frontiers | Research Topics

Bioinformatics and computational approaches for the development of innovative genetic and cellular therapies

Edited by
Aurélie Carlier, Ping Zhang, Giulia Russo, Francesco Pappalardo and Liesbet Geris

Published in
Frontiers in Bioengineering and Biotechnology

ISSN 1664-8714
ISBN 978-2-83251-136-7
DOI 10.3389/978-2-83251-136-7

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

Bioinformatics and computational approaches for the development of innovative genetic and cellular therapies

Topic editors

Aurélie Carlier — Maastricht University, Netherlands
Ping Zhang — Griffith University, Australia
Giulia Russo — University of Catania, Italy
Francesco Pappalardo — University of Catania, Italy
Liesbet Geris — University of Liège, Belgium

Citation

Carlier, A., Zhang, P., Russo, G., Pappalardo, F., Geris, L., eds. (2023). *Bioinformatics and computational approaches for the development of innovative genetic and cellular therapies*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-136-7

Table of contents

frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 08 October 2021
doi: 10.3389/fbioe.2021.717234

Prioritization of Diagnostic and Prognostic Biomarkers for Lupus Nephritis Based on Integrated Bioinformatics Analyses

Zhimin Chen[1], Ruilong Lan[2], Keng Ye[1], Hong Chen[3], Caiming Chen[1] and Yanfang Xu[1]**

[1]Department of Nephrology, Blood Purification Research Center, First Affiliated Hospital of Fujian Medical University, Fuzhou, China, [2]Central Laboratory, First Affiliated Hospital of Fujian Medical University, Fuzhou, China, [3]Department of Pathology, First Affiliated Hospital of Fujian Medical University, Fuzhou, China

Lupus nephritis (LN) is an important driver of end-stage renal disease (ESRD). However, few biomarkers are available for evaluating the diagnosis and prognosis of LN. For this study, we downloaded microarray data of multiple LN expression profiles from the GEO database. We used the WGCNA and R limma packages to identify LN hub genes and differentially-expressed genes (DEGs). We identified nine co-DEGs in the intersection with LN-related genes from the Genecards database. We found DEGs that are primarily associated with immune-related functions and pathways (including with the complement pathway, primary immunodeficiency markers, and MHC-like protein complexes) through our comprehensive GSEA, GO, and KEGG enrichment analyses. We used other LN and SLE validation datasets and discovered six explicitly expressed co-DEGs: *HLA-DMA*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DRA*, *IL10RA*, and *IRF8* in the LN set; ROC and Precision-Recall curve analyses revealed that these six genes have a good diagnostic efficacy. The correlation analysis with prognostic data from the Nephroseq database indicates that the differential expression of these co-DEGs is associated with a low glomerular filtration rate in that cohort. Additionally, we used a single-cell LN database of immune cells (for the first time) and discovered these co-DEGs to be predominantly distributed in different types of macrophages and B cells. In conclusion, by integrating multiple approaches for DEGs discovery, we identified six valuable biomarkers that are strongly correlated with the diagnosis and prognosis of LN. These markers can help clarify the pathogenesis and improve the clinical management of LN.

Keywords: lupus nephritis, bioinformatics, biomarkers, diagnosis, prognosis, WGCNA

OPEN ACCESS

Edited by:
Giulia Russo,
University of Catania, Italy

Reviewed by:
Michele Carrabba,
University of Bristol, United Kingdom
Tim Herpelinck,
KU Leuven, Belgium

****Correspondence:***
Caiming Chen
chencm419@126.com
Yanfang Xu
xuyanfang99@hotmail.com

Specialty section:
This article was submitted to Preclinical Cell and Gene Therapy, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *30 May 2021*
Accepted: *28 September 2021*
Published: *08 October 2021*

Citation:
Chen Z, Lan R, Ye K, Chen H, Chen C and Xu Y (2021) Prioritization of Diagnostic and Prognostic Biomarkers for Lupus Nephritis Based on Integrated Bioinformatics Analyses. Front. Bioeng. Biotechnol. 9:717234. doi: 10.3389/fbioe.2021.717234

INTRODUCTION

Systemic lupus erythematosus (SLE) is a common chronic autoimmune disease with multifactorial causes. SLE affects mainly women of childbearing age and its progression and prognosis are highly heterogeneous. The characteristics of SLE include the production of autoantibodies, the deposition of immune complexes, and impairment of multiple organ systems (Kiriakidou and Ching, 2020). A genetic predisposition, environmental factors, apoptosis abnormalities, infections, the use of certain drugs, and sex hormone levels are factors thought to play a role in the pathogenesis of SLE (Durcan et al., 2019). The kidney is the most commonly affected organ in patients with SLE, and renal biopsies

show a nearly 100% involvement with approximately 45–85% of patients presenting clinical symptoms of lupus nephritis (LN) (Furie et al., 2020). The pathogenesis of LN includes a process of kidney damage caused by immune complex deposition in renal tissues (Anders et al., 2020) due to inflammatory cell recruitment, cytokine production, oxidative damage, complement activation, and abnormal fibroblast proliferation (Davidson et al., 2019). Inflammation and fibrosis are critical for the development of LN because they include the interaction of innate and adaptive immune cells with resident renal cells. Although glucocorticoids and immunosuppressants have been shown to improve survival in patients with LN, the current treatment for this disease remains unsatisfactory (Lech and Anders, 2013). Additionally, the adverse effects of non-specific treatments (including those against infection and renal failure) make a more effective and targeted approach urgent. Thus, conducting additional research into the etiology and pathogenesis of the disease is necessary to further improve the survival of patients with LN.

Bioinformatics is a branch of computer science that focuses on the storage, retrieval, and analysis of biological data. The analysis of massive amounts of data generated by biochips has provided helpful information to help understand molecular disease mechanisms (Wooller et al., 2017). Bioinformatics has been widely used to obtain disease gene expression profiles, to identify disease-related genes and drug targets, and to analyze complex disease pathogenic mechanisms. Craciun *et al.*, for example, used RNA sequencing to characterize the renal transcriptomic profile of specimens in a mouse model of folic acid-induced nephropathy (Craciun et al., 2016). As a result, they identified 10 molecules associated with renal fibrosis, with the levels of *CDH11*, *MRC1*, and *PLTP* being significantly increased in the urine of patients with chronic kidney disease. Köttgen *et al.* used genetic and genotype-population analyses to obtain 67093 European genome-wide SNPs and then performed a GWAS analysis to identify SNP mutations in *UMOD*, a susceptibility gene for chronic kidney disease (Köttgen et al., 2010). Additionally, GWAS studies have helped identify susceptibility genes for diabetic nephropathy and IgA nephropathy (Salem et al., 2019; Sallustio et al., 2019).

In this study, we used bioinformatics approaches to screen for co-differentially-expressed genes (co-DEGs) in LN from multiple LN dataset sources. The purpose of this study was to identify and prioritize diagnostic and prognostic biomarkers for lupus nephritis and to explore the potential pathways and immune cells that are related to the pathogenesis of LN.

MATERIALS AND METHODS

Data Download

We searched the GEO database (https://www.ncbi.nlm.nih.gov/geo/) (Barrett et al., 2013) for human SLE- and LN-related expression profiles using the keywords "lupus nephritis" and "systemic lupus erythematosus." GSE32591 is based on the GPL14663 platform, which includes 29 normal kidney biopsy and 64 LN kidney biopsy samples (Berthier et al., 2012). We extracted data from 17 normal blood samples and 29 peripheral blood samples from SLE patients with lupus nephritis from the GSE99967 dataset (based on the GPL21970 platform) for subsequent analysis (Wither et al., 2018). In addition, we extracted data from seven normal and 14 lupus nephritis samples from the GSE112943 dataset (built on the GPL10558 platform). GSE81622 is based on the GPL10558 platform and includes 25 normal and 15 SLE samples of patients without LN (Zhu et al., 2016). The GSE60681 dataset based on the GPL13497 platform includes data from 11 patients with LN and from 37 control samples in stable phase (Magnusson et al., 2017). Finally, we also searched the Genecards database (https://www.genecards.org/) using the keyword "lupus nephritis" to identify differential genes associated with LN. **Figure 1** illustrates the specific applications used and the workflow for all data in this study.

Data Pre-Processing

We used the Perl language to process the original matrix of GSE32591, GSE99967, GSE112943, GSE81622, and GSE60681. The probe IDs were converted to gene symbols, and empty probes were removed based on the annotation information contained in each platform file. When multiple probes matched the same gene, the average expression value was used to determine the gene's expression level. The Perl script we used to pre-process is detailed in **Supplementary Data S1**.

Weighted Gene Co-expression Network Analysis

We used the R language package WGCNA (Langfelder and Horvath, 2008) to evaluate the GSE99967 expression matrix. We extracted the LN grouping, SLEDIA-2K score, age, and gender data from the original set as input data for the WGCNA. The sample clustering dendrogram was constructed with the hcluster function, and the TOM matrix was constructed using the pickSoftThreshold function to determine the optimal soft threshold. We used candidate power values (1–30) to determine the average connectivity and independence of various modules. Dynamic shear trees were used to identify gene modules. Next, we measured the association between modules and sample traits using gene significance values (GS) and module membership values (MM), and key modules were identified. We set |GS| to >0.3 and |MM| to >0.7 to filter hub genes in accordance with the official WGCNA guidelines and prior application examples to obtain the most relevant genes to the traits in the key module (Langfelder and Horvath, 2008; Tang et al., 2018).

Identification of Co-Differentially-Expressed Genes

The annotated GSE32591 expression matrix was analyzed for differentially-expressed genes (DEGs) using the R limma package (http://www.bioconductor.org/packages/release/bioc/html/limma.html) (Smyth, 2004) setting the |logFC| to >0.75 and the adjusted *p*-value to <0.05 as the criterion. The DEGs were then intersected with the module hub genes identified using WGCNA in the GSE99967 set and the LN-related differential genes used in

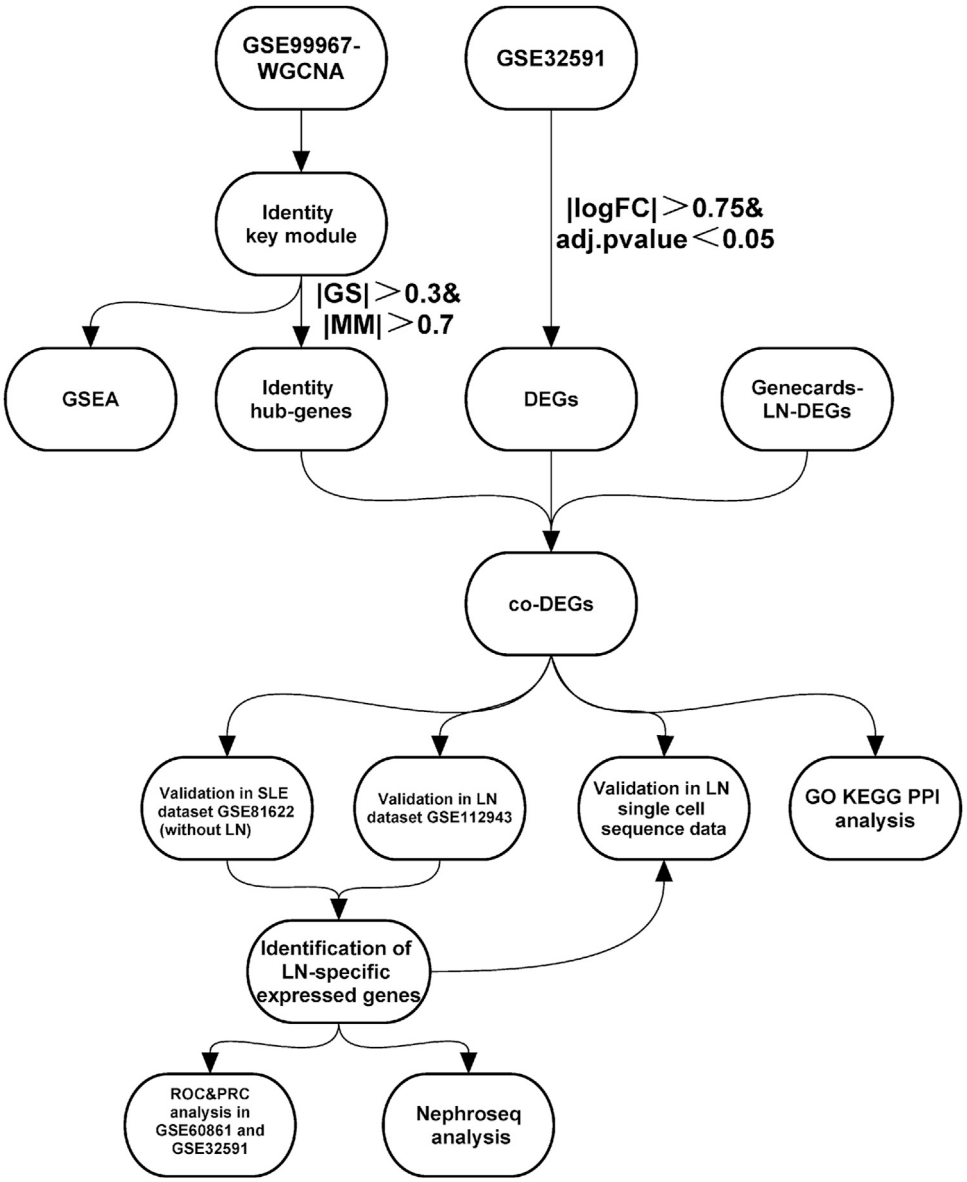


FIGURE 1 | | Study workflow. Abbreviations: Systemic lupus erythematosus (SLE); Lupus nephritis (LN); Weighted gene co-expression network analysis (WGCNA); Gene set enrichment analysis (GSEA); Differentially-expressed genes (DEGs); Gene ontology (GO); Kyoto encyclopedia of genes and genomes (KEGG); Protein-Protein interaction (PPI); Receiver operating characteristic curve (ROC); Precision-Recall curve (PRC).

the Genecards database to identify co-expressed LN differential genes (co-DEGs) across multiple source datasets.

Gene Set Enrichment Analysis

We performed a gene set enrichment analysis (GSEA) (Subramanian et al., 2005) using the KEGG and REACTOME gene sets in the GSEA C2 dataset (c2.cp.kegg.v7.4.symbols.gmt c2.cp.reactome.v7.4.symbols.gmt), the GO gene set in the C5 dataset (c5.go.bp.v7.4.symbols.gmt, c5.go.cc.v7.4.symbols.gmt, c5.go.mf.v7.4.symbols.gmt), and the hallmarker gene set h.all.v7.4.symbols.gmt). |NES| > 1, NOM p-value < 0.05, FDR q-value < 0.25 were set as the screening criteria for enrichment pathways. The results were visualized using the OmicShare (http://www.omicshare.com) cloud platform tool.

Gene Ontology, Kyoto Encyclopedia of Genes and Genomes Enrichment Analysis and Protein-Protein Interaction Network Construction

We performed gene ontology (GO) (Ashburner et al., 2000) and Kyoto encyclopedia of genes and genomes (KEGG) (http://www.genome.jp/kegg/) (Kanehisa and Goto, 2000) enrichment analyses of co-DEGs using the ClueGo (Bindea et al., 2009)

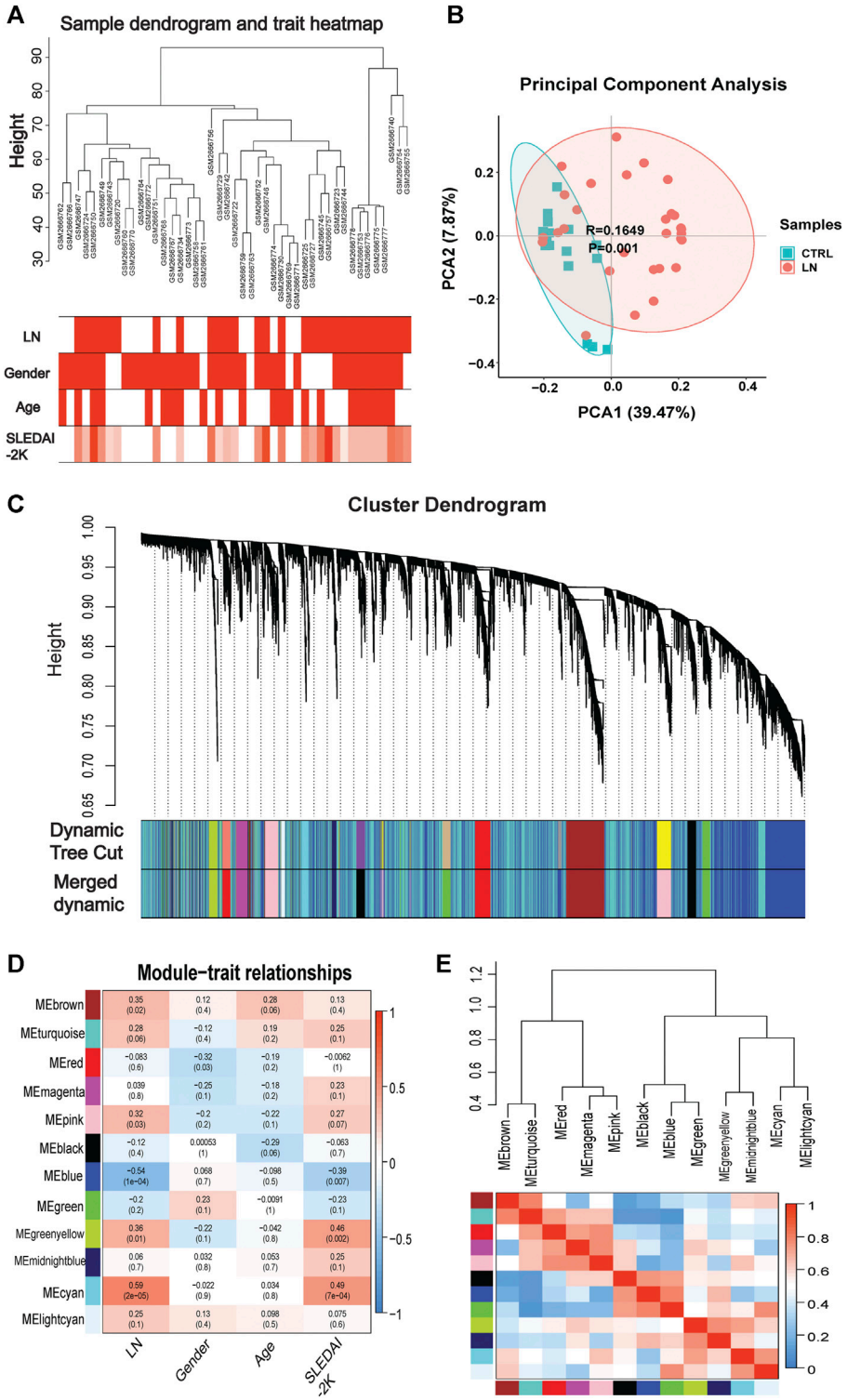


FIGURE 2 | | WGCNA analysis in GSE99968. **(A)** Sample-trait clustering heatmap. **(B)** Principal component analysis (PCA) shows the dimensionality reduction distribution of control and LN sets. **(C)** Dynamic shearing tree merging similar module genes. **(D)** Module-trait correlation heatmap. **(E)** Module-module clustering tree and correlation heatmap.

and the Cluepedia (Bindea et al., 2013) plugins within the CytoScape software (V 3.7.2, http://www.cytoscape.org/) (Cline et al., 2007), setting p to <0.05 to screen the results and construct the target-pathway network. The co-DEGs were submitted to the STRING database (https://www.string-db.org/) (Szklarczyk et al., 2015) to evaluate the interaction between co-DEGs from the protein level; we obtained protein interaction network co-DEGs by setting the confidence level to 0.4.

Validation of Co-Differentially-expressed genes

We validated the expression levels of co-DEGs in the GSE112943 LN expression profile dataset and the GSE81622 SLE expression profile dataset without lupus nephritis to verify that co-DEGs have similar expression profiles in different LN datasets and to verify whether the co-DEGs were specific to LN. Significance tests were performed using the Wilcoxon–Mann–Whitney test, with results visualized using the ggplot2 package (Villanueva and Chen, 2019). We used the PRROC package (Grau et al., 2015) to examine the co-DEGs' diagnostic efficacy by performing Receiver Operating Characteristic (ROC) and Precision-Recall curves (PRC) analyses in the GSE60681 dataset. We validated the distribution of co-DEGs in the published LN single-cell transcriptome sequencing database (https://singlecell.broadinstitute.org/) to explore co-DEGs' distribution in LN immune cells) (Arazi et al., 2019). Finally, we analyzed the association between co-DEGs and clinical features using the Nephroseq database (http://v5.nephroseq.org/) (Zheng et al., 2017). A scatter plot was constructed after calculating the Pearson correlation coefficient between co-DEGs and glomerular filtration rates (GFRs). The Kruskal–Wallis test was used to test the significance of co-DEGs and lupus pathological staging.

RESULTS

Weighted Gene Co-Expression Network Analysis Identifies Key Lupus Nephritis Genes

Genes of interest can be identified by combining gene and clinical trait data and dividing the gene co-expression network of complex biological processes into several highly correlated signature modules that can detect the genes that perform critical functions. As shown by the hierarchical clustering in **Figure 2A**, potential differences between control and LN clusters exist between the different clinical phenotypes. Our principal component analysis (PCA) results show the dimensionality reduction distribution of control and LN sets (**Figure 2B**). We used WGNCA to analyze the GSE99967 expression matrix, with the shear height of the function hcluster set to 100 and an outlier sample GSM266765 excluded (**Supplementary Figure S1**). We calculated the pickSoftThreshold parameter to determine the optimal soft threshold, which is 4 (**Supplementary Figure S2**). The dynamic shear tree's merged shear height was 0.25 for module identification and module merging (**Figure 2C**). The minimum number of genes in each network module was set to 120, resulting in a total of 12 gene modules. The most strongly correlated positive and negative modules were chosen as critical modules for the pathogenesis of the LN and SLEDAI-2K traits. Our results indicate that the blue module was significantly negatively correlated with the LN trait, while the cyan module was significantly positively correlated with the SLEDAI-2K trait (**Figures 2D,E**). These two modules were identified as critical modules, and when |GS|>0.3 and |MM|>0.7 were used to screen for essential genes, we found 2255 genes in the LN trait and 1,388 genes in the SLEDAI-2K trait.

Gene Set Enrichment Analysis Enrichment Analysis

We performed a comprehensive enrichment analysis of the screened blue and cyan modules to discover the functions or pathways associated with LN using the GSEA software and exploring the functions and pathways of the key modules. Our results show that the key modules were mainly enriched for GO entries (including antimicrobial humoral response, defense response to fungus), and KEGG pathways (including cell cycle, P53 signaling pathway, systemic lupus erythematosus, and primary immunodeficiency). The hallmark entries included G2M checkpoint, mitotic spindle, and REACTOME pathways including the RNA pol-I promoter opening and meiotic recombination pathways (**Supplementary Figure S3**). **Figures 3A,B** present two GSEA pathways that are highly associated with lupus pathogenesis. The complete GSEA enrichment results are presented in **Supplementary Table S1**.

Identification of Differentially-Expressed Genes and Screening of Co-Differentially-Expressed Genes

In our screen for critical LN genes, we used a variety of methods to obtain differentially-expressed LN genes. The differential genes in GSE32591 were screened using the R limma package, setting the |logFC| to >0.75 and the adjusted p-value to <0.05 as criteria, yielding a total of 216 up-regulated and 63 down-regulated differential genes. The volcano plot in **Figure 3C** depicts the distribution of DEGs. Additionally, we entered the keyword "lupus nephritis" to search the Genecards database for LN-related differential genes and found 1,248 genes (**Supplementary Table S2**). After using the VennDiagram package (Chen and Boutros, 2011) to intersect the critical genes of WGCNA, the DEGs of GSE32591, and the LN-related genes in the Genecards database, we identified nine co-expressed differential genes in multiple source datasets: *TLR2*, *LTF*, *IL10RA*, *IRF8*, *CD163*, *HLA-DMA*, *HLA-DRA*, *HLA-DPA1*, and *HLA-DPB1* (**Figure 3D**).

Co-Differentially-Expressed Genes Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Enrichment Analysis and Protein-Protein Interaction Network Construction

We deduced the specific functions and pathways of co-DEGs via GO and KEGG enrichment analyses. Additionally, the PPI

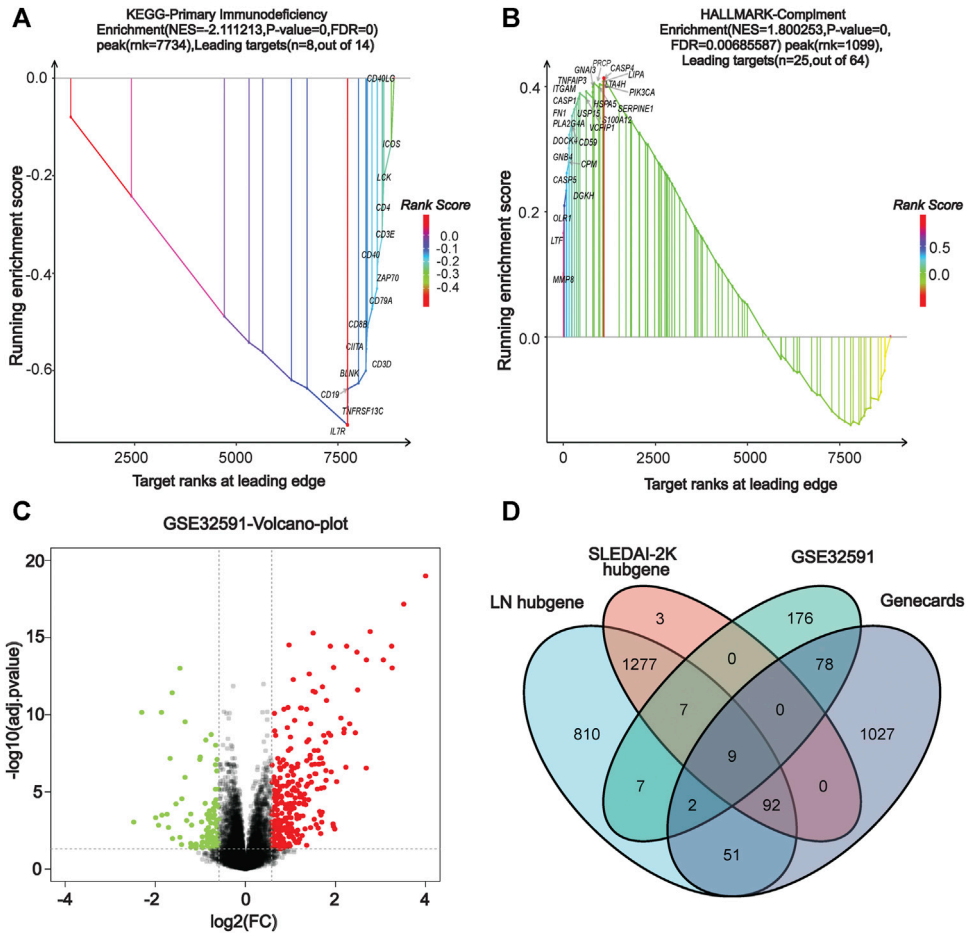


FIGURE 3 | | GSEA enrichment analysis and identification of co-DEGs. **(A)** KEGG dataset enrichment results. **(B)** HALLMARKER dataset enrichment results. **(C)** Volcano map of GSE32591 differentially-expressed genes. **(D)** Venn diagram screening for co-differentially-expressed genes.

network results uncovered intrinsic co-DEG connections. The ClueGo plugin in Cytoscape software can show a network of the association between pathways and the enrichment of genes among them. We imported co-DEGs into ClueGo for analysis, and the results show that the main GO entries enriched by co-DEGs included those for the MHC and MHC class II protein complexes (**Figure 4A**). The KEGG pathways involved include those for toxoplasmosis, inflammatory bowel disease, and others (**Figure 4B**). Our STRING database results showed the protein interaction associations of co-DEGs; we imported the results into CytoScape software to calculate the degree values between the networks (visualized using the cytoHubba plugin). *IRF8* was the hub gene in the network (**Figure 4C**).

Validation of Co-Differentially-Expressed Genes in Other Lupus Nephritis and Systemic lupus erythematosus Datasets

We used other LN and SLE expression profile data to validate the expression of co-DEGs in different datasets. Our results indicated that all co-DEGs were differentially expressed between control and LN groups in the LN GSE112943 set (**Figure 5A**). The expression levels of *IRF8*, *IL10RA*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DMA*, and *HLA-DRA* were not significantly different between the control and SLE groups (without LN) in the GSE81622 expression profile data. (**Figure 5B**). Therefore, these six genes highly expressed in LN samples indicated that they may serve as novel biomarkers for the disease.

Validation of Co-Differentially-Expressed Genes' Distribution in Single-Cell Datasets

Due to the close association between LN pathogenesis and immune disorders, we investigated the distribution of co-DEGs in immune cells using the published LN single-cell sequencing database. We found that among the six co-DEGs explicitly expressed in LN, HLA-related genes were overexpressed in a variety of macrophage subtypes (inflammatory $CD16^+$ macrophages, tissue-resident macrophages, phagocytic $CD16^+$ macrophages) and B cells (naive B cells, ISG-high B cells, activated B cells). However, *IRF8* and *IL10RA* were expressed at a relatively low level in macrophages and B cells. Additionally,

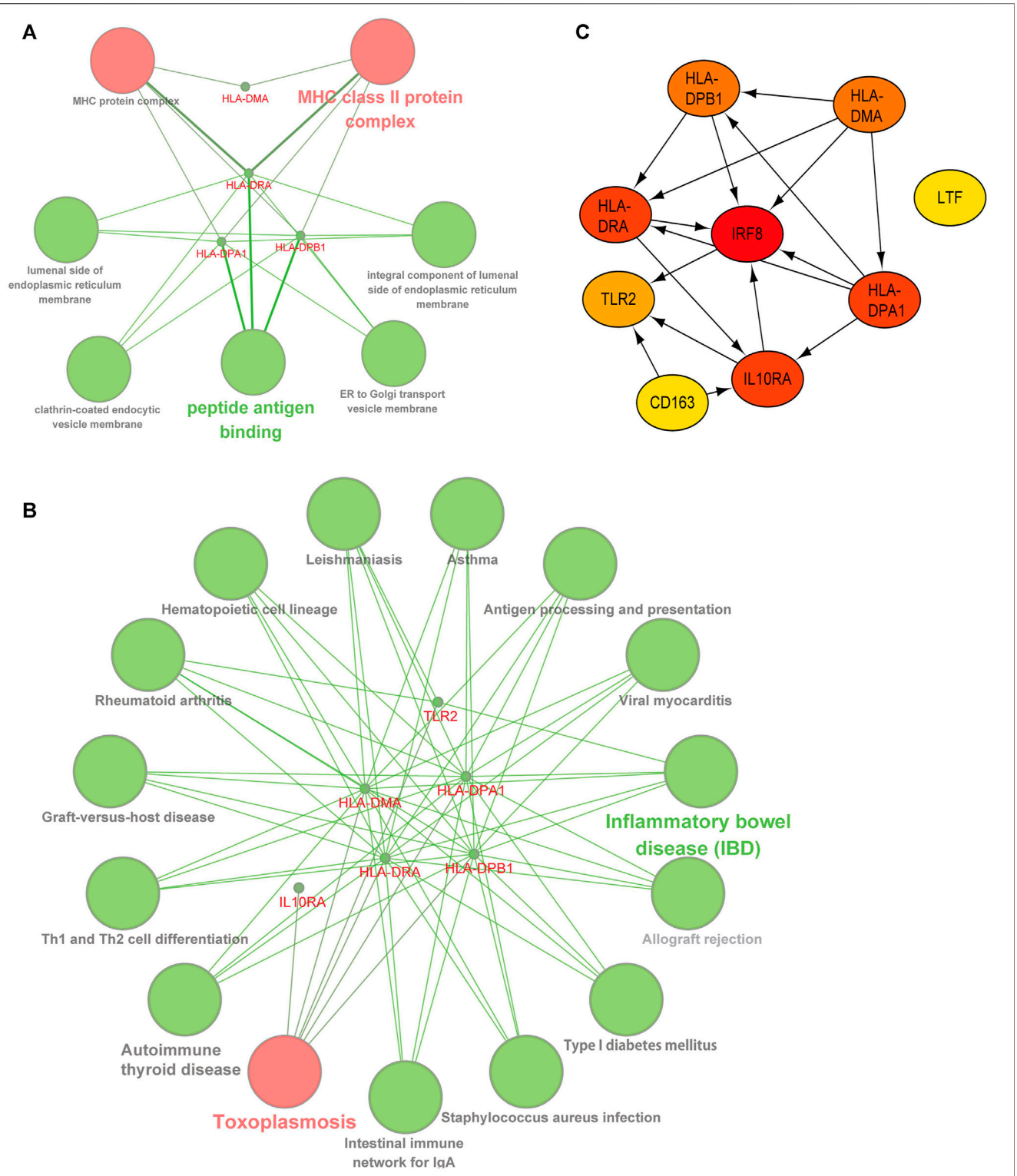


FIGURE 4 | | Functional and pathway analysis of co-DEGs. **(A)** GO enrichment analysis of co-DEGs. **(B)** KEGG enrichment analysis of co-DEGs. **(C)** Protein-protein interaction network of co-DEGs, color shades represent the degree size calculated with the cytoHubba plugin.

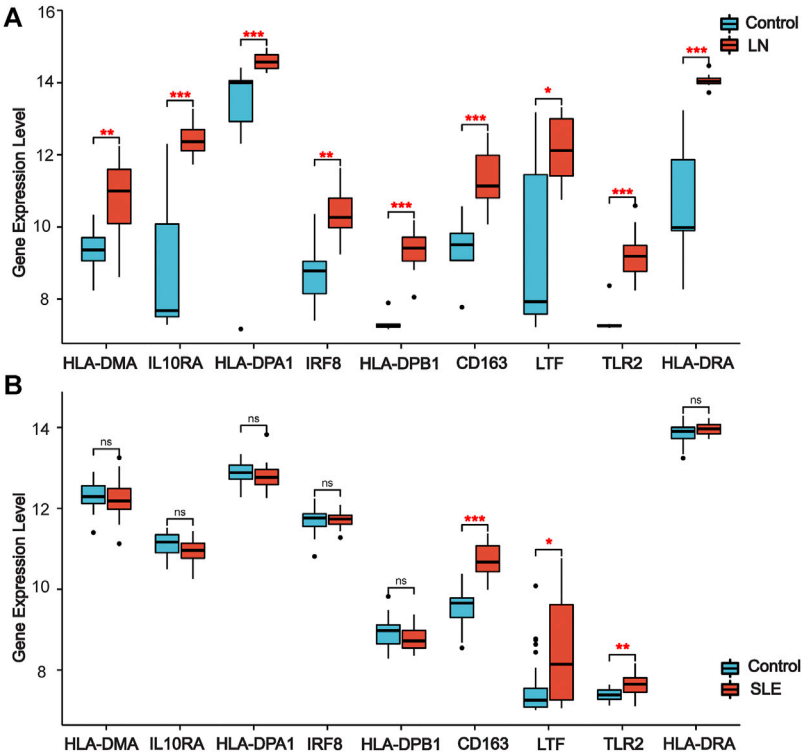


FIGURE 5 | | Validation of co-DEGs in additional LN and SLE datasets. **(A)** Expression of co-DEGs in the additional LN dataset GSE112943. **(B)** Expression of co-DEGs in the GSE81622 SLE dataset. (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$)

three other co-DEGs, *TLR2*, *LTF*, and *CD163*, were only partially expressed in macrophages (tissue-resident macrophages, phagocytic CD16$^+$ macrophages, M2-like CD16$^+$macrophages) (**Figure 6A**).

Diagnostic and Prognostic Analysis of Six Co-Differentially-Expressed Genes

We further evaluated the six LN-specific co-DEGs as diagnostic biomarkers by using ROC curves and a PRC analysis in a new LN dataset GSE60681 to assess the sensitivity and specificity of co-DEGs for LN diagnosis. As shown in **Figure 6B**, the ROC-AUC values for all six genes were greater than 0.78 (*IRF8*, 0.919; *HLA-DPB1*, 0.875; *HLA-DRA*, 0.882; *IL10RA*, 0.867; *HLA-DMA*, 0.803; and *HLA-DPA1*, 0.781) suggesting that these six genes have good diagnostic efficacy as LN markers. To obtain the minimum set of genes with the greatest predictive power, we tested combinations of the six genes; we found that the combination of *IRF8*, *IL10RA*, *HLA-DMA*, and *HLA-DPA1* had the greatest AUC value (0.94) among the 15 tested combinations (**Figure 6C**). We also performed a PRC analysis to compensate for the imbalance of the selected samples, and our results showed that *IRF8* (PR-AUC, 0.707) and *IL10RA* (PR-AUC, 0.605) retained a good diagnostic sensitivity despite the unevenness of the samples (**Figure 6D**). However, considering that the results of this method are affected by the number of positive and negative samples, analysis using a dataset with a different number of imbalanced samples would allow for a more comprehensive assessment of the results (that is, both ROC-AUC and PR-AUC showed good diagnostic efficacy when analyzed using the GSE32591 dataset; please see **Supplementary Figures S4, S5**). Thus, we believe that all six genes have good diagnostic efficacy after combining the ROC and PRC results from different datasets. To assess the association between co-DEGs and LN prognostic factors, we validated the association between co-DEGs and clinical traits in LN samples from the Nephroseq database. Our findings indicate that a high expression of different co-DEGs was correlated with a low glomerular filtration rate in kidney disease samples (*IL10RA*, *HLA-DPA1*, r = −0.490, $p < 0.001$; *IRF8*, *HLA-DRA*, r = −0.500, $p < 0.001$; *HLA-DPB1*, r = −0.510, $p < 0.001$; *HLA-DMA*, r = −0.480, $p < 0.001$) (**Figure 7A**). Further analysis revealed that *HLA-DPA1*, *IL10RA*, and *IRF8* were differentially expressed in different pathological staging samples of lupus nephritis (**Figure 7B**).

DISSCUSSION

We identified nine differential genes that were simultaneously significant in datasets from multiple sources. Six of the genes are LN-specific and are associated with a poor prognosis, and their good diagnostic efficacy suggests that these genes can serve as novel LN biomarkers. In addition, a comprehensive functional and pathway enrichment approach revealed that the biological

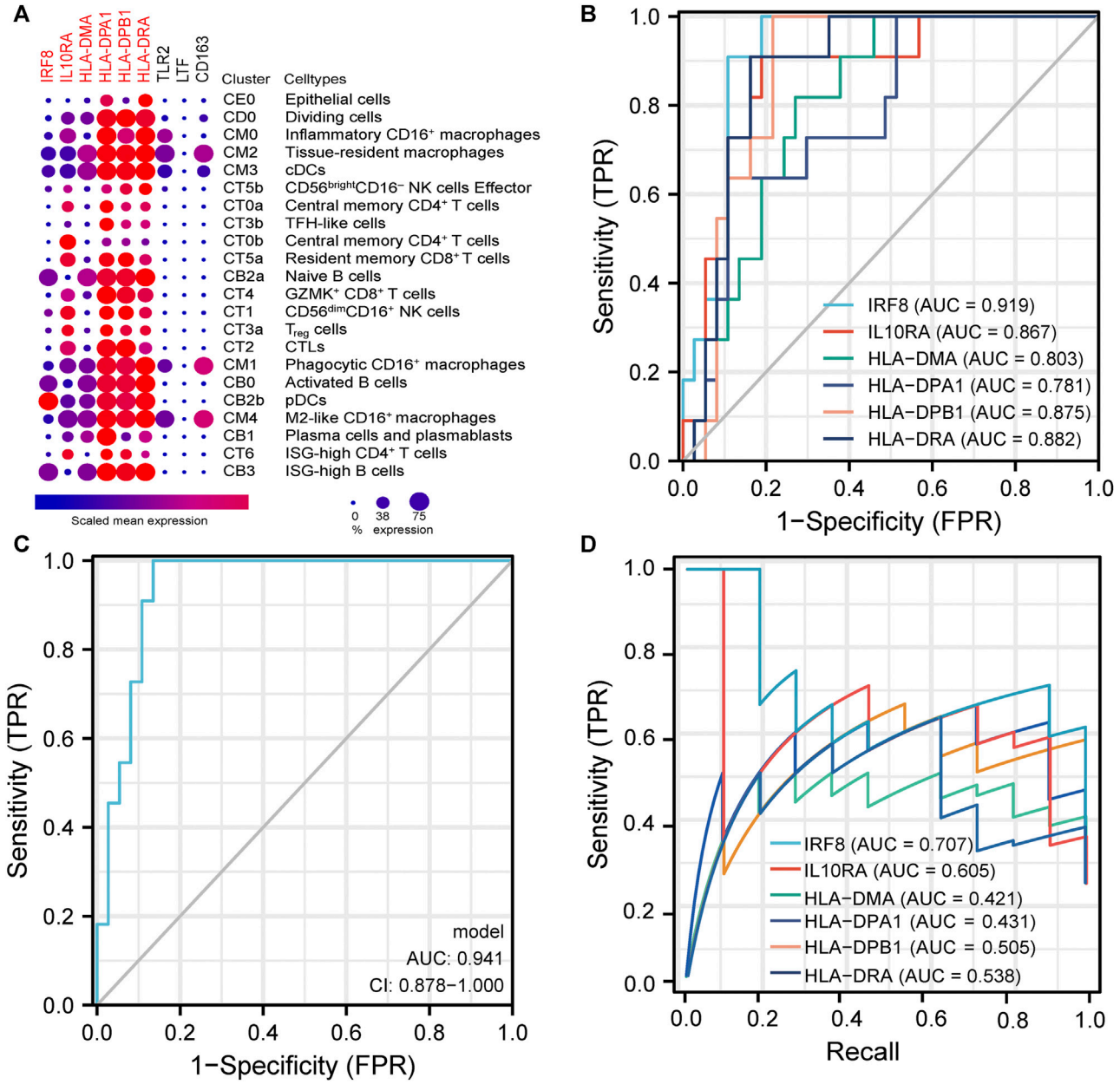


FIGURE 6 | | Analysis of immune cell distribution and diagnostic efficacy of LN biomarkers **(A)**. Expression of co-DEGs in different immune cell types in the single cell LN sequencing database. **(B)** ROC curves of six differentially-expressed genes specific to LN in GSE60861. **(C)** ROC curve with optimal minimal gene set (*IRF8*, *IL10RA*, *HLA-DMA*, *HLA-DPA1*) in GSE60861. **(D)** Precision-Recall Curves of six differentially-expressed genes specific to LN in GSE60861.

mechanisms mediating LN development are interrelated. A single DEG identification approach may miss some DEGs, but not those that are significantly differentially expressed; thus, the co-expressed genes identified in this study by integrating multiple approaches may be critical for elucidating the pathogenesis of LN.

The SLEDIA-2K score is a tool for assessing the severity of SLE (Touma et al., 2018). Our WGCNA analysis revealed that the critical modules derived from the LN trait are consistent with those derived from the SLEDIA trait, indicating that the module genes obtained are genuinely involved in the pathogenesis of LN. LN is primarily caused by an antigen-antibody complex immune response that results in large amounts of autoantibodies in the intrarenal space (Qiu et al., 2019). These atypical antigen-antibody reactions result in vascular damage, abnormal complement activation, complex deposition, and an imbalance of the oxidative/antioxidant and cytokine systems (Davidson et al., 2019; Anders et al., 2020). Our results on the function and pathways of critical gene enrichments confirm these mechanisms (**Supplementary Table S1**).

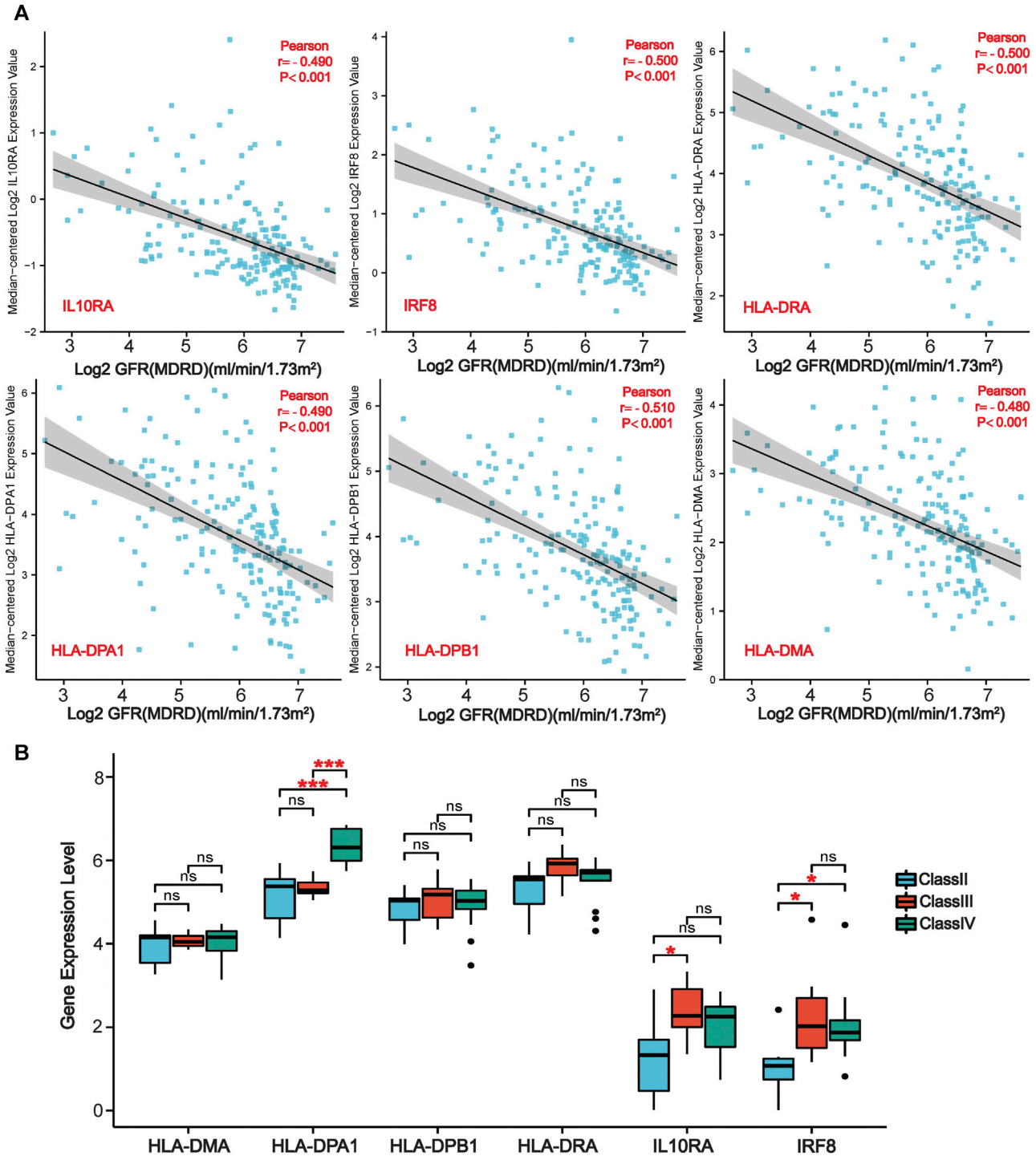


FIGURE 7 | | Association of co-DEGs with LN clinical traits in the Nephroseq database. **(A)** Correlation analysis of six co-DEGs with the glomerular filtration rate of patients (GFR). **(B)** Variations of six co-DEGs in samples from patients with different pathological staging of lupus nephritis. (*p < 0.05, ***p < 0.001).

Among the LN-specific DEGs identified, *HLA-DMA*, *HLA-DPA1*, *HLA-DPB1*, and *HLA-DRA* all belong to the human major histocompatibility complex (MHC) and the HLA class II region, alternatively referred to as the HLA-D region. Most genes in this region are involved in immune responses and are classified into several subregions (*HLA-DR*, *DQ*, *DP*, *DO*, and *DM*) (Wieczorek et al., 2017; Wang et al., 2019). SLE has been linked to polymorphisms in the HLA-D region that vary by race and geographical region. Alleles at the same HLA locus, which may differ structurally by a few nucleotides, can result in

completely different disease susceptibility or resistance profiles (Xue et al., 2018). This explains the differential expression of the four HLA genes we identified in the SLE and LN datasets, with genetic polymorphisms resulting in systemic and local pathological changes (Xu et al., 2017). *IRF8* is a member of the interferon regulatory factor (IRF) family that regulates the signaling pathway for Toll-like receptors (Salem et al., 2020). Alternatively, *IRF8* regulates Th cell differentiation, thereby regulating immune cell development and inhibiting tumor cell growth. Silencing the *IRF8* gene in SLE mice has been suggested to inhibit DC-mediated activation of NF-κβ or MAPKs, thereby impairing type I interferon induction (Salem et al., 2020). IL-10 is a multifunctional cytokine derived from multicellular organisms that functions only when bound to a specific receptor (Moore et al., 2001). IL-10 interacts with *IL10RA* and delivers excitatory or inhibitory signals to cells via the JAK-STAT signal transduction pathway (Geginat et al., 2019). Increased IL-10 expression in LN kidney tissues is associated with an increase in macrophage infiltration and is highly correlated with the severity of kidney damage (Saraiva et al., 2020). Additionally, our results suggest that co-DEGs are differentially expressed primarily in macrophages and B cells, a finding consistent with the previous view that different types of macrophages and B cells play more important roles in LN (through complex interactions) than T cells (Arazi et al., 2019).

In summary, we identified several valuable biomarkers associated with the diagnosis and prognosis of lupus nephritis. These biomarkers are involved in a variety of different molecular pathways expressed in various immune cells. However, additional research is necessary to determine the association between specific HLA alleles and LN because of the presence of HLA gene polymorphisms. In addition, the lack of a definitive experimental validation represents a limitation of our study. We will focus on establishing more conclusive and robust evidence for the validity of these identified co-DEGs as novel biomarkers in subsequent studies.

CONCLUSION

We found nine differentially expressed genes closely associated with LN diagnosis and prognosis by integrating multiple DEG identification methods. Next, we identified six biomarkers that may be LN-specific by expression validation in LN and SLE datasets. A comprehensive gene enrichment analysis revealed that the molecular mechanisms associated with LN pathogenesis are linked to multiple critical immune pathways. Finally, we explored the distribution of co-DEGs in LN immune cells by analyzing data from a single-cell transcriptome sequencing database of LN. Our prioritized biomarkers should be helpful for the diagnosis and prognosis of LN and they should deepen our understanding of its pathogenesis.

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

ZC, RL,KY, HC, CC, and YX were involved in the design of the study and the acquisition, analysis, and interpretation of the data. ZC, RL and YX were involved in the interpretation of the data and drafting the manuscript. CC and YX supervised the entire study and reviewed and edited the article. All authors read and approved the final paper.

FUNDING

This work was supported by grants from National Natural Science Foundation of China (No. 82070720), Joint Funds for the Innovation of Science and Technology (No. 2019Y9019) and Natural Science Foundation of Fujian province (No. 2020J02020), and Fujian Province Finance Project (2020B009). YX was supported by Outstanding Young Talents Program of the First Affiliated Hospital of Fujian Medical University (YJCQN-A-XYF2021).

ACKNOWLEDGMENTS

We thank Jianfeng Wu in School of Life Sciences of Xiamen University for research information, helpful discussion and valuable suggestions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2021.717234/full#supplementary-material

REFERENCES

Anders, H.-J., Saxena, R., Zhao, M.-h., Parodis, I., Salmon, J. E., and Mohan, C. (2020). Lupus Nephritis. *Nat. Rev. Dis. Primers* 6, 7. doi:10.1038/s41572-019-0141-9

Arazi, A., Rao, D. A., Rao, D. A., Berthier, C. C., Davidson, A., Liu, Y., et al. (2019). The Immune Cell Landscape in Kidneys of Patients with Lupus Nephritis. *Nat. Immunol.* 20, 902–914. doi:10.1038/s41590-019-0398-x

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene Ontology: Tool for the Unification of Biology. *Nat. Genet.* 25, 25–29. doi:10.1038/75556

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: Archive for Functional Genomics Data Sets-Update. *Nucleic Acids Res.* 41, D991–D995. doi:10.1093/nar/gks1193

Berthier, C. C., Bethunaickan, R., Gonzalez-Rivera, T., Nair, V., Ramanujam, M., Zhang, W., et al. (2012). Cross-Species Transcriptional Network Analysis

Defines Shared Inflammatory Responses in Murine and Human Lupus Nephritis. *J. Immunol.* 189, 988–1001. doi:10.4049/jimmunol.1103031

Bindea, G., Galon, J., and Mlecnik, B. (2013). CluePedia Cytoscape Plugin: Pathway Insights Using Integrated Experimental and In Silico Data. *Bioinformatics* 29, 661–663. doi:10.1093/bioinformatics/btt019

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., et al. (2009). ClueGO: A Cytoscape Plug-In to Decipher Functionally Grouped Gene Ontology and Pathway Annotation Networks. *Bioinformatics* 25, 1091–1093. doi:10.1093/bioinformatics/btp101

Chen, H., and Boutros, P. C. (2011). VennDiagram: A Package for the Generation of Highly-Customizable Venn and Euler Diagrams in R. *BMC Bioinformatics* 12, 35. doi:10.1186/1471-2105-12-35

Cline, M. S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., et al. (2007). Integration of Biological Networks and Gene Expression Data Using Cytoscape. *Nat. Protoc.* 2, 2366–2382. doi:10.1038/nprot.2007.324

Craciun, F. L., Bijol, V., Ajay, A. K., Rao, P., Kumar, R. K., Hutchinson, J., et al. (2016). RNA Sequencing Identifies Novel Translational Biomarkers of Kidney Fibrosis. *J. Am. Soc. Nephrol.* 27, 1702–1713. doi:10.1681/asn.2015020225

Davidson, A., Aranow, C., and Mackay, M. (2019). Lupus Nephritis: Challenges and Progress. *Curr. Opin. Rheumatol.* 31, 682–688. doi:10.1097/bor.0000000000000642

Durcan, L., O'dwyer, T., and Petri, M. (2019). Management Strategies and Future Directions for Systemic Lupus Erythematosus in Adults. *Lancet* 393, 2332–2343. doi:10.1016/s0140-6736(19)30237-5

Furie, R., Rovin, B. H., Houssiau, F., Malvar, A., Teng, Y. K. O., Contreras, G., et al. (2020). Two-Year, Randomized, Controlled Trial of Belimumab in Lupus Nephritis. *N. Engl. J. Med.* 383, 1117–1128. doi:10.1056/NEJMoa2001180

Geginat, J., Vasco, M., Gerosa, M., Tas, S. W., Pagani, M., Grassi, F., et al. (2019). IL-10 Producing Regulatory and Helper T-Cells in Systemic Lupus Erythematosus. *Semin. Immunol.* 44, 101330. doi:10.1016/j.smim.2019.101330

Grau, J., Grosse, I., and Keilwagen, J. (2015). PRROC: Computing and Visualizing Precision-Recall and Receiver Operating Characteristic Curves in R. *Bioinformatics* 31, 2595–2597. doi:10.1093/bioinformatics/btv153

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30. doi:10.1093/nar/28.1.27

Kiriakidou, M., and Ching, C. L. (2020). Systemic Lupus Erythematosus. *Ann. Intern. Med.* 172, Itc81–itc96. doi:10.7326/aitc202006020

Köttgen, A., Pattaro, C., Böger, C. A., Fuchsberger, C., Olden, M., Glazer, N. L., et al. (2010). New Loci Associated with Kidney Function and Chronic Kidney Disease. *Nat. Genet.* 42, 376–384. doi:10.1038/ng.568

Langfelder, P., and Horvath, S. (2008). WGCNA: An R Package for Weighted Correlation Network Analysis. *BMC Bioinformatics* 9, 559. doi:10.1186/1471-2105-9-559

Lech, M., and Anders, H.-J. (2013). The Pathogenesis of Lupus Nephritis. *J. Am. Soc. Nephrol.* 24, 1357–1366. doi:10.1681/asn.2013010026

Magnusson, R., Mariotti, G. P., Köpsén, M., Lövfors, W., Gawel, D. R., Jörnsten, R., et al. (2017). LASSIM-A Network Inference Toolbox for Genome-wide Mechanistic Modeling. *Plos Comput. Biol.* 13, e1005608. doi:10.1371/journal.pcbi.1005608

Moore, K. W., De Waal Malefyt, R., Coffman, R. L., and O'garra, A. (2001). Interleukin-10and Theinterleukin-10 Receptor. *Annu. Rev. Immunol.* 19, 683–765. doi:10.1146/annurev.immunol.19.1.683

Qiu, C. C., Caricchio, R., and Gallucci, S. (2019). Triggers of Autoimmunity: The Role of Bacterial Infections in the Extracellular Exposure of Lupus Nuclear Autoantigens. *Front. Immunol.* 10, 2608. doi:10.3389/fimmu.2019.02608

Salem, R. M., Todd, J. N., Sandholm, N., Cole, J. B., Chen, W. M., Andrews, D., et al. (2019). Genome-Wide Association Study of Diabetic Kidney Disease Highlights Biology Involved in Glomerular Basement Membrane Collagen. *J. Am. Soc. Nephrol.* 30, 2000–2016. doi:10.1681/asn.2019030218

Salem, S., Salem, D., and Gros, P. (2020). Role of IRF8 in Immune Cells Functions, Protection against Infections, and Susceptibility to Inflammatory Diseases. *Hum. Genet.* 139, 707–721. doi:10.1007/s00439-020-02154-2

Sallustio, F., Curci, C., Di Leo, V., Gallone, A., Pesce, F., and Gesualdo, L. (2019). A New Vision of IgA Nephropathy: The Missing Link. *Int. J. Mol. Sci.* 21, 189. doi:10.3390/ijms21010189

Saraiva, M., Vieira, P., and O'Garra, A. (2020). Biology and Therapeutic Potential of Interleukin-10. *J. Exp. Med.* 217, e20190418. doi:10.1084/jem.20190418

Smyth, G. K. (2004). Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Stat. Appl. Genet. Mol. Biol.* 3, 1–25. doi:10.2202/1544-6115.1027

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

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING V10: Protein-Protein Interaction Networks, Integrated over the Tree of Life. *Nucleic Acids Res.* 43, D447–D452. doi:10.1093/nar/gku1003

Tang, J., Kong, D., Cui, Q., Wang, K., Zhang, D., Gong, Y., et al. (2018). Prognostic Genes of Breast Cancer Identified by Gene Co-Expression Network Analysis. *Front. Oncol.* 8, 374. doi:10.3389/fonc.2018.00374

Touma, Z., Gladman, D. D., Su, J., Anderson, N., and Urowitz, M. B. (2018). A Novel Lupus Activity Index Accounting for Glucocorticoids: SLEDAI-2K Glucocorticoid Index. *Rheumatology (Oxford)* 57, 1370–1376. doi:10.1093/rheumatology/key103

Villanueva, R. A. M., and Chen, Z. J. (2019). ggplot2: Elegant Graphics for Data Analysis (2nd ed.). *Meas. Interdiscip. Res. Perspect.* 17, 160–167. doi:10.1080/15366367.2019.1565254

Wang, L.-Y., Chen, C.-F., Wu, T.-W., Lai, S.-K., Chu, C.-C., and Lin, H. H. (2019). Response to Hepatitis B Vaccination Is Co-Determined by HLA-DPA1 and -DPB1. *Vaccine* 37, 6435–6440. doi:10.1016/j.vaccine.2019.09.001

Wieczorek, M., Abualrous, E. T., Sticht, J., Álvaro-Benito, M., Stolzenberg, S., Noé, F., et al. (2017). Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation. *Front. Immunol.* 8, 292. doi:10.3389/fimmu.2017.00292

Wither, J. E., Prokopec, S. D., Noamani, B., Chang, N.-H., Bonilla, D., Touma, Z., et al. (2018). Identification of a Neutrophil-Related Gene Expression Signature that Is Enriched in Adult Systemic Lupus Erythematosus Patients with Active Nephritis: Clinical/Pathologic Associations and Etiologic Mechanisms. *PLoS One* 13, e0196117. doi:10.1371/journal.pone.0196117

Wooller, S. K., Benstead-Hume, G., Chen, X., Ali, Y., and Pearl, F. M. G. (2017). Bioinformatics in Translational Drug Discovery. *Biosci. Rep.* 37, BSR20160180. doi:10.1042/bsr20160180

Xu, R., Li, Q., Liu, R., Shen, J., Li, M., Zhao, M., et al. (2017). Association Analysis of the MHC in Lupus Nephritis. *J. Am. Soc. Nephrol.* 28, 3383–3394. doi:10.1681/asn.2016121331

Xue, K., Niu, W. Q., and Cui, Y. (2018). Association of HLA-DR3 and HLA-DR15 Polymorphisms with Risk of Systemic Lupus Erythematosus. *Chin. Med. J. (Engl)* 131, 2844–2851. doi:10.4103/0366-6999.246058

Zheng, X., Soroush, F., Long, J., Hall, E. T., Adishesha, P. K., Bhattacharya, S., et al. (2017). Murine Glomerular Transcriptome Links Endothelial Cell-specific Molecule-1 Deficiency with Susceptibility to Diabetic Nephropathy. *PLoS One* 12, e0185250. doi:10.1371/journal.pone.0185250

Zhu, H., Mi, W., Luo, H., Chen, T., Liu, S., Raman, I., et al. (2016). Whole-Genome Transcription and DNA Methylation Analysis of Peripheral Blood Mononuclear Cells Identified Aberrant Gene Regulation Pathways in Systemic Lupus Erythematosus. *Arthritis Res. Ther.* 18, 162. doi:10.1186/s13075-016-1050-x

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 Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 22 November 2021
doi: 10.3389/fbioe.2021.772206

Identification of Inflammatory Response-Related Gene Signature Associated With Immune Status and Prognosis of Lung Adenocarcinoma

Weijie Zou[1,2†], *Li Chen*[3†], *Wenwen Mao*[1,2†], *Su Hu*[1,2], *Yuanqing Liu*[1,2] *and Chunhong Hu*[1,2*]

[1]Department of Radiology, The First Affiliated Hospital of Soochow University, Suzhou, China, [2]Institute of Medical Imaging of Soochow University, Suzhou, China, [3]Department of Interventional Radiology, The First Affiliated Hospital of Soochow University, Suzhou, China

Background: Lung adenocarcinoma (LUAD) is an exceedingly diverse disease, making prognostication difficult. Inflammatory responses in the tumor or the tumor microenvironment can alter prognosis in the process of the ongoing cross-talk between the host and the tumor. Nonetheless, Inflammatory response-related genes' prognostic significance in LUAD, on the other hand, has yet to be determined.

Materials and Methods: The clinical data as well as the mRNA expression patterns of LUAD patients were obtained from a public dataset for this investigation. In the TCGA group, a multigene prognostic signature was built utilizing LASSO Cox analysis. Validation was executed on LUAD patients from the GEO cohort. The overall survival (OS) of low- and high-risk cohorts was compared utilizing the Kaplan-Meier analysis. The assessment of independent predictors of OS was carried out utilizing multivariate and univariate Cox analyses. The immune-associated pathway activity and immune cell infiltration score were computed utilizing single-sample gene set enrichment analysis. GO keywords and KEGG pathways were explored utilizing gene set enrichment analysis.

Results: LASSO Cox regression analysis was employed to create an inflammatory response-related gene signature model. The high-risk cohort patients exhibited a considerably shorter OS as opposed to those in the low-risk cohort. The prognostic gene signature's predictive ability was demonstrated using receiver operating characteristic curve analysis. The risk score was found to be an independent predictor of OS using multivariate Cox analysis. The functional analysis illustrated that the immune status and cancer-related pathways for the two-risk cohorts were clearly different. The tumor stage and kind of immune infiltrate were found to be substantially linked with the risk score. Furthermore, the cancer cells' susceptibility to anti-tumor medication was substantially associated with the prognostic genes expression levels.

Conclusion: In LUAD, a new signature made up of 8 inflammatory response-related genes may be utilized to forecast prognosis and influence immunological state. Inhibition of these genes could also be used as a treatment option.

Keywords: lung adenocarcinoma, inflammatory response, gene signature, immune status, tumor microenvironment, drug sensitivity

OPEN ACCESS

Edited by:
Ping Zhang,
Griffith University, Australia

Reviewed by:
Shining Ma,
Stanford University, United States
Young Hee Choi,
Dongguk University Seoul, South Korea

***Correspondence:**
Chunhong Hu
sdhuchunhong@sina.com

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Preclinical Cell and Gene Therapy, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *07 September 2021*
Accepted: *29 October 2021*
Published: *22 November 2021*

Citation:
Zou W, Chen L, Mao W, Hu S, Liu Y and Hu C (2021) Identification of Inflammatory Response-Related Gene Signature Associated With Immune Status and Prognosis of Lung Adenocarcinoma. Front. Bioeng. Biotechnol. 9:772206. doi: 10.3389/fbioe.2021.772206

INTRODUCTION

Lung cancer has become the second commonly occurring type of cancer in the world, and it is also the leading contributor of cancer mortality in both men and women, as per the global cancer statistics reported in 2020 (Sung et al., 2021). There are two distinct subtypes of lung cancer that include small cell lung cancer as well as non-small cell lung cancer (NSCLC). The commonly occurring NSCLC is lung adenocarcinoma (LUAD) (Cancer, 2014). The number of LUAD patients is increasing as air pollution and smoking rates decline. Research evidence has suggested that although the prognosis of patients with early LUAD is relatively good, approximately 10–44% of LUAD patients still die within 5 years following the surgical intervention (Goldstraw et al., 2016). However, the 5-year overall survival (OS) rate of patients with advanced LUAD is less than 15% (Kris et al., 2014). Therefore, in addition to standard clinical factors, a new prognostic signature for individualized survival risk assessment must be devised.

The association between inflammation and cancer is well recognized, and its function in the onset and progression of cancer has always been a research topic (Balkwill and Mantovani, 2001; Koliaraki et al., 2020). The inflammation acts as a two-edged sword that can either inhibit or promote tumor development (Greten and Grivennikov, 2019). People can investigate the association between tumor and inflammatory markers by evaluating commonly available measures in the blood. For instance, the Glasgow prognostic score, which includes albumin and C-reactive protein, has independent predictive significance in cancer patients (McMillan, 2013). The clinical systemic inflammation markers such as platelet-lymphocyte ratio (PLR), neutrophil-lymphocyte ratio (NLR), and lymphocyte-monocyte ratio (LMR) were assessed in newly diagnosed lung cancer that has not been previously treated, and these markers portrayed considerable prognostic ability for OS that was independent of formerly identified prognostic factors for lung cancer (Moik et al., 2020). Studies are increasingly supporting the utilization of combined acute-phase proteins to create an inclusive predictive score for cancer premised on inflammation. Some inflammatory response-related genes, in addition to serum indicators, are utilized to predict tumor prognosis and metastatic potential (Budhu et al., 2006; Lin et al., 2021). Current research on the association between inflammation-related genes and LUAD is limited.

LUAD patients' clinical information and mRNA expression profiles were collected from a public database for this investigation. Then, using differentially expressed genes (DEGs) that were linked to an inflammatory response from The Cancer Genome Atlas (TCGA) cohort, we built a prognostic signature and confirmed the reliability and stability of the model utilizing the Gene Expression Omnibus (GEO) cohort. To investigate its potential mechanism, we employed functional enrichment analysis. Moreover, we also looked at the link between the types of immune infiltrates and the expression of prognostic genes. Finally, we searched into the link between the expression of prognostic genes, cancer chemoresistance, as well as tumor stemness.

MATERIALS AND METHODS

Data Collection (TCGA-LUAD Cohort and GEO(GSE68465) Cohort)

The TCGA (https://www.cancer.gov/tcga) (n = 594) and GEO (http://www.ncbi.nlm.nih.gov/geo/) (n = 442) databases were used to obtain clinical data and RNA sequencing data. Patients who did not have a survival status or whose follow-up duration was less than a day were omitted from the study. Both TCGA and GEO made their data available to the public, in accordance with their respective publication requirements and data access policies. Moreover, in the Molecular Signatures database, 200 inflammatory response-related genes were discovered (**Supplementary Table S1**).

Construction and Validation of a Prognostic Inflammatory Response-Related Gene Signature

In the TCGA cohort, DEGs between non-tumor as well as tumor tissues were detected utilizing the "limma" R package with a false discovery rate <0.05 and a fold change >2. The inflammatory response-related genes with prognostic significance were evaluated utilizing Univariate Cox analysis. To reduce the overfitting risk, LASSO-penalized Cox regression analysis was utilized to build a prognostic model (Simon et al., 2011). With the "glmnet" R package, the LASSO algorithm was utilized to choose and compress variables so that some regression coefficients were stringently equivalent to 0, resulting in a model that was interpretable. The prognostic model's penalty parameter (λ) was assessed utilizing tenfold cross-validation, and we adhered to the minimum requirements. The expression levels of each inflammatory response-related gene, as well as the matching regression coefficient, were utilized to identify the patient's risk scores. The algorithm was defined as score = $e^{\text{sum (each gene's expression} \times \text{corresponding coefficient)}}$. The participants were categorized into low- and high-risk cohorts premised on their median risk score. To investigate the distribution of distinct cohorts with reference to gene expression levels in the built model, t-SNE and PCA analyses were done using the "Rtsne" and "ggplot2" R packages. The "survminer" R package was utilized to undertake a survival analysis on the OS of low- and high-risk cohorts. The time-dependent ROC curve analysis so as to assess the prognostic signature's predictive ability. In addition, multivariate and univariate Cox analyses were employed to investigate the signature's independent prognostic value.

Functional Enrichment Analysis

Gene Set Enrichment Analysis (GSEA) software 4.1 was employed to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses based on the DEGs between the

low- and high-risk cohorts using GSEA. Single-sample GSEA (ssGSEA) with the "GSVA" R package was used to compare the activity of 13 immune-related pathways and the infiltration scores of 16 immune cells between the low- and high-risk cohorts.

Immune Response Analysis and Tumor Microenvironment

Stromal cells and immune cells infiltration levels in distinct tumor tissues were assessed utilizing stromal and immune scores (Yoshihara et al., 2013). The link between the risk score and the stromal/immune scores was verified using the Spearman correlation. A two-way ANOVA analysis was used to see if there was a connection between the subtype of immune infiltration and risk score. To quantify tumor stem cell-like traits, researchers analyzed data collected from the epigenetics and transcriptome of TCGA tumor samples (Dib et al., 2017). The relationship between risk score and tumor stemness was investigated using the Spearman correlation test.

Analysis of Sensitivity to Chemotherapy

The CellMiner interface (https://discover.nci.nih.gov/cellminer) was utilized to access the NCI-60 dataset, which included 60 distinct cancer cell lines from 9 distinct kinds of malignancies. The link between medication sensitivity and prognostic gene expression was investigated utilizing Pearson correlation analysis. A correlation analysis was performed to investigate the effectiveness of 263 medications authorized by the FDA or currently in clinical studies (**Supplementary Table S2**).

Statistical Analysis

The DEGs between tumor samples and surrounding tissues were assessed utilizing the WilCoxon test. Subsequently, the Chi-squared test was performed for comparison of the various proportions. Next, ssGSEA scores of immunological pathways or immune cells were compared between low- and high-risk cohorts utilizing the Mann-Whitney, and the adjustment of the *p*-value was done utilizing the Benjamini and Hochberg technique. The Kaplan-Meier method was employed for the comparison of the differences in OS among distinct cohorts. To screen the independent determinants of OS, multivariate and univariate Cox analyses were done. Spearman or Pearson correlation analysis was utilized to examine the connection between prognostic gene expression level or prognostic model risk score and stemness score, drug sensitivity, immune score, and stromal score, and. Plots were made using R software (Version 3.6.3) and the tools survminer, corrplot, venn, ggplot2, pheatmap, igraph, and ggpubr. A two-tailed $p < 0.05$ depicted statistical significance in all statistical results.

RESULTS

Figure 1 depicts this study's flowchart. The research comprised 522 LUAD patients from the TCGA-LUAD group and 442 LUAD participants from the GEO (GSE68465) group. The specific clinical characteristics of these individuals were presented in **Table 1**.

Identification of Prognostic Inflammation-Related DEGs in the TCGA Cohort

In the tumor tissues as well as the adjacent non-tumor tissues, 59 inflammatory response-related genes were found to be differently expressed. Based on the outcomes of the univariate Cox analysis, 13 of them were associated with OS (**Figures 2A,B**). The prognostic indicators consisted of 13 inflammatory response-related genes, and the PCDH7 gene risk ratio was 1.29 (95% CI = 1.130–1.472, $p < 0.001$, **Figure 2C**). **Figure 2D** depicts the relationship among these genes.

Designing a Prognostic Model in the TCGA Group

The LASSO-Cox regression was utilized to evaluate the expression profiles of the aforementioned 13 genes, which was ensued by the development of a prognostic model. Premised on the ideal value of λ, an 8-gene marker was identified (**Supplementary Figure S1**). The risk score was calculated as follows: $e^{(\text{expression level of BTG2} * -0.088 + \text{expression level of CCL20} * 0.083 + \text{expression level of CD69} * -0.044 + \text{expression level of GPC3} * -0.012 + \text{expression level of IL7R} * -0.119 + \text{expression level of MMP14} * 0.068 + \text{expression level of NMUR1} * -0.090 + \text{expression level of PCDH7} * 0.148)}$. As per the median cut-off value, patients were separated into two cohorts (**Figure 3A**). The high-risk category in the TCGA cohort was determined to be strongly related to the more advanced TNM stage (**Table 2**). The patients in distinct risk categories were dispersed in two ways, according to PCA and t-SNE analysis (**Figures 3E,F**). Furthermore, the scatter chart revealed that high-risk patients had an increased likelihood of dying earlier as opposed to low-risk ones (**Figure 3B**). The Kaplan-Meier curve constantly demonstrated that high-risk patients had a considerably shorter OS contrasted with those with low risk (**Figure 3I**, $p < 0.001$). For the investigation of the prognostic model's survival prediction, time-dependent ROC curves were created, with the area under the curve (AUC) ranging between 0.695, 0.666, and 0.694 at 1-, 2-, and 3 years in that order (**Figure 3J**). Survival analysis was carried out premised on the optimal cut-off expression value for each of the prognostic genes to investigate the relationship between the prognostic genes and prognosis, which revealed that an elevated expression of these genes was all considerably linked to poor OS except GPC3 (**Supplementary Figures S2A–H**, $p < 0.05$).

Validation of the 8-Gene Signature in the GEO Cohort

The participants in the GEO group were also grouped into low-risk or high-risk cohorts premised on the median value from the TCGA group to assess the stability of the model established from the TCGA group (**Figure 3C**). t-SNE and PCA analysis indicated a distinct dispersal of patients in the two groupings, similar to the TCGA cohort results (**Figures 3G,H**). Correspondingly, as contrasted with the low-risk cohort, patients in the high-risk

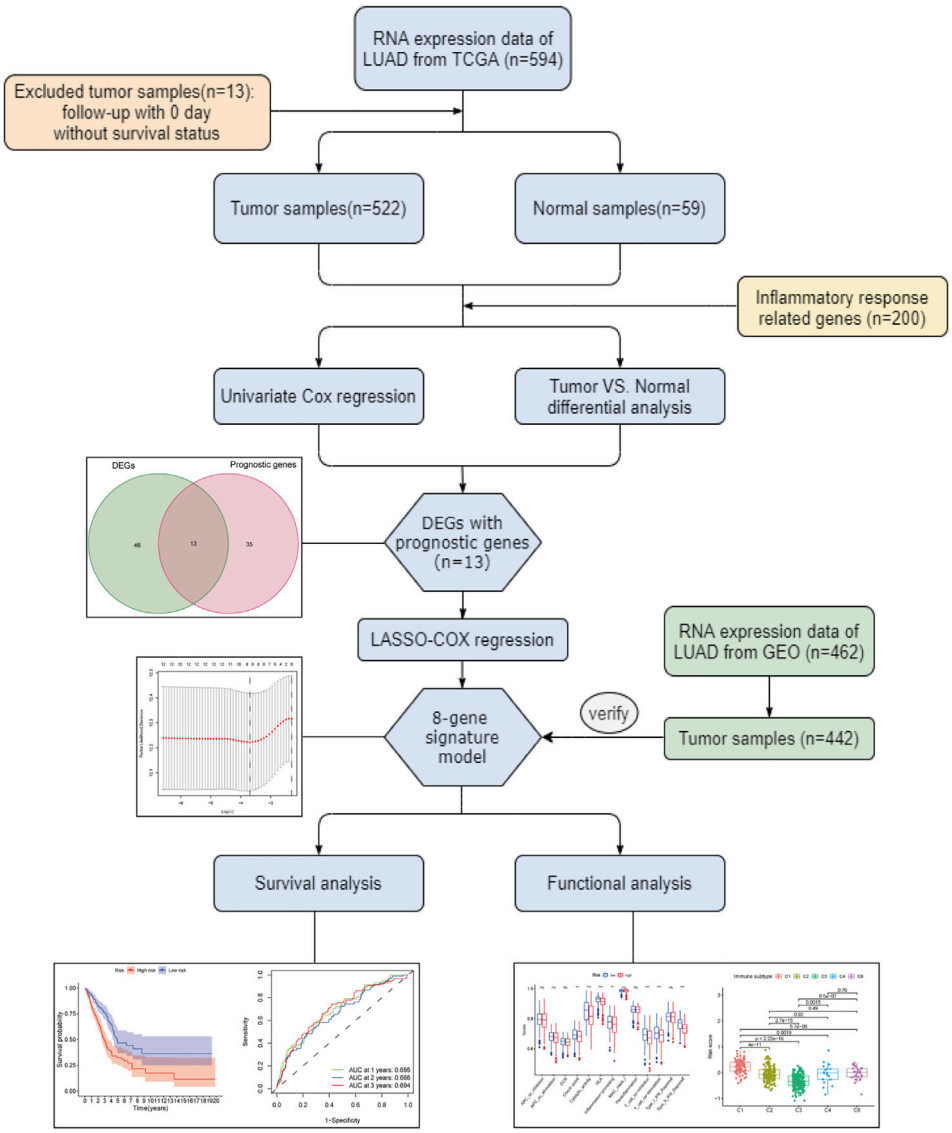


FIGURE 1 | Flow chart of the whole study.

cohort had an increased likelihood of dying prematurely (**Figure 3D**) and a lower survival period (**Figure 3K**). Furthermore, the AUC of the eight gene signature was 0.647, 0.642, and 0.638 at 1-, 2-, and 3-year in that order (**Figure 3L**).

Independent Prognostic Value of the 8-Gene Signature

To see if the risk score could be an independent predictive factor for OS, we used both multivariate and univariate Cox analysis of factors. The univariate Cox analysis illustrated that the risk scores in the GEO and TCGA groups had a significant correlation with OS (GEO group: HR = 2.115, 95% CI = 1.404–3.186, $p < 0.001$; TCGA group: HR = 4.184, 95% CI = 2.703–6.474, $p < 0.001$) (**Figures 4A,B**). Multivariate Cox analysis revealed that the risk score remained to be an independent predictor of OS after accounting for other confounding factors (TCGA group: HR = 3.475, 95% CI = 2.240–5.390, $p < 0.001$; GEO group: HR = 1.519, 95% CI = 1.012–2.278, $p = 0.043$) (**Figures 4C,D**). The ROC curve research revealed that the risk score exhibited a better prognostic predictive precision and that when paired with the tumor stage, it offered a highly precise 3-year OS forecast in LUAD patients, regardless of whether they were in the TCGA (AUC = 0.702) or GEO dataset (AUC = 0.685) (**Figures 4E,F**). As a result, LUAD's predictive value was outstanding when the risk score and clinicopathological parameters were combined.

Prognostic Model Risk Score and Clinical Features

By analyzing the association of risk score with the clinical characteristics of LUAD patients, we showed that the risk

TABLE 1 | Clinical characteristics of the LUAD patients used in this study.

	TCGA cohort	GEO cohort
No. of patients	522	442
Age (median, range)	66 (33–88)	65 (33–87)
Gender		
Female	280 (53.6%)	219 (49.5%)
Male	242 (46.4%)	223 (50.5%)
Stage		
I	279 (53.4%)	115 (26%)
II	124 (23.8%)	257 (58.1%)
III	85 (16.3%)	70 (15.9%)
IV	26 (5.0%)	0 (0%)
Unknown	8 (1.5%)	0 (0%)
Survival status		
Alive	334 (64%)	207 (46.8%)
Deceased	188 (36%)	235 (53.2%)

score was significantly higher in patients' age ≤65 (p = 0.0013) compared with >65 in TCGA cohort, while the risk score was no significant difference between the ages ≤65 and >65 years in GEO cohort (p = 0.16) (**Figures 5A,D**). The risk score was significantly higher in male (p = 0.0013) compared with female in GEO cohort, while there was no significant difference in TCGA cohort (p = 0.0019) (**Figures 5B,E**). Our findings depicted risk scores that were considerably greater in tumor stage III-IV as opposed to tumor stage I-II in both the GEO and TCGA datasets when we investigated the relationship between risk scores and clinical characteristics of patients with LUAD (GEO: p = 0.051; TCGA: p = 0.001) (**Figures 5C,F**). Furthermore, the expression of BTG2, IL7R, and NMUR1 was dissimilar between age ≤65 years and age >65 years, and the level of expression of CD69, GPC3, and IL7R was distinct between males and females (p < 0.05, **Supplementary Figures S3A,B**). In addition, the expression of BTG2, CD69, GPC3, and IL7R were considerably elevated in tumor stages III-IV as opposed to tumor stage I-II (p < 0.05, **Supplementary Figure S3C**).

Immune Status and TME Analysis

ssGSEA was employed in quantifying the enrichment scores of various immune cell subpopulations, related pathways, and functions in order to further assess the connection between immune status and risk score. The antigen presentation

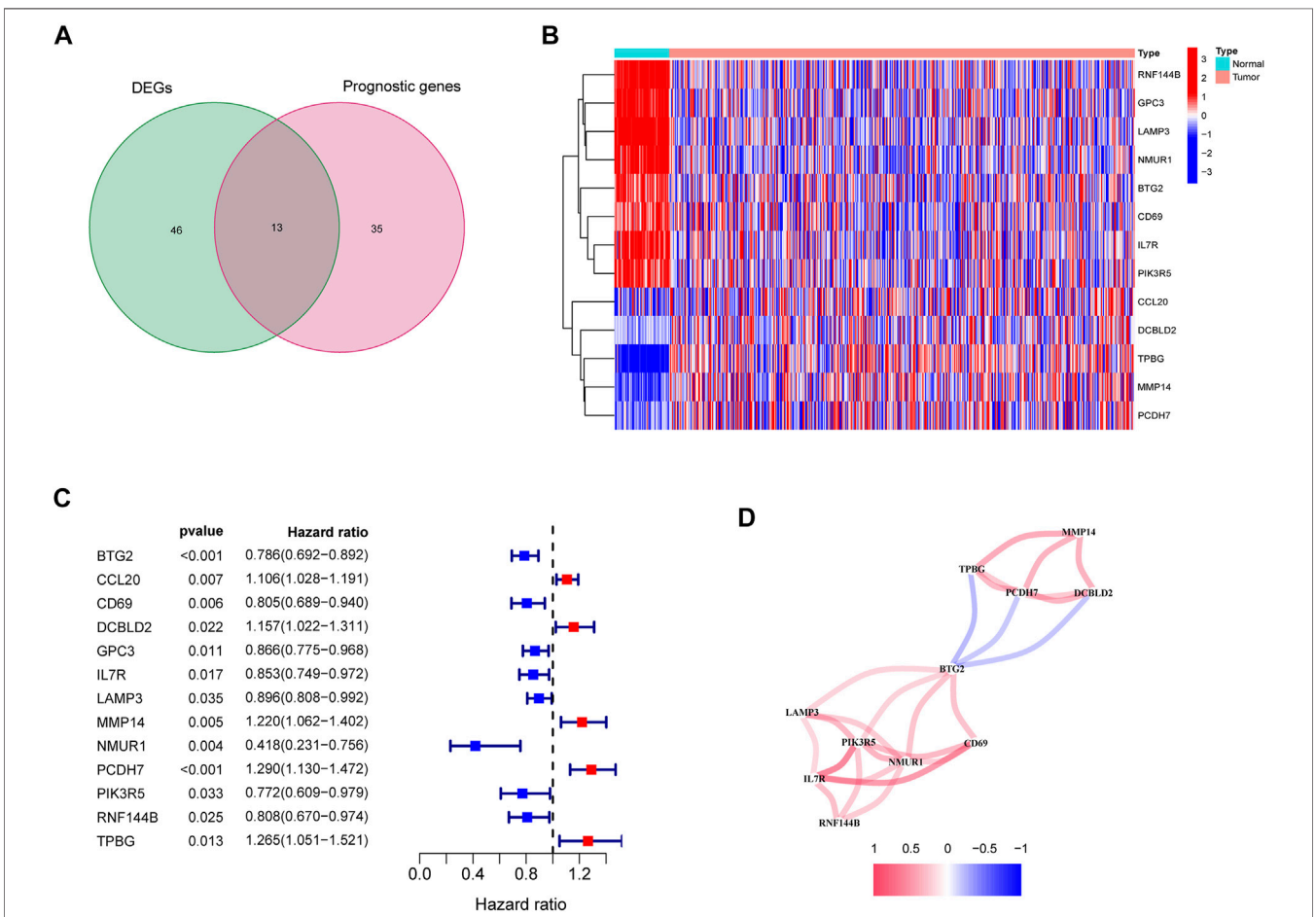


FIGURE 2 | Identification of the candidate inflammatory response-related genes in the TCGA cohort. **(A)** Venn diagram to identify DEGs between LUAD tissues and adjacent normal tissues. **(B)** The 13 overlapping genes expression between LUAD tissues and adjacent normal tissues. **(C)** Forest plots showing the results of the association between 13 overlapping gene expression and OS. **(D)** The correlation network of candidate genes.

