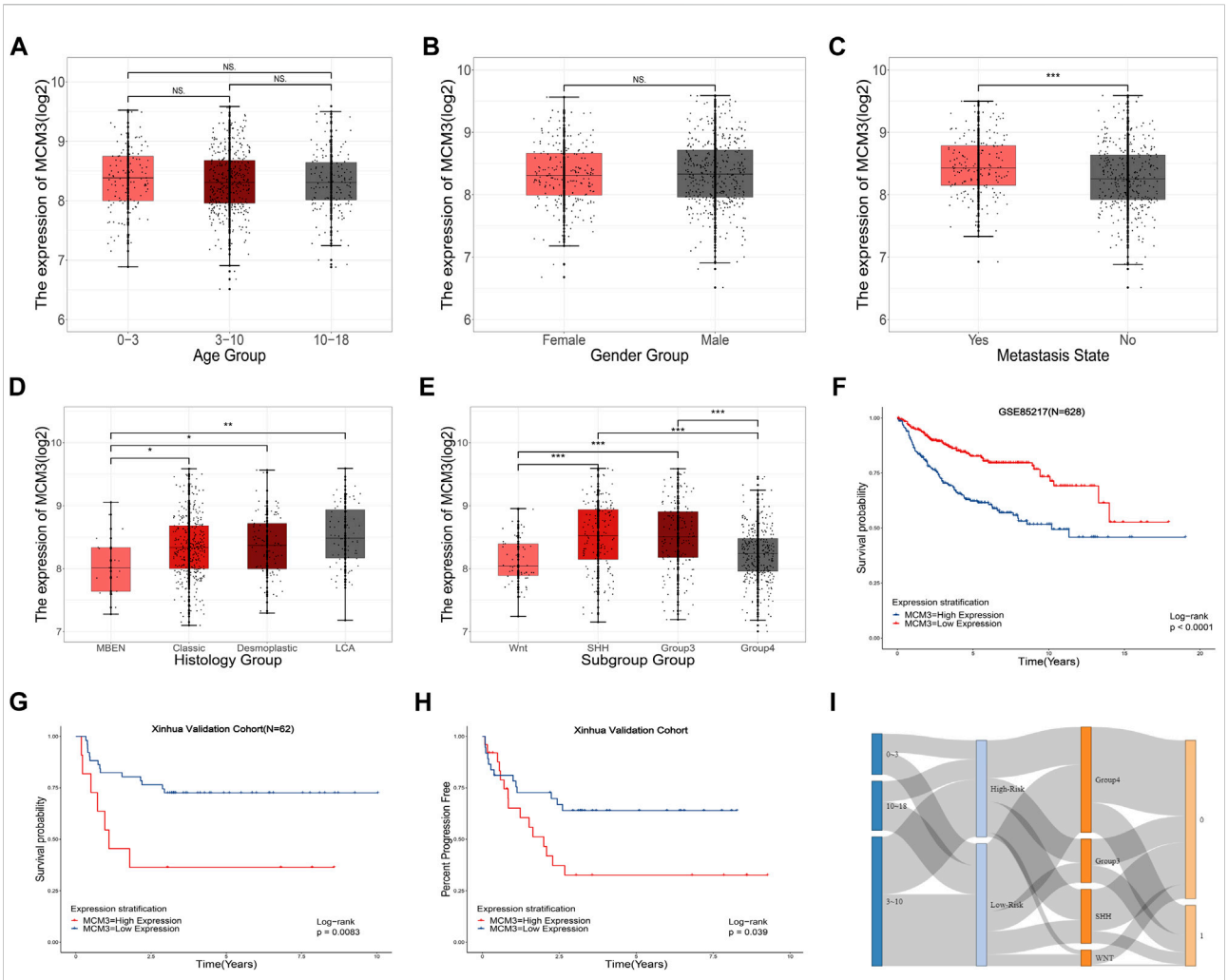


FIGURE 4
Relationships between MCM3 expression and clinicopathologic features. Correlation of MCM3 expression with age **(A)**, sex **(B)**, metastasis status **(C)**, histology **(D)** and molecular subtypes **(E)**. MCM3 expression is correlated with OS in samples from children (GSE85217, *n* = 628) **(F)**. Validation of the prognostic value of MCM3 in our patient cohort **(G)**. Correlation of MCM3 expression level with metastasis in our cohort **(H)**. Relationships among molecular subtype, MCM3 risk classification, and survival state **(I)**.

in Group4 MB(Supplementary Figures S3D,E). In addition, other biological processes were also analyzed, including apoptosis, autophagy, and ferroptosis (Figure 5C). First, the DEGs and MCM3-correlated genes were analyzed with the three process-related genes to obtain the intersection genes; for example, CDK5RAP3 and TOP2A in apoptosis, DNM3, GABARAPL1, GABBR2, OPTN (Supplementary Figures S3F–I), SCNA and MAPK10 in autophagy (Figure 5C), and MT3 in ferroptosis (Figure 5C). Only DNM3, GABARAPL1, GABBR2, SNCA (Supplementary Figures S3J -L), CDK5RAP3, TOP2A, SCNA and MAPK10 (Figure 5C) were associated with OS. Therefore, MCM3 might affect the development of MB *via* these genes.