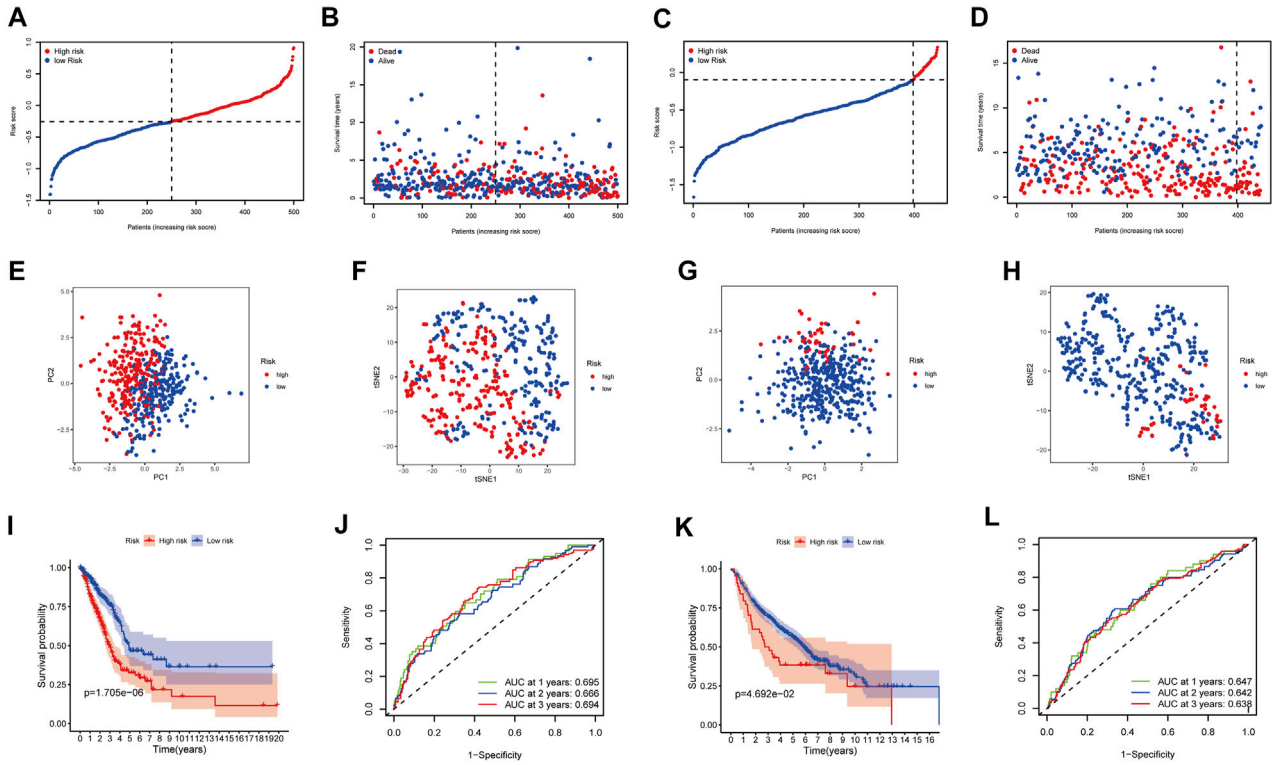


FIGURE 3 | Prognostic analysis of the 8-gene signature model in the TCGA cohort and GEO cohort. TCGA cohort **(A,B,E,F,I,J)**, GEO cohort **(C,D,G,H,K,L)**. **(A,C)** The median value and distribution of the risk scores. **(B,D)** The distribution of OS status. **(E,G)** PCA plot. **(F,H)** t-SNE analysis. **(I,K)** Kaplan-Meier curves for OS of patients in the high- and low-risk groups. **(J,L)** AUC time-dependent ROC curves for OS.

TABLE 2 | Baseline characteristics of the patients in different risk groups.

Characteristics	TCGA cohort			GEO cohort		
	High risk	Low risk	*p* value	High risk	Low risk	*p* value
Age						
≤65	136 (54.4%)	101 (40.4%)	0.0021	21 (47.73%)	210 (52.76%)	0.6343
>65	109 (43.6%)	144 (57.6%)		23 (52.27%)	188 (47.24%)	
unknow	5 (2%)	5 (2%)		—	—	
Gender						
Female	127 (50.8%)	143 (57.2%)	0.1783	20 (45.45%)	199 (50%)	0.6793
Male	123 (49.2%)	107 (42.8%)		24 (54.55%)	199 (50%)	
Stage						
I-II	182 (72.8%)	205 (82%)	0.0095	32 (72.73%)	353 (88.69%)	0.0058
III-IV	65 (26%)	40 (16%)		12 (27.27%)	45 (11.31%)	
unknow	3 (1.2%)	5 (2%)		—	—	

process, involving aDCs, DCs, iDCs, pDCs, and HLA, was shown to be considerably higher in the low-risk cohort in the TCGA class ($p < 0.05$, **Figures 6A,C**). The fractions of T helper cells, $CD8^+$ T cells, Th1 cells, Tfh cells, TIL, T cell co-inhibition, and T cell co-stimulation in the low-risk cohort were greater than in the high-risk cohort, showing variances in T cell regulation between the two cohorts. In addition, the high-risk cohort had greater scores for B cells, mast cells, neutrophils, inflammation-promoting, check-point, type II IFN response activity, and cytolytic activity ($p < 0.05$). Comparing the two risk subcategories in the GEO group yielded results that were similar to those in the TCGA ($p < 0.05$, **Figures 6B,D**).

We investigated the link between immune infiltrates and risk scores to find out more about the link between risk scores and immune components. In human tumors, six types of immune infiltrate were established that ranges from tumor-promotion to tumor-suppression (Tamborero et al., 2018), including C1 (wound healing), C2 (INF-γ dominant), C3 (inflammatory), C4 (lymphocyte depleted), C5 (immunologically quiet), and C6 (TGF-β dominant). Because none of the patient samples in LUAD corresponded to the

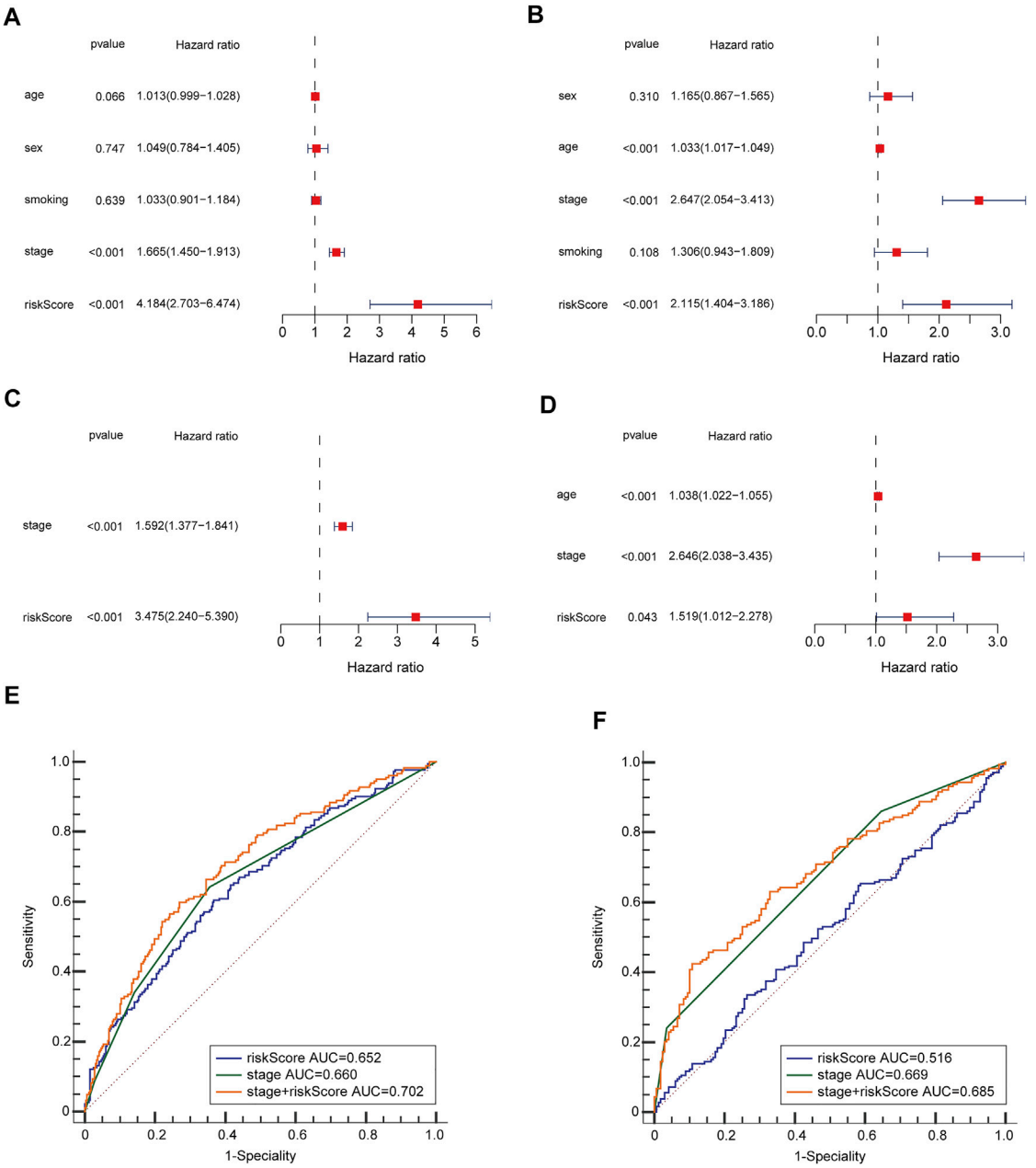


FIGURE 4 | OS-related factors were screened, and the prognostic accuracy of risk score and clinicopathological factors were compared. TCGA cohort **(A,C,E)**, GEO cohort **(B,D,F)**. **(A,B)** OS-related factors were screened by Univariate Cox regression analyses. **(C,D)** OS-related factors were screened by Multivariate Cox regression analysis. **(E,F)** Time-dependent ROC curve was used to compare the prognostic accuracy of risk score, tumor stage, and the combination of risk score. and tumor stage in 3-year.

C5 immune subtype, C5 immune subtypes were excluded from the study. The immune infiltration of LUAD in the TCGA dataset was evaluated and contrasted with a risk score, with the outcomes revealing that an elevated risk score was considerably linked to C1, whereas a reduced risk score was considerably linked to C3 (**Figure 6E**). High expression of BTG2, CD69, GPC3, IL7R, MMP14, and NMUR1 were significantly related with C3, and high expression of CCL20, MMP14, and PCDH7 was clearly connected with C1, as shown in **Supplementary Figure S4**.

In cancer immune evasion, the PD-1/PD-L2 and PD-1/PD-L1 pathways are important regulators. Immune checkpoints's expression such as PD-L2 and PD-L1 are essential markers for personalized treatment. In the cohort categorized as high-risk, the PD-1 and PD-L2 expression levels were distinct from those in the low-risk cohort (**Figures 7A,C**) and revealed a negative relationship with the risk score (**Figures 7F,H**). The PD-L1 expression levels and immunological checkpoints were not substantially related to the low- and high-risk cohorts (**Figures 7B,G**). With regards

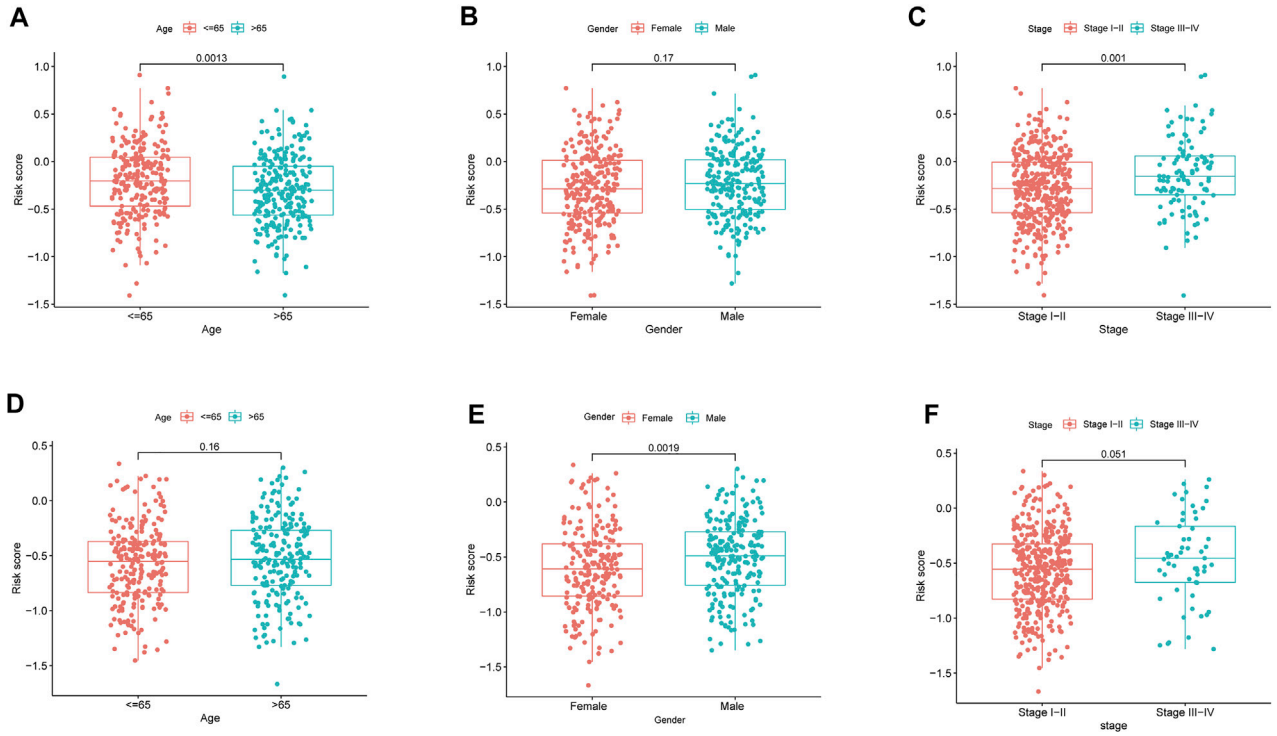


FIGURE 5 | The risk score in different groups divided by clinical characteristics. TCGA cohort **(A–C)**, GEO cohort **(E–F)**. **(A,D)** Age. **(B,E)** Gender. **(C,F)** Tumor stage.

to tumor medication resistance genes, patients in the high-risk cohort had elevated MRP1 expression as opposed to low-risk patients, which was positively connected with risk scores, but MRP3 was the inverse. With regards to tumor drug resistance genes, MRP1 expression was elevated in the high-risk cohort as opposed to the low-risk cohort and had a positive correlation with risk scores, whereas MRP3 was the opposite (**Figures 7D,E,I,J**).

The RNA stemness (RNAss) score, which is premised on mRNA expression, and the DNA stemness (DNAss) score, which is premised on DNA methylation pattern, can both be used to determine tumor stemness (Malta et al., 2018). The tumor immune microenvironment was estimated using stromal and immune scores. The goal of the correlation analysis was to see if the risk score was linked to the immune microenvironment and tumor stem cells. The outcomes illustrated that the risk score was significantly and positively linked to RNAss and DNAss but significant and negatively linked to immune and stromal score ($p < 0.05$) (**Figure 6F**).

Pathway Analyses and Biological Function

The GSEA was employed for comparison between the low- and high-risk cohorts in terms of GO function and KEGG pathway enrichment. Cell cycle phase transition regulation was considerably enriched in the high-risk cohort, while intracellular activity was considerably enriched in the low-risk cohort, according to GO function enrichment analysis (**Figure 8A**). Also, enrichment of 12 KEGG pathways took place in the high- and low-risk cohorts with a $p < 0.05$ (**Figure 8B**). Some cancer-related pathways such as Cell Cycle, Proteasome, and P53, were found to be enriched in the high-risk cohort, while JAK-STAT, MAPK, and VEGF were revealed to be enriched in the low-risk cohort. In addition, FcεRI receptor, calcium, and T cell receptor were also revealed in the KEGG pathways, which were associated with inflammatory responses. PI3K-AKT-mTOR-Signaling, mTORC1-Signaling G2/M checkpoint, hypoxia, and epithelial-mesenchymal transition pathways were statistically significant programs, according to GSEA utilizing TCGA data from the Hallmarks gene sets (**Figure 8C**).

Expression of Prognostic Genes and Sensitivity of Cancer Cells to Drugs

The prognostic genes expression in NCI-60 cell lines was explored, as well as the association between their levels of expression and medication sensitivity. The findings revealed that all prognostic genes were linked to some chemotherapy drug sensitivity ($p < 0.05$) (**Figure 9**). For instance, higher expression of CD69, BTG2, MMP14, PCDH7, GPC3, and NMUR1 has been related to greater cancer cell drug sensitivity to a variety of chemotherapeutic agents, including oxaliplatin, vemurafenib, trametinib, paclitaxel, and vinblastine, etc. In contrast, elevated expression of IL7R and CCL20 was linked to

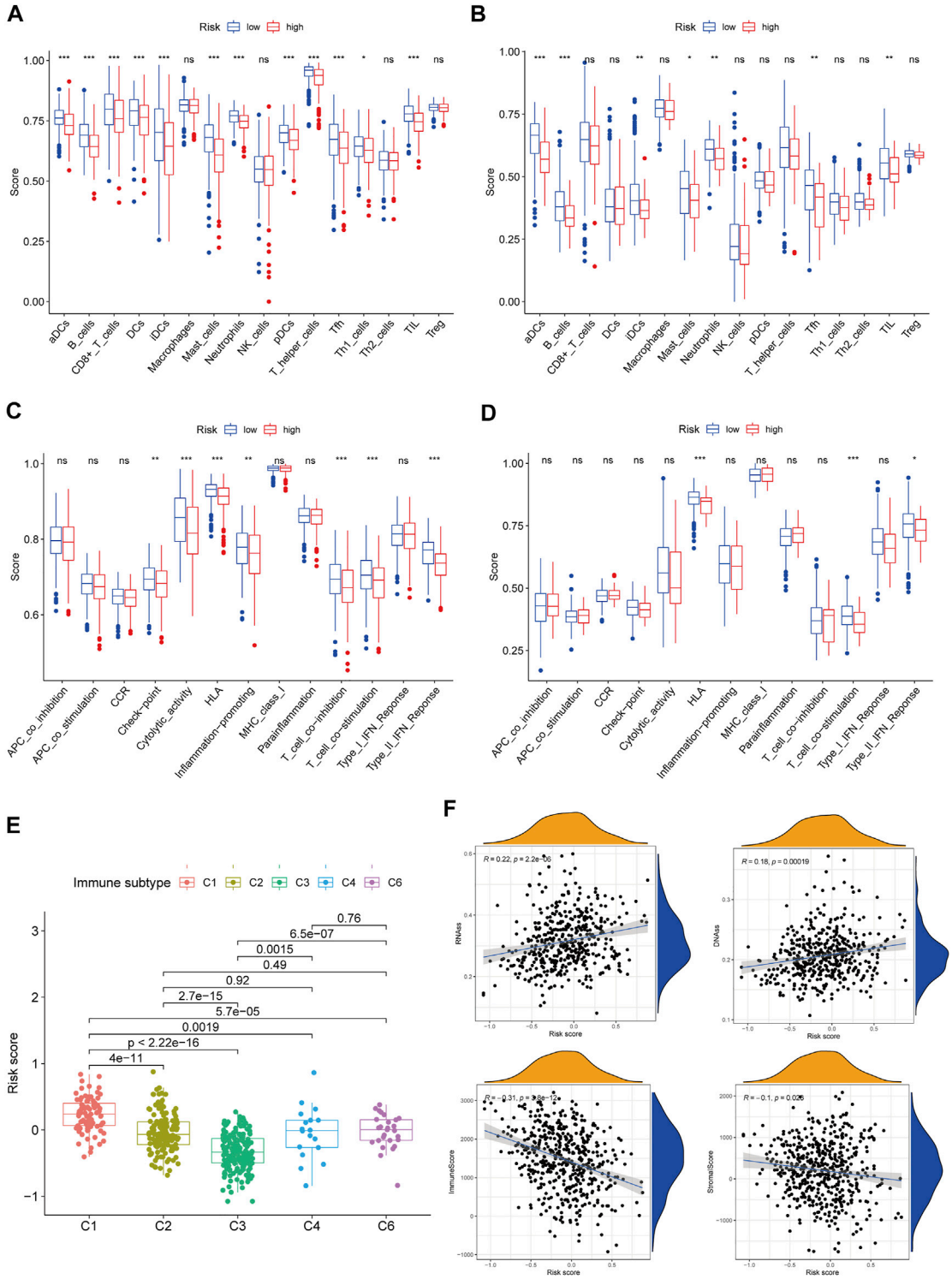


FIGURE 6 | Immune status between different risk groups and the association between risk score and tumor microenvironment. TCGA cohort **(A,C)**, GEO cohort **(B,D)**. **(A,B)** The scores of 16 immune cells and **(C,D)** 13 immune-related functions were showed in boxplots. **(E)** Comparison of the risk score in different immune infiltration subtypes. **(F)** The relationship between risk score and RNAss, DNAss, Stromal Score and Immune Score. *p* values were showed as: ns, not significant; **p* < 0.05; ***p* < 0.01; ****p* < 0.001.

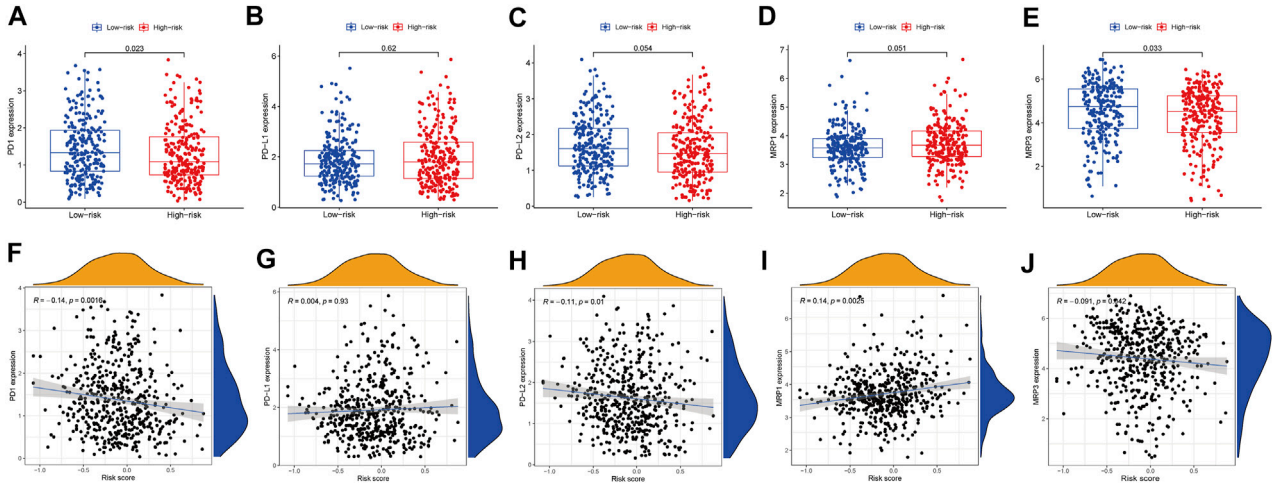


FIGURE 7 | The comparison of the expression levels of PD-1, PD-L1, PD-L2, MRP1 and MRP3 between different risk groups and correlation analysis between risk score and the expression levels of PD-1, PD-L1, PD-L2, MRP1 and MRP3. **(A,F)** PD-1. **(B,G)** PD-L1. **(C,H)** PD-L2. **(D,I)** MRP1. **(E,J)** MRP3.

greater cancer cell drug resistance to bosutinib, lapatinib, tamoxifen, IPI−145, and idelalisib.

DISCUSSION

The treatment of LUAD has advanced dramatically as a result of the development of sequencing technology as well as the dawn of a period of precision medicine, but lung cancer remains the most fatal cancer on a global scale. We can't always diagnose and forecast the therapeutic effects of LUAD because there are few reliable indicators. Therefore, developing a novel approach for reliably identifying LUAD is critical for the disease's diagnosis and prognosis. Accompaniment fragments, circulating blood protein profiling, DNA methylation, tumor DNA, miRNAs (Seijo et al., 2019), circulating tumor cells (Maly et al., 2019), and a plasma miRNA panel (Wadowska et al., 2020) have all been shown to have high accuracy in LUAD prognosis in previous research. Moreover, inflammatory response-related serum biomarkers including PLR, LMR, NLR, and SII (systemic immune-inflammation index) also portray better performance in forecasting the prognosis of LUAD (Nost et al., 2021). Though, there hasn't been any research on the inflammatory response-related gene signature as a predictive predictor for LUAD. Studies have shown that gene signatures that are immune-related (Yi et al., 2021), hypoxia-related (Sun et al., 2020), ferroptosis-related (Zhang et al., 2021), and energy metabolism-related (Zhang et al., 2019) forecast 3-year OS in a similar way as the findings in this research. Besides the excellent predictive performance of LUAD, the inflammatory response-related gene signature established in this research has greater advantages as opposed to the gene signature aforementioned. For instance, it can differentiate tumor resistance genes and immune checkpoint genes into low- and high-expression categories, and it has been demonstrated that risk scores are connected to resistance to several chemotherapeutic medicines. High-throughput sequencing was used in our investigation to evaluate the prognostic signature genes expression levels, which is a common technology that can produce accurate results.

For this research, we explored the expression of 200 inflammatory response-related genes in LUAD tissues as well as how they relate to OS. From the TCGA cohort, 59 DEGs were selected. Univariate Cox analysis depicted that 13 of the DEGs were related to OS. Subsequently, a prognostic model was created utilizing the LASSO regression analysis. The model included 8 inflammatory response-related genes and was verified in the GEO group. The participants were grouped into low- and high-risk cohorts premised on their median risk score. The outcomes illustrated that the high-risk cohort was linked to the shorter OS and higher tumor grade. Independent prognostic analysis illustrated that risk score was an independent predictor for OS.

This research created a prognostic model comprised of 8 inflammatory response-related genes. CCL20, MMP14, and PCDH7 were upregulated in LUAD tumor tissues and linked to poor clinical outcomes, while BTG2, CD69, GPC3, IL7R, and NMUR1 are the opposite. CCL20 is one of the important members of chemokine family, which was reported that may be a protective factor or prognostic risk factor for LUAD (Bao et al., 2016). CCL20-encoded proteins are capable of chemotaxis of lymphocytes, allowing the tumor to form an immune tolerance state (Schutyser et al., 2003). In patients with NSCLC, the CCL20 gene and protein are overexpressed, and autocrine of CCL20 can promote the migration and proliferation of lung cancer cells (Mao et al., 2021). MMP14 is a member of the matrix metalloproteinase family and contributes to a key function in cancer metastasis, its expression is significantly correlated with poor OS (Stawowczyk et al., 2017; Infante et al., 2018; Zhang et al., 2020). PCDH7 is a member of the cadherin superfamily and has been found to promote the metastasis of lung cancer cells (Chen et al., 2016). BTG2 is a recently recognized tumor suppressor belonging to the TOB/BTG family, and a study shows that BTG2 expression is reduced in NSCLC tissues and is linked to shorter OS for patients (Chen et al., 2020). Kim et al. (2003) demonstrated GPC3 to be a

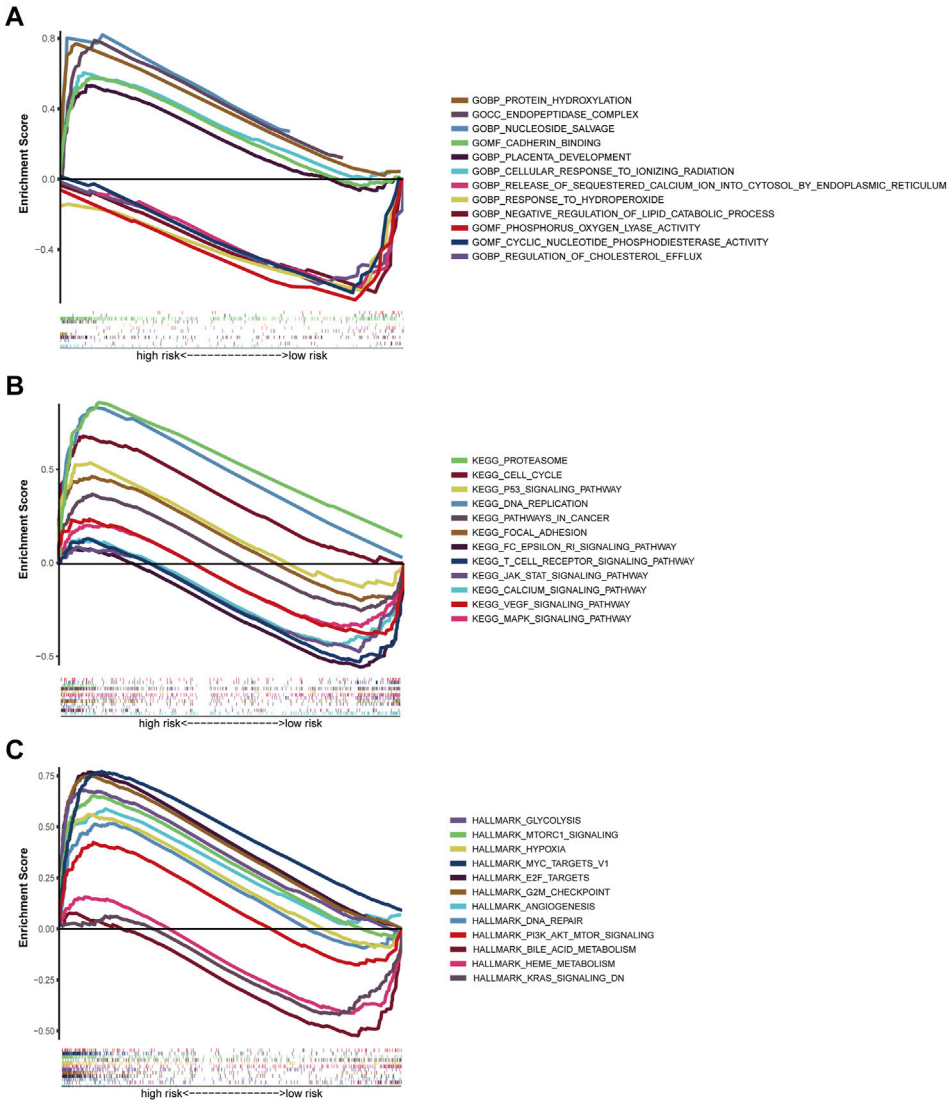


FIGURE 8 | Gene set enrichment analysis of Biological functions and pathways. **(A)** GO, Gene Ontology. **(B)** KEGG, Kyoto Encyclopedia of Genes and Genomes. **(C)** Hallmark gene set.

candidate gene for lung tumor suppression. Ning et al. (2021) found GPC3 expression was significantly correlated with gender and tumour stage in LUAD samples. IL7R has been explored as aggressive tumor features for patients with LUAD: KRAS mutation, larger tumor size, lymphovascular invasion, high-grade morphology, and more frequent recurrence (Suzuki et al., 2013). NMUR1 is a receptor for the neuropeptide NMU, and NMUR1 signaling promotes inflammatory ILC2 responses, highlighting the importance of neuro-immune crosstalk in allergic inflammation at mucosal surfaces (Wallrapp et al., 2017).

We investigated the function of the risk score in immune infiltration type to acquire a comprehensive understanding of the interplay between immune components and risk score. We found that high-risk score was closely linked to C1, whereas low-risk score was strongly linked to C3, implied that C1 stimulates tumor occurrence and progression while C3 was a good protective factor. Because highly cytotoxic immunophenotypes can limit the occurrence and progression of tumors, this study was consistent with the findings of prior investigations (Tamborero et al., 2018). With regard to the relationship between clinical features and risk score, a high-risk score was found to be strongly connected with tumor stages III-IV, implying that a high-risk score is unquestionably linked to a poor prognosis.

However, because there has been little research on these genes, it is unclear if they influence LUAD patients' prognosis through an inflammatory response. Tumor-related signal pathways like MAPK, p53, and JAK-STAT were considerably enriched in the GSEA analysis, and incessant activation of these pathways is connected to LUAD, which could be new treatment targets (Chou et al., 2019; Mohrherr et al., 2019; Stutvoet et al., 2019). Inflammation-related signal pathways including Calcium, T cell

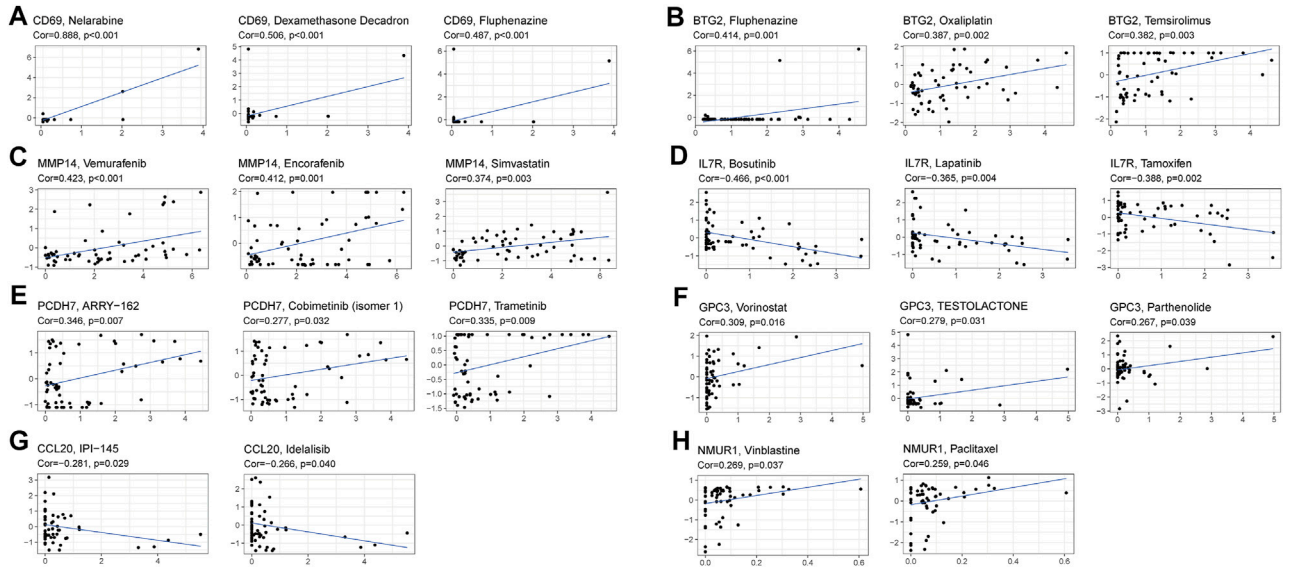


FIGURE 9 | Scatter plot of relationship between prognostic gene expression and drug sensitivity. **(A)** CD69. **(B)** BTG2. **(C)** MMP14. **(D)** IL7R. **(E)** PCDH7. **(F)** GPC3. **(G)** CCL20. **(H)** NMUR1.

receptor, and FcεRI receptor pathways were significantly enriched, indicating that the inflammatory response is associated with tumor progression. Besides, the low-risk cohort exhibited higher fractions of mast cells, Dcs, and neutrophils. Existing evidence has illustrated that the presence of tumor-associated t mast cells, DCs, and some neutrophils might have a protective influence on the progression of the tumor, they were thought to be beneficial for survival in NSCLC (Welsh et al., 2005; Engblom et al., 2017; Li et al., 2017). We show that anti-PD-L2 antibodies were increased in the low-risk group, this finding was aligned with the outcomes of former studies (Shinchi et al., 2019). In our research, the low-risk cohort had a greater immune checkpoint score as opposed to the high-risk cohort, and the risk score was negatively connected with PD-1 and PD-L2 expression. Hence, the created prognostic model has the potential to guide treatment decisions by predicting immune checkpoint expression levels. Furthermore, a high-risk score was linked to reductions in the type II IFN response activity, which is critical for promoting tumor elimination, promoting anti-tumor immunity, as well as tumor immune surveillance (Shankaran et al., 2001; Street et al., 2001; Wang et al., 2015). Finally, higher T helper cell, $CD8^+$ T cell, TIL, Tfh cell, Th1 cell, T cell co-inhibition, and T cell co-stimulation activities in the low-risk cohort implied that the immune regulatory role in the high-risk cohort has been inhibited, and this may be the main reason behind its poor prognosis.

Currently, cancer biology is shifting from a "cancer cell-centered" perspective to one whereby the cancer cells are embedded in a stromal cells network that includes inflammatory immune cells, vascular cells, and fibroblasts. The TME is comprised of these cells (Greten and Grivennikov, 2019). Cancer stem cell-like cells (CSCs) could be formed from a variety of sources such as progenitor cells, long-lived stem cells, and dedifferentiation of non-stem cells (Malta et al., 2018). Because of their propensity to undergo self-renewal and invasion, CSCs promote tumor growth. This is why treatment-induced drug resistance is a problem for these cells (Huang et al., 2010). The risk score was substantially linked to RNAss, DNAss, immune and stromal score in a correlation analysis between the immune microenvironment, tumor stem cells, and the risk score.

We discovered that higher expression of several prognostic genes was linked to greater medication resistance against a variety of FDA-authorized chemotherapeutic medicines, including bosutinib, lapatinib, tamoxifen, IPI145, and idelalisib, using data from NCI-60 cell lines. Of course, some prognostic genes have been connected with higher drug sensitivity for a variety of medications. For instance, increased expression of CD69, BTG2, MMP14, PCDH7, GPC3, and NMUR1 was linked to greater drug sensitivity of cancer cells to drugs such as oxaliplatin, vemurafenib, trametinib, paclitaxel, vinblastine, etc. The MRP family consists of 13 members, with MRP1–MRP9 being the primary transporters implicated in multidrug resistance *via* extruding anticancer medicines from tumor cells (Sodani et al., 2012). Therefore, the relationship between medication resistance genes including MRP1 and risk score indicated that targeting tumor medication resistance genes might be a potential therapeutic option for high-risk patients, while MRP3 is the opposite. These findings established that various prognostic genes might be employed therapeutically as targets for overcoming adjuvant drug sensitivity or drug resistance.

CONCLUSION

To summarize, our research identified an 8-gene inflammatory response signature as a novel predictive factor. In both the

TCGC and GEO validation cohorts, the signature was identified as being independently linked with OS, and it was also found to be useful in treatment sensitivity, functional analysis, and TME, and providing understanding in forecasting the prognosis of LUAD. The exact process linking inflammatory response-related genes to tumor immunity in LUAD is unknown, and more research is needed. Our research will go an extra mile toward elucidating their function in carcinogenesis, especially in the fields of drug resistance, TME, and immune response, which is critical for developing individualized cancer therapeutics.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.cancer.gov/tcga http://www.ncbi.nlm.nih.gov/geo/. GSE68465.

ETHICS STATEMENT

The study was reviewed and approved by the Ethics Committee of The First Affiliated Hospital of Soochow University.

AUTHOR CONTRIBUTIONS

Conception and design: WZ, LC, and CH; Collection and assembly of data: WZ, LC, WM, and SH; Data analysis and interpretation: WZ, LC, WM, and YL; Manuscript writing: All authors; Final approval of manuscript: All authors.

FUNDING

This work has been supported by the Gusu health talent project of Suzhou (GSWS2020003) Suzhou.

ACKNOWLEDGMENTS

Thanks to all the peer reviewers and editors for their opinions and suggestions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2021.772206/full#supplementary-material

REFERENCES

Balkwill, F., and Mantovani, A. (2001). Inflammation and Cancer: Back to Virchow. *The Lancet* 357 (9255), 539–545. doi:10.1016/s0140-6736(00)04046-0

Bao, L., Zhang, Y., Wang, J., Wang, H., Dong, N., Su, X., et al. (2016). Variations of Chromosome 2 Gene Expressions Among Patients with Lung Cancer or Non-cancer. *Cell Biol Toxicol.* 32 (5), 419–435. doi:10.1007/s10565-016-9343-z

Budhu, A., Forgues, M., Ye, Q.-H., Jia, H.-L., He, P., Zanetti, K. A., et al. (2006). Prediction of Venous Metastases, Recurrence, and Prognosis in Hepatocellular Carcinoma Based on a Unique Immune Response Signature of the Liver Microenvironment. *Cancer Cell* 10 (2), 99–111. doi:10.1016/j.ccr.2006.06.016

Cancer, N. (2014). Comprehensive Molecular Profiling of Lung Adenocarcinoma. *Nature* 511 (7511), 543–550. doi:10.1038/nature13385

Chen, Q., Boire, A., Jin, X., Valiente, M., Er, E. E., Lopez-Soto, A., et al. (2016). Carcinoma-astrocyte gap Junctions Promote Brain Metastasis by cGAMP Transfer. *Nature* 533 (7604), 493–498. doi:10.1038/nature18268

Chen, Z., Chen, X., Lu, B., Gu, Y., Chen, Q., Lei, T., et al. (2020). Up-regulated LINC01234 Promotes Non-small-cell Lung Cancer Cell Metastasis by Activating VAV3 and Repressing BTG2 Expression. *J. Hematol. Oncol.* 13 (1), 7. doi:10.1186/s13045-019-0842-2

Chou, C.-W., Lin, C.-H., Hsiao, T.-H., Lo, C.-C., Hsieh, C.-Y., Huang, C.-C., et al. (2019). Therapeutic Effects of Statins against Lung Adenocarcinoma via P53 Mutant-Mediated Apoptosis. *Sci. Rep.* 9 (1), 20403. doi:10.1038/s41598-019-56532-6

Dib, L., San-Jose, L. M., Ducrest, A. L., Salamin, N., and Roulin, A. (2017). Selection on the Major Color Gene Melanocortin-1-Receptor Shaped the Evolution of the Melanocortin System Genes. *Int. J. Mol. Sci.* 18 (12). doi:10.3390/ijms18122618

Engblom, C., Pfirschke, C., Zilionis, R., Da Silva Martins, J., Bos, S. A., Courties, G., et al. (2017). Osteoblasts Remotely Supply Lung Tumors with Cancer-Promoting SiglecFhigh Neutrophils. *Science* 358 (6367). doi:10.1126/science.aal5081

Goldstraw, P., Chansky, K., Crowley, J., Rami-Porta, R., Asamura, H., Eberhardt, W. E., et al. (2016). The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *J. Thorac. Oncol.* 11 (1), 39–51. doi:10.1016/j.jtho.2015.09.009

Greten, F. R., and Grivennikov, S. I. (2019). Inflammation and Cancer: Triggers, Mechanisms, and Consequences. *Immunity* 51 (1), 27–41. doi:10.1016/j.immuni.2019.06.025

Huang, Z., Cheng, L., Guryanova, O. A., Wu, Q., and Bao, S. (2010). Cancer Stem Cells in Glioblastoma-Molecular Signaling and Therapeutic Targeting. *Protein Cell* 1 (7), 638–655. doi:10.1007/s13238-010-0078-y

Infante, E., Castagnino, A., Ferrari, R., Monteiro, P., Agüera-Gonza´lez, S., Paul-Gilloteaux, P., et al. (2018). LINC Complex-Lis1 Interplay Controls MT1-MMP Matrix Digest-On-Demand Response for Confined Tumor Cell Migration. *Nat. Commun.* 9 (1), 2443. doi:10.1038/s41467-018-04865-7

Kim, H., Xu, G.-L., Borczuk, A. C., Busch, S., Filmus, J., Capurro, M., et al. (2003). The Heparan Sulfate ProteoglycanGPC3Is a Potential Lung Tumor Suppressor. *Am. J. Respir. Cel Mol Biol.* 29 (6), 694–701. doi:10.1165/rcmb.2003-0061oc

Koliaraki, V., Prados, A., Armaka, M., and Kollias, G. (2020). The Mesenchymal Context in Inflammation, Immunity and Cancer. *Nat. Immunol.* 21 (9), 974–982. doi:10.1038/s41590-020-0741-2

Kris, M. G., Johnson, B. E., Berry, L. D., Kwiatkowski, D. J., Iafrate, A. J., Wistuba, II, et al. (2014). Using Multiplexed Assays of Oncogenic Drivers in Lung Cancers to Select Targeted Drugs. *JAMA* 311 (19), 1998–2006. doi:10.1001/jama.2014.3741

Li, R., Fang, F., Jiang, M., Wang, C., Ma, J., Kang, W., et al. (2017). STAT3 and NF-Kb Are Simultaneously Suppressed in Dendritic Cells in Lung Cancer. *Sci. Rep.* 7, 45395. doi:10.1038/srep45395

Lin, Z., Xu, Q., Miao, D., and Yu, F. (2021). An Inflammatory Response-Related Gene Signature Can Impact the Immune Status and Predict the Prognosis of Hepatocellular Carcinoma. *Front. Oncol.* 11, 644416. doi:10.3389/fonc.2021.644416

Malta, T. M., Sokolov, A., Gentles, A. J., Burzykowski, T., Poisson, L., Weinstein, J. N., et al. (2018). Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* 173 (2), 338–354. doi:10.1016/j.cell.2018.03.034

Maly, V., Maly, O., Kolostova, K., and Bobek, V. (2019). Circulating Tumor Cells in Diagnosis and Treatment of Lung Cancer. *In Vivo* 33 (4), 1027–1037. doi:10.21873/invivo.11571

Mao, K., Lin, F., Zhang, Y., and Zhou, H. (2021). Identification of Key Genes and Pathways in Gefitinib-Resistant Lung Adenocarcinoma Using Bioinformatics Analysis. *Evol. Bioinform Online* 17, 11769343211023767. doi:10.1177/11769343211023767

McMillan, D. C. (2013). The Systemic Inflammation-Based Glasgow Prognostic Score: a Decade of Experience in Patients with Cancer. *Cancer Treat. Rev.* 39 (5), 534–540. doi:10.1016/j.ctrv.2012.08.003

Mohrherr, J., Haber, M., Breitenecker, K., Aigner, P., Moritsch, S., Voronin, V., et al. (2019). JAK-STAT Inhibition Impairs K-RAS-driven Lung Adenocarcinoma Progression. *Int. J. Cancer* 145 (12), 3376–3388. doi:10.1002/ijc.32624

Moik, F., Zöchbauer-Müller, S., Posch, F., Pabinger, I., and Ay, C. (2020). Systemic Inflammation and Activation of Haemostasis Predict Poor Prognosis and Response to Chemotherapy in Patients with Advanced Lung Cancer. *Cancers (Basel)* 12 (6). doi:10.3390/cancers12061619

Ning, J., Jiang, S., Li, X., Wang, Y., Deng, X., Zhang, Z., et al. (2021). GPC3 Affects the Prognosis of Lung Adenocarcinoma and Lung Squamous Cell Carcinoma. *BMC Pulm. Med.* 21 (1), 199. doi:10.1186/s12890-021-01549-9

Nost, T. H., Alcala, K., Urbarova, I., Byrne, K. S., Guida, F., Sandanger, T. M., et al. (2021). *Systemic Inflammation Markers and Cancer Incidence in the UK Biobank*. Langnes, troms Norway: Eur J Epidemiol.

Schutyser, E., Struyf, S., and Van Damme, J. (2003). The CC Chemokine CCL20 and its Receptor CCR6. *Cytokine Growth Factor. Rev.* 14 (5), 409–426. doi:10.1016/s1359-6101(03)00049-2

Seijo, L. M., Peled, N., Ajona, D., Boeri, M., Field, J. K., Sozzi, G., et al. (2019). Biomarkers in Lung Cancer Screening: Achievements, Promises, and Challenges. *J. Thorac. Oncol.* 14 (3), 343–357. doi:10.1016/j.jtho.2018.11.023

Shankaran, V., Ikeda, H., Bruce, A. T., White, J. M., Swanson, P. E., Old, L. J., et al. (2001). IFNγ and Lymphocytes Prevent Primary Tumour Development and Shape Tumour Immunogenicity. *Nature* 410 (6832), 1107–1111. doi:10.1038/35074122

Shinchi, Y., Komohara, Y., Yonemitsu, K., Sato, K., Ohnishi, K., Saito, Y., et al. (2019). Accurate Expression of PD-L1/L2 in Lung Adenocarcinoma Cells: A Retrospective Study by Double Immunohistochemistry. *Cancer Sci.* 110 (9), 2711–2721. doi:10.1111/cas.14128

Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2011). Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J. Stat. Softw.* 39 (5), 1–13. doi:10.18637/jss.v039.i05

Sodani, K., Patel, A., Kathawala, R. J., and Chen, Z.-S. (2012). Multidrug Resistance Associated Proteins in Multidrug Resistance. *Chin. J. Cancer* 31 (2), 58–72. doi:10.5732/cjc.011.10329

Stawowczyk, M., Wellenstein, M. D., Lee, S. B., Yomtoubian, S., Durrans, A., Choi, H., et al. (2017). Matrix Metalloproteinase 14 Promotes Lung Cancer by Cleavage of Heparin-Binding EGF-like Growth Factor. *Neoplasia* 19 (2), 55–64. doi:10.1016/j.neo.2016.11.005

Street, S. E. A., Cretney, E., and Smyth, M. J. (2001). Perforin and Interferon-γ Activities Independently Control Tumor Initiation, Growth, and Metastasis. *Blood* 97 (1), 192–197. doi:10.1182/blood.v97.1.192

Stutvoet, T. S., Kol, A., Vries, E. G., Bruyn, M., Fehrmann, R. S., Terwisscha van Scheltinga, A. G., et al. (2019). MAPK Pathway Activity Plays a Key Role in PD-L1 Expression of Lung Adenocarcinoma Cells. *J. Pathol.* 249 (1), 52–64. doi:10.1002/path.5280

Sun, J., Zhao, T., Zhao, D., Qi, X., Bao, X., Shi, R., et al. (2020). Development and Validation of a Hypoxia-Related Gene Signature to Predict Overall Survival in Early-Stage Lung Adenocarcinoma Patients. *Ther. Adv. Med. Oncol.* 12, 1758835920937904. doi:10.1177/1758835920937904

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

Suzuki, K., Kadota, K., Sima, C. S., Nitadori, J.-i., Rusch, V. W., Travis, W. D., et al. (2013). Clinical Impact of Immune Microenvironment in Stage I Lung Adenocarcinoma: Tumor Interleukin-12 Receptor β2 (IL-12Rβ2), IL-7R, and Stromal FoxP3/CD3 Ratio Are Independent Predictors of Recurrence. *Jco* 31 (4), 490–498. doi:10.1200/jco.2012.45.2052

Tamborero, D., Rubio-Perez, C., Muiños, F., Sabarinathan, R., Piulats, J. M., Muntasell, A., et al. (2018). A Pan-Cancer Landscape of Interactions between Solid Tumors and Infiltrating Immune Cell Populations. *Clin. Cancer Res.* 24 (15), 3717–3728. doi:10.1158/1078-0432.ccr-17-3509

Wadowska, K., Bil-Lula, I., Trembecki, Ł., and Śliwińska-Mossoń, M. (2020). Genetic Markers in Lung Cancer Diagnosis: A Review. *Int. J. Mol. Sci.* 21 (13). doi:10.3390/ijms21134569

Wallrapp, A., Riesenfeld, S. J., Burkett, P. R., Abdulnour, R.-E. E., Nyman, J., Dionne, D., et al. (2017). The Neuropeptide NMU Amplifies ILC2-Driven Allergic Lung Inflammation. *Nature* 549 (7672), 351–356. doi:10.1038/nature24029

Wang, L., Wang, Y., Song, Z., Chu, J., and Qu, X. (2015). Deficiency of Interferon-Gamma or its Receptor Promotes Colorectal Cancer Development. *J. Interferon Cytokine Res.* 35 (4), 273–280. doi:10.1089/jir.2014.0132

Welsh, T. J., Green, R. H., Richardson, D., Waller, D. A., O'Byrne, K. J., and Bradding, P. (2005). Macrophage and Mast-Cell Invasion of Tumor Cell Islets Confers a Marked Survival Advantage in Non-small-cell Lung Cancer. *Jco* 23 (35), 8959–8967. doi:10.1200/jco.2005.01.4910

Yi, M., Li, A., Zhou, L., Chu, Q., Luo, S., and Wu, K. (2021). Immune Signature-Based Risk Stratification and Prediction of Immune Checkpoint Inhibitor's Efficacy for Lung Adenocarcinoma. *Cancer Immunol. Immunother.* 70 (6), 1705–1719. doi:10.1007/s00262-020-02817-z

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring Tumour Purity and Stromal and Immune Cell Admixture from Expression Data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612

Zhang, A., Yang, J., Ma, C., Li, F., and Luo, H. (2021). Development and Validation of a Robust Ferroptosis-Related Prognostic Signature in Lung Adenocarcinoma. *Front. Cel Dev. Biol.* 9, 616271. doi:10.3389/fcell.2021.616271

Zhang, H., Guo, W., Zhang, F., Li, R., Zhou, Y., Shao, F., et al. (2020). Monoacylglycerol Lipase Knockdown Inhibits Cell Proliferation and Metastasis in Lung Adenocarcinoma. *Front. Oncol.* 10, 559568. doi:10.3389/fonc.2020.559568

Zhang, L., Zhang, Z., and Yu, Z. (2019). Identification of a Novel Glycolysis-Related Gene Signature for Predicting Metastasis and Survival in Patients with Lung Adenocarcinoma. *J. Transl Med.* 17 (1), 423. doi:10.1186/s12967-019-02173-2

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

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 Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 23 December 2021
doi: 10.3389/fbioe.2021.792911

Downregulated ADARB1 Facilitates Cell Proliferation, Invasion and has Effect on the Immune Regulation in Ovarian Cancer

*Wei Zhu[1], Zhijie Xu[1,2], Meiyuan Huang[3], Xiang Wang[4], Xinxin Ren[5], Yuan Cai[1], Bi Peng[1], Qiuju Liang[4], Xi Chen[4] and Yuanliang Yan[4]**

[1]Department of Pathology, Xiangya Hospital, Central South University, Changsha, China, [2]National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China, [3]Department of Pathology, The Affiliated Zhuzhou Hospital Xiangya Medical College, Central South University, Zhuzhou, China, [4]Department of Pharmacy, Xiangya Hospital, Central South University, Changsha, China, [5]Key Laboratory of Molecular Radiation Oncology of Hunan Province, Xiangya Hospital, Central South University, Changsha, China

Ovarian cancer (OC) is typically diagnosed at an advanced stage and poses a significant challenge to treatment and recovery. Rencently, Adenosine deaminase RNA-specific B1 (ADARB1), an adenosine-to-inosine (A-to-I) RNA-editing enzyme, has been found to play an essential role in the development of cancer. However, the specific function of ADARB1 in ovarian cancer is still not fully understood. Here, we investigated the effects of ADARB1 on OC biology. By conducting bioinformatics analyses of several public databases, we found significantly decreased ADARB1 expression in OC cells and tissues. Moreover, RT-PCR and western blot showed lower ADARB1 expression in OVCAR3, HO8910pm and A2780 OC cells compared to human normal ovarian epithelial cell IOSE. Cell proliferation assay and clone formation assay showed that overexpression of ADARB1 (ADARB1-OE) inhibited the proliferation of tumor cells. Wound healing and transwell assay indicated that ADARB1-OE could suppress OC cell invasion and metastasis. Kaplan-Meier methods revealed that the patients with low level of ADARB1 displayed poor prognosis. TISIDB databases were further used to analyze the roles of ADARB1 in tumor-immune system interactions in OC patients. Furthermore, ADARB1-OE down-regulated the expression of phosphorylated AKT. Combination of ADARB1-OE and AKT inhibitor MK2206 exerted stronger cell growth inhibition. Thus, our investigation demonstrated that low levels of ADARB1 might be a potential target in the tumorigenesis and prognostic evaluation of OC patients.

Keywords: ADARB1, ovarian cancer, Akt, cell growth, immune regulation

OPEN ACCESS

Edited by:
Giulia Russo,
University of Catania, Italy

Reviewed by:
Jes-niels Boeckel,
University Hospital Leipzig, Germany
Rui Liu,
Sichuan University, China

****Correspondence:***
Yuanliang Yan
yanyuanliang@csu.edu.cn

Specialty section:
This article was submitted to Preclinical Cell and Gene Therapy, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *11 October 2021*
Accepted: *08 December 2021*
Published: *23 December 2021*

Citation:
Zhu W, Xu Z, Huang M, Wang X, Ren X, Cai Y, Peng B, Liang Q, Chen X and Yan Y (2021) Downregulated ADARB1 Facilitates Cell Proliferation, Invasion and has Effect on the Immune Regulation in Ovarian Cancer. Front. Bioeng. Biotechnol. 9:792911. doi: 10.3389/fbioe.2021.792911

INTRODUCTION

Ovarian cancer (OC) is one of the most common gynecological tumors, and the onset age is mainly concentrated in people between 40 and 60 years old, especially in people around 50 years old. It accounts for 2.4–6.5% of the common female malignant tumors, and ranks third in female reproductive system malignant tumors, next only to cervical cancer and endometrial cancer (Brunty et al., 2020; Hidayat et al., 2020). The standard treatment for OC patients is surgery

and chemotherapy. Due to the high recurrence rate and therapeutic resistance, the 5-year survival rate for OC patients is about 47% (Demircan et al., 2020; Moschetta et al., 2020). Therefore, it is necessary to search for novel target molecules for improving the treatment and prognosis of OC.

Adenosine deaminase RNA specific B1 (ADARB1) catalyzes the conversion of adenosine (A) to inosine (I) in double-stranded RNAs (Ma et al., 2020). Recently, emerging researches have revealed the relationship between ADARB1 and cancer. In glioma, aberrant expression of ADARB1 can affect the proliferation and invasion of glioma cells by regulating microRNAs (Cesarini et al., 2018) or CDC14B-Skp2 signaling axis (Galeano et al., 2013). The down-regulation of ADARB1 may be related to the pathogenesis of non-small cell lung cancer (Wang et al., 2019; Wang et al., 2020), and low levels of ADARB1 in lung cancer are correlated with shorter first progression (FP), overall survival time (OS) and post-progression survival time (PPS). Moreover, ADAR2 suppresses tumor growth and induces apoptosis by editing and stabilizing IGFBP7 in esophageal squamous cell carcinoma (Chen et al., 2017). However, the association between ADARB1 and OC has not been investigated.

The purpose of our study was to evaluate the role and mechanism of ADARB1 in human OC. Through bioinformatics data analysis and *in vitro* experiments, we found that ADARB1 was downregulated in OC tissues and cell lines, and was correlated with poor prognoses of OC patients. Furthermore, overexpression of ADARB1 (ADARB1-OE) significantly inhibited the proliferation, invasion and metastasis of OC cells by down-regulating AKT phosphorylation.

MATERIALS AND METHODS

Data Acquisition and Reanalysis

As mentioned previously (Yan et al., 2018; Zhou et al., 2019) several bioinformatics network resources were used to reanalyze the molecular profiles of ADARB1 in OC patients (**Supplementary Table S1**). Gene Expression Profiling Interactive Analysis (GEPIA) (Tang et al., 2017) and University of Alabama Cancer Database (UALCAN) (Chandrashekar et al., 2017) were used to identify the expression profiles of ADARB1 in OC tissues. For the prognostic analysis, the Kaplan–Meier Plotter (Lánczky et al., 2016) and DRUGSURV database (Amelio et al., 2014) were utilized to describe the relationship between ADARB1 expression and patients' prognosis.

From the cBioPortal web tool (Gao et al., 2013), the genes coexpressed with ADARB1 in OC were downloaded (**Supplementary Table S2**). Then, the STRING (Szklarczyk et al., 2017) and Cytoscape software (Reimand et al., 2019) were used to complete the protein–protein interaction (PPI) network of these coexpression genes. Then, we utilized the WebGestalt (Wang et al., 2017) to conduct the GO and KEGG pathway analysis of ADARB1 coexpression genes in OC samples.

Immunohistochemistry

The ovarian cancer tissues and paired adjacent tissues were obtained from the Xiangya hospital, central south university. The ethical approval number is 20210205. Specimens were deparaffinized in xylene and rehydrated in a series of graded alcohol. Antigen retrieval was completed after heating in citrate buffer. Endogenous peroxidase was blocked using 3% H_2O_2 for 20 min. The sections were incubated with the primary antibody against ADARB1, p-AKT, CXCL12 and KDR at 37°C for 1 h. Horseradish peroxidase-conjugated secondary antibody was added and incubated at room temperature for 30 min, and 3,3′-diaminobenzitine (DAB) solution was used for color reaction (ZSGB-BIO, China).

Cells and Reagents

Human OC cells OVCAR3, HO8910pm, A2780 and normal ovarian epithelial cells IOSE were obtained from department of Pathology, School of Basic Medical Sciences, Central South University. OVCAR3, HO8910pm and IOSE were maintained in Roswell Park Memorial Institute (RPMI)-1640 medium (BI, Israel Beit Haemek Ltd.) with 10% fetal bovine serum (FBS), while A2780 was grown in DMEM (BI, Israel Beit Haemek Ltd.) supplemented with 10% fetal bovine serum FBS at 37°C and 5% CO2. MK2206 was purchased from Selleck Chemicals and dissolved in dimethylsulfoxide (DMSO) (Sigma, United States), The exposed concentrations of MK2206 were 5 mM. Antibodies against ADARB1 and p-AKT were purchased from Proteintech (United States, 22248-1-AP and 66444-1-Ig). Antibodies against GAPDH and CXCL12 were purchased from Abclonal (China, AC002 and A1325), Antibody against KDR was purchased from Affinity (China, AF6281). Antibodies against AKT and p-AKT were from Cell Signaling Technology (United States). The second antibody were purchased from ZSGB-BIO.

Western Blot

Cells were lysed in RIPA buffer and protein concentrations were determined by a BCA protein assay kit (PP1002, BioTeck, Beijing, China). A total of 30–40 μg of protein was separated by 10% SDS-PAGE and electroblotted onto polvinylidene fluoride membranes (Millipore, Merck, United States) for detection using indicated antibodies. Immunoreactions were detected by an imaging system (Alpha FluorChem FC3).

RNA Extraction and Reverse Transcription PCR

Total RNA was extracted employing TRIzol reagent (Invitrogen) according to the manufacturter's instruction and reverse transcribed to cDNA using the PrimeScript™ RT reagent kit (Abclonal, China). The RT-PCR assay was conducted through iTaqTM Universal SYBR green Supermix (Abclonal, China), with GAPDH as the internal control. The forward and reverse primer sequences were used as follows: ADARB1: 5′-CGCGCCTTTGTT TGTCATGTC-3′ and 5′-GGAAACTGAACGAAAGACCTC AA-3′; GAPDH: 5′-CAGCAAGAGCACAAGAGGAA-3′ and 5′-ATGGTACATGACAAGGTGCGG-3′. Relative expression levels were decided using the 2-ΔΔCT method.

MTT Asaay

The logarithmic growth phase cells were cultured in a 96-well culture plate for 6 days with five replicates per group. MTT assay was performed 48 h after transfection. Add 10 μl MTT solution to

each well at the end of each day. After 4 h of continuous culture, the culture medium was discarded, and DMSO (150 μl/well) was added 6 days later, and the medium was shaken for 10 min. The cell proliferation rate was calculated by measuring the optical density (OD) of each well at 490 nm on a Microplate Reader.

Colony Formation Assay

A total of 1,000 cells were seeded in 6-well plates and incubated at 37°C with 5% CO2 for 10 days. Culture plates were performed in duplicates. After a wash with phosphate buffer saline (PBS), cultures were fixed with methanol for 20 min and stained with crystal violet for 15 min. Colonies were examined and calculated.

Wound Healing Assay

The cells were inoculated into 6-well plates and cultured in complete medium (37°C, 5% CO2) to at least 95% confluence before wounds were created. To measure the cell migration, a plastic 100 ul pipette tip was used to scrape cells in a monolayer to creating wounds. Then, washed them three times in PBS and incubated with FBS-free RPMI-1640 medium. Subsequently, cells were cultured in either medium for 0, 12, 24 h. At the end of the incubation period, phase-contrast microscopy was employed to photograph the wounded area and migration cells at the wounded area. Finally, the extent to which the wound had closed over 24 h was calculated and expressed as a percentage of the difference between times 0 and 24 h using ImageJ software.

Transwell Invasion Assays

For invasion assays, Transwells (corning, United States) with 8-μm pore size filters covered with matrigel were inserted into 24-well plates. The cells were serum-starved overnight and then added in the upper chamber (5×10^4 cells per insert) and the complete culture medium supplemented with 10% FBS was used as a chemoattractant in the lower chamber. After 24 h of incubation, non-invading cells that remained on the upper surface of the filter were removed, and the cells that had passed through the filter and attached to the bottom of the membrane were fixed in methanol and stained with 0.2% Crystal violet. Numbers of the invasive cells in seven randomly selected fields from triplicate chambers were counted in each experiment under a phase-contrast microscope.

Statistical Analysis

Statistical analyses were performed using SPSS 22.0 for Windows. Data were presented as mean ± standard deviation (SD) of at least three independent experiments. Statistical analysis was performed using Student's t test and one-way ANOVA. The Statistical significance is shown in the figures with $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$ and $^{****}p < 0.0001$.

RESULTS

ADARB1 Was Downregulated in OC Tissues and Cell Lines

To evaluate changes in ADARB1 expression in OC and normal ovarian tissues, we analyzed the expression levels of ADARB1 through several bioinformatics databases. The GEPIA analysis revealed that ADARB1 mRNA expression was significantly decreased in OC tissues (**Figure 1A**). From TCGA-OV, we discovered the ADARB1 expression was clearly reduced in OC compared with the normal tissues (**Figure 1B**). Similarly, using the TNM plot (**Figure 1C**) and UALCAN (**Figure 1D**), we further confirmed the downregulation of ADARB1 in OC tissues. In addition, we found that OC cells OVCAR3, HO8910pm and A2780 displayed a lower level of ADARB1 compared to the normal ovarian cell IOSE (**Figures 1E,F**). Moreover, the expression of ADARB1 in tumor tissues was lower than that in adjacent tissues (**Supplementary Figure S1**). Next, we analyzed the effects of ADARB1 expression on the clinical characteristics of patients with OC. From DRUGSURV database, we found that the patients with low level of ADARB2 displayed poor prognosis in GSE14764 (**Figure 2A**). In addition, Kaplan-Meier Plotter database was employed to further evaluate the effects of ADARB1 expression on survival, revealing that low levels of ADARB1 expression were correlated with shorter progression-free survival (PFS) (**Figure 2B**). Therefore, these results suggested that ADARB1 expression levels might serve as an indicator for the clinical prognosis of OC patients. Taken together, the decreased expression of ADARB1 in OC tissues and cell lines suggested its anti-oncogenic roles.

ADARB1 Could Suppress the Malignant Biological Behaviors of OC Cells

To further identify the function of ADARB1 in OC, we overexpressed of ADARB1 in HO8910pm and OVCAR3 ovarian cancer cells (**Figures 3A,B**). After then, we tested the roles of ADARB1 overexpression in cell proliferation, migration and invasion. The MTT assay and colony formation assay showed that ADARB1 overexpressed significantly decreased the proliferation ability in HO8910pm and OVCAR3 cells (**Figures 3C–F**). Similarly, we observed that ADARB1-OE inhibited the metastasis and invasion ability in HO8910pm and OVCAR3 cells, as indicated by the wound healing and transwell assays (**Figures 4A–D**). These results suggest that ADARB1 could suppress the malignant biological behaviors of OC cells.

ADARB1 Suppressed OC Proliferation and Metastasis by Inhibiting AKT Phosphorylation

The PI3K-AKT pathway is a well-known tumor signaling pathway that is involved in the development of various tumors, including glioma (Dai et al., 2017), ovarian cancer (Wu et al., 2021), etc. Moreover, phosphorylation activation of AKT signaling can promote the malignant progression of ovarian cancer (Samartzis et al., 2020; Zhai et al., 2020). We found that the phosphorylation levels of AKT (p-AKT) in tumor tissues was higher than that in adjacent tissues (**Supplementary Figure S1**). And ADARB1-OE can effectively inhibit p-AKT without affecting the total AKT level. Then, we investigated whether

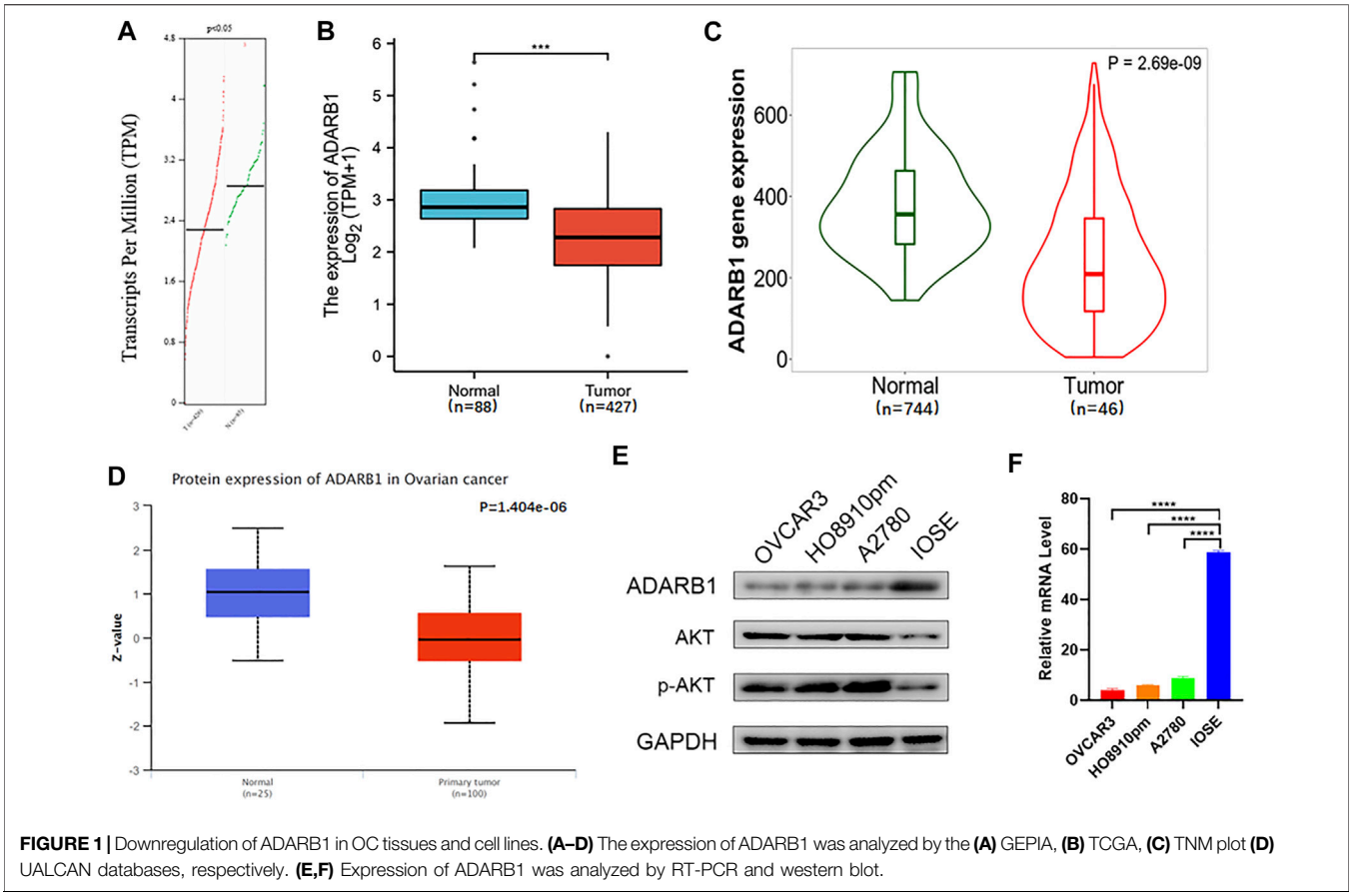


FIGURE 1 | Downregulation of ADARB1 in OC tissues and cell lines. **(A–D)** The expression of ADARB1 was analyzed by the **(A)** GEPIA, **(B)** TCGA, **(C)** TNM plot **(D)** UALCAN databases, respectively. **(E,F)** Expression of ADARB1 was analyzed by RT-PCR and western blot.

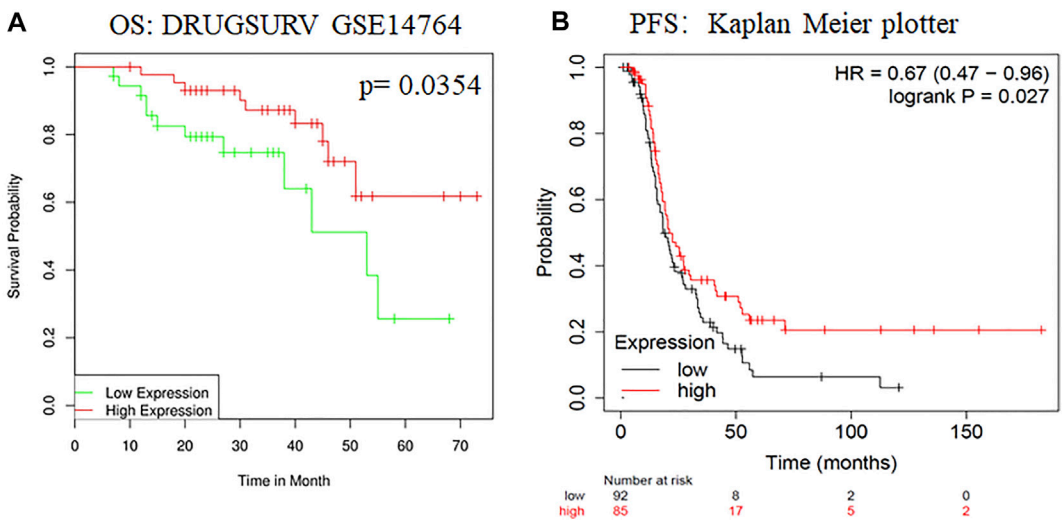


FIGURE 2 | The effects of ADARB1 expression on the prognosis of OC patients. **(A)** The relationship between ADARB1 expression and OS analyzed by DRUGSURV database. **(B)** The relationship between ADARB1 expression and PFS analyzed by Kaplan–Meier Plotter.

the effect of ADARB1 on OC cell proliferation and metastasis was associated with AKT activity. Western blot analysis showed that p-AKT was obviously decreased when OC cells were overexpressed ADARB1 or treated with MK2206, a specific inhibitor of AKT. The inhibitory effect on p-AKT levels was further increased by combinational treatment (**Figure 5A**). We

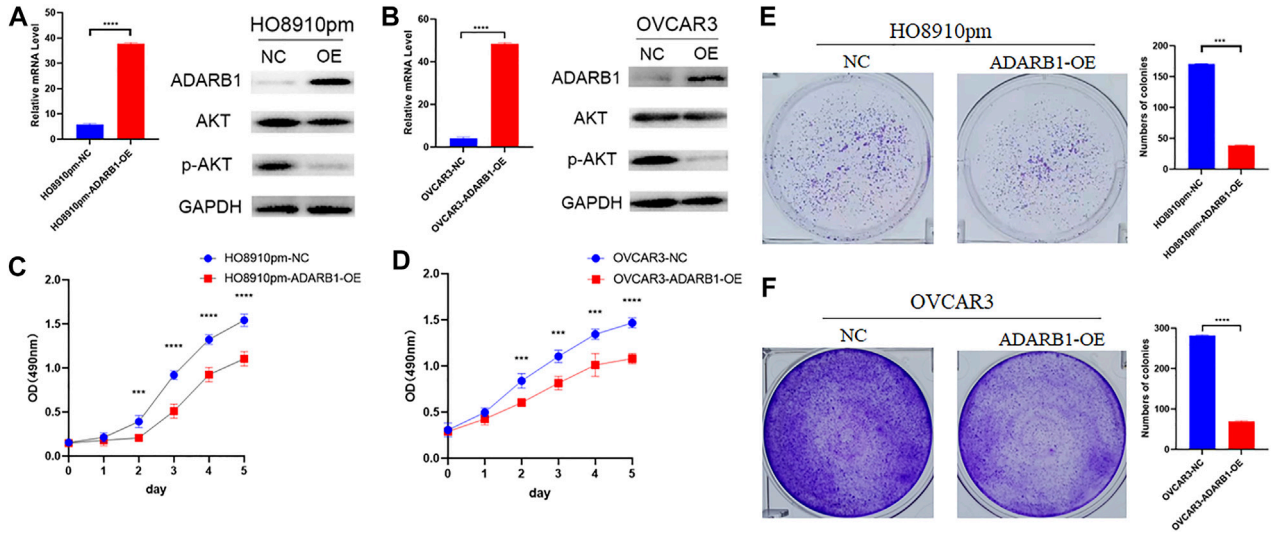


FIGURE 3 | Overexpression of ADARB1 suppressed the proliferation ability of OC cells. **(A,B)** The mRNA levels of ADARB1 after transfected with ADARB1-OE in HO8910pm and OVCAR3 cells. **(C,D)** The results of MTT assay after transfected with ADARB1-OE in HO8910pm and OVCAR3 cells. **(E,F)** The results of colony formation assay after transfected with ADARB1-OE in HO8910pm and OVCAR3 cells.

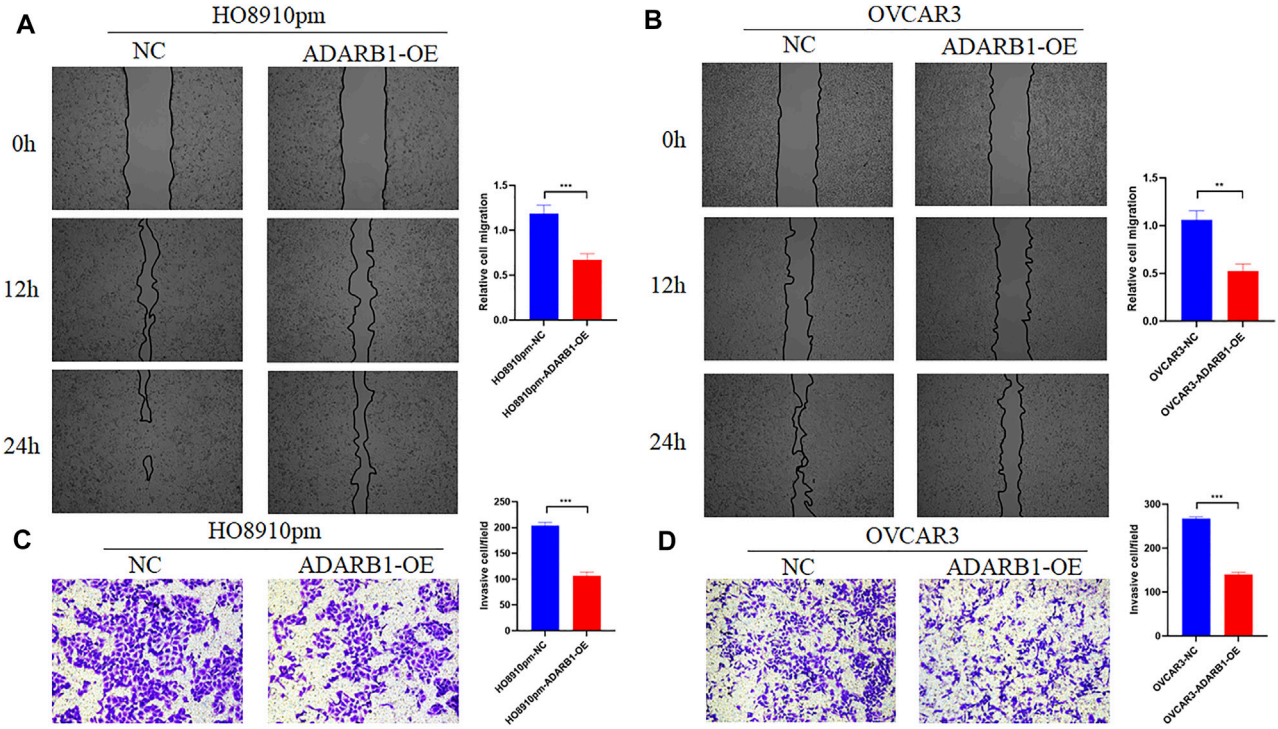


FIGURE 4 | Overexpression of ADARB1 inhibited the migration and invasion ability of OC cells. **(A,B)** The results of wound healing assay after transfected with ADARB1-OE in HO8910pm and OVCAR3 cells. **(C,D)** The results of transwell assay after transfected with ADARB1-OE in HO8910pm and OVCAR3 cells.

performed the MTT assay and colony formation assay and observed that OC cells treated with both MK2206 and ADARB1-OE had a significantly suppression effect of cell proliferation compared with the control group and either of the individual treatment groups (**Figures 5B–D**). Moreover, we examined cell migration and invasion capacity of HO8910pm and OVCAR3 cells under ADARB1-OE and MK2206 treatment. Compared to the control group, OC cells

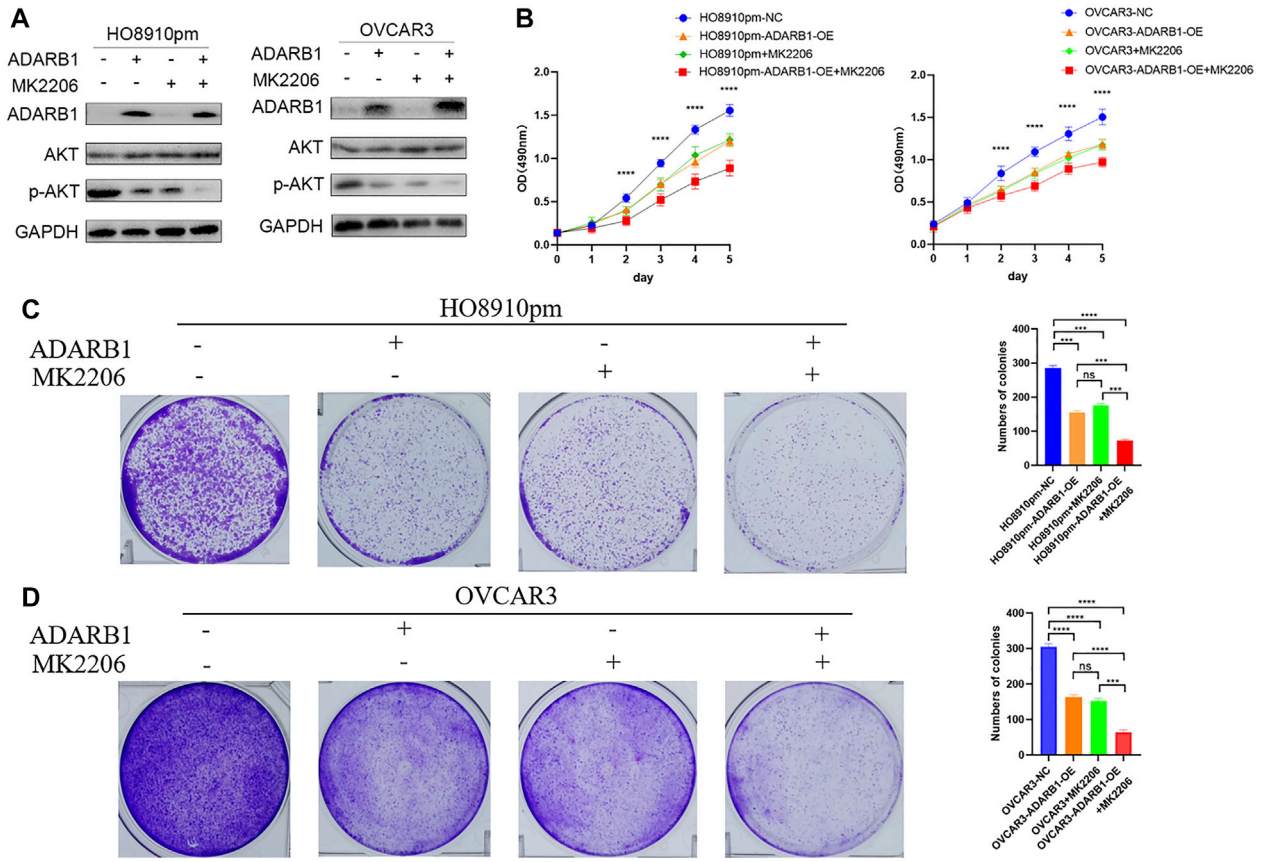


FIGURE 5 | AKT pathway was involved in ADARB1-reduced cell proliferation. **(A)** HO8910pm and OVCAR3 cells treated with ADARB1-OE and/or AKT inhibitors MK2206 subjected to Western blot assays. **(B)** The MTT assay were performed after HO8910pm and OVCAR3 cells treated with ADARB1-OE and/or AKT inhibitors MK2206. **(C,D)** The colony formation assay were performed after HO8910pm and OVCAR3 cells treated with ADARB1-OE and/or AKT inhibitors MK2206.

showed a significantly weaker migration and invasion ability when treated with MK2206 or ADARB1-OE (**Figures 6A–D**). The above results indicated that anti-cancer activity by ADARB1 overexpression might be due to the inhibition of AKT signaling.

Functional Enrichment Analysis of ADARB1-Associated Coexpression Genes

To further investigate the biological functions of ADARB1, functional enrichment analysis was performed using ADARB1 related coexpressed genes. By using the cBioPortal database, we screened 1399 ADARB1-associated coexpressed genes with the criteria of p value $\leq$ 0.05 and |Spearman's correlation|$\geq$0.3 (**Supplementary Table S2**). Next, a PPI network was performed using STRING and Cytoscape software (**Figure 7A**). Meanwhile, the GO annotation (**Figure 7B**) and KEGG pathway (**Figure 7C**) were further analyzed using WebGestalt database. Biological processes indicated that these coexpressed genes are mainly involved in metabolic processes and biological regulation. For cell component, these coexpressed genes were mainly located on the nucleus and membrane. In the aspect of molecular function, the coexpressed genes were primarily enriched in protein binding and ion binding. The KEGG pathway demonstrated that these coexpressed genes were mainly related to the ribonucleotide metabolic process and mitochondrial transport.

Regulation of Immune Molecules by ADARB1

Increasing evidence demonstrated that ADAR family members play important roles in immune regulation (Heraud-Farlow and Walkley, 2020; Yanai et al., 2020). Therefore, from TISIDB database, we investigated the relationship between ADARB1 expression and immune infiltration. As shown in **Figure 8A**, the expression levels of ADARB1 positively correlated with several immune cells, such as natural killer cell (NK cells), central Memory T cell (Tcm), T effector memory (Tem) and eosinophils. **Figure 8B** further confirmed the positive correlation between ADARB1 expression and NK cells (Spearman r = 0.546), Tcm (Spearman r = 0.45), Tem (Spearman r = 0.35) and eosinophils (Spearman r = 0.24). Similarly, ADARB1 was positively correlated with eosinophil and NK cells by using the TISIDB database (**Supplementary Figure S2**). Moreover, we analyzed the associations between ADARB1 expression and the immunomodulators and chemokines. **Supplementary**

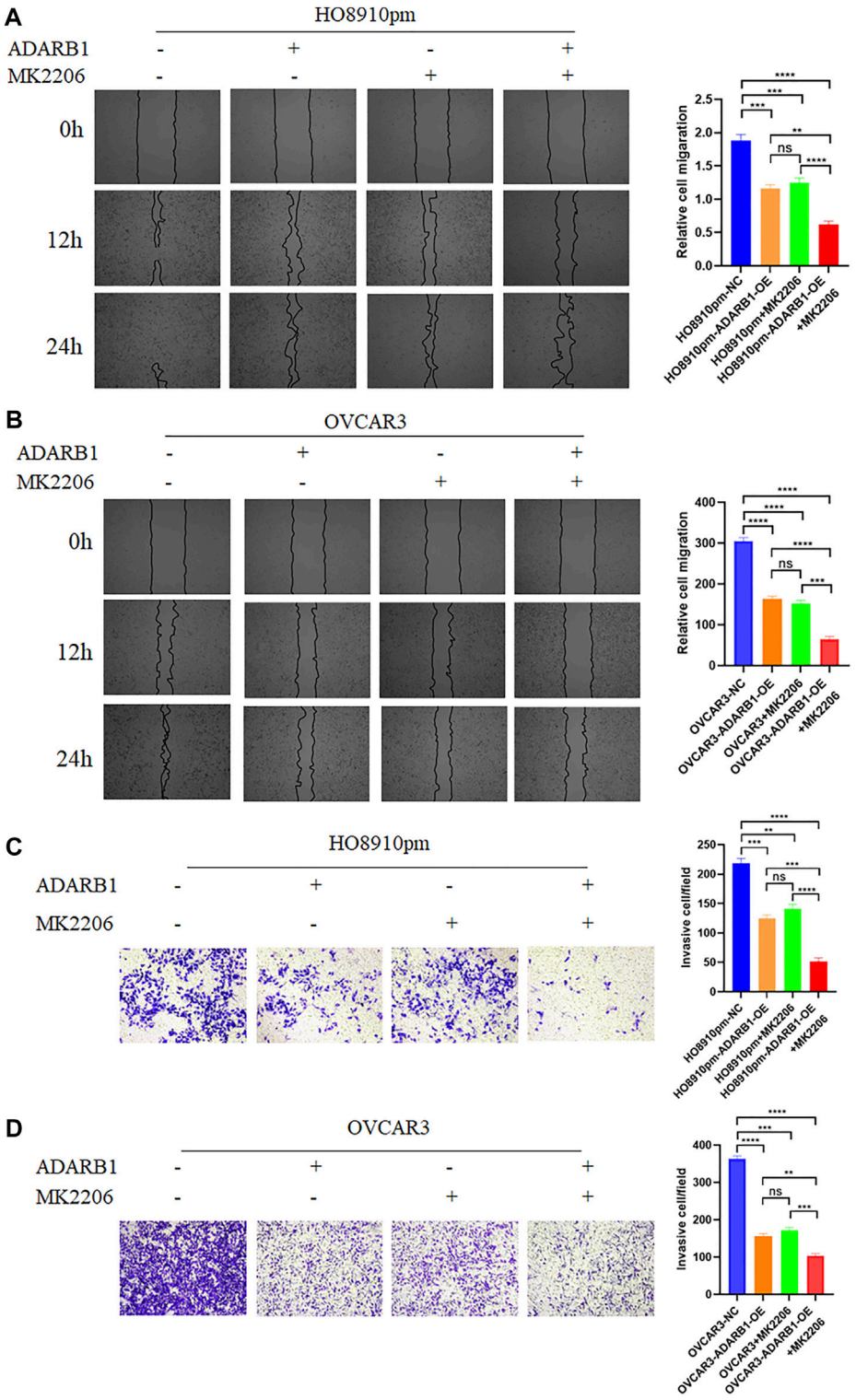


FIGURE 6 | ADARB1 inhibits OC cell Metastasis and invasion by regulating AKT pathway. **(A,B)** The wound healing assays were performed after HO8910pm and OVCAR3 cells treated with ADARB1-OE and/or AKT inhibitors MK2206. **(C,D)** The transwell assay were performed after HO8910pm and OVCAR3 cells treated with ADARB1-OE and/or AKT inhibitors MK2206.

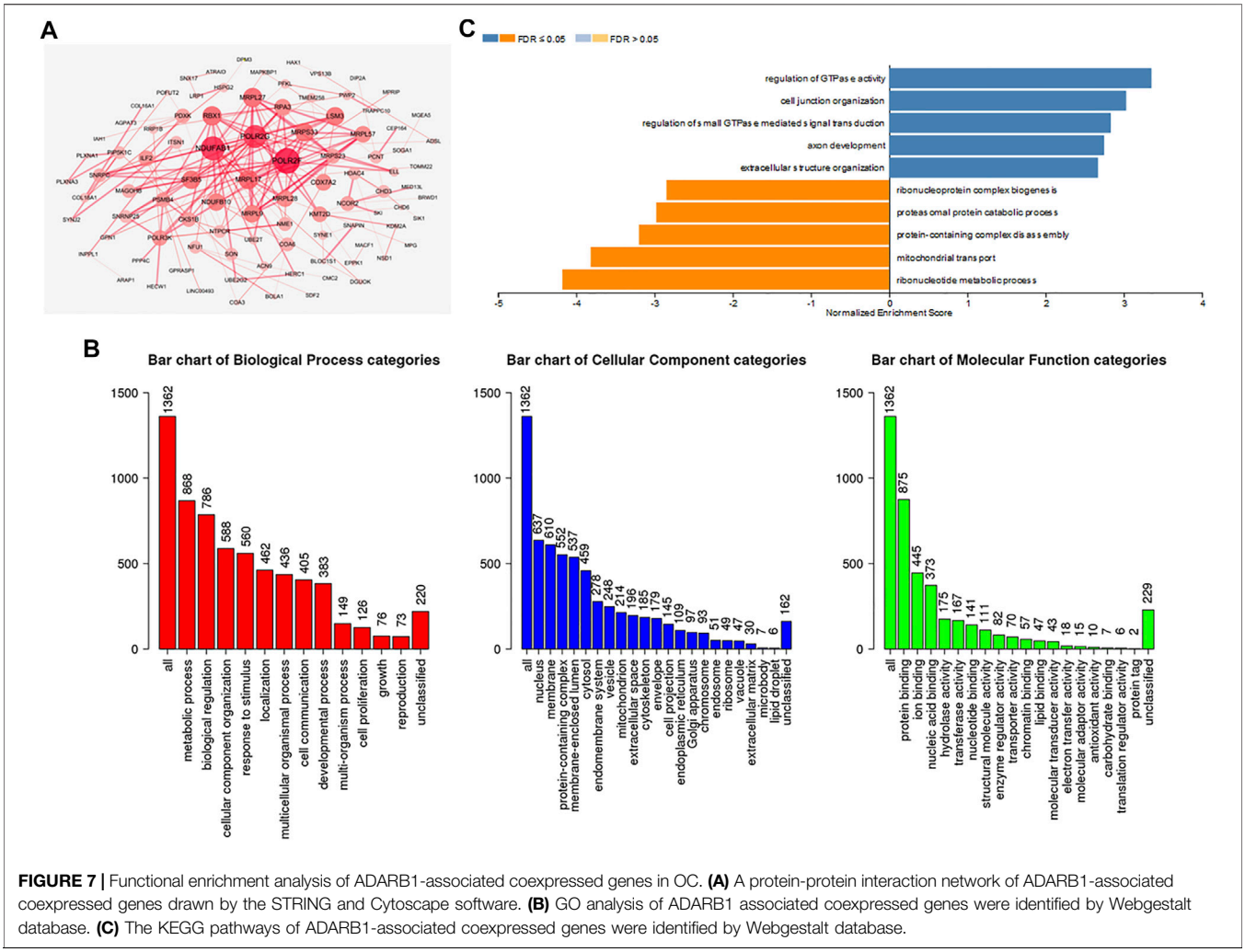


FIGURE 7 | Functional enrichment analysis of ADARB1-associated coexpressed genes in OC. **(A)** A protein-protein interaction network of ADARB1-associated coexpressed genes drawn by the STRING and Cytoscape software. **(B)** GO analysis of ADARB1 associated coexpressed genes were identified by Webgestalt database. **(C)** The KEGG pathways of ADARB1-associated coexpressed genes were identified by Webgestalt database.

Figure S3 showed the positive correlations between ADARB1 expression and several immunomodulators, including CD160 (Spearman r = 0.211), CSF1R (Spearman r = 0.231), KDR (Spearman r = 0.3) and TGFBR1 (Spearman r = 0.208). **Supplementary Figures S4A,B** showed the positive correlations between ADARB1 expression and several chemokines, including CCL14 (Spearman r = 0.153), CCL21 (Spearman r = 0.136), CXCL12 (Spearman r = 0.224) and CXCL14 (Spearman r = 0.18), while the ADARB1 expression levels were negative correlation with CCL13 (Spearman r = −0.13), CXCL8 (Spearman r = −0.143), CXCL13 (Spearman r = −0.129) and XCL1 (Spearman r = −0.194). In addition, we found the low level of ADARB1, KDR and CXCL2 in tumor tissues compared with the adjacent tissues by immunohistochemistry, which indicated that the expression of ADARB1 was positively correlated with the expression of KDR and CXCL12 (**Supplementary Figure S1**). We hypothesized that ADARB1 might have a significant effect on immune regulation in OC since it is obviously associated with various types of tumor-infiltrating lymphocytes, immunomodulators, and chemokines in OC.

DISCUSSION

This is the first study to investigate the expression and function of ADARB1 in OC and its association with clinical features from the perspective of bioinformatics. The results confirmed that the expression of ADARB1 was significantly reduced in OC tissues and cell lines. ADARB1 might display an anti-cancer role through inhibiting AKT phosphorylation. Moreover, patients with low expression of ADARB1 had shorter OS and PFS from DRUGSURV database and Kaplan-Meier Plotter database. However, we found the inconsistent prognostic data by using the Human Protein Altas database. These conflicting data may be due to different patients' characteristics, such as age and race, in different databases. Thus, more OC patients from different regions will be needed for further confirmation the roles of ADARB1 in the future.

To date, ADAR has been widely found in all multicellular animal species. The mammalian genome has five genes encoding ADAR protein. Among these, ADAR1 and ADARB1 are active deaminases, ADAR3 has no known editing activity, and two other closely related testicle-specific ADAD1 and ADAD2 proteins lack

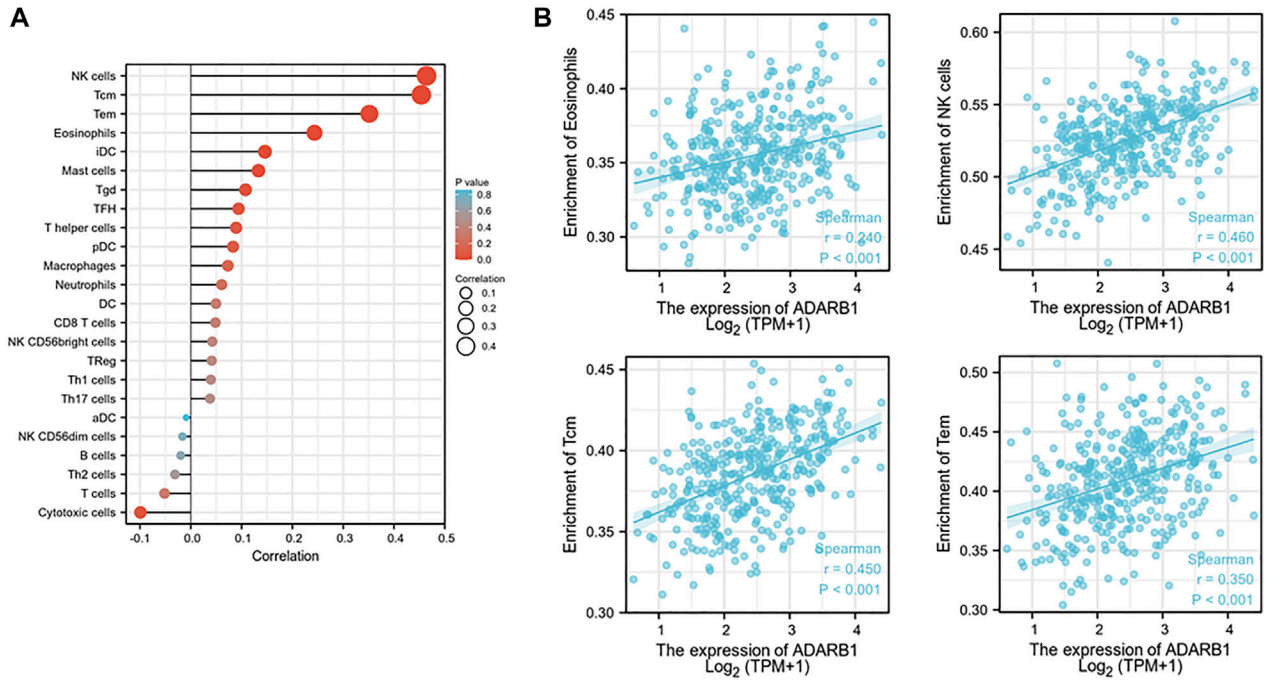


FIGURE 8 | Correlation of ADARB1 expression with tumor-infiltrating lymphocytes in OC. **(A)** The correlation between ADARB1 expression and tumor-infiltrating lymphocytes. **(B)** The top four tumor-infiltrating lymphocytes showing the most significant correlations with ADARB1 expression. **Supplementary Figure S1**. Immunohistochemistry was used to examine the ADARB1,p-AKT, CXCL12 and KDR expression in ovarian cancer tissue and paired adjacent tissue (magnification 200×).

the key catalytic residues (Goncharov et al., 2019). Recently, studies have demonstrated the key roles of ADARB1 in cancer development (Galeano et al., 2012; Goncharov et al., 2019). For example, hepatoma carcinoma patients with down-regulated ADARB1 expression have a poor prognosis (Chan et al., 2014). RNA editing of SLC22A3 regulated by ADARB1 promotes tumor invasion and metastasis in early familial esophageal carcinoma (Fu et al., 2017). In addition, ADARB1-OE promotes cell growth, motility, and invasion of malignant pleural mesothelioma cells independent of its RNA-editing activity (Sakata et al., 2020). However, the functional roles ADARB1 in OC pathology has not been investigated. In the present study, we demonstrated that ADARB1 expression was significantly down-regulated in OC tissues and cells, and might act as a tumor suppressor gene.

Alterations in AKT kinase activity are associated with a variety of pathologies. Increased AKT1 activity is associated with a variety of cancers, including OC (Gu et al., 2020; Samartzis et al., 2020; Zhai et al., 2020). Previous study has demonstrated that changes in AKT-dependent phosphorylation of ADARB1 have the potential to affect cellular programming (Piazzi et al., 2020). Similarly, the results in our study indicated that ADARB1-OE significantly inhibited tumor proliferation and metastasis through down-regulating AKT phosphorylation.

In addition, GO and KEGG pathway analysis indicated that genes coexpressed with ADARB1 were mainly enriched in ribonucleotide metabolic process, which suggested that ADARB1 might also be involved in OC progression by regulating these pathways. The metabolism was thought to sustained cancer cell malignant behaviors, including proliferation and metastasis, and anti-tumor immune function (Vaupel et al., 2019; Bergers and Fendt, 2021). To date, increasingly anti-tumor metabolism targets have been discovered, and have been used clinically for certain cancers (Luengo et al., 2017; Vander Heiden and DeBerardinis, 2017). Therefore, ADARB1 may be a useful molecular target for OC therapeutics.

In recent years, immunotherapy has become the focus of tumor therapy (Huang et al., 2020; Yan et al., 2020). T cell-rich tumor patients have longer PFS and OS (Zhang et al., 2003), while immune avoidance-associated mechanisms are associated with low survival (Kandalaft et al., 2009). All this evidence suggested that OC patients may benefit from immunotherapy (Ghisoni et al., 2019). In this study, TISIDB database was used to analyze the correlation between ADARB1 and immune system, and the results showed that ADARB1 had most significant correlation with several tumor-infiltrating lymphocytes (NK cells, Tcm, Tem, and eosinophils), immunomodulators (KDR, CSF1R, CD160 and TGFBR1), and chemokines (CXCL12, CCL14, CXCL14 and CCL21). ADARB1 knockout induces an antiviral immune response, leading to the suppression of infection (Yanai et al., 2020). Although these results suggested that ADARB1 may be

related to tumor immunity, clinical research is needed for further study.

In summary, our study illustrated that low expression of ADARB1 is associated with poor prognosis in OC, suggesting that ADARB1 may be an anti-oncogene and could serve as a promising biomarker in the tumorigenesis of OC. Moreover, ADARB1 was found to be involved in AKT-mediated malignant biological properties of OC cells. In addition, ADARB1 expression is related to tumor-infiltrating lymphocytes and immunomodulators. Therefore, our findings suggested that ADARB1 likely plays a pivotal role in immune cell infiltration and could serve as a promising biomarker for prognosis in OC patients.

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

Conception and design: YY and ZX. Writing, review, and/or revision of the manuscript: WZ, XC, XR, and XW. Administrative, technical, or material support: BP, YC, and QL. All authors approved final version of manuscript.

FUNDING

This study is supported by grants from the Natural Science Foundation of Hunan Province (2020JJ5934), and the Postdoctoral Science Foundation of Central South University (248485)

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2021.792911/full#supplementary-material

REFERENCES

Amelio, I., Gostev, M., Knight, R. A., Willis, A. E., Melino, G., and Antonov, A. V. (2014). DRUGSURV: a Resource for Repositioning of Approved and Experimental Drugs in Oncology Based on Patient Survival Information. *Cell Death Dis* 5, e1051. doi:10.1038/cddis.2014.9

Bergers, G., and Fendt, S.-M. (2021). The Metabolism of Cancer Cells during Metastasis. *Nat. Rev. Cancer* 21 (3), 162–180. doi:10.1038/s41568-020-00320-2

Brunty, S., Mitchell, B., Bou-Zgheib, N., and Santanam, N. (2020). Endometriosis and Ovarian Cancer Risk, an Epigenetic Connection. *Ann. Transl Med.* 8 (24), 1715. doi:10.21037/atm-20-2449

Cesarini, V., Silvestris, D. A., Tassinari, V., Tomaselli, S., Alon, S., Eisenberg, E., et al. (2018). ADAR2/miR-589-3p axis Controls Glioblastoma Cell Migration/invasion. *Nucleic Acids Res.* 46 (4), 2045–2059. doi:10.1093/nar/gkx1257

Chan, T. H. M., Lin, C. H., Qi, L., Fei, J., Li, Y., Yong, K. J., et al. (2014). A Disrupted RNA Editing Balance Mediated by ADARs (Adenosine DeAminases that Act on RNA) in Human Hepatocellular Carcinoma. *Gut* 63 (5), 832–843. doi:10.1136/gutjnl-2012-304037

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

Chen, Y.-B., Liao, X.-Y., Zhang, J.-B., Wang, F., Qin, H.-D., Zhang, L., et al. (2017). ADAR2 Functions as a Tumor Suppressor via Editing IGFBP7 in Esophageal Squamous Cell Carcinoma. *Int. J. Oncol.* 50 (2), 622–630. doi:10.3892/ijo.2016.3823

Dai, S., Yan, Y., Xu, Z., Zeng, S., Qian, L., Huo, L., et al. (2017). SCD1 Confers Temozolomide Resistance to Human Glioma Cells via the Akt/GSK3β/β-Catenin Signaling Axis. *Front. Pharmacol.* 8, 960. doi:10.3389/fphar.2017.00960

Demircan, N. C., Boussios, S., Tasci, T., and Öztürk, M. A. (2020). Current and Future Immunotherapy Approaches in Ovarian Cancer. *Ann. Transl Med.* 8 (24), 1714. doi:10.21037/atm-20-4499

Fu, L., Qin, Y.-R., Ming, X.-Y., Zuo, X.-B., Diao, Y.-W., Zhang, L.-Y., et al. (2017). RNA Editing of SLC22A3 Drives Early Tumor Invasion and Metastasis in Familial Esophageal Cancer. *Proc. Natl. Acad. Sci. USA* 114 (23), E4631–E4640. doi:10.1073/pnas.1703178114

Galeano, F., Rossetti, C., Tomaselli, S., Cifaldi, L., Lezzerini, M., Pezzullo, M., et al. (2013). ADAR2-editing Activity Inhibits Glioblastoma Growth through the Modulation of the CDC14B/Skp2/p21/p27 axis. *Oncogene* 32 (8), 998–1009. doi:10.1038/onc.2012.125

Galeano, F., Tomaselli, S., Locatelli, F., and Gallo, A. (2012). A-to-I RNA Editing: the "ADAR" Side of Human Cancer. *Semin. Cel Developmental Biol.* 23 (3), 244–250. doi:10.1016/j.semcdb.2011.09.003

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

Ghisoni, E., Imbimbo, M., Zimmermann, S., and Valabrega, G. (2019). Ovarian Cancer Immunotherapy: Turning up the Heat. *Ijms* 20 (12), 2927. doi:10.3390/ijms20122927

Goncharov, A. O., Kliuchnikova, A. A., Nasaev, S. S., and Moshkovskii, S. A. (2019). RNA Editing by ADAR Adenosine Deaminases: From Molecular Plasticity of Neural Proteins to the Mechanisms of Human Cancer. *Biochem. Mosc.* 84 (8), 896–904. doi:10.1134/S0006297919080054

Gu, H., Tu, H., Liu, L., Liu, T., Liu, Z., Zhang, W., et al. (2020). RSPO3 Is a Marker Candidate for Predicting Tumor Aggressiveness in Ovarian Cancer. *Ann. Transl Med.* 8 (21), 1351. doi:10.21037/atm-20-3731

Heraud-Farlow, J. E., and Walkley, C. R. (2020). What Do Editors Do? Understanding the Physiological Functions of A-To-I RNA Editing by Adenosine Deaminase Acting on RNAs. *Open Biol.* 10 (7), 200085. doi:10.1098/rsob.200085

Hidayat, Y. M., Munizar, M., Harsono, A. B., Winarno, G. N. A., Hasanuddin, H., and Salima, S. (2020). Chemokine Ligand 5 to Predict Optimal Cytoreduction in Ovarian Cancer. *Ijgm* Vol. 13, 1201–1206. doi:10.2147/IJGM.S280858

Huang, H., Bai, Y., Lu, X., Xu, Y., Zhao, H., and Sang, X. (2020). N6-methyladenosine Associated Prognostic Model in Hepatocellular Carcinoma. *Ann. Transl Med.* 8 (10), 633. doi:10.21037/atm-20-2894

Kandalaft, L. E., Facciabene, A., Buckanovich, R. J., and Coukos, G. (2009). Endothelin B Receptor, a New Target in Cancer Immune Therapy. *Clin. Cancer Res.* 15 (14), 4521–4528. doi:10.1158/1078-0432.CCR-08-0543

Lánczky, A., Nagy, Á., Bottai, G., Munkácsy, G., Szabó, A., Santarpia, L., et al. (2016). miRpower: a Web-Tool to Validate Survival-Associated miRNAs Utilizing Expression Data from 2178 Breast Cancer Patients. *Breast Cancer Res. Treat.* 160 (3), 439–446. doi:10.1007/s10549-016-4013-7

Luengo, A., Gui, D. Y., and Vander Heiden, M. G. (2017). Targeting Metabolism for Cancer Therapy. *Cel Chem. Biol.* 24 (9), 1161–1180. doi:10.1016/j.chembiol.2017.08.028

Ma, P., Yue, L., Zhang, S., Hao, D., Wu, Z., Xu, L., et al. (2020). Target RNA Modification for Epigenetic Drug Repositioning in Neuroblastoma: Computational Omics Proximity between Repurposing Drug and Disease. *Aging* 12 (19), 19022–19044. doi:10.18632/aging.103671

Moschetta, M., Boussios, S., Rassy, E., Samartzis, E. P., Funingana, G., and Uccello, M. (2020). Neoadjuvant Treatment for Newly Diagnosed Advanced Ovarian

Cancer: where Do We Stand and where Are We Going? *Ann. Transl Med.* 8 (24), 1710. doi:10.21037/atm-20-1683

Piazzi, M., Bavelloni, A., Gallo, A., and Blalock, W. L. (2020). AKT-dependent Phosphorylation of ADAR1p110 and ADAR2 Represents a New and Important Link between Cell Signaling and RNA Editing. *DNA Cel Biol.* 39 (3), 343–348. doi:10.1089/dna.2020.5351

Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., et al. (2019). Pathway Enrichment Analysis and Visualization of Omics Data Using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat. Protoc.* 14 (2), 482–517. doi:10.1038/s41596-018-0103-9

Sakata, K.-I., Maeda, K., Sakurai, N., Liang, S., Nakazawa, S., Yanagihara, K., et al. (2020). ADAR2 Regulates Malignant Behaviour of Mesothelioma Cells Independent of RNA-Editing Activity. *Anticancer Res.* 40 (3), 1307–1314. doi:10.21873/anticanres.14072

Samartzis, E. P., Labidi-Galy, S. I., Moschetta, M., Uccello, M., Kalaitzopoulos, D. R., Perez-Fidalgo, J. A., et al. (2020). Endometriosis-associated Ovarian Carcinomas: Insights into Pathogenesis, Diagnostics, and Therapeutic Targets-A Narrative Review. *Ann. Transl Med.* 8 (24), 1712. doi:10.21037/atm-20-3022a

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING Database in 2017: Quality-Controlled Protein-Protein Association Networks, Made Broadly Accessible. *Nucleic Acids Res.* 45 (D1), D362–D368. doi:10.1093/nar/gkw937

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

Vander Heiden, M. G., and DeBerardinis, R. J. (2017). Understanding the Intersections between Metabolism and Cancer Biology. *Cell* 168 (4), 657–669. doi:10.1016/j.cell.2016.12.039

Vaupel, P., Schmidberger, H., and Mayer, A. (2019). The Warburg Effect: Essential Part of Metabolic Reprogramming and central Contributor to Cancer Progression. *Int. J. Radiat. Biol.* 95 (7), 912–919. doi:10.1080/09553002.2019.1589653

Wang, J., Vasaikar, S., Shi, Z., Greer, M., and Zhang, B. (2017). WebGestalt 2017: a More Comprehensive, Powerful, Flexible and Interactive Gene Set Enrichment Analysis Toolkit. *Nucleic Acids Res.* 45 (W1), W130–W137. doi:10.1093/nar/gkx356

Wang, X., Ren, X., Liu, W., Chen, X., Wei, J., Gong, Z., et al. (2020). Role of Downregulated ADARB1 in Lung Squamous Cell Carcinoma. *Mol. Med. Rep.* 21 (3), 1517–1526. doi:10.3892/mmr.2020.10958

Wang, X., Xu, Z., Ren, X., Chen, X., Wei, J., Lin, W., et al. (2019). Function of Low ADARB1 Expression in Lung Adenocarcinoma. *PLoS One* 14 (9), e0222298. doi:10.1371/journal.pone.0222298

Wu, H., Liu, J., Zhang, Y., Li, Q., Wang, Q., and Gu, Z. (2021). miR-22 Suppresses Cell Viability and EMT of Ovarian Cancer Cells via NLRP3 and Inhibits PI3K/AKT Signaling Pathway. *Clin. Transl Oncol.* 23 (2), 257–264. doi:10.1007/s12094-020-02413-8

Yan, Y., Xu, Z., Hu, X., Qian, L., Li, Z., Zhou, Y., et al. (2018). SNCA Is a Functionally Low-Expressed Gene in Lung Adenocarcinoma. *Genes* 9 (1), 16. doi:10.3390/genes9010016

Yan, Y., Zeng, S., Gong, Z., and Xu, Z. (2020). Clinical Implication of Cellular Vaccine in Glioma: Current Advances and Future Prospects. *J. Exp. Clin. Cancer Res.* 39 (1), 257. doi:10.1186/s13046-020-01778-6

Yanai, M., Kojima, S., Sakai, M., Komorizono, R., Tomonaga, K., and Makino, A. (2020). ADAR2 Is Involved in Self and Nonself Recognition of Borna Disease Virus Genomic RNA in the Nucleus. *J. Virol.* 94 (6). doi:10.1128/JVI.01513-19

Zhai, Y., Lu, Q., Lou, T., Cao, G., Wang, S., and Zhang, Z. (2020). MUC16 Affects the Biological Functions of Ovarian Cancer Cells and Induces an Antitumor Immune Response by Activating Dendritic Cells. *Ann. Transl Med.* 8 (22), 1494. doi:10.21037/atm-20-6388

Zhang, L., Conejo-Garcia, J. R., Katsaros, D., Gimotty, P. A., Massobrio, M., Regnani, G., et al. (2003). Intratumoral T Cells, Recurrence, and Survival in Epithelial Ovarian Cancer. *N. Engl. J. Med.* 348 (3), 203–213. doi:10.1056/NEJMoa020177

Zhou, S., Yan, Y., Chen, X., Wang, X., Zeng, S., Qian, L., et al. (2019). Roles of Highly Expressed PAICS in Lung Adenocarcinoma. *Gene* 692, 1–8. doi:10.1016/j.gene.2018.12.064

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 Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 11 January 2022
doi: 10.3389/fbioe.2021.798584

Identification and Validation of HOTAIRM1 as a Novel Biomarker for Oral Squamous Cell Carcinoma

Yixiu Yu[1], Jiamei Niu[2], Xingwei Zhang[1], Xue Wang[1], Hongquan Song[1], Yingqun Liu[3], Xiaohui Jiao[1] and Fuyang Chen[4]**

[1]Department of Oral Maxillofacial Surgery, The First Affiliated Hospital of Harbin Medical University, Harbin, China, [2]Department of Abdominal Ultrasonography, The First Affiliated Hospital of Harbin Medical University, Harbin, China, [3]Pediatric Dentistry Department, The First Affiliated Hospital of Harbin Medical University, Harbin, China, [4]Department of Stomatology, The Second Affiliated Hospital of Harbin Medical University, Harbin, China

ORAL squamous cell carcinoma (OSCC) is a malignant tumor with the highest incidence among tumors involving the oral cavity maxillofacial region, and is notorious for its high recurrence and metastasis potential. Long non-coding RNAs (lncRNAs), which regulate the genesis and evolution of cancers, are potential prognostic biomarkers. This study identified HOTAIRM1 as a novel significantly upregulated lncRNA in OSCC, which is strongly associated with unfavorable prognosis of OSCC. Systematic bioinformatics analyses demonstrated that HOTAIRM1 was closely related to tumor stage, overall survival, genome instability, the tumor cell stemness, the tumor microenvironment, and immunocyte infiltration. Using biological function prediction methods, including Weighted gene co-expression network analysis (WGCNA), Gene set enrichment analysis (GSEA), and Gene set variation analysis (GSVA), HOTAIRM1 plays a pivotal role in OSCC cell proliferation, and is mainly involved in the regulation of the cell cycle. *In vitro*, cell loss-functional experiments confirmed that HOTAIRM1 knockdown significantly inhibited the proliferation of OSCC cells, and arrested the cell cycle in G1 phase. At the molecular level, PCNA and CyclinD1 were obviously reduced after HOTAIRM1 knockdown. The expression of p53 and p21 was upregulated while CDK4 and CDK6 expression was decreased by HOTAIRM1 knockdown. *In vivo*, knocking down HOTAIRM1 significantly inhibited tumor growth, including the tumor size, weight, volume, angiogenesis, and hardness, monitored by ultrasonic imaging and magnetic resonance imaging In summary, our study reports that HOTAIRM1 is closely associated with tumorigenesis of OSCC and promotes cell proliferation by regulating cell cycle. HOTAIRM1 could be a potential prognostic biomarker and a therapeutic target for OSCC.

Keywords: oral squamous cell carcinoma, lncRNA HOTAIRM1, biomarker, cell proliferation, cell cycle

OPEN ACCESS

Edited by:
Francesco Pappalardo, University of Catania, Italy

Reviewed by:
Aarón Vázquez Jiménez, Instituto Nacional de Medicina Genómica (INMEGEN), Mexico
Guijun Zhao, Shanghai Jiao Tong University, China

****Correspondence:***
Xiaohui Jiao
jxhhyd@hrbmu.edu.cn
Fuyang Chen
Cfy666@hrbmu.edu.cn

Specialty section:
This article was submitted to Preclinical Cell and Gene Therapy, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *20 October 2021*
Accepted: *24 December 2021*
Published: *11 January 2022*

Citation:
Yu Y, Niu J, Zhang X, Wang X, Song H, Liu Y, Jiao X and Chen F (2022) Identification and Validation of HOTAIRM1 as a Novel Biomarker for Oral Squamous Cell Carcinoma. Front. Bioeng. Biotechnol. 9:798584. doi: 10.3389/fbioe.2021.798584

Abbreviations: OSCC, oral squamous cell carcinoma; HNSC, head and neck squamous carcinoma; lncRNAs, long non-coding RNAs; CNV, copy number variation; TMB, tumor mutation burden; TME, tumor microenvironment; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; WGCNA, Weighted gene co-expression network analysis; GSEA, Gene set enrichment analysis; GSVA, Gene set variation analysis; USI, Ultrasonic imaging; MRI, magnetic resonance imaging.

INTRODUCTION

Oral squamous cell carcinoma (OSCC) is recognized for its high recurrence and metastasis rate, and is the most frequent malignancy in oral tumors, and in particular, presents constantly increasing morbidity (Ferlay et al., 2015; Economopoulou et al., 2017). Although surgeons are constantly adjusting surgical approaches and improving multidisciplinary therapy, the 5-years survival rate of 50% still represents a very poor prognosis, which is related to the specific tissue involved, clinical features, histopathological grading, or other factors (Bray et al., 2018). From another standpoint, the cure rate increases to above 90% through early interventional treatment, in individuals who are diagnosed with early-stage cancer (Lingen et al., 2008; Qiu et al., 2021a). Therefore, it is an urgent task to explore valuable biomarkers, which can provide predictive evidence for the identification of OSCC high-risk groups, early diagnosis, selection of treatment methods, and monitoring of prognosis monitoring.

With the achievements of genome engineering, scientists have no longer been limited to the development of a coding genome, alternatively, they now focus on the non-coding genome, which accounts for 98% of the whole genome (Rong et al., 2021). Long non-coding RNAs (lncRNAs) have emerged as non-coding RNAs, which are defined as a type of RNAs with transcription length not exceed 200 bp, and structural diversity with no coding ability (Salmena et al., 2011). LncRNAs that regulate the expression pattern of mammalian coding genes at the transcriptional or post-transcriptional levels directly or indirectly have been demonstrated on the basis of numerous studies (Winkle et al., 2021). LncRNAs participate in complicated biological processes and are vital regulators in regulating or deregulating expression of genes that determine the cellular fate. LncRNAs have been identified to be involved in cancer formation and development in thyroid cancer, head and neck cancer, pancreatic cancer, and glioma (Ahadi 2020; Song and Kim 2021). Studies have confirmed that abnormal expression of some functional lncRNAs could affect the progression of cancer. For example, LINC00941 is abnormally upregulated in pancreatic cancer, promoting glycolysis by regulating the Hippo pathway, thereby promoting the malignant biological behavior of pancreatic cancer (Xu et al., 2021a). H19 is a significantly down-regulated lncRNA in prostate cancer confirmed to inhibit the invasion of tumor cells by targeting TGFBI via regulating miR-675, which could be a biomarker to benefit diagnosis and therapy of advanced prostate cancer (Zhu et al., 2014). Thus, it is of the utmost importance to explore the expression, distribution, and molecular biological function of OSCC-related lncRNAs as valuable markers in diagnosis, treatment, and prognosis. It has been reported that the expression of HOTAIRM1 is evidently downregulated in lung adenocarcinoma tissues, and the growth of H1650 and PC-9 cell lines were accelerated by promoting cell-cycle progression when HOTAIRM1 was silenced (Chen et al., 2020). In glioblastoma (BGM), researcher confirmed the upregulation of HOTAIRM1, and further demonstrated that HOTAIRM1 acted as a facilitator of malignant-behavior in BGM, which was shown to accelerate the occurrence of migration and invasion (Xie et al., 2020). It's a sign that HOTAIRM1 may be a cancer-associated lncRNA and influences the malignant progression of some cancers. Nevertheless, the expression pattern and molecular function of HOTAIRM1 in OSCC have not been elucidated to date.

In our study, HOTAIRM1 was identified as an upregulated lncRNA and was found to be closely associated with OSCC overall survival (OS), tumor stage, genomic instability, the tumor cell stemness, the tumor microenvironment (TME), and immune inflammation. In addition, our *in vitro* studies revealed that HOTAIRM1 could accelerate the cell proliferation by regulating the cell cycle of OSCC cells. *In vivo*, the tumor growth was significantly inhibited by HOTAIRM1 knockdown, including tumor size, weight, volume, angiogenesis, and tumor hardness, as assessed by ultrasonic imaging (USI) and magnetic resonance imaging (MRI). In conclusion, our findings have demonstrated that HOTAIRM1 is a risk factor for OSCC, and confirmed its function, and we suggest that HOTAIRM1 could be a novel biomarker for the early diagnosis and therapy of OSCC.

MATERIALS AND METHODS

Data Extraction

The data used for our analysis was obtained from TCGA-HNSC dataset, which can be downloaded through the Genomic Data Commons (GDC) (https://portal.gdc.cancer.gov/), and including the following information: Exp (expression data gathered in HTSeq-Counts, HTSeq-FPKM, copy number variation (CNV) and somatic mutations data. From the UCSC Xena database (https://xenabrowser.net/), we obtained the clinical phenotype and survival data. Patient samples with lesion sites in the oral cavity (lip, gum, palate, jaw, floor of mouth, tongue, maxilla, mandible, and buccal mucosa) were selected for our study and those without clinical data information were excluded.

Data Processing and Differential Expression Analysis

Firstly, the genes whose 80% of samples' count value was at least one were retained for the following analyses. Meanwhile, the human gene transfer format (GTF) annotation file was acquired from the GENCODE project (http://www.gencodegenes.org, release 35) to convert the Ensembl gene ID into the gene symbol and extracted the lncRNAs and protein-coding genes profiles. Then, 4324 lncRNAs and 15952 mRNAs and their corresponding count and FPKM expression profiles were obtained.

Secondly, differential expression analysis of lncRNAs between OSCC and control samples was performed using the R "DESeq2" package. It was considered to be significant when DEGs meeting the following conditions: I. false discovery rate (FDR) < 0.01; II. | log2Fold Change (FC)|> 1 (Zhao and Ruan 2020; Xu et al., 2021b; Markert, et al., 2021). Here, we focused on the lncRNA HOTAIRM1, who is associated with multiple cancers, for subsequent analysis. The samples were divided into high- and low-HOTAIRM1 groups after the median HOTAIRM1

expression value (1.5948) was set as the cut-off point (Ozawa et al., 2017).

Survival Analysis

The R package "survival" and "survminer" were used to perform univariate and multivariate Cox regression analyses and Kaplan-Meier survival analysis. To retian the samples with complete clinical data, we screened 288 samples from 324 OSCC patients for overall survival (OS) analysis. The low and high groups were cut-off through the median of HOTAIRM1 expression value. p-value < 0.05 was considered statistically significant based on log-rank test.

Genomic Variation Analysis

Firstly, the CNV data relevant to HOTAIRM1 expression were analyzed using "gistic2.0" software, setting 0.1 as the q-value, and was specific to chromosomes alterationsbetween HOTAIRM1-high and HOTAIRM1-low groups; the results are presented using the R package "maftools". Secondly, to compare the tumor mutation burden (TMB) between HOTAIRM1-high and HOTAIRM1-low groups, the somatic raw variant counts identified by TCGA were calculated, while whole-exome 38 Mb size was regarded as the estimate. Finally, to analyze the correlation between the HOTAIRM1 and genetic mutation, the Pearson correlation coefficient (PCC) was calculated on the HOTAIRM1 expression and genetic mutations vector. It is considered to be significantly correlated when p-value of PCC< 0.05.

Correlation Analyses Between Tumor Microenviroment and HOTAIRM1

For analyzing the association between tumor microenvironment (TME) and HOTAIRM1 expression, the R packages "ESTIMATE" was employed to get the stromal score, immune score and ESTIMATE score (Yoshihara et al., 2013). And, to assess the state of infiltration of immune cells, the CIBERSORT algorithm was used to analyze (Newman et al., 2015). The PCC was conducted to analyze the correlation between HOTAIRM1 and tumor cell stemness, TME and immune-infiltrating cells. A significant correlation was considered to be when p-value of PCC< 0.05.

Weighted Gene Co-Expression Network Analysis

According to absolute median difference of expression values, the first 5000 mRNAs in all the OSCC samples were selected to conduct WGCNA using the R package "WGCNA" (Langfelder and Horvath 2008) to aquire the gene-sets associated with HOTAIRM1. The power parameters were selected through the "pickSoftThreshold." We use topology overlap similarity (TOM) matrix to represent the similarity of two genes in network structure and modules containing at least 30 genes were retained. We next calculated the correlation coefficient between the modules and HOTAIRM1 expression to determine the modules most related to HOTAIRM1. To identify the key genes in this module having gene significances (GS) ≥0.3, the module membership (MM) in the top 10% after calculating were obtained.

Next, we further analyzed the biological functions involved in these key genes. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were completed by the R package "ClusterProfiler." After the Benjaminiand Hochberg (BH) correction, p-values< 0.05 were considered to indicate significantly enriched function and the results were displayed by R package "ggplot2" and "enrichplot", respectively. In addition, we obtained the protein-protein interaction (PPI) information of key genes from the STRING database (https://string-db.org/) and visualized the PPI network using "cytoscape" software. In this study, the interaction pairs with combined score> 0.4 were retained.

Analysis of Biological Function Associated With HOTAIRM1 Up-Regulation

To further identify the biological changes caused by abnormal HOTAIRM1 expression, we choose the top 20% samples of the high-HOTAIRM1 group (highest 20% samples) and the bottom 20% samples of the low-HOTAIRM1 group (lowest 20% samples) to analyze. The rank of each mRNA was determined based on their Fold Change value (DESeq2) in the above samples. Meanwhile, the gene sets of Hallmark, KEGG pathways, GO-BP terms and Reactome pathways were all collected from Molecular Signatures Database (MSigDB). Then, applying the R software package "clusterProfiler" to conduct Gene Set Enrichment Analysis (GSEA), and the gene sets were significant with adjust p-values (BH) < 0.05. Finally, to confirm that HOTAIRM1 was indeed involved in the biological processes of the cell cycle and cell proliferation in OSCC, the Gene Set Variation Analysis (GSVA) was conducted in low- and high- HOTAIRM1 expression groups. This process was completed by R "GSVA" package.

Cell Culture

Human OSCC cell lines (Cal27 and SCC9) were obtained from the Harbin Medical University (Harbin, China), and human oral epithelial cells (HOEC) were originally supplied by the American Type Culture Collection (ATCC, United States). Cal27 and HOEC cells were maintained in a 37°C incubator with humidified atmosphere containing 5% CO_2 and cultured with Dulbecco's Modified Eagle's Medium (DMEM) containing 10% Fetal Bovine Serum (FBS), while Roswell Park Memorial Institute-1640 (RPMI 1640) containing 10% FBS was used for SCC9 cells.

Cell Transfection

Specific sh-RNAs (sh-HOTAIRM1) and the control sh-RNAs (sh-NC) were obtained from General Biosystems (Anhui, China). The sh-RNAs were transfected using the Lipofectamine 3000 reagent (following the manufacturer's instructions) to knockdown HOTAIRM1 expression in Cal27 and SCC9 cells.

Quantitative Real-Time Polymerase Chain Reaction

Total RNA of OSCC cells was isolated using a Total RNA Extraction Kit (Suzhou, China) and then transcribed into

cDNA using Prime Script RT Reagent Kit (Takara, Shiga, Japan). Subsequently, the SYBR Green Master Mix (TOYOBO, Japan) was applied to conduct quantitative real-time polymerase chain reaction (qRT-PCR) on the ABIPRISM 7900HT instrument (Applied Biosystems, United States). The above experiments were repeated 3 times. Primer sequences used were as follows:

HOTAIRM1
F: 5′-TTGACCTGGAGACTGGTAGC-3′
R: 5′-TTCAGTGCACAGGTTCAAGC-3′;
β-actin
F:5′ ATGAACTGGCGAGAGGTCTGT3′
R:5′ CCAGGAATGAGTAACACGGAGT3′.

Cell Counting Kit-8 Assay

The Cell counting kit-8 (CCK-8) (Dojindo, Rockville, MD, United States) was used to estimate the proliferation ability. The CCK-8 solution (10 μL) was added at specified culture time points (24, 48, 72, and 96 h), to each well of Cal27 and SCC9 cells seeded in a 96-well plate (density: 2.0×10^3 cells/well), and then cultured for an additional 2 h at 37°C in the dark. Finally, the optical absorption value, which represented the number of viable cells, was evaluated at 450 nm.

Colony Formation Assay

One thousand treated OSCC cells were plated in a culture dish whose diameter was 6-cm. After a 2-week culture, visible colonies were fixed with methanol after staining with 0.5% crystal violet for 15 min, and were counted using a microscope (IX51, Olympus, Tokyo, Japan).

Western Blotting

First, the total proteins were extracted from the cells using the radioimmunoprecipitation assay buffer (Pierce, Rockford, IL, United States). The proteins at equivalent amount were employed for the sodium dodecyl sulfate-polyacrylamide gel electro-phoresis (SDS-PAGE), and then transferred to the polyvinylidene fluoride membranes. After blocking membranes with 5% non-fat milk/TBST for 1 hour, incubation with the primary antibodies, including PCNA, Cyclin D1, CDK4, CDK6, *β*-actin (Proteintech, Wuhan, Hubei, China), p53 and p21 (Cell Signaling Technology, Danvers, MA, United States), were performed at 4°C overnight. Next, blots were exposed to the corresponding horseradish per-oxidase-conjugated secondary antibodies for an additional hour at room temperature. The secondary antibodies used in this study, including anti-mouse IgG and anti-rabbit IgG, were provided by Proteintech (Wuhan, Hubei, China). Finally, images of protein bands were obtained using the BioSpectrum 600 Imaging System (UVP, United States).

Cell Cycle Analysis

Transfected OSCC cells were washed twice by cold PBS and fixed using precooled 75% ethanol. Thereafter, the cell-cycle phase distribution of samples that had been digested using RNase (10 mg/ml) and stained using propidium iodide (PI, 1 mg/ml), were detected, and analyzed on FACSCalibur flow cytometer (BD Biosciences, United States). The above process was repeated 3 times.

Animal Experiments *in vivo*

The female BALB/c nude mice (aged, 4–5 weeks) were purchased from Charles River Japan (Beijing, China) and maintained in a sterile environment. This animal experiment was approved by the Committee on Animals of the First Affiliated Hospital of Harbin Medical University (Harbin, China). After animals were anesthetized, Cal27 cells (sh-NC or sh-HOTAIRM1) were injected subcutaneously into the bilateral flanks of BALB/c nude mice. For each mouse, the OSCC cells transfected with sh-NC were inoculated into the left flank, while the OSCC cells transfected with sh-HOTAIRM1 were inoculated into the right flank. Every week, the animal models were weighed and measured, and subjected to USI and MRI to observed neoplastic formation and growth, for a total treatment period of 4 weeks. The tumor volume was calculated using the following formula: $V = 0.5 \times \text{length} \times \text{Width}^2$. At the end of the experiment, the tumors in mice were excised after the animals were euthanized.

Ultrasonic Imaging

The Ultrasound System (Aplio 500, Canon, Japan) was applied to observe the tumors *in vivo* by USI. The specific imaging modes used were B-mode ultrasonography to evaluate neoplastic growth, Color doppler flow imaging (CDFI) and Color Power Angio (CPA) were used to evaluate angiopoiesis and Ultrasonic elastosonography (USE) to evaluate the tumor stiffness.

Magnetic Resonance Imaging

MRI soft tissue imaging was performed using an Achieva 3.0T TX MRI System (Philip, Netherlands) to observe the neoplastic growth and signal intensity with clinical oral floor imaging sequences, containing T1WI, T2WI and T2-FLAIR images.

Statistical Analysis

SPSS software (Version 22.0, Armonk, NY, United States) was employed to conduct the statistic analysis. All the experimental data, which were obtained after three repetitions of independent experiments, were presented as means ± standard deviation (SD). To compare the difference between two groups, the Student's *t*-test was performed, and one-way ANOVA was conducted to compare the difference between multiple groups. It was regarded as statistically significant when p-values < 0.05.

RESULTS

HOTAIRM1 was identified as abnormally upregulated in OSCC and was associated with tumor stage and worse prognosis

Figure 1 shows the complete workflow of our study to explore LncRNA HOTAIRM1 in OSCC. According to the conditions in "Data Extraction" screening, we finally acquired 342 OSCC patients and a control group comprised of 32 para-cancerous samples (**Supplementary Table S1**). Following calculation and processing of the sample data, finally, 4324 lncRNAs and 15,952

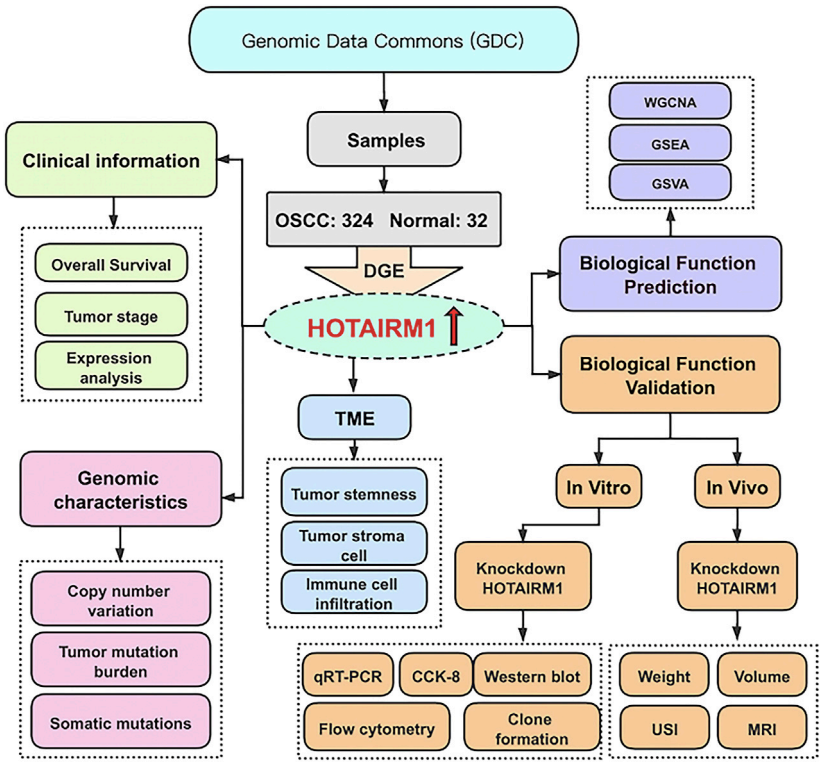


FIGURE 1 | The flowchart of this study.

mRNAs were obtained. Then the obtained lncRNAs were employed for the DEG analysis. Eventually, we identified 1069 abnormally expressed lncRNAs in the OSCC dataset (**Supplementary Table S2**), which are illustrated in the volcano plot in **Figure 2A**, including 717 upregulated lncRNAs and 352 downregulated lncRNAs. Among the DEGs, HOTAIRM1 was observed to be a markedly upregulated lncRNA, which has not been reported in previous studies on OSCC. Further, we found that HOTAIRM1 was indeed highly expressed in cancer samples but not in normal samples, even in the paired-samples comparison (**Figures 2B**,C). In addition, the expression of HOTAIRM1 in OSCC cell lines (SCC9 and Cal27) was also significantly higher than in human oral epithelial cells (HOEC), as assessed by qRT-PCR (**Figure 2D**). The power of prediction by HOTAIRM1 as a potential biomarker was evaluated by a receiver operating characteristic (ROC) curve and the area under the ROC (AUC). The result showed that the AUC value was 0.838, indicating the excellent power of prediction by HOTAIRM1 as a potential biomarker (**Supplementary Figure S1**). Next, we conducted an analysis to evaluate the relationship between OS and clinical data, including HOTAIRM1 expression, age, gender, tumor stage, smoking as variables. In the univariate Cox-regression analysis, HOTAIRM1 expression and tumor stage were revealed as the risk factors for OSCC patient prognosis. While, in the multivariate Cox-regression analysis, age, HOTAIRM1 expression and tumor stage were all significantly independent predictors of OS of OSCC patients (**Table 1**). Thus, it was reasonable to suspect that the elevation of HOTAIRM1 may be a risk factor for OSCC. In addition, we identified a correlation between HOTAIRM1 expression and tumor stage, indicating that more advanced OSCC stage was associated with the higher expression of HOTAIRM1 (**Figure 2E**). With regard to OS analysis, OSCC patients with high-HOTAIRM1 expression were characterized by a low survival rate and poorer bleak prognosis (**Figure 2F**). Combined with the above findings, we considered HOTAIRM1 as a risk factor likely to promote tumorigenesis of OSCC.

Upregulation of HOTAIRM1 Was Associated With Genomic Instability

To explore the inherent factors associated with abnormal HOTAIRM1-expression and OSCC carcinogenesis, we investigated different genomic alterations in terms of somatic variation and CNVs. First, we evaluated the copy number variation status between groups with high-HOTAIRM1 and low-HOTAIRM1 levels (**Figure 3A**), and determined that amplification of CHR3 and deficiency of CHR2 were significant variations in the high-HOTAIRM1 expression group, whereas in the low-HOTAIRM1 expression group, CHR11 amplification and the CHR13 deficiency were observed. These findings associated HOTAIRM1 expression to CNVs of specific genes. Next, we evaluated tumor mutation burden (TMB) status. In the high-HOTAIRM1 group, the TMB was evidently higher

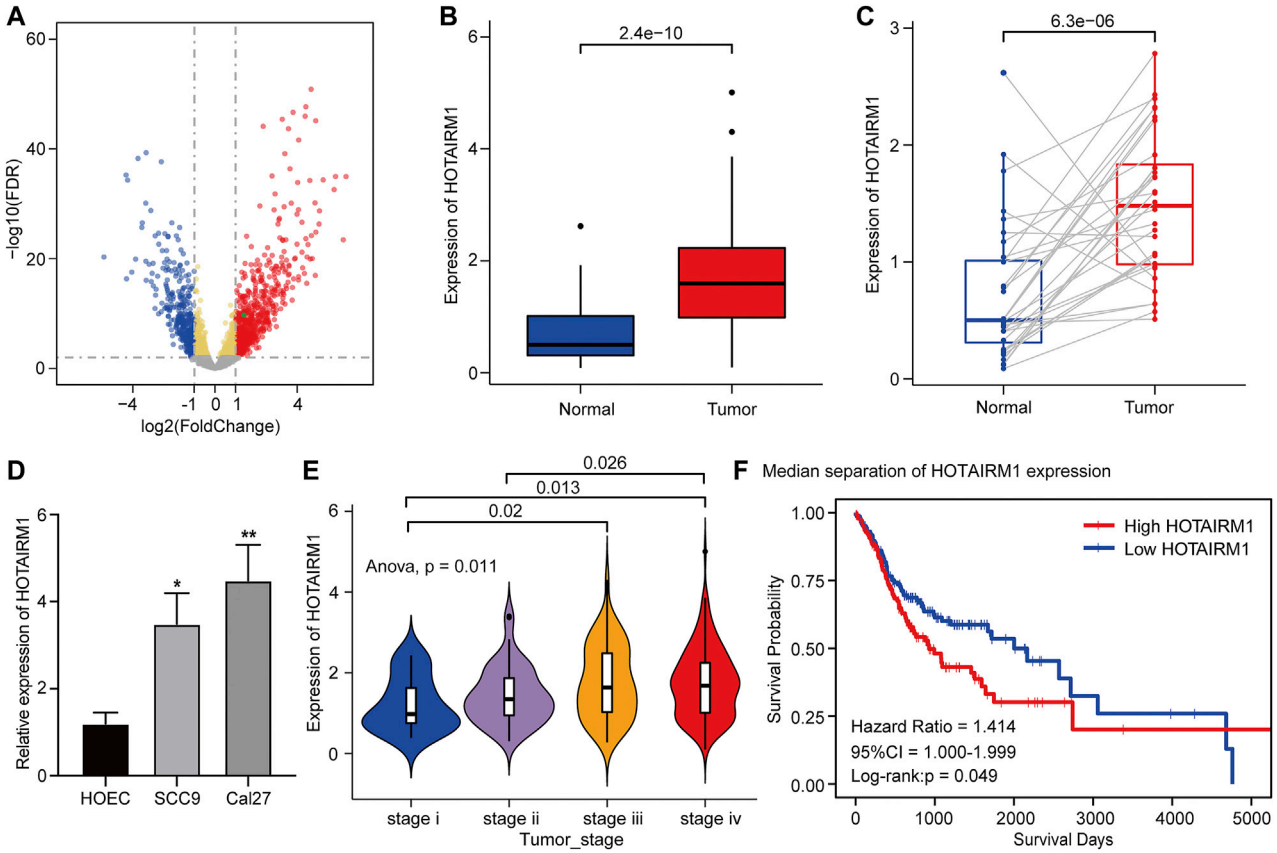


FIGURE 2 | Upregulation and clinical significance of HOTAIRM1 in OSCC. **(A)** Volcano plot for DEGs. The blue represented the significantly down-regulated genes; the red dots represented the significantly up-regulated genes; the green dot was the HOTAIRM1. **(B)** The comparison of HOTAIRM1 expression between all OSCC samples and control samples. **(C)** The comparison of HOTAIRM1 expression between OSCC tissues and corresponding adjacent normal tissues. **(D)** qRT-PCR analysis confirming HOTAIRM1 upregulation in OSCC cell lines. **(E)** Correlation analysis between HOTAIRM1 expression and OSCC tumor stage indicating the higher HOTAIRM1 expression was associated with a higher tumor stage. **(F)** Kaplan-Meier analysis suggested that high HOTAIRM1 expression was associated with unfavorable prognosis in OSCC. *$p < 0.05$, **$p < 0.01$.

TABLE 1 | Univariate and multivariate Cox regression analyses for overall survival in OSCC patients.

Variables	Status	Number	Univariate analysis			Multivariate analysis		
			DSBA	95% CI of HR	Pvalue	HR	95% CI of HR	P- value
HOTAIRM1	low/high	146/142	1.414	1.000–1.999	0.049[a]	1.458	1.019–2.087	0.039[a]
Age	≤60/>60	126/162	1.314	0.924–1.870	0.129	1.505	1.031–2.197	0.034[a]
Gender	female/male	90/198	0.943	0.655–1.359	0.754	0.940	0.639–1.382	0.753
Tumor stage	stage I	18	1 (ref)	—	—	1 (ref)	—	—
	Stage II	51	1.9773	0.584–6.693	0.273	1.738	0.510–5.921	0.377
	Stage III	60	2.6724	0.801–8.920	0.110	2.231	0.663–7.507	0.195
	Stage IV	159	4.2115	1.329–13.343	0.015[a]	3.637	1.137–11.658	0.030[a]
Smoking	no/yes	197/91	1.3787	0.956–1.989	0.086	1.230	0.839–1.803	0.288

[a]p < *0.05*.

compared to that in the low-HOTAIRM1 expression group (**Figure 3B**), which indicated that non-synonymous mutations may be involved in the dysregulation of HOTAIRM1 (*t*-test, $p < 0.05$). To better understand the underlying link between genetic mutations and the dysregulation of HOTAIRM1, we identified the first five most relevant mutated genes (TP53, FAT1, MUC16, DNAH5, HRAS), which provided evidence for the association between genome aberrations and upregulation of HOTAIRM1 in OSCC (**Figure 3C**).

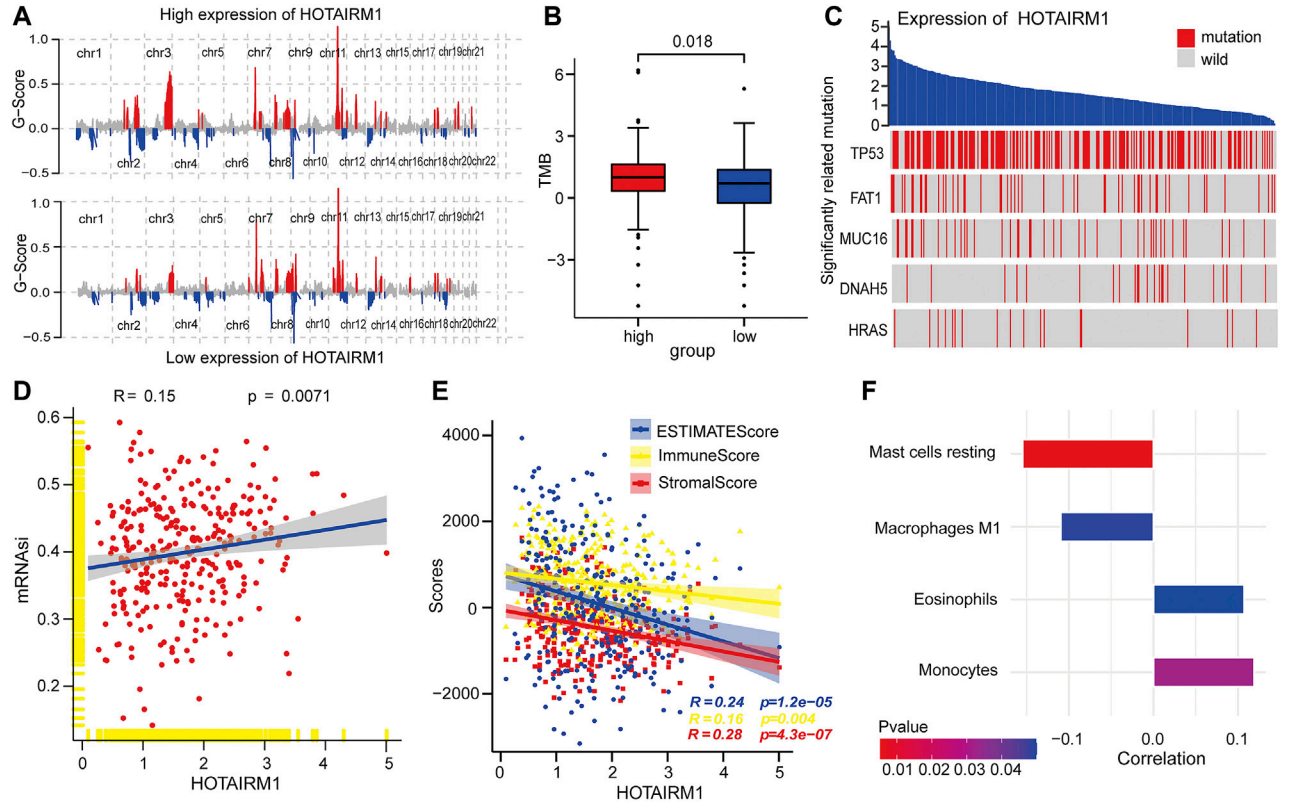


FIGURE 3 | Correlation analyses between HOTAIRM1 and tumorigenesis-relevant variables. **(A)** The copy number variation of each chromosome in the samples with high-HOTAIRM1 and low-HOTAIRM1 expression of OSCC is shown. The red peaks represent recurring focal amplification of chromosomes, while the blue areas represent deletions. **(B)** TMB was higher in OSCC patients with high-HOTAIRM1 expression than in patients with low-HOTAIRM1 expression. **(C)** Relevance between HOTAIRM1 expression and status of gene mutations and the five most correlated genes. **(D)** The relative association between the stemness index and HOTAIRM1 expression in OSCC. **(E)** The correlation analysis between HOTAIRM1 expression and cellular components related to the tumor microenvironment in OSCC, as assessed by the stromal score, immune score, and ESTIMATE score. **(F)** Several immune-infiltrating cells markedly associated with HOTAIRM1 expression.

HOTAIRM1 Expression Was Related With the Tumor Cell Stemness and Tumor Microenvironment

To explore the relationship between HOTAIRM1 expression and tumor cell dryness, pearson correlation analysis was performed between HOTAIRM1 expression and stemness indices from mRNA level (mRNAsi). The results showed that the higher HOTAIRM1 expression was, the higher mRNAsi was (**Figure 3D**), implying that HOTAIRM1 may contribute to the progression of OSCC (p = 0.0071). It is well known that tumors and the TME are an indivisible whole, therefore, we analysed the association between HOTAIRM1 and stromal cells, and immunocyte infiltration. The pearson correlation analysis showed that HOTAIRM1 expression level was negatively correlated with Stromal Score, ESTIMATE Score and Immune Score (**Figure 3E**). With regard to immune-infiltrating cells, there were four closely related immune cells associated with HOTAIRM1 expression. Among these, there was a negative correlation between HOTAIRM1 expression and resting mast cell and M1 macrophages levels, while a positive correlation was observed between HOTAIRM1 expression and eosinophils and monocytes infiltration (**Figure 3F**). The above results suggested that HOTAIRM1 might facilitate tumor cell proliferation in OSCC.

Biological Functions Related to HOTAIRM1 in ORAL Squamous Cell Carcinoma

WGCNA was used to screen genes interacting with HOTAIRM1 in OSCC. A total of 14 modules were obtained following data processing (**Figure 4A**), in which the brown module was identified as the most significantly related to HOTAIRM1 as it presented the highest Pearson's coefficient, which suggested signifying that HOTAIRM1 may could regulate or be regulated by these genes in brown module (**Figure 4B**). A total of 57 key genes whose expression was recognized as being influenced by HOTAIRM1 and were used to speculate the functions of HOTAIRM1 in OSCC (**Figure 4C**). Following functional enrichment analysis on these key genes, the 57 genes were determined to be active in biological processes regulating cell proliferation, including DNA replication, mitotic nuclear division, and changes in cell cycle phases (**Figure 4D**). In the KEGG pathways analysis, the DNA replication, Mismatch repair,

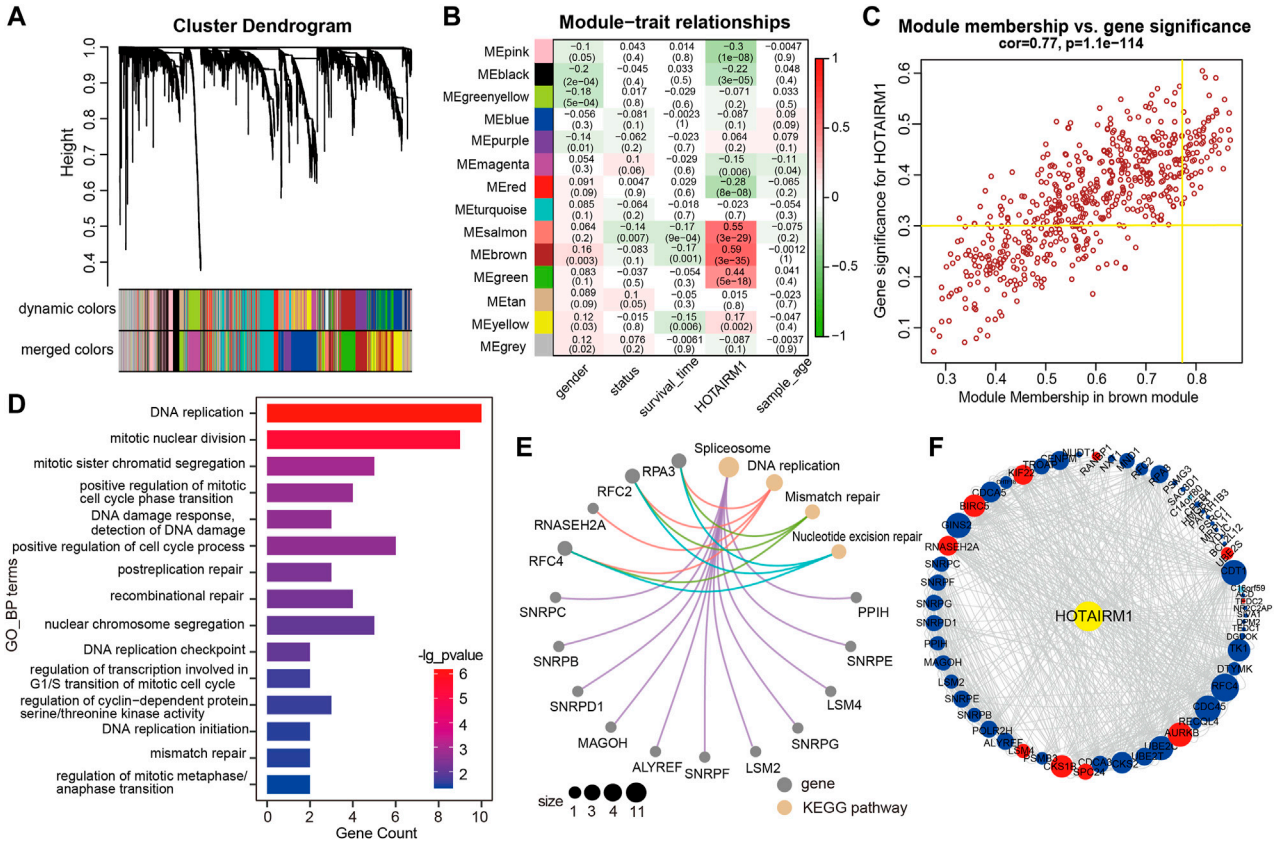


FIGURE 4 | WGCNA module analysis showing that HOTAIRM1 was mainly involved in the biological process of cell proliferation. **(A)** A total of 14 related modules were identified through WGCNA, represented by different colors, which were displayed in a hierarchical clustering tree diagram. **(B)** The heatmap showed the correlation between HOTAIRM1 expression and the modular eigengenes. The brown module showed the most positive significant correlation with HOTAIRM1. **(C)** Scatter plot of module eigengenes in the brown module. The genes meeting the module criteria, *p*-value ≥ 0.3 and in the top 10% of the brown module, are identified as key driving genes associated with high HOTAIRM1 expression. **(D)** Functional enrichment analysis by GO demonstrated that key driver genes related to HOTAIRM1 were enriched in cell proliferation processe. **(E)** The KEGG signaling pathway analysis showing driver genes related to HOTAIRM1 expression enriched in the pathways regulating cell proliferation. **(F)** The interrelationship between the driver genes illustrated in the PPI network.

and Nucleotide excision repair pathways were significantly enriched and were associated with cellular proliferation (**Figure 4E**). The above results indicated that HOTAIRM1 was involved in OSCC-cell proliferation and specifically in cell cycle regulation. A PPI-network-analysis was performed, which demonstrated the interactive relationship between the 57 critical genes, which supported the functional prediction regarding HOTAIRM1 activity in OSCC (**Figure 4F**). In addition, we retained the interaction pairs with combined score> 0.4 (**Supplementary Table S3**).

Gene Set Enrichment Analysis and Gene Set Variation Analysis Indicated HOTAIRM1 Promoted Cell Proliferation

GSEA was used to identify the potential biomolecular alterations caused by up-regulation of HOTAIRM1, focusing on the gene sets of Hallmark, GO, KEGG, and REACTOME. The gene sets enriched were mostly related to cell proliferative processes, for example, DNA repair and E2F targets in Hallmark (**Figure 5A**), regulation of the cell cycle G2/M phase transition and nucleotide excision repair in GO (**Figure 5B**), DNA replication and homologous recombination in KEGG (**Figure 5C**), and cell cycle checkpoints and G1/S DNA damage checkpoints in REACTOME (**Figure 5D**). GSVA was conducted to further confirm the GSEA results by comparing the cellular proliferation activity between groups with high-HOTAIRM1 and low-HOTAIRM1 expression, and a more active state of cellular proliferation was observed in high-HOTAIRM1 expression group (**Figures 5E–H**).

Knockdown of HOTAIRM1 Inhibited Proliferation and Arrested the Cell Cycle of ORAL Squamous Cell Carcinoma Cells

To evaluated the influence on cell proliferation induced by the upregulation of HOTAIRM1 expression in OSCC, we used a sh-HOTAIRM1 strategy to knockdown HOTAIRM1 expression in SCC9 and Cal27 cell lines. The knockdown efficiency of sh-HOTAIRM1 in SCC9 and Cal27 cells were validated by qRT-PCR

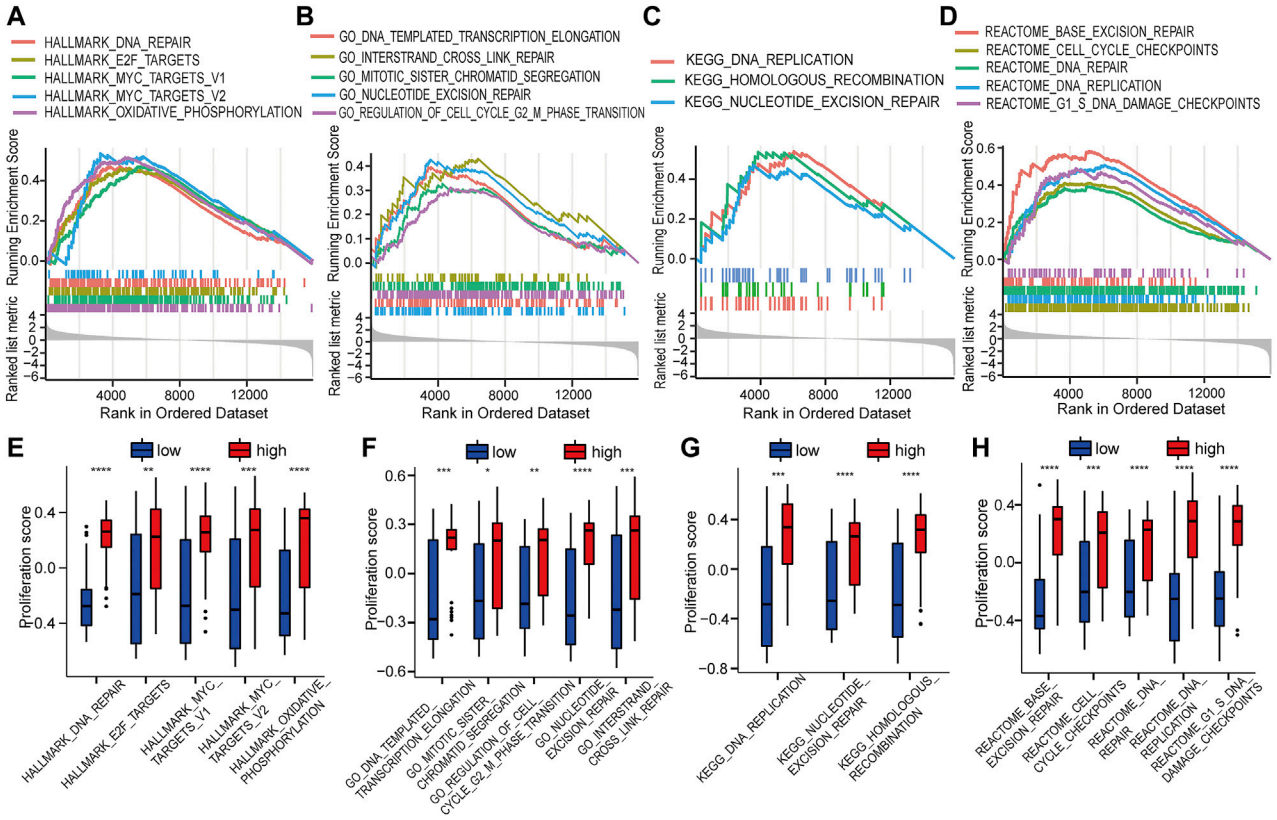


FIGURE 5 | The function of HOTAIRM1 mainly involves cell proliferation in OSCC further indicated by GSEA and GSVA analyses. **(A)** GSEA identified the gene sets significantly enriched in Hallmark. **(B)** GSEA identified the gene sets significantly enriched in GO. **(C)** GSEA identified the gene sets significantly enriched in KEGG. **(D)** GSEA identified the gene sets significantly enriched in REACTOME. **(E–H)** The scores related to cell proliferation assessed in the GSVA in the high-HOTAIRM1 expression group was higher than in the low expression group, suggesting the involvement of HOTAIRM1 in regulating cell proliferation. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$.

(**Figure 6A**), and the results demonstrated that the sh-HOTAIRM1 significantly decreased the expression of HOTAIRM1 in both cell lines compared with the control groups (sh-NC). The follow-up CCK-8 assay indicated that sh-HOTAIRM1 cells presented much slower growth potential than control cells with sh-NC (**Figure 6B**). Correspondingly, colonies produced by sh-HOTAIRM1 cells showed a weak competitiveness in quantity and size compared with sh-NC cells (**Figure 6C**). Cell cycle analysis was carried out using flow cytometry to determine whether the inhibition caused by HOTAIRM1 knockdown in OSCC cells was due to a block in the cell cycle. The analysis results indicated a significant G1 phase arrest in SCC9 and Cal27 cells with HOTAIRM1 knockdown (**Figure 6D**). Furthermore, to provide evidence supporting the role of HOTAIRM1 activity in OSCC at the molecular level, western blotting was performed to evaluate typical regulatory factors and markers associated with cell proliferation and the cell cycle, including the cell proliferation marker PCNA; G1 phase regulatory factor CyclinD1; and key molecules of the cell cycle signaling pathway p53, p21, CDK4, and CDK6. Compared to control groups, the expression of PCNA, Cyclin D1, CDK4, and CDK6 was reduced in sh-HOTAIRM1 OSCC cells, while p53 and p21 expression was markedly increased (**Figure 6E**). The above results confirmed that the knockdown of HOTAIRM1 induced the inhibition of OSCC cell proliferation, altered the cell-cycle distribution, and arrested cells in G1 phase due to the inhibition of the cell cycle signaling pathway.

Knockdown of HOTAIRM1 Inhibited Tumor Growth *in vivo*

To investigate the effect of HOTAIRM1 on OSCC *in vivo*, tumor growth of xenografts in nude mice was monitored. HOTAIRM1 knockdown significantly inhibited the growth of xenograft tumors formed by OSCC cells (**Figure 7A**). As shown in **Figures 7B**,C, both the tumor weight and volume of the sh-HOTAIRM1 group were obviously smaller than those of the sh-NC group. In addition, the USI and MRI findings provided valuable information on tumor progression and provided information on parameters that are not detected by visual inspection. In sh-HOTAIRM1 groups, both slower growth and slightly weaker internal echogenicity of tumors were observed compared to the control groups, provided by B-mode ultrasonography. Poor angiogenesis and micro-angiogenesis in

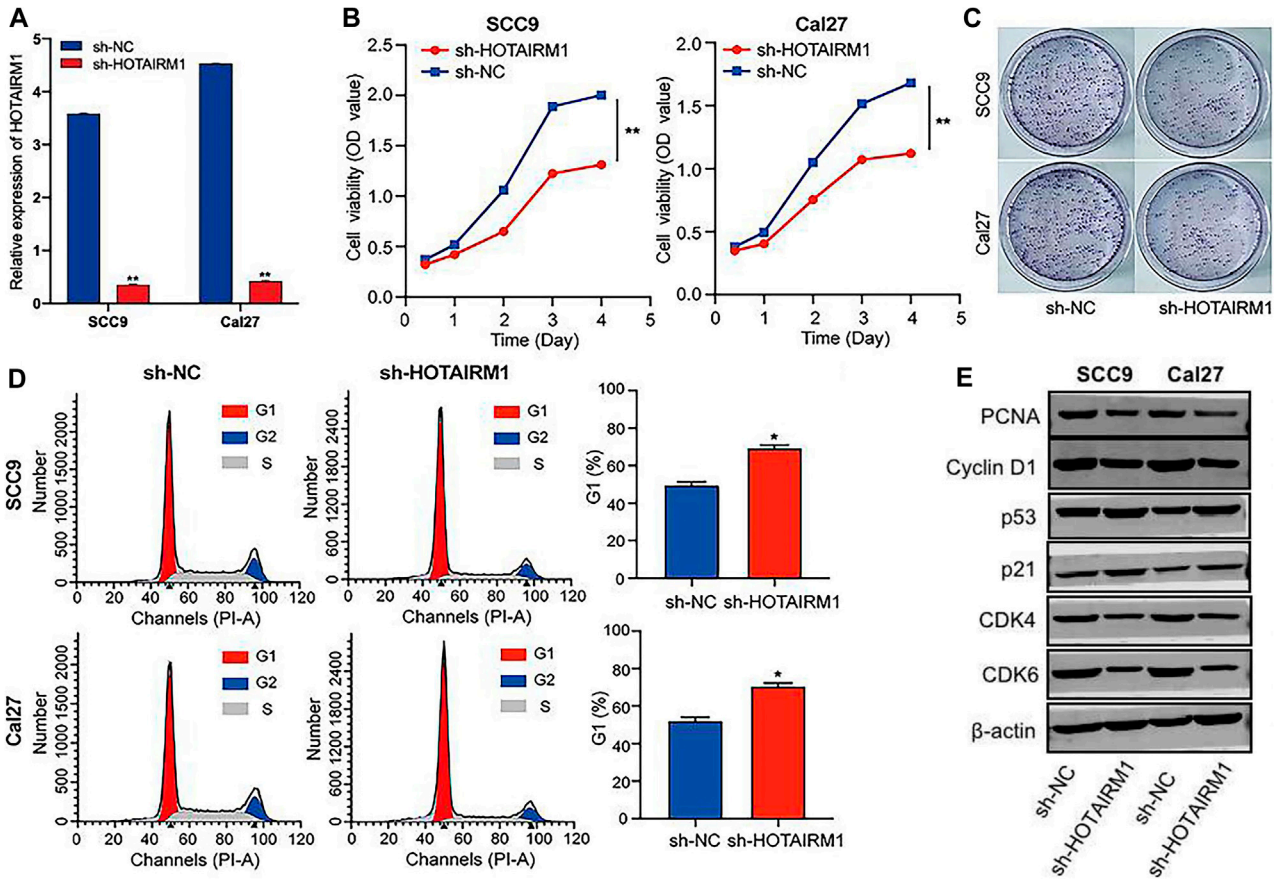


FIGURE 6 | *In vitro* knockdown of HOTAIRM1 restrained cell proliferation and regulated the cell cycle in OSCC cells. **(A)** Knockdown efficiency by sh-HOTAIRM1 in SCC9 and Cal27 cells were validated by qRT-PCR. **(B)** Knockdown of HOTAIRM1 expression in OSCC cells suppressed cell growth measured via the CCK-8 assay. **(C)** Colony formation assays evaluated the influence of HOTAIRM1 knockdown on cell proliferation of OSCC cells. **(D)** Flow cytometry was used to evaluate the impact on cell cycle distribution by knockdown of HOTAIRM1 expression in OSCC cells lines. **(E)** Western blotting showing changes in key biomolecules related to cell proliferation and cell cycle regulation, including PCNA, CyclinD1, p53, p21, CDK4, and CDK6. $^{*}p < 0.05$, $^{**}p < 0.01$.

sh-HOTAIRM1 groups were revealed by Color Doppler Flow Imaging (CDFI) and Color Power Angiography (CPA), respectively. Ultrasonic elastosonography (USE) revealed the slightly weaker hardness of tumors than that in sh-NC groups (**Figure 7D**). Observed by the MRI, the tumor size of the sh-HOTAIRM1 group was obviously smaller than that of the sh-NC group, which was consistent with the results of USI. Furthermore, in both the sh-NC group and sh-HOTAIRM1 group, the tumors presented low signal intensity in T1WI and high signal intensity in T2WI/T2-FLAIR, that were concordant with the conventional MRI performance of OSCC (**Figure 7E**). In summary, these findings convincingly demonstrated that HOTAIRM1-knockdown inhibited tumor growth *in vivo*.