Abnormal MCM3 demethylation may contribute to its expression dysregulation

Considering the epigenetic disorders in pediatric brain tumors (Lin et al., 2016; Northcott et al., 2017; Petralia et al., 2020), we investigated the cause of MCM3 dysregulation at the methylation level. The PCA and heatmap revealed significant differences between the tumor and normal groups (Figure 6A), while the heatmaps demonstrated the acceptable quality control (Figures 6B–C). MCM3 probes were then obtained according to the beadchip annotation. Among the 14 probes for MCM3, probes 7 and 13 were filtered, respectively, after data normalization of GSE85212 and GSE54880. “cg02243303” and

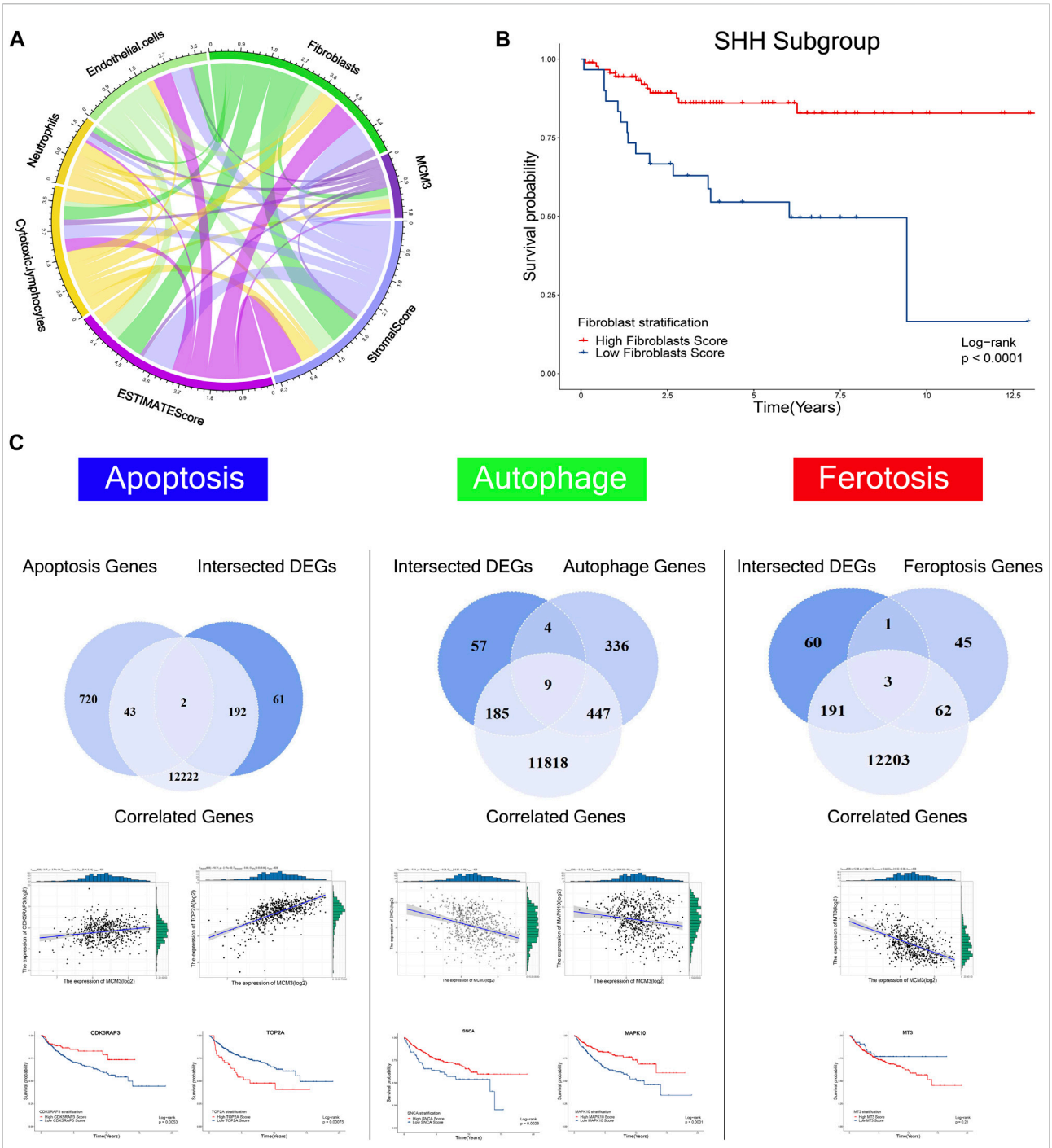


FIGURE 5
Correlations of MCM3 with fibroblasts, endothelial cells, neutrophils, cytotoxic lymphocytes, estimate score, and stromal score **(A)**. Association of fibroblasts with OS in the SHH subtype **(B)**. Kaplan–Meier analysis of genes related to apoptosis, autophagy, and ferroptosis that were correlated with MCM3 and differentially expressed. CDK5RAP3, TOP2A, SNCA, MAPK10 are correlated with MCM3 and associated with OS **(C)**.