DISCUSSION

As the most common malignant tumor involving the oral and maxillofacial regions, OSCC exerts a severe negative impact on patients due to its poor prognosis and high recurrence rate (Ferlay et al., 2015; Economopoulou et al., 2017; Speight et al., 2017). Currently, the standard measures for treating OSCC involve the surgery to resect the tumor followed by adjunctive therapy such as chemotherapy and radiation. Nonetheless these treatment approaches are not effective in improving the clinical outcome of OSCC patients (Zanoni et al., 2019). Because traditional surgical treatment cannot effectively improve the prognosis of OSCC patients, searching for reliable biomarkers is essential to improve the diagnosis and treatment of OSCC in the early stages of the disease (Li et al., 2020; Jia et al., 2021). Based on previous studies, abnormally expressed lncRNAs have been recognized as moderators of tumor evolution in different cancers, exerting procarcinogenic function or anticancer function by regulating the malignant biological behaviors of tumors directly or indirectly, including processes such as proliferation, migration, invasion, and glycolysis (Schmitt and Chang 2016; Ahadi 2020; Song and Kim 2021). For example, FOXD2-AS1 is upregulated in OSCC and has been reported to inhibit the progression of gallbladder cancer by mediating methylation of MLH1 (Liang et al., 2020). LINC01410 has been reported to be abnormally

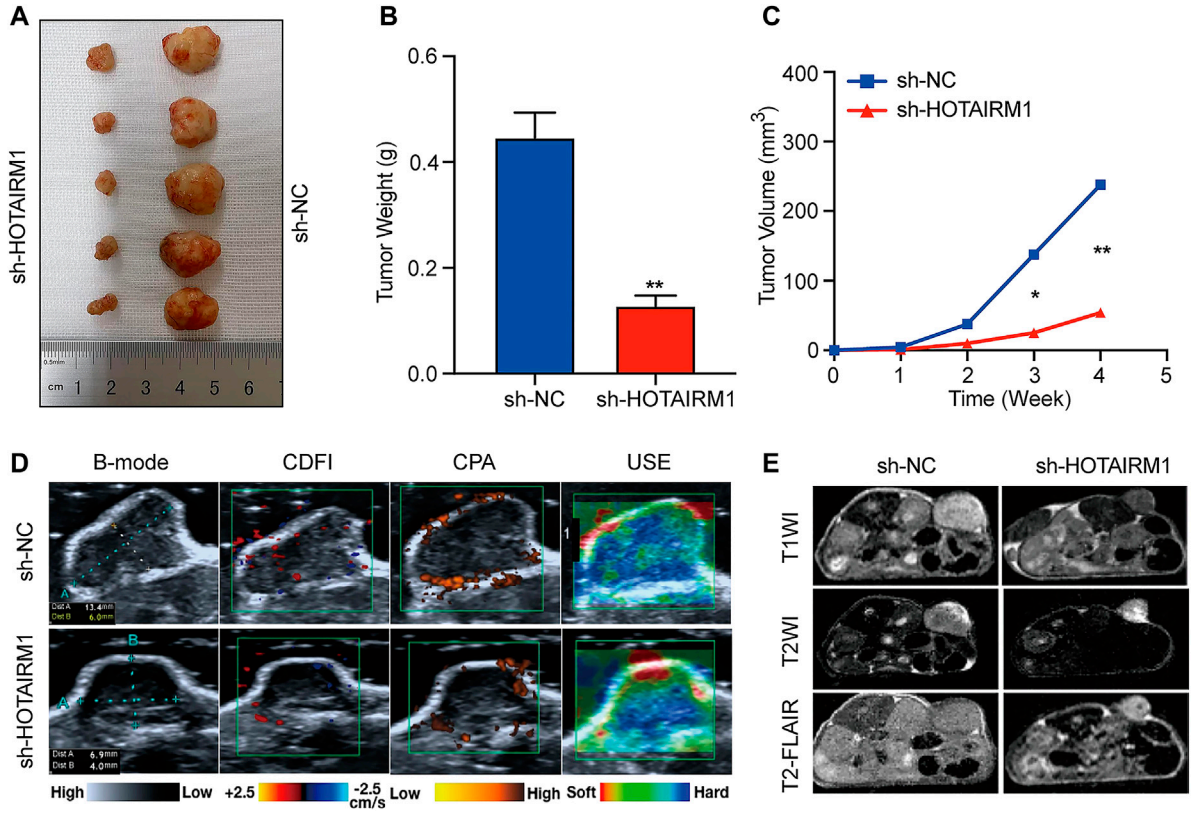


FIGURE 7 | Knockdown of HOTAIRM1 inhibits tumor growth *in vivo*. **(A)** Tumors were excised from nude mice at day 28 after tumor graft inoculation ($n = 5$). **(B–C)** Weight of tumors and Volume were measured and calculated. **(D)** USI evaluation of tumors in mice using B-mode, CDFI, CPA, and USE. **(E)** MRI evaluation of tumors using T1WI, T2WI and T2- FLAIR. $^{*}p < 0.05$, $^{**}p < 0.01$.

expressed in cervical cancer and promotes the growth and invasion of cancer cells by regulating the miR-2467/VOPP1 axis (Liu and Wen 2020).

In this study, our aim was to identify the clinically valuable lncRNAs pivotal to the initiation and progression of OSCC and to further elucidate their functions and clinical significance. Our study was the first to determine that HOTAIRM1 was an upregulated lncRNA in OSCC according to the differential expression analysis, which was subsequently confirmed by its high expression in OSCC cell lines, confirming the relevance between HOTAIRM1 expression and OSCC. However, a review of the literatures relative to HOTAIRM1 showed that HOTAIRM1 was previously explored only in thyroid, ovarian cancer, non-small cell lung cancer, leukemia, and clear cell renal cell carcinoma (Zhang et al., 2014; Chao et al., 2020; Hamilton et al., 2020; Chen D. et al., 2021; Li et al., 2021; Liang et al., 2021). In leukemia, HOTAIRM1 was considered to be unfavorable and it could activate RHOA/ROCK1 pathway to enhance glucocorticoid resistance by inhibiting ARHGAP18 (Liang et al., 2021). In ovarian cancer, silencing HOTAIRM1 promoted cell proliferation and inhibited cell apoptosis by regulating the Wnt pathway and its downstream target gene MMP9 (Ye et al., 2021). Thus, it was worthy to further investigate the functional effects of HOTAIRM1 given the lack of studies evaluating HOTAIRM1 in OSCC. We analyzed the correlation between upregulated HOTAIRM1 expression and clinical data of OSCC patients, in which OS and tumor clinical stages were associated with HOTAIRM1 expression, indicating an unfavorable prognosis as well as the prognostic value of HOTAIRM1 in OSCC.

Some scholars proposed that a succession of genomic alterations in neoplastic cells are crucial triggering factors for cancers (Wu and Li 2021), therefore, we further assessed the correlation between HOTAIRM1 expression patterns and tumorigenic characteristics. In our study, the top five mutant genes most related to HOTAIRM1 are TP53, FAT1, MUC16, DNAH5 and HRAS. TP53 gene mutation was found to be most correlated with abnormal expression of HOTAIRM1. Increasing evidence has shown that the mutational status and the functional regulation of TP53 in cancers are related to the disordered expression of lncRNAs directly or indirectly (Aubrey et al., 2018). TP53, is considered the "guardian of the genome" and triggers cell apoptosis by inhibiting multiple pathways, and participates in the identification of DNA damage via DNA repair processes, thereby playing an important role in maintaining genomic stability (Chatterjee and Viswanathan 2021). In other words, disrupting the "guardian of the genome" could force damaged cells into senescence or

apoptosis, thereby accelerating the accumulation of somatic mutations. Moreover, FAT1 mutation information has been repeatedly detected in human cancers, especially in squamous cell carcinoma (Mountzios et al., 2014); MUC16 was recognized as the third most common mutant oncogene (Bafna et al., 2010); HRAS was a member of the RAS family of oncogenes, and its mutation has been confirmed to be closely related to the initiation of OSCC (Chai et al., 2020). This was also consistent with our analysis, in which in the high-HOTAIRM1 expression group, the somatic mutation rate was significantly higher than that of the low-HOTAIRM1 expression group. In addition, our study also identified obvious genome amplification and deletion patterns (CNVs) in the high-HOTAIRM1 expression group. Similarly, in anaplastic thyroid cancer (ATC), the amplification of HOTAIRM1 genome copy number increased the expression of HOTAIRM1, and drove the occurrence of ATC by inhibiting the biosynthesis of miR-144 (Zhang et al., 2021). Our study indicated that, at the genetic level, HOTAIRM1 may play a stimulatory role in driving OSCC occurrence.

With the progressive revelation of tumour heterogeneity, cancer stem cells (CSCs) are generally considered to be the components of cancer initiation, which are able to form tumors and influence the progression and malignancy of cancers (Clarke 2019). In Chen's study of esophageal squamous cell carcinoma, up-regulated LINC-POU3F3 promoted the upregulation of tumor cell stem cell markers (CD133, CD44 and CD90), thereby enhancing the radiotherapy resistance of tumor cells and increasing the degree of malignancies of esophageal cancer (Chen et al., 2021b). mRNAsi is considered to be an indicator to describe the stemness of tumor cells, which could quantify the CSCs to a certain extent (Malta et al., 2018). In our study, the high expression of HOTAIRM1 was associated with higher mRNAsi, indicating that the higher the expression of HOTAIRM1, the stronger the dryness of OSCC cells, suggesting that HOTAIRM1 may promote the malignant potential of OSCC. The TME is described as the environment of tumor cell production and growth, which includes not only the tumor cells themselves but also the surrounding fibroblasts, immune and inflammatory cells, and other cells, as well as the adjacent intercellular signals, microvessels, and relevant infiltrating biological molecules (Balaji et al., 2021). In our study, the HOTAIRM1 expression was inversely proportional to stromal cell and immune cell content, suggesting that HOTAIRM1 may promote the proliferation of OSCC cells and improve tumor purity. Studies accumulating over the past few decades have shown that, tumor-antagonizing areas are often accompanied by chronic inflammation and on the other hand, the infiltration of immune cells in tumor tissues serve to promote tumor evolution (Denaro et al., 2019). To date the identified tumor-promoting immune cells consist of macrophages, mast cells, and neutrophils, as well as T lymphocytes and B lymphocytes, which are confirmed to contribute to induce and help maintain tumor angiogenesis, to promote tumor cell proliferation, and to facilitate metastasis and dissemination *via* seeding of cancer cells at the edge of the tumor (Anderson 2014). In our study, the HOTAIRMI expression was negatively correlated with macrophage M1 (anti-tumor phenotype) and mast cells, and positively correlated with monocytes and eosinophils. Eosinophils infiltrate multiple cancers and have the ability to regulate tumor progression and promote tumor growth either directly by interacting with cancer cells or indirectly by regulating TME (Grisaru-Tal et al., 2020). In the tumor immune microenvironment, monocytes mainly support tumor cells to escape host immune response by infiltrating tumor and differentiating into tumor-related macrophages, thus affecting tumor progression (Chittezhath et al., 2014). In this research, HOTAIRM1 expression levels were correlated with the types of the immune cells, and we speculated that HOTAIRM1 might influence the tumor immune microenvironment and thus promote the progression of OSCC.

To further explore the biological function of HOTAIRM1 in OSCC, WGCNA was used to identify genes most closely related to HOTAIRM1 expression. We determined that nearly all the crucial genes were involved in cell proliferation and the cell cycle, such as G1/S checkpoint, DNA replication, DNA mismatch repair, and DNA damage detection. Subsequently, the function of HOTAIRM1 was verified using GSEA and GSVA. Moreover, *in vitro* cell function experiments showed the inhibition of proliferation and the arrest in G1 phase cell cycle following HOTAIRM1 knockdown in OSCC cells. As a classical cell proliferation marker at the molecular level (Lu et al., 2019), PCNA protein expression was significantly decreased in the sh-HOTAIRM1 knockdown cells, which also suggested that HOTAIRM1 knockdown inhibited the proliferation of OSCC cells. In addition, the detection of molecular markers related to cell cycle pathway showed that p53 and p21 were significantly increased, while the CyclinD1, CDK4 and CDK6 were decreased in OSCC cells with HOTAIRM1 knockdown. The smooth progression of the cell cycle is a guarantee of cell growth and is mainly controlled by p53 via monitoring checkpoints progression at G1/S and G2/M phases, which is closely related to the transcriptional activation of cell cycle-related proteins (Li et al., 2019). As a p53 downstream gene, p21 is a cyclin-dependent kinases (CDKs) inhibitor, which participates in cell cycle processes, plays a key role in tumors activity through the p53 signal pathway, and can bind with a series of cyclin/CDKs complexes to inhibit the activity of corresponding protein kinases, mainly in the G1 phase (Xiong et al., 2021). CyclinD1, a G1/S-specific cyclin-D1, is a marker of G1 phase, and mainly forms complexes with CDK4 or CDK6 to regulate the cell cycle and is indispensable for G1 phase entry to the S phase (Yan et al., 2021). In various cancers, like in lung cancer, gastric cancer, thyroid cancer, and ovarian cancer (Liang et al., 2019; Qiu et al., 2021b; Yan et al., 2021), p53 and p21 jointly constitute the G1 checkpoint of the cell cycle to ensure that the cancer cells at G1 phase smoothly enter S phase. FKBP11 has been described as a regulator of the cell cycle and apoptosis via p53/p21/p27 and p53/Bcl-2/Bax signaling pathways in OSCC, thereby promoting the proliferation of cancer cells (Qiu et al., 2021b). BCAR4 is upregulated in esophageal squamous cell carcinoma acting on the G1 phase of cell cycle to promote cell proliferation by regulating the miR-139–3p/ELAVL1 axis and the p53/p21 signaling pathway (Yan et al., 2021). In human acute promyelocytic leukemia, HOTAIRM1 knockdown inhibits NB4 granulocyte cells differentiation by maintaining cells in the G1 phase and by regulating integrin gene expression levels (Zhang et al., 2014). In brief, our results revealed that in OSCC, the knockdown of HOTAIRM1 upregulated the expression of p53

and p21, maintaining the cells at G1 phase, likely due to the inhibition of the p53/p21-mediated cell cycle signaling pathway, thereby inhibiting cell proliferation. Finally, in xenograft tumor experiments in nude mice, tumor growth was effectively inhibited when HOTAIRM1 was knockdown in OSCC xenografted cells, and resulted in tumors with smaller weight and volume, weaker blood flow signals, and reduced tumor stiffness, as detected using ultrasound imaging and MRI. It is well known that ultrasound imaging and MRI are authoritative tools used to evaluate tumors activity in the clinic (Smiley et al., 2019; Lalfamkima et al., 2021). Altogether, it can be concluded that HOTAIRM1 exhibits tumorigenic properties and facilitates tumor growth in OSCC.

In conclusion, our study was the first to identify HOTAIRM1 as a significantly upregulated lncRNA in OSCC, which is closely related to poor prognosis of patients and may represent a new potential biomarker for OSCC. Moreover, HOTAIRM1 was shown to be carcinogenic in OSCC, by promoting cell proliferation and by accelerating cell cycle progression, via regulatory factors in the p53/p21 pathway induced by abnormal HOTAIRM1 expression. Overall, our study revealed that HOTAIRM1 acts as a novel possible prognostic biomarker and therapeutic target for OSCC.

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 animal study was reviewed and approved by the Committee on Animals of the First Affiliated Hospital of Harbin Medical University.

AUTHOR CONTRIBUTIONS

YY performed the experiments, data analysis, and wrote the manuscript. JN downloaded and assembled the data. XZ and XW checked the results. HS and YL performed the statistical analysis and manuscript revision. XJ and FC designed and guided the research. All the authors have read and approved the manuscript.

FUNDING

This research was funded by the Postgraduate Research & Practice Innovation Program of Harbin Medical University (Grant No. YJSSJCX 2019-13HYD).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2021.798584/full#supplementary-material

REFERENCES

Ahadi, A. (2021). Functional Roles of lncRNAs in the Pathogenesis and Progression of Cancer. *Genes Dis.* 8 (4), 424–437. doi:10.1016/j.gendis.2020.04.009

Anderson, C. C. (2014). Application of central Immunologic Concepts to Cancer: Helping T Cells and B Cells Become Intolerant of Tumors. *Eur. J. Immunol.* 44 (7), 1921–1924. doi:10.1002/eji.201444826

Aubrey, B. J., Janic, A., Chen, Y., Chang, C., Lieschke, E. C., Diepstraten, S. T., et al. (2018). Mutant TRP53 Exerts a Target Gene-Selective Dominant-Negative Effect to Drive Tumor Development. *Genes Dev.* 32 (21-22), 1420–1429. doi:10.1101/gad.314286.118

Bafna, S., Kaur, S., and Batra, S. K. (2010). Membrane-bound Mucins: the Mechanistic Basis for Alterations in the Growth and Survival of Cancer Cells. *Oncogene* 29 (20), 2893–2904. doi:10.1038/onc.2010.87

Balaji, S., Kim, U., Muthukkaruppan, V., and Vanniarajan, A. (2021). Emerging Role of Tumor Microenvironment Derived Exosomes in Therapeutic Resistance and Metastasis through Epithelial-To-Mesenchymal Transition. *Life Sci.* 280, 119750. doi:10.1016/j.lfs.2021.119750

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: a Cancer J. clinicians* 68 (6), 394–424. doi:10.3322/caac.21492

Chai, A. W. Y., Lim, K. P., and Cheong, S. C. (2020). Translational Genomics and Recent Advances in Oral Squamous Cell Carcinoma. *Semin. Cancer Biol.* 61, 71–83. doi:10.1016/j.semcancer.2019.09.011

Chao, H., Zhang, M., Hou, H., Zhang, Z., and Li, N. (2020). HOTAIRM1 Suppresses Cell Proliferation and Invasion in Ovarian Cancer through Facilitating ARHGAP24 Expression by Sponging miR-106a-5p. *Life Sci.* 243, 117296. doi:10.1016/j.lfs.2020.117296

Chatterjee, M., and Viswanathan, P. (2021). Long Noncoding RNAs in the Regulation of P53-mediated Apoptosis in Human Cancers. *Cell Biol Int* 45 (7), 1364–1382. doi:10.1002/cbin.11597

Chen, D., Li, Y., Wang, Y., and Xu, J. (2021a). LncRNA HOTAIRM1 Knockdown Inhibits Cell Glycolysis Metabolism and Tumor Progression by miR-498/ABCE1 axis in Non-small Cell Lung Cancer. *Genes Genom* 43 (2), 183–194. doi:10.1007/s13258-021-01052-9

Chen, T.-j., Gao, F., Yang, T., Li, H., Li, Y., Ren, H., et al. (2020). LncRNA HOTAIRM1 Inhibits the Proliferation and Invasion of Lung Adenocarcinoma Cells via the miR-498/WWOX Axis. *Cmar* Vol. 12, 4379–4390. doi:10.2147/CMAR.S244573

Chen, Y., Tang, J., Li, L., and Lu, T. (2021b). Effect of Linc-Pou3f3 on Radiotherapy Resistance and Cancer Stem Cell Markers of Esophageal Cancer Cells. Linc-Pou3f3. *Zhong Nan da Xue Xue Bao. Yi Xue Ban = J. Cent. South UniversityMedical Sci.* 46 (6), 583–590. doi:10.11817/j.issn.1672-7347.2021.190758

Chittezhath, M., Dhillon, M. K., Lim, J. Y., Laoui, D., Shalova, I. N., Teo, Y. L., et al. (2014). Molecular Profiling Reveals a Tumor-Promoting Phenotype of Monocytes and Macrophages in Human Cancer Progression. *Immunity* 41 (5), 815–829. doi:10.1016/j.immuni.2014.09.014

Clarke, M. F. (2019). Clinical and Therapeutic Implications of Cancer Stem Cells. *N. Engl. J. Med.* 380, 2237–2245. doi:10.1056/NEJMra1804280

Denaro, N., Merlano, M. C., and Lo Nigro, C. (2019). Long Noncoding RNA S as Regulators of Cancer Immunity. *Mol. Oncol.* 13 (1), 61–73. doi:10.1002/1878-0261.12413

Economopoulou, P., Kotsantis, I., Kyrodimos, E., Lianidou, E. S., and Psyrri, A. (2017). Liquid Biopsy: An Emerging Prognostic and Predictive Tool in Head and Neck Squamous Cell Carcinoma (HNSCC). Focus on Circulating Tumor Cells (CTCs). *Oral Oncol.* 74, 83–89. doi:10.1016/j.oraloncology.2017.09.012

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., et al. (2015). Cancer Incidence and Mortality Worldwide: Sources, Methods and

Major Patterns in GLOBOCAN 2012. *Int. J. Cancer* 136 (5), E359–E386. doi:10.1002/ijc.29210

Grisaru-Tal, S., Itan, M., Klion, A. D., and Munitz, A. (2020). A New Dawn for Eosinophils in the Tumour Microenvironment. *Nat. Rev. Cancer* 20 (10), 594–607. doi:10.1038/s41568-020-0283-9

Hamilton, M. J., Young, M., Jang, K., Sauer, S., Neang, V. E., King, A. T., et al. (2020). HOTAIRM1 lncRNA Is Downregulated in clear Cell Renal Cell Carcinoma and Inhibits the Hypoxia Pathway. *Cancer Lett.* 472, 50–58. doi:10.1016/j.canlet.2019.12.022

Jia, H., Wang, X., and Sun, Z. (2021). Screening and Validation of Plasma Long Non-coding RNAs as B-iomarkers for the E-arly D-iagnosis and S-taging of O-ral S-quamous C-ell C-arcinoma. *Oncol. Lett.* 21 (2), 172. doi:10.3892/ol.2021.12433

Lalchhuanawma, T., Lalfamkima, F., Georgeno, G., Rao, N., Selvakumar, R., Devadoss, V., et al. (2021). Clinical Diagnostic Criteria versus Advanced Imaging in Prediction of Cervical Lymph Node Metastasis in Oral Squamous Cell Carcinomas: A Magnetic Resonance Imaging Based Study. *J. Carcinog* 20, 3. doi:10.4103/jcar.JCar_27_20

Langfelder, P., and Horvath, S. (2008). WGCNA: an R Package for Weighted Correlation Network Analysis. *BMC Bioinformatics* 9, 559. doi:10.1186/1471-2105-9-559

Li, C., Chen, X., Liu, T., and Chen, G. (2021). lncRNA HOTAIRM1 Regulates Cell Proliferation and the Metastasis of Thyroid Cancer by Targeting Wnt10b. *Oncol. Rep.* 45 (3), 1083–1093. doi:10.3892/or.2020.7919

Li, V. D., Li, K. H., and Li, J. T. (2019). TP53 Mutations as Potential Prognostic Markers for Specific Cancers: Analysis of Data from the Cancer Genome Atlas and the International Agency for Research on Cancer TP53 Database. *J. Cancer Res. Clin. Oncol.* 145 (3), 625–636. doi:10.1007/s00432-018-2817-z

Li, Y., Cao, X., and Li, H. (2020). Identification and Validation of Novel Long Non-coding RNA Biomarkers for Early Diagnosis of Oral Squamous Cell Carcinoma. *Front. Bioeng. Biotechnol.* 8, 256. doi:10.3389/fbioe.2020.00256

Liang, L., Gu, W., Li, M., Gao, R., Zhang, X., Guo, C., et al. (2021). The Long Noncoding RNA HOTAIRM1 Controlled by AML1 Enhances Glucocorticoid Resistance by Activating RHOA/ROCK1 Pathway through Suppressing ARHGAP18. *Cell Death Dis* 12 (7), 702. doi:10.1038/s41419-021-03982-4

Liang, Q., Yao, Q., and Hu, G. (2019). CyclinD1 Is a New Target Gene of Tumor Suppressor MiR-520e in Breast Cancer. *Open Med. (Warsaw, Poland)* 14, 913–919. doi:10.1515/med-2019-0108

Liang, X., Chen, Z., and Wu, G. (2020). FOXD2-AS1 Predicts Dismal Prognosis for Oral Squamous Cell Carcinoma and Regulates Cell Proliferation. *Cel Transpl.* 29, 096368972096441. doi:10.1177/0963689720964411

Lingen, M. W., Kalmar, J. R., Karrison, T., and Speight, P. M. (2008). Critical Evaluation of Diagnostic Aids for the Detection of Oral Cancer. *Oral Oncol.* 44 (1), 10–22. doi:10.1016/j.oraloncology.2007.06.011

Liu, F., and Wen, C. (2020). LINC01410 Knockdown Suppresses Cervical Cancer Growth and Invasion via Targeting miR-2467-3p/VOPP1 Axis. *Cmar* 12, 855–861. doi:10.2147/CMAR.S236832

Lu, E. M.-C., Ratnayake, J., and Rich, A. M. (2019). Assessment of Proliferating Cell Nuclear Antigen (PCNA) Expression at the Invading Front of Oral Squamous Cell Carcinoma. *BMC oral health* 19 (1), 233. doi:10.1186/s12903-019-0928-9

Malta, T. M., Sokolov, A., Gentles, A. J., Burzykowski, T., Poisson, L., Weinstein, J. N., et al. (2018). Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* 173, 338–e15. doi:10.1016/j.cell.2018.03.034

Markert, L., Holdmann, J., Klinger, C., Kaufmann, M., Schork, K., Turewicz, M., et al. (2021). Small RNAs as Biomarkers to Differentiate Benign and Malign Prostate Diseases: An Alternative for Transrectal Punch Biopsy of the Prostate?. *PloS one* 16 (3), e0247930. doi:10.1371/journal.pone.0247930

Mountzios, G., Rampias, T., and Psyrri, A. (2014). The Mutational Spectrum of Squamous-Cell Carcinoma of the Head and Neck: Targetable Genetic Events and Clinical Impact. *Ann. Oncol.* 25 (10), 1889–1900. doi:10.1093/annonc/mdu143

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

Ozawa, T., Matsuyama, T., Toiyama, Y., Takahashi, N., Ishikawa, T., Uetake, H., et al. (2017). CCAT1 and CCAT2 Long Noncoding RNAs, Located within the 8q.24.21 'gene Desert', Serve as Important Prognostic Biomarkers in Colorectal Cancer. *Ann. Oncol.* 28 (8), 1882–1888. doi:10.1093/annonc/mdx248

Qiu, H., Cao, S., and Xu, R. (2021a). Cancer Incidence, Mortality, and burden in China: a Time-trend Analysis and Comparison with the United States and United Kingdom Based on the Global Epidemiological Data Released in 2020. *Cancer Commun.* 41, 1037–1048. doi:10.1002/cac2.12197

Qiu, L., Liu, H., Wang, S., Dai, X.-H., Shang, J.-W., Lian, X.-L., et al. (2021b). FKBP11 Promotes Cell Proliferation and Tumorigenesis via P53-Related Pathways in Oral Squamous Cell Carcinoma. *Biochem. biophysical Res. Commun.* 559, 183–190. doi:10.1016/j.bbrc.2021.04.096

Rong, D., Sun, G., Wu, F., Cheng, Y., Sun, G., Jiang, W., et al. (2021). Epigenetics: Roles and Therapeutic Implications of Non-coding RNA Modifications in Human Cancers. *Mol. Ther. - Nucleic Acids* 25, 67–82. doi:10.1016/j.omtn.2021.04.021

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P. (2011). A ceRNA Hypothesis: the Rosetta Stone of a Hidden RNA Language?. *Cell* 146 (3), 353–358. doi:10.1016/j.cell.2011.07.014

Schmitt, A. M., and Chang, H. Y. (2016). Long Noncoding RNAs in Cancer Pathways. *Cancer cell* 29 (4), 452–463. doi:10.1016/j.ccell.2016.03.010

Smiley, N., Anzai, Y., Foster, S., and Dillon, J. (2019). Is Ultrasound a Useful Adjunct in the Management of Oral Squamous Cell Carcinoma?. *J. Oral Maxillofac. Surg.* 77 (1), 204–217. doi:10.1016/j.joms.2018.08.012

Song, H. K., and Kim, S. Y. (2021). The Role of Sex-specific Long Non-coding RNAs in Cancer Prevention and Therapy. *J. Cancer Prev.* 26 (2), 98–109. doi:10.15430/JCP.2021.26.2.98

Speight, P. M., Epstein, J., Kujan, O., Lingen, M. W., Nagao, T., Ranganathan, K., et al. (2017). Screening for Oral Cancer-A Perspective from the Global Oral Cancer Forum. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* 123 (6), 680–687. doi:10.1016/j.oooo.2016.08.021

Winkle, M., El-Daly, S. M., Fabbri, M., and Calin, G. A. (2021). Noncoding RNA Therapeutics - Challenges and Potential Solutions. *Nat. Rev. Drug Discov.* 20 (8), 629–651. doi:10.1038/s41573-021-00219-z

Wu, S., and Li, X. (2021). A Genomic Instability-Derived Risk index Predicts Clinical Outcome and Immunotherapy Response for clear Cell Renal Cell Carcinoma. *Bioengineered* 12 (1), 1642–1662. doi:10.1080/21655979.2021.1922330

Xie, P., Li, X., Chen, R., Liu, Y., Liu, D., Liu, W., et al. (2020). Upregulation of HOTAIRM1 Increases Migration and Invasion by Glioblastoma Cells. *Aging* 13 (2), 2348–2364. doi:10.18632/aging.202263

Xiong, L., Deng, N., Zheng, B., Li, T., and Liu, R. H. (2021). Goji berry (Lycium spp.) Extracts Exhibit Antiproliferative Activity via Modulating Cell Cycle Arrest, Cell Apoptosis, and the P53 Signaling Pathway. *Food Funct.* 12 (14), 6513–6525. doi:10.1039/d1fo01105g

Xu, F., Shen, J., and Xu, S. (2021b). Integrated Bioinformatical Analysis Identifies GIMAP4 as an Immune-Related Prognostic Biomarker Associated with Remodeling in Cervical Cancer Tumor Microenvironment. *Front. Cel Dev. Biol.* 9, 637400. doi:10.3389/fcell.2021.637400

Xu, M., Cui, R., Ye, L., Wang, Y., Wang, X., Zhang, Q., et al. (2021a). LINC00941 Promotes Glycolysis in Pancreatic Cancer by Modulating the Hippo Pathway. *Mol. Ther. - Nucleic Acids* 26, 280–294. doi:10.1016/j.omtn.2021.07.004

Yan, S., Xu, J., Liu, B., Ma, L., Feng, H., Tan, H., et al. (2021). Long Non-coding RNA BCAR4 Aggravated Proliferation and Migration in Esophageal Squamous Cell Carcinoma by Negatively Regulating P53/p21 Signaling Pathway. *Bioengineered* 12 (1), 682–696. doi:10.1080/21655979.2021.1887645

Ye, L., Meng, X., Xiang, R., Li, W., and Wang, J. (2021). Investigating Function of Long Noncoding RNA of HOTAIRM1 in Progression of SKOV3 Ovarian Cancer Cells. *Drug Dev. Res.* 82, 1162–1168. doi:10.1002/ddr.2182110.1002/ddr.21821

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring Tumour Purity and Stromal and Immune Cell Admixture from Expression Data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612

Zanoni, D. K., Montero, P. H., Migliacci, J. C., Shah, J. P., Wong, R. J., Ganly, I., et al. (2019). Survival Outcomes after Treatment of Cancer of the Oral Cavity (1985-2015). *Oral Oncol.* 90, 115–121. doi:10.1016/j.oraloncology.2019.02.001

Zhang, L., Zhang, J., Li, S., Zhang, Y., Liu, Y., Dong, J., et al. (2021). Genomic Amplification of Long Noncoding RNA HOTAIRM1 Drives Anaplastic Thyroid Cancer Progression via Repressing miR-144 Biogenesis. *RNA Biol.* 18 (4), 547–562. doi:10.1080/15476286.2020.1819670

Zhang, X., Weissman, S. M., and Newburger, P. E. (2014). Long Intergenic Non-coding RNA HOTAIRM1 Regulates Cell Cycle Progression during Myeloid Maturation in NB4 Human Promyelocytic Leukemia Cells. *RNA Biol.* 11 (6), 777–787. doi:10.4161/rna.28828

Zhao, Y., and Ruan, X. (2020). Identification of PGRMC1 as a Candidate Oncogene for Head and Neck Cancers and its Involvement in Metabolic Activities. *Front. Bioeng. Biotechnol.* 7, 438. doi:10.3389/fbioe.2019.00438

Zhu, M., Chen, Q., Liu, X., Sun, Q., Zhao, X., Deng, R., et al. (2014). lncRNA H19/miR-675 axis Represses Prostate Cancer Metastasis by Targeting TGFBI. *Febs J.* 281 (16), 3766–3775. doi:10.1111/febs.12902

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 Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 11 January 2022
doi: 10.3389/fbioe.2021.814813

A Novel Model Based on Necroptosis-Related Genes for Predicting Prognosis of Patients With Prostate Adenocarcinoma

*Xin-yu Li[1,2†], Jian-xiong You[1†], Lu-yu Zhang[3†], Li-xin Su[1] and Xi-tao Yang[1]**

[1]Department of Interventional Radiotherapy, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [2]Department of Neurosurgery, Shanghai Ninth People's Hospital, Shanghai JiaoTong University School of Medicine, Shanghai, China, [3]Department of Urologic Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

Background: Necroptosis is a newly recognized form of cell death. Here, we applied bioinformatics tools to identify necroptosis-related genes using a dataset from The Cancer Genome Atlas (TCGA) database, then constructed a model for prognosis of patients with prostate cancer.

Methods: RNA sequence (RNA-seq) data and clinical information for Prostate adenocarcinoma (PRAD) patients were obtained from the TCGA portal (http://tcga-data.nci.nih.gov/tcga/). We performed comprehensive bioinformatics analyses to identify hub genes as potential prognostic biomarkers in PRAD u followed by establishment and validation of a prognostic model. Next, we assessed the overall prediction performance of the model using receiver operating characteristic (ROC) curves and the area under curve (AUC) of the ROC.

Results: A total of 5 necroptosis-related genes, namely *ALOX15, BCL2, IFNA1, PYGL* and *TLR3*, were used to construct a survival prognostic model. The model exhibited excellent performance in the TCGA cohort and validation group and had good prediction accuracy in screening out high-risk prostate cancer patients.

Conclusion: We successfully identified necroptosis-related genes and constructed a prognostic model that can accurately predict 1- 3-and 5-years overall survival (OS) rates of PRAD patients. Our riskscore model has provided novel strategy for the prediction of PRAD patients' prognosis.

Keywords: prostate adenocarcinoma, necroptosis, prognosis, model introduction, cancer

OPEN ACCESS

Edited by:
Francesco Pappalardo, University of Catania, Italy

Reviewed by:
Ming Yi, Huazhong University of Science and Technology, China
Quan Wang, Peking University, China

****Correspondence:***
Xi-tao Yang
xitao123456@126.com

†These authors share first authorship

Specialty section:
This article was submitted to Preclinical Cell and Gene Therapy, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *14 November 2021*
Accepted: *15 December 2021*
Published: *11 January 2022*

Citation:
Li X-y, You J-x, Zhang L-y, Su L-x and Yang X-t (2022) A Novel Model Based on Necroptosis-Related Genes for Predicting Prognosis of Patients With Prostate Adenocarcinoma. Front. Bioeng. Biotechnol. 9:814813. doi: 10.3389/fbioe.2021.814813

Abbreviations: AUC, Area under the curve; APC, Antigen presenting cell; CCR, Cytokine-cytokine receptor; CI, Confidence interval; DEGs, Differentially expressed genes; FDR, False discovery rate; GO, Gene Ontology; HLA, Human leukocyte antigen; HR, Hazard ratio; KEGG, Kyoto Encyclopedia of Genes and Genomes; OS, Overall survival; PRAD, Prostate adenocarcinoma; ROC, Receiver operating characteristic; ssGSEA, Single-sample gene set enrichment analysis; PCA, Principal component analysis; aDC, Activated dendritic cell; iDC, Immature dendritic cell; pDC, Plasmacytoid dendritic cell; TCGA, The Cancer Genome Atlas; Tcm, T central memory; Tfh, T follicular helper cell; TIL, Tumor Infiltrating Lymphocyte; Tgd, T gamma delta.

BACKGROUND INFORMATION

Prostate adenocarcinoma (PRAD) is a complex but common malignancy that accounts for about 1300000 new cases and 360,000 deaths every year worldwide. Notably, PRAD accounts for 15% of all new tumor-related cases, making it the second most common neoplasia in elderly males, and the fifth most frequent cause of cancer-related deaths worldwide (Bray et al., 2018; Siegel et al., 2020). In China, PRAD is fast increasing, owing to the rapid socio-economic development in the country, coupled with changes in people's living and eating habits, as well as an increase in the aging population. Consequently, the disease has become one of the most common urogenital malignancies among elderly Chinese males (Wei et al., 2020). Current treatment options for PRAD are limited, while patient prognosis remains unsatisfactory. Therefore, prospecting for novel prognostic markers is imperative to development of effective treatment strategies and enhanced prognosis of PRAD patients. Previous studies have shown that necroptosis, which was first identified and named in 2005, is regulated by intracellular signaling pathways (Degterev et al., 2005). Notably, this phenomenon can be unmediated by caspases, thus functioning upon inhibition of apoptotic pathways. Moreover, its cellular morphology is consistent with that of conventional necrosis. Additional research evidences have shown that necroptosis is not only an important mechanism underlying cell death, but also plays a crucial role in development and progression of many immune system diseases, including chronic kidney diseases, cerebral ischemia, myocardial ischemia, acute and chronic neurodegenerative diseases, as well as tumors and many other human pathological activities (Jouan-Lanhouet et al., 2014; Barbosa et al., 2018). Although necroptosis exerts important functions in oncogenesis and anticancer processes, only a handful of studies have described its significance in PRAD. In the present study, we systematically analyzed differential expression profiles of necroptosis-related genes between normal and PRAD tissues, then developed a novel risk-score-based model for predicting prognosis of PRAD patients.

MATERIAL AND METHODS

Data Acquisition and Differential Gene Expression Analysis

Necroptosis-related genes were extracted from previous studies (Fan et al., 2014; Frank and Vince, 2019; Gong et al., 2019; Yuan et al., 2019; Tang et al., 2020). RNA sequence dataset belonging to 499 PRAD patients and 52 normal controls, together with corresponding clinical information were accessed and downloaded from the TCGA database. Next, we employed the "limma" package implemented in R software to identify differentially expressed genes (DEGs) between the tumor and adjacent normal tissues, based on FDR <0.05 and $|\log2FC| \geq 1$. Thereafter, we recruited a total of 80 PRAD patients at The First Affiliated Hospital of Zhengzhou University, as the validation cohort. All patients voluntarily signed a written informed consent prior to inclusion in the study, and ethical approval for the study protocol was obtained from The First Affiliated Hospital of Zhengzhou University.

Construction and Validation of a Prognostic Model

The prognostic value of each DEG was first assessed by univariate Cox regression analysis, then genes that were significantly correlated with OS in PRAD patients identified. To avoid overfitting of the model, we performed Lasso regression analysis to further select significant prognostic genes for OS in PRAD patients, using a penalty parameter tuning (λ) that was conducted by 10-fold cross-validation based on minimum criteria. The identified significantly expressed genes were then incorporated into a multivariate Cox regression model, and the risk score of each patient calculated using the following formula: risk score = esum (each gene's expression × corresponding coefficient). Median risk-scores were then used to stratify patients into either high- or low-risk groups, and validation of model feasibility and accuracy conducted by generating AUC of the ROC as well as calibration plots. Next, we applied the Cox proportional hazards regression model to analyze these risk factors in PRAD, targeting risk scores, gender, age, as well as the T, N, and M stages. To validate the established model, based on TCGA, we used the median scores to divide patients in the validation cohort into high- and low-risk groups. We also validated the model by stratifying patients in the validation cohort into low- and high-risk subgroups based on the median value of risk scores using the same formula as in the TCGA cohort.

Validation of Gene Expression

Next, we performed quantitative real-time polymerase chain reaction (qRT-PCR) analysis to quantify expression of DEGs used for model construction in the validation cohort. Briefly, total RNA was extracted from thoroughly ground (under liquid nitrogen) target tissues using the Trizol reagent (Life Technology, Grand Island, NY, United States). The RNA was reverse transcribed to complementary DNA (cDNA) using the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific, Lithuania), then subjected to qRT-PCR using the SYBR® Green qPCR mix 2.0 kit performed on the Roche LightCycler® 480 Real-Time PCR System. The primers for the genes targeted in this study were obtained from TsingKe Biological Technology (Nanjing, China), and their sequences are as follows: PYGL (forward 5′- CAGCCTATG GATACGGCATTC -3′, reverse 5′- CGGTGTTGGTGTGTT CTACTTT-3′), ALOX15 (forward 5′-GGGCAAGGAGAC AGAACTCAA-3′, reverse 5′- CAGCGGTAACAAGGGAAC CT-3′), BCL2 (forward 5′- GGTGGGGTCATGTGTGTGG -3′, reverse 5′-CGGTTCAGGTACTCAGTCATCC-3′), IFNA1 (forward 5′-GCCTCGCCCTTTGCTTTACT-3′, reverse 5′-CTGTGGGTCTCAGGGAGATCA -3′), TLR3 (forward 5′- TTGCCTTGTATCTACTTTTGGGG -3′, reverse 5′-TCAACACTGTTATGTTTGTGGGT -3′) β-actin (Forward: 5′-CGAGCACAGAGCCTCGCCTTTGCC-3′, Reverse: 5′-TGTCGACGACGAGCGCGGCGATAT-3').

TABLE 1 | Basic clinical characteristics of PRAD patients in TCGA PRAD cohort.

Characteristic	Levels	Overall
n		499
T stage, n (%)	T2	189 (38.4%)
	T3	292 (59.3%)
	T4	11 (2.2%)
N stage, n (%)	N0	347 (81.5%)
	N1	79 (18.5%)
M stage, n (%)	M0	455 (99.3%)
	M1	3 (0.7%)
Race, n (%)	Asian	12 (2.5%)
	Black or African American	57 (11.8%)
	White	415 (85.7%)
Age, n (%)	≤60	224 (44.9%)
	>60	275 (55.1%)
Age, median (IQR)		61 (56, 66)

Relative mRNA expression was calculated using the 2-ΔΔCt method.

Functional Enrichment and Drug Sensitivity Analyses

We first used the "cluster Profiler" package in R to perform Gene Ontology (GO) functional and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses on the screened necroptosis-related DEGs (Zeng et al., 2021). Next, we applied the "org.Hs.eg.db" package to identify significantly enriched genes and classify gene clusters. All analyses were performed at a statistical significance of $p < 0.05$. The obtained DEG list was compared with a reference dataset from the CMap database, to obtain a correlation score based on enrichment of DEGs in the reference gene expression profile. A positive number indicated that the DEG was similar to the reference gene expression profile, thus allowed analysis of similar interventions that produce the above effects. Conversely, a negative number indicated that the DEGs had an opposite gene expression profile to that of the reference, suggesting that the drug may antagonize the DEGs' expression. Ultimately, potentially effective drugs that can counteract drug resistance were inferred from the genetic changes in drug-resistant cell lines, and ranked according to correlation scores obtained on the reference gene expression profile.

Analysis of Immune Infiltration and Tumor Microenvironment

Considering that immune infiltration levels are correlated with survival and prognosis of cancer patients, we evaluated the relationship between risk-scores and immune infiltration levels. Specifically, we applied multiple algorithms implemented in the IOBR package in R (Zeng et al., 2021) to estimate proportions of tumor-infiltrating immune cells. Next, we calculated stromal and immune scores for each sample using the ESTIMATE algorithm implemented in the 'limma' and 'estimate' packages in R.

RESULTS

Profiles of DEGs

A total of 499 and 80 PRAD patients from the TCGA and validation cohorts, respectively, were included in this study. Detailed clinical characteristics of these patients are shown in **Tables 1**, **2**. A total of 16 necroptosis-related genes were differentially expressed between tumor and adjacent non-tumor tissues. Profiles of these expression patterns are presented using heatmaps in **Figure 1A**, while the relationships among the DEGs are shown in **Figure 1B**. Next, we performed LASSO regression analysis to screen covariates, then applied a 10-fold cross validation with minimum criteria to obtain an optimal λ. The final value of λ yielded a minimum cross validation error. Consequently, a total of 5 DEGs, namely *ALOX15, BCL2, IFNA1, PYGL* and *TLR3*, were identified (**Figures 2A,B**). Incorporation of these significant DEGs into a multivariate Cox regression model revealed coefficients of included each gene. Finally, each PRAD patient was assigned a separate risk score according to the aforementioned formula.

Construction and Validation of a Prognostic Model

Risk scores = 0.05834* expression level of *ALOX15*–0.0004914* expression level of *BCL2*+ 0.665* expression level of *IFNA1*+0.03093* expression level of *PYGL* −0.0008311* expression level of *TLR3*. The resulting median cut-off value was used to stratify patients into high-risk and low-risk groups, comprising 249 and 250 individuals, respectively (**Figure 3A**). A principal component analysis (PCA) plot showed that patients in both groups were distributed in different directions, indicating that the established model had excellent predictive ability to distinguish between high and low-risk PRAD (**Figure 3B**). Kaplan-Meier curves showed that patients in the low-risk group were significantly associated with higher survival rates compared to their high-risk counterparts ($p < 0.05$) (**Figure 3C**). In addition, we generated time-dependent ROC curves to estimate performance of the risk prediction model. AUC values for the prognostic model were 0.822, 0.856, and 0.795

TABLE 2 | Basic clinical characteristics of PRAD patients in validation cohort.

Characteristic	Levels	Overall
n		80
T stage, n (%)	T2	32 (40%)
	T3	40 (50%)
	T4	8 (10%)
N stage, n (%)	N0	64 (80%)
	N1	16 (20%)
M stage, n (%)	M0	72 (90%)
	M1	8 (10%)
Age, n (%)	≤60	37 (46%)
	>60	43 (54%)
OS event, n (%)	Alive	64 (80%)
	Dead	16 (20%)
Age, median (IQR)		59 (52, 69)

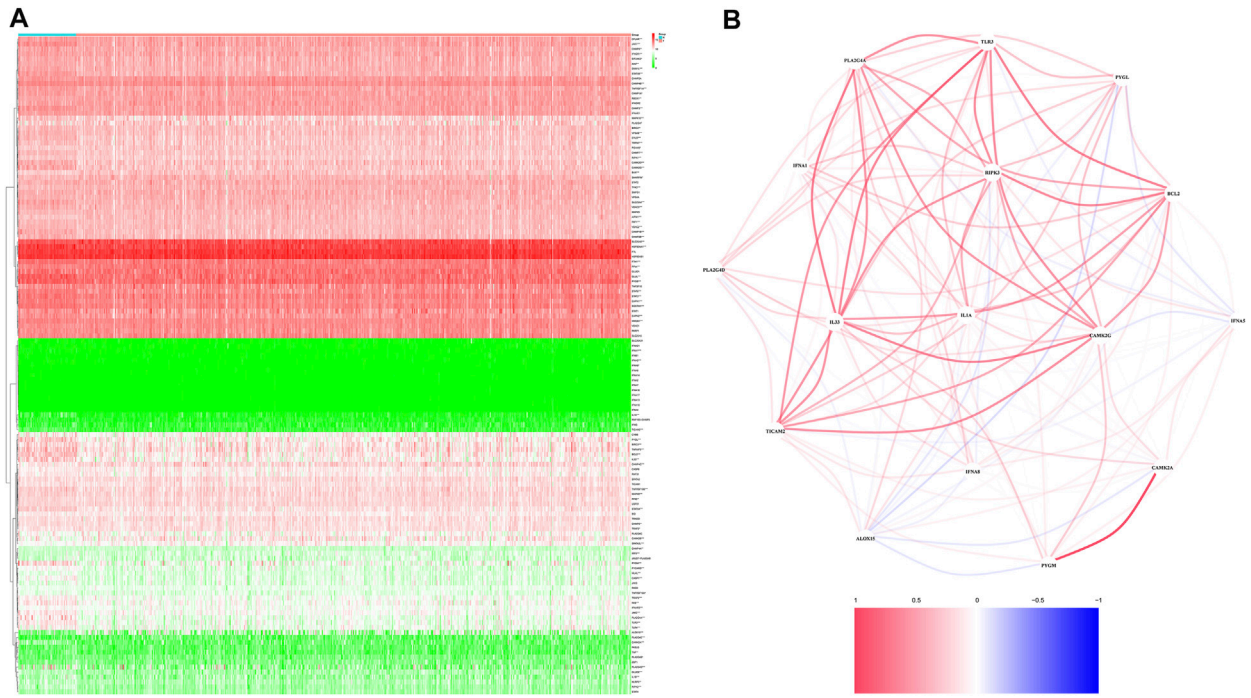


FIGURE 1 | Identification of the candidate genes. **(A)**. Heatmap showing differentially expressed genes between the two groups. **(B)**. The relationship among necroptosis-related DEGs.

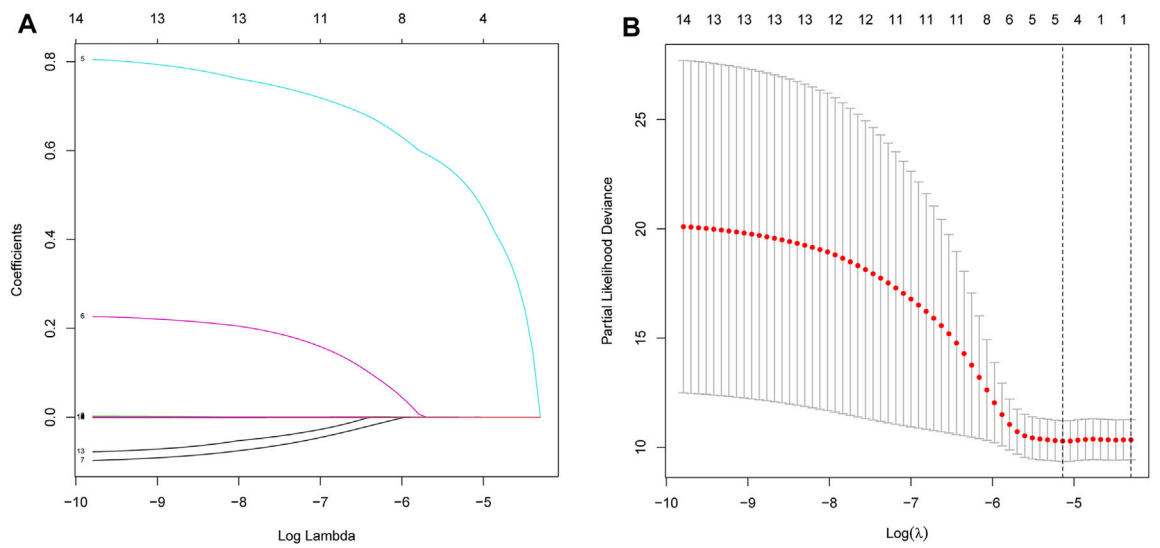


FIGURE 2 | Processes of LASSO Cox model fitting. **(A)**. Profiles of coefficients in the model at varying levels of penalization were plotted against the log(lambda) sequence. **(B)**. Ten-fold cross-validated error (the first and second vertical lines denote the minimum and cross-validated errors, respectively, within 1 standard error of the minimum).

for 1-, 3-, and 5-years survival, respectively (**Figure 3D**). The calibration curve was close to 45°, indicating that the model had good prognostic performance (**Figure 3E**).

qRT-PCR results showed that *BLC2*, *PYGL*, and *TLR3* were significantly downregulated in PRAD relative to normal adjacent tissues in the validation cohort (**Figure 4A**). Results from risk score

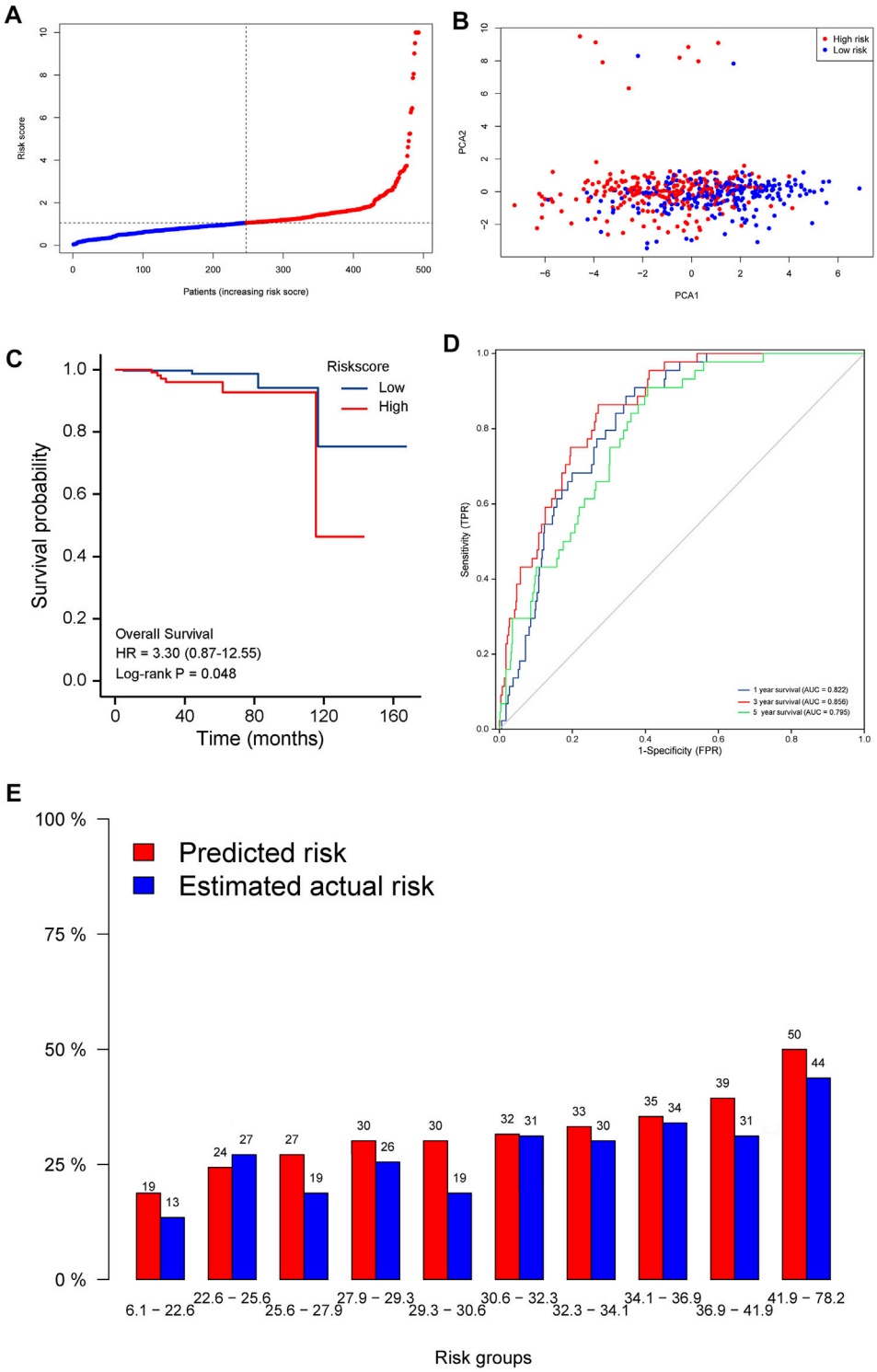


FIGURE 3 | Prognostic value of the 5-gene signature model in the test (TCGA) cohort. **(A)**. Distribution and median values of the risk scores. **(B)**. PCA plot. **(C)**. Kaplan-Meier curves showing OS of patients in the high-risk and low-risk groups. **(D)**. AUC values of the time-dependent ROC curves showing prognostic performance of the risk score. **(E)**. Calibration plots for the established model.

analysis for each patient, calculated using the same formula and critical values, indicated that patients in the high-risk group had significantly worse overall survival rates than those in the low-risk group ($p < 0.05$) (**Figure 4B**). AUC values for the 5 necroptosis-related gene signature were 0.836, 0.669, and 0 0.726 at 1, 3 and, 5 years, respectively (**Figure 4C**). **Figure 4D** shows the statistically

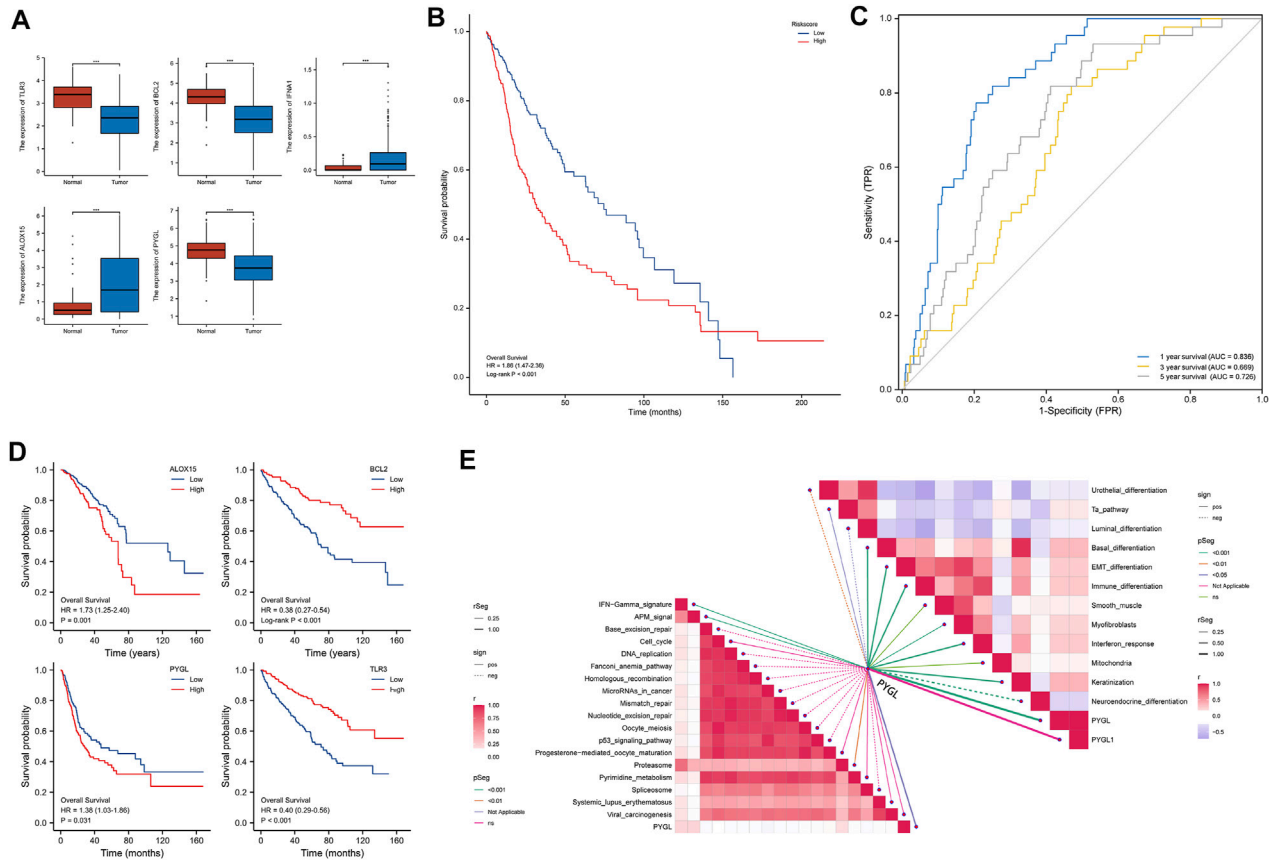


FIGURE 4 | Validation of the 5-gene signature using a validation cohort. **(A)**. Results of qRT-PCR analysis. **(B)**. Kaplan-Meier curves showing OS of patients in the high-risk and low-risk groups. **(C)**. AUCs of time-dependent ROC curves indicating the prognostic performance of the risk score. **(D)**. Kaplan-Meier curves showing OS of patients in the high- and low-expression groups of 5 DEGs. **(E)** Pathway analysis targeting PYGL.

significant results for the genes that make up the model. Multivariate cox regressionreveals that the most significant in 5 DEGs was PYGL. Based on these findings, we further explored the relationship between *PYGL* and multiple pathways, and found that *PYGL* expression was positively correlated with IFN−Gamma signature, APM signaling, Proteasome, Basal differentiation, EMT differentiation, Immune differentiation, Myofibroblasts, Interferon response, and Keratinization, but had a negative correlation with Urothelial differentiation, Luminal differentiation and Neuroendocrine differentiation (**Figure 4E**). Collectively, these results suggested that *PYGL* expression might be involved in cancer progression.

Independent Prognostic Value of the Risk Model

Next, we investigated the prognostic significance of different clinical features in PRAD patients. Results from univariate Cox proportional hazard regression analysis revealed that the risk score had prognostic significance in OS (HR = 1.67, 95% CI: 1.23–2.79, **Table 3**). These results remained significant (HR = 1.42, 95% CI: 1.14–1.75, **Table 3**) in PRAD patients. Validation of this signature's prognostic value revealed that it was an independent prognostic factor for PRAD (**Table 3**). A heatmap of DEGs and distribution of patients' age as well as survival status across both low- and high-risk subgroups are shown in **Supplementary Figure S1**.

High-Risk Scores Were Associated With a Hot Tumor Microenvironment

Previous studies have shown that the TME plays a crucial role in tumor extension, progression, migration and invasion (Chalmin et al., 2018). GO terms showed that the DEGs were significantly enriched in immune-related GO entries, including lymphocyte differentiation, and macrophage aggregation. Moreover, KEGG pathway enrichment results revealed that these genes were enriched in cell adhesion molecules, cytokine and cytokine receptor interactions, and necroptosis (**Figure 5A**). Overall, these results indicated that the DEGs were significantly enriched in immune-related activities, which implies that immune factors represent the main feature of the TME in PRAD. Furthermore, risk scores were positively correlated with the tumor microenvironment (ESTIMATEScore) (**Figure 5B**). Furthermore, multiple algorithms revealed that *PYGL* expression was strongly associated with macrophages, eosinophils and activated $CD4^+$ T cells

TABLE 3 | univariate cox and multivariate cox regression analyses.

Characteristics	Univariate analysis		Multivariate analysis	
	Hazard ratio (95% CI)	*p* value	Hazard ratio (95% CI)	*p* value
TCGA PRAD cohort				
Age (>60 vs. ≤60)	5.28 (3.48–6.86)	<0.001	4.69 (2.36–5.47)	<0.001
Race (Black or African American vs. Asian)	1.74 (0.254–5.84)	0.951	—	—
Riskscore (High vs. Low)	2.29 (1.63–3.65)	<0.001	1.67 (1.23–2.79)	<0.001
T (T1,T2vs. T3,T4)	0.37 (0.18–0.65)	0.041	0.46 (0.24–1.65)	0.21
N(N0vs.N1)	0.62 (0.42–0.81)	0.032	0.72 (0.54–0.89)	0.042
M(M0vs.M1)	0.76 (0.51–1.21)	0.63	—	—
Independent validation cohort				
Age (>60 vs<=60)	4.81 (2.96–5.91)	<0.001	3.51 (2.19–5.78)	0.025
T (T2vs. T3,T4)	0.77 (0.59–0.89)	0.03	0.82 (0.53–0.98)	0.046
N (N0vs.N1)	0.68 (0.31–0.87)	0.02	0.63 (0.25–0.86)	0.031
M (M0vs.M1)	0.87 (0.36–1.86)	0.75	—	—
Riskscore (High vs. Low)	1.87 (1.36–2.05)	<0.001	1.42 (1.14–1.75)	0.034

among others, indicating that the established prognostic model was associated with a hot TME (**Figure 6A**). ssGSEA results revealed significant enrichment of many types of immune cells, such as macrophages, and Tregs, among others, between the groups, while immune-related functions were also significantly different between the groups (**Figure 6B**).

DEG-Based Tumor Classification

Next, we performed K-means cluster analysis to stratify PRAD patients into subtypes according necroptosis-related DEGs, and obtained the highest intra- and low inter-group correlations when k = 2 (**Figure 7A**). Based on results of the 16 DEGs, the 499 PRAD patients were divided into 2 clusters. We generated a heatmap to illustrate the resulting gene expression profiles and found that they were matched to the clinical data (**Figure 7B**). A comparison of OS revealed significant differences between the clusters (**Figure 7C**).

Drug Sensitivity Analysis

The highest negative correlation score was obtained in Thioridazine (-0.703), a drug used for treatment of acute schizophrenia, mania and depression. This suggests that this drug has potential therapeutic effects in PRAD. On the other hand, the highest scores were obtained in trichostatin A (an antitumor inhibitor), LY-294002 (the first synthetic protein kinase inhibitor), Sirolimus (an immunosuppressant), Tanespimycin (an antitumor agent), Monorden (an antibiotic), Sirolimus (an immunosuppressant), Tanespimycin (an antitumor agent), and Monorden (an antibiotic). Estradiol (estradiol) is a transdermal estrogenic therapeutic agent that has been used to treat advanced PRAD. The above-mentioned drugs exhibited a strong negative correlation suggesting that they may have potential form treatment of PRAD (**Figure 8**).

DISCUSSION

Tumor growth occurs due to an imbalance between tumor cell death and growth (Meier et al., 2000). Notably, inhibition of excessive cell proliferation or normal cell death in the body markedly exacerbates incidence of malignant tumors. Therefore, some researchers believe that unlimited cell proliferation and death inhibition represent the two distinguishing features of malignant tumors. Necroptosis is a newly discovered form of cell death that morphologically manifests with similar characteristics to those observed in necrosis. However, the 2 phenomena differ in that while necrosis is a passive death caused by external physicochemical stresses, such as infection or inflammation, and is not regulated by signaling pathways, necroptosis is a regulated by programmed death. Numerous studies have demonstrated that necroptosis plays an important role in cancer initiation, progression, and metastasis (Jouan-Lanhouet et al., 2014; Barbosa et al., 2018). In the present study, we evaluated the relationship between necroptosis-related genes and prognosis of PRAD patients. Summarily, we screened for key genes that can independently predict prognosis of PRAD patients, then used them to construct a prognostic prediction model. Next, we verified the predictive power of the established model, then applied a multifactorial Cox regression model to assess the effect of other clinicopathological parameters on the signature's prognostic value in PRAD patients.

A total of 5 DEGs were included in the current model, with ROC and calibration curves in both training and validation cohorts showing that the model had excellent power in predicting PRAD patients. In addition, we selected *PYGL,* a gene located on chromosome 14q22.1 with a total of 20 exons that has been widely used as a building block for predictive models (Luo et al., 2020), and investigated its role in PRAD. Results showed that high *PYGL* expression was an independent predictor of poor prognosis in PRAD patients, consistent with a pervious study that reported similar findings in glioma patients (Luo et al., 2020). Our results further demonstrated that high *PYGL* expression was closely associated with infiltration of immune cells in tumors, with this expression pattern also positively correlated with cancer associated fibroblasts (CAFs). Previous studies have shown that CAFs inhibit T cell infiltration by secreting peritumoral collagen and TGF-Beta/PD-L1 specific

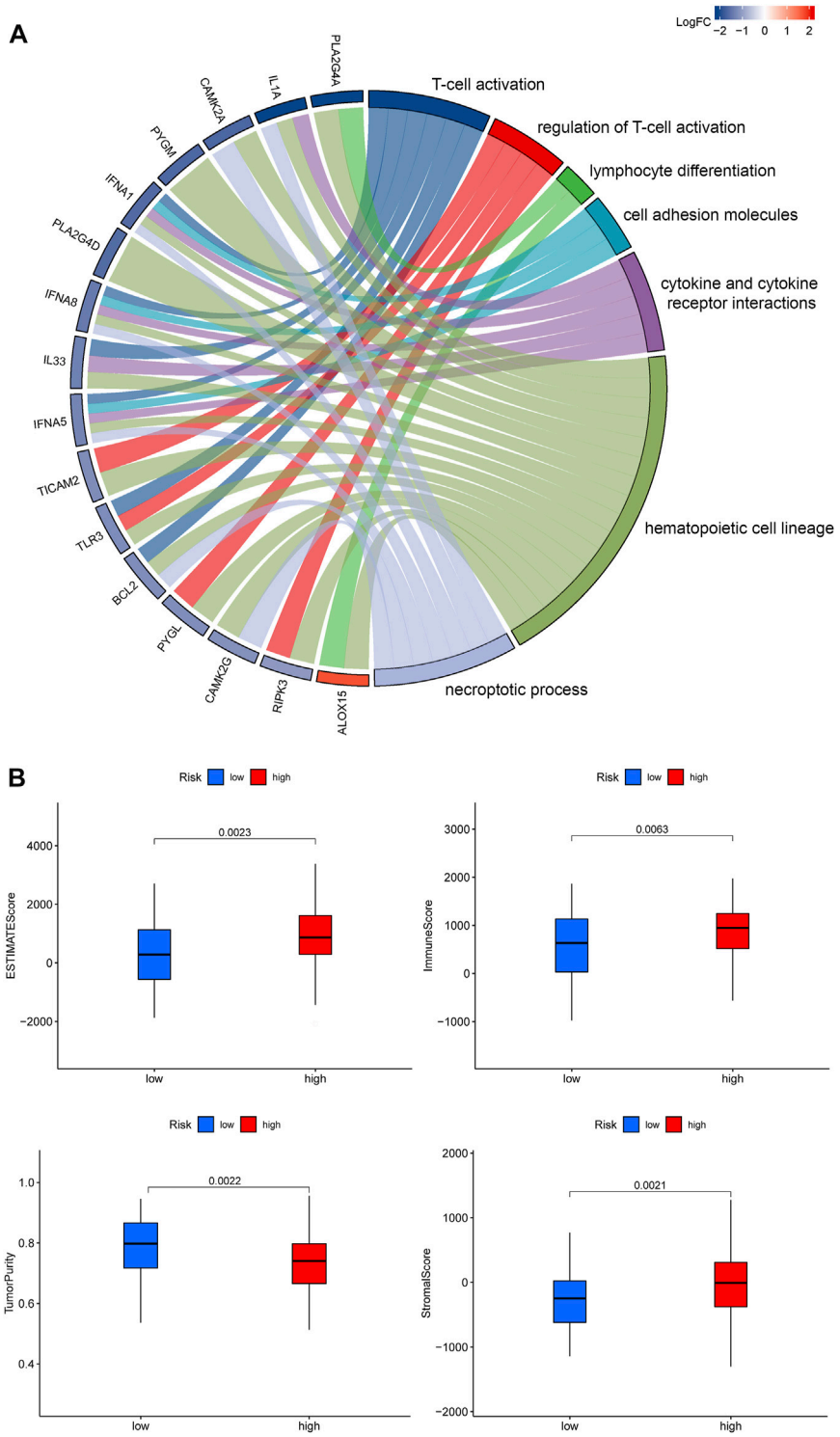


FIGURE 5 | (A). Profiles of GO functional terms and KEGG pathway enrichment for the identified DEGs **(B)**. An overview of the ESTIMATE algorithm.

antibody YM101, while M7824 could effectively suppress CAFs activity thereby promoting T cell infiltration (Lan et al., 2018; Yi et al., 2021a; Yi et al., 2021b). Additional research evidences have shown that the extracellular matrix produced by CAFs can also limit efficacy of tumor therapy (Ziani et al., 2018; Dou et al., 2020). Therefore, CAFs have become a new target for tumor therapy. For example, inhibition of tumor progression by targeting and regulating CAFs or overcoming their barrier

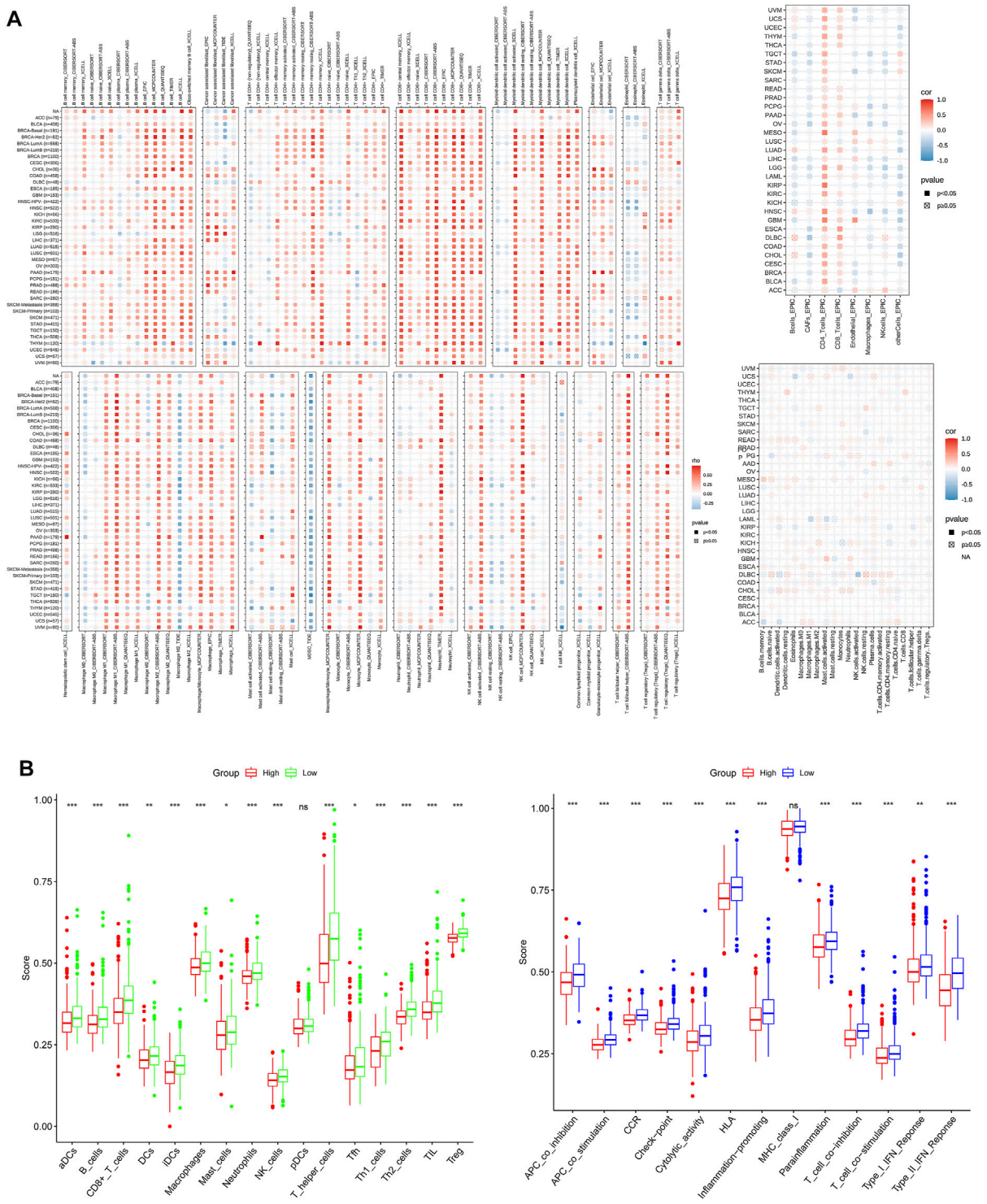


FIGURE 6 | (A) The relationship between PYGL expression and immune cell infiltration. **(B)**. Comparison of ssGSEA scores between the two risk groups in the test (TCGA) cohort.

effect has become a new approach for improving efficacy of tumor therapy (Ziani et al., 2018). Since only a handful of reports have described the specific function of *PYGL* in tumorigenesis of PRAD, results of the present study provide new insights into the relationship between necroptosis and PRAD based on *PYGL* functions.

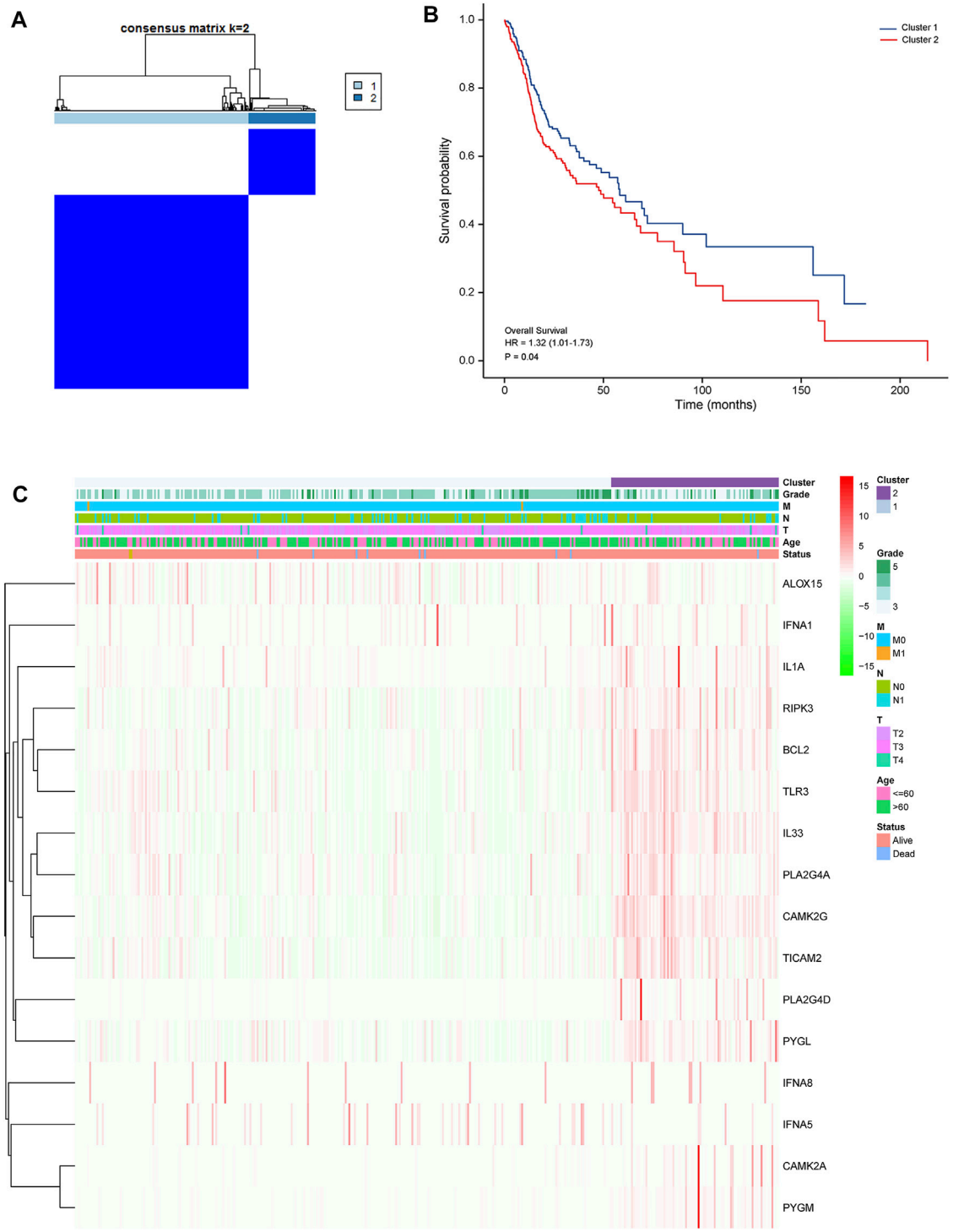


FIGURE 7 | Tumor classification based on the identified pyroptosis-related DEGs **(A)**. Patients were grouped into 2 clusters according to the consensus clustering matrix (k = 2). **(B)**. Kaplan–Meier curves showing OS of patients in the 2 clusters. **(C)**. A heatmap and clinicopathologic characteristics of patients in the 2 clusters based on the identified DEGs.

Previous studies have demonstrated the role of the tumor microenvironment, especially the immune microenvironment, in tumor biology (Sumida et al., 2011; Xu et al., 2020; Li et al., 2021). Results of the present study showed that the identified DEGs were mainly enriched in immune-related pathways. Consequently, we performed an immune cell infiltration analysis to assess the

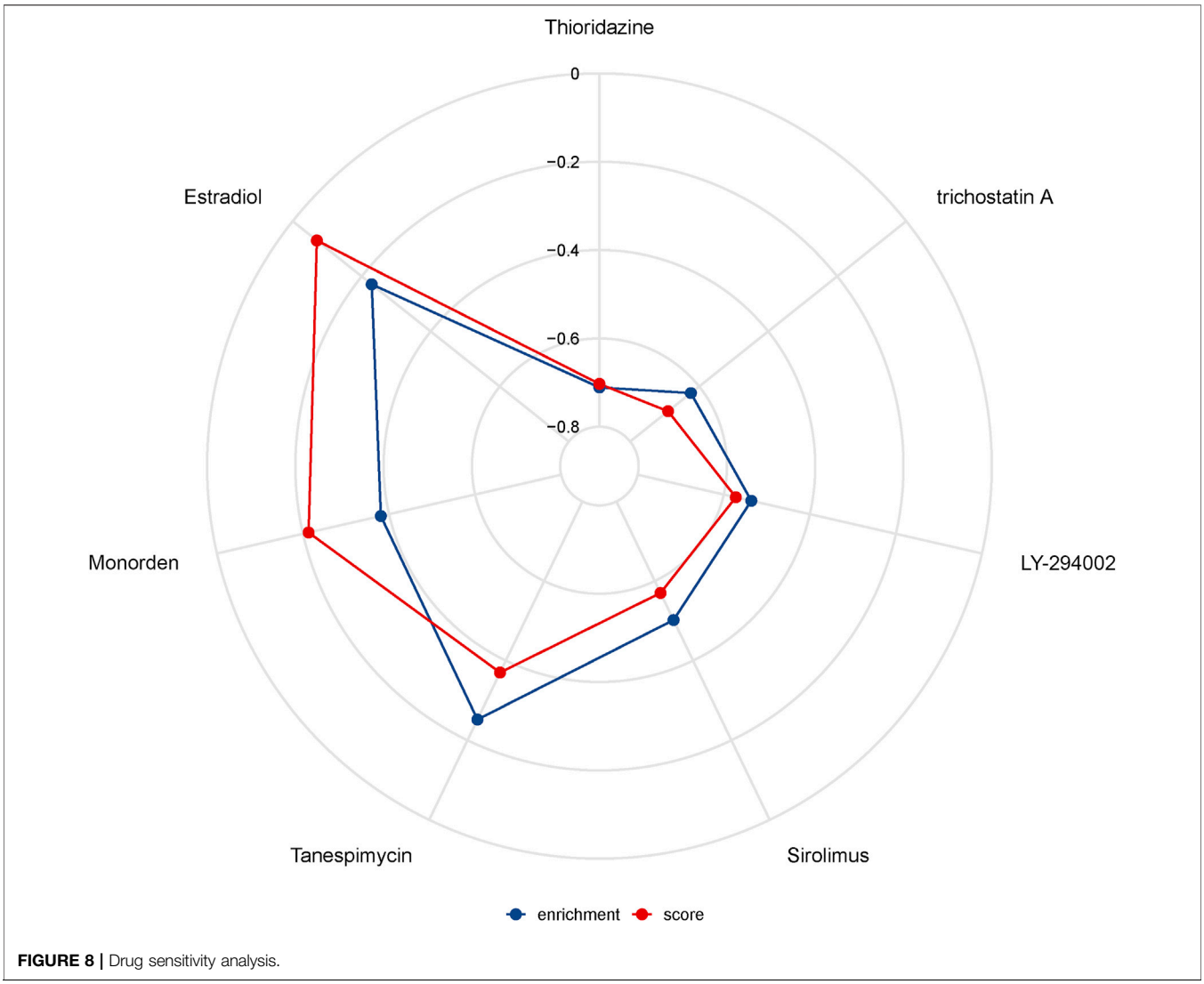


FIGURE 8 | Drug sensitivity analysis.

relationship between risk scores and overall survival of patients in the context of immune system response. Our results indicated that low-risk patients recruited more immune cells and triggered higher activation of immune pathways than their high-risk counterparts. Notably, the former group exhibited marked enrichment of NK cells. Previous studies have showed that NK cells have a strong cytocidal activity and do not require MHC activation to kill tumor cells (Shu and Cheng, 2020). In addition, tumors have been found to induce production of type II NKT cells, which in turn secrets IL-13, thereby not only causing aggregation of MDSC in the tumor microenvironment but also activating the STAT6 signaling pathway and suppressing $CD8^+$ T cell function (Terabe et al., 2000; Terabe et al., 2005). Consequently, the use of NK cells for tumor immunotherapy has gained numerous attention from all over the world. Some of the current NK cell-based clinical approaches for tumor therapy include cytokine-mediated NK cell activation, autologous or allogeneic NK cell transfer, CAR gene-modified NK cells and memory cells. In addition to hematological tumors, modified NK and memory-like NK cells have shown great potential for treatment of liver, non-small cell lung, colorectal, ovarian and breast cancers (Li et al., 2015). Results from our drug sensitivity analysis revealed several potential drugs that might modulate the necroptosis-related genes. Notably, expression of these necroptosis-related genes was negatively correlated with thioridazine, trichostatin A, LY-294002, sirolimus, tanespimycin and monorden, suggesting that these could be potential new options as therapeutic drugs.

CONCLUSION

In summary, we identified differentially expressed necroptosis-related genes, between PRAD and normal adjacent tissues, and used them to establish a model for predicting prognosis of PRAD patients. Moreover, we revealed the correlation between risk scores and immune activities. However, further studies are needed to elucidate the mechanisms underlying necroptosis in tumor immunology.

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 All patients signed a written informed consent prior to inclusion in the study, and ethical approval for the study protocol was obtained from The First Affiliated Hospital of Zhengzhou University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

X-TY, J-XY, L-XS designed experiments; L-YZ, X-YL, carried out experiments, and wrote the manuscript, X-TY, L-XS performed manuscript review. All authors read and approved the final manuscript.

FUNDING

This study received Fundamental research program funding of Ninth People's Hospital affiliated to Shanghai Jiao Tong university School of Medicine (No. JYZZ076), Clinical Research Program of Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine (No. JYLJ201801, JYLJ201911), the China Postdoctoral Science Foundation (No. 2017M611585) and the National Natural Science Foundation of China (No. 81871458).

ACKNOWLEDGMENTS

Home for Researchers editorial team (www.home-for-researchers.com).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2021.814813/full#supplementary-material

REFERENCES

Barbosa, L. A., Fiuza, P. P., Borges, L. J., Rolim, F. A., Andrade, M. B., Luz, N. F., et al. (2018). RIPK1-RIPK3-MLKL-Associated Necroptosis Drives Leishmania Infantum Killing in Neutrophils. *Front. Immunol.* 9, 1818. doi:10.3389/fimmu.2018.01818

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: a Cancer J. clinicians* 68 (6), 394–424. doi:10.3322/caac.21492

Chalmin, F., Bruchard, M., Vegran, F., and Ghiringhelli, F. (2018). Regulation of T Cell Antitumor Immune Response by Tumor Induced Metabolic Stress. *Cst* 3 (1), 9–18. doi:10.15698/cst2019.01.171

Degterev, A., Huang, Z., Boyce, M., Li, Y., Jagtap, P., Mizushima, N., et al. (2005). Chemical Inhibitor of Nonapoptotic Cell Death with Therapeutic Potential for Ischemic Brain Injury. *Nat. Chem. Biol.* 1 (2), 112–119. doi:10.1038/nchembio711

Dou, D., Ren, X., Han, M., Xu, X., Ge, X., Gu, Y., et al. (2020). Cancer-Associated Fibroblasts-Derived Exosomes Suppress Immune Cell Function in Breast Cancer via the miR-92/pd-L1 Pathway. *Front. Immunol.* 11, 2026. doi:10.3389/fimmu.2020.02026

Fan, H., Liu, F., Dong, G., Ren, D., Xu, Y., Dou, J., et al. (2014). Activation-induced Necroptosis Contributes to B-Cell Lymphopenia in Active Systemic Lupus Erythematosus. *Cell Death Dis* 5 (9), e1416. doi:10.1038/cddis.2014.375

Frank, D., and Vince, J. E. (2019). Pyroptosis versus Necroptosis: Similarities, Differences, and Crosstalk. *Cell Death Differ* 26 (1), 99–114. doi:10.1038/s41418-018-0212-6

Gong, Y., Fan, Z., Luo, G., Yang, C., Huang, Q., Fan, K., et al. (2019). The Role of Necroptosis in Cancer Biology and Therapy. *Mol. Cancer* 18 (1), 100. doi:10.1186/s12943-019-1029-8

Jouan-Lanhouet, S., Riquet, F., Duprez, L., Vanden Berghe, T., Takahashi, N., and Vandenabeele, P. (2014). Necroptosis, *In Vivo* Detection in Experimental Disease Models. *Semin. Cell Dev. Biol.* 35, 2–13. doi:10.1016/j.semcdb.2014.08.010

Lan, Y., Zhang, D., Xu, C., Hance, K. W., Marelli, B., Qi, J., et al. (2018). Enhanced Preclinical Antitumor Activity of M7824, a Bifunctional Fusion Protein Simultaneously Targeting PD-L1 and TGF-β. *Sci. Transl. Med.* 10 (424), 1. doi:10.1126/scitranslmed.aan5488

Li, X.-Y., Zhang, L.-Y., Li, X.-Y., Yang, X.-T., and Su, L.-X. (2021). A Pyroptosis-Related Gene Signature for Predicting Survival in Glioblastoma. *Front. Oncol.* 11, 697198. doi:10.3389/fonc.2021.697198

Li, Y., Yin, J., Li, T., Huang, S., Yan, H., Leavenworth, J., et al. (2015). NK Cell-Based Cancer Immunotherapy: from Basic Biology to Clinical Application. *Sci. China Life Sci.* 58 (12), 1233–1245. doi:10.1007/s11427-015-4970-9

Luo, X., Hu, J., Gao, X., Fan, Y., Sun, Y., Gu, X., et al. (2020). Novel PYGL Mutations in Chinese Children Leading to Glycogen Storage Disease Type VI: Two Case Reports. *BMC Med. Genet.* 21 (1), 74. doi:10.1186/s12881-020-01010-4

Meier, P., Silke, J., Leevers, S. J., and Evan, G. I. (2000). The Drosophila Caspase DRONC Is Regulated by DIAP1. *Embo J.* 19 (4), 598–611. doi:10.1093/emboj/19.4.598

Shu, Y., and Cheng, P. (2020). Targeting Tumor-Associated Macrophages for Cancer Immunotherapy. *Biochim. Biophys. Acta (Bba) - Rev. Cancer* 1874 (2), 188434. doi:10.1016/j.bbcan.2020.188434

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

Sumida, T., Kitadai, Y., Shinagawa, K., Tanaka, M., Kodama, M., Ohnishi, M., et al. (2011). Anti-stromal Therapy with Imatinib Inhibits Growth and Metastasis of Gastric Carcinoma in an Orthotopic Nude Mouse Model. *Int. J. Cancer* 128 (9), 2050–2062. doi:10.1002/ijc.25812

Tang, R., Xu, J., Zhang, B., Liu, J., Liang, C., Hua, J., et al. (2020). Ferroptosis, Necroptosis, and Pyroptosis in Anticancer Immunity. *J. Hematol. Oncol.* 13 (1), 110. doi:10.1186/s13045-020-00946-7

Terabe, M., Matsui, S., Noben-Trauth, N., Chen, H., Watson, C., Donaldson, D. D., et al. (2000). NKT Cell-Mediated Repression of Tumor Immunosurveillance by IL-13 and the IL-4R-STAT6 Pathway. *Nat. Immunol.* 1 (6), 515–520. doi:10.1038/82771

Terabe, M., Swann, J., Ambrosino, E., Sinha, P., Takaku, S., Hayakawa, Y., et al. (2005). A Nonclassical Non-vα14jα18 CD1d-Restricted (Type II) NKT Cell Is Sufficient for Down-Regulation of Tumor Immunosurveillance. *J. Exp. Med.* 202 (12), 1627–1633. doi:10.1084/jem.20051381

Wei, W., Zeng, H., Zheng, R., Zhang, S., An, L., Chen, R., et al. (2020). Cancer Registration in China and its Role in Cancer Prevention and Control. *Lancet Oncol.* 21 (7), e342–e349. doi:10.1016/S1470-2045(20)30073-5

Xu, D., Wang, Y., Zhou, K., Wu, J., Zhang, Z., Zhang, J., et al. (2020). Identification of an Extracellular Vesicle-Related Gene Signature in the Prediction of Pancreatic Cancer Clinical Prognosis. *Biosci. Rep.* 40 (12), R20201087. doi:10.1042/BSR20201087

Yi, M., Niu, M., Zhang, J., Li, S., Zhu, S., Yan, Y., et al. (2021). Combine and Conquer: Manganese Synergizing Anti-TGF-β/pd-L1 Bispecific Antibody YM101 to Overcome Immunotherapy Resistance in Non-inflamed Cancers. *J. Hematol. Oncol.* 14 (1), 146. doi:10.1186/s13045-021-01155-6

Yi, M., Zhang, J., Li, A., Niu, M., Yan, Y., Jiao, Y., et al. (2021). The Construction, Expression, and Enhanced Anti-tumor Activity of YM101: a Bispecific Antibody Simultaneously Targeting TGF-β and PD-L1. *J. Hematol. Oncol.* 14 (1), 27. doi:10.1186/s13045-021-01045-x

Yuan, J., Amin, P., and Ofengeim, D. (2019). Necroptosis and RIPK1-Mediated Neuroinflammation in CNS Diseases. *Nat. Rev. Neurosci.* 20 (1), 19–33. doi:10.1038/s41583-018-0093-1

Zeng, D., Ye, Z., Shen, R., Yu, G., Wu, J., Xiong, Y., et al. (2021). IOBR: Multi-Omics Immuno-Oncology Biological Research to Decode Tumor Microenvironment and Signatures. *Front. Immunol.* 12, 687975. doi:10.3389/fimmu.2021.687975

Ziani, L., Chouaib, S., and Thiery, J. (2018). Alteration of the Antitumor Immune Response by Cancer-Associated Fibroblasts. *Front. Immunol.* 9, 414. doi:10.3389/fimmu.2018.00414

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

frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 11 February 2022
doi: 10.3389/fbioe.2022.790605

Clinical Significance of TET2 in Female Cancers

*Fang Wan, Fangfang Chen, Yangfan Fan and Deqin Chen**

Department of Surgery, The Women's Hospital, School of Medicine, Zhejiang University, Hangzhou, China

Female cancers refer to malignant tumors of the female reproductive system and breasts, which severely affect the physical and mental health of women. Although emerging experiment-based studies have indicated a potential correlation between ten-eleven translocation methylcytosine dioxygenase (TET2) and female cancers, no comprehensive studies have been conducted. Therefore, this study aimed to summarize the clinical value and underlying oncogenic functions of TET2 in female cancers, such as breast invasive carcinoma (BRCA), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), ovarian serous cystadenocarcinoma (OV), uterine corpus endometrial carcinoma (UCEC), and uterine carcinosarcoma (UCS), based on the data obtained from The Cancer Genome Atlas. The expression of TET2 was decreased in most female cancers, and its high expression was distinctly associated with the favorable prognosis of most female cancers. Furthermore, $CD8^+$ T-cell infiltration was not correlated with TET2 in OV, UCEC, and UCS, whereas tumor-associated fibroblast infiltration was significantly correlated with TET2 in BRCA, CESC, and OV. TET2 was co-expressed with the immune checkpoint molecules ADORA2A, CD160, CD200, CD200R1, CD44, CD80, NRP1 TNFSF4, and TNFSF15 in most female cancers. Enrichment analysis revealed that some signaling pathways involving TET2 and related genes were related to tumorigenesis. Immunohistochemical and immunofluorescence staining confirmed the results of cancer immune infiltration analysis in BRCA tissues. Therefore, this study provides evidence for the oncogenic functions and clinical value of TET2 in female cancers.

Keywords: TET2, female cancers, immune infiltration, TCGA, GEO

OPEN ACCESS

Edited by:
Ping Zhang,
Griffith University, Australia

Reviewed by:
Abhinita S. Mohanty,
Memorial Sloan Kettering Cancer Center, United States
Weida Gong,
University of North Carolina at Chapel Hill, United States

****Correspondence:***
Deqin Chen
5515001@zju.edu.cn

Specialty section:
This article was submitted to Preclinical Cell and Gene Therapy, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *07 October 2021*
Accepted: *05 January 2022*
Published: *11 February 2022*

Citation:
Wan F, Chen F, Fan Y and Chen D (2022) Clinical Significance of TET2 in Female Cancers. Front. Bioeng. Biotechnol. 10:790605. doi: 10.3389/fbioe.2022.790605

INTRODUCTION

Female cancers refer to malignant tumors of the female reproductive system and breasts, which seriously threaten the physical and mental health of women. The incidence of breast, cervical, uterine body, and ovarian cancers is gradually increasing, and ranking at the forefront of the incidence of female malignant tumors (Chen et al., 2014). Among these cancers, breast cancer is a major malignant tumor threatening the health of women worldwide. In 2018, more than 2 million new cases of breast cancers were reported worldwide, with the incidence and mortality of breast cancer ranking first among all female cancers (Ferlay et al., 2019). In addition, the incidence of cervical cancer is increasing in the younger population (Lin et al., 2021). Therefore, effective diagnosis and treatment of female cancers are of great clinical significance worldwide.

DNA methylation is a widely recognized epigenetic modification. In mammals, the addition of guanine (linear dinucleotide) (CpG) to the promoter region can inhibit the activity of many genes and promote stable heredity. However, the underlying mechanisms of active DNA demethylation remain unclear, which is a fundamental question in epigenetics research (Greenberg and Bourc'his,

2019). In 2009, TET protein was found to oxidize 5-methylcytosine (5 mC) to 5-hydroxymethylcytosine (5 hmC) *in vitro*, and since then, the Tet methylcytosine dioxygenase (TET2) protein family has been extensively investigated. TET2 protein is one of the evolutionarily highly conserved members of the Tet family. Several studies have demonstrated that TET2 protein can be used as an invertase to convert 5 mC to 5 hmC (Chowdhury et al., 2014). Furthermore, TET2 can also convert 5 mC to 5-formylcytosine (5 fC) and 5-carboxylcytosine (5 caC) through continuous oxidation reaction (Cadet and Wagner, 2014; Zhang et al., 2014). Although the underlying mechanisms of dynamic regulation of 5 mC and its active and passive demethylation processes remain unclear, the discovery of new enzymatic activities of TET2 enhances the understanding of its potential mechanisms. For example, DNA methyltransferase 1, an enzyme used to maintain DNA methylation, does not recognize 5 hmC, and the transformation of 5 mC to 5 hmC may result in replication-dependent DNA-passive demethylation. Oxidative derivatives of 5 hmC may be involved in active DNA demethylation that is not dependent on replication. Thymine DNA glycosylase (TDG) can shear 5fC or 5caC on the CpG island (TDG has minimal activity for 5 hmC), and the resulting baseless sites are then repaired through the base excision repair (BER) pathway, producing unmethylated cytosine (Zhang et al., 2013). Owing to its role in promoting hematopoietic stem cell self-renewal, cell line typification, and monocyte differentiation during hematopoietic processes, TET2 protein has been mainly investigated in hematologic malignancies, such as myeloproliferative disease and myelodysplastic syndromes (acute leukemia) (Yan et al., 2017; Yue and Rao, 2020). According to recent studies, TET2 has been closely associated with solid tumors, and plays an important role in tumor occurrence and development.

With the advancement of high-throughput and high-resolution sequencing technologies, the open-access The Cancer Genome Atlas (TCGA) (Ganini et al., 2021) and Gene Expression Omnibus (GEO) (Clough and Barrett, 2016) databases provide functional genomic data of tumorigenesis in various cancers. Therefore, combined analyses of a single gene in multiple tumors can be performed. In this study, a conjoint analysis of TET2 was performed using data from TCGA and GEO to assess the following aspects: transcriptional level, clinical outcomes, DNA methylation, genetic mutations, cancer immune infiltration, and related signaling pathways. In addition, tumor tissue validation was performed to consolidate the results. Therefore, this study aimed to comprehensively analyze the clinical significance of TET2 in major female cancers, which, to the best of our knowledge, has never been reported previously.

MATERIALS AND METHODS

Gene Expression Analysis

Data on TET2 expression in normal and tumor cell types and tissues were obtained from the Human Protein Atlas (HPA) (Godin and Eichler, 2017). High specificity was defined as a normalized expression of ≥1 in at least one type of tissue or cell, without the elevation of expression in any tissue or cell type. The major female cancers included for analysis were breast invasive carcinoma (BRCA), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), ovarian serous cystadenocarcinoma (OV), uterine corpus endometrial carcinoma (UCEC), and uterine carcinosarcoma (UCS), which were included in TCGA database. Subsequently, the TET2 expression levels in various types and subtypes of cancers in TCGA database and healthy controls were obtained from the Tumor Immune Estimation Resource version 2 (TIMER2, timer. cistrome.org) database (Li et al., 2017; Li et al., 2020). Gene Expression Profiling Interactive Analysis version 2 (GEPIA2, gepia2. cancer-pku.cn) was used for expression analysis using box plots for cancer types with limited information of healthy controls (Tang et al., 2019). Furthermore, TET2 expression in different cancer stages was summarized via violin plots using GEPIA2. The transformed transcriptional levels were calculated as log2 of transcripts per million (TPM) +1 and were used in box and violin plots. A Sanguini diagram was created using the R package "ggalluval".

The data of phosphorylated protein levels from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) (Edwards et al., 2015) were analyzed using the UALCAN portal (ualcan.path.uab.edu/analysis-prot.html) (Chandrashekar et al., 2017). The expression level ssof phosphoprotein (with phosphorylation at the S99 and S38 loci) of TET2 (NP_001120680) (with the absence of data of total protein) was compared between the primary tumor and normal tissues. The correlation between TET2 phosphoprotein and disease stage and between the age of patients and histological features of tumors was further analyzed. The available datasets of two tumors, BRCA and UCEC, were selected. The PhosphoNET database (phosphonet.ca) was used to analyze CPTAC-identified TET2 phosphorylation, indicating that various parameters should be considered while selecting putative P-sites. Confirmed P-sites have lower hydrophobicity scores. A lower P-site similarity score typically resembles confirmed corresponding P-Ser, P-Thr, or P-Tyr sites. The maximum KInase score provides the calculated score for the highest match of 500 human protein kinases for amino acid sequence surrounding the target P-site as determined using the kinase substrate predictor V2. The total KInase score provides the sum of the positive individual kinase substrate predictor V2 scores from 500 human protein kinases. The conservation score is defined as the average percentage similarity between the human P-site and equivalent P-sites of 20 other diverse species. This observation warrants further molecular studies to explore the potential role of S38 phosphorylation in tumorigenesis.

Survival and Prognosis Analyses

The association of overall survival (OS) and disease-free survival (DFS) with TET2 in various TCGA cancers was assessed using the survival map section of GEPIA2, and the threshold for different expression levels was set as high (50%) and low (50%). The hypothesis was validated using the log-rank test. In addition, OS, distant metastasis-free survival (DMFS), relapse-free survival

(RFS), post-progression survival (PPS), first progression (FP), disease-specific survival (DSS), and progression-free survival (PFS) were assessed using the interactive operation interface of the KM plotter (kmplot.com) using data extracted from the GEO database (BRCA and OV data were available [**Supplementary Table 1**]). Multiple clinical variables were included in the prognostic analysis, such as immunohistochemical (IHC) staining of tumor tissues, lymph node involvement, pathological grading, TP53 status, and treatment. The median was used to analyze the association between BRCA and OV. The hazard ratio (HR), 95% confidence intervals (CIs), and log-rank *p*-values were computed, and the Kaplan–Meier survival plots were generated.

The pooled analysis (univariate) of OS, disease-free interval (DFI), progression-free interval (PFI), and DSS was conducted for all tumors. A forest plot was drawn without merging the HRs. In addition, a receiver operator characteristic curve (ROC) was generated to determine whether TET2 expression could accurately predict the 1-, 3-, and 5-years OS in selected types of tumors in TCGA. The best cut-off value of TET2 expression was calculated, and patients from each tumor type were divided into the high- and low-expression groups based on the cut-off value. Survival analysis was performed to validate OS between groups stratified according to the cut-off value.

Genetic Alteration and DNA Methylation Analysis

The genetic alteration features of TET2 were analyzed using TCGA pan-cancer data from cBioPortal for Cancer Genomics (cbioportal.org) (Gao et al., 2013). Alteration frequencies, mutation types, and copy number alterations in selected types of TCGA cancers were analyzed using the cancer-type summary. In addition, mutation sites and three-dimensional TET2 structures were analyzed and represented in a schematic illustration. Differences in TET2 genetic alteration-associated OS, DFS, and PFS among various TCGA cancers were analyzed under the comparison section. Furthermore, KM plots with *p*-values were generated based on the log-rank test, and the association between different DNA methylation probes of gene susceptibility and TET2 expression was assessed using MEXPRESS (mexpress.be) (Koch et al., 2019; Koch et al., 2015). In addition, the beta value of each sample, *p*-value (adjusted using the Benjamini–Hochberg procedure), and Pearson correlation coefficients (R) were evaluated. MethSurv (Modhukur et al., 2018) (https://biit.cs.ut.ee/methsurv/) and SurvivalMeth (Zhang et al., 2021) (http://bio-bigdata.hrbmu.edu.cn/survivalmeth/) were used to analyze the prognostic significance of single CpG methylation in TET2 in patients with cancer.

Immune Infiltration Analysis

The association of TET2 levels with $CD8^+$ T-cell infiltration and cancer-associated fibroblasts in selected types of TCGA cancers was analyzed using the Immune-Gene section of TIMER2. To estimate immune infiltration, algorithms such as TIMER, CIBERSORT, CIBERSORT-ABS, QUANTISEQ, XCELL, MCPCOUNTER, and EPIC were used. Both *p* and partial correlation (*r*) values were evaluated using the purity-adjusted Spearman's rank correlation test. All results were presented in heatmaps and scatter plots.

Cancer Immune Analysis

The potential association of TET2 expression with tumor mutational burden (TMB), microsatellite instability (MSI), checkpoint expression, number of neoantigens, tumor microenvironment (ESTIMATE algorithm), and immune response pathways in cancers was analyzed using Sangerbox (sangerbox.com/Tool). The Spearman's rank correlation test was performed, and *p* and partial correlation (*r*) values were evaluated.

TET2-Related Gene Enrichment Analysis

The STRING website (string-db.org) was used for TET2-related gene enrichment analysis, including the following parameters: organism, "*Homo sapiens*"; minimum required interaction score, "low confidence (0.150)"; meaning of network edges, "evidence"; maximum number of interactors to show, "no more than 50 interactors in the first shell", and active interaction sources, "experiments". As a result, proteins that bound to or interacted with TET2 were screened out based on published experimentally confirmed data. Moreover, the top 100 genes associated with TET2 were obtained using the similar gene detection section of GEPIA2 based on the data of selected cancers and healthy controls. Pearson's correlation coefficient was used to assess the correlation between TET2 and potential TET2-related genes. The log2 of TPM was used to analyze dot plots, and *p* and r values were calculated. In addition, a heatmap representing related genes was generated using the Gene-Corr section of TIMER2, with r and *p* values calculated using the purity-adjusted Spearman's rank correlation test. Intersection analyses were conducted to compare TET2-related genes using Jvenn (Bardou et al., 2014). Thereafter, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was conducted using combined results of previous analyses. The potential TET2-related genes were analyzed using the Database for Annotation, Visualization, and Integrated Discovery (Dennis et al., 2003). The enriched signaling pathways were listed in "tidyr" (cran.rproject org/web/packages/tidyr) and "ggplot2" (cran.r-project.org/web/packages/ggplot2) of the R package. The "clusterProfiler" R package (http://www.bioconductor.org/packages/release/bioc/html/clusterProfiler) was used for gene ontology (GO) enrichment analysis. The results of biological processes, cellular components, and molecular functions were represented in cnetplots. The R language software (R-4.0.4, 64-bit) (www.r-project.org) was used for analysis, and *p*-values < 0.05 were set as statistical significance for two-tailed analyses.

Histological Analysis

This study was approved by the ethics committee of The Women's Hospital, School of Medicine, Zhejiang University. TET2 was detected in formalin-fixed paraffin-embedded (FFPE) BRCA tissues and paired paracancerous tissues using IHC staining, which was performed as previously described

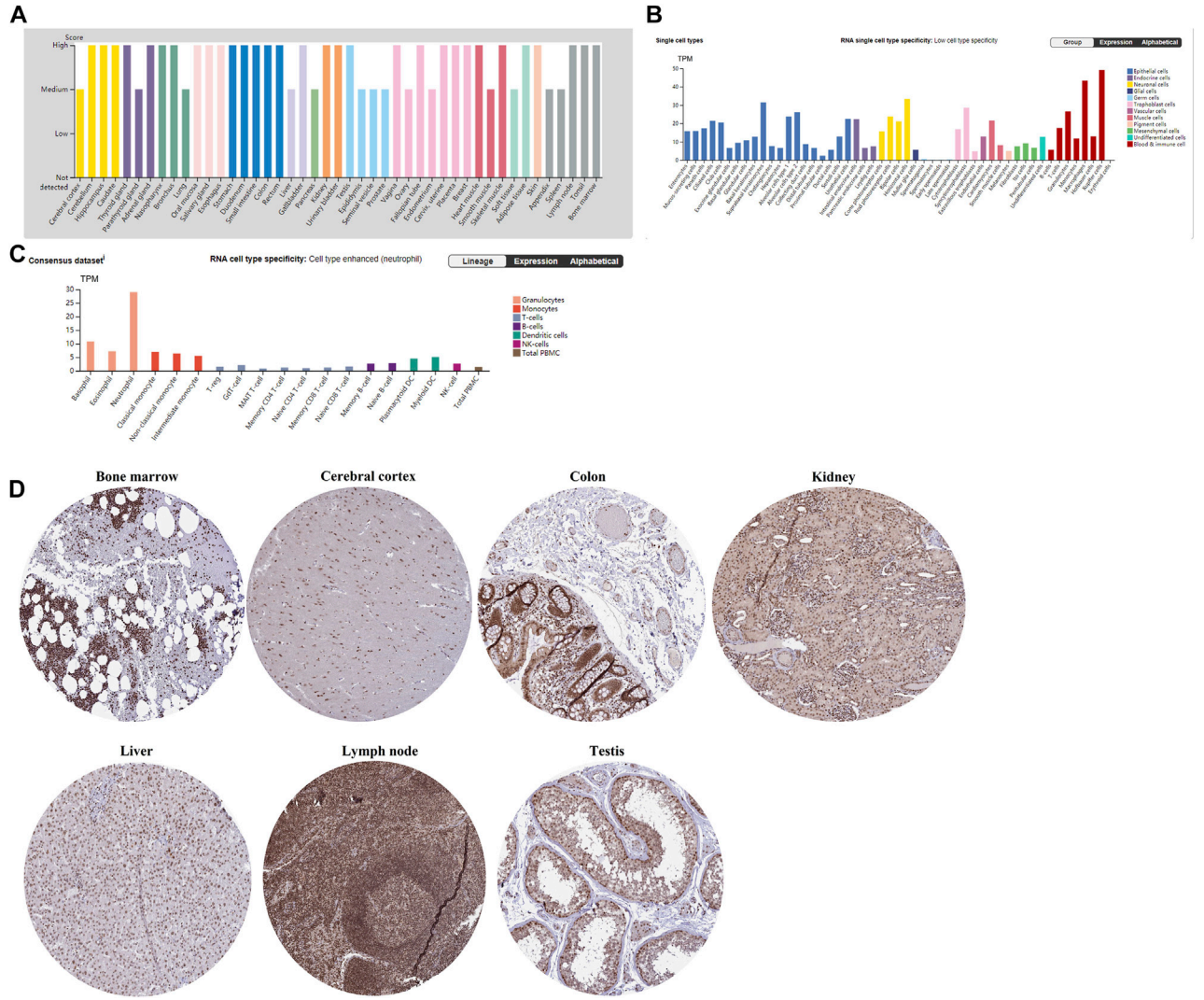


FIGURE 1 | (A) TET2 transcriptional levels in different normal tissues analyzed using HPA, GTEx, and FANTOM5. **(B)** TET2 transcriptional levels in different normal cells analyzed using HPA, Monaco, and Schmiedel. **(C)** TET2 expression in immune cells at the single-cell resolution extracted from the HPA. **(D)** IHC staining reveals the protein expression of TET2 in normal tissues (mainly located in the nucleus, cited from the HPA).

(Tsutsumi, 2021). Three pairs of tissues were examined. Briefly, the tumor tissues were cut into 4-mm-thick sections, dewaxed in xylene, and rehydrated in a graded series of alcohols. The antigen was retrieved by heating the tissue sections in the EDTA solution (1 mM, pH 9.0) at 100°C for 30 min. Cooled tissue sections were immersed in 0.3% hydrogen peroxide for 15 min to block the endogenous peroxidase activity, rinsed with phosphate-buffered saline for 5 min, and blocked with 3% bovine serum albumin at room temperature for 30 min. Subsequently, the sections were incubated with mouse monoclonal antibody against human TET2 (ab243323) (1:30) at 4°C overnight, followed by incubation with HRP-conjugated goat anti-rabbit secondary antibody. Diaminobenzene and hematoxylin were used as a chromogenic substrate and nuclear counterstain, respectively, and representative images were captured. To verify the correlation between TET2 expression and checkpoints in BRCA, immunofluorescence (IF) staining was performed in FFPE BRCA tissues. CD276 (B7-H3), LAG3, and PDCD1 were selected for verification. Briefly, the tissue sections were incubated with mouse monoclonal antibody against human TET2 (10 μg/ml), rabbit monoclonal antibody against CD276 (ab134161) (1 μg/ml), rabbit monoclonal antibody against LAG3 (ab209236) (1: 100), and rabbit monoclonal antibody against PDCD1 (ab237728) (1:50) at 4°C overnight, followed by incubation with Alexa®488-conjugated goat anti-mouse secondary antibody or Alexa®549-conjugated goat anti-rabbit secondary antibody (Thermo Fisher Scientific, CA, United States). The nuclear stain Hoechst 34,580 (5 μg/ml; Molecular Probes, Thermo Fisher Scientific, CA, United States) was added before washing the incubated tissues. Finally, the sections were dehydrated, cleared, and mounted using a confocal

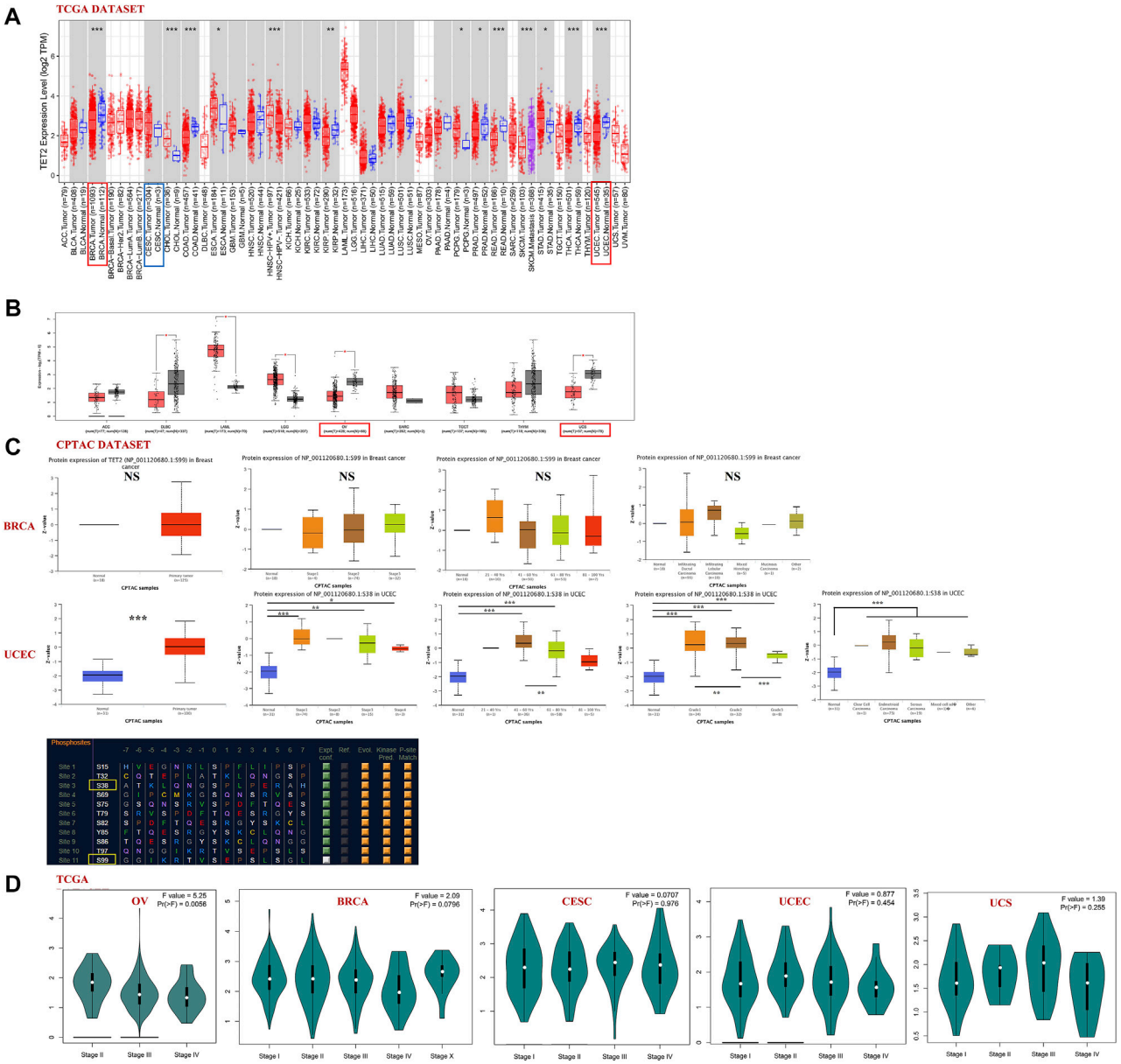


FIGURE 2 | (A,B) Analysis of TET2 expression in various types or subtypes of cancers and normal tissues using TIMER2 and GEPIA2. **(C)** Analysis of phosphorylated protein expression of TET2 in breast cancer (S99 site) and uterine corpus endometrial carcinoma (S38 site) and their corresponding normal tissues combined with clinical parameters using CPTAC, **(D)** TET2 expression in different pathological stages of female cancers in TCGA database (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$). **(E)** Sanguini diagram representing the correlation between TET2 expression and the age, TMN stage (or tumor grade), and prognosis of patients.

microscope. The resulting area was measured and cells were quantified using the ImageJ software.

RESULTS

Gene Expression Analysis Using Open-Access Databases

The gene expression patterns of TET2 in normal tissues and immune cells are shown in **Figures 1A,B**. According to the combined analysis of data from the HPA, GTEx, and Functional Annotation of the Mammalian Genome 5 (FANTOM5), TET2 was found to be moderately or highly expressed in all detected tissues (normalized expression value >1), demonstrating low tissue-specific RNA expression (**Figure 1A**, Human Protein Atlas). However, a contradictory result was observed in immune cells. High cell-specific RNA expression was observed in neutrophils (**Figure 1C**, Human Protein Atlas). Moreover, according to results extracted from the HPA, Monaco, and Schmiedel databases at the single-cell

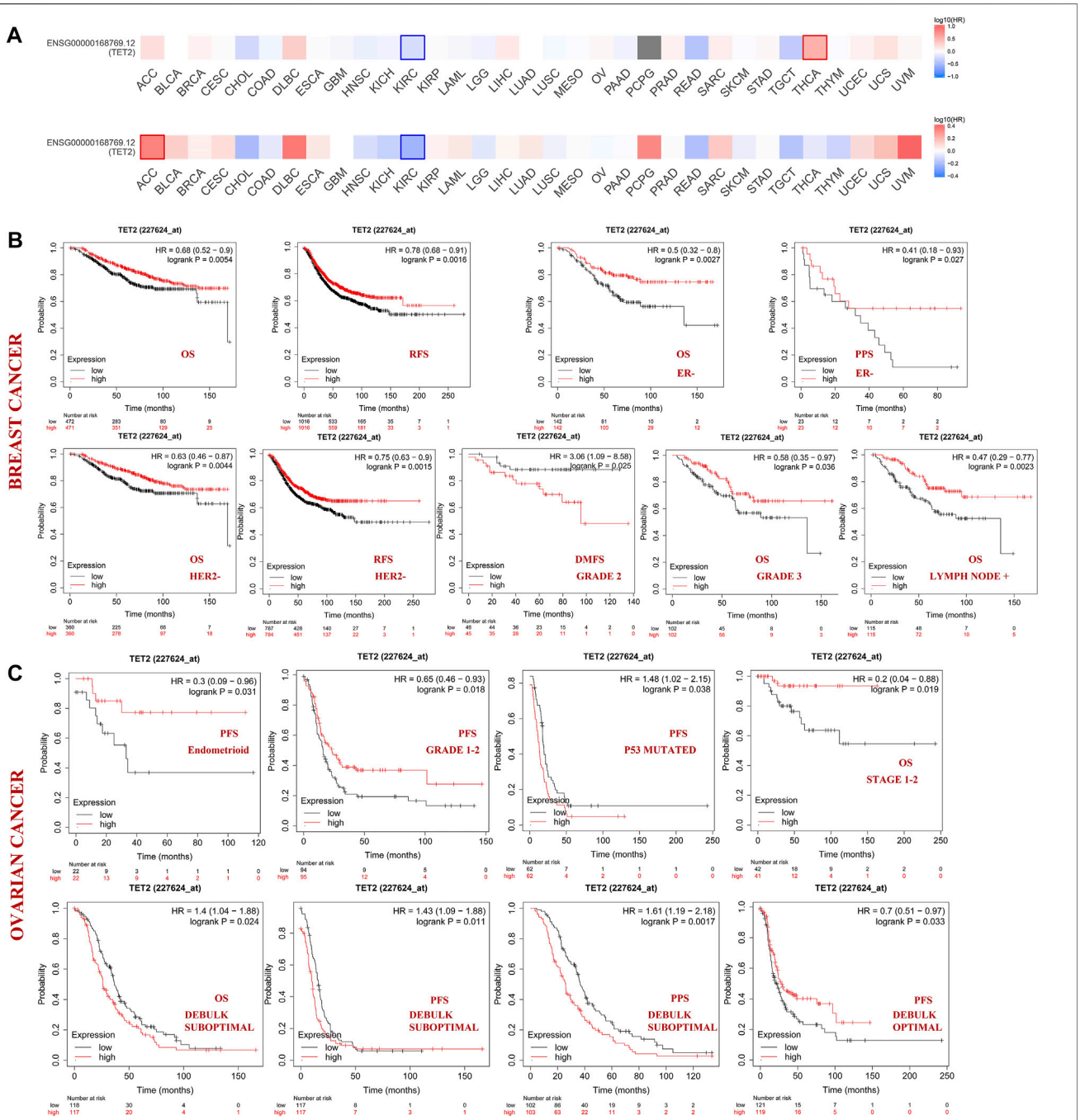


FIGURE 3 | (A) Correlation between TET2 and clinical prognosis analyzed using GEPIA2. The bold border indicates statistical significance. The Kaplan–Meier plotter was also used to examine the association between TET2 and clinical prognosis. Survival studies related to TET2 in breast **(B)** and ovarian cancers **(C)**.

resolution, RNA single-cell-type specificity was low (**Figure 1B**, Human Protein Atlas). **Figure 1D** demonstrates the IHC staining (nuclear staining) of several normal tissues (the bone marrow, cerebral cortex, colon, kidney, liver, lymph node, and testis).

TET2 expression in major female cancers in TCGA database (including BRCA, CESC, OV, UCEC, and UCS) was analyzed using GEPIA2 and TIMER2 (**Figures 2A,B**). The TET2 expression levels were lower in patients with BRCA, UCEC, OV, and UCS than in healthy controls ($p < 0.05$). Furthermore, the total TET2 protein levels were not available in the CPTAC database, and differences in TET2 phosphorylation levels were observed between normal and primary tumor tissues. The data of patients with BRCA and UCEC in the CPTAC dataset were further assessed. The S38 locus exhibited a higher phosphorylation level in UCEC tissues than in normal tissues ($p < 0.001$), followed by a non-

TABLE 1 | Survival analysis between TET2 high expression and low expression groups in patients with breast cancer (Kaplan-Meier Plotter).

Condition/Clinical outcome	Case number	HR (95%CI)	*P*
All			
RFS	4934	**0.78 (0.68–0.91)**	**0.0016**
OS	1880	**0.68 (0.52–0.90)**	**0.0054**
DMFS	2767	0.68 (0.75–1.27)	0.85
PPS	458	0.80 (0.56–1.14)	0.22
ER (IHC)			
Positive			
RFS	902	0.91 (0.69–1.21)	0.51
OS	221	0.69 (0.35–1.34)	0.27
DMFS	248	1.09 (0.52–2.28)	0.83
PPS	34	0.64 (0.26–1.57)	0.33
Negative			
RFS	470	0.96 (0.71–1.31)	0.82
OS	284	**0.50 (0.32–0.80)**	**0.0027**
DMFS	270	0.87 (0.57–1.33)	0.52
PPS	46	**0.41 (0.18–0.93)**	**0.027**
PR (IHC)			
Positive			
RFS	511	0.96 (0.67–1.4)	0.85
OS	0	-	-
DMFS	144	2.02 (0.69–5.9)	0.19
PPS	0	—	—
Negative			
RFS	436	0.97 (0.69–1.36)	0.86
OS	148	0.82 (0.43–1.56)	0.55
DMFS	266	0.87 (0.56–1.35)	0.52
PPS	3	—	—
HER2 (Array)			
Positive			
RFS	461	0.87 (0.64–1.18)	0.36
OS	223	0.92 (0.56–1.52)	0.75
DMFS	260	1.23 (0.78–1.94)	0.37
PPS	54	1.15 (0.63–2.12)	0.65
Negative			
RFS	1,571	**0.75 (0.63–0.90)**	**0.0015**
OS	720	**0.63 (0.46–0.87)**	**0.0044**
DMFS	698	0.90 (0.65–1.25)	0.54
PPS	126	0.71 (0.46–1.11)	0.13
Lymph node			
Positive			
RFS	814	0.83 (0.65–1.06)	0.13
OS	230	**0.47 (0.29–0.77)**	**0.0023**
DMFS	261	0.81 (0.51–1.30)	0.38
PPS	76	0.57 (0.31–1.06)	0.073
Negative			
RFS	574	0.98 (0.67–1.42)	0.90
OS	180	0.66 (0.29–1.46)	0.30
DMFS	240	1.22 (0.61–2.44)	0.58
PPS	23	0.70 (0.23–2.15)	0.53
Grade			
1			
RFS	113	1.80 (0.60–5.36)	0.29
OS	26	0.73 (0.06–8.33)	0.80
DMFS	44	—	—
PPS	6	—	—
2			
RFS	243	0.77 (0.47–1.27)	0.30

(Continued in next column)

TABLE 1 | (*Continued*) Survival analysis between TET2 high expression and low expression groups in patients with breast cancer (Kaplan-Meier Plotter).

Condition/Clinical outcome	Case number	HR (95%CI)	*P*
OS	64	0.42 (0.13–1.40)	0.15
DMFS	91	**3.06 (1.09–8.58)**	**0.025**
PPS	13	—	—
3			
RFS	481	0.92 (0.68–1.23)	0.56
OS	204	**0.58 (0.35–0.97)**	**0.036**
DMFS	234	1.21 (0.72–2.02)	0.47
PPS	72	0.61 (0.33–1.10)	0.099
TP53 status			
Mutated			
RFS	132	0.92 (0.51–1.65)	0.77
OS	56	0.96 (0.25–3.64)	0.96
DMFS	56	1.10 (0.38–3.13)	0.86
PPS	11	—	—
Wild type			
RFS	82	0.71 (0.30–1.66)	0.43
OS	6	—	—
DMFS	6	—	—
PPS	0	—	—
Systemically untreated patients			
Included			
RFS	61	0.41 (0.14–1.19)	0.09
OS	0	—	—
DMFS	6	—	—
PPS	0	—	—
Excluded			
RFS	751	0.90 (0.73–1.11)	0.32
OS	107	0.72 (0.25–2.01)	0.52
DMFS	324	1.34 (0.81–2.22)	0.25
PPS	17	—	—
Endocrine therapy			
Included			
RFS	385	1.01 (0.63–1.60)	0.98
OS	50	—	—
DMFS	205	0.86 (0.37–1.99)	0.72
PPS	0	—	—
Excluded			
RFS	275	0.87 (0.56–1.34)	0.53
OS	107	0.72 (0.25–2.01)	0.52
DMFS	177	1.59 (0.88–2.84)	0.12
PPS	17	—	—
Chemotherapy			
Adjuvant only			
RFS	255	1.27 (0.79–2.03)	0.32
OS	0	—	—
DMFS	118	2.03 (0.90–4.57)	0.08
PPS	0	—	—
Neoadjuvant only			
RFS	111	0.75 (0.35–1.58)	0.45
OS	107	0.72 (0.25–2.01)	0.52
DMFS	107	0.73 (0.32–1.65)	0.45
PPS	17	—	—

Abbreviation: OS, overall survival; DMFS, distant metastasis-free survival; RFS, relapse-free survival; PPS, post-progression survival; IHC, immunohistochemical staining; ER, estrogen receptor; HER2, human epithelial growth factor receptor 2; PR, progesterone receptor. Note: High expression and low expression groups were defined by the median expression of TET2. Bold values are statistically significant ($p < 0.05$).

TABLE 2 | Survival analysis between TET2 high expression and low expression groups in patients with ovarian cancer (Kaplan-Meier Plotter).

Condition/Clinical outcome	Number of patients with available clinical data	HR (95%CI)	*P*
All			
PFS	614	0.98 (0.81–1.18)	0.81
OS	655	0.68 (0.81–1.21)	0.93
PPS	382	0.80 (0.92–1.48)	0.19
Histology			
Endometrioid			
PFS	44	**0.30 (0.09–0.96)**	**0.031**
OS	30	0.68 (0.10–4.84)	0.70
PPS	10	—	—
Serous			
PFS	483	0.98 (0.80–1.21)	0.88
OS	523	1.00 (0.80–1.25)	1.00
PPS	346	1.09 (0.85–1.40)	0.49
Stage			
1 + 2			
PFS	115	0.57 (0.27–1.20)	0.13
OS	83	**0.20 (0.04–0.88)**	**0.019**
PPS	20	2.44 (0.65–9.14)	0.18
3 + 4			
PFS	494	1.10 (0.91–1.34)	0.32
OS	487	1.01 (0.81–1.27)	0.91
PPS	361	1.12 (0.88–1.43)	0.35
Grade			
1 + 2			
PFS	189	0.65 (0.46–0.93)	0.018
OS	203	0.78 (0.52–1.16)	0.22
PPS	118	1.26 (0.81–1.96)	0.3
3			
PFS	315	1.23 (0.96–1.58)	0.10
OS	392	1.05 (0.82–1.35)	0.71
PPS	240	1.15 (0.86–1.55)	0.34
4		—	
PFS	18	—	—
OS	18	—	0.53
PPS	18	0.73 (0.27–1.98)	—
TP53 status			
Mutated			
PFS	124	1.48 (1.02–2.15)	0.038
OS	124	1.24 (0.85–1.81)	0.26
PPS	116	1.19 (0.81–1.74)	0.37
Wild type			
PFS	19	1.57 (0.56–4.37)	0.38
OS	19	0.94 (0.34–2.65)	0.91
PPS	17	0.79 (0.27–2.32)	0.67
Debulk			
Optimal			
PFS	240	**0.70 (0.51–0.92)**	**0.033**
OS	243	0.96 (0.64–1.43)	0.83
PPS	139	1.18 (0.77–1.81)	0.43
Suboptimal			
PFS	234	**1.43 (1.09–1.88)**	**0.011**
OS	235	**1.40 (1.04–1.88)**	**0.024**
PPS	205	**1.61 (1.19–2.18)**	**0.0017**
Chemotherapy			
Contains Platin			
PFS	502	1.14 (0.94–1.38)	0.19
OS	478	1.13 (0.90–1.43)	0.29

(Continued on following page)

TABLE 2 | (*Continued*) Survival analysis between TET2 high expression and low expression groups in patients with ovarian cancer (Kaplan-Meier Plotter).

Condition/Clinical outcome	Number of patients with available clinical data	HR (95%CI)	*P*
PPS	373	1.20 (0.95–1.53)	0.12
Contains Taxol			
PFS	381	1.04 (0.83–1.30)	0.73
OS	357	1.08 (0.81–1.44)	0.59
PPS	274	1.18 (0.88–1.57)	0.26
Contains Platin + Taxol			
PFS	380	0.87 (0.67–1.12)	0.27
OS	356	1.18 (0.88–1.59)	0.27
PPS	273	1.23 (0.92–1.65)	0.17
Contains Avastin			
PFS	0	—	—
OS	0	—	—
PPS	0	—	—
Contains Docetaxel		—	—
PFS	0	—	—
OS	0	—	—
PPS	0	—	—
Contains Gemcitabine		—	—
PFS	0	—	—
OS	0	—	—
PPS	0	—	—
Contains Paclitaxel		—	—
PFS	0	—	—
OS	0	—	—
PPS	0	—	—
Contains Topotecan		—	—
PFS	0	—	—
OS	0	—	—
PPS	0	—	—

Abbreviation: OS, overall survival; PFS, progression-free survival; PPS, post-progression survival. Note: High-expression and low-expression groups were defined by the median expression of TET2.

Bold values are statistically significant (p < 0.05).

statistical phosphorylation level of the S99 locus in BRCA tissues. In patients with BRCA, the phosphorylation level of the S99 locus was not correlated with stage, age, and histological features of tumors. In patients with UCEC, the phosphorylation level of S38 was higher in patients aged 21–60 years than in those aged 61–80 years, except for the difference between tumor and normal tissues. The phosphorylation level of the S38 locus was significantly higher in patients with grade 2 UCEC than in patients with grade 1 and 3 UCEC (**Figure 2C**). The PhosphoNET database was used to analyze CPTAC-identified phosphorylation of TET2, and it was found that S38 and S99 phosphorylation ranked 3rd and 11th among phosphorylation levels of all TET2 loci. In addition, S99 phosphorylation was experimentally supported, whereas S38 phosphorylation was not (S99 source: Courtesy of Dr. Leonard Foster and Lindsay Rogers, University of British Columbia, hydrophobicity = −0.260, P-site similarity score = −52.8, maximum kinase specificity = 522, sum of kinase specificity scores = 22,740 and conservation score = 15.5; S38: hydrophobicity = −1.120, P-site similarity score = −55.7, maximum kinase specificity = 451, sum of kinase specificity scores = 19,431 and conservation score = 17.3).

The "pathological stage plot" module of GEPIA2 was used to observe the correlation between TET2 expression and pathological stages of cancer. TET2 was found to be significantly correlated with OV stages among selected tumors (**Figure 2D**, $p < 0.05$). The Sanguini diagram revealed the correlation between TET2 and the age, TMN stage (or tumor grade), and prognosis of patients (**Figure 2E**).

TET2-Related Survival Analysis Based on GEO and TCGA Databases

Patients with different cancers were divided into the high- and low-expression groups based on the median TET2 expression level. The association between TET2 expression and prognosis of patients with various cancer types was analyzed using data obtained from TCGA and GEO. TET2 expression was not associated with prognosis in the selected major female cancers in TCGA database (**Figure 3A**).

Subsequently, patients with BRCA and OV were analyzed using the KM plotter (GEO database). In patients with BRCA, high TET2 expression was associated with favorable OS (HR = 0.68, $p = 0.0054$) and RFS (HR = 0.78, $p = 0.0016$). In patients with estrogen receptor (ER)-negative and human epithelial growth factor receptor 2 (HER2)-negative BRCA, high TET2 expression was correlated with a better prognosis (OS: HR = 0.50, $p = 0.0027$;

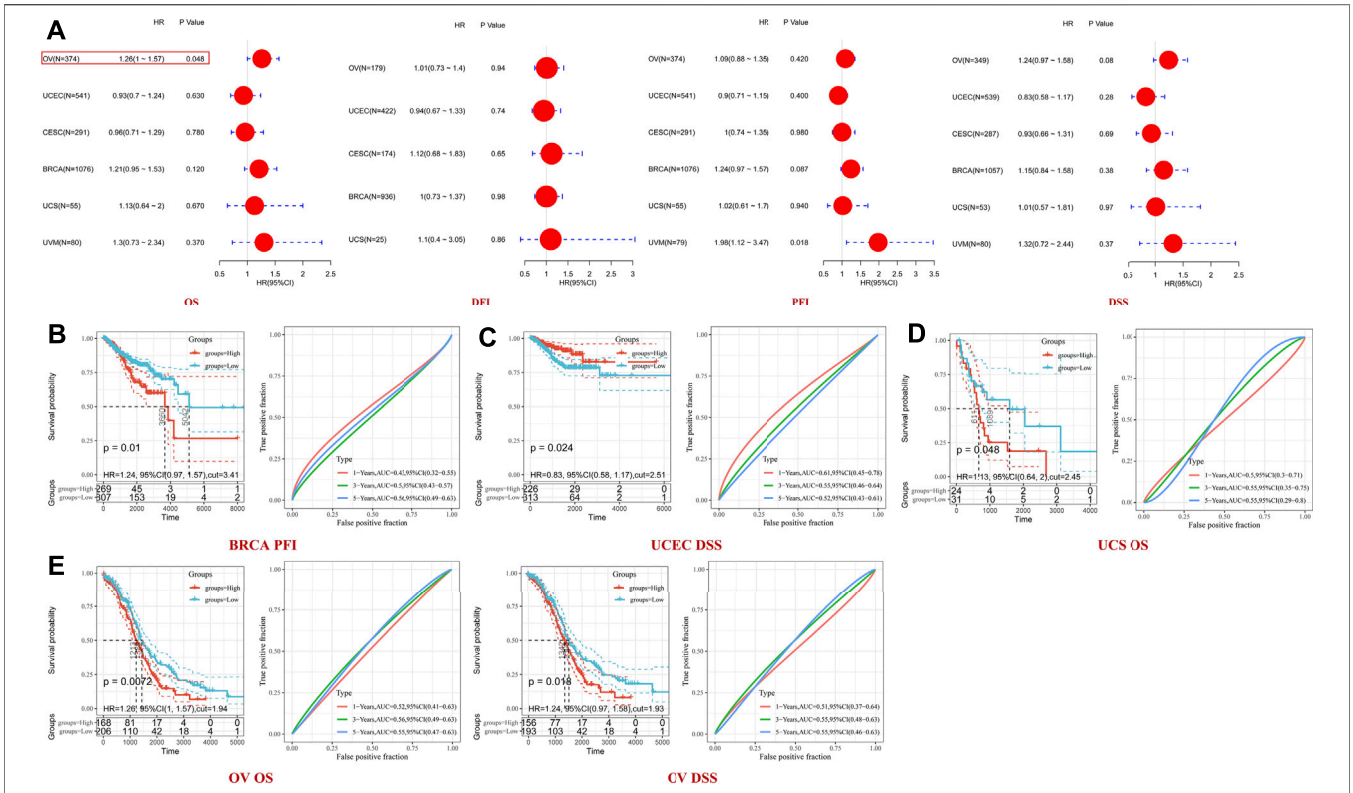


FIGURE 4 | (A) Pooled analysis between TET2 expression and OS, disease-free interval (DFI), disease-specific survival (DSS), and progression-free interval (PFI) based on data of female cancers obtained from TCGA; receiver operator characteristic curve analysis was performed to determine the efficiency of TET2 in predicting the 1-, 3-, and 5-years OS, PFI, or DSS of patients with breast cancer **(B)**, uterine corpus endometrial carcinoma **(C)**, uterine carcinosarcoma **(D)**, and ovarian cancer **(E)**.

PPS: HR = 0.41, $p = 0.027$ in ER-negative patients; OS: HR = 0.63, $p = 0.0044$; RFS: HR = 0.75, $p = 0.0015$ in HER2-negative patients). In addition, high TET2 expression was associated with good OS in patients with grade 3 BRCA (HR = 0.58, $p = 0.036$) and BRCA with lymph node involvement (HR = 0.47, $p = 0.0023$). High TET2 expression was only associated with a poor prognosis in patients with grade 2 BRCA (DMFS: HR = 3.06, $p = 0.025$) (**Figure 3B**). Detailed results of univariate prognostic analysis are summarized in **Table 1**. The TET2 expression level was associated with the prognosis of patients with endometrioid OV (PFS: HR = 0.3, $p = 0.038$), grade 1–2 OV (PFS: HR = 0.65, $p = 0.018$), P53-mutated OV (PFS: HR = 1.48, $p = 0.038$), and stage 1–2 OV (OS: HR = 0.2, $p = 0.019$). TET2 expression was also associated with the prognosis of patients with OV treated with suboptimal debulking (OS: HR = 1.4, $p = 0.024$; PFS: HR = 1.43, $p = 0.011$; PPS: HR = 1.61, $p = 0.0017$) and optimal debulking (OS: HR = 0.7, $p = 0.033$) (**Figure 3C**). Detailed results are summarized in **Table 2**. Therefore, TET2 expression indicating clinical outcomes differed between patients with BRCA and OV.

A pooled analysis was performed to examine the association between TET2 expression and OS, DFI, DSS, and PFI in major female cancers in TCGA. The forest plot revealed that TET2 was an independent prognostic factor for OS in OV (HR = 1.26, $p = 0.048$) but not for DFI and DSS in any selected tumors (**Figure 4A**). Furthermore, ROC analysis was performed to determine the efficiency of TET2 in predicting the 1-, 3-, and 5-years OS, DFI, PFI, and DSS in patients with selected tumor types from TCGA database. Subsequently, the optimal cut-off value of TET2 expression was calculated and used to divide patients into the high- and low-expression groups. A survival analysis was performed to validate prognosis between the groups. The acceptable prediction efficiency of TET2 was defined as area under the curve (AUC) > 0.6 and p-value < 0.05. TET2 was found to have statistically significant efficiency in predicting survival in several tumors (OS in UCS [cut-off value, 2.45; $p = 0.048$; 3-years AUC, 0.55; 95% CI, 0.35–0.75; 5-years AUC 0.55; 95% CI, 0.29–0.8] and OV [cut-off value, 1.94; $p = 0.072$; 1-year AUC, 0.52; 95% CI, 0.41–0.63; 3-years AUC, 0.56; 95% CI, 0.49–0.63; 5-years AUC, 0.55; 95%CI, 0.47–0.63], PFI in BRCA [cut-off value, 3.41; $p = 0.01$; 5-years AUC, 0.56; 95% CI, 0.49–0.63), and DSS in OV [cut-off value, 1.93; $p = 0.072$; 1-year AUC, 0.51; 95% CI, 0.37–0.64; 3-years AUC, 0.55; 95% CI, 0.48–0.63; 5-years AUC, 0.55; 95% CI, 0.46–0.63]). However, the efficiency of TET2 in predicting DSS in UCEC was acceptable (cut-off value, 2.51; $p = 0.024$; 1-year AUC, 0.61; 95% CI, 0.45–0.78) (**Figures 4B–E**).

TET2 Genetic Alteration and DNA Methylation Analysis

Genetic alterations of TET2 in common female cancers (UCEC, BRCA, CESC, UCS, and OV) were investigated. The results

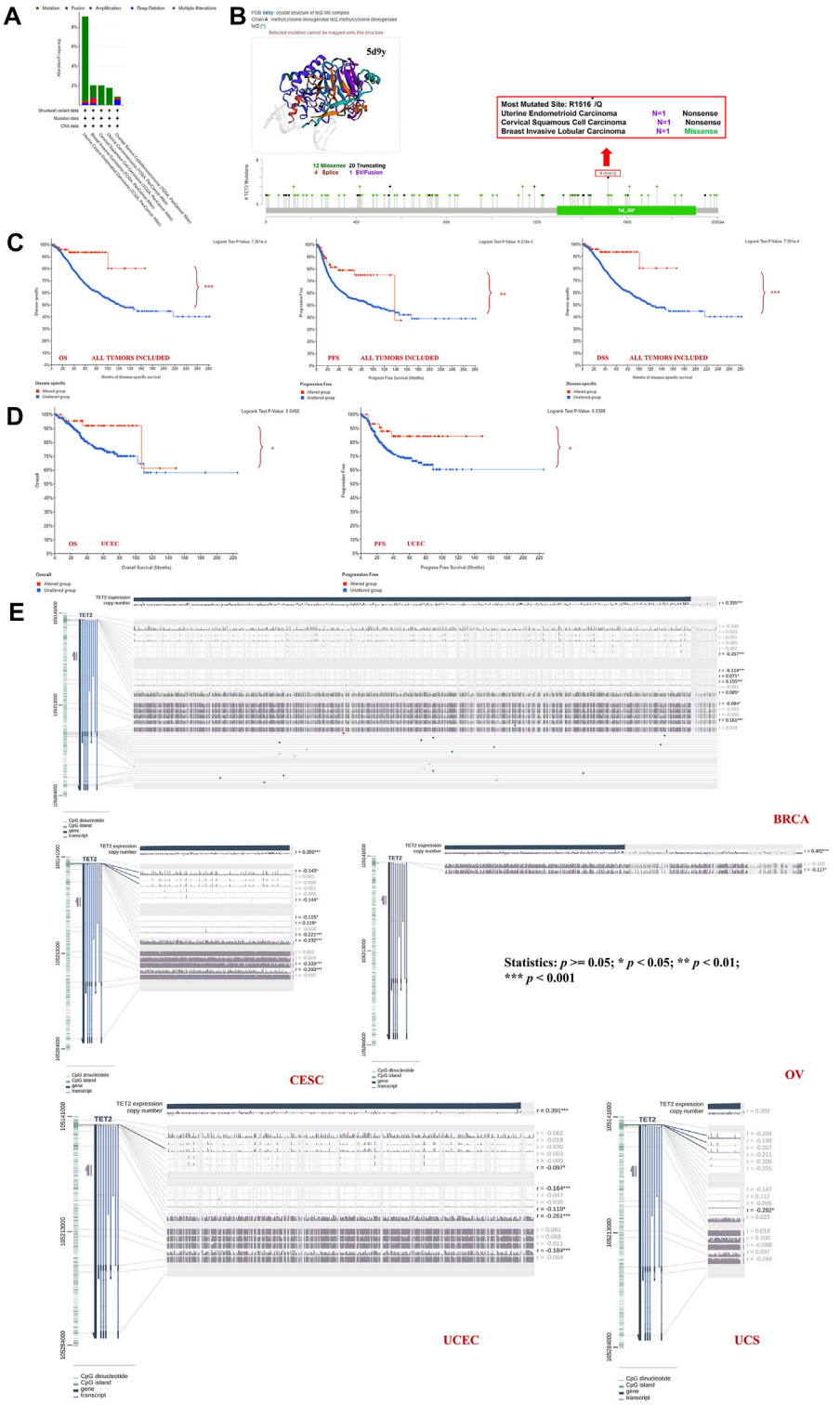


FIGURE 5 | TET2 mutation features analyzed using cBioPortal for Cancer Genomics based on TCGA database. Mutation patterns with cancer types **(A)**, most altered sites and three-dimensional structure of TET2 **(B)**. Survival analysis for assessing the correlation between TET2 mutation and clinical prognosis of all female cancers **(C)** and uterine corpus endometrial carcinoma **(D)**. **(E)** Correlation between DNA methylation and TET2 expression in female cancers in TCGA database (*p < 0.05; **p < 0.01; ***p < 0.001).

TABLE 3 | Correlation of TET2 DNA methylation and gene expression at multiple probes.

Variable	*p*_value	pearson_*r*
BRCA		
cg13440296	3.57E-06	−0.15663
cg13365781	0.000731	−0.11437
cg02382073	0.035197	0.071456
cg09295382	4.52E-06	0.154988
cg09666717	0.0126	0.084798
cg22794775	0.013024	−0.08421
cg17862558	1.92E-06	0.161007
CESC		
cg08924430	0.01208	−0.14263
cg13440296	0.011217	−0.14433
cg13365781	0.0433	−0.11505
cg02382073	0.036848	0.118813
cg00911488	8.75E-05	−0.22136
cg09666717	3.89E-05	−0.23224
cg20586654	2.55E-09	−0.33288
cg17862558	1.71E-07	−0.29336
OV		
cg07360692	0.053061	−0.10038
cg08924430	0.024524	−0.11659
UCEC		
cg13440296	0.036476	−0.09724
cg13365781	0.000396	−0.164
cg00911488	0.017602	−0.11029
cg09666717	1.53E-08	−0.26089
cg17862558	7.12E-05	−0.18357
UCS		
cg00911488	0.029101599	−0.291904522

Abbreviation: BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; OV, ovarian serous cystadenocarcinoma; UCEC, uterine corpus endometrial carcinoma; UCS uterine carcinosarcoma.

revealed that mutation in the TET2 gene was the most common genetic alteration found in most patients with selected cancers. Patients with UCEC had the highest proportion (9.07%) of TET2 genetic alterations, with gene mutations constituting the largest proportion (46/529, 8.7%) (**Figure 5A**). Detailed information on TET2 genetic alterations is presented in **Figure 5B**. Furthermore, a missense mutation in TET2 was the most common type of alteration, whereas the R1516*/Q was the most common mutation site found in one patient with UCEC, one with CESC, and one with BRCA. This mutation site was found to induce missense or nonsense mutation in TET2. The possible mutated protein structure of TET2 is presented in **Figure 5B**. Consequently, the association between TET2 genetic alterations and the prognosis of patients with selected cancers was assessed. As demonstrated in **Figure 5C**, TET2 genetic alterations were associated with favorable OS, PFS, and DSS in selected tumors. Moreover, favorable OS and PFS were only observed in patients with UCEC (**Figure 5D**). Furthermore, the association between the DNA methylation level and gene expression of TET2 in selected cancers in TCGA database was analyzed using MEXPRESS. The DNA methylation level was found to be significantly correlated with the gene transcriptional level at different probes. Data are summarized in **Table 3** and **Figure 5E**. The heatmaps of single CpG site methylation of TET2 in BRCA, CESC, UCEC, and UCS (analyzed using MethSurv) are demonstrated in **Figure 6A**. It was found that cg12306086, cg20586654, cg08530497, and cg22794775 in BRCA, CESC, UCEC, and UCS, respectively, had the highest DNA methylation level and were mainly associated with the prognosis of patients with BRCA. Detailed results are provided in **Table 4**. As demonstrated in **Figure 6B**, the heatmap (created using SurvivalMeth) revealed the single CpG methylation level of TET2 in BRCA (TCGA and the GSE37754 dataset of the GEO project) and UCEC (TCGA) samples. Moreover, significant differences in the expression patterns of single CpG methylation of TET2 were found between the low- and high-risk groups in BRCA and UCEC. In addition, a significant prognostic correlation was observed between the high- and low-risk groups. These results supported the conclusions drawn from MethSurv analysis.

TET2-Related Immune Infiltration Analysis Based on TCGA Database

As one of the main characteristics of the tumor microenvironment, immune cell infiltration is strongly associated with oncogenesis. Tumor-associated fibroblasts and CD 8^+ T cells located in the stroma of the tumor microenvironment contribute to regulating the functions of different tumor-infiltrating immune cells. In this study, the association between immune cell infiltration and TET2 transcriptional level was analyzed using data from TCGA database. The results revealed that $CD8^+$ T-cell infiltration was not correlated with TET2 expression in OV, UCEC, and UCS (**Figure 7A**). TET2 expression was significantly positively correlated with immune-infiltrating tumor-associated fibroblasts in BRCA, CESC, and OV (**Figure 7B**). The representative plots of correlation analysis of the aforementioned cancers estimated by a single algorithm are demonstrated in **Figures 7A,B**. For instance, estimation using TIMER revealed that TET2 expression was found to be positively correlated with $CD8^+$ T-cell infiltration in BRCA ($r = 0.315$, $p =$ 2.92e-24).

Moreover, TET2 expression was found to be negatively correlated with TMB in UCEC ($p = 0.0081$) (**Figure 8A**). No correlation was found between TET2 expression and MSI (**Figure 8B**). Furthermore, a positive correlation was observed between TET2 expression and ADORA2A, CD160, CD200, CD200R1, CD44, CD80, NRP1 TNFSF4, and TNFSF15 in most female cancers in TCGA database, such as BRCA, OV, UCEC, and CESC (**Figure 8C**). However, TET2 expression was not associated with any checkpoints in UCS. Moreover, based on the correlation analyses of immune response pathways, TET2 expression was negatively correlated with activated $CD8^+$ T cells, $CD56^+$ natural killer cells, gamma delta T cells, macrophages, MDSC, and monocytes but positively correlated with memory B cells in BRCA, CESC, and UCEC in TCGA database (**Figure 8D**). Furthermore, TET2 expression was not associated with the number of neoantigens in female cancers (**Figure 8E**). In addition, the correlation between TET2 expression and the tumor microenvironment was quantitatively assessed using the R package "ESTIMATE", and

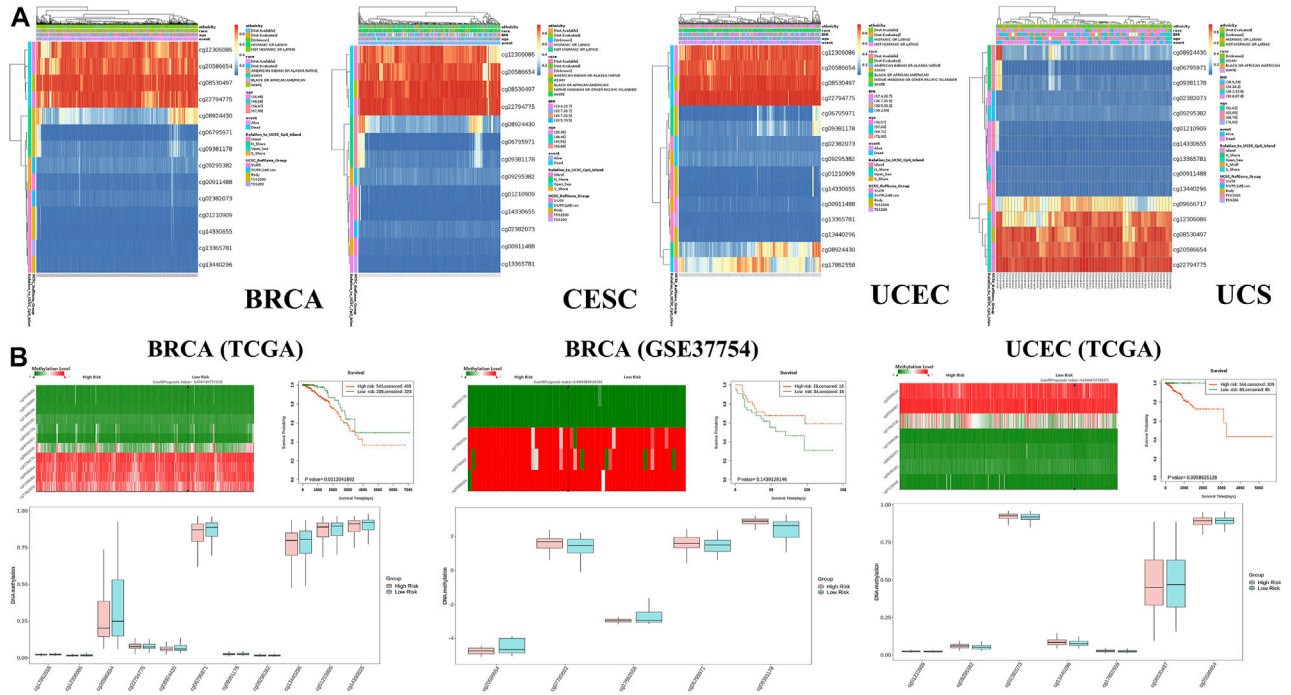


FIGURE 6 | (A) Heatmap of the clustered expression level of TET2 in single CpG methylation (MethSurv) in breast cancer, cervical squamous cell carcinoma and endocervical adenocarcinoma, uterine corpus endometrial carcinoma, uterine carcinosarcoma, and uveal melanoma. **(B)** Heatmap of single CpG methylation level (upper left panel); methylation difference in CpG between the high- and low-risk groups (lower panel) and the Kaplan–Meier survival plot (upper right panel). These analyses were performed using SurvivalMeth based on the data of TCGA and GEO projects.

TABLE 4 | The prognostic significance of single CpG methylation of TET2 in patients with female cancer.

	HR	95% CI	P
BRCA			
cg12306086	0.607	(0.402; 0.916)	**0.017**
cg20586654	0.67	(0.452; 0.992)	**0.045**
cg08530497	0.417	(0.26; 0.669)	**0.00029**
cg22794775	0.584	(0.392; 0.87)	**0.0081**
CESC			
cg12306086	0.785	(0.464; 1.328)	0.37
cg20586654	0.865	(0.544; 1.377)	0.54
cg08530497	0.316	(0.162; 0.618)	**0.00075**
cg22794775	0.657	(0.41; 1.054)	0.081
UCEC			
cg12306086	1.447	(0.817; 2.564)	0.21
cg20586654	1.441	(0.884; 2.349)	0.14
cg08530497	0.635	(0.342; 1.18)	0.15
cg22794775	1.453	(0.897; 2.353)	0.13
UCS			
cg12306086	1.128	(0.569; 2.238)	0.73
cg20586654	1.392	(0.627; 3.089)	0.42
cg08530497	2.122	(0.946; 4.761)	0.068
cg22794775	0.84	(0.431; 1.637)	0.61

Abbreviation: BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; OV, ovarian serous cystadenocarcinoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma; HR, hazard ratio; CI, confidence interval.

Bold values are statistically significant ($p < 0.05$).

the correlation coefficients of TET2 expression and ESTIMATE, stromal, and immune scores were evaluated separately. The ESTIMATE score was negatively associated with TET2 expression in UCEC and CESC, the stromal score was negatively associated with TET2 expression in UCEC, CESC, and BRCA and the immune score was positively associated with TET2 expression in OV and BRCA (**Figure 8F**).

Enrichment Analysis of TET2-Related Parameters

To investigate the potential mechanisms of action of TET2 in oncogenesis, potential TET2-binding proteins and TET2-expression-associated genes were identified for pathway enrichment analyses. A total of 50 TET2-binding proteins were extracted using STRING. The interaction networks of TET2 and the 50 proteins are demonstrated in **Figure 9A**. Thereafter, GEPIA2 was used to combine the expression data of selected tumors obtained from TCGA, and the top 100 related genes were found to be associated with TET2 expression. The results revealed that TET2 expression was positively correlated with StAR-related lipid transfer domain containing 3 (R = 0.12), proteasome 26S subunit, non-ATPase 3 (R = 0.11), post-GPI attachment to proteins 3 (R = 0.084), ORMDL sphingolipid biosynthesis regulator 3 (R = 0.12), and growth factor

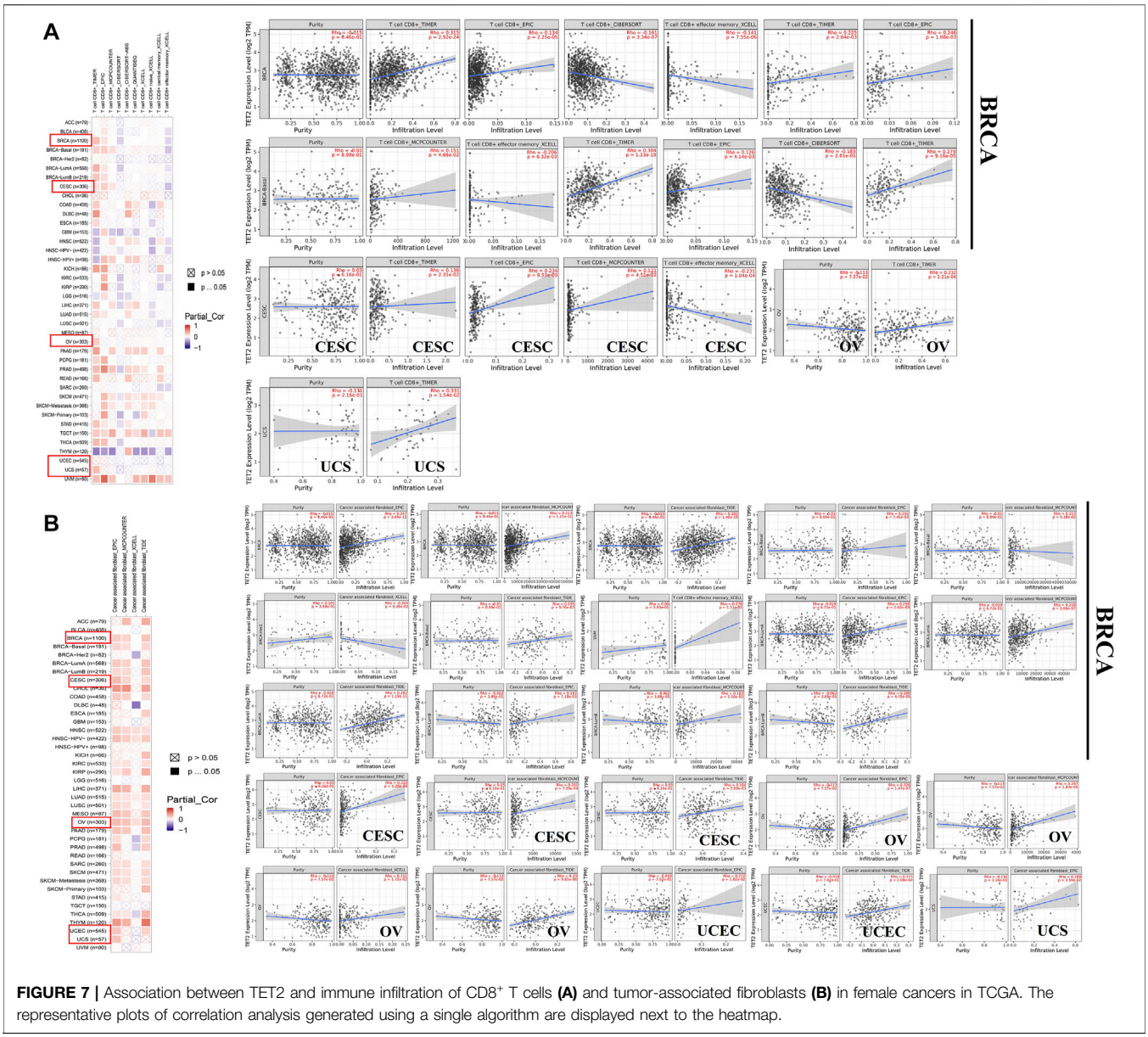


FIGURE 7 | Association between TET2 and immune infiltration of CD8$^+$ T cells **(A)** and tumor-associated fibroblasts **(B)** in female cancers in TCGA. The representative plots of correlation analysis generated using a single algorithm are displayed next to the heatmap.

receptor-bound protein 7 (R = 0.039) genes (**Figure 9B**). As demonstrated in the heatmap representing these genes (**Figure 9C**), they were found to be positively correlated with TET2 expression mainly in OV and CESC. Furthermore, KEGG and GO enrichment analyses were conducted using the combined results of the aforementioned analyses. Based on the GO analysis, TET2-associated genes were mostly involved in processes associated with collagen catabolism, histone deacetylase activity, extracellular matrix organization, collagen type IV trimer, extracellular matrix structural constituent, Sin3 complex, endoplasmic reticulum lumen, nucleoplasm, collagen-activated tyrosine kinase receptor signaling pathway, and platelet-derived growth factor binding (top 10) (**Figure 9D**). Based on the KEGG analysis, the top 10 signaling pathways were enriched in protein digestion and absorption, ECM–receptor interaction, human papillomavirus infection, amebiasis, AGE–RAGE signaling in complications associated with diabetes, relaxin signaling, PI3K–Akt signaling, focal adhesion, thyroid hormone signaling, and cell cycle. In addition, several STAR signaling pathways were found to be associated with oncogenesis, such as the PI3K–Akt signaling pathway (p = 0.00017), Notch signaling pathway (p = 0.0078), transcriptional misregulation in cancer (p = 0.021), and Hippo signaling pathway (p = 0.042) (**Figure 9E**).

Furthermore, we performed a differential expression analysis between the high- and low-TET2-expression groups to identify TET2-associated genes. The high-throughput sequencing data of female cancers obtained from TCGA were stratified into the high- and low-expression groups based on the median expression of TET2.

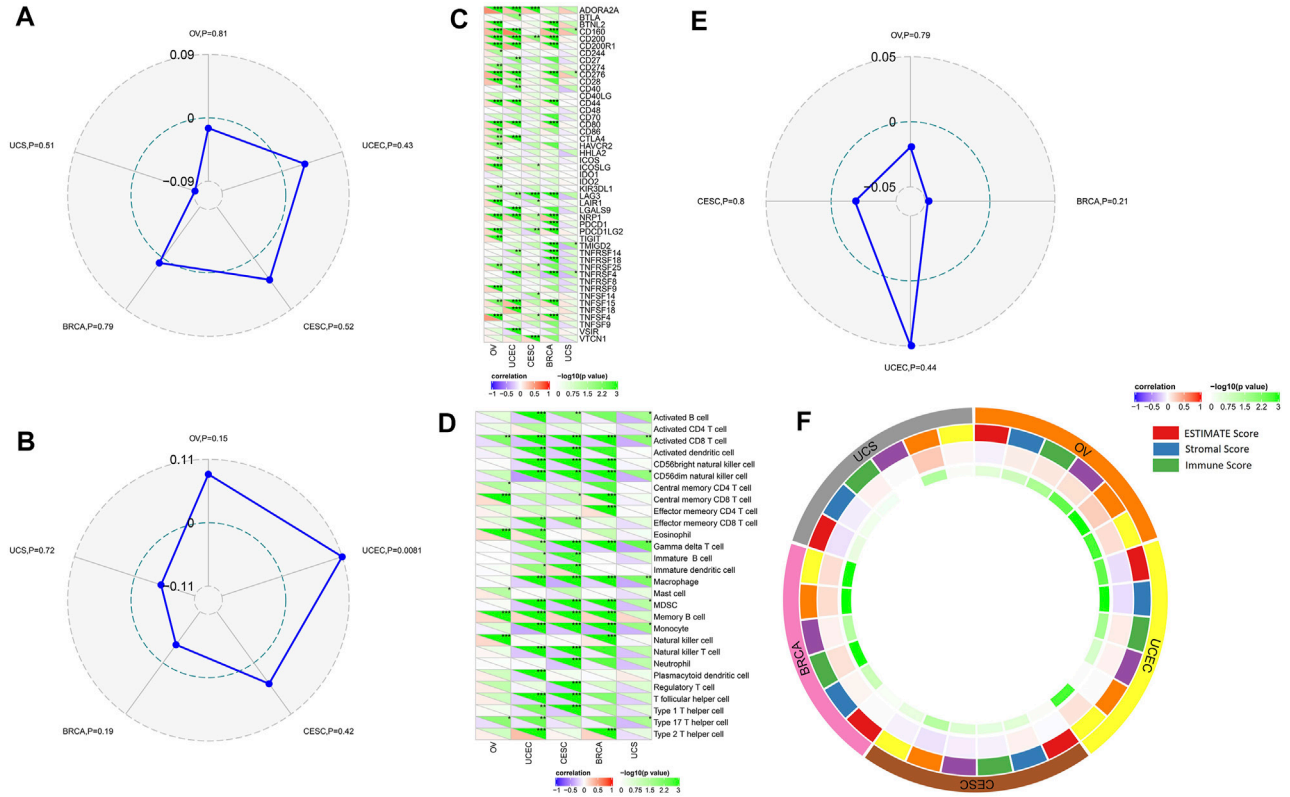


FIGURE 8 | Analysis between TET2 expression and cancer immune infiltration. Correlation between TET2 expression and tumor mutational burden **(A)** and microsatellite instability **(B)**. Correlation between TET2 expression and checkpoint molecules **(C)**, immunization routes **(D)**, and the number of neoantigens **(E)**. Correlation between TET2 expression and ESTIMATE, stromal and immune scores in female cancers **(F)** (*p < 0.05; **p < 0.01; ***p < 0.001).

Differentially expressed genes were screened based on the criteria of fold change >2 and p < 0.05. The result revealed 37 upregulated genes in BRCA, 16 upregulated genes and 1 downregulated gene in CESC, 147 upregulated and 5 downregulated genes in OV, 74 upregulated and 11 downregulated genes in UCEC, and 179 upregulated and 11 downregulated genes in UCS (**Figure 10A**). The Venn diagram revealed no overlapping genes in all types of female cancers (**Figure 10B**). Eventually, 419 genes were identified as TET2-associated genes. Based on the GO analysis, these genes were mostly involved in processes associated with anatomical structure morphogenesis, tissue development, cell adhesion, biological adhesion, collagen-containing extracellular matrix, extracellular matrix, extracellular matrix structural constituent, system development extracellular matrix organization, and animal organ morphogenesis (**Figure 10D**). Based on the KEGG analysis, the associated pathways were enriched in ECM–receptor interaction, human papillomavirus infection, proteoglycans in cancer, arrhythmogenic right ventricular cardiomyopathy (ARVC), Hippo signaling, regulation of the pluripotency of stem cells, breast cancer, Wnt signaling, aldosterone-regulated sodium reabsorption, and small cell lung cancer were enriched. Some enriched pathways are represented in **Figure 9E**. In addition, the STAR signaling pathways, such as the PI3K–Akt signaling (p = 0.00323) and Notch signaling (p = 0.00196) pathways, were associated with oncogenesis (**Figure 10C**). Therefore, these results suggest that TET2 plays a critical role in cancers. Furthermore, two different methods to identify TET2-associated genes revealed seven overlapping TET2-associated genes (TMEM178A, ARSJ, COL4A5, COL1A1, COL3A1, COL5A1, and COL5A2) (**Figure 10E**). As demonstrated in the heatmap (**Figure 10F**), these genes were found to be positively correlated with TET2 expression in all female cancers except UCS.