"cg21858961" differed significantly between normal and tumor groups and were associated with OS (Figures 6D–E). Moreover, MCM3 expression was correlated with the methylation level of the two sites (Figures 7F–G). Therefore, "cg02243303" and "cg21858961" hypomethylation might be involved in MCM3 regulation in MB.

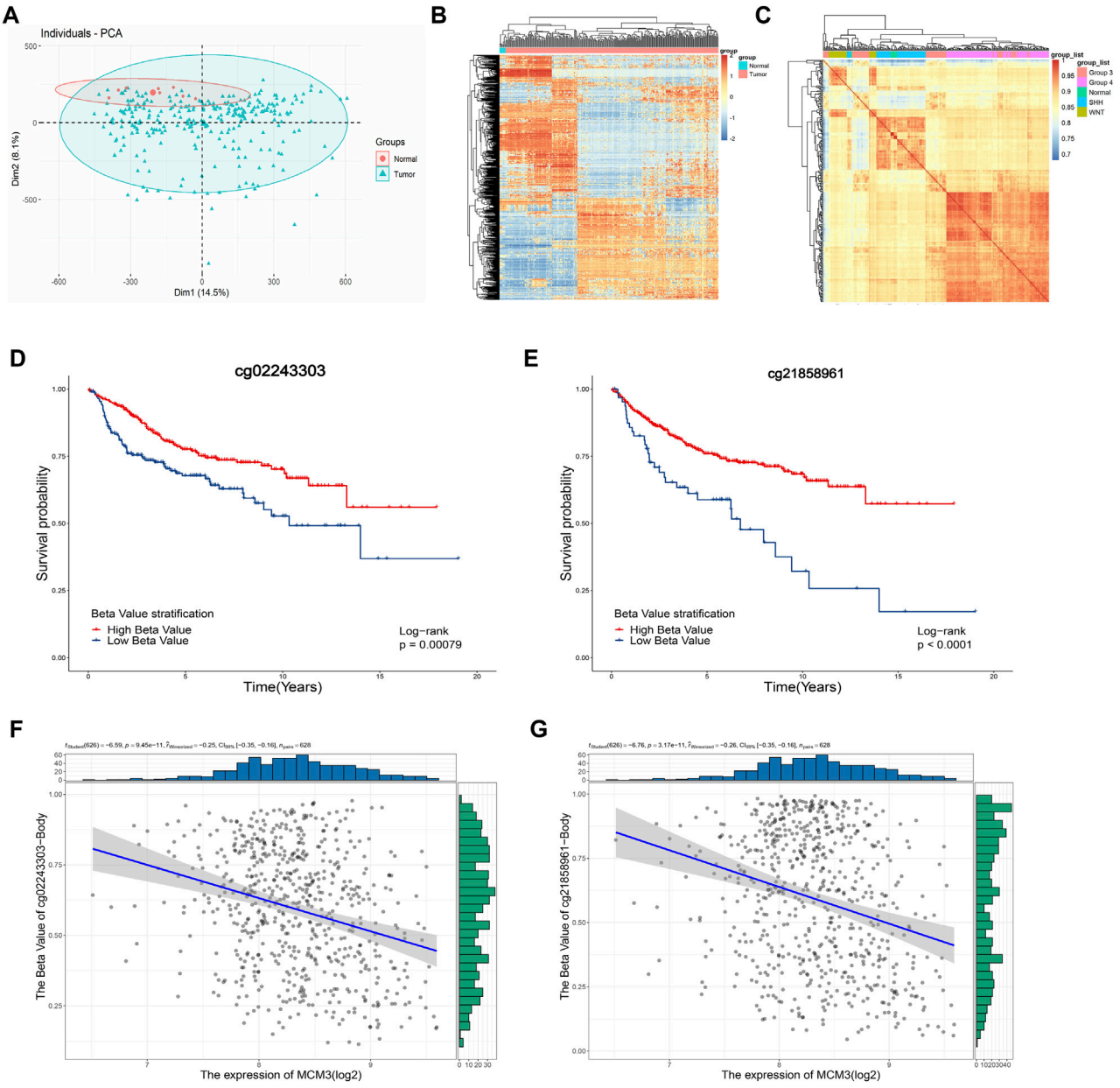


FIGURE 6
The PCA and heatmap showing the significant difference between normal and tumor samples **(A,B)** and the high correlation within groups **(C)**. The "cg02243303" and "cg2185896" probes are significantly associated with OS **(D,E)** and negatively correlated with MCM3 expression **(F,G)**.