Histological Analysis of BRCA Tissues

IHC staining was used to detect TET2 in FFPE tissues. As shown in **Figures 11A,B**, TET2 expression was slightly higher in paired paracancerous tissues than in BRCA tissues. In addition, IF staining revealed that TET expression was negatively associated with LAG3 and PDCD1 expression but positively associated with CD276 expression, which was consistent with the results of the previous bioinformatic analyses.

DISCUSSION

Studies have demonstrated that 5-hydroxymethyluracil (5hmU) can be formed by active deamination of 5 hmC through the activation-induced deaminase/apolipoprotein B mRNA editing

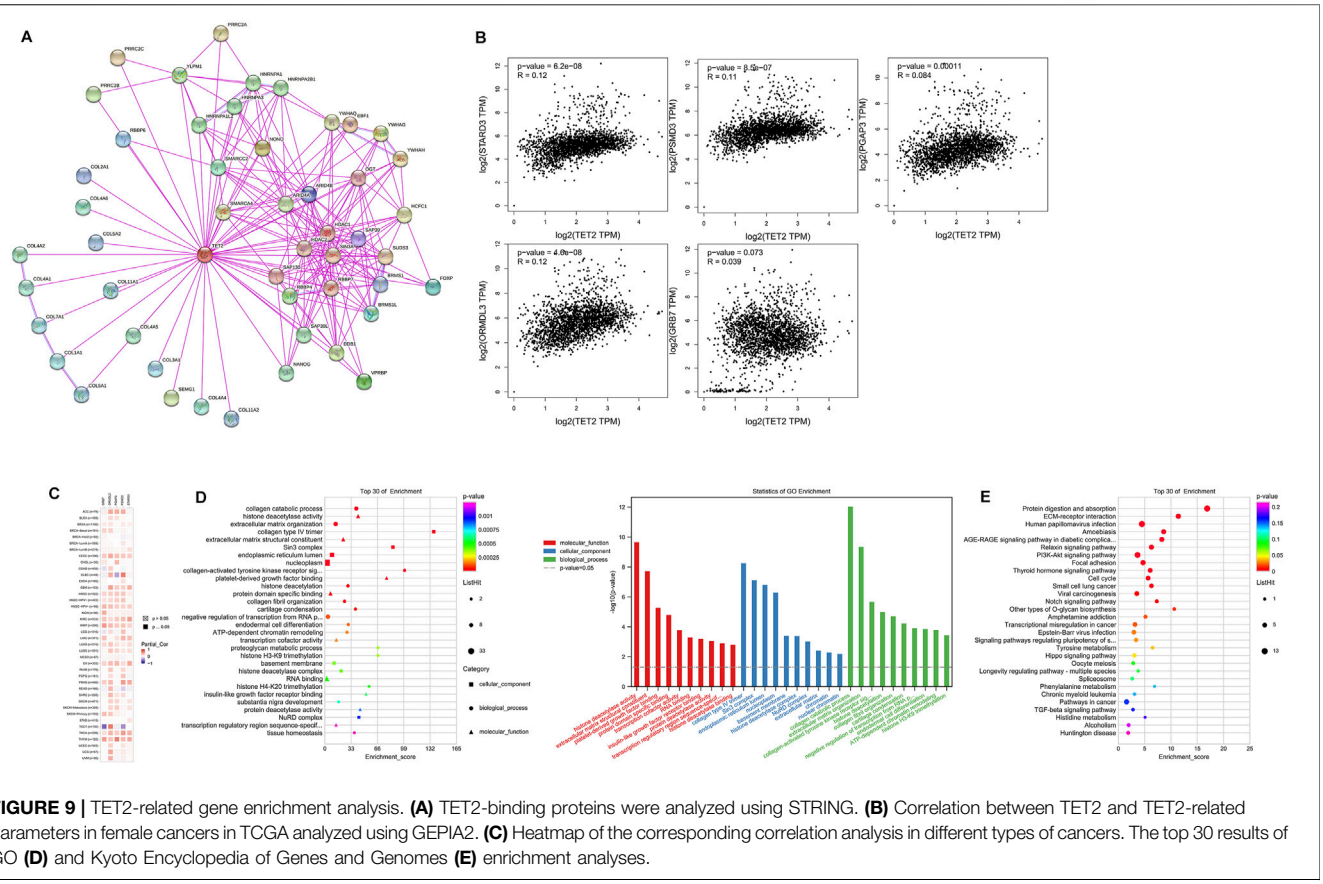


FIGURE 9 | TET2-related gene enrichment analysis. **(A)** TET2-binding proteins were analyzed using STRING. **(B)** Correlation between TET2 and TET2-related parameters in female cancers in TCGA analyzed using GEPIA2. **(C)** Heatmap of the corresponding correlation analysis in different types of cancers. The top 30 results of GO **(D)** and Kyoto Encyclopedia of Genes and Genomes **(E)** enrichment analyses.

enzyme complex (Pfaffeneder et al., 2014). The 5 hmU produced can be removed via the action of DNA glycosylase and the BER pathway. TET2 is considered a tumor suppressor gene, and its haploid deficiency can initiate myeloid and lymphatic transformation. TET2 mutations, such as deletion, insertion, and code shift mutations, are usually accompanied by a significant reduction in the total amount of 5 hmC. Loss of TET2 function caused by the TET2 gene and IDH1/2 mutations is common in myeloid malignancies and lymphomas. One of the most common causes of onco-suppressor gene inactivation in tumors is hypermethylation in the promoter region related to onco-suppressor gene silencing. In recent years, the role of TET2 in the occurrence and development of solid tumors has been gradually revealed. Decreased TET2 expression has been reported in solid malignant tumors, such as thyroid (Jia et al., 2020), gastric (Deng et al., 2016; He et al., 2018), colorectal (Huang et al., 2016; Ma et al., 2018; Zhang et al., 2019), ovarian (Zhang et al., 2015), breast (Yang et al., 2015; Chen et al., 2017; Shen et al., 2021), and prostate cancers (Nickerson et al., 2017). *In vitro* studies have demonstrated decreased TET2 expression is associated with solid tumor development. *In vitro* studies on breast cancer have reported that TET2 knockout upregulates the expression of programmed death-ligand 1 (PD-L1) in MCF7 cells. However, ectopic TET2 expression inhibits the gene expression of PD-L1 in MDA-MB-231 cells (Shen et al., 2021). Furthermore, KDM2A knockout significantly increases TET2 expression in various breast cancer cell lines. In cells with KDM2A deletion, TET2 expression is inhibited owing to ectopic KDM2A expression, indicating that TET2 is the transcriptional inhibition target of KDM2A. Therefore, KDM2A interacting with RelA co-occupies the gene promoter region of TET2, thus inhibiting TET2 transcription. Moreover, RelA or KDM2A depletion can restore TET2 expression. In KDM2A-deficient cells, TET2 upregulation can induce the reactivation of two tumor-suppressive genes, epithelial cell adhesion molecule and e-cadherin located downstream of TET, and thus inhibiting cell migration and invasion (Chen et al., 2017).

This study aimed to demonstrate the overview of the TET2 gene in multiple female cancers and provide a foundation for the study of such cancers. Whether TET2 plays a role in the development of various female cancers mainly through a well-known DNA methylation signaling pathway remains unknown. To the best of our knowledge, studies on TET2 in female cancers have not yet been reported. Therefore, this study comprehensively evaluated the potential functions of TET2 in different female cancers using data extracted from TCGA, CPTAC, and GEO databases, such as TET2 gene expression features, genetic mutations, DNA methylation, immune infiltration, and gene interactions. TET2 was found to be widely expressed in different tissues, with low tissue and cell specificity, and high TET2 expression was found in multiple normal organs. Low TET2 expression was found in female

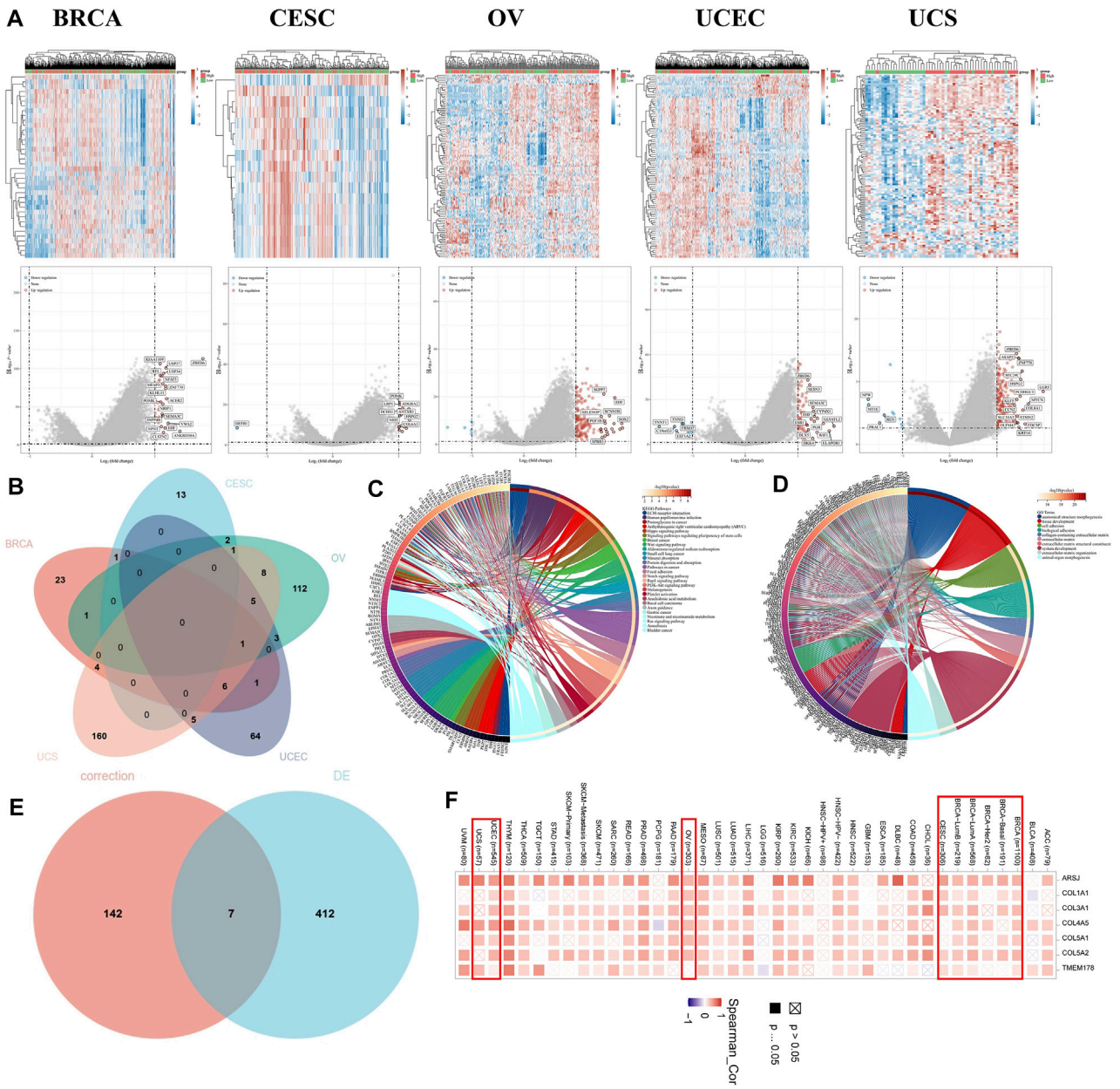


FIGURE 10 | (A) Heatmap (upper panel) and volcano plot (lower panel) representing differentially expressed genes between the high- and low-TET2-expression groups in female cancers. **(B)** Venn diagram demonstrating the overlap between differentially expressed genes among female cancers. Enrichment analyses using Kyoto Encyclopedia of Genes and Genomes **(C)** and GO **(D)**. **(E)** Venn diagram demonstrating seven overlapping genes between two methods (based on correlation analysis and differential expression analysis) of identification of TET2-associated genes. **(F)** Heatmap demonstrating the correlation between TET2 expression and the seven overlapped genes in different types of cancers.

cancers, except for CESC, suggesting its potential role in the development of female cancers. However, this finding could not be supported by protein-level analysis based on data extracted from the CPTAC database owing to the unavailability of data on the total protein level of TET2. The expression of S99-phosphorylated TET2 protein in BRCA and S38-phosphorylated TET2 protein in UCEC was also analyzed. S38-phosphorylated TET2 was found to be elevated in tumor tissues and negatively correlated with the tumor pathological grade. A significant association was observed between TET2 expression and the tumor stage in OV in TCGA database. In a study, Zhang reported that TET2 expression in OV was significantly lower than that in normal ovarian tissues and was correlated with the pathologic stage, tumor grade, lymph node metastasis, and vascular thrombosis (Zhang et al., 2015), which is consistent with our findings.

Survival analyses were conducted using GEPIA2 and KM plotter, suggesting different outcomes in female cancers. TCGA prognostic

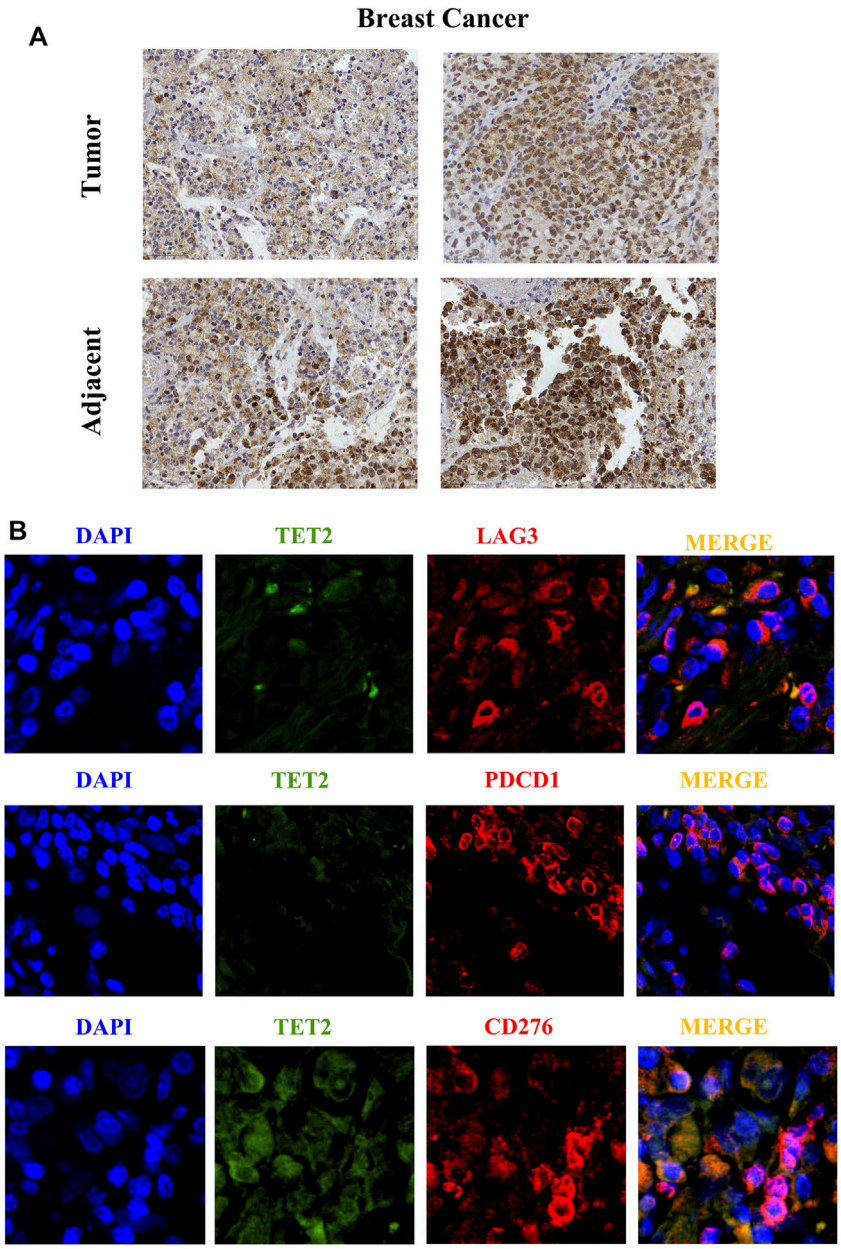


FIGURE 11 | (A) Immunohistochemical staining reveals the protein expression of TET2 in breast cancer and the adjacent tissues. **(B)** Immunofluorescence staining indicates subcellular localization of TET2 and the checkpoint molecules LAG3, CD276, and PDCD1 in breast cancer tissues. (1 cm in figure represents 50 μm in actual).

analysis revealed that TET2 expression was not associated with prognosis in female cancers. However, based on the survival analysis of the KM plotter with Affymetrix HGU133A and HGU133 + 2 microarrays (Győrffy et al., 2010), high TET2 expression was associated with favorable OS and RFS in all patients with BRCA. The results were partially confirmed in the subgroup analysis stratified by clinical parameters, including IHC staining, lymph node involvement, pathological grade, TP53 mutation status, and treatment. In patients with ER-positive, HER2-positive, grade 2–3, and lymph node-involved BRCA, increased TET2 expression indicated good survival. Similar results were observed in patients with OV in the GEO database. However, high TET2 expression was not always associated with a favorable prognosis. Elevated TET2 expression was considered a risk factor in patients with OV characterized by p53 mutation and in those who underwent suboptimal debulking. To further confirm the effects of TET2 on the prognosis of female cancers, studies with larger sample sizes and those focusing on other types of female cancers such as UCEC and CESE should be conducted.

More exhaustive studies are required to determine the exact role of TET2 in tumor development—whether it initiates oncogenesis or acts as a result of tumor development. Analysis of patient data extracted

from TCGA revealed that patients with UCEC had the highest TET2 mutation frequency, at approximately 9%. According to the analysis, TET2 mutation was correlated with a poor prognosis in TET2-mutated female cancers. Furthermore, low DNA methylation levels at multiple regions and upregulated TET2 genetic expression were found to be significantly correlated in female cancers. However, the detailed clinical value should be further identified.

The correlation of TET2 with MSI and TMB in female cancers remains unclear. In this study, TET2 expression was significantly correlated with checkpoints and immunization routes in UCEC and BRCA. In OV, TET2 expression was associated with a majority of checkpoints but not with immunization routes, whereas contradictory results were obtained in CESC. Several checkpoints were examined in FFPE BRCA tissues via IF staining, and bioinformatic analyses were used to confirm the results. Moreover, based on the combined data of TET2-interacting elements and TET2 transcription-associated genes in female cancers, enrichment analysis revealed some potential effects of cancer-associated signaling pathways, such as the PI3K–Akt signaling pathway, Notch signaling pathway, transcriptional misregulation in cancer, and Hippo signaling pathway, on tumor etiologies or oncogenesis. Some of these findings have been reported (Xu, 2020); however, further validation is required *in vitro* and *in vivo*.

We found no correlation between TET2 expression and $CD8^+$ T-cell infiltration in BRCA, CESC, OV, UCEC, and UCS. Jiang (2020) reported that the loss of TET2 enhances the differentiation memory of $CD8^+$ T cells. In addition, the loss of TET2 promotes early acquisition of memory $CD8^+$ T-cell features and strengthens the formation of memory $CD8^+$ T cells in a cell-inherent manner without affecting antigen-driven cell proliferation or effector behavior (Carty et al., 2018). Whether TET2 participates in immune checkpoint suppression mediated by the CTLA-4 or PD-1 pathway, whether and how TET2 acts on $CD8^+$ cytotoxic T lymphocytes, and coordination of PD-L1/CTLA-4 to regulate the anti-cancer $CD8^+$ T-cell response remain to be studied further. To the best of our knowledge, this is the first study to report a positive correlation between TET2, and immune infiltration in cancer-associated fibroblasts in BRCA, CESC, OV, and UCEC. However, detailed transcriptome or genome analyses are required to reveal more evidence on the potential association between TET2 and female cancer immunology. In this study, the histological analysis of BRCA tissues revealed some signals of TET2 localization in the cytoplasm. We speculated that such a result was observed because of non-specific signaling of TET2 antibodies or because some unexpected interstitial tissues were stained. However, it did not affect the conclusion.

Therefore, this study suggested that TET2 expression was significantly correlated with clinical prognosis, DNA methylation, gene mutation, and cancer immunology in female cancers. Our findings provide a relatively comprehensive understanding of the role of TET2 in the oncogenesis of female cancers.

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 Women's Hospital, School of Medicine, Zhejiang University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

WF, FC, and YF performed the analysis, prepared the figures, and wrote the manuscript; DC edited the manuscript.

ACKNOWLEDGMENTS

All authors collaborated in the collection and interpretation of the data and contributed to the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2022.790605/full#supplementary-material

REFERENCES

Bardou, P., Mariette, J., Escudié, F., Djemiel, C., and Klopp, C. (2014). Jvenn: An Interactive Venn Diagram Viewer. *BMC Bioinformatics* 15, 293. doi:10.1186/1471-2105-15-293

Cadet, J., and Wagner, J. R. (2014). TET Enzymatic Oxidation of 5-methylcytosine, 5-hydroxymethylcytosine and 5-formylcytosine. *Mutat. Research/Genetic Toxicol. Environ. Mutagenesis* 764-765, 18–35. doi:10.1016/j.mrgentox.2013.09.001

Carty, S. A., Gohil, M., Banks, L. B., Cotton, R. M., Johnson, M. E., Stelekati, E., et al. (2018). The Loss of TET2 Promotes CD8+ T Cell Memory Differentiation. *J. Immuno.* 200, 82–91. doi:10.4049/jimmunol.1700559

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

Chen, J.-Y., Luo, C.-W., Lai, Y.-S., Wu, C.-C., and Hung, W.-C. (2017). Lysine Demethylase KDM2A Inhibits TET2 to Promote DNA Methylation and Silencing of Tumor Suppressor Genes in Breast Cancer. *Oncogenesis* 6, e369. doi:10.1038/oncsis.2017.71

Chen, W., Zheng, R., Zhang, S., Zhao, P., Zeng, H., and Zou, X. (2014). Report of Cancer Incidence and Mortality in China, 2010. *Ann. Transl Med.* 2, 61. doi:10.3978/j.issn.2305-5839.2014.04.05

Chowdhury, B., Cho, I.-H., Hahn, N., and Irudayaraj, J. (2014). Quantification of 5-methylcytosine, 5-hydroxymethylcytosine and 5-carboxylcytosine from the

Blood of Cancer Patients by an Enzyme-Based Immunoassay. *Analytica Chim. Acta* 852, 212–217. doi:10.1016/j.aca.2014.09.020

Clough, E., and Barrett, T. (2016). The Gene Expression Omnibus Database. *Methods Mol. Biol.* 1418, 93–110. doi:10.1007/978-1-4939-3578-9_5

Deng, W., Wang, J., Zhang, J., Cai, J., Bai, Z., and Zhang, Z. (2016). TET2 Regulates LncRNA-ANRIL Expression and Inhibits the Growth of Human Gastric Cancer Cells. *IUBMB Life* 68, 355–364. doi:10.1002/iub.1490

Dennis, G., Jr., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., et al. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 4, P3. doi:10.1186/gb-2003-4-5-p3

Edwards, N. J., Oberti, M., Thangudu, R. R., Cai, S., McGarvey, P. B., Jacob, S., et al. (2015). The CPTAC Data Portal: A Resource for Cancer Proteomics Research. *J. Proteome Res.* 14, 2707–2713. doi:10.1021/pr501254j

Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Piñeros, M., et al. (2019). Estimating the Global Cancer Incidence and Mortality in 2018: GLOBOCAN Sources and Methods. *Int. J. Cancer* 144, 1941–1953. doi:10.1002/ijc.31937

Ganini, C., Amelio, I., Bertolo, R., Bove, P., Buonomo, O. C., Candi, E., et al. (2021). Global Mapping of Cancers: The Cancer Genome Atlas and beyond. *Mol. Oncol.* 15, 2823–2840. doi:10.1002/1878-0261.13056

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, pl1. doi:10.1126/scisignal.2004088

Godin, N., and Eichler, J. (2017). The Mitochondrial Protein Atlas: A Database of Experimentally Verified Information on the Human Mitochondrial Proteome. *J. Comput. Biol.* 24, 906–916. doi:10.1089/cmb.2017.0011

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

Győrffy, B., Lanczky, A., Eklund, A. C., Denkert, C., Budczies, J., Li, Q., et al. (2010). An Online Survival Analysis Tool to Rapidly Assess the Effect of 22,277 Genes on Breast Cancer Prognosis Using Microarray Data of 1,809 Patients. *Breast Cancer Res. Treat.* 123, 725–731. doi:10.1007/s10549-009-0674-9

He, Z., Wang, X., Huang, C., Gao, Y., Yang, C., Zeng, P., et al. (2018). The FENDRR/miR-214-3P/TET2 axis Affects Cell Malignant Activity via RASSF1A Methylation in Gastric Cancer. *Am. J. Transl Res.* 10, 3211–3223.

Huang, Y., Wang, G., Liang, Z., Yang, Y., Cui, L., and Liu, C.-Y. (2016). Loss of Nuclear Localization of TET2 in Colorectal Cancer. *Clin. Epigenet* 8, 9. doi:10.1186/s13148-016-0176-7

Jia, M., Li, Z., Pan, M., Tao, M., Wang, J., and Lu, X. (2020). LINC-PINT Suppresses the Aggressiveness of Thyroid Cancer by Downregulating miR-767-5p to Induce TET2 Expression. *Mol. Ther. - Nucleic Acids* 22, 319–328. doi:10.1016/j.omtn.2020.05.033

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

Koch, A., De Meyer, T., Jeschke, J., and Van Criekinge, W. (2015). MEXPRESS: Visualizing Expression, DNA Methylation and Clinical TCGA Data. *BMC Genomics* 16, 636. doi:10.1186/s12864-015-1847-z

Koch, A., Jeschke, J., Van Criekinge, W., van Engeland, M., and De Meyer, T. (2019). MEXPRESS Update 2019. *Nucleic Acids Res.* 47, W561–W565. doi:10.1093/nar/gkz445

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

Li, T., Fu, J., Zeng, Z., Cohen, D., Li, J., Chen, Q., et al. (2020). TIMER2.0 for Analysis of Tumor-Infiltrating Immune Cells. *Nucleic Acids Res.* 48, W509–W514. doi:10.1093/nar/gkaa407

Lin, S., Gao, K., Gu, S., You, L., Qian, S., Tang, M., et al. (2021). Worldwide Trends in Cervical Cancer Incidence and Mortality, with Predictions for the Next 15 Years. *Cancer* 127, 4030–4039. doi:10.1002/cncr.33795

Ma, H., Gao, W., Sun, X., and Wang, W. (2018). STAT5 and TET2 Cooperate to Regulate FOXP3-TSDR Demethylation in CD4+ T Cells of Patients with Colorectal Cancer. *J. Immunol. Res.* 2018, 1–8. doi:10.1155/2018/6985031

Modhukur, V., Iljasenko, T., Metsalu, T., Lokk, K., Laisk-Podar, T., and Vilo, J. (2018). MethSurv: a Web Tool to Perform Multivariable Survival Analysis Using DNA Methylation Data. *Epigenomics* 10, 277–288. doi:10.2217/epi-2017-0118

Nickerson, M. L., Das, S., Im, K. M., Turan, S., Berndt, S. I., Li, H., et al. (2017). TET2 Binds the Androgen Receptor and Loss Is Associated with Prostate Cancer. *Oncogene* 36, 2172–2183. doi:10.1038/onc.2016.376

Pfaffeneder, T., Spada, F., Wagner, M., Brandmayr, C., Laube, S. K., Eisen, D., et al. (2014). Tet Oxidizes Thymine to 5-hydroxymethyluracil in Mouse Embryonic Stem Cell DNA. *Nat. Chem. Biol.* 10, 574–581. doi:10.1038/nchembio.1532

Shen, Y., Liu, L., Wang, M., Xu, B., Lyu, R., Shi, Y. G., et al. (2021). TET2 Inhibits PD-L1 Gene Expression in Breast Cancer Cells through Histone Deacetylation. *Cancers* 13, 2207. doi:10.3390/cancers13092207

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

Tsutsumi, Y. (2021). Pitfalls and Caveats in Applying Chromogenic Immunostaining to Histopathological Diagnosis. *Cells* 10, 1501. doi:10.3390/cells10061501

Xu, Y. (2020). TET2 Expedites Coronary Heart Disease by Promoting microRNA-126 Expression and Inhibiting the E2F3-PI3K-AKT axis. *Biochem. Cell Biol.* 98, 698–708. doi:10.1139/bcb-2020-0297

Yan, H., Wang, Y., Qu, X., Li, J., Hale, J., Huang, Y., et al. (2017). Distinct Roles for TET Family Proteins in Regulating Human Erythropoiesis. *Blood* 129, 2002–2012. doi:10.1182/blood-2016-08-736587

Yang, L., Yu, S.-J., Hong, Q., Yang, Y., and Shao, Z.-M. (2015). Reduced Expression of TET1, TET2, TET3 and TDG mRNAs Are Associated with Poor Prognosis of Patients with Early Breast Cancer. *PLoS One* 10, e0133896. doi:10.1371/journal.pone.0133896

Yue, X., and Rao, A. (2020). TET Family Dioxygenases and the TET Activator Vitamin C in Immune Responses and Cancer. *Blood* 136, 1394–1401. doi:10.1182/blood.2019004158

Zhang, C., Zhao, N., Zhang, X., Xiao, J., Li, J., Lv, D., et al. (2021). SurvivalMeth: a Web Server to Investigate the Effect of DNA Methylation-Related Functional Elements on Prognosis. *Brief Bioinform* 22. doi:10.1093/bib/bbaa162

Zhang, J., Tan, P., Guo, L., Gong, J., Ma, J., Li, J., et al. (2019). p53-dependent Autophagic Degradation of TET2 Modulates Cancer Therapeutic Resistance. *Oncogene* 38, 1905–1919. doi:10.1038/s41388-018-0524-5

Zhang, L.-y., Li, P.-l., Wang, T.-z., and Zhang, X.-c. (2015). Prognostic Values of 5-hmC, 5-mC and TET2 in Epithelial Ovarian Cancer. *Arch. Gynecol. Obstet.* 292, 891–897. doi:10.1007/s00404-015-3704-3

Zhang, L., Chen, W., Iyer, L. M., Hu, J., Wang, G., Fu, Y., et al. (2014). A TET Homologue Protein from Coprinopsis Cinerea (CcTET) that Biochemically Converts 5-methylcytosine to 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine. *J. Am. Chem. Soc.* 136, 4801–4804. doi:10.1021/ja500979k

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

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 Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 14 February 2022
doi: 10.3389/fbioe.2022.820092

Molecular Patterns Based on Immunogenomic Signatures Stratify the Prognosis of Colon Cancer

Cong Shen[1,2†], Cong Luo[2,3†], Zhijie Xu[2,4], Qiuju Liang[5], Yuan Cai[4], Bi Peng[4], Yuanliang Yan[5] and Fada Xia[1]**

[1]Department of Thyroid Surgery, Xiangya Hospital, Central South University, Changsha, China, [2]National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China, [3]Department of Urology, Xiangya Hospital, Central South University, Changsha, China, [4]Department of Pathology, Xiangya Hospital, Central South University, Changsha, China, [5]Department of Pharmacy, Xiangya Hospital, Central South University, Changsha, China

Background: Colon cancer is an aggressive and heterogeneous disease associated with high morbidity and mortality. The immune system is intimately involved in tumorigenesis and can influence malignant properties at the protein, epigenetic, and even genomic levels by shaping the tumor immune microenvironment (TIM). However, immune-related molecules that can effectively predict the prognosis of colon cancer remain under exploration.

Methods: A total of 606 patients from TCGA and GEO databases were employed in our study, in which 429 cases were set as the training cohort and 177 were defined as the validation cohort. The immune infiltration was evaluated by ESTIMATE, TIMER, and CIBERSORT algorithms. The risk signature was constructed by LASSO Cox regression analysis. A nomogram model was generated subsequent to the multivariate Cox proportional hazards analysis to predict 1-, 3-, and 5-year survival of patients with colon cancer.

Results: Infiltrating immune cell profiling identified two colon cancer clusters (Immunity_L group and Immunity_H group). The abundances of immune cells were higher in the Immunity_H group, which indicated a better prognosis. Through further statistical analysis, we identified four genes which were highly correlated with prognosis and representative of this gene set, namely *ARL4C*, *SERPINE1*, *BST2*, and *AXIN2.* When the patients were divided into low- and high-risk groups based on their risk scores, we found that patients in the high-risk group had shorter overall survival time. Moreover, a nomogram including clinicopathologic features and the established risk signature could robustly predict 1-, 3-, and 5-year survival in patients with colon cancer.

Conclusion: We identified two distinct immune patterns by analyzing clinical and transcriptomic information from colon cancer patients. A subsequently constructed immune-related gene-based prognostic model as well as a nomogram model can be used to predict the prognosis of colon cancer, thereby guiding risk stratification and treatment regimen development for colon patients.

Keywords: immune, colon cancer, signature, risk score, prognosis

OPEN ACCESS

Edited by:
Giulia Russo,
University of Catania, Italy

Reviewed by:
Debashis Sahoo,
University of California, San Diego, United States
Qian Chen,
Guangxi Medical University Cancer Hospital, China

****Correspondence:***
Yuanliang Yan
yanyuanliang@csu.edu.cn
Fada Xia
xiafada@csu.edu.cn

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

Specialty section:
This article was submitted to Preclinical Cell and Gene Therapy, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *22 November 2021*
Accepted: *11 January 2022*
Published: *14 February 2022*

Citation:
Shen C, Luo C, Xu Z, Liang Q, Cai Y, Peng B, Yan Y and Xia F (2022) Molecular Patterns Based on Immunogenomic Signatures Stratify the Prognosis of Colon Cancer. Front. Bioeng. Biotechnol. 10:820092. doi: 10.3389/fbioe.2022.820092

INTRODUCTION

Recent cancer statistics reveals a high incidence of 10.2% for human colon cancer, while the mortality rate is up to 9.2%, rising from the fourth to the second place in the oncological field, seriously threatening human life health (Bray et al., 2018; Siegel et al., 2020a). Surgical removal of cancerous tissue, combined with radiotherapy and chemotherapy (if necessary), has been the mainstay of combating colon malignancies (Nie et al., 2020; Liao et al., 2021). Due to advances in systemic drug targeting and surgical techniques, the prognosis of colon cancer patients will be significantly improved if they are diagnosed at an early stage (Labianca et al., 2010). Therefore, accurate grading/staging of colon cancer is helpful for the development of treatment options as well as the prognosis of patients.

Current prognostic prediction mainly relies on the Tumor, Nodes, Metastases (TNM) classification system (clinical level) and histopathological criteria (histological level) (Pagès et al., 2018). Unfortunately, patients still demonstrated an unpleasing survival outcome due to the recurrence, metastasis, and resistance to the agents. Clinicians and researchers have been searching for novel treatment strategies in the hope for achieving better results, and they focused on the cellular and molecular levels to identify valuable markers and tumor-cell differentiation events. A growing number of studies have demonstrated the role of gene mutation status, gene expression levels, and signaling pathway alterations in tumor initiation and progression, but accurately identifying prognostic factors that can provide targets for therapy remains difficult at the genomic level (Beane et al., 2009; Nagy et al., 2018). The immune system is intimately involved in tumorigenesis and can influence malignant properties at the protein, epigenetic, and even genomic levels by shaping the tumor immune microenvironment (TIM) (Deng et al., 2021). Recently, a large body of evidence has shown that immune-related molecules are of great value in predicting prognosis and assessing therapeutic efficacy (Binnewies et al., 2018).

The TIM is a complex system composed of multiple immune cells infiltrating into tumor tissue and various cytokines and chemokines secreted by them (Mallmann-Gottschalk et al., 2019). In there, natural killer (NK) cells can secrete cytokines, such as

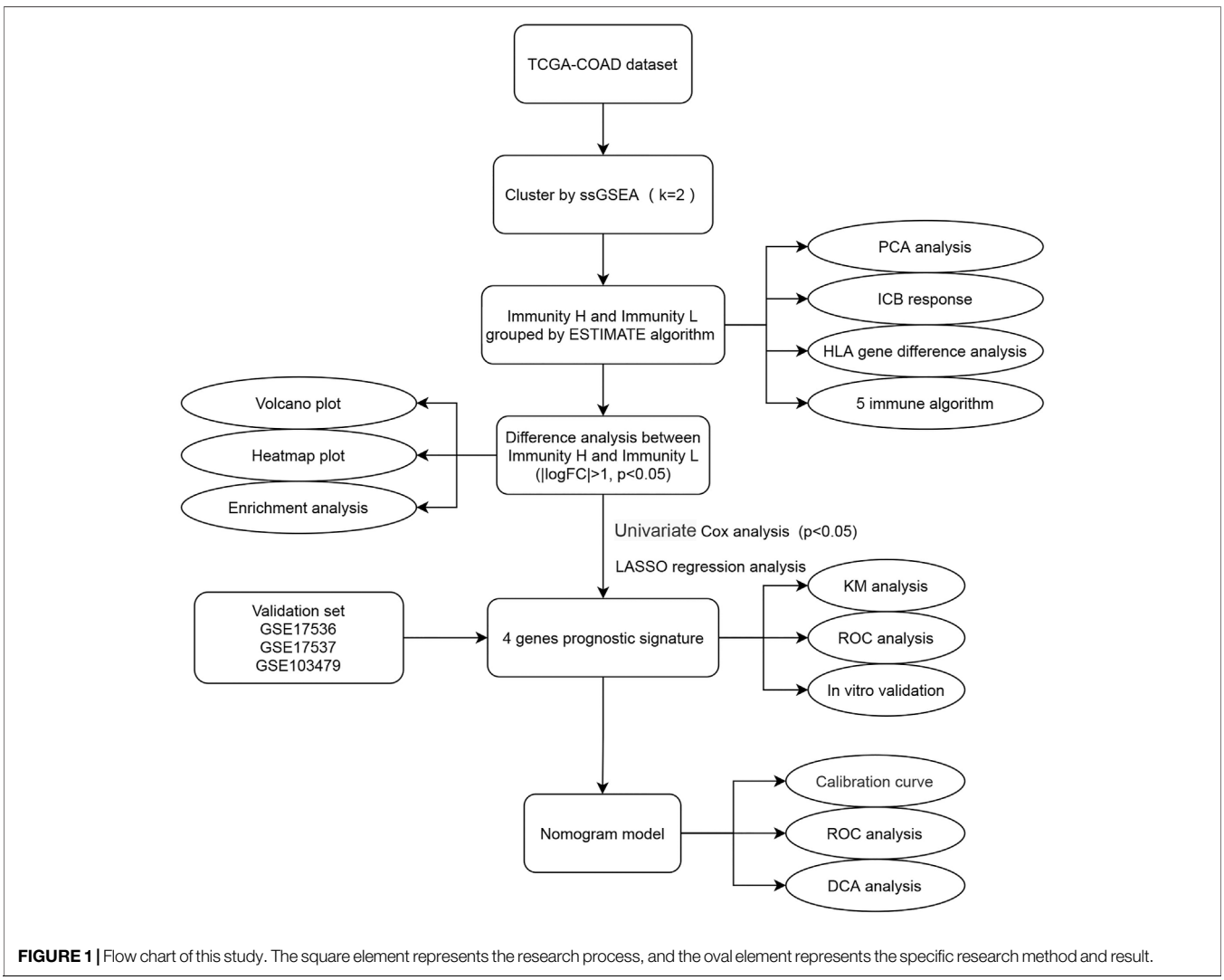


FIGURE 1 | Flow chart of this study. The square element represents the research process, and the oval element represents the specific research method and result.

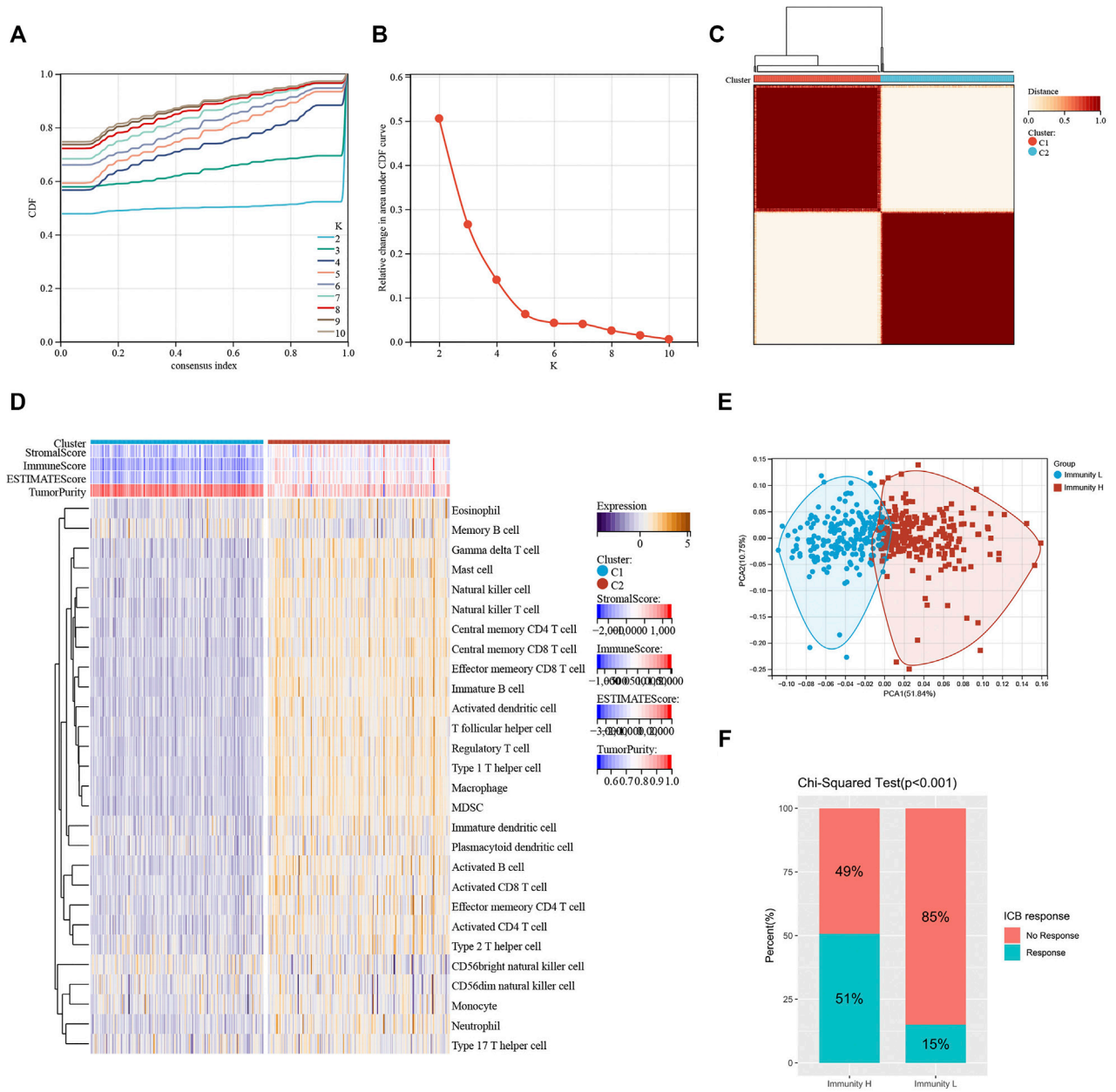


FIGURE 2 | Two patterns of colon cancer based on immune cell clustering. **(A)** Cumulative distribution function (CDF) under different clustering numbers. **(B)** Relative change in area under CDF curve with different clustering numbers. **(C)** Clustering heatmap when $k = 2$. **(D)** Heatmap showing the differences in immune cell ssGSEA scores between two clusters. **(E)** PCA regarding two clusters. **(F)** The differences in response rates to immune checkpoint blockades between two clusters. **(G)** The expressions of HLA genes between two clusters.

interferon (IFN)-gamma and tumor necrosis factor (TNF)-α, to exert an immunosuppressive phenotype *via* inhibiting tumor cell proliferation and tumor angiogenesis (Liu et al., 2018). Moreover, tumor associated macrophages (TAMs) and regulatory T cells (Tregs) mediate a suppressed tumor microenvironment which helps tumor cells achieve immune escape, and promotes the development of malignancy (Pan et al., 2020).

To address these suppressive immune phenotypes, targeting immune-tumor cell interactions has become more intensively studied, and immunotherapy has emerged as a promising area of cancer treatment and has demonstrated its impressive clinical value in colon cancer. This is mostly attributed to the face of immune checkpoint inhibitors as antitumor agents, such as programmed death 1 receptor (PD1) and cytotoxic T lymphocyte antigen 4 (CTLA-4) inhibitors (Yaghoubi et al.,

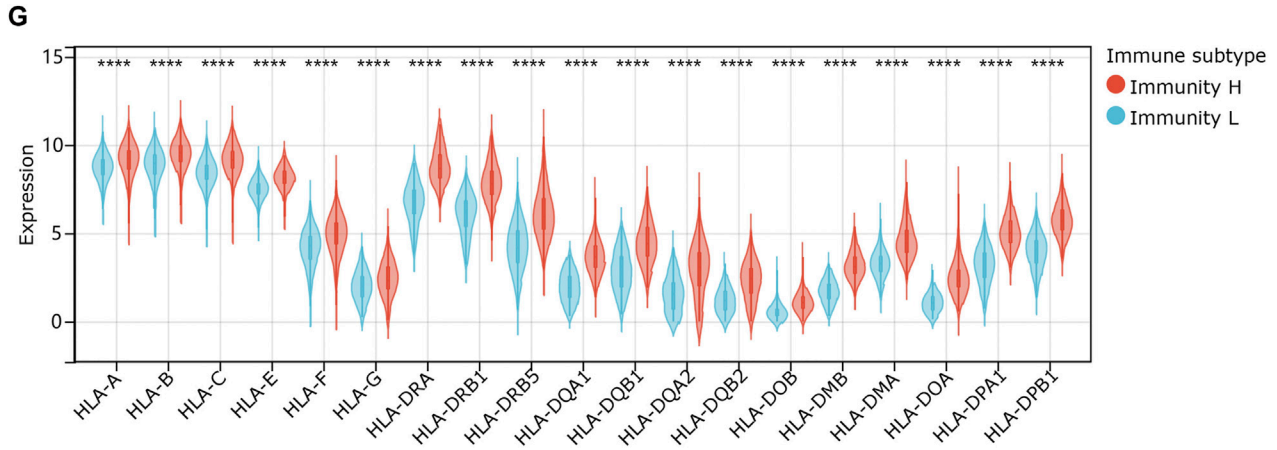


FIGURE 2 | (continue)

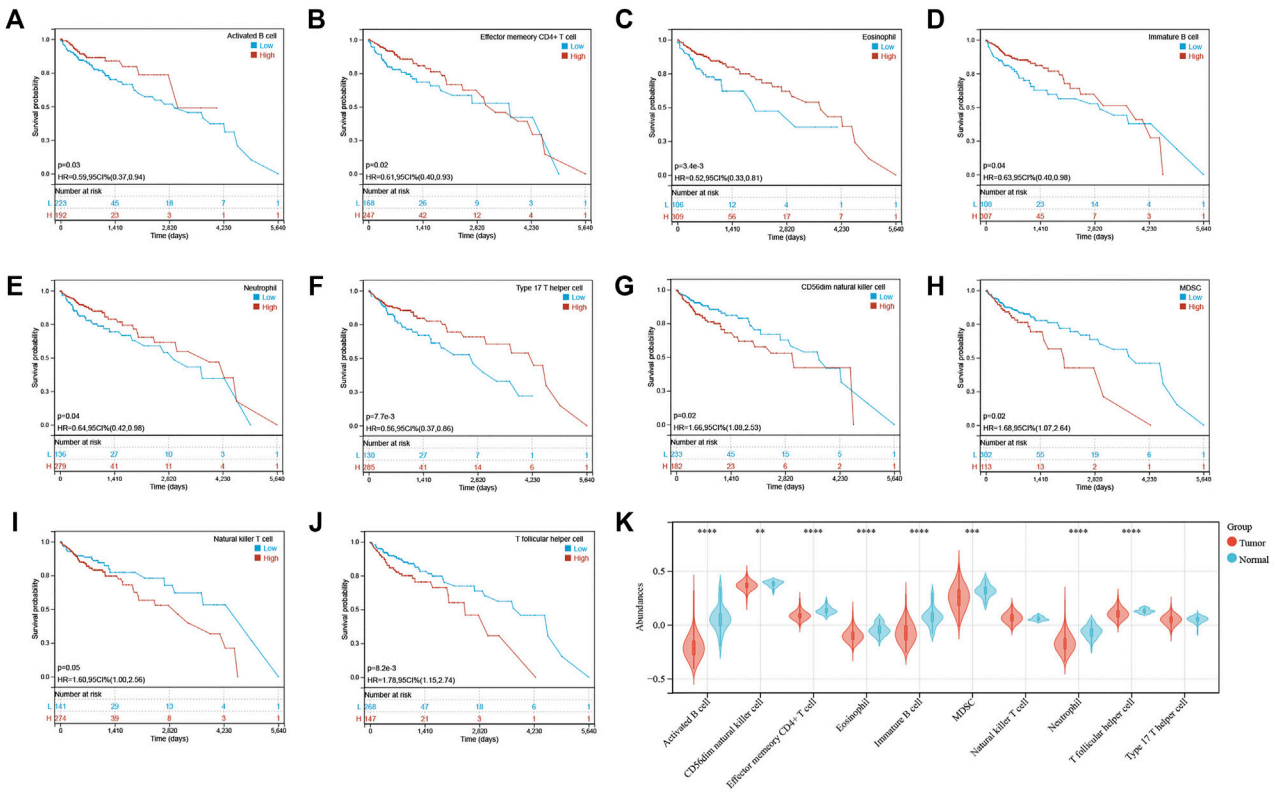


FIGURE 3 | (A–J) Kaplan–Meier survival curves of individual cell types with significant prognostic values. **(K)** Differences in abundance of 10 immune cells between normal and tumor tissues. **(L)** Immune cell abundances estimated by five algorithms (CIBERSORT, EPIC, IPS, MCPcounter, and TIMER) between two clusters.

2019). However, immunotherapy for colon cancer is still imperfect, for example, the effect of immunotherapy cannot be evaluated in advance. In conclusion, immunogenomic classification will help to guide the identification and effective treatment of early colon cancer and improve the accuracy of prognosis evaluation.

In this study, we performed an immunogenomic profiling of patients with colon cancer and divided them into two distinct subtypes: high immunity (Immunity_H) and low immunity (Immunity_L). We focused on analyzing two independent cohorts of colon cancer patients to identify genes highly associated with prognosis; based on their expression levels, the

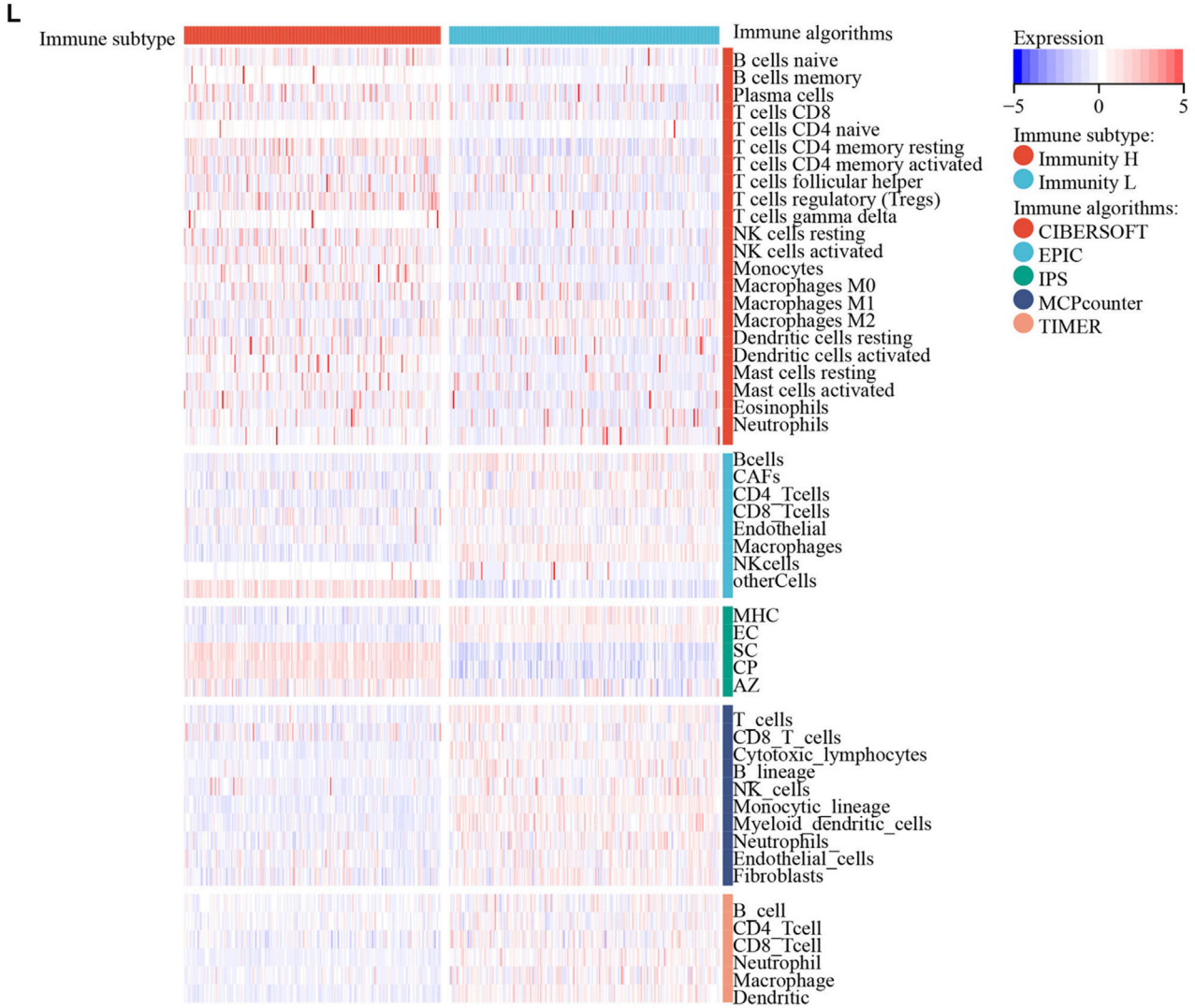


FIGURE 3 | (continue)

patients were allocated risk scores. We then combined clinicopathological characteristics and the risk scores to establish a model to accurately predict the survival rate of patients with colon cancer. This analysis is of great significance for the survival prediction of patients with colon cancer and provides a potential target for its treatment.

MATERIALS AND METHODS

Data Source and Extraction

The data we used for analysis were obtained from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/) (The Cancer Genome Atlas Network, 2012). In the dataset, 429 colon adenocarcinoma (COAD) patients had complete clinicopathological and transcriptomic expression data. Thus, they were enrolled as the training set. In addition, we also downloaded three datasets, including GSE17536 (n = 177), GSE17537 (n = 55), and GSE103479 (n = 156) in the Gene Expression Omnibus (GEO, https:/www.ncbi.nlm.nih.gov/geo/) to validate the results (Smith et al., 2010; Dienstmann et al., 2019). The crude RNA expression data Fragments Per Kilobase of exon model per Million mapped fragments (FPKM) were transformed into Transcripts Per Kilobase of exon model per Million mapped reads (TPM) for a better statistical evaluation.

Clustering for Distinct Immune Patterns in Colon Cancer

We screened and quantified 28 immune cell types in the TCGA-COAD dataset using the single sample gene set enrichment analysis (ssGSEA) algorithm (Barbie et al., 2009). For each independent dataset, an enrichment score was calculated to represent the enrichment level of 28 immune cell types for

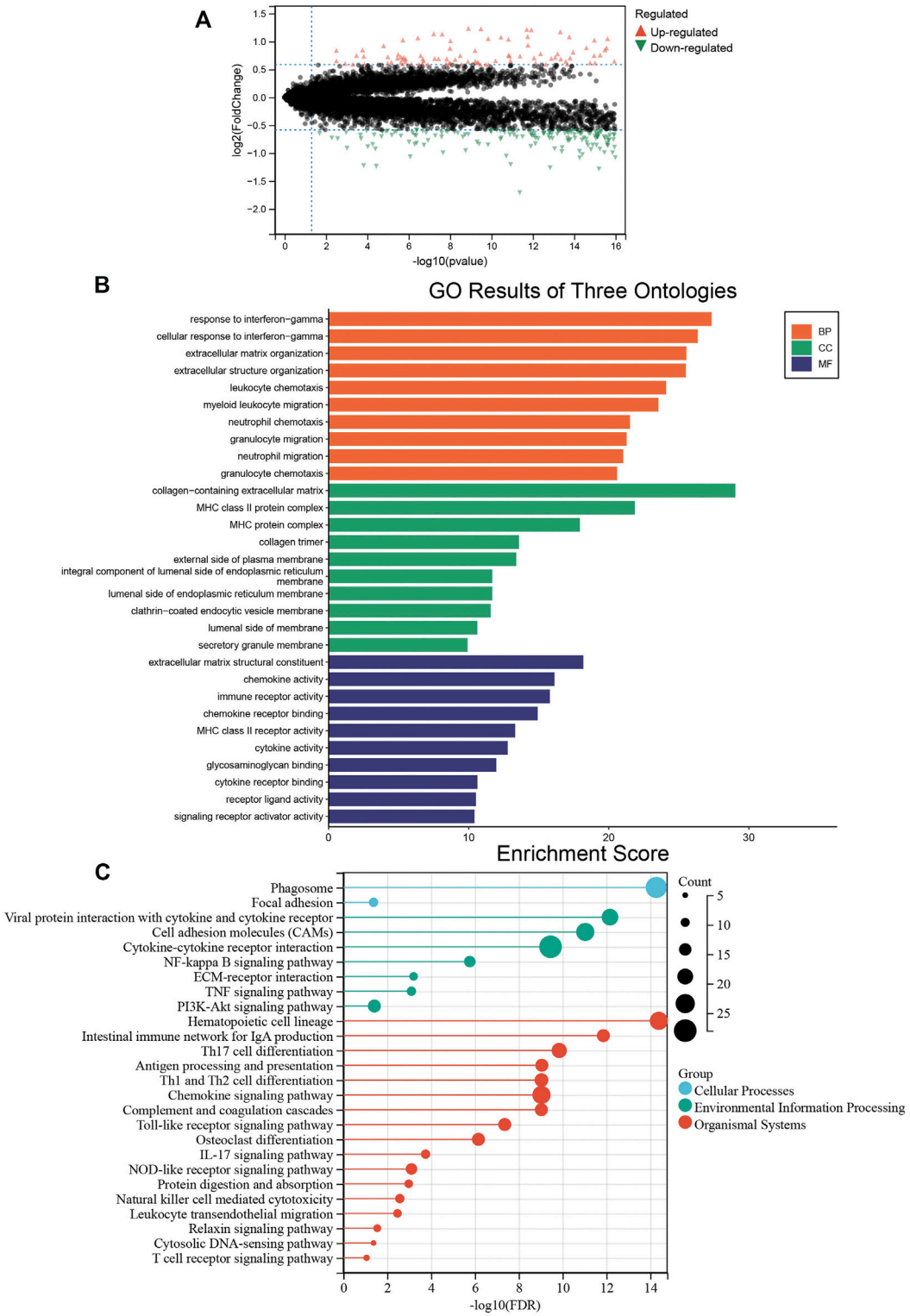


FIGURE 4 | The biological functions of differentially expressed genes between two clusters. **(A)** The differentially expressed genes between the Immunity_L and Immunity_H groups. Red triangles represent up-regulated genes and green triangles represent down-regulated genes. **(B)** Enrichment results of GO terms. Orange: biological process, green: cellular component, purple: molecular function. **(C)** Enrichment results of KEGG pathways. Blue: cellular process, green: environmental information processing, red: organismal systems.

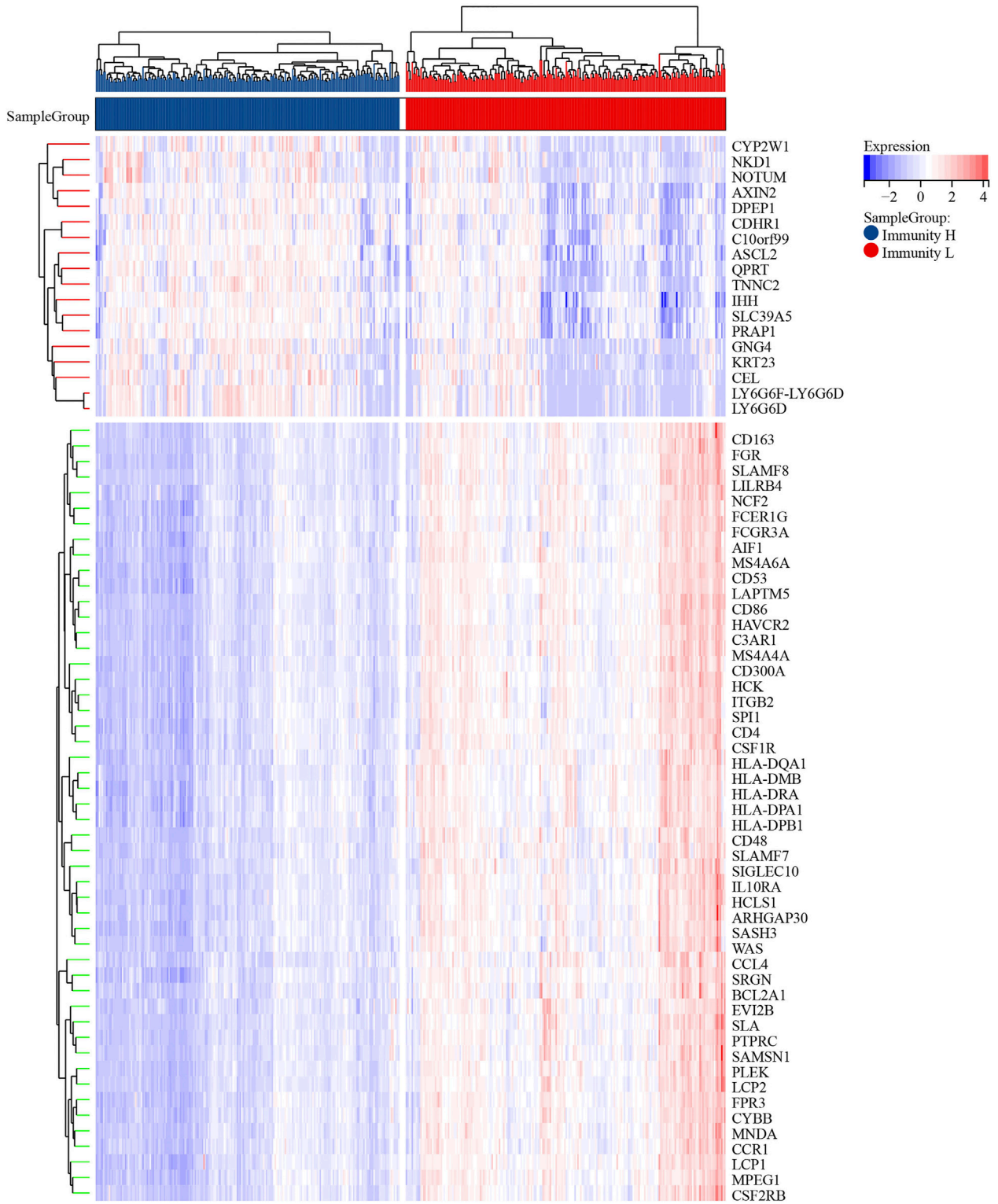


FIGURE 5 | The heatmap of differentially expressed genes.

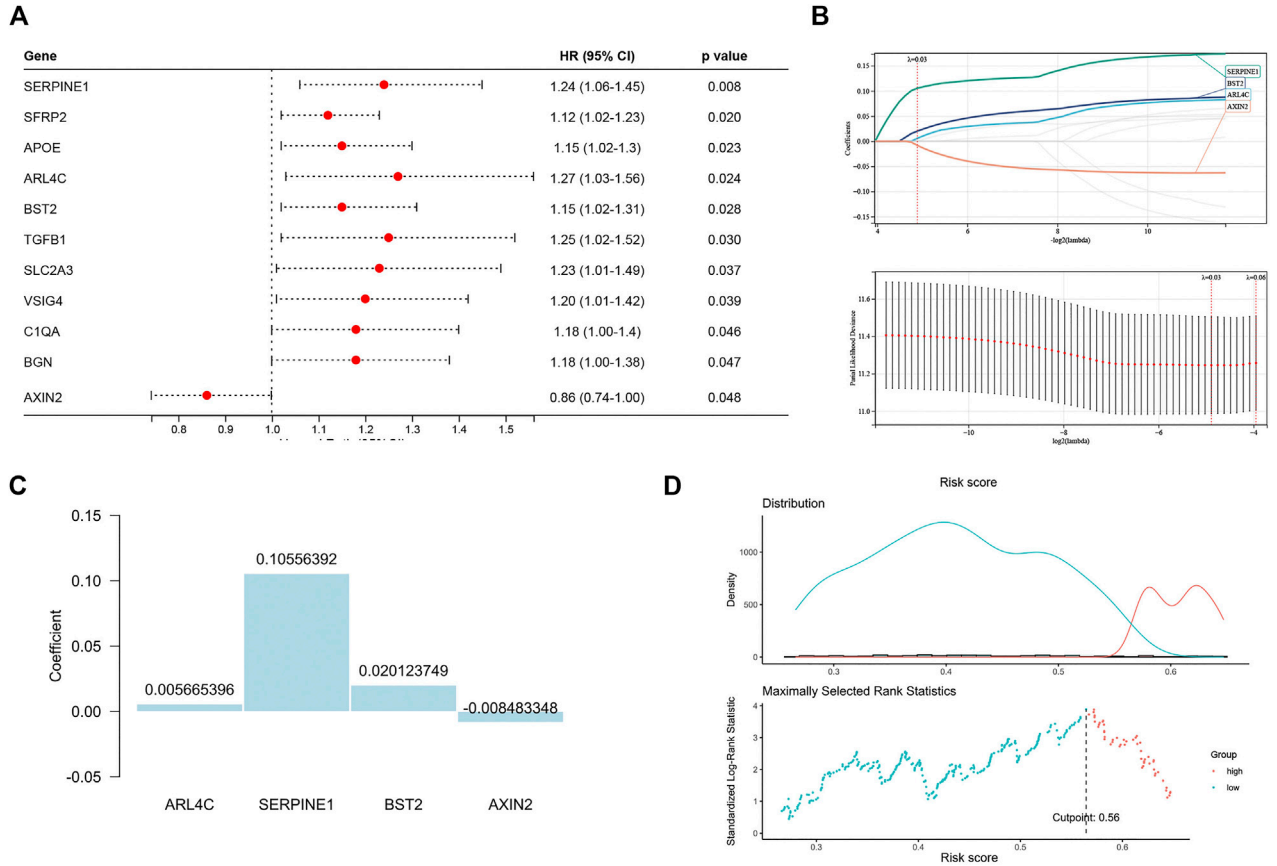


FIGURE 6 | Construction of risk signature based on differentially expressed genes. **(A)** The forest plot of differentially expressed genes with prognostic significance. **(B)** The screening of coefficients and likelihood deviance under LASSO analysis. **(C)** The coefficients of model genes. **(D)** Best cutoff value selected by the log-rank test. **(E)** The relationship between risk score, survival status, and model gene expression. **(F)** The Kaplan–Meier survival curve showing the difference between *high risk* and *low risk* groups in TCGA-COAD. **(G)** The ROC curves regarding 1-, 3-, and 5-year survival outcomes in TCGA-COAD. **(H)** The differences in response rates to immune checkpoint blockades between two groups.

each tumor sample. Based on these ssGSEA scores, we performed consensus clustering on TCGA-COAD. In brief, cluster analysis was performed using "ConsensusClusterPlus" (Wilkerson and Hayes, 2010), using agglomerative k-means clustering with a 1-Pearson correlation distance and resampling 80% of the samples for 10 repetitions. The optimal number of clusters was determined using the empirical cumulative distribution function plot. A principal component analysis (PCA) was conducted to analyze the distinguishing ability of the clustering.

Quantification of Immunotherapy Response, Major Histocompatibility Complex, and Immune Microenvironment

We used the "ESTIMATE" package to calculate the Immune, Stromal, and ESTIMATE scores for each sample to show the component fractions and tumor purity (Yoshihara et al., 2013). ImmuCellAI (Immune Cell Abundance Identifier) is a tool to estimate the abundance of 24 immune cells from gene expression dataset including RNA-Seq and microarray data (Miao et al., 2020). We applied it to predict patients' response to immune checkpoint blockade therapy. The gene expression of the major histocompatibility complex (MHC), human leukocyte antigen (HLA), has also been explored in different clusters. Moreover, the abundances of diverse immune and stromal components were calculated using the "IOBR" R package (https://github.com/IOBR/IOBR), which is designed for multi-omics immuno-oncology biological research to decode tumor microenvironment and signatures (Zeng et al., 2021). Specifically, five algorithms built into the tool, including CIBERSORT (Newman et al., 2015), immunophenoscore (IPS) (Charoentong et al., 2017), MCP-counter (Becht et al., 2016), xCell (Aran et al., 2017), and EPIC (Racle et al., 2017), were used to calculate the scores of 51 infiltrating (immune and stromal) cells in each sample.

Profiling of the Differentially Expressed Prognostic Genes Related to Colon Cancer Subtype-specific Immunity

We performed difference analysis between the Immunity_Low (Immunity_L) group and Immunity_High (Immunity_H) group using the "limma" package (Ritchie et al., 2015). The absolute

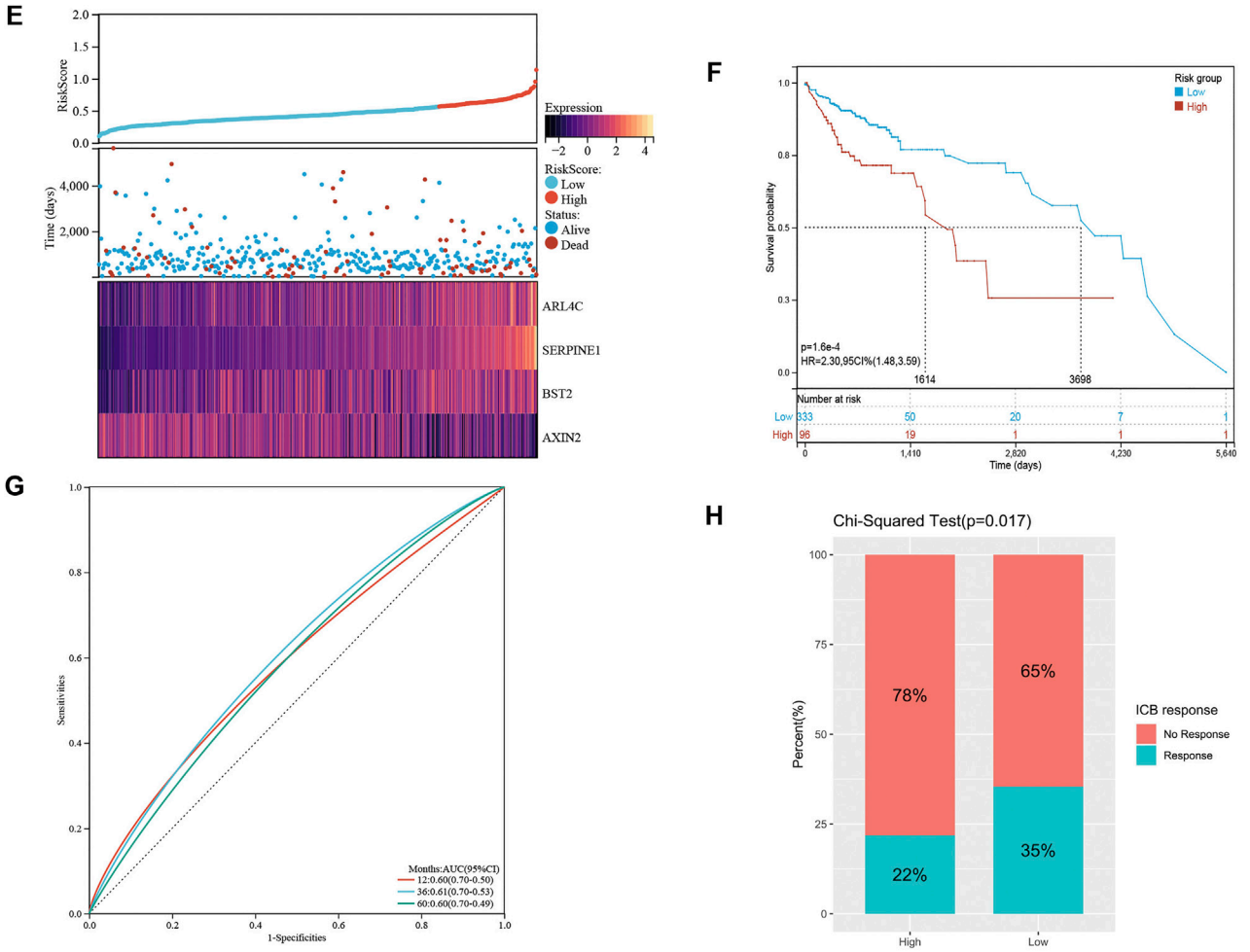


FIGURE 6 | (continue)

values of differential expression multiples >1.5 and $p < 0.05$ were used as the criteria for screening differentially expressed genes. The up- and down-regulated immune related genes (IRGs) in colon cancer were displayed in volcano plot. For functional analysis, we used the "org.Hs.eg.db" package (version 3.1.0) to perform the gene ontology (GO; including biological process, molecular function, and cellular component) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotation. A Benjamini–Hochberg false discovery rate (BH-FDR) <0.05 was considered statistically significant. The enrichment results were displayed as histogram and lollipop charts.

Identification of Prognostic Genes and Construction of Immune Signature

Univariate Cox regression analysis based on differentially expressed genes was used to screen immune related genes significantly related to the prognosis of colon cancer with $p < 0.05$ as the threshold. Subsequently, a Least Absolute Shrinkage and Selection Operator (LASSO) Cox regression analysis with "glmnet" package (Tibshirani, 1996) was used to further identify essential genes and allocate coefficients for them. The risk score of each sample was calculated using the following formula:

$$\text{Risk score} = \sum_{i=1}^{n} Coef i * xi$$

where $Coef_i$ is the risk coefficient of each factor and x_i is the mRNA expression value (logarithmic transformed TPM) of each factor. After determining the optimal cut-off value of risk score through "survival" and "survminer" packages, patients were divided into *low risk* and *high risk* groups correspondingly. Survival curves were used to show the differences in survival time and survival probability between *high risk* and *low risk* patients based on the Kaplan–Meier method. The area under curve (AUC) of receiver operating characteristic (ROC) curves represented the predictive accuracy (Heagerty et al., 2000). Univariate and multivariate Cox regression models were used to analyze whether the risk score was able to independently predict survival in patients with colon cancer.

TABLE 1 | Clinicopathological characteristics between the *low risk* and *high risk* groups

Characteristics	Total (N = 429)	*Low risk* (N = 333)	*High risk* (N = 96)	p Value[a]
Age				0.59
Mean ± SD	66.70 ± 12.77	66.52 ± 12.75	67.31 ± 12.88	
Median [min-max]	69.00 [31.00,90.00]	68.00 [31.00,90.00]	69.00 [34.00,89.00]	
Gender				0.87
Female	202 (47.09%)	158 (36.83%)	44 (10.26%)	
Male	227 (52.91%)	175 (40.79%)	52 (12.12%)	
AJCC stage				<0.01
Stage I	74 (17.25%)	68 (15.85%)	6 (1.40%)	
Stage II	170 (39.63%)	131 (30.54%)	39 (9.09%)	
Stage III	123 (28.67%)	89 (20.75%)	34 (7.93%)	
Stage IV	62 (14.45%)	45 (10.49%)	17 (3.96%)	
T stage				<0.01
T1	9 (2.10%)	9 (2.10%)	0 (0.0e + 0%)	
T2	75 (17.48%)	69 (16.08%)	6 (1.40%)	
T3	297 (69.23%)	221 (51.52%)	76 (17.72%)	
T4	48 (11.19%)	34 (7.93%)	14 (3.26%)	
N stage				0.01
N0	253 (58.97%)	208 (48.48%)	45 (10.49%)	
N1	99 (23.08%)	73 (17.02%)	26 (6.06%)	
N2	77 (17.95%)	52 (12.12%)	25 (5.83%)	
M stage				0.39
M0	367 (85.55%)	288 (67.13%)	79 (18.41%)	
M1	62 (14.45%)	45 (10.49%)	17 (3.96%)	
Disease type				0.35
Adenocarcinoma	372 (86.71%)	292 (68.07%)	80 (18.65%)	
Mucinous adenocarcinoma	57 (13.29%)	41 (9.56%)	16 (3.73%)	
Survival status				<0.01
Alive	340 (79.25%)	276 (64.34%)	64 (14.92%)	
Dead	89 (20.75%)	57 (13.29%)	32 (7.46%)	

[a]*p Value between the* low risk *group and* high risk *group.*
AJCC, American Joint Committee on Cancer.

Establishment of Nomogram Prognosis Prediction Model

Nomograms are widely used to predict the prognosis of the disease, so we drew a nomogram based on all independent prognostic factors identified by multivariate Cox regression analysis to predict the survival probability of patients within 1, 3, and 5 years by using the "rms" package. For practical application, we created a dynamic nomogram through "DynNom" package and built an interactive web-based tool with Shiny (https://shiny.rstudio.com/) (**Supplementary Material S1**). A nomogram calibration curve was plotted to judge nomogram accuracy by observing the relationship between predicted probability and actual incidence. The practicability of 1-, 3-, and 5-years OS was evaluated by ROC curves. The prognostic ability of the nomogram and other predictors (risk score, N stage, and M stage) for survival were evaluated by decision curve analysis (DCA) curves using the "rmda" R package.

Cell Culture

Normal human colon epithelial cells (NCM-460) and two human colon cancer cell lines (HCT116, HCT8) were obtained from the cell bank at the Chinese Academy of Sciences (Shanghai, China). All cells were authenticated by short tandem repeat (STR) profiling upon receipt and were propagated for less than 6 months after resuscitation. All cell lines were cultured in Dulbecco's modified eagle medium (DMEM) medium with 10% fetal bovine serum (FBS; Thermo Fisher Scientific, Waltham, MA, USA). These cell lines were maintained in a humidified chamber containing 5% CO_2 at 37°C.

RNA Extraction and qPCR

Total RNA from cultured cells and fresh tissues was extracted with Trizol regent (Thermo Fisher Scientific). We used NanoDrop and an Agilent 2,100 bioanalyzer (Thermo Fisher Scientific) to determine the concentration of extracted total RNA. cDNA was obtained by reverse transcription using a reverse transcription kit (Hiscrip II Q RT SuperMix for qPCR; Vazyme, Nanjing, China) according to the manufacturer's protocol. Quantitative real-time polymerase chain reaction amplification was performed with SYBR Green PCR master mix (Takara, Japan) according to the manufacturer's protocol. Primers were designed as follows: SERPINE1, forward, 5′-AGT GGACTTTTCAGAGGTGGA-3′, reverse, 5′-GCCGTTGAA GTAGAGGGCATT-3′; ARL4C, forward, 5′-CCAGTCCCT GCATATCGTCAT-3′, reverse, 5′-TTCACGAACTCGTTG AACTTGA-3′; BST2, forward, 5′-CACACTGTGATGGCC CTAATG-3′, reverse, 5′-GTCCGCGATTCTCACGCTT-3′; AXIN2, forward, 5′-TACACTCCTTATTGGGCGATCA-3′, reverse, 5′-TTGGCTACTCGTAAAGTTTTGGT-3′. GAPDH

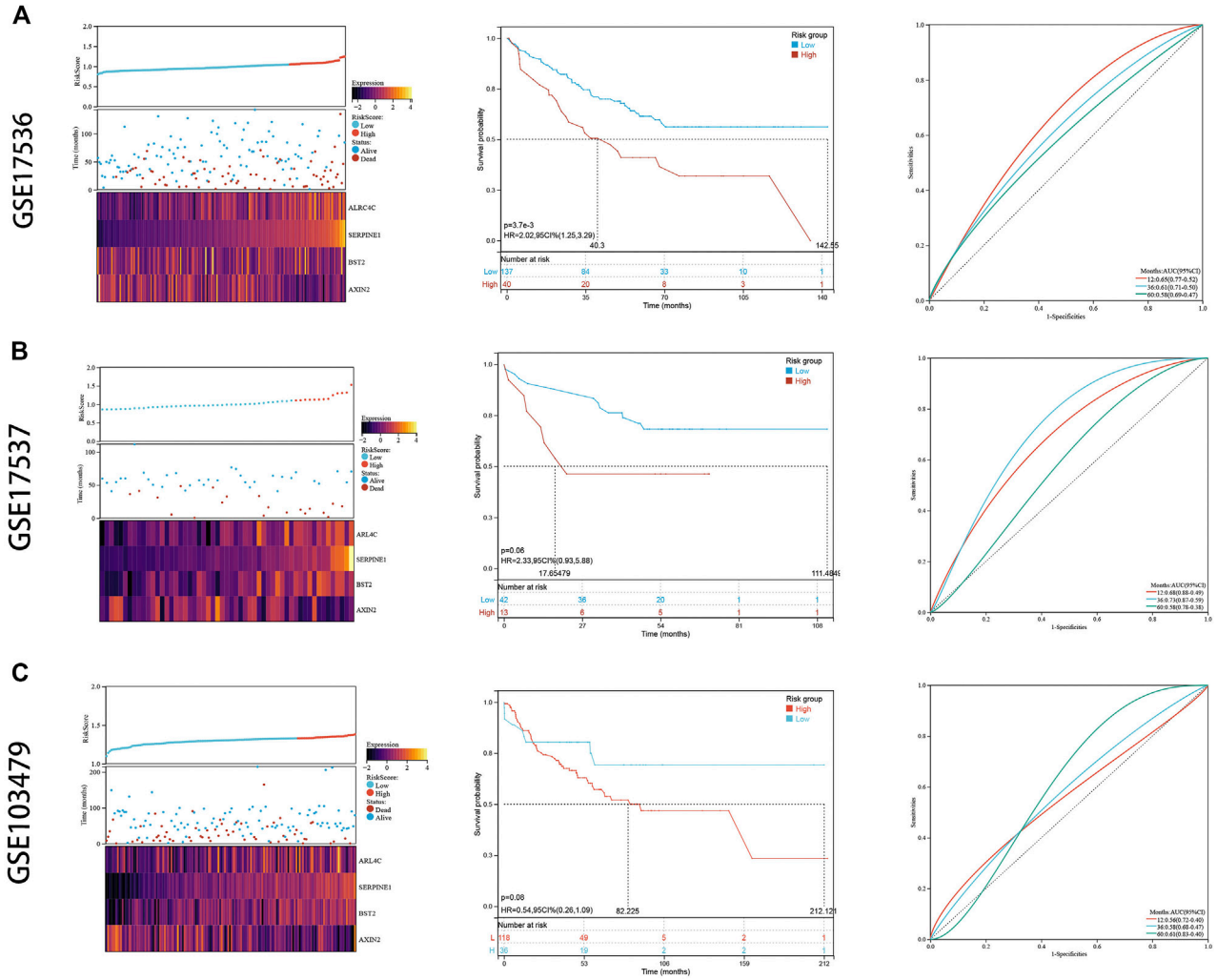


FIGURE 7 | External validation. **(A)** The risk distribution, the Kaplan–Meier survival curve, and the ROC curves in GSE17536. **(B)** The risk distribution, the Kaplan–Meier survival curve, and the ROC curves in GSE17537. **(C)** The risk distribution, the Kaplan–Meier survival curve, and the ROC curves in GSE103479.

was used as an endogenous control, and relative gene expression was determined by the comparative $2^{-\Delta\Delta CT}$ method.

Immunohistochemistry

The Human Protein Atlas (HPA) (https://www.proteinatlas.org/) is a program for mapping human proteins in cells, tissues, and organs using integration of various omics technologies (Uhlen et al., 2017; Sun et al., 2018). We obtained representative immunohistochemistry results of the four target proteins in colon cancer and normal colon tissues from the tissue atlas and pathology atlas in HPA database, respectively.

Statistical Analysis

In this study, Kaplan–Meier method was used to estimate the overall survival rate of different groups, and log-rank was used to test the significance. The inter-group comparisons were achieved by using Wilcoxon rank sum test. The chi-square test was used to compare the clinicopathologic features (age, gender, TNM stage, and AJCC stage) between the *low risk* group and *high risk* group. Univariate and multivariate Cox regression analyses were utilized to evaluate the independent prognostic value of the risk signature regarding OS. Statistical analyses were made using R software (version 4.0.3). Most visualizations were achieved by "ggplot2" package. In most situations, $p < 0.05$ was used as a significant threshold if not otherwise specified **Figure 1**.

RESULTS

Infiltrating Immune Cell Profiling Identified Two Colon Cancer Clusters

We screened and analyzed 28 immune cell types by ssGSEA for every tumor sample (**Supplementary Table S2**). Then, we performed consensus clustering on TCGA-COAD dataset based on the ssGSEA scores, which represented the activity or infiltration levels of immune cells in the tumor sample. Finally,

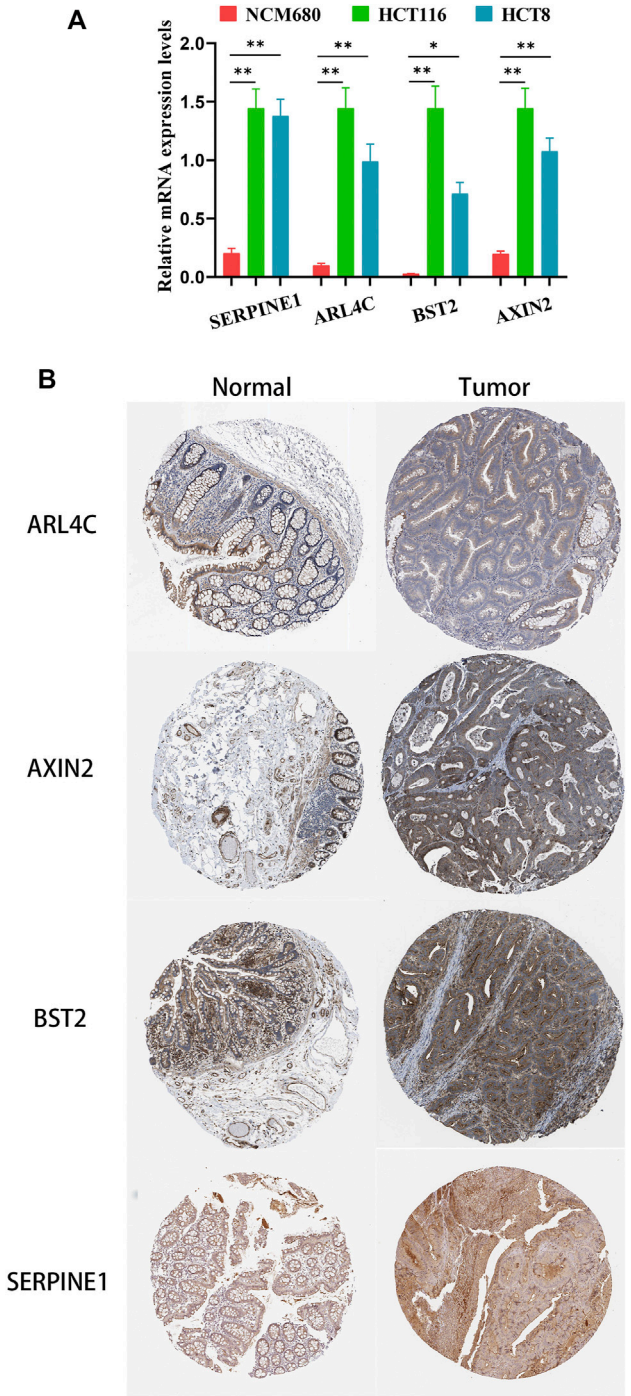


FIGURE 8 | The expression of these four genes in colon cancer cells and tissues. **(A)** The relative mRNA levels of the four genes in normal colon epithelial cells (NCM-460) and colon cancer cells (HCT116, HCT8) by qPCR. **(B)** Representative IHC images of the four genes in colon cancer and normal colon tissues.

the patients were divided into two unique immune clusters (C1 and C2; **Figures 2A–C**). We found that the ssGSEA scores of all the immune cell types were higher in C2 than those in C1 (**Figure 2D**). The C1 and C2 were thus defined as the Immunity_Low (Immunity_L) group and Immunity_High (Immunity_H) group, respectively (**Supplementary Table S3**). Furthermore, the heatmap clearly demonstrated that the Immunity_H group possessed higher stromal scores, immune scores, and ESTIMATE scores when compared with the Immunity_L group (**Figure 2D**). Correspondingly, the Immunity_H group exhibited lower tumor purity.

PCA result further confirmed the reliability of this clustering, as the subgroups could be significantly distinguished (**Figure 2E**). More importantly, the Immunity_H group showed a higher response rate to immune checkpoint blockers than the Immunity_L group (51 vs. 15%; $p < 0.001$; **Figure 2F**). At the same time, we found that most HLA genes were expressed at higher levels in the Immunity_H group than in the Immunity_L group (**Figure 2G**). These results collaborated that our clustering significantly distinguished colon cancers into two groups with distinct immune landscapes.

Next, we examined the relationship between individual immune cell ssGSEA score and COAD patient's overall survival and found that 10/28 cell types were significantly correlated with patient prognosis. Specifically, activated B cell ($p = 0.03$; **Figure 3A**), effector memory $CD4^+$ T cell ($p = 0.02$; **Figure 3B**), eosinophil ($p = 3.4e\text{-}3$; **Figure 3C**), immature B cell ($p = 0.04$; **Figure 3D**), neutrophil ($p = 0.04$; **Figure 3E**), and type 17 T helper cell (Th17; $p = 7.7e\text{-}3$; **Figure 3F**) were beneficial to patient survival, as a patient with higher scores of these cell types would have a higher survival probability, whereas CD56dim NK cell ($p = 0.02$; **Figure 3G**), myeloid-derived suppressor cell (MDSC; $p = 0.02$; **Figure 3H**), natural killer T cell ($p = 0.05$; **Figure 3I**), and T follicular helper cell (Tfh; $p = 8.2e\text{-}3$; **Figure 3J**) were unfavorable to patient survival, as a patient with higher scores of these cell types would have a worse survival. We then compared the abundances of these immune cells between normal and tumor tissues. Except for the natural killer T cell and type 17 T helper cell, the remaining eight immune cells had generally lower abundance in tumor tissues than in normal tissues (**Figure 3K**). We used five algorithms to demonstrate differences in the infiltrating cells between the two groups. Overall, the abundances of immune cells were higher in the Immunity_H group, especially for those calculated by CIBERSOFT algorithm (**Figure 3L**).

Differentially Expressed Genes Related to Colon Cancer Cluster-Specific Immunity Were Identified and Verified

We analyzed the differentially expressed genes between the two colon cancer immune subtypes in TCGA-COAD cohort. A total of 18 genes were up-regulated in the immunty_H group and down-regulated in the Immunity_L group. Conversely, 259 genes were down-regulated in the immunty_H group and up-regulated in the Immunity_L group (**Supplementary Table S4**; **Figure 4A**). Through GO enrichment analysis, we found these differentially expressed genes were mostly located on collagen-containing extracellular matrix, involved in

TABLE 2 | Univariate and multivariate analyses in TCGA-COAD cohort.

Variable	N	Univariate analysis		Multivariate analysis	
		HR (95% CI)	p Value	HR (95% CI)	p Value
Age (years)	429	1.02 (1.00, 1.04)	0.091		
Gender					
Female	202	1 (ref)			
Male	227	1.11 (0.73, 1.70)	0.627		
T stage					
T1	9	1 (ref)			
T2	75	0.48 (0.05, 4.68)	0.531		
T3	297	1.83 (0.25, 13.21)	0.551		
T4	48	6.01 (0.80, 45.04)	0.081		
N stage					
N0	253	1 (ref)		1 (ref)	
N1	99	1.70 (0.98, 2.97)	0.060	0.30 (0.10, 0.87)	0.027
N2	77	4.63 (2.82, 7.58)	<0.001	0.69 (0.25, 1.90)	0.477
M stage					
M0	367	1 (ref)		1 (ref)	
M1	62	4.65 (2.98, 7.24)	<0.001	21.87 (5.21, 91.76)	<0.001
AJCC stage					
Stage I	74	1 (ref)		1 (ref)	
Stage II	170	2.42 (0.72, 8.10)	0.153	2.23 (0.66, 7.50)	0.195
Stage III	123	4.77 (1.45, 15.69)	0.010	9.99 (2.15, 46.42)	0.003
Stage IV	62	13.72 (4.19, 44.94)	<0.001	NA	NA
Risk score	429	7.16 (1.97, 26.00)	0.003	3.12 (0.81, 12.00)	0.011

HR, hazard ratio; CI, confidence interval; AJCC: American Joint Committee on Cancer; NA: not applicable.

biological processes including response to interferon-gamma and extracellular matrix organization. In terms of molecular functions, these genes were mainly involved in extracellular matrix structural constituent and chemokine activity (**Figure 4B**). The results were consistent to the well-established immune processes. Furthermore, KEGG enrichment analysis concluded that these genes mainly participated in the cellular processes of phagosome, environmental information processing such as cytokine-cytokine receptor interaction. Besides, they were enriched in the organismal systems including hematopoietic cell lineage, Th17 cell differentiation, and chemokine signaling pathway (**Figure 4C**). These results largely indicated that these differentially expressed genes derived from immunogenomic clusters were closely linked to immune-related pathways.

A Four-Gene Risk Signature was Constructed in the Training Set

These differentially expressed genes showed a completely distinct expression patterns between the immunty_H group and the Immunity_L group (**Figure 5**). By univariate Cox regression analysis, we screened 11 differentially expressed genes that were significantly associated with prognosis (**Figure 6A**), of which *AXIN2* was a protective factor because it harbored a hazard ratio (HR) of less than 1 (HR = 0.861, p = 0.0481), while the remaining 10 genes were risk factors: *SERPINE1* (HR = 1.24, p = 0.0076), SFRP2 (HR = 1.12, p = 0.0203), APOE (HR = 1.15, p = 0.0230), *ARL4C* (HR = 1.27, p = 0.0236), *BST2* (HR = 1.15, p = 0.0283), TGFB1 (HR = 1.25, p = 0.0302), SLC2A3 (HR = 1.23, p = 0.0369), VSIG4 (HR = 1.2, p = 0.0389), C1QA (HR = 1.18, p = 0.0455), and BGN (HR = 1.18, p = 0.0466). We performed further screening of these genes by LASSO regression analysis and identified four genes which were highly correlated with prognosis and representative of this gene set, namely *ARL4C*, *SERPINE1*, *BST2*, and *AXIN2* (**Figure 6B**).

The LASSO analysis also allocated coefficients for these four genes (**Figure 6C**), thus facilitating the assignment of risk score to each patient. Under the best cut-off value of 0.56 detected by log-rank test (**Figure 6D**), we displayed risk scores and survival status distributions as **Figure 6E**. The *high risk* group had more mortality events. And it clearly showed that *ARL4C*, *SERPINE1*, and *BST2* expression increased with increasing risk score, whereas the expression of the *AXIN2* decreased with elevating risk score. The clinicopathological characteristics between the high and low risk groups are summarized in **Table 1**. It shows that the risk signature was an independent indicator of prognosis as the characteristics were not significantly different between the two groups, including age (p = 0.097), gender (p = 0.36), AJCC pathologic stage (p = 0.22), T stage (p = 0.052), N stage (p = 0.46), M stage (p = 0.48), and disease type (p = 0.40). Kaplan-Meier survival curves depicted that colon cancer patients with higher risk scores significantly had worse clinical outcomes (HR = 2.17, 95% CI 1.41–3.35, p < 0.001; **Figure 6F**). The ROC curves demonstrated that the risk signature harbored a promising ability to predict OS in the TCGA-COAD cohort (AUC: 1 year = 0.60, 3 years = 0.61, 5 years = 0.60; **Figure 6G**). Besides, we also analyzed the response to immune checkpoint blockers in the *high risk* and *low risk* groups. The result showed that the *low risk* group showed higher response rate to immune checkpoint blockers than the *high risk* group (35 vs. 22%; p = 0.017; **Figure 6H**).

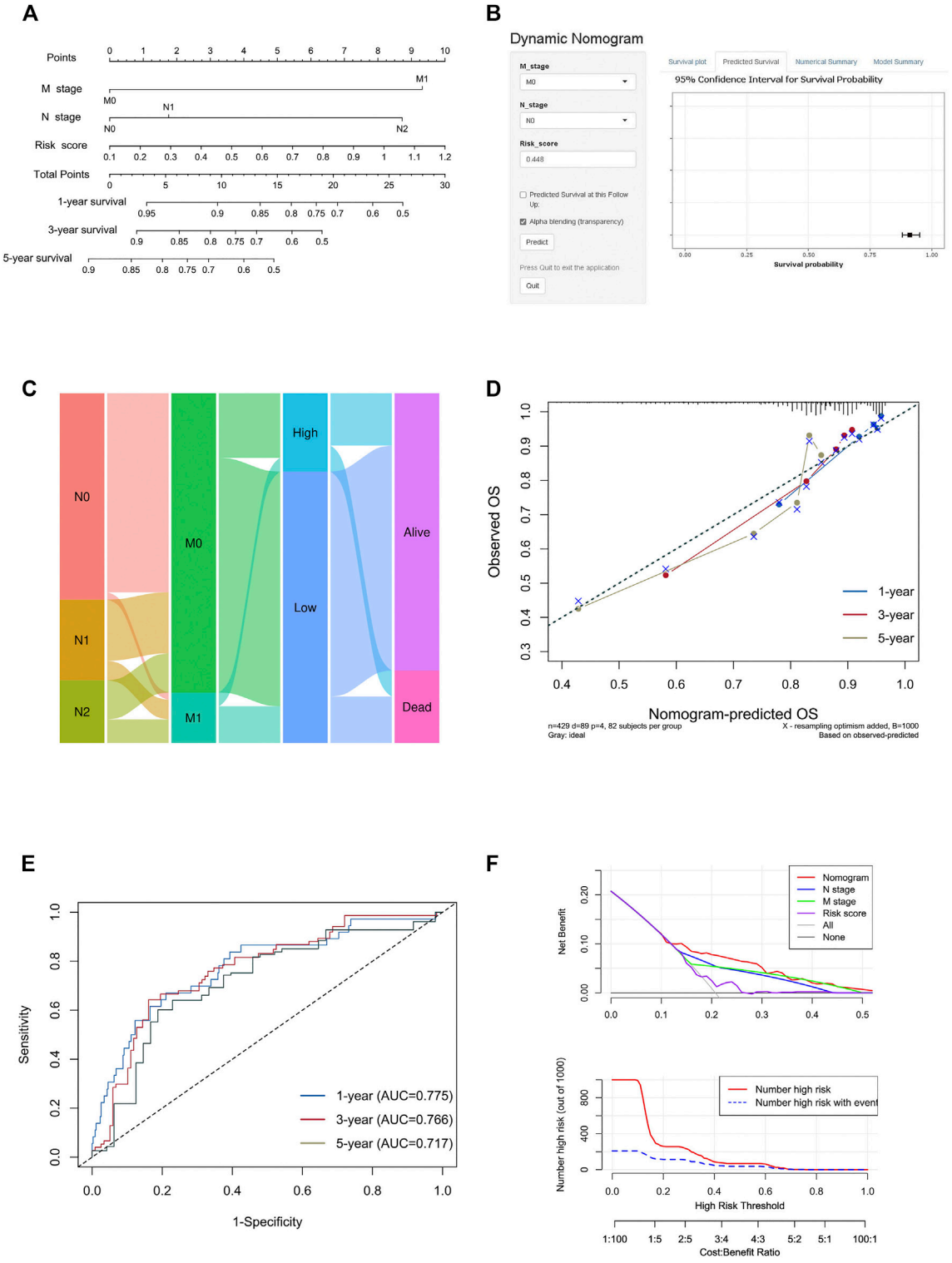


FIGURE 9 | Construction and validation of the nomogram model. **(A)** Construction of the nomogram model combining risk score and prognostic clinicopathological indicators in TCGA-COAD. **(B)** Online dynamic nomogram accessible at https://scxiangya.shinyapps.io/DynNom/, depicting an example for predicting the survival probability of a patient with TxN0M0 colon cancer and a risk score of 0.448. **(C)** Sankey diagram showing the distribution of patients in the *high risk* and *low risk* groups. **(D)** The calibration curve judging nomogram accuracy by comparing the relationship between predicted and observed overall survival. **(E)** Time dependent ROC curves showing the predictive ability of the nomogram. **(F)** Decision curve analysis curves and clinical impact curves evaluating the benefit when using the nomogram model.

Validation Cohort Demonstrated Stability of the Risk Signature

To externally validate the prognostic ability of the established risk signature, we calculated risk scores for patients in another three independent cohorts (GSE17536, GSE17537, and GSE103479) using the same formula (**Figures 7A–C**). Consistently, colon cancer patients with higher risk scores had lower OS rate and shorter OS time in the validation cohorts. The ROC analysis also indicated that the risk signature had a promising prognostic value for patients with colon cancer in the validation cohort (AUC for GSE17536: 1 year = 0.65, 3 years = 0.61, 5 years = 0.58; AUC for GSE17537: 1 year = 0.68, 3 years = 0.73, 5 years = 0.58; AUC for GSE103479: 1 year = 0.56, 3 years = 0.58, 5 years = 0.61). These results showed that the risk signature had an effective and stable OS-predictive ability for colon cancer patients.

In Vitro Validation Demonstrated High Expression of Model Genes in Colon Cancer at Transcriptional and Protein Levels

To confirm that these four genes are indeed highly expressed in colon cancer, the expressions of these four genes were detected by quantitative PCR (qPCR) in normal colon epithelial cells (NCM-460) and colon cancer cells (HCT116 and HCT8). The results showed that the expression levels of these four genes in colon cancer cells were obviously higher than those in normal colon epithelial cells (**Figure 8A**). Furthermore, we confirmed the protein expression profiles of these four genes in human tissues. As showed in immunohistochemistry results (**Figure 8B**), these four proteins were mainly distributed in the cytoplasm or membrane and were upregulated in colon cancer tissues compared with corresponding normal tissues. These results evidently demonstrated that *ARL4C*, *SERPINE1*, *BST2,* and *AXIN2* were upregulated in colon cancer cells and tissues at the transcriptional and protein levels, implying the importance of these four genes in colon cancer pathogens.

The Risk Signature-based Nomogram Had Better Prediction Ability and Practical Value

Firstly, we used univariate and multivariate Cox analyses to assess whether the established risk signature was an independent prognostic factor for patients with colon cancer (**Table 2**). Based on the data of colon cancer samples in the TCGA data set, univariate Cox analysis indicated that N stage, M stage, AJCC stage, and risk score were remarkably associated with OS ($p < 0.05$). Subsequent multivariate Cox analysis further showed that N, M stages, and risk score were independent predictors of OS ($p < 0.05$). These results indicated that our risk signature, as an independent prognostic indicator, might be useful for clinical prognosis evaluation.

To create a clinically applicable quantitative tool to predict the OS of colon cancer patients, we constructed a nomogram model including the risk score, N stage, and M stage in the TCGA data set (**Figure 9A**), which was available online (https://scxiangya.shinyapps.io/DynNom/) as screenshot in **Figure 9B**. The Sankey diagram exhibited the distribution of the clinicopathological features of the patients in different groups (**Figure 9C**). Calibration plots using 1,000 booted resampling revealed perfect concordance regarding the observed vs. predicted rates of 1-, 3- and 5-year OS in the TCGA-COAD cohort (**Figure 9D**). The ROC analysis also indicated that the nomogram had a stable and robust power in predicting the OS for colon cancer patients (AUC: 1 year = 0.775, 3 years = 0.766, 5 years = 0.717; **Figure 9E**). The DCA result indicated that the model combining prognosis-related clinicopathologic characteristics and risk signature conferred a better predictive potency than the three-factor model alone (**Figure 9F**).

DISCUSSION

Colon cancer is one of the major malignant tumors of the gastrointestinal tract, and approximately 600,000 people die from it every year (Bray et al., 2018; Siegel et al., 2020b), although 5-year survival for colon cancer has approached 65% with improved surgical methods and subsequent treatments in developed countries (Miller et al., 20192019). For colon cancer patients who present with local invasion or distant metastasis, the mortality rate is very high (Misale et al., 2012; Edwards et al., 2014; Fang et al., 2017). Therefore, there is an urgent need to find some new predictive parameters or therapeutic targets that are highly correlated with prognosis. This helps us to establish an early warning system in advance to rapidly identify patients with more critical conditions in clinical work to guide the development of subsequent treatment regimens and the prediction of survival outcomes.

Although it has long been recognized that the immune cells play an important role in tumor initiation and development (Fridman et al., 2012), these insights have not made a major influence on routine clinical practice. Moreover, the transcriptomic correlation of immune infiltration in cancer tissues on diagnosis and prognosis has attracted substantial interest. However, very few of these studies focused on the association between the difference of the immune cell composition and prognosis in colon cancer.

In this study, we first analyzed 28 immune cell types to divide TCGA-COAD into two unique immune patterns: the Immunity_H group and Immunity_L group. The two groups showed significant differences in anti-tumor immune activity, immune cell infiltration, and response to immune checkpoint blockades. Regarding the specific cell types, the complex and diverse immune cells in the TIM include T lymphocytes (70–80%), B lymphocytes (10–20%), macrophages (5–10%), NK cells (<5%), and dendritic cells (1–2%) (Frankel et al., 2017). Additionally, Tregs and TAMs contribute to tumor escape with immune suppressive activity and inhibit anti-tumor responses. Immune cells infiltrating tumors mediate the TIM and thereby influence tumor prognosis (Wu et al., 2013). In this study, we found that effector memory CD4 T cells, activated B cells, eosinophils, and Th17 cells were positively correlated with patient survival prognosis, while the immunosuppressive cells MDSCs were negatively correlated with survival prognosis.

Tumor immunotherapy is known to act as a tumor suppressor by acting through these immune cells, so we could screen out colon cancer patients who could benefit from immunotherapy based on different expression levels of immune cells.

It is of great significance to find which immune-related genes that play important roles in the development of the colon cancer and the prognosis of the patient. By specific algorithms, we identified a group of immune-related genes that predict the prognosis of colon cancer patients. With further screening and model construction, four genes including *ARL4C*, *SERPINE1*, *BST2*, and *AXIN2* were singled out to be highly associated with prognosis. It has been found that overexpression of *ARL4C* might contribute to the tumorigenesis and lead to worse prognosis in colorectal cancer (Chen et al., 2016), which supported the result of our study. A recent study has found that *SERPINE1* participated in colon cancer microenvironment remodeling and immune cell infiltration (Wang et al., 2021). This explained why patients with overexpression of *SERPINE1* had poor prognosis in our study. Moreover, the prognostic significance of *BST2* in colon cancer has been put forward as early as 2015 (Chiang et al., 2015). And finally, *AXIN2* has been consistently classified as a tumor suppressor gene in colorectal cancers both *in vivo* and *in vitro* (Church and Fazio, 2005; Waaler et al., 2012). However, it should be noted that *AXIN2* was previously identified as a potent tumor promoter instead of a suppressor, as it was found to exert global control over gene expression networks which were critical for tumor-invasive and metastatic behavior (Wu et al., 2012). Its precise function in the carcinomatous state may require further studies. In general, these four genes are promising prognostic molecules in colon cancer. The constructed model can well distinguish colon cancer patients and predict prognosis, thereby helping to develop individualized treatment options based on survival risk.

The aim of this study is to construct a model composed of prognostic immune related genes, which can robustly predict prognosis. The multivariate Cox regression analysis result showed that the survival time of the *high risk* group was significantly lower than that of the *low risk* group. This shows that our model can be used as an independent prognostic factor for colon cancer patients. According to the nomogram model, the survival rate of colon cancer patients is consistent with the actual situation. This indicates that the model can well distinguish colon cancer patients and outperform clinical parameters alone. Combining the fact that there was a significant difference in the immune cell constitutions between the Immunity_H and Immunity_L groups, we hypothesized that the immune-related genes may affect the tumor prognosis by affecting the immune infiltration. Consequently, the responses rates to immune checkpoint blockade were significantly different between the groups with distinct immune landscapes. It is thus suggested that the poor prognosis of patients in the Immunity_L group may be due to the immunosuppressive microenvironment.

However, there are still some limitations in our study. Due to insufficient clinical information in the three GEO cohorts, the nomogram model failed to be validated, and the AUC values of ROC curves for the risk signature were not high due to the limited sample size. Therefore, we will further validate this prognostic model in other independent large cohorts to ensure the reliability of our model. Moreover, functional experiments are also needed to further reveal the interplay between immune related genes and tumors.

CONCLUSION

We identified two distinct immune patterns by analyzing clinical and transcriptomic information from colon cancer patients, which exhibited distinct tumor purity and immune composition. A subsequently constructed immune-related gene-based prognostic model as well as a nomogram model was closely related to the prognosis of colon cancer patients to predict prognosis more precisely, thereby guiding risk stratification and treatment regimen development for colon patients.

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

Acquisition of Data: CS and CL. Analysis and Interpretation of Data: ZX and QL. Conception and Design: YY and FX. Data Curation: YY and FX. Development of Methodology: YC and BP. Writing the Manuscript: CS and CL.

FUNDING

This work was supported by grants from the Science and Technology Innovation Program of Hunan Province (No. 2021RC3029), the China Postdoctoral Science Foundation (Nos. 2021T140754 and 2020M672521), and the Natural Science Foundation of Hunan Province (No. 2020JJ5934).

ACKNOWLEDGMENTS

All authors thank TCGA and GEO databases for providing the worthy data and the Sangerbox tools which is a free online platform for data analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2022.820092/full#supplementary-material

REFERENCES

Aran, D., Hu, Z., and Butte, A. J. (2017). xCell: Digitally Portraying the Tissue Cellular Heterogeneity landscapePubMed Central PMCID: PMCPMC5688663 Authors Declare that They Have No Competing Interests. PUBLISHER'S NOTE: Springer Nature Remains Neutral with Regard to Jurisdictional Claims in Published Maps and Institutional Affiliations. *Genome Biol.* 18 (1), 220. Epub 2017/11/17PubMed PMID: 29141660;. doi:10.1186/s13059-017-1349-1

Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., et al. (2009). Systematic RNA Interference Reveals that Oncogenic KRAS-Driven Cancers Require TBK1. *Nature* 462 (7269), 108–112. Epub 2009/10/23PubMed PMID: 19847166; PubMed Central PMCID: PMCPMC2783335. doi:10.1038/nature08460

Beane, J., Spira, A., and Lenburg, M. E. (2009). Clinical Impact of High-Throughput Gene Expression Studies in Lung Cancer. *J. Thorac. Oncol.* 4 (1), 109–118. Epub 2008/12/20PubMed PMID: 19096318; PubMed Central PMCID: PMCPMC2731413. doi:10.1097/JTO.0b013e31819151f8

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. Epub 2016/10/22PubMed PMID: 27765066; PubMed Central PMCID: PMCPMC5073889. doi:10.1186/s13059-016-1070-5

Binnewies, M., Roberts, E. W., Kersten, K., Chan, V., Fearon, D. F., Merad, M., et al. (2018). Understanding the Tumor Immune Microenvironment (TIME) for Effective Therapy. *Nat. Med.* 24 (5), 541–550. Epub 2018/04/25PubMed PMID: 29686425; PubMed Central PMCID: PMCPMC5998822. doi:10.1038/s41591-018-0014-x

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer J. Clinicians* 68 (6), 394–424. Epub 2018/09/13PubMed PMID: 30207593. doi:10.3322/caac.21492

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017). Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cel Rep.* 18 (1), 248–262. Epub 2017/01/05PubMed PMID: 28052254. doi:10.1016/j.celrep.2016.12.019

Chen, X., Su, Z., Wang, S., and Xu, H. (2016). Clinical and Prognostic Significance of Arl4c Expression in Colorectal Cancer. *Cbm* 16 (2), 253–257. Epub 2016/01/13PubMed PMID: 26756615. doi:10.3233/cbm-150562

Chiang, S.-F., Kan, C.-Y., Hsiao, Y.-C., Tang, R., Hsieh, L.-L., Chiang, J.-M., et al. (2015). Bone Marrow Stromal Antigen 2 Is a Novel Plasma Biomarker and Prognosticator for Colorectal Carcinoma: A Secretome-Based Verification Study. *Dis. Markers* 2015, 1–10. Epub 2015/10/27PubMed PMID: 26494939; PubMed Central PMCID: PMCPMC4606116. doi:10.1155/2015/874054

Church, J. M. (2005). "Molecular Genetics of Colorectal Cancer," in *Current Therapy in Colon and Rectal Surgery*. Editor V. W. Fazio. Second Edition (Hamilton, Canada: Decker), 343–347. doi:10.1016/b978-1-55664-480-1.50064-2

Deng, D., Luo, X., Zhang, S., and Xu, Z. (2021). Immune Cell Infiltration-Associated Signature in colon Cancer and its Prognostic Implications. *Aging* 13 (15), 19696–19709. Epub 2021/08/06PubMed PMID: 34349038; PubMed Central PMCID: PMCPMC8386549. doi:10.18632/aging.203380

Dienstmann, R., Villacampa, G., Sveen, A., Mason, M. J., Niedzwiecki, D., Nesbakken, A., et al. (2019). Relative Contribution of Clinicopathological Variables, Genomic Markers, Transcriptomic Subtyping and Microenvironment Features for Outcome Prediction in Stage II/III Colorectal Cancer. *Ann. Oncol.* 30 (10), 1622–1629. Epub 2019/09/11PubMed PMID: 31504112; PubMed Central PMCID: PMCPMC6857614. doi:10.1093/annonc/mdz287

Edwards, B. K., Noone, A. M., Mariotto, A. B., Simard, E. P., Boscoe, F. P., Henley, S. J., et al. (2014). Annual Report to the Nation on the Status of Cancer, 1975-2010, Featuring Prevalence of Comorbidity and Impact on Survival Among Persons with Lung, Colorectal, Breast, or Prostate Cancer. *Cancer* 120 (9), 1290–1314. Epub 2013/12/18PubMed PMID: 24343171; PubMed Central PMCID: PMCPMC3999205. doi:10.1002/cncr.28509

Fang, C., Fan, C., Wang, C., Huang, Q., Meng, W., Yu, Y., et al. (2017). Prognostic Value of CD133+CD54+CD44+circulating Tumor Cells in Colorectal Cancer with Liver Metastasis. *Cancer Med.* 6 (12), 2850–2857. Epub 2017/11/07PubMed PMID: 29105339; PubMed Central PMCID: PMCPMC5727299. doi:10.1002/cam4.1241

Frankel, T., Lanfranca, M. P., and Zou, W. (2017). The Role of Tumor Microenvironment in Cancer Immunotherapy. *Adv. Exp. Med. Biol.* 1036, 51–64. Epub 2017/12/25PubMed PMID: 29275464. doi:10.1007/978-3-319-67577-0_4

Fridman, W. H., Pagès, F., Sautès-Fridman, C., and Galon, J. (2012). The Immune Contexture in Human Tumours: Impact on Clinical Outcome. *Nat. Rev. Cancer* 12 (4), 298–306. Epub 2012/03/16PubMed PMID: 22419253. doi:10.1038/nrc3245

Heagerty, P. J., Lumley, T., and Pepe, M. S. (2000). Time-dependent ROC Curves for Censored Survival Data and a Diagnostic Marker. *Biometrics* 56 (2), 337–344. Epub 2000/07/06PubMed PMID: 10877287. doi:10.1111/j.0006-341x.2000.00337.x

Labianca, R., Beretta, G. D., Kildani, B., Milesi, L., Merlin, F., Mosconi, S., et al. (2010). Colon Cancer. *Crit. Rev. Oncology/Hematology* 74 (2), 106–133. Epub 2010/02/09PubMed PMID: 20138539. doi:10.1016/j.critrevonc.2010.01.010

Liao, Z., Nie, H., Wang, Y., Luo, J., Zhou, J., and Ou, C. (2021). The Emerging Landscape of Long Non-coding RNAs in Colorectal Cancer Metastasis. *Front. Oncol.* 11, 641343. Epub 2021/03/16PubMed PMID: 33718238; PubMed Central PMCID: PMCPMC7947863. doi:10.3389/fonc.2021.641343

Liu, X., Hu, J., Li, Y., Cao, W., Wang, Y., Ma, Z., et al. (2018). Mesenchymal Stem Cells Expressing Interleukin-18 Inhibit Breast Cancer in a Mouse Model. *Oncol. Lett.* 15 (5), 6265–6274. Epub 2018/05/05PubMed PMID: 29725393; PubMed Central PMCID: PMCPMC5920279. doi:10.3892/ol.2018.8166

Mallmann-Gottschalk, N., Sax, Y., Kimmig, R., Lang, S., and Brandau, S. (2019). EGFR-specific Tyrosine Kinase Inhibitor Modifies NK Cell-Mediated Antitumoral Activity against Ovarian Cancer Cells. *Ijms* 20 (19), 4693. Epub 2019/09/25PubMed PMID: 31546690; PubMed Central PMCID: PMCPMC6801374. doi:10.3390/ijms20194693

Miao, Y. R., Zhang, Q., Lei, Q., Luo, M., Xie, G. Y., Wang, H., et al. (2020). ImmuCellAI: A Unique Method for Comprehensive T-Cell Subsets Abundance Prediction and its Application in Cancer Immunotherapy. *Adv. Sci.* 7 (7), 1902880. Epub 2020/04/11PubMed PMID: 32274301; PubMed Central PMCID: PMCPMC7141005. doi:10.1002/advs.201902880

Miller, K. D., Nogueira, L., Mariotto, A. B., Rowland, J. H., Yabroff, K. R., Alfano, C. M., et al. (20192019). Cancer Treatment and Survivorship Statistics, 2019. *CA A. Cancer J. Clin.* 69 (5), 363–385. Epub 2019/06/12PubMed PMID: 31184787. doi:10.3322/caac.21565

Misale, S., Yaeger, R., Hobor, S., Scala, E., Janakiraman, M., Liska, D., et al. (2012). Emergence of KRAS Mutations and Acquired Resistance to Anti-EGFR Therapy in Colorectal Cancer. *Nature* 486 (7404), 532–536. Epub 2012/06/23PubMed PMID: 22722830; PubMed Central PMCID: PMCPMC3927413. doi:10.1038/nature11156

Nagy, Á., Lánczky, A., Menyhárt, O., and Győrffy, B. (2018). Validation of miRNA Prognostic Power in Hepatocellular Carcinoma Using Expression Data of Independent Datasets. *Sci. Rep.* 8 (1), 9227. Epub 2018/06/17PubMed PMID: 29907753; PubMed Central PMCID: PMCPMC6003936. doi:10.1038/s41598-018-27521-y

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. Epub 2015/03/31PubMed PMID: 25822800; PubMed Central PMCID: PMCPMC4739640. doi:10.1038/nmeth.3337

Nie, H., Wang, Y., Liao, Z., Zhou, J., and Ou, C. (2020). The Function and Mechanism of Circular RNAs in Gastrointestinal Tumours. *Cell Prolif* 53 (7), e12815. Epub 2020/06/10PubMed PMID: 32515024; PubMed Central PMCID: PMCPMC7377939. doi:10.1111/cpr.12815

Pagès, F., Mlecnik, B., Marliot, F., Bindea, G., Ou, F. S., Bifulco, C., et al. (2018). International Validation of the Consensus Immunoscore for the Classification of colon Cancer: a Prognostic and Accuracy Study. *Lancet* 391 (10135), 2128–2139. Epub 2018/05/15PubMed PMID: 29754777. doi:10.1016/s0140-6736(18)30789-x

Pan, Y., Yu, Y., Wang, X., and Zhang, T. (2020). Tumor-Associated Macrophages in Tumor Immunity. *Front. Immunol.* 11, 583084. Epub 2020/12/29PubMed PMID: 33365025; PubMed Central PMCID: PMCPMC7751482. doi:10.3389/fimmu.2020.583084

Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E., and Gfeller, D. (2017). Simultaneous Enumeration of Cancer and Immune Cell Types from Bulk Tumor Gene Expression Data. *Elife* 6, e26476. Epub 2017/11/14PubMed PMID: 29130882; PubMed Central PMCID: PMCPMC5718706. doi:10.7554/eLife.26476

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* 43 (7), e47. Epub 2015/01/22PubMed PMID: 25605792; PubMed Central PMCID: PMCPMC4402510. doi:10.1093/nar/gkv007

Siegel, R. L., Miller, K. D., Goding Sauer, A., Fedewa, S. A., Butterly, L. F., Anderson, J. C., et al. (2020). Colorectal Cancer Statistics, 2020. *CA A. Cancer J. Clin.* 70 (3), 145–164. Epub 2020/03/07PubMed PMID: 32133645. doi:10.3322/caac.21601

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer Statistics, 2020. *CA A. Cancer J. Clin.* 70 (1), 7–30. Epub 2020/01/09PubMed PMID: 31912902. doi:10.3322/caac.21590

Smith, J. J., Deane, N. G., Wu, F., Merchant, N. B., Zhang, B., Jiang, A., et al. (2010). Experimentally Derived Metastasis Gene Expression Profile Predicts Recurrence and Death in Patients with colon Cancer. *Gastroenterology* 138 (3), 958–968. Epub 2009/11/17PubMed PMID: 19914252; PubMed Central PMCID: PMCPMC3388775. doi:10.1053/j.gastro.2009.11.005

Sun, B. B., Maranville, J. C., Peters, J. E., Stacey, D., Staley, J. R., Blackshaw, J., et al. (2018). Genomic Atlas of the Human Plasma Proteome. *Nature* 558 (7708), 73–79. Epub 2018/06/08PubMed PMID: 29875488; PubMed Central PMCID: PMCPMC6697541. doi:10.1038/s41586-018-0175-2

The Cancer Genome Atlas Network (2012). Comprehensive Molecular Characterization of Human colon and Rectal Cancer. *Nature* 487 (7407), 330–337. Epub 2012/07/20PubMed PMID: 22810696; PubMed Central PMCID: PMCPMC3401966. doi:10.1038/nature11252

Tibshirani, R. J. (1996). Regression Shrinkage and Selection via the LASSO. *J. R. Stat. Soc. Ser. B: Methodological* 73 (1), 273–282. doi:10.1111/j.2517-6161.1996.tb02080.x

Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., et al. (2017). A Pathology Atlas of the Human Cancer Transcriptome. *Science* 357 (6352), eaan2507. Epub 2017/08/19PubMed PMID: 28818916. doi:10.1126/science.aan2507

Waaler, J., Machon, O., Tumova, L., Dinh, H., Korinek, V., Wilson, S. R., et al. (2012). A Novel Tankyrase Inhibitor Decreases Canonical Wnt Signaling in colon Carcinoma Cells and Reduces Tumor Growth in Conditional APC Mutant Mice. *Cancer Res.* 72 (11), 2822–2832. Epub 2012/03/24PubMed PMID: 22440753. doi:10.1158/0008-5472.Can-11-3336

Wang, S., Pang, L., Liu, Z., and Meng, X. (2021). SERPINE1 Associated with Remodeling of the Tumor Microenvironment in colon Cancer Progression: a Novel Therapeutic Target. *BMC Cancer* 21 (1), 767. Epub 2021/07/04PubMed PMID: 34215248; PubMed Central PMCID: PMCPMC8254339. doi:10.1186/s12885-021-08536-7

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: a Class Discovery Tool with Confidence Assessments and Item Tracking. *Bioinformatics* 26 (12), 1572–1573. Epub 2010/04/30PubMed PMID: 20427518; PubMed Central PMCID: PMCPMC2881355. doi:10.1093/bioinformatics/btq170

Wu, F. Y., Fan, J., Tang, L., Zhao, Y. M., and Zhou, C. C. (2013). Atypical Chemokine Receptor D6 Inhibits Human Non-small Cell Lung Cancer Growth by Sequestration of Chemokines. *Oncol. Lett.* 6 (1), 91–95. Epub 2013/08/16PubMed PMID: 23946783; PubMed Central PMCID: PMCPMC3742824. doi:10.3892/ol.2013.1358

Wu, Z.-Q., Brabletz, T., Fearon, E., Willis, A. L., Hu, C. Y., Li, X.-Y., et al. (2012). Canonical Wnt Suppressor, Axin2, Promotes colon Carcinoma Oncogenic Activity. *Proc. Natl. Acad. Sci.* 109 (28), 11312–11317. Epub 2012/06/30PubMed PMID: 22745173; PubMed Central PMCID: PMCPMC3396472. doi:10.1073/pnas.1203015109

Yaghoubi, N., Soltani, A., Ghazvini, K., Hassanian, S. M., and Hashemy, S. I. (2019). PD-1/PD-L1 Blockade as a Novel Treatment for Colorectal Cancer. *Biomed. Pharmacother.* 110, 312–318. Epub 2018/12/07PubMed PMID: 30522017. doi:10.1016/j.biopha.2018.11.105

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring Tumour Purity and Stromal and Immune Cell Admixture from Expression Data. *Nat. Commun.* 4, 2612. Epub 2013/10/12PubMed PMID: 24113773; PubMed Central PMCID: PMCPMC3826632. doi:10.1038/ncomms3612

Zeng, D., Ye, Z., Shen, R., Yu, G., Wu, J., Xiong, Y., et al. (2021). IOBR: Multi-Omics Immuno-Oncology Biological Research to Decode Tumor Microenvironment and Signatures. *Front. Immunol.* 12, 687975. Epub 2021/07/20PubMed PMID: 34276676; PubMed Central PMCID: PMCPMC8283787. doi:10.3389/fimmu.2021.687975

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 Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 21 February 2022
doi: 10.3389/fbioe.2022.818905

DesA Prognostic Risk Model of LncRNAs in Patients With Acute Myeloid Leukaemia Based on TCGA Data

Weidong Ding[1†], *Yun Ling*[1†], *Yuan Shi*[2*] and *Zhuojun Zheng*[1*]

[1]*Department of Hematology, The Third Affiliated Hospital of Soochow University, Soochow, China,* [2]*Laboratory of Hematology, The Third Affiliated Hospital of Soochow University, Soochow, China*

Purpose: This study aimed to combine the clinical data of acute myeloid leukaemia (AML) from The Cancer Genome Atlas (TCGA) database to obtain prognosis-related biomarkers, construct a prognostic risk model using long non-coding RNAs (lncRNAs) in AML and help patients with AML make clinical treatment decisions.

Methods: We analysed the transcriptional group information of 151 patients with AML obtained from TCGA and extracted the expressions of lncRNAs. According to the mutation frequency, the patients were divided into the high mutation group (genomic unstable group, top 25% of mutation frequency) and low mutation group (genomic stable group, 25% after mutation frequency). The 'limma' R package was used to analyse the difference in lncRNA expressions between the two groups, and the "survival," "caret," and "glmnet" R packages were used to screen lncRNAs that are related to clinical prognosis. Subsequently, a prognosis-related risk model was constructed and verified through different methods.

Results: According to the lncRNA expression data in TCGA, we found that seven lncRNAs (i.e. AL645608.6, LINC01436, AL645608.2, AC073534.2, LINC02593, AL512413.1, and AL645608.4) were highly correlated with the clinical prognosis of patients with AML, so we constructed a prognostic risk model of lncRNAs based on LINC01436, AC073534.2, and LINC02593. Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway analyses of differentially expressed lncRNA-related target genes were performed, receiver operating characteristic (ROC) curves were created, the applicability of the model in children was assessed using the TARGET database and the model was externally verified using the GEO database. Furthermore, different expression patterns of lncRNAs were validated in various AML cell lines derived from Homo sapiens.

Conclusions: We have established a lncRNA prognostic model that can predict the survival of patients with AML. The Kaplan-Meier analysis showed that this model distinguished survival differences between patients with high- and low-risk status. The ROC analysis confirmed this finding and showed that the model had high prediction accuracy. The Kaplan-Meier analysis of the clinical subgroups showed that this model can

OPEN ACCESS

Edited by:
Giulia Russo,
University of Catania, Italy

Reviewed by:
Natasha Andressa Jorge,
Leipzig University, Germany
Tiansheng Zeng,
Guangzhou Medical University, China

***Correspondence:**
Zhuojun Zheng
zenki_zheng@163.com
Yuan Shi
cz-shiyuan@163.com

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Preclinical Cell and Gene Therapy, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *20 November 2021*
Accepted: *03 February 2022*
Published: *21 February 2022*

Citation:
Ding W, Ling Y, Shi Y and Zheng Z (2022) DesA Prognostic Risk Model of LncRNAs in Patients With Acute Myeloid Leukaemia Based on TCGA Data. Front. Bioeng. Biotechnol. 10:818905. doi: 10.3389/fbioe.2022.818905

predict prognosis independent of clinicopathological factors. Therefore, the proposed prognostic lncRNA risk model can be used as an independent biomarker of AML.

Keywords: acute myeloid leukaemia, cox regression, lncRNA, prognostic risk model, TCGA

1 INTRODUCTION

Acute myeloid leukaemia (AML) is a malignant clonal disease of the haematopoietic stem and progenitor cells (Pelcovits and Niroula, 2013). In China, leukaemia affects 3–4 individuals per 100,000 populations. Among deaths due to malignant tumours, leukaemia ranked sixth in men, seventh in women and first in children and adults aged <35 years. In China, the incidence of acute leukaemia (AL) was significantly higher than that of chronic leukaemia, and AML was the most common (1.62/100,000). In recent years, intensive chemotherapy, haematopoietic stem cell transplantation and rigorous supportive treatment have greatly improved the prognosis of patients with AML aged <60 years. Moreover, 30% of patients with non-acute promyelocytic leukaemia are expected to survive for a long time (Sasine and Schiller, 2016). However, for these patients, a good quantitative model for predicting survival time is still not established (Song et al., 2018).

With the development of sequencing technology, the detection of leukaemia-related genes is becoming increasingly impeccable, which has increasingly attracted the attention of researchers. Among them, the long noncoding RNA (lncRNA) has gradually become a research hotspot. LncRNA is an RNA molecule with a length of >200 bp that originates from the noncoding region of the genome. It regulates gene expression at the transcriptional and post-transcriptional levels and participates in various biological functions (Al-Kershi et al., 2019). Some studies have shown that lncRNAs play an important role in many life activities, such as dose compensation effect, epigenetic regulation, cell cycle regulation and cell differentiation regulation (Peng et al., 2017). Recent studies have confirmed that changes in lncRNAs are related to the occurrence and development hematological malignancy, especially in AML. Myeloid-specific and polyadenylated lncRNA LOUP was found to induce myeloid differentiation and inhibits cell growth, acting as a transcriptional inducer of the myeloid master regulator. LOUP recruits RUNX1 to both the LOUP enhancer and the promoter, leading to the formation of an active chromatin loop (Trinh et al., 2021). Yin et al.(Yin et al., 2021) found that lncRNA DUBR highly expressed in AML, resulting in poor prognosis, especially in M4 AML. *In vitro* studies elucidated that knockdown of DUBR suppress the survival colony formation ability in AML cells. Academics pointed out that the activation of HOXBLINC, a HOXB locus-associated lncRNA, is a critical downstream mediator of NPM1c(+)-associated leukemic transcription program and leukemogenesis. HOXBLINC loss attenuates NPM1c(+)-driven leukemogenesis by rectifying the signature of NPM1c(+) leukemic transcription programs. Overexpression of HOXBLINC in mice enhances hematopoietic stem cell self-renewal and expands myelopoiesis, leading to the development of AML-like disease, reminiscent of the phenotypes seen in the Npm1 mutant knock-in (Npm1(c/+)) mice (Zhu et al., 1956). Gourvest et al.(Wei and Wang, 2017; Zhao et al., 2018; Liang et al., 2020; Gourvest et al., 2021) report an identification of lncRNA LONA overexpressed in NPM1-mutated AML patients. While NPM1 is nuclear and LONA cytoplasmic in wild-type NPM1 AML cells, LONA becomes nuclear as mutant NPM1 moves toward the cytoplasm. Gain or loss of function combined with a genome-wide RNA-seq identified a set of LONA mRNA targets encoding proteins involved in myeloid cell differentiation and interaction with its microenvironment. LONA overexpression exerts an anti-myeloid differentiation effect in mutant NPM1 established cell lines and primary AML cells. *In vivo*, LONA overexpression acts as an oncogenic lncRNA reducing the survival of mice transplanted with AML cells and rendering AML tumors more resistant to cytarabine chemotherapy.

This study aimed to combine the clinical data of AML from TCGA to obtain prognosis-related biomarkers, construct a prognostic risk model related to lncRNAs in AML and help patients with AML make clinical treatment decisions. In this study, transcription data of 151 patients with AML were downloaded from The Cancer Genome Atlas (TCGA). Perl language was used to collate the data, and R language was used for data analysis in an attempt to determine effective prognostic biomarkers for AML and construct a prognostic risk model using lncRNA in patients with AML.

2 METHODS

2.1 Research Objects and Data Acquisition

Transcription group data (transcription profiling) and corresponding clinical data of 151 patients with AML were obtained through TCGA website (https://tcga-data.nci.nih.gov/tcga/). The database contains clinical data such as patient number, age, survival time and survival status. The genome mutation data (simple nucleotide variation) of 149 patients were also downloaded from TCGA.

2.2 Acquisition of the Expression Matrix of lncRNAs

On the official website of TCGA (https://gdc.nci.nih.gov/), the TCGA-LAML transcriptional group data (transcription profiling) were checked on the GDC Data Portal to download relevant raw htseq-count data, manifest and metadata files. In the CMD environment, the Perl language script was used to extract the original count data to form an expression matrix. The identification transformation of the transcription expression profile was implemented as Homo_sapiens.GRCh38.95. chr.gtf

and downloaded from the Ensemble website, the gene expression profile matrix was obtained after comparison, and lncRNA was extracted using Perl language script to obtain the lncRNA expression matrix of patients with myeloid leukaemia.

2.3 Sample Mutation Frequency and Grouping

In the CMD environment, Perl language scripts were used to calculate the mutation frequency of the samples. In this study, 31 patients with the top 25% mutation frequency were assigned to the genomic instability group (high mutation group, genomic unstable [GU]) and 24 patients with the bottom 25% mutation frequency to the genomic stability group (low mutation group, genomic stable [GS]).

2.4 Screening of Differentially Expressed lncRNAs and mRNAs in the High and Low Mutation Groups and Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Analyses

To determine potential AML biomarkers, R language was used to calculate the mean value of lncRNA in the GS and GU groups, and the "limma" R package was then used to set the threshold to logFC >1.0 and $p < 0.05$ to screen the differentially expressed IncRNAs between the two groups. Thereafter, data of the upregulated and downregulated differentially expressed IncRNAs and their corresponding expressions were saved, and the "pheatmap" R package was used to draw a heat map. We used the "limma" R package to test the correlation between the expressions of mRNA and lncRNA in the samples and obtained the correlation coefficient and p value. The mRNAs of the first 10 related genes were selected as the target genes of the corresponding differentially expressed lncRNAs. Subsequently, GO and KEGG enrichment analyses of the target genes were performed using "clusterProfiler" (version 3.14.3). The minimum and maximum genes were set to 5 and 5,000, respectively. A p value of <0.05 and a false discovery rate (FDR) of <0.25 were considered significant.

2.5 Cox Regression Analysis

To determine which of the differentially expressed genes were related to prognosis, we extracted the survival time and survival status of patients with AML and excluded data of patients whose survival time was less than 30 days. The survival time data were compared individually with differential gene expression data, and duplicated samples of expression data were removed; finally, 128 samples with complete prognosis were screened. Thereafter, the survival time data and differential gene expression data were combined into a matrix. Furthermore, we used the random function of R language to randomly divide the 128 samples into the train and test groups. First, we determined the prognosis-related lncRNAs from the train group for Cox analysis. Specifically, the univariate regression analysis was performed on the train group using the "survival" and "survminer" R packages. Seven lncRNAs related to prognosis were obtained, and the corresponding forest map was drawn by R packages "survival," "caret," "glmnet," and "survminer." Thereafter, the "glmnet" and "survival" R packages were used in the multivariate Cox regression analyses of the seven lncRNAs.

2.6 Construction of the lncRNA Prognostic Model

Using multivariate regression, a survival-related prognosis model was constructed based on three lncRNAs out of 7 that screened through multivariate regression. The model formula is as follows:

$$\textbf{Risk score} = \Sigma^{3}_{i=1}\boldsymbol{\beta}\mathbf{i}.$$

In this equation, $\boldsymbol{\beta}$ is the regression analysis coefficient of each IncRNA after multifactor Cox regression analysis, and $\mathbf{i}$ is the correlation of lncRNA of the multifactor regression. We defined the risk score of single samples lower than the median risk score of the training group as low risk and conversely as high risk. Accordingly, samples were divided into the high-risk subgroup and the low-risk subgroup in the train and test groups, respectively.

2.7 Visualization, Evaluation and Testing of the lncRNA Prognostic Model

2.7.1 Testing the Clinical Characteristic Preference Between the Train Group and the Test Group Randomly Divided Into Subgroups

To test whether the clinical characteristics of patients with AML were consistent, patients with AML were divided into the elderly (≥60 years) and non-elderly (<60 years) groups and male and female groups. The chi-square test was used to process the percentage of age and sex in the two groups.

2.7.2 Survival Analysis Verification

To test the applicability of the prognostic model in the identification of patients with high- and low-risk status, we used the "survival" and "survminer" R packages to draw the high- and low-risk Kaplan–Meier curves of the train, test and overall groups.

2.7.3 Receiver Operating Characteristic (ROC) Curve

To verify the accuracy of the model in predicting the prognosis of patients with AML in different periods, we used the "survival," "survminer," and "timeROC" R packages to draw the ROC curves of the 1-year and 5-year survival rates of the train, test and overall groups and calculated the area under their ROC curves (AUCs).

2.7.4 Relationship Between Known Clinical Prognostic Genes and Gene Mutation Frequency and Prognostic Model

Relationship Between Patient Risk Score and Gene Expression and Between Patient Score and Genomic Instability

To examine the relationship between the prognostic model and other factors, we ranked the samples according to the risk score from low to high. Thereafter, we used the "limma" and "pheatmap" R packages to visualise the expression of three

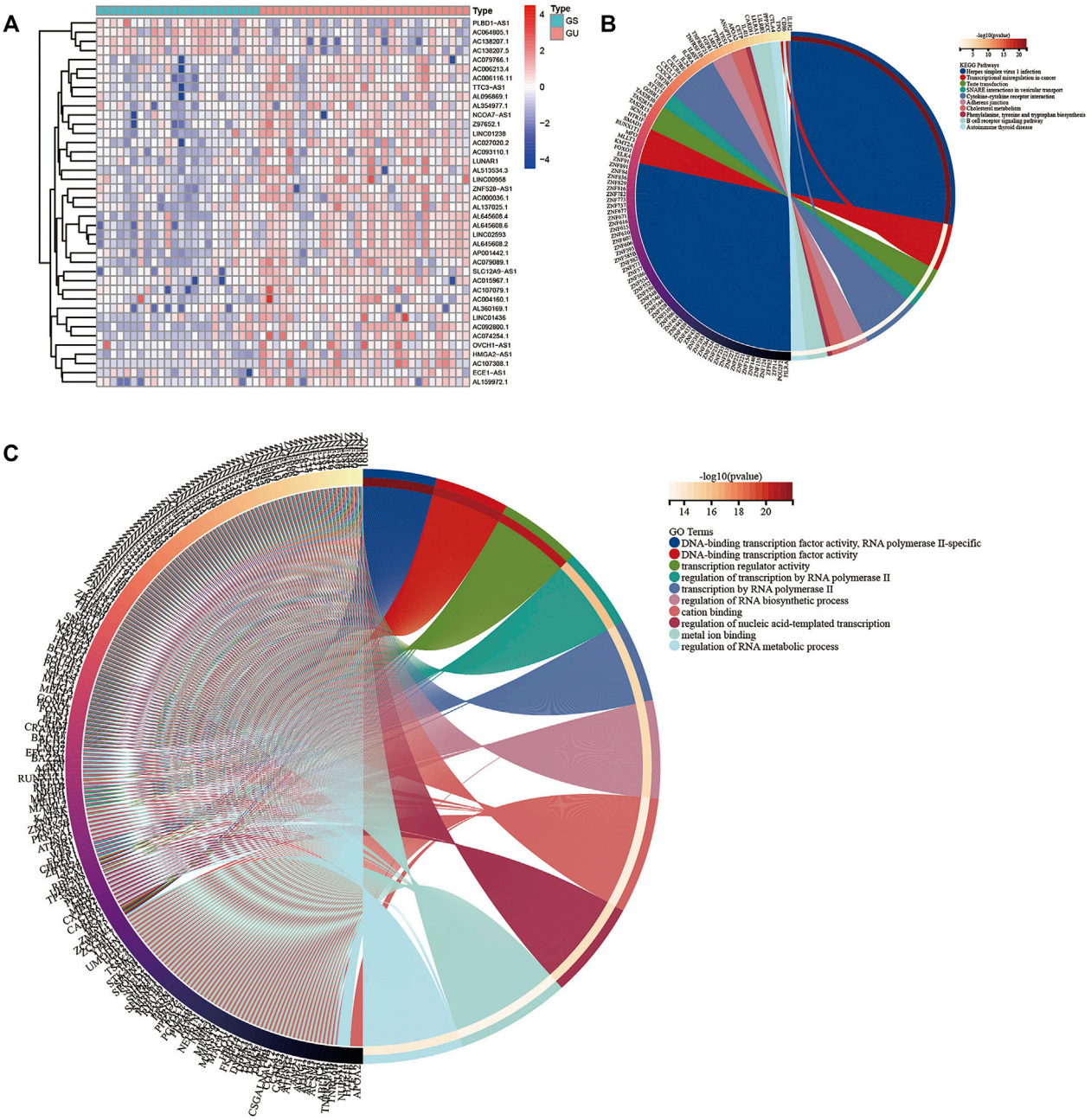


FIGURE 1 | (A) Heat map of the top 20 upregulated and downregulated lncRNAs. The upregulated and downregulated genes are marked red and blue, respectively. The top 10 mRNAs related to differentially expressed lncRNAs were selected as target genes. The Kyoto Encyclopedia of Genes **(B)** and Genomes and Gene Ontology **(C)** analysis maps of the target genes were drawn.

lncRNAs related to multivariate regression in the train, test and overall groups. Moreover, we visualised the mutation frequency of each sample and the known mutation-driven genes.

Relationship Between the Prognostic Model and Known Mutation-Driven Genes Such as TP53

We used the "limma" and "sparcl" R packages to extract the differentially expressed lncRNAs and mutation statistics of all samples. Thereafter, the samples were divided into GS (GS-like) and GU (GU-like) groups by cluster analysis. Then, the mutation frequency between the two types and the expression of the mutant gene LUNAR1 were evaluated, and the "ggpubr" R package was used to draw the corresponding box diagram. To examine the effect of common mutations on this prognostic model, we used the "survival" and "survminer" R packages to draw the Kaplan–Meier curves of the first six genes with the highest mutation frequency in the GS (GS-like) and GU (GU-like) groups. To test whether there was a significant difference in

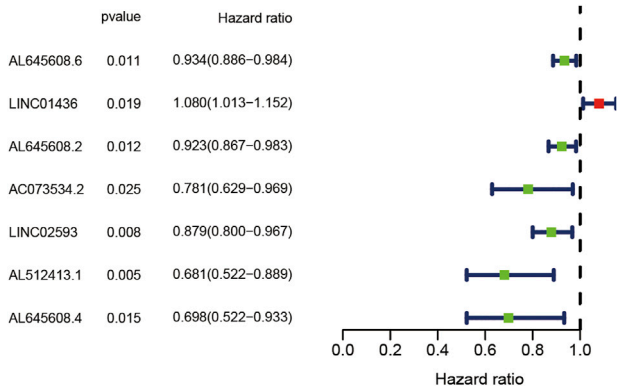


FIGURE 2 | Forest plot of the univariate Cox regression analysis of seven lncRNAs with prognostic value. The *X*-axis represents the risk ratio, and the *Y*-axis represents the lncRNAs with prognostic value. Red represents a high-risk factor, which is negatively correlated with the survival time of the patient. Blue represents a low-risk factor, which is positively correlated with the survival time of the patient.

gene mutations between the high- and low-risk groups, we used "plyr" and "ggplot2" R packages to draw single mutation frequency histograms of common mutations in the train, test and overall groups to verify whether known common mutations affect the prognostic model of high- and low-risk scores.

2.7.5 Verification of the Prognostic Model With Different Clinical Characteristics

To test the ability of the prognostic model to evaluate high- and low-risk groups with different clinical traits, we divided the samples into the elderly and non-elderly groups, male and female groups, non-M3 groups and high mutation frequency and low mutation frequency groups according to the clinical traits. Note that: 1. the survival analysis of M3 (acute promyelocytic leukaemia) group is not going to be performed as the sample size in this group is relatively small. 2. TCGA-AML data contains mutation data of 134 patients and complete clinical information 128 patients, taking the intersection of them, there are 88 patients with complete mutation data and clinical information. High mutation frequency was defined as single sample mutation counts ≥ median mutation counts, and low mutation frequency as single sample mutation counts < median mutation counts. Then, the "survival" and "survminer" R packages were used to draw the Kaplan–Meier curves of above-mentioned groups at high and low risks.

2.7.6 Applicability of the Prognostic Model in Children

We downloaded data of 155 child patients with AML with complete prognosis-related lncRNA expression and clinical information from the TARGET database (https://ocg.cancer.gov/programs/target). Then, Microsoft Excel software was used to sort out the data into the children group and calculated the corresponding risk score. Finally, we used the "survival" and "survminer" R packages to draw the corresponding Kaplan–Meier curves between different groups of children and adults to test the applicability of the model in children.

2.7.7 External Verification

GSE106291 (Herold et al., 2018) data from the GEO database, which included survival information and transcriptome data of 250 patients with AML, were downloaded for further analysis. As not every lncRNA was clearly annotated in this data (especially for RNA-seq data of early year), we calculated the risk score by determining the RNA-related expression level according to LNCipedia (Volders et al., 2019) and catRAPID (Armaos et al., 2021), in which different lncRNA transcripts were considered to belong to a certain gene if they share at least one (partially) overlapping exon and reside on the same DNA strand. In this way, transcripts were clustered into genes. The risk score was thus calculated according to the gene expression when lncRNA was not clearly annotated. Then, we used the "survival" and "survminer" packages for prognostic analysis. Finally, 123 patients with complete survival information and transcriptome data were included in the validation cohort.

2.7.8 LncRNA Expression Verification in AML Cells

The LncRNAs were detected in AML cell lines derived from Homo sapiens with fluorescence *in situ* hybridisation (FISH) based on the protocol (Duncan et al., 2019). The methods were as follows: The specimens were permeabilised with cold 0.1% Triton X-100. The pre-hybridisation buffer was discarded, and hybridisation was performed using the LINC01436, AC073534.2 and LINC02593 probe overnight, respectively. After washing with SCC buffer, the coverslip was dyed with 4′,6-diamidino-2-phenylindole (DAPI), and the fluorescence test was conducted with laser scanning confocal microscope. AML cell lines were selected as follows: HL-60 (derived from the peripheral blood of patients with acute promyelocytic leukaemia), U937 (derived from the peripheral blood of patients with acute monocytic leukaemia), MV4-11 (derived from the peripheral blood of patients with biphenotypic B myelomonocytic leukaemia with FLT3-ITD mutation) and Kasumi-1 (derived from the peripheral blood of patients with acute myeloblastic leukaemia with AML1-ETO fusion gene positive). All cell lines were ATCC sources. Cells were maintained in RPMI-1640 medium supplemented with 10% foetal calf serum (HyClone Laboratories, Logan, UT, United States), 100 U/mL penicillin and 100 μg/ml streptomycin at 37°C in a humidified atmosphere of 5% CO_2.

3 STATISTICAL ANALYSIS

Data analyses were performed in R language (R4.0.2), and the difference was considered significant when $p < 0.05$.

4 RESULTS

4.1 Establishment of the lncRNA Prognostic Model for AML

4.1.1 Differentially Expressed lncRNAs in TCGA Patients With AML

We obtained 149 AML samples with transcriptome data from TCGA database. A total of 31 patients with the first 25% mutation frequency were classified as the GU group (high mutation group,

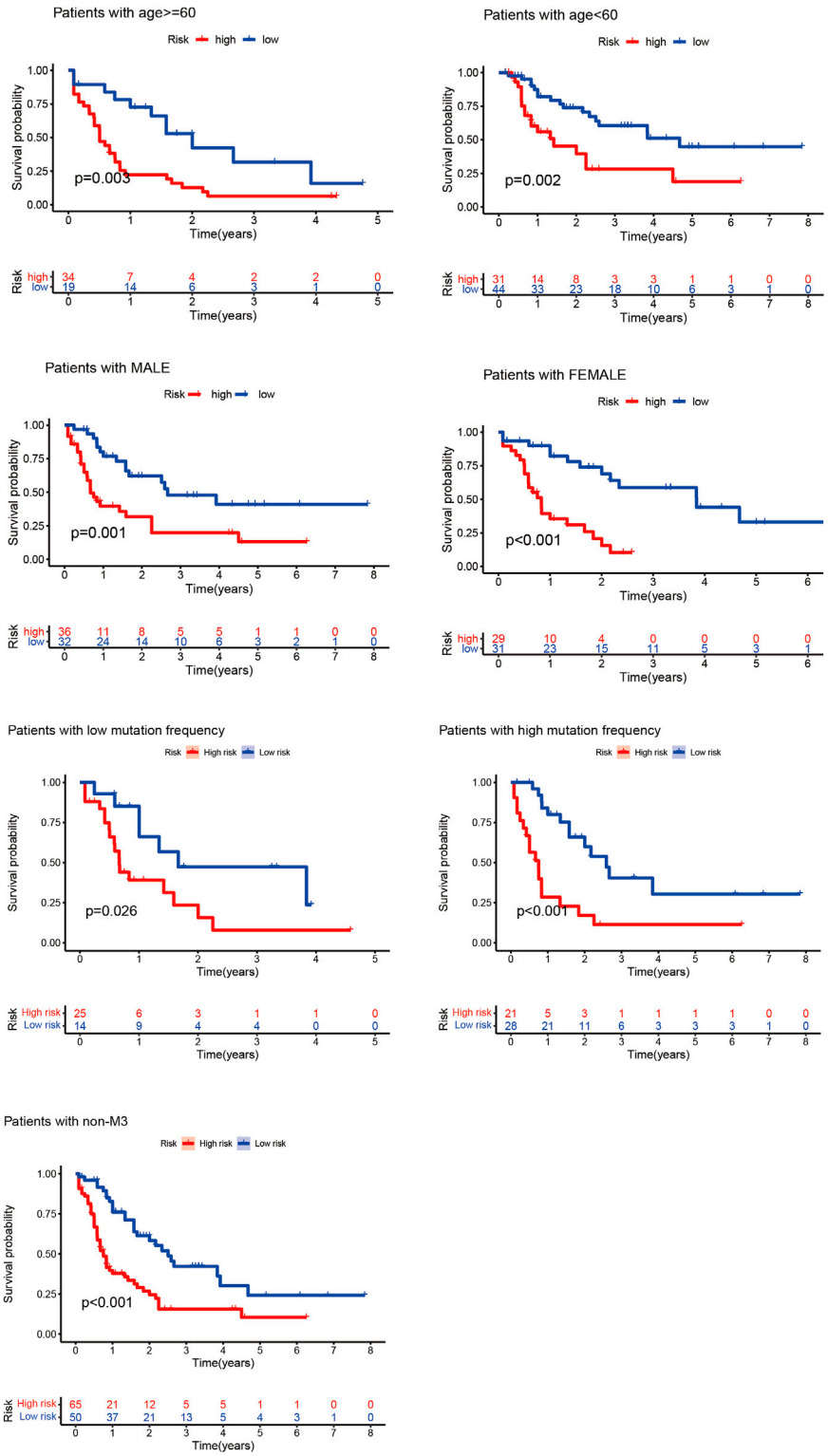


FIGURE 3 | Kaplan–Meier curves of high- and low-risk groups of elderly and non-elderly patients and male and female patients. In the Kaplan–Meier curve of different clinical subgroups of patients with AML, the X- and *Y*-axes represent the time and probability of survival, whereas the red and blue lines represent the high- and low-risk groups, respectively.

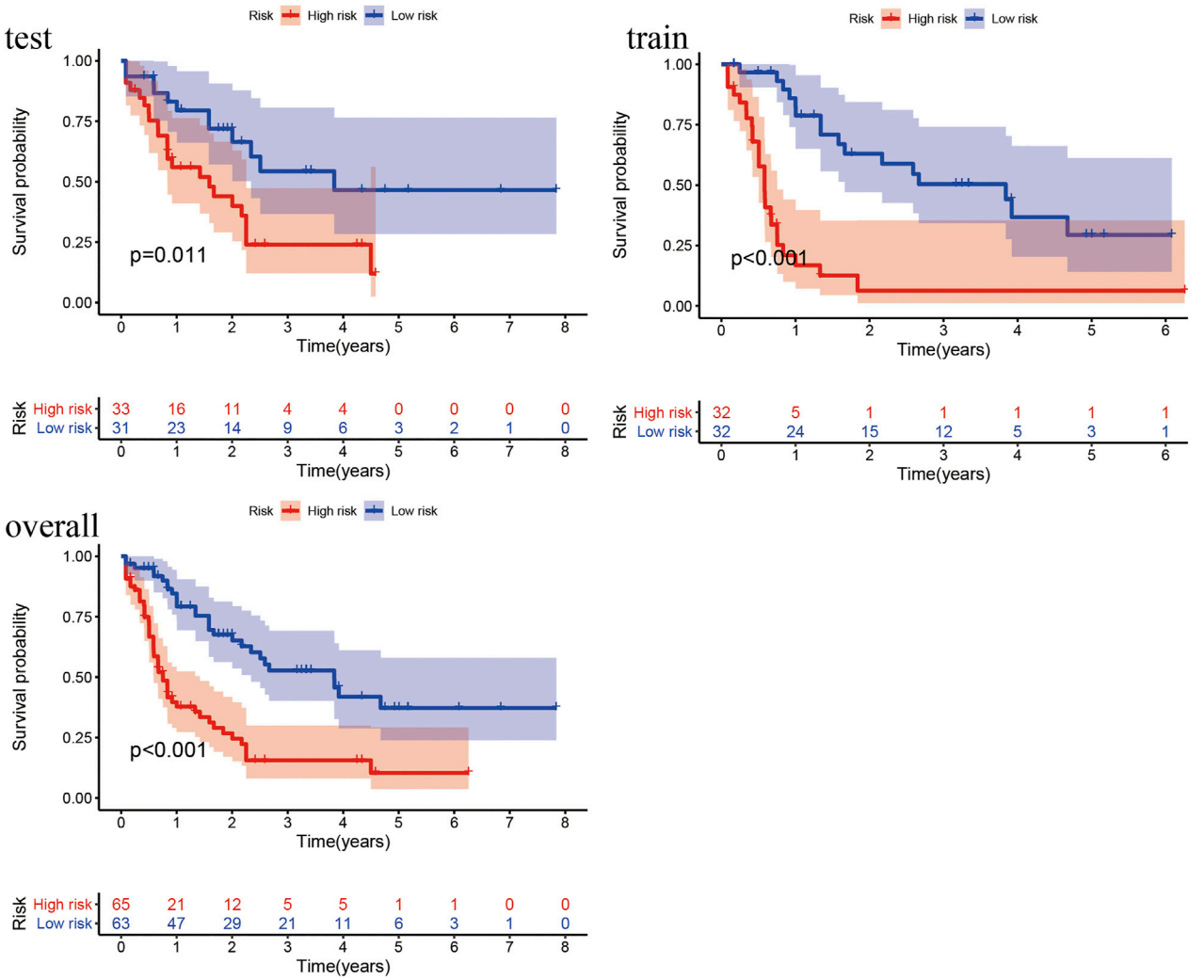


FIGURE 4 | Kaplan–Meier curves of high- and low-risk groups in the train, test, and overall groups. The X- and *Y*-axes represent the time and probability of survival, whereas the red and blue lines represent the high- and low-risk groups, respectively.

mutCount ≥19), and 24 patients with post 25% mutation frequency were classified as the GS group (low mutation group, mutCount ≤3). The average expressions of lncRNAs in the GS and GU groups were calculated by R language, differentially expressed lncRNAs were screened using the "limma" R package and the threshold was set to (logFC >1.0 and $p < 0.05$). Finally, 59 differentially expressed lncRNAs were obtained (**Supplementary Table S1**). Among them, heat maps of 20 upregulated and 20 downregulated differentially expressed lncRNAs were drawn (**Figure 1A**). To further examine the human functions these lncRNAs are involved in, we performed KEGG and GO enrichment analyses of 59 target genes of differentially expressed lncRNAs. The results (**Figures 1B**,C) revealed that the most abundant genes in the KEGG were enriched in "Herpes simplex virus 1 infection" pathway. Regarding GO, the most enriched genes were in the 'DNA-binding transcription factor activity, RNA polymerase II-specific' process.

4.1.2 lncRNA Prognostic Risk Model for AML

To evaluate the prognostic value of lncRNAs in AML, 128 samples were randomly divided into the train and test groups by R language random function. Thereafter, the differentially expressed lncRNAs in the train group were analysed by Cox regression analysis. Seven lncRNAs (AL645608.6, LINC01436, AL645608.2, AC073534.2, LINC02593, AL512413.1, and AL645608.4) related to prognosis were obtained, and their corresponding forest maps were drawn. As shown in **Figure 2**, among the seven lncRNAs, only the hazard ratio (HR) of LINC01436 is greater than 1, which means that LINC01436 is a risk factor for patients with AML and has a negative correlation with the clinical prognosis; therefore, the higher the expression, the worse the prognosis. The rest of the lncRNAs (AL645608.6 [HR = 0.934], AL645608.2 [HR = 0.923], AC073534.2 [HR = 0781], LINC02593 [HR = 0.879], AL512413.1 [HR = 0.681] and AL645608.4 [HR = 0698]) appeared as protective factors, which

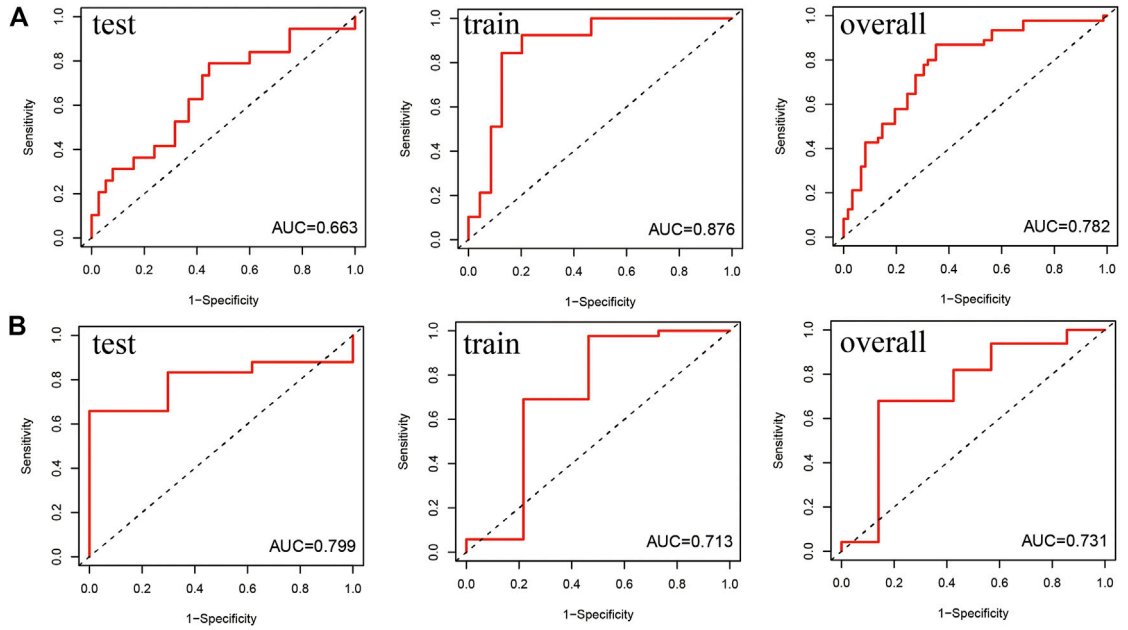


FIGURE 5 | Receiver operating characteristic (ROC) curve of the prediction efficiency of the train, test, and overall groups. The *X*- and *Y*-axes represent 1-specificity and sensitivity, respectively. Corresponding ROC curve of the **(A)** 1-year survival rate and **(B)** 5-year survival rate.

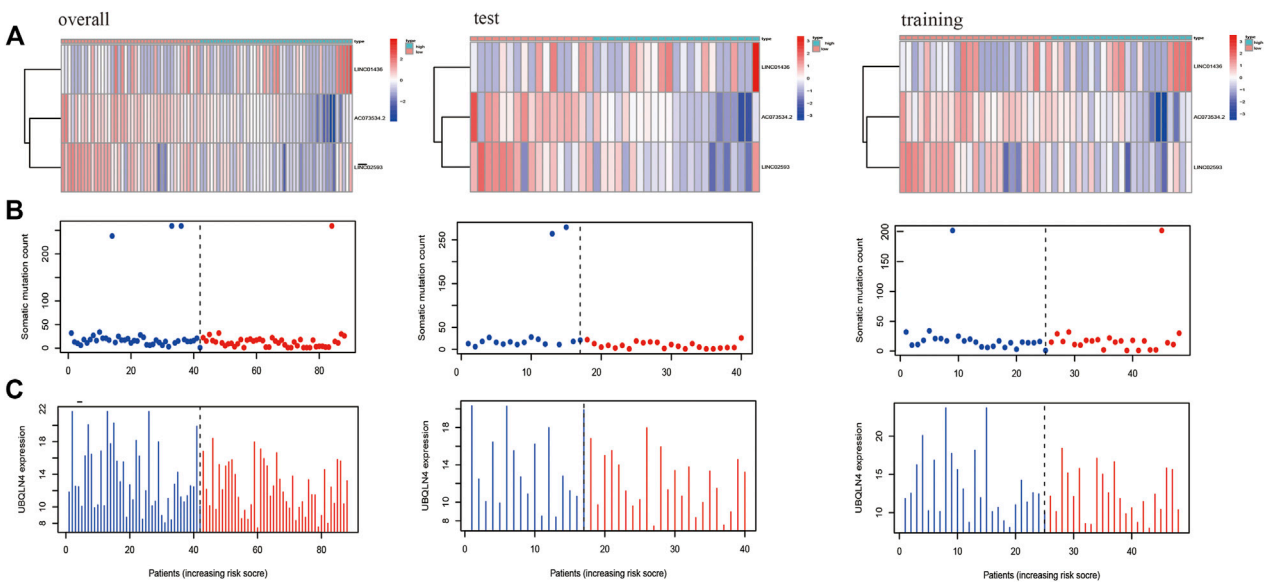


FIGURE 6 | The *X*-axes of A, B, and C are all samples sorted according to the increasing values at risk. **(A)** Heat map of the lncRNAs related to the prognostic model; the ordinate represents the three lncRNAs that make up the prognostic model. Red represents upregulation of gene expression, while blue represents downregulation of gene expression. **(B)** Distribution of gene mutation frequency; the ordinate represents the mutation frequency of each sample. **(C)** Expression frequency distribution of UBQLN4, the driver of genomic instability; the ordinate is the expression of UBQLN4 in each sample.

were positively correlated with the clinical prognosis of the patients, i.e. the higher the expression, the better the prognosis. Furthermore, we used "glmnet" and "survival" R packages to perform multivariate Cox regression analysis of the seven lncRNAs, and the key result of model construction was obtained. As shown in **Supplementary Tables S2, S3** of 7 lncRNAs (LINC01436, AC073534.2 and LINC02593) were selected as major parameters to build the model, and they were identified as the independent prognostic factors through multivariate Cox regression analysis (coef [LINC01436] =

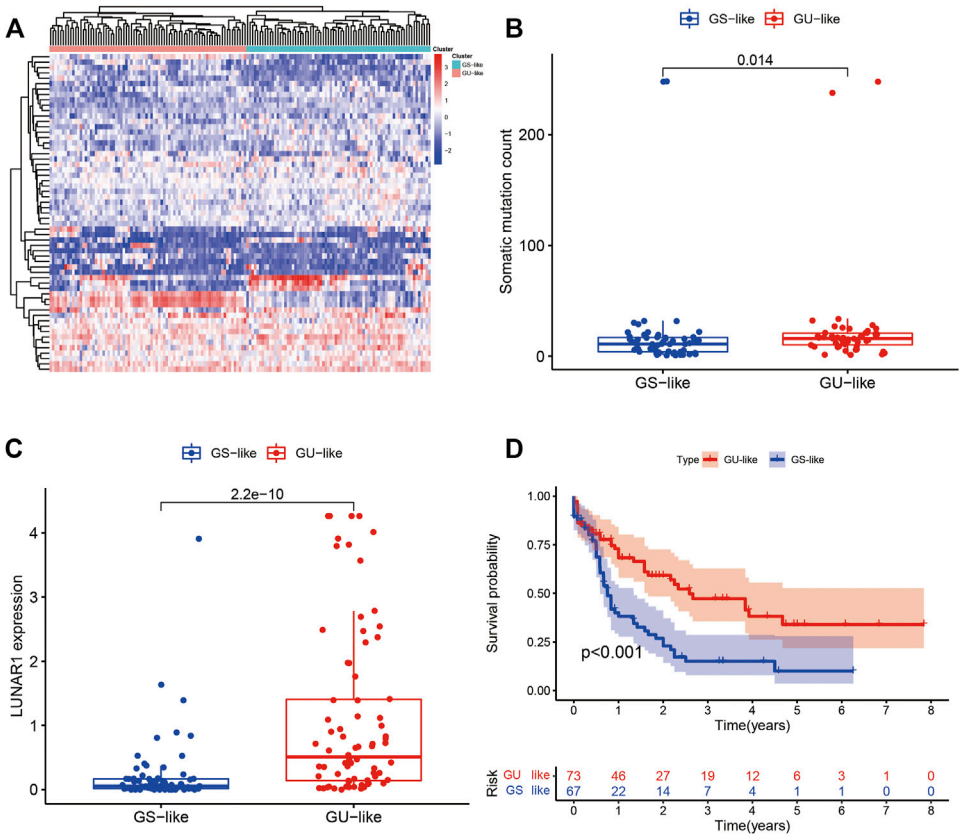


FIGURE 7 | (A) Heat map of all samples after the cluster analysis. The *X*-axis is the sample type, blue represents the genomic stable type and red represents the genomic unstable type. The *Y*-axis represents differentially expressed lncRNAs, red represents upregulation, and blue represents downregulation. **(B)** Box diagram of the mutation frequencies of the two types. **(C)** Expression map of the two types of LUNAR1. **(D)** Kaplan–Meier curves of the GU-like and GS-like groups. In the Kaplan–Meier curve of different somatic mutation count of patients with AML, the X- and *Y*-axes represent the time and probability of survival, whereas the red and blue lines represent the GU-like and GS-like groups, respectively.

0.070402, coef [AC073534.2] = −0.303302254, coef [LINC02593] = −0.139241309). The risk score of each patient could be calculated according to the regression coefficient and expression value of the three lncRNAs.

4.2 Visualization, Evaluation, and Testing of the lncRNA Prognostic Model

4.2.1 Analysis of the Clinical Characteristic Preference Among Random Groups

To determine the consistency of the clinical characteristics during model construction, clinical features of all samples were compared by the chi-square test. As shown in **Supplementary Table S3**, the *p* value between the train and test groups were all >0.05, indicating that our model grouping has no characteristic preference.