The prognostic model based on MCM3-related genes performs better than current tumor classifications

We next investigated the clinical significance of MCM3. First, given its extensive influence, genes correlated with MCM3 (11,008 genes) and associated with OS (RAP2B, ARHGEF40, ADGRG6, ALS2CL, FZD4, TJP2, EIF2AK3, FBLIM1, DGLUCY, C6orf141, APLN, FCRL1, ZCCHC13, ZMYND15, FAM163B, LBHD1, UEVLD) were included in univariate and lasso regression analyses to construct gene signatures (Supplementary Figures S4A,B). Kaplan–Meier analysis revealed that a low risk score showed a dramatically longer OS compared to that for a high risk score (Figure 7A). The testing data set for internal validation (Supplementary Figure S4C) and our cohort data (Supplementary Figure S4D) also confirmed the prognostic value of the gene

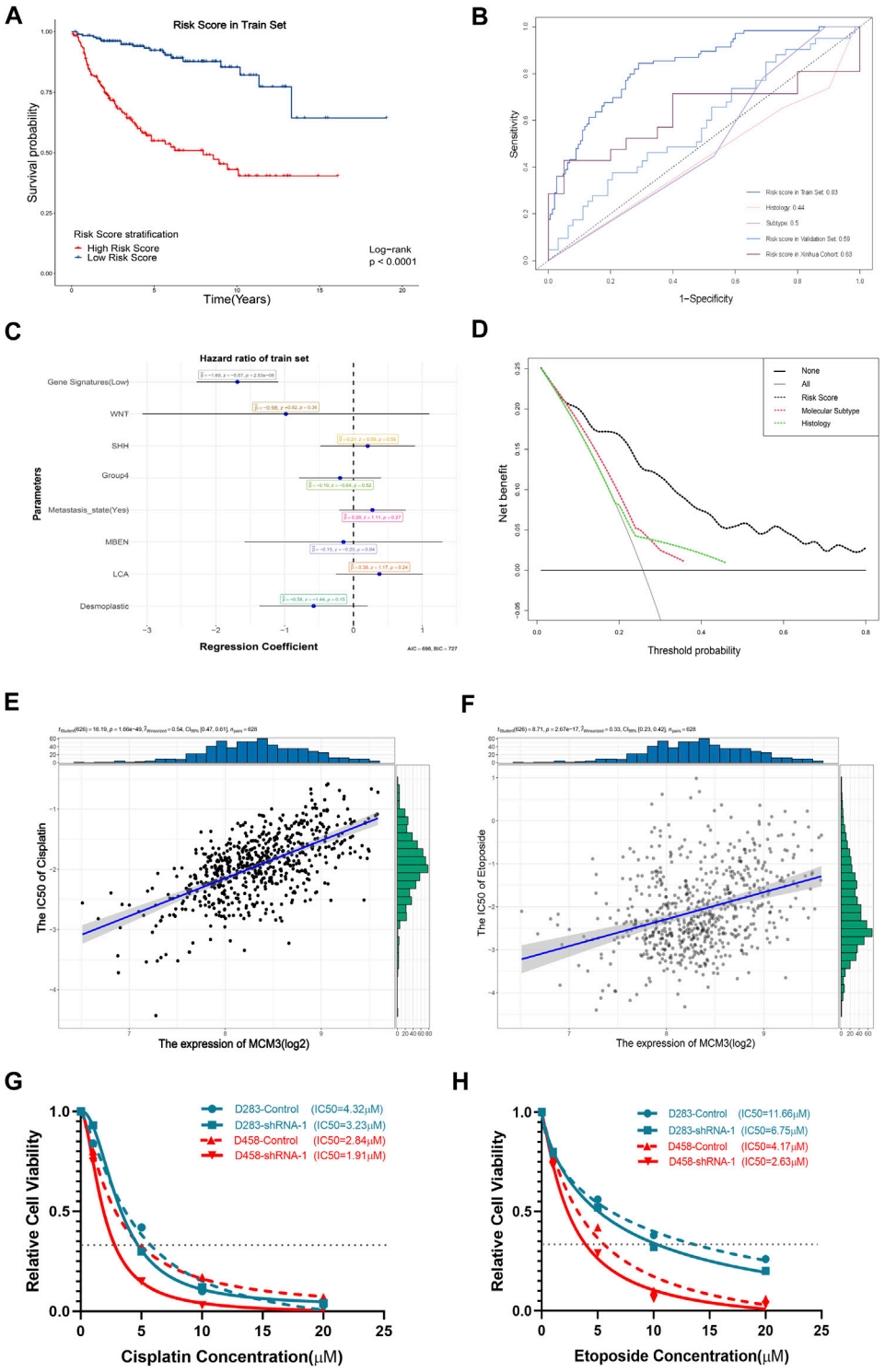


FIGURE 7
Significant association of the gene signature based on MCM3-related genes with OS **(A)**. The AUC values of the training set, internal validation set, and our cohort are higher than those for histological classification and molecular subtyping **(B)**. The forest plot showed that the gene signature was the only independent factor **(C)**. DCA demonstrating the better performance of the risk score compared to the current histological classification and molecular subtyping **(D)**. Positive expression of MCM3 expression with sensitivity to etoposide **(E)** and cisplatin **(F)**. Decreased IC_{50} values of cisplatin **(G)** and etoposide **(H)** in the sh-MCM3 groups in both D283 and D458 cell lines.

signature. The AUC of 5-year survival was 0.83 (Figure 7B). Although the AUC values of the validation set and our cohort were relatively lower, all were significantly higher than those of histological and molecular subtypes (Figure 7B). Moreover, the multivariate Cox regression analysis revealed that only the risk score was the independent factor (Figure 7C). In addition, the nomogram model predicting 1-year and 5-year probabilities can be explored on the website (https://cll12345.shinyapps.io/DynNomapp/). The C-index of the nomogram was 0.784 (CI: 0.723–0.831), suggesting its reliability. Moreover, decision curve analysis (DCA) of the nomogram showed that the prognostic model performed better than the current strategies for MB classification (Figure 7D). Finally, owing to the predominance of conventional chemotherapies in the treatment of MB, we evaluated the correlation between MCM3 expression and drug sensitivities. Patients with low MCM3 expression were more sensitive to etoposide and cisplatin treatment (Figures 7E,F). Furthermore, we validated the decrease in the IC50 of etoposide and cisplatin by knocking down MCM3 expression in both D283 and D458 cell lines (Figures 7G,H). Therefore, MCM3 might be used to guide prognostic assessment and MCM3 targeted therapy might be a new potential strategy to reduce chemotherapy doses, which is of great significance for individualized chemotherapy in children.

Discussion

The results of this study demonstrated the dysregulation of MCM3 in most common cancers and showed its expression level in MB cell lines of pediatric brain tumors, which indicated its potential correlation with tumorigenesis. The CRISPR screening data, covering multiple cancer cell lines from the BioGRID ORCS database, also indicated its essential role in malignant cells and potential relationships with apoptosis, autophagy, and ferroptosis. We then showed MCM3 overexpression in MB, its core place in DEGs, and the biological process it drove at bulk and single-cell RNA-seq levels. We also investigated the correlations between MCM3 and clinical features, finding that the MCM3 expression level was related to high-risk clinicopathologic and molecular subtypes and poor prognosis. We further performed GSEA based on high and low MCM3 expression levels and found that MCM3 might promote tumorigenesis through the dysfunction of the pre-replicative complex and CDC45-MCM-GINS (CMG) helicases formed by MCM2, MCM3, MCM7, and CDC45. Considering the reported relationships with immune response, apoptosis, autophagy, and ferroptosis, we also found that MCM3 was correlated with immune microenvironment components and might affect genes related to the above biological processes, such as CDK5RAP3, TOP2A, OPTN, MAPK10, etc., which indicated that MCM3 might affect tumorigenesis through a variety of mechanisms. In addition, we investigated the mechanisms of MCM3 dysregulation at the DNA methylation level and identified two differential sites that were associated with OS and correlated with MCM3 expression, which indicated that abnormal methylation might result in MCM3 dysregulation. Finally, we discovered the correlation of MCM3 expression with sensitivity to chemotherapy medications, including etoposide and cisplatin.

MCM3, a member of the MCM family of DNA-dependent ATPases that bind to replication origins and support a single round of DNA replication, has demonstrated dysfunction in most cancers. As shown in Figure 1, MCM3 is upregulated in various tumors compared to normal tissues. MB is a highly heterogeneous tumor with the highest incidence and malignancy. Molecular subtypes have been described as a reference for prognosis and individual therapy; however, the high costs of this analysis limit its popularity in primary medical care. Thus, there is an urgent need to identify low-cost biomarkers to guide clinical decision-making. As reported in other tumors, MCM3 is also upregulated in MB based on bulk RNA-seq data. Moreover, we also investigated its expression at the single-cell level. As shown in Figure 2, MCM3 was overexpressed in tumor cells and might be related to malignant transformation.

The present study also evaluated the impact of MCM3 dysregulation according to previous reports (Söling et al., 2005; Puustinen et al., 2020). First, we studied the association of MCM3 expression within the MB microenvironment via different methods. We found the differential immune infiltration between high and low MCM3 expression levels, including higher infiltration levels of cytotoxic lymphocytes in the high MCM3 group and higher infiltration levels of fibroblasts and M2 macrophages in the low MCM3 group. Moreover, immune components such as neutrophils and fibroblasts, as well as stromal and estimate scores, were associated with OS in MB subtypes, indicating the potential role of MCM3 by affecting immune infiltration. We also evaluated the association of MCM3 with apoptosis, autophagy, and ferroptosis-related genes, in which some differentially expressed genes were correlated with MCM3 and associated with OS. Therefore, MCM3 might also influence prognosis via these genes, which requires further robust experimental verification.

Considering the low mutation burden of pediatric brain tumors, we investigated the mechanism of MCM3 expression dysregulation. As expected, 2 of 13 sites in MCM3 showed hypermethylation in MB and were associated with OS. Moreover, the hypermethylation state was significantly correlated with MCM3overexpression. Previous studies also reported that MCM3 hypomethylation could increase its expression in hepatocellular carcinoma (Hua et al., 2020) and osteosarcoma (Zhou et al., 2021), and was negatively associated with prognosis. We also identified that "double-strand break repair *via* break-induced replication" might be the most affected biological processes in GSEA of MCM3 correlated and differential genes between the high and low MCM3 expression groups (Figure 3B). Meanwhile, MCM7 and CDC45, which were

involved in this process, were correlated with MCM3 and differentially expressed between normal and tumor tissues.

Our results also verified that MCM3 expression was correlated with other clinical parameters, especially molecular subtypes and OS, with low-risk classification based on MCM3 expression commonly observed in the WNT MB and surviving groups (Figure 4). This indicated that MCM3 might be a low-cost biomarker for MB risk classification. MCM3 could also guide for chemotherapy selection according to its robust correlation ($r^2 = 0.54$ and $r^2 = 0.33$) with cisplatin and etoposide, which might contribute to individualized therapy to reduce their toxic side effects in children (Figures 7E,F).

Despite the results, this study has several limitations. First, while our results demonstrated the correlations between MCM3 and multiple biological processes through bioinformatics analysis, further validation studies are needed to reveal the crosstalk between them. Second, further studies of the MCM-associated mechanisms via CMG helicases might also provide a direction for novel target development. In addition, although our prognostic model demonstrated a better performance than current histological and molecular subtypes, further clinical translation is needed.

Conclusion

Overall, we systematically studied abnormal MCM3 expression and affected biological processes in bulk and single-cell RNA-seq. Our results showed MCM3 overexpression in MB and its relationships to clinical parameters. We also discovered that MCM3 might affect varied biological processes by aberrant methylation of MCM3 and dysregulation of the complex consisting of MCM2, MCM3, MCM7, and CDC45. Importantly, our results indicated that MCM3 might serve as a potential biomarker of prognosis prediction and better guidance compared to current histological and molecular classification systems for clinical decision-making.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of Xin Hua Hospital affiliated with Shanghai Jiao Tong University School of Medicine (Approval No. XHEC-D-2021-076). Written informed consent to participate in this study obtained from the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin for the publication of any potentially identifiable images or data included in this article.

Author contributions

CL: methodology, visualization, and writing of the original draft. FJ: survival analysis and methodology; ZL: survival analysis; JY: methodology, formal analysis, writing review and editing. JW: methodology, formal analysis, writing review and editing. ST: software and visualization. QW: collection of tumor samples; BW: writing review and editing; YZ: writing review and editing, funding acquisition; JM: writing review and editing, funding acquisition.

Funding

This research was supported by the National Science Foundation of China for Young Scholars (No. 81702453 to YZ), Shanghai Science and Technology Committee (No. 17411965700 to YZ, No. 17411951800 and No. 19411952100 to JM), the Joint Research of Medicine and Industry of Shanghai Jiao Tong University (No. YG2015QN42 to YZ), Shanghai Xin Hua Hospital (No. JZPI201701 to JM), the Shanghai Shen Kang Hospital Development Center (No. 16CR 2031B to JM) and Shanghai Leading Talent Program.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.815260/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Dysregulation of MCM3 in most cancers compared to corresponding normal tissues . MCM3 upregulation in the tumor group compared to adult normal samples from GSE42656 **(B)**. MCM3 upregulation in the tumor group compared to fetal normal samples from GSE42656 **(C)**. The adult normal and medulloblastoma samples from GSE42656 clustered according to the expression of 255-DEGs **(D)**. The fetal normal and medulloblastoma samples from GSE42656 clustered respectively according to the expression of 255-DEGs **(E)**. The core position of MCM3 among DEG interactions **(F)**. Ranking of MCM3 within the top three hub genes **(G)**.

SUPPLEMENTARY FIGURE S2
Estimation of tumor-infiltrating cells via three algorithms, including ESTIMATE, CIBERSORT, and XCELL, showing differential infiltration of immune cells, including cytotoxic lymphocytes, fibroblasts, CD4 T cells, macrophages, etc.

SUPPLEMENTARY FIGURE S3
Association of neutrophils with OS in the SHH subtype . Association of stromal score with OS in the SHH subtype **(B)**. Association of neutrophils with OS in WNT MB **(C)**. Association of stromal score with OS in the Group4 subtype **(D)**. Associations of estimate score with OS in Group4 MB **(E)**. Negative correlations of the expression of autophagy-related genes (DNM3, GABARAPL1 GABBR2, and OPTN) with MCM3 expression **(F–I)**. Relationship of the expression of autophagy-related genes (DNM3, GABBR2 and GABARAPL1) with OS **(J–L)**.

SUPPLEMENTARY FIGURE S4
Lasso regression analysis with MCM3-correlated genes in GSE85217, in which 17 genes constituted the multi-gene signature . Association of the risk score with OS in the internal validation set **(C)**. Association of the risk score with OS in our cohort **(D)**.

References

Andrew, F., Paul, G., Lechner, M., Fenton, T., Wilson, G. A., Thirlwell, C., et al. (2014). Using high-density DNA methylation arrays to profile copy number alterations. *Genome Biol.* 15, R30. doi:10.1186/gb-2014-15-2-r30

Aran, D. A.-O., Looney, A. P., Liu, L., Wu, E., Fong, V., Hsu, A., et al. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* 20, 163–172. doi:10.1038/s41590-018-0276-y

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

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. A.-O. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36, 411–420. doi:10.1038/nbt.4096

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

Duc, N. A.-O., Huy, H. Q., Nadarajan, C., and Keserci, B. (2020). The role of predictive model based on quantitative basic magnetic resonance imaging in differentiating medulloblastoma from ependymoma. *Anticancer Res.* 40, 2975–2980. doi:10.21873/anticanres.14277

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

Ghandi, M., Huang, F. W., Jane-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

Gu, C., Wang, W., Tang, X., Xu, T., Zhang, Y., Guo, M., et al. (2021). CHEK1 and circCHEK1_246aa evoke chromosomal instability and induce bone lesion formation in multiple myeloma. *Mol. Cancer* 20, 84. doi:10.1186/s12943-021-01380-0

Hovestadt, V., Ayrault, O., Swartling, F. J., Robinson, G. W., Pfister, S. M., and Northcott, P. A. (2020). Medulloblastomics revisited: biological and clinical insights from thousands of patients. *Nat. Rev. Cancer* 20, 42–56. doi:10.1038/s41568-019-0223-8

Hua, S., Ji, Z., Quan, Y., Zhan, M., Wang, H., Li, W., et al. (2020). Identification of hub genes in hepatocellular carcinoma using integrated bioinformatic analysis. *Aging (Albany NY)* 12, 5439–5468. doi:10.18632/aging.102969

Iglesias-Gato, D., Thysell, E., Tyanova, S., Crnalic, S., Santos, A., Lima, T. S., et al. (2018). The proteome of prostate cancer bone metastasis reveals heterogeneity with prognostic implications. *Clin. Cancer Res.* 24, 5433–5444. doi:10.1158/1078-0432.CCR-18-1229

Lannering, B., Rutkowski S Fau - Doz, F., Doz F Fau - Pizer, B., Pizer B Fau - Gustafsson, G., Gustafsson G Fau - Navajas, A., Navajas A Fau - Massimino, M., et al. (2012). Hyperfractionated versus conventional radiotherapy followed by chemotherapy in standard-risk medulloblastoma: results from the randomized multicenter HIT-SIOP PNET 4 trial. *J. Clin. Oncol.* 30, 3187–3193. doi:10.1200/JCO.2011.39.8719

Lin, C. Y., Erkek, S., Tong, Y., Yin, L., Federation, A. J., Zapatka, M., et al. (2016). Active medulloblastoma enhancers reveal subgroup-specific cellular origins. *Nature* 530, 57–62. doi:10.1038/nature16546

Madine, M. A., Khoo, C. Y., Mills, A. D., and Laskey, R. A. (1995). MCM3 complex required for cell cycle regulation of DNA replication in vertebrate cells. *Nature* 375, 421–424. doi:10.1038/375421a0

Madine, M. A., Swietlik, M., Pelizon, C., Romanowski, P., Mills, A. D., and Laskey, R. A. (2000). The roles of the MCM, ORC, and Cdc6 proteins in determining the replication competence of chromatin in quiescent cells. *J. Struct. Biol.* 129, 198–210. doi:10.1006/jsbi.2000.4218

McCarthy, C., Gupta, N., Johnson, S. R., Yu, J. J., McCormack, F. X., Gu, C., et al. (2021). Lymphangioleiomyomatosis: pathogenesis, clinical features, diagnosis, and management. *Lancet. Respir. Med.* 9, 1313–1327. doi:10.1016/S2213-2600(21)00228-9

Minh Thong, P., and Minh Duc, N. A.-O. (2020). The role of apparent diffusion coefficient in the differentiation between cerebellar medulloblastoma and brainstem glioma. *Neurol. Int.* 12, 34–40. doi:10.3390/neurolint12030009

Musahl, C., Holthoff Hp Fau - Lesch, R., Lesch R Fau - Knippers, R., and Knippers, R. (1998). Stability of the replicative Mcm3 protein in proliferating and differentiating human cells. *Exp. Cell Res.* 241, 260–264. doi:10.1006/excr.1998.4041

Newman, A. C., Falcone, M., Huerta Uribe, A., Zhang, T., Athineos, D., Pietzke, M., et al. (2021). Immune-regulated IDO1-dependent tryptophan metabolism is source of one-carbon units for pancreatic cancer and stellate cells. *Mol. Cell* 81, 2290–2302. doi:10.1016/j.molcel.2021.03.019

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

Northcott, P. A., Buchhalter, I., Morrissy, A. S., Hovestadt, V., Weischenfeldt, J., Ehrenberger, T., et al. (2017). The whole-genome landscape of medulloblastoma subtypes. *Nature* 547, 311–317. doi:10.1038/nature22973

Oughtred, R., Stark, C., Breitkreutz, B. J., Rust, J., Boucher, L., Chang, C., et al. (2019). The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* 47, D529–D541. doi:10.1093/nar/gky1079

Park, H., Lee, S., Shrestha, P., Kim, J., Park, J. A., Ko, Y., et al. (2015). AMIGO2, a novel membrane anchor of PDK1, controls cell survival and angiogenesis via Akt activation. *J. Cell Biol.* 211, 619–637. doi:10.1083/jcb.201503113

Petralia, F., Tignor, N., Reva, B., Koptyra, M., Chowdhury, S., Rykunov, D., et al. (2020). Integrated proteogenomic characterization across major histological types of pediatric brain cancer. *Cell* 183, 1962–1985. doi:10.1016/j.cell.2020.10.044

Puustinen, P., Keldsbo, A., Corcelle-Termeau, E., Ngoei, K. A.-O., Sønder, S. L., Farkas, T., et al. (2020). DNA-dependent protein kinase regulates lysosomal AMP-dependent protein kinase activation and autophagy. *Autophagy* 16, 1871–1888. doi:10.1080/15548627.2019.1710430

Qiu, Y., Xu, M., and Huang, S. (2021). Long noncoding RNAs: emerging regulators of normal and malignant hematopoiesis. *Blood* 138 (23), 2327–2336. doi:10.1182/blood.2021011992

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

Schwab, B. L., Leist, M., Knippers, R., and Nicotera, P. (1998). Selective proteolysis of the nuclear replication factor MCM3 in apoptosis. *Exp. Cell Res.* 238, 415–421. doi:10.1006/excr.1997.3850

Sedlackova, H., Rask, M. B., Gupta, R., Choudhary, C., Somyajit, K., and Lukas, J. (2020). Equilibrium between nascent and parental MCM proteins protects replicating genomes. *Nature* 587, 297–302. doi:10.1038/s41586-020-2842-3

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

Shrestha, R., Llaurado Fernandez, M., Dawson, A., Hoenisch, J., Volik, S., Lin, Y. Y., et al. (2021). Multiomics characterization of low-grade serous ovarian carcinoma identifies potential biomarkers of MEK inhibitor sensitivity and therapeutic vulnerability. *Cancer Res.* 81, 1681–1694. doi:10.1158/0008-5472.CAN-20-2222

Söling, A., Sackewitz, M., Volkmar, M., Schaarschmidt, D., Jacob, R., Holzhausen, H-J., et al. (2005). Minichromosome maintenance protein 3 elicits a cancer-restricted immune response in patients with brain malignancies and is a strong independent predictor of survival in patients with anaplastic astrocytoma. *Clin. Cancer Res.* 11, 249–258. doi:10.1158/1078-0432.249.11.1

Sonja, H., Robert, C., and Justin, G. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* 14, 7. doi:10.1186/1471-2105-14-7

Stewart, P. A., Khamis, Z. I., Zhau, H. E., Duan, P., Li, Q., Chung, L. W. K., et al. (2017). Upregulation of minichromosome maintenance complex component 3 during epithelial-to-mesenchymal transition in human prostate cancer. *Oncotarget* 8, 39209–39217. doi:10.18632/oncotarget.16835

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. U. S. A.* 102, 15545–15550. doi:10.1073/pnas.0506580102

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

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

Tsherniak, A., Vazquez, F., Montgomery, P. G., Weir, B. A., Kryukov, G., Cowley, G. S., et al. (2017). Defining a cancer dependency map. *Cell* 170, 564–576. doi:10.1016/j.cell.2017.06.010

von Bueren, A. O., Kortmann, R. D., von Hoff, K., Friedrich, C., Mynarek, M., Müller, K., et al. (2016). Treatment of children and adolescents with metastatic medulloblastoma and prognostic relevance of clinical and biologic parameters. *J. Clin. Oncol.* 34, 4151–4160. doi:10.1200/JCO.2016.67.2428

Wang, L., Guo, J., Zhou, J., Wang, D., Kang, X., and Zhou, L. (2020). NF-κB maintains the stemness of colon cancer cells by downregulating miR-195-5p/497-5p and upregulating MCM2. *J. Exp. Clin. Cancer Res.* 39, 225. doi:10.1186/s13046-020-01704-w

Wang, N. N., Dong, J., Zhang, L., Ouyang, D., Cheng, Y., Chen, A. F., et al. (2018). HAMdb: a database of human autophagy modulators with specific pathway and disease information. *J. Cheminform.* 10, 34. doi:10.1186/s13321-018-0289-4

Yoshihara, K., Shahmoradgoli, M., Martinez, 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

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 15, 284–287. doi:10.1089/omi.2011.0118

Zhang, M., Wong, S. W., Wright, J. N., Wagner, M. W., Toescu, S., Han, M., et al. (2022). MRI radiogenomics of pediatric medulloblastoma: a multicenter study. *Radiology* 304 (2). doi:10.1148/radiol.212137

Zhao, Y., Wang, Y., Zhu, F., Zhang, J., Ma, X., and Zhang, D. (2020). Gene expression profiling revealed MCM3 to be a better marker than Ki67 in prognosis of invasive ductal breast carcinoma patients. *Clin. Exp. Med.* 20, 249–259. doi:10.1007/s10238-019-00604-4

Zhou, J., Wang, M., Zhou, Z., Wang, W., Duan, J., and Wu, G. (2021). Expression and prognostic value of MCM family genes in osteosarcoma. *Front. Mol. Biosci.* 8, 668402. doi:10.3389/fmolb.2021.668402

Zhou, N., and Bao, J. (2020). FerrDb: a manually curated resource for regulators and markers of ferroptosis and ferroptosis-disease associations. *Database (Oxford)* 2020, baaa021. doi:10.1093/database/baaa021

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