4.2.2 Clinical Grouping Verification of the Prognostic Model

According to the clinical characteristics, patients were divided into the elderly and non-elderly groups and male and female groups. The "survival" and "survminer" R packages were used to draw the Kaplan–Meier curves of the high- and low-risk patients in the elderly and non-elderly groups and male and female groups. As presented in **Figure 3**, the Kaplan–Meier curve showed that the survival time of patients in the low-risk group was significantly prolonged, and the median survival time in the non-elderly group was 2.17 years, which was higher than that in the high-risk group (0.84 years). In the elderly group, the median survival time in the low-risk group (1.58 years) was higher than that in the high-risk group (0.50 years). In the male group, the median survival time of the low-risk group (1.84 years) was higher than that of the high-risk group (0.63 years). In the female group, the median survival time of the low-risk group (1.84 years) was higher than that of the high-risk group (0.75 years). In the low mutation frequency group, the median survival time of the low-risk group (1.17 years) was higher than that of the high-risk group (0.66 years). In the high mutation frequency group, the median survival time of the low-risk group (1.66 years) was higher than that of the high-risk group (0.67 years). In the non-M3 group, the median survival time of the low-risk group (1.71 years) was higher than that of the high-risk group (0.67 years). These results showed that the prognostic model

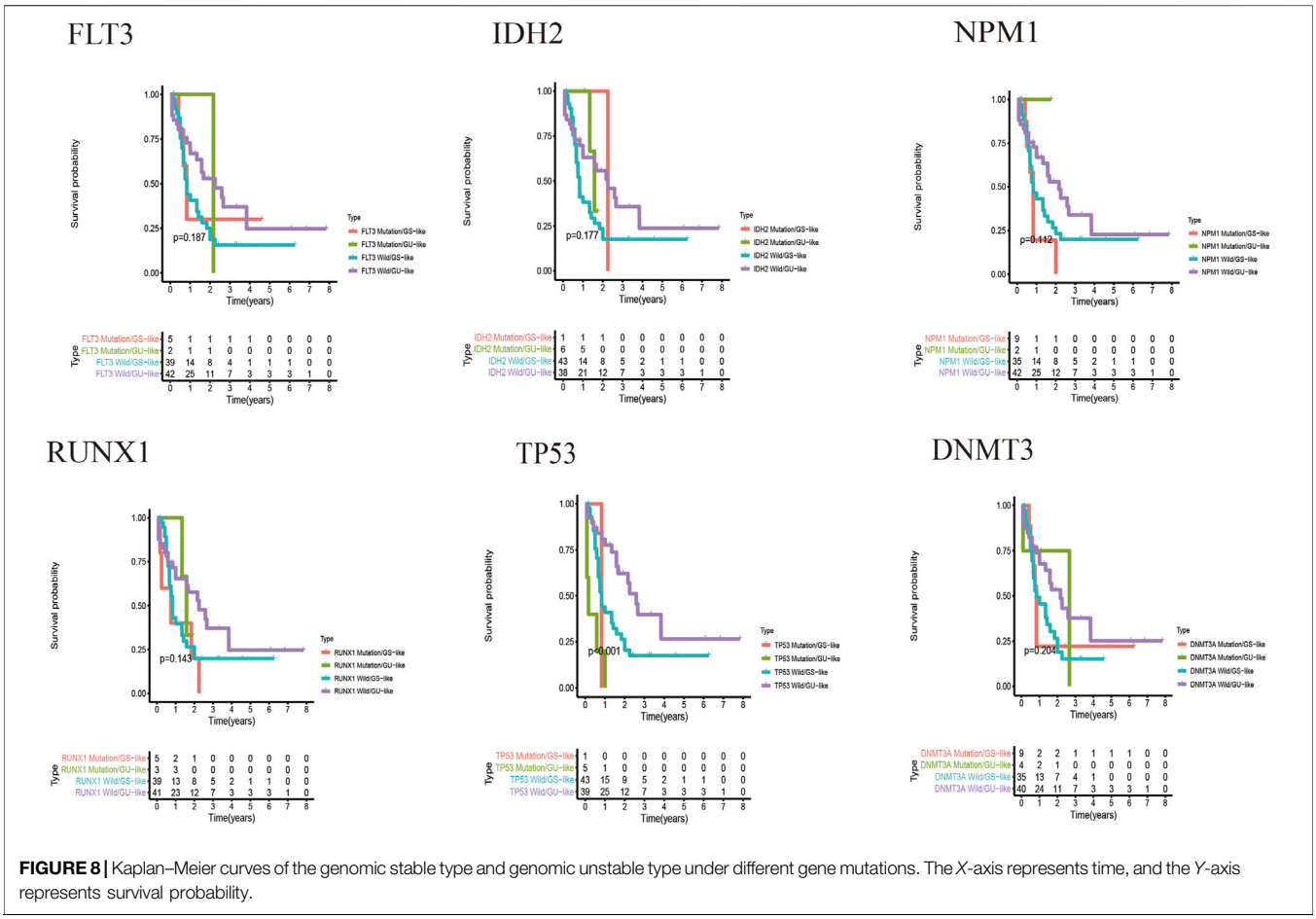


FIGURE 8 | Kaplan–Meier curves of the genomic stable type and genomic unstable type under different gene mutations. The *X*-axis represents time, and the *Y*-axis represents survival probability.

was not affected by the gender, age, FAB subtypes and mutation frequency of the patients, and the lncRNA prognostic prediction model demonstrated good applicability when patients were divided into high- and low-risk groups according to the clinicopathological characteristics, suggesting that the model is an independent index for predicting the prognosis of patients with AML.

4.2.3 Verification of Survival Prediction

According to the prognostic model, we constructed the high- and low-risk Kaplan–Meier curves of the train, test and overall groups and calculated the *p* value of the high and low survival curves. The median survival times were 1.84 and 0.67 years in the low- and high-risk groups of the overall group (**Figure 4**), 1.715 and 0.58 years in the train group, and 1.92 and 0.92 years in the test group, respectively. These results revealed that the survival rate of the low-risk group was higher than that of the high-risk group, indicating that the lncRNA prognostic prediction model showed good applicability for the survival prediction of patients with AML.

4.2.4 ROC Curve

To analyse the prognostic model, we constructed the ROC curves (**Figure 5**) of the 1-year and 5-year survival rates in the train, test and overall groups. As shown in **Figure 5**, the AUCs of the 1-year and 5-year survival rates were 0.876 and 0.713 in the train group, 0.663 and 0.799 in the test group, and 0.782 and 0.731 in the overall group, respectively. The results revealed that the lncRNA model had good prediction accuracy within 1–5 years and can predict the survival of patients with AML in other independent cohorts, and the accuracy of this model in predicting the 1-year survival rate (AUC = 0.782) was better than that of the 5-year survival rate (AUC = 0.731). The *p* values of the above-mentioned analyses were all less than 0.05.

4.2.5 Relationship Between the Prognostic Model and Clinically Known Prognostic Genes

Relationship Between Sample Risk and Gene Expression, Patient Risk and Genomic Instability

To evaluate the relationship between predictive model scores and gene expression, gene mutation frequency, and mutation-driven genes, we ranked the samples of the train, test and overall groups according to the risk scores from low to high (from left to right). Thereafter, the expression heat maps, gene mutation frequency distribution maps and box maps of related genes of the three lncRNAs in the prognostic model were drawn. As shown in **Figure 6**, the abscissa presents all samples sorted according to the increasing risk prediction value of the model. The expressions of AC073534.2 and LINC02593 decreased gradually with the

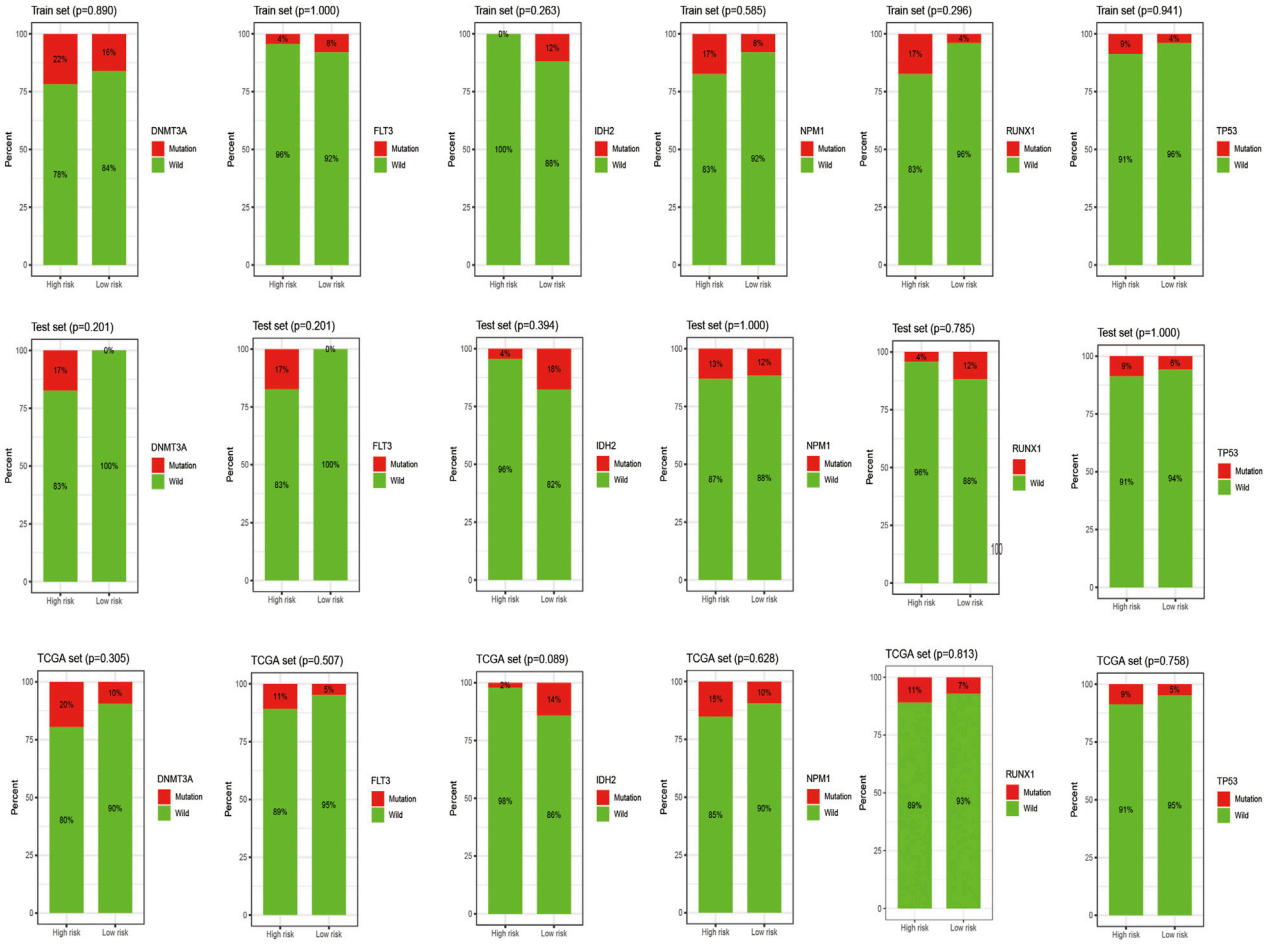


FIGURE 9 | Histogram of the mutation frequency of common mutation genes in the train, test and overall groups. The *X*-axis represents the high- and low-risk groups, and the red on the *Y*-axis represents the ratio of samples with corresponding gene mutations.

TABLE 1 | Clinical characteristics of patients from the TARGET database included in the validation study.

	Children (n = 155)
Sex	
Male	79
Female	76
NPM mutation	
Yes	7
No	143
FLT3 PM	
Yes	11
No	144
FLT3/ITD positive?	
Yes	13
No	142
CNS disease	
Yes	10
No	145
Life status	
Alive	79
Dead	76

Abbreviation: PM, point mutation; CNS, central nervous system

increase in the risk score (**Figure 6A**). The median frequency of gene mutation in the low-risk group (i.e. 16) was higher than that in the high-risk group (i.e. 12) (**Figure 6B**). The median frequency of gene mutation in the low-risk group was higher than that in the high-risk group. The median UBQLN4 expression of the genes driven by genomic instability in the low-risk group was 12.39 (mean, 13.12), whereas that of the high-risk group was 11.85 (mean, 12.16) (**Figure 6C**). From **Figure 6**, we can conclude that the frequency of the gene mutation and genomic instability are negatively correlated with the risk score of the patients.

Relationship Between the Prognostic Model and Known Mutant Genes Such as TP53

Figure 7 shows the heat map after the cluster analysis. **Figure 7A** presents the lncRNA expression heat map of the GS type (GS-like) and GU type (GU-like). **Figure 7B** presents a significant difference in the mutation frequency between the two groups ($p < 0.014$). A significant difference was found in the expression of the visible gene *LUNAR1* ($p < 0.05$) (**Figure 7C**). As presented in **Figure 7D**, the Kaplan–Meier curve showed that the survival time of patients in the GU-like group was significantly prolonged, with

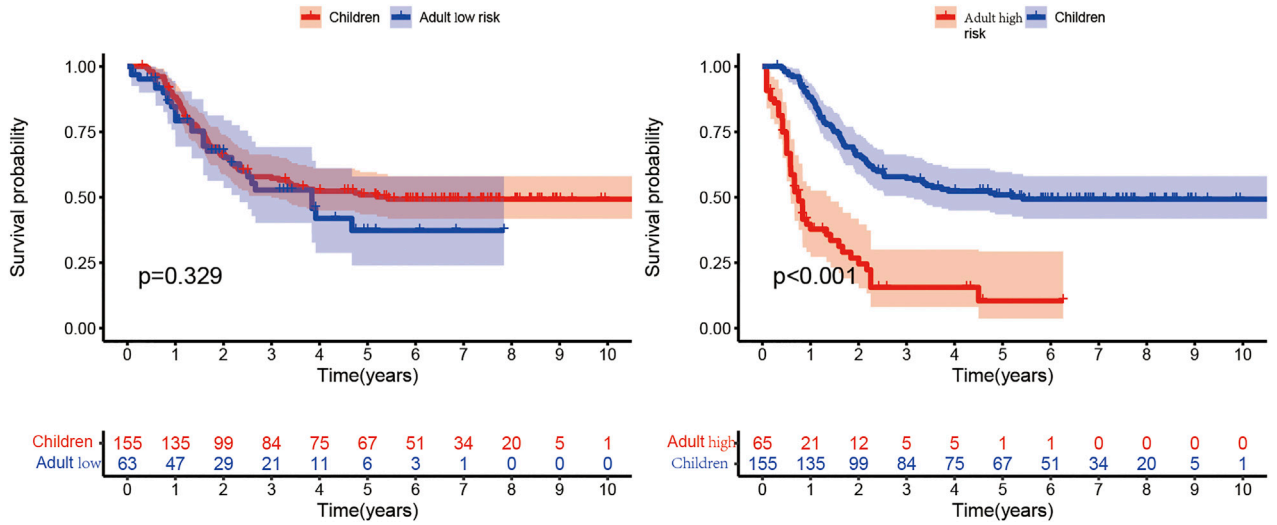


FIGURE 10 | Comparison of the children group from the TARGET database with the high- and low-risk subgroups of the adult group. In the Kaplan–Meier curve, the X- and *Y*-axes represent the time and probability of survival, respectively.

TABLE 2 | Clinical characteristics of patients from GSE106291 included in the validation study.

	High risk (n = 56)	Low risk (n = 67)	P
Age			0.908562
≥60	27	33	
<60	29	34	
Sex			0.192443
Male	26	39	
Female	30	28	
RUNX1_mutation			0.642099
Yes	11	11	
No	45	56	
RUNX1-RUNX1T1_fusion			0.476547
Yes	1	4	
No	55	63	
Treatment response			0.641233
resistant	19	20	
sensitive	34	43	
Life status			0.115523
Alive	15	27	
Dead	41	40	

Abbreviation: RUNX1, runt-related transcription factor 1; RUNX1T1, rUNX1 partner transcriptional co-repressor 1. Significant P value is in bold typeface.

a median survival time of 1.340 years, which was higher than that in the GS-like group (0.666 years). This means that patients with higher mutation frequencies have better overall survival (OS). Thereafter, the Kaplan–Meier curves of different gene mutations in the GS (GS-like) and GU (GU-like) groups were drawn. In **Figure 8**, among the first six genes with the highest mutation frequency, only the unstable gene group and the mutant TP53-positive group had a synergistic effect on the survival curve ($p < 0.001$). The specific reason is not clear, which may be related to the chemical resistance and high risk of recurrence of TP53 mutation (Döhner et al., 2010; Stengel et al., 2017; Welch, 2018; Barbosa et al., 2019). The effect of other gene mutations on the survival rate of gene instability had no clinical significance. Thereafter, we drew the bar chart of the mutation frequency of common genes in the train, test and overall groups. **Figure 9** shows no significant difference in the expressions of mutant *DNMT3A*, *FLT3*, *IDH2*, *NPM1*, *RUNX1*, and *TP53* among the groups, indicating that the known prognostic genes do not affect the prediction of high and low risk in the prognostic model and that the risk score can be used as a prediction tool independent of the current prognostic-related genes.

4.3 Applicability of the Model in Child Patients

The clinical characteristics of child samples from the TARGET database are summarized in **Table 1**. The Kaplan–Meier curve in **Figure 10** shows that the median survival time of children is 3.67 years, which is higher than that of adults. In conclusion, the prognosis of the children group was significantly better than that of the adult high-risk group, and no significant difference was found between the children and adult low-risk groups. This finding suggested that the prediction model was not suitable for disease prediction and evaluation in children.

4.4 External Verification

Patients with relatively short survival time (OS ≤ 30 days) were excluded. As mentioned in 4.2.5.2 of the Results section, high mutant TP53 expression could most independently affect OS. Patients with high TP53 mutation expression were also removed from the analysis. Finally, we performed a prognostic analysis in the validation cohort, which consisted of 123 samples with complete survival information and transcriptome data. The clinical characteristics are summarized in **Table 2**. The model was applied to stratify these patients into low-risk and high-risk groups. The median survival times of the low- and high-risk groups were 2.89 and 1.19 years, respectively, which indicated

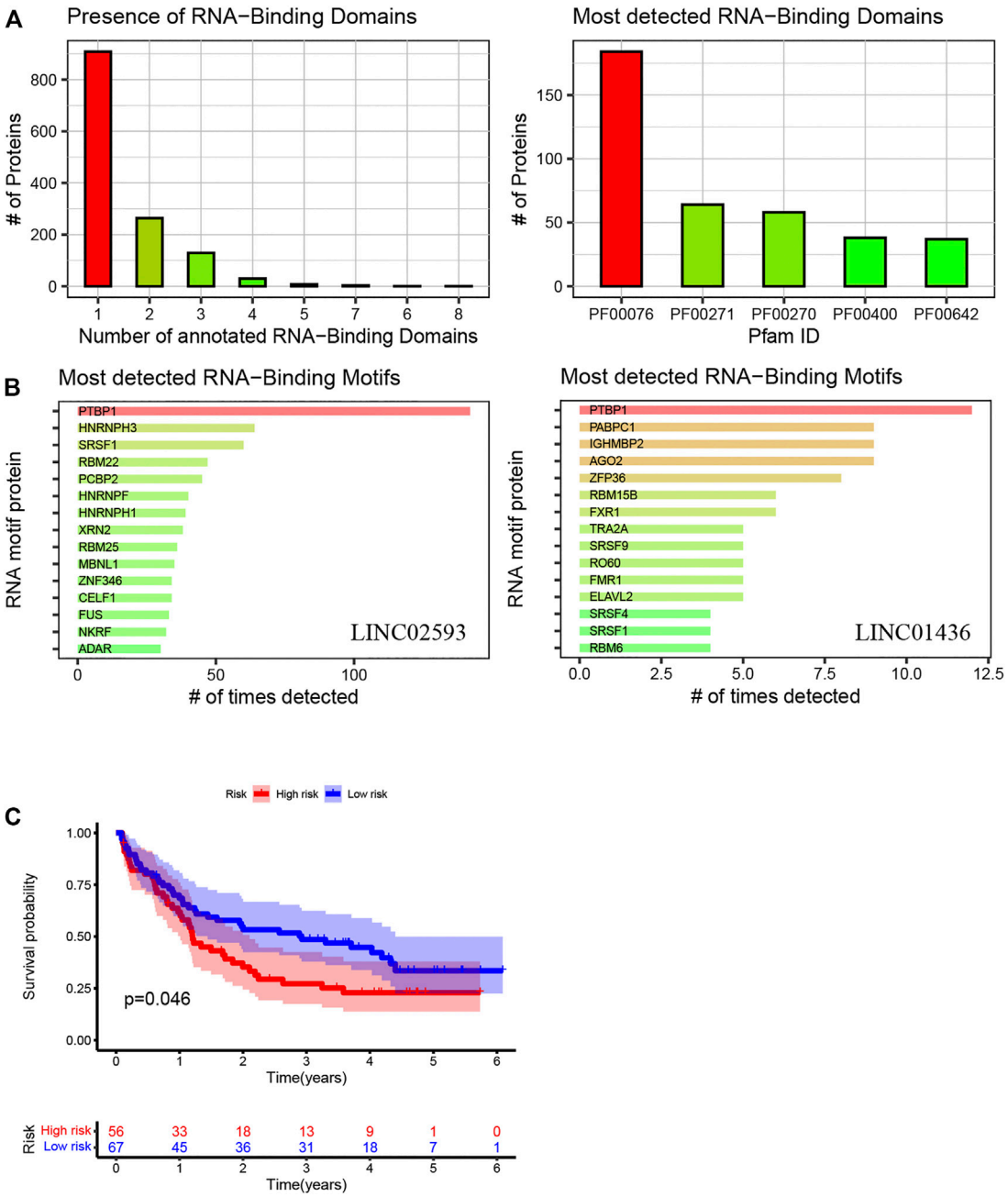


FIGURE 11 | Using the method of LNCipedia (Volders et al., 2019) and catRAPID (Armaos et al., 2021), transcripts are clustered into genes. **(A,B)** indicate the presence of protein nucleic-acid-binding domains or RNA recognition motifs in LINC01436 and LINC02593. **(C)** Kaplan–Meier curves of high- and low-risk groups in the validation cohort. The X- and *Y*-axes represent the time and probability of survival, whereas the red and blue lines represent the high- and low-risk groups, respectively.

that the validation efficiency of the model in an external cohort is acceptable (**Figure 11**).

4.5 LncRNA Expression Pattern Verification in AML Cell Lines

To determine the expression patterns of LINC01436, AC073534.2 and LINC02593 in various AML cells, we performed a FISH assay in four AML cell lines that may represent common clinical conditions. The results showed that LINC01436 was more expressed in MV4-11, but less expressed in Kasumi-1, AC073534.2 was significantly reduced in THP-1 and MV4-11 cells and LINC02593 was less found in all cell lines, but in Kasumi-1 (**Figure 12**). FLT3-ITD mutation is correlated with poor prognosis, and *AML1–ETO* fusion gene is associated with good survival in clinical practice. The above

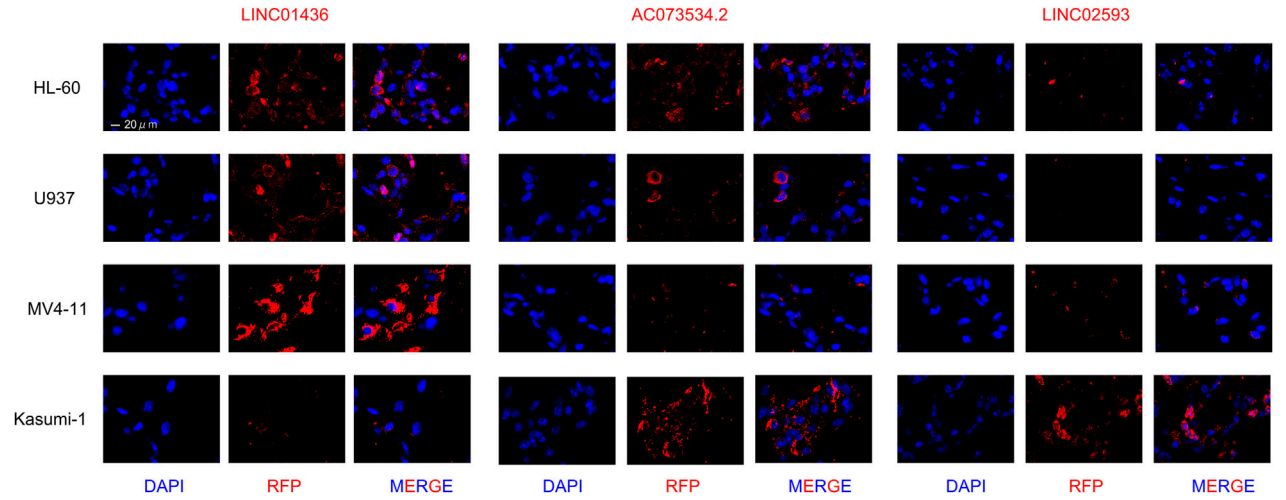


FIGURE 12 | FISH staining in LINC01436, AC073534.2, and LINC02593 in AML cell lines HL-60, U937, MV4-11, and Kasumi-1 (original magnification ×630).

expression pattern of lncRNA may indicate their role in prognosis.

5 DISCUSSION

AML is a serious threat to human health. However, a quantitative index to predict the prognosis of AML is still lacking. Previous studies on AML have mainly focused on mutant genes, namely, *FLT3*, *IDH2*, *NPM1*, *RUNX1*, *TP53,* and *DNMT3A* (Ofran, 2014; Sudhindra and Smith, 2014; Ohgami et al., 2015; Papaemmanuil et al., 2016), miRNA (Ali Syeda et al., 2020) and mRNA (Blagden and Willis, 2011). Researchers have extensively investigated the expression patterns of mRNAs and miRNAs, and many mRNAs or miRNAs have been identified as prognostic markers for patients with AML (Ge et al., 2019; Ibraheem et al., 2019; Elsayed et al., 2020). In recent years, a new class of noncoding RNA (lncRNA) has gradually become a research hotspot in various cancer fields (Ferrè et al., 2016; Paraskevopoulou and Hatzigeorgiou, 2016; Qian et al., 2019). RNA, which lacks protein-coding ability, is defined as noncoding RNA, which accounts for >98% of the human gene sequence. Approximately 90% of the noncoding sequences are transcribed, producing numerous noncoding transcripts, in which RNA with a length of more than 200 nucleotides is also known as lncRNA (Jathar et al., 2017). Some studies have reported the abnormal expression of lncRNAs in the occurrence and development of AML and found that some lncRNAs could be highly related to the prognosis (Garzon et al., 2014). Although some previous studies have confirmed a series of differences in the expression of lncRNAs in AML, the research on the value of lncRNA in predicting the clinical prognosis of AML is still limited. Except for our study, according to the expression and mutation degrees of *FLT3*, *DNMT3A*, *TP53* and other genes and chromosome changes, cases are classified as low, middle and high types; as a result, the survival time of patients was roughly estimated (Infante et al., 2018). However, a good quantitative model to analyse the survival of patients has not yet been established. Therefore, this study attempted to construct a prognostic risk model of lncRNA in patients with AML, to determine a potentially clinically applicable lncRNA prognostic model and to examine its high repeatability and practicability in different clinical groups.

Based on the mutation data of 149 samples, we used R language to divide the patients into a high and a low mutation group. Thereafter, differential lncRNA expressions of the high and low mutation groups were screened. We screened out seven lncRNAs related to prognosis, namely, AL645608.6, LINC01436, AL645608.2, AC073534.2, LINC02593, AL512413.1, and AL645608.4, using univariate Cox regression analysis and established the prognostic risk model formula based on multivariate Cox regression. Among them, the regression coefficient of LINC01436 was greater than 0, which was negatively correlated with the survival time, whereas the regression coefficients of two lncRNAs (AC073534.2 and LINC02593) were less than 0, which were positively correlated with the survival time. We extracted the regression coefficients of lncRNAs through multivariate Cox analysis and constructed three prognostic risk models composed of lncRNAs. Of the three lncRNAs, LINC01436 has been reported in gastric cancer, lung cancer and other diseases (Yuan et al., 2019; Lu et al., 2020; Xu et al., 2020; Zhang et al., 2020). However, no studies have investigated AC073534.2 and LINC02593. Furthermore, we used the prognostic model to calculate the risk score of each sample according to the median risk score. The sample was divided into high- and low-risk groups. We used R language to draw relevant heat maps, ROC curves and Kaplan–Meier curves. The ROC curve showed that the prognostic model was stable for predicting the 1-year and 5-year survival of patients with AML, indicating that the model has a good predictive ability. Moreover, under different clinical characteristics, the OS rate of the high-risk group was significantly lower than that of the low-risk group, indicating that the prognostic model can distinguish patients with different prognoses. Therefore, we are certain that the prognostic model can be used as an independent prognostic marker with a high clinical value.

However, this study has some limitations. First, our AML sample size and clinical data are limited. Second, our research results are preliminary, mainly based on a previously published dataset for secondary mining and analysis and thus lacks functional verification of lncRNAs. Therefore, further prospective studies are needed to verify our findings.

In conclusion, we have developed a lncRNA prognostic model that is significantly related to the prognosis of patients with AML. This model can accurately stratify patients and help determine whether patients are more active in treatment. Moreover, the predictive ability of the prognostic model is not influenced by clinicopathological factors such as age and sex; therefore, it has good applicability. Compared with the known prognostic biomarkers, the developed model is more convenient and intuitive in predicting the prognosis of patients with AML. If our prognosis model can be combined with the known biomarkers of AML molecules like FLT3-ITD, C-KIT mutation, et al. we can further screen high-risk groups and guide the clinical formulation of individualised treatment plans. Therefore, we believe that the prognostic model has a wide application prospect.

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

WD wrote the manuscript and prepared figures. YL reviewed the data, ZZ and YS edited the manuscript.

FUNDING

This work was supported by the General Program of National Natural Science Foundation of China (82070148), General project of China Postdoctoral Science Foundation (2019M650122), The Youth talent science and technology project of Changzhou Health Commission (QN202013) and Science and Technology Bureau foundation application project of Changzhou (CJ20210061).

ACKNOWLEDGMENTS

All authors collaborated in the collection and interpretation of the data and contributed to the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2022.818905/full#supplementary-material

REFERENCES

Al-Kershi, S., Bhayadia, R., Ng, M., Verboon, L., Emmrich, S., Gack, L., et al. (2019). The Stem Cell-specific Long Noncoding RNA HOXA10-AS in the Pathogenesis of KMT2A-Rearranged Leukemia. *Blood Adv.* 3, 4252–4263. doi:10.1182/bloodadvances.2019032029

Ali Syeda, Z., Langden, S. S. S., Munkhzul, C., Lee, M., and Song, S. J. (2020). Regulatory Mechanism of MicroRNA Expression in Cancer. *Ijms* 21, 1723. doi:10.3390/ijms21051723

Armaos, A., Colantoni, A., Proietti, G., Rupert, J., and Tartaglia, G. G. (2021). catRAPID Omics v2.0: Going Deeper and Wider in the Prediction of Protein-RNA Interactions. *Nucleic Acids Res.* 49, W72–w79. doi:10.1093/nar/gkab393

Barbosa, K., Li, S., Adams, P. D., and Deshpande, A. J. (2019). The Role of TP53 in Acute Myeloid Leukemia: Challenges and Opportunities. *Genes Chromosomes Cancer* 58, 875–888. doi:10.1002/gcc.22796

Blagden, S. P., and Willis, A. E. (2011). The Biological and Therapeutic Relevance of mRNA Translation in Cancer. *Nat. Rev. Clin. Oncol.* 8, 280–291. doi:10.1038/nrclinonc.2011.16

Döhner, H., Estey, E. H., Amadori, S., Appelbaum, F. R., Büchner, T., Burnett, A. K., et al. (2010). Diagnosis and Management of Acute Myeloid Leukemia in Adults: Recommendations from an International Expert Panel, on Behalf of the European LeukemiaNet. *Blood* 115, 453–474. doi:10.1182/blood-2009-07-235358

Duncan, D. J., Vandenberghe, M. E., Scott, M. L. J., and Barker, C. (2019). Fast Fluorescence *In Situ* Hybridisation for the Enhanced Detection of MET in Non-small Cell Lung Cancer. *PLoS One* 14, e0223926. doi:10.1371/journal.pone.0223926

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

Ferrè, F., Colantoni, A., and Helmer-Citterich, M. (2016). Revealing Protein-lncRNA Interaction. *Brief Bioinform* 17, 106–116. doi:10.1093/bib/bbv031

Garzon, R., Volinia, S., Papaioannou, D., Nicolet, D., Kohlschmidt, J., Yan, P. S., et al. (2014). Expression and Prognostic Impact of lncRNAs in Acute Myeloid Leukemia. *Proc. Natl. Acad. Sci. USA* 111, 18679–18684. doi:10.1073/pnas.1422050112

Ge, Y., Schuster, M. B., Pundhir, S., Rapin, N., Bagger, F. O., Sidiropoulos, N., et al. (2019). The Splicing Factor RBM25 Controls MYC Activity in Acute Myeloid Leukemia. *Nat. Commun.* 10, 172. doi:10.1038/s41467-018-08076-y

Gourvest, M., De Clara, E., Wu, H. C., Touriol, C., Meggetto, F., De Thé, H., et al. (2021). A Novel Leukemic Route of Mutant NPM1 through Nuclear Import of the Overexpressed Long Noncoding RNA LONA. *Leukemia* 35, 2784–2798. doi:10.1038/s41375-021-01307-0

Herold, T., Jurinovic, V., Batcha, A. M. N., Bamopoulos, S. A., Rothenberg-Thurley, M., Ksienzyk, B., et al. (2018). A 29-gene and Cytogenetic Score for the Prediction of Resistance to Induction Treatment in Acute Myeloid Leukemia. *Haematologica* 103, 456–465. doi:10.3324/haematol.2017.178442

Ibraheem, F. M., Badawy, R., Ayoub, M. A., Hassan, N. M., and Mostafa, M. N. (2019). SALL4 Gene Expression in Acute Myeloid Leukemia. *Asian Pac. J. Cancer Prev.* 20, 3121–3127. doi:10.31557/apjcp.2019.20.10.3121

Infante, M. S., Piris, M. Á., and Hernández-Rivas, J. Á. (2018). Molecular Alterations in Acute Myeloid Leukemia and Their Clinical and Therapeutical Implications. *Medicina Clínica (English Edition)* 151, 362–367. doi:10.1016/j.medcli.2018.05.00210.1016/j.medcle.2018.05.044

Jathar, S., Kumar, V., Srivastava, J., and Tripathi, V. (2017). Technological Developments in lncRNA Biology. *Adv. Exp. Med. Biol.* 1008, 283–323. doi:10.1007/978-981-10-5203-3_10

Liang, Y., Song, X., Li, Y., Chen, B., Zhao, W., Wang, L., et al. (2020). LncRNA BCRT1 Promotes Breast Cancer Progression by Targeting miR-1303/PTBP3 axis. *Mol. Cancer* 19, 85. doi:10.1186/s12943-020-01206-5

Lu, M.-d., Liu, D., and Li, Y.-x. (2020). LINC01436 Promotes the Progression of Gastric Cancer via Regulating miR-513a-5p/APE1 Axis. *Ott* Vol. 13, 10607–10619. doi:10.2147/ott.S257747

Ofran, Y. (2014). Concealed Dagger in FLT3/ITD+ AML. *Blood* 124, 2317–2319. doi:10.1182/blood-2014-09-597690

Ohgami, R. S., Ma, L., Merker, J. D., Gotlib, J. R., Schrijver, I., Zehnder, J. L., et al. (2015). Next-generation Sequencing of Acute Myeloid Leukemia Identifies the Significance of TP53, U2AF1, ASXL1, and TET2 Mutations. *Mod. Pathol.* 28, 706–714. doi:10.1038/modpathol.2014.160

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

Paraskevopoulou, M. D., and Hatzigeorgiou, A. G. (2016). Analyzing MiRNA-LncRNA Interactions. *Methods Mol. Biol. (Clifton, N.J.)* 1402, 271–286. doi:10.1007/978-1-4939-3378-5_21

Pelcovits, A., and Niroula, R. (2013). Acute Myeloid Leukemia: A Review. *R. Med. J.* 103, 38–40.

Peng, W.-X., Koirala, P., and Mo, Y.-Y. (2017). LncRNA-mediated Regulation of Cell Signaling in Cancer. *Oncogene* 36, 5661–5667. doi:10.1038/onc.2017.184

Qian, X., Zhao, J., Yeung, P. Y., Zhang, Q. C., and Kwok, C. K. (2019). Revealing lncRNA Structures and Interactions by Sequencing-Based Approaches. *Trends Biochemical Sciences* 44, 33–52. doi:10.1016/j.tibs.2018.09.012

Sasine, J. P., and Schiller, G. J. (2016). Acute Myeloid Leukemia: How Do We Measure Success? *Curr. Hematol. Malig Rep.* 11, 528–536. doi:10.1007/s11899-016-0346-x

Song, X., Peng, Y., Wang, X., Chen, Y., Jin, L., Yang, T., et al. (2018). Incidence, Survival, and Risk Factors for Adults with Acute Myeloid Leukemia Not Otherwise Specified and Acute Myeloid Leukemia with Recurrent Genetic Abnormalities: Analysis of the Surveillance, Epidemiology, and End Results (SEER) Database, 2001-2013. *Acta Haematol.* 139, 115–127. doi:10.1159/000486228

Stengel, A., Kern, W., Haferlach, T., Meggendorfer, M., Fasan, A., and Haferlach, C. (2017). The Impact of TP53 Mutations and TP53 Deletions on Survival Varies between AML, ALL, MDS and CLL: an Analysis of 3307 Cases. *Leukemia* 31, 705–711. doi:10.1038/leu.2016.263

Sudhindra, A., and Smith, C. C. (2014). FLT3 Inhibitors in AML: Are We There yet? *Curr. Hematol. Malig Rep.* 9, 174–185. doi:10.1007/s11899-014-0203-8

Trinh, B. Q., Ummarino, S., Zhang, Y., Ebralidze, A. K., Bassal, M. A., Nguyen, T. M., et al. (2021). Myeloid lncRNA LOUP Mediates Opposing Regulatory Effects of RUNX1 and RUNX1-ETO in T(8;21) AML. *Blood* 138, 1331–1344. doi:10.1182/blood.2020007920

Volders, P.-J., Anckaert, J., Verheggen, K., Nuytens, J., Martens, L., Mestdagh, P., et al. (2019). LNCipedia 5: towards a Reference Set of Human Long Non-coding RNAs. *Nucleic Acids Res.* 47, D135–d139. doi:10.1093/nar/gky1031

Wei, G. H., and Wang, X. (2017). lncRNA MEG3 Inhibit Proliferation and Metastasis of Gastric Cancer via P53 Signaling Pathway. *Eur. Rev. Med. Pharmacol. Sci.* 21, 3850–3856.

Welch, J. S. (2018). Patterns of Mutations in TP53 Mutated AML. *Best Pract. Res. Clin. Haematol.* 31, 379–383. doi:10.1016/j.beha.2018.09.010

Xu, Y., Dong, M., Wang, J., Zhao, W., and Jiao, M. (2020). LINC01436 Inhibited miR-585-3p Expression and Upregulated MAPK1 Expression to Promote Gastric Cancer Progression. *Dig. Dis. Sci.* 66, 1885–1894. doi:10.1007/s10620-020-06487-w

Yin, Z., Shen, H., Gu, C. m., Zhang, M. q., Liu, Z., Huang, J., et al. (2021). MiRNA-142-3P and FUS Can Be Sponged by Long Noncoding RNA DUBR to Promote Cell Proliferation in Acute Myeloid Leukemia. *Front. Mol. Biosci.* 8, 754936. doi:10.3389/fmolb.2021.754936

Yuan, S., Xiang, Y., Wang, G., Zhou, M., Meng, G., Liu, Q., et al. (2019). Hypoxia-sensitive LINC 01436 Is Regulated by E2F6 and Acts as an Oncogene by Targeting miR-30a-3p in Non-small Cell Lung Cancer. *Mol. Oncol.* 13, 840–856. doi:10.1002/1878-0261.12437

Zhang, Y., Yang, G., He, X., Chen, S., Zhang, F., and Fang, X. (2020). LINC01436, Regulating miR-585 and FBXO11 , Is an Oncogenic lncRNA in the Progression of Gastric Cancer. *Cell Biol Int* 44, 882–893. doi:10.1002/cbin.11287

Zhao, W., Geng, D., Li, S., Chen, Z., and Sun, M. (2018). LncRNA HOTAIR Influences Cell Growth, Migration, Invasion, and Apoptosis via the miR-20a-5p/HMGA2axis in Breast Cancer. *Cancer Med.* 7, 842–855. doi:10.1002/cam4.1353

Zhu, G., Luo, H., Feng, Y., Guryanova, O. A., Xu, J., Chen, S., et al. (1956). HOXBLINC Long Non-coding RNA Activation Promotes Leukemogenesis in NPM1-Mutant Acute Myeloid Leukemia. *Nat. Commun.* 12. doi:10.1038/s41467-021-22095-2

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

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 Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 01 March 2022
doi: 10.3389/fbioe.2022.823619

A Prognostic Model Based on mRNA Expression Analysis of Esophageal Squamous Cell Carcinoma

Ke Liu[1,2], Ye-Lin Jiao[3], Liu-Qing Shen[2], Pan Chen[2], Ying Zhao[2], Meng-Xiang Li[1,2], Bian-Li Gu[2], Zi-Jun Lan[2], Hao-Jie Ruan[1], Qi-Wei Liu[2], Feng-Bo Xu[2], Xiang Yuan[2], Yi-Jun Qi[2] and She-Gan Gao[1,2]**

[1]School of Information Engineering, Henan University of Science and Technology, Luoyang, China, [2]Henan Key Laboratory of Microbiome and Esophageal Cancer Prevention and Treatment, Henan Key Laboratory of Cancer Epigenetics, Cancer Hospital, The First Affiliated Hospital (College of Clinical Medicine) of Henan University of Science and Technology, Luoyang, China, [3]Department of Pathology, Luo Yang First Peoples's Hospital, Luoyang, China

Background: The aim of this study was to identify prognostic markers for esophageal squamous cell carcinoma (ESCC) and build an effective prognostic nomogram for ESCC.

Methods: A total of 365 patients with ESCC from three medical centers were divided into four cohorts. In the discovery phase of the study, we analyzed transcriptional data from 179 cancer tissue samples and identified nine marker genes using edgeR and rbsurv packages. In the training phase, penalized Cox regression was used to select the best marker genes and clinical characteristics in the 179 samples. In the verification phase, these marker genes and clinical characteristics were verified by internal validation cohort (n = 58) and two external cohorts (n = 81, n = 105).

Results: We constructed and verified a nomogram model based on multiple clinicopathologic characteristics and gene expression of a patient cohort undergoing esophagectomy and adjuvant radiochemotherapy. The predictive accuracy for 4-year overall survival (OS) indicated by the C-index was 0.75 (95% CI, 0.72–0.78), which was statistically significantly higher than that of the American Joint Committee on Cancer (AJCC) seventh edition (0.65). Furthermore, we found two marker genes (TM9SF1, PDZK1IP) directly related to the OS of esophageal cancer.

Conclusion: The nomogram presented in this study can accurately and impersonally predict the prognosis of ESCC patients after partial resection of the esophagus. More research is required to determine whether it can be applied to other patient populations. Moreover, we found two marker genes directly related to the prognosis of ESCC, which will provide a basis for future research.

Keywords: esophageal squamous cell carcinoma, marker gene, nomogram, prognosis, dimensionality reduction (DR)

OPEN ACCESS

Edited by:
Giulia Russo,
University of Catania, Italy

Reviewed by:
Amirhosein Kefayat,
Isfahan University of Medical Sciences, Iran
Lei-Lei Wu,
Tongji University, China

****Correspondence:***
She-Gan Gao
gsg112258@163.com
Yi-Jun Qi
qiyijun@haust.edu.cn

Specialty section:
This article was submitted to Preclinical Cell and Gene Therapy, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *27 November 2021*
Accepted: *31 January 2022*
Published: *01 March 2022*

Citation:
Liu K, Jiao Y-L, Shen L-Q, Chen P, Zhao Y, Li M-X, Gu B-L, Lan Z-J, Ruan H-J, Liu Q-W, Xu F-B, Yuan X, Qi Y-J and Gao S-G (2022) A Prognostic Model Based on mRNA Expression Analysis of Esophageal Squamous Cell Carcinoma. Front. Bioeng. Biotechnol. 10:823619. doi: 10.3389/fbioe.2022.823619

INTRODUCTION

Esophageal cancer (EC) is a very common digestive tract tumor with the sixth highest mortality rate in the world, and there are about 150,000 deaths from EC in China every year. (Rubenstein and Shaheen, 2015; Liang et al., 2017) The histological types of EC mainly include esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma (EAC). In 2018, more than 570,000 people

worldwide were diagnosed with EC, and more than 500,000 people died of EC in the same year. (Enzinger and Mayer, 2003) Most of the new cases and deaths in the world occur in less developed areas. (Gao et al., 2018) Histologically, approximately 90% of EC cases in China are ESCC. (International Agency for Research on Cnacer, 2012) ESCC is characterized by high aggressiveness and poor prognosis. (Qi et al., 2012) Despite various comprehensive treatments, including surgery, radiotherapy, and chemotherapy, the 5-year survival rate of patients is still less than 22%. (Zhao et al., 2012) The significant geographical variation in incidence means that environmental and genetic factors could play major roles in the development of EC. Smoking and drinking are as known as risk factors for EC, whereas high consumption of vegetables and fruits is likely to prevent EC. (Ando, 2015; Owen et al., 2018; Wu et al., 2018)

At present, the tumor-node-metastasis (TNM) staging system ignores the important clinical factors of tumor prognosis, and the great difference in clinical course leads to the inaccuracy of TNM staging, so it is necessary to establish a new ESCC prognosis grading system. (Cao et al., 2016; Duan et al., 2016) A nomogram can successfully quantify risk prediction by incorporating and illustrating important factors for tumor prognosis. (Wierda et al., 2007; Zhang et al., 2018; Sun et al., 2019) Compared with the TNM staging system, a nomogram can not only predict the survival of all types of cancer patients more accurately but also quantify the outcome of survival prediction by using clinical factors and other factors affecting the prognosis of cancer. Thus, the nomogram is a new prognostic criterion that produces a quantified risk probability of clinical survival by creating a linear graph of the prediction model instead of the traditional method. (Mariani et al., 2005; Sternberg, 2006; Wang et al., 2006) We hypothesized that combining multiple clinicopathologic characteristics and signature gene expression levels can improve the prediction result of ESCC, but there remains a paucity of reliable genetic markers. TGF-β1 is an efficient prognostic biomarker for ESCC patients. HER-2 can be used as a potential molecular marker for ESCC molecular typing. But, HER-2 is not an efficient prognostic biomarker and potential therapeutic target for Iranian ESCC patients. (Heidarpour et al., 2020) Using the partial likelihood of the Cox model, we recently excavated a gene set that is closely related to the overall survival (OS) of ESCC patients.

To the best of the authors' knowledge, this paper presents the first ESCC nomogram model based on multiple clinicopathologic characteristics and gene expression of a patient cohort undergoing esophagectomy and adjuvant radiochemotherapy. Furthermore, we used an independent cohort from The Cancer Genome Atlas (TCGA) database for external validation. Another independent cohort of 105 ESCC patients was employed to verify the effectiveness of the gene that we found. This study also compared the nomogram with the TNM staging system, proving that the model is more effective in survival prediction.

MATERIALS AND METHODS

Patient Selection

We downloaded transcriptome sequencing data of 179 ESCC samples from the GEO database (GSE53625) (Li et al., 2014). This set of samples, which served as the primary cohort, was from the Chinese Academy of Medical Sciences (CAMS). For internal validation, we used a computer to randomly select 58 samples from the primary cohort and denoted this set as the internal validation cohort.

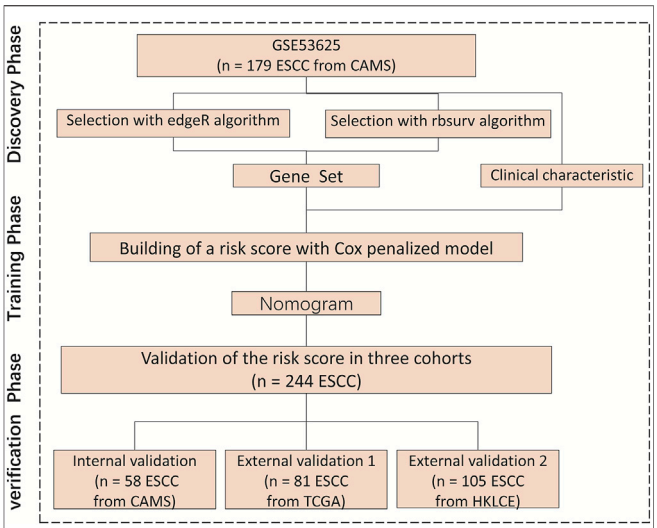


FIGURE 1 | Research flowchart, HKLCE, The First Affiliated Hospital of Henan University of Science and Technology. EdgeR and rbsurv algorithm, School of Information Engineering, Henan University of Science and Technology.

For external validation, we first selected samples from the open-access and public TCGA database. Transcriptome sequencing data and follow-up data of 81 samples were downloaded from the TCGA database and denoted as the external validation one cohort. A second cohort of 105 ESCC samples for external validation included 38 samples from Anyang Cancer Hospital (ACH), The Fourth Affiliated Hospital of Henan University of Science and Technology, and 67 samples from Henan Key Laboratory of Cancer Epigenetics (HKLCE), The First Affiliated Hospital of Henan University of Science and Technology. This cohort was denoted as external validation two cohort.

Follow-Up and Classification of Cause of Death

Most of the patients were followed up for 48–72 months. In this study, the statistics were made according to the 4-year survival period. Survivors over 48 months after surgery were counted as living, and survival periods greater than 48 months were calculated as 48 months.

Study Design

We divided this study into three phases to identify and validate OS-related clinical characteristics and gene sets in ESCC patients. During the discovery phase, we processed the transcriptome sequencing data of the primary cohort to obtain 16,738 genes and then selected important gene sets through two algorithms.

During the training phase, a penalized Cox regression model was used to identify the best gene sets (**Figure 1**). During the verification phase, the gene sets that we chose were validated in the internal validation cohort, external validation one cohort, and external validation two cohort. These three cohorts included patients from multiple medical centers.

Meanwhile, We used qPCR to verify external validation sample 2, in order to obtain results widely used in clinical practice, formalin fixed paraffin embedded samples were used with a minimum tumor cell composition of 80%. Our research program was approved by the ethics committees of the research centers, and 105 samples were reassessed and confirmed by pathologists.

Dimension Reduction Process

Initially, GSE53625 data were processed with the annotation package to obtain expression spectrum and probe ID. The gene expression value was combined with the annotation file to obtain the complete gene expression value, and the edgeR package was used to find the differential gene. Then, the rbsurv package was used to calculate the dimensionality reduction of the differential gene (max.n.genes = 10, n. iter = 10, n. fold = 3, n. seq = 3, seed = 1,234).

Construction and Validation of the Nomogram

For the development of the nomogram, we found a number of clinical characteristics that have been shown to be associated with survival as a prognostic characteristic. (Sun et al., 2019) These clinical characteristics ($p < 0.05$) included Age, Sex, Smoking, Drinking, Tumor grade, Tumor location, T stage, N stage, TNM stage, Arrhythmia, Pneumonia, Anastomotic leak, and Adjuvant therapy. For each clinical characteristic, we used a multivariate Cox proportional risk model to evaluate the projected 4-year OS. Nomogram validation was divided into the following three stages. First, internal validation was conducted using the internal validation cohort, and the C-index was estimated by analyzing the area under the curve of the receiver operating characteristic curve. Next, by means of regression analysis, the correction curve was obtained to judge whether the predicted survival probability was consistent with the observed survival probability. (Qiu et al., 2017) The calibration curve adopted Bootstrap resampling (1,000 resampling). Last, external validation was conducted using the external validation cohorts, and Cox regression analysis was performed using the total score of each patient as an independent characteristic. The C-index and calibration curve were obtained by regression analysis.

Statistical Analyses

Overall survival was defined as the time from the date of initial treatment until the date of death because of ESCC or the date of the last follow-up. The OS curve was estimated by the Kaplan-Meier method and compared with the log-rank test stratified by prognostic factors. The rbsurv packages were used to reduce the dimensions of the data within RStudio Version 1.1.463 software. We built a nomogram in a previous study. On the basis of multivariate Cox analysis results, this nomogram was compiled by R through survival and the RMS software package. (Frank and Harrell, 2014)

TABLE 1 | Demographics and clinicopathologic characteristics of patients with ESCC.

Characteristics	*N*	Primary cohort (*n* = 179)		
		Hazard ratio	CI95	*p*-value
Age	—	1.62	1.1–2.39	0.015
<60	91	References	—	—
≥60	88	1.6	1.1–2.3	0.02
Sex	—	0.81	0.5–1.31	0.393
Female	33	Reference	—	—
Male	146	0.78	0.49–1.3	0.307
Smoking	—	0.75	0.5–1.11	0.147
No	65	Reference	—	—
Yes	114	0.75	0.37–1.1	0.145
Drinking	—	0.86	0.58–1.27	0.449
No	73	Reference	—	—
Yes	106	1.43	0.79–2.6	0.455
Tumor location	—	1.17	0.86–1.6	0.309
Lower	62	Reference	—	—
Middle	97	1.1	0.74–1.7	0.562
Upper	20	1.7	0.9–3.1	0.101
Tumor grade	—	1.24	1.03–1.48	0.020
Moderately	98	References	—	—
Poorly	49	1.63	1.07–2.5	—
Well	32	0.99	0.57–1.7	—
T stage	—	1.28	0.97–1.69	0.077
T1	12	References	—	—
T2	27	1.1	1.25–2.0	0.863
T3	110	1	0.7–2.3	0.935
T4	30	1.7	0.72–4.0	0.226
N stage	—	1.44	1.18–1.76	<0.001
N0	83	Reference	—	—
N1	62	1.30	0.64–2.6	0.002
N2	22	1.27	0.49–3.3	0.017
N3	12	1.82	0.66–5	0,004
TNM stage	—	2.12	1.47–3.05	<0.001
I	10	Reference	—	—
II	77	1.75	0.55–5.8	—
III	92	3.6	1.14–11.5	—
Arrhythmia	—	1.10	0.71–1.71	0.667
No	126	Reference	—	—
Yes	43	1.1	0.73–0.17	—
Pneumonia	—	1.46	0.74–2.89	0.278
No	164	Reference	—	—
Yes	15	1.4	0.72–2.8	031
Anastomotic leak	—	1.34	0.62–2.9	0.450
No	127	Reference	—	—
Yes	12	1.3	0.63–2.7	0.504
Adjuvant therapy	—	1.38	0.92–2.07	0.115
No	45	Reference	—	—
Unknown	30	2.7	1.4–5.1	0.003
Yes	104	2.3	1.3–3.9	0.003

RESULTS

Clinicopathologic Characteristics of Patients

The research flowchart is shown in **Figure 1**. The demographics and clinical characteristics of patients with ESCC are presented in **Table 1**. In the primary cohort, the median follow-up time was

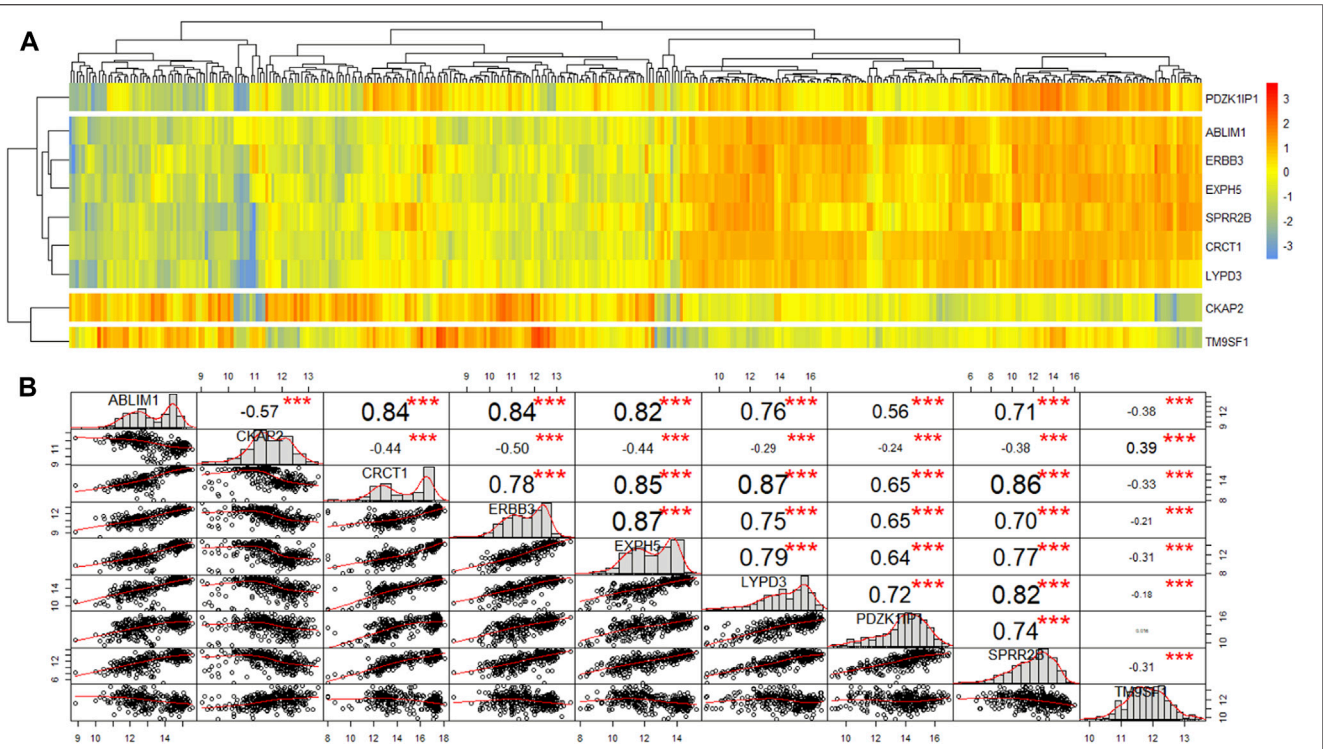


FIGURE 2 | The heatmap and similarity analysis of the expression of nine genes for patients from primary cohort. **(A)** Heatmap demonstrating unsupervised hierarchical clustering of nine genes for patients from primary cohort. **(B)** Similarity analysis of the expression of nine genes for patients from primary cohort.

34.7 months (range, 23.4–45.9 months). For the external validation two cohort, of 140 patients with ESCC who received partial esophagectomy during the study period, 105 met the inclusion criteria to enter this study. The median survival time of these 105 patients was 35 months (95% CI, 27.9–42 years). **Table 1** lists the demographic and clinicopathological characteristics of patients in the primary cohorts.

Independent Prognostic Factors in the Training Set

The results of the univariable analysis are shown in **Table 1**. Younger age (<60 vs. ≥60 $p = 0.015$) and TNM stage I (I vs. II vs. III $p < 0.001$) were associated with better prognosis. In addition, age ($p = 0.015$ vs. $p = 0.037$) and TNM stage ($p < 0.001$ vs. $p = 0.001$) were correlated with OS in both the primary cohort ($n = 179$) and the internal validation cohort ($n = 58$).

Selection of Gene Set

In total, nine genes were excavated by two algorithms, the edgeR package and rbsurv package. First, heatmap and similarity analysis were performed for these genes. Heatmap demonstrating unsupervised hierarchical clustering of nine genes for patients from primary cohort (**Figure 2A**). These nine genes were clustered into four groups with obvious differences, among which TM9SF1, CKAP2 and PDZK1IP1 were independently grouped. CKAP2 and TM9SF1 were highly expressed in the left half, while PDZK1IP1 was highly expressed in the right half. Correlation analysis of the nine genes found that they were directly related to each other except for TM9SF1 and PDZK1IP1 (**Figure 2B**). Three stars in **Figure 2B** represent p values less than 0.001. **Figure 2** shows that TM9SF1 and PDZK1IP1 are independent prognostic factors and can be used as marker genes.

Next, we used univariate and multivariate Cox analyses to discriminate the marker genes in the primary cohort ($n = 179$). The results are shown in **Table 2**. In the univariate cox analysis, all nine genes were correlated with survival, and six genes with p-value less than 0.001 were selected for multivariate Cox analysis. Multivariate Cox analysis demonstrated that PDZK1IP1 (high expression vs. low expression, $p = 0.031$) and TM9SF1 (high expression vs. low expression, $p < 0.001$) were independent risk factors for OS (**Table 2**). Therefore, PDZK1IP1 and TM9SF1 were defined as gene sets directly related to prognosis.

To classify gene expression values, we determine cutoff values using ggplot2 packages in the primary cohort. The results are listed in Appendix **Figure 3A** and **Figure 3C** (TM9SF1 cutpoint = 12.4, PDZK1IP1 cutpoint = 14.78). A violin plot of the marker genes was drawn using primary cohort. The expression value of PDZK1IP1 gene was high in normal, which is significantly different from cancer (**Figure 3D**, $p < 0.001$). TM9SF1 gene has high expression value in cancer, which is significantly different from normal (**Figure 3B**, $p < 0.001$).

TABLE 2 | Univariate and multivariate Cox analysis of the expression of nine genes for patients from primary cohort (n = 179).

Characteristics	Univariate cox analysis			Multivariate cox analysis		
	Hazard ratio	95% CI	*p*-value	Hazard ratio	95% CI	*p*-value
ABLIM1	0.5	0.34–0.73	<**0.001**	0.75	0.48–1.16	0.191
CKAP2	2.14	1.24–3.71	0.006	—	—	—
CRCT1	0.46	0.30–0.70	<**0.001**	0.80	0.50–1.28	0.349
ERBB3	0.46	0.31–0.68	<**0.001**	0.67	0.43–1.04	0.075
EXPH5	0.4	0.22–0.72	0.002	—	—	—
LYPD3	0.35	0.19–0.64	<**0.001**	0.52	0.26–1.03	0.059
PDZK1IP1	0.3	0.14–0.61	<**0.001**	0.38	0.16–0.91	**0.031**
SPRR2B	0.45	0.27–0.75	0.002	—	—	—
TM9SF1	2.13	1.44–3.14	<**0.001**	2.09	1.38–3.17	<**0.001**

*The bold values denoted in **Table 2** represent selected marker genes in this study.*

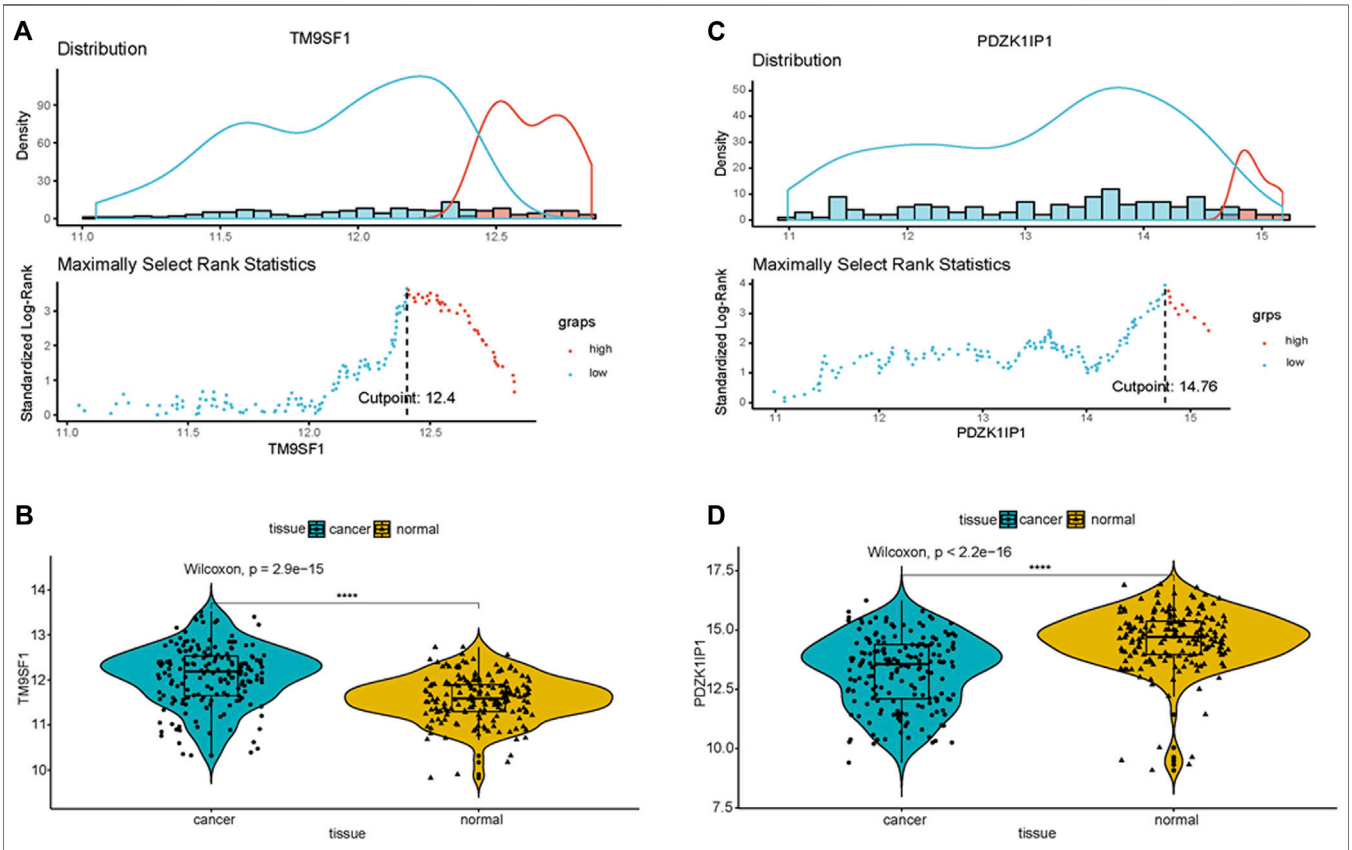


FIGURE 3 | PDZK1IP1 and TM9SF1 were defined as marker genes **(A,C)** Cutpoint of the marker genes was obtained. **(B,D)** Violin plot showed that the expression value of PDZK1IP1 gene was high in normal, TM9SF1 gene has high expression value in cancer.

Independent Prognostic Factors in the Primary Cohort

Univariate Cox analysis results show that, in the primary cohort, the prognostic factors that predicted poor OS were age ≥60 years, TNM stages II and III, high expression of TM9SF1, and low expression of PDZK1IP1. We performed a multivariate Cox analysis of gene set and clinical clinicopathologic characteristics. The results are listed in Appendix **Figure 4**. Multivariate Cox analysis demonstrated that age (p = 0.031), TNM (p = 0.004), PDZK1IP1 (p = 0.001), and TM9SF1 (p = 0.001) were independent risk factors for OS.

Nomogram Development and Validation

The results of multivariate cox analysis were used to establish a nomogram for the predicted 4-year OS (**Figure 5A**). As age, TNM, PDZK1IP1, and TM9SF1 were independent risk factors for survival in multivariate Cox analysis, these variables were incorporated into the nomogram. In the internal validation, the predictive accuracy for 4-year OS as indicated by the C-index was 0.75. The 4-year OS probabilistic calibration chart shows that the actual observation results have a high correlation with the prediction results of the nomogram (**Figure 5B**).

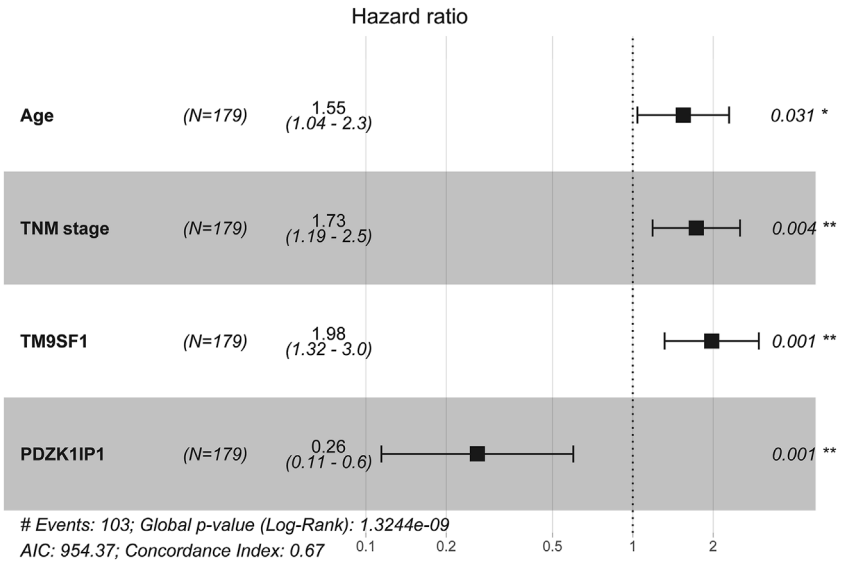


FIGURE 4 | Multivariate cox analysis of the gene set and clinicopathologic characteristics from primary Cohort (*N* = 179).

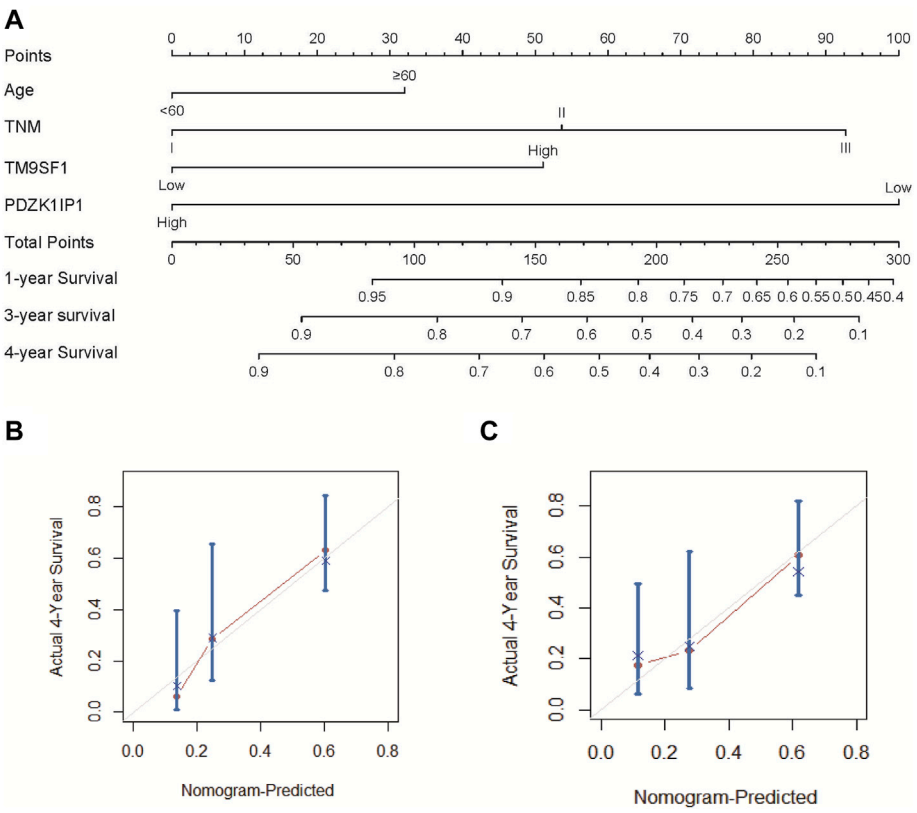


FIGURE 5 | Nomogram and calibration curve of ESCC patients. **(A)** To use the nomogram, each variable has a patient's assignment on its axis, and a line is drawn upward to determine the number of points for each variable's value. The sum of these points is located on the total points axis, and then a perpendicular line is then drawn downwards to the survival axis to determine the 1-year, 3-year, and 4-year OS probability. **(B)** The calibration curve for the prediction of 4-year OS based on the internal validation cohort. **(C)** The calibration curve for the prediction of 4-year OS based on the external validation two cohort. In **(B)** and **(C)**, the OS prediction of nomogram probability is plotted on the *x*-axis, and the real OS is plotted on the *y*-axis.

TABLE 3 | Point assignment and prognostic score.

Variable name	Score	Estimated 4-year OS (%)
Age, years	—	—
<60	0	—
≥60	32	—
TNM stage		—
I	0	—
II	54	—
III	93	—
TM9SF1 expression	—	—
High	0	—
Low	51	—
PDZK1IP1 expression		—
High	100	—
Low	0	—
Total prognostic score		
0–126	—	70
127–189	—	44
190–229	—	24
≥230	—	12

The nomogram was externally verified by the calibration plot in **Figure 5C** as well as by calculating the bootstrap C statistics of 105 patients in the external validation two cohort. In the external verification stage, the C-index of the 4-year OS nomogram was predicted to be 0.72 (**Figure 5C**), indicating that the model has a higher discrimination ability. The calibration curve shows that the calibration effect of the nomogram is good; the 4-year OS showed maximum consistency between the actual observation and the nomogram prediction.

Performance of the Nomogram in Stratifying Risk of Patients

We determined the truncation by dividing the patients in the training cohort into four subgroups based on their total scores (score: 0 to 126, 127 to 189, 190 to 229, and ≥230), where each subgroup corresponds to a different prognosis (**Table 3**). After applying the cutoff values to sort patients in each cohort, stratification into different risk subgroups resulted in significantly different Kaplan–Meier curves for survival outcomes in each group (**Figures 6A–C**). The survival curve for subgroups sorted according to TNM stage showed worse performance as shown by the survival rate of stage I being lower than that of stage II (**Figure 6F**). The grouping result by the nomogram score was observably better than that by TNM stage ($p < 0.0001$ vs. 0.00019, 0.0093 vs. 0.01, <0.0001 vs. 0.025).

At the same time, we performed sub-group analysis of adjuvant therapy and TNM stage in the primary cohort, stratification into different risk subgroups allowed significant distinction between Kaplan-Meier curves for survival outcomes within each TNM stage, age and adjuvant therapy (**Figure 7**).

Validating the Marker Genes

In order to verify the stability of the marker genes, we performed validation analysis in the primary cohort and two external validation cohorts. In the primary cohort, we divided the 179 samples into a high TM9SF1 expression group with 56 samples and a low TM9SF1 expression group with 123 samples ($p < 0.001$; **Figure 6G**). Similar analyses showed that 18 samples with high TM9SF1 expression had poorer prognosis than 63 samples with low TM9SF1 expression in the external validation one cohort ($p = 0.0021$; **Figure 6H**), and similar results were found in the external validation two cohort ($p = 0.0077$; **Figure 6I**).

For another marker gene, in the primary cohort, we successfully divided the 179 samples into a high PDZK1IP1 expression group with 28 samples and a low PDZK1IP1 expression group with 151 samples ($p = 0.0014$; **Figure 6J**). Similar analyses showed that 71 high PDZK1IP1 expression patients had better prognosis than 10 low PDZK1IP1 expression patients in the external validation one cohort ($p = 0.071$; **Figure 6K**), and 52 high PDZK1IP1 expression samples had better prognosis than 53 low PDZK1IP1 expression samples in the external validation two cohort ($p = 0.0085$; **Figure 6L**). Together, these results indicate that PDZK1IP1 and TM9SF1 can be defined as marker genes directly related to prognosis.

DISCUSSION

Due to the significant heterogeneity of ESCC in individual patient survival, the prediction of survival using the TNM staging system is inaccurate. (Sun et al., 2019) It is necessary to develop a new ESCC prognosis grading system. Therefore, we aimed to develop a nomogram model that uses gene expression values to predict long-term survival in patients with operable ESCC.

In this study, we determined that age and TNM stage were independent prognostic factors through univariate and multivariate Cox analyses of clinical characteristics. Age and TNM staging were highly consistent with previous studies on ESCC risk factors. Meanwhile we used edgeR package to screen differential genes and selected nine significant genes by using the rbsurv package in the discovery phase. Then we narrowed the selection down to two marker genes in the training phase. Finally, we verified the marker genes using an internal validation cohort and two external validation cohorts, which included samples from the GEO database and the TCGA database. According to the cutoff value of marker genes, we can divide ESCC cases into two subgroups with significantly different high or low risk of death. ESCC patients with high expression of PDZK1IP1 had worse prognosis than those with low expression, suggesting that PDZK1IP1 was a negative factor for EC mortality (**Figures 6G–I**). In contrast, TM9SF1 expression levels were correlated with better prognosis, suggesting that TM9SF1 was a positive factor for EC mortality (**Figures 6J–L**). There have been reports that the TGF-β1,HER-2 and Smad4 are associated with the development of ESCC, and the low quality of HER-2 as a prognostic biomarker in ESCC. (Heidarpour et al., 2020) HER-2 expressed in a variety of tumor tissues including primary breast tumors and tumors from small bowel, esophagus, kidney and mouth. The effect of PDZK1IP1 and TM9SF1 on ESCC is not clear. PDZK1IP1 expressed at significant levels only in a single epithelial cell population, the proximal tubular epithelial cells of the kidney as well as diffusely expressed in various carcinomas originating from kidney, colon, lung and breast. There are reports that PDZK1IP1 interacts with Smad4 and thereby suppresses the

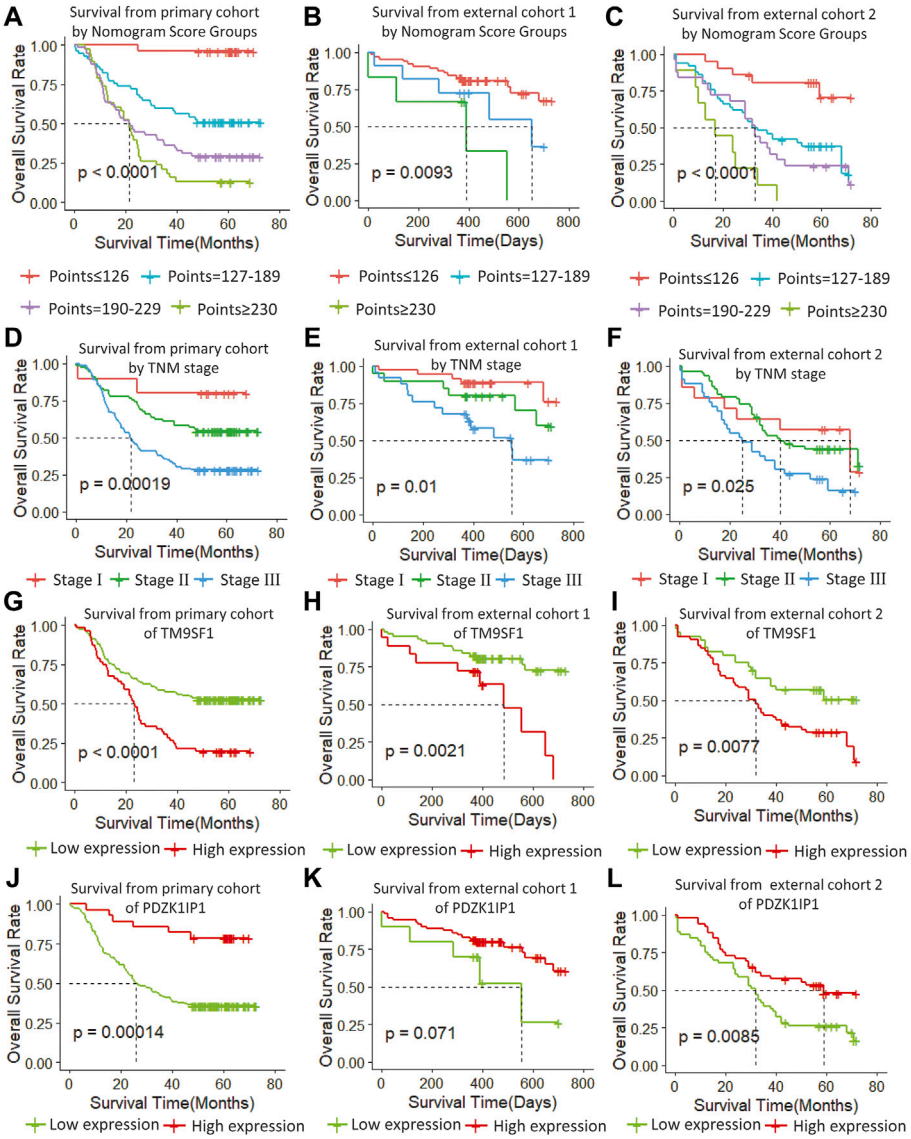


FIGURE 6 | (A–C) Survival from primary cohort by nomogram score groups, **(D–F)** survival from each cohort by TNM stage, **(G–I)** survival from each cohort of TM9SF1, and **(J–L)** survival from each cohort of PDZK1IP1.

TGF-β signaling pathway. (Ikeno et al., 2019) TM9SF1 Plays an essential role in autophagy. There are reports that TM9SF1 as a collaborative EBAG9 interactor, which regulates epithelial-mesenchymal transition (EMT) in cancer cells (Miyazaki et al., 2018).

Therefore, PDZK1IP1 and TM9SF1 are defined as marker genes directly related to prognosis. These two marker genes may have clinical significance for customized follow-up and treatment of ESCC patients. With these two marker genes, low-risk patients can avoid the toxic side effects of adjuvant therapy whereas high-risk ESCC patients can receive more rigorous monitoring and treatment regimens to prevent their condition from worsening (Cao et al., 2016).

The nomogram aimed to estimate 1-year, 3-year, and 4-year OS probabilities based on the multivariate Cox proportional risk model, which includes TNM staging, age, and two mRNA expression values for postoperative measurements of cancer tissue. In the validation phase, we demonstrated that the nomogram was an excellent model for predicting 1-year, 3-year, and 4-year OS for ESCC patients, and we demonstrated that the accuracy was better than TNM staging. The predicative accuracy of our nomogram model in primary cohort is the best one among PDZK1ID1, TM9SF1, Age, TNM as single predictors with C-index of 0.75, 0.701, 0.659 and 0.65 (**Table 4**). All of the aforementioned predictors have greater C-index compared with age with a C-index of 0.532. In external cohort one and cohort two, similar results were obtained. In addition, the two clinicopathologic characteristic and the mRNA expression values of two genes incorporated into nomogram should be recorded by every clinician for ESCC patients to increase their clinical effectiveness.

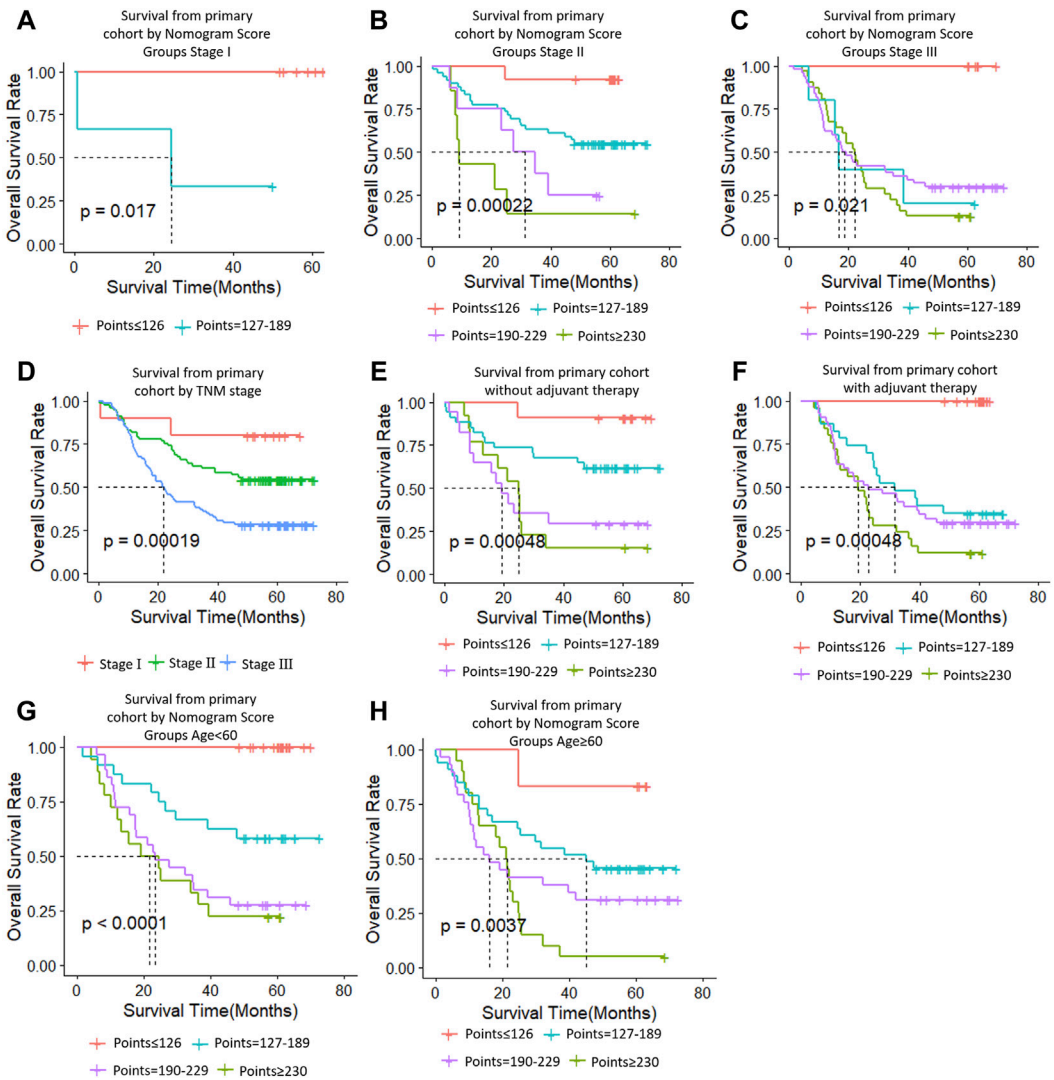


FIGURE 7 | Risk group stratification within each TNM stage (stages, **(A–C)**; **(D)**, all patients), group according to whether receive adjuvant therapy **(E,F)** and group by age **(G,H)**.

In this study, the proposed nomogram was verified to avoid overfitting of the model and to determine the generalization of the model. (Iasonos et al., 2008) The predicted value of the calibration diagram was highly consistent with the actual value, which proved the accuracy and repeatability of the nomogram. Furthermore, the model was validated using the TCGA database (external validation one cohort), which included patients from Asia, North America, South America, and Europe, making it possible to use and promote the model globally, regardless of race, regional lifestyle, and economic factors. In the verification phase, the C-index of the model was obviously better than the TNM staging system. Using the model, the OS rate prediction ability was slightly worse in the external validation one cohort (**Figure 6B**) compared to the primary cohort (**Figure 6A**) and the external validation two cohort (**Figure 6C**). Subsequently, applying threshold values to divide each cohort into four different risk subgroups resulted in significant differences in Kaplan–Meier curves for survival results in every group. The discrimination ability of the primary cohort (C-index, 0.75 for nomogram vs. 0.68 for TNM staging system; 0.07 difference) and external validation two cohort (0.72 for nomogram vs. 0.64 for TNM staging system; 0.08 difference) were similar. When using the TNM staging system, the survival curve of patients with stage II did not reach the significance level (**Figures 6D–F**). Additionally, there were some intersections among the OS rate lines of different nomogram score groups (**Figure 6C**) and TNM stage groups (**Figure 6F**). We think that sample size is the most important reason for these indistinctions between different nomogram score groups and TNM stage groups.

In recent years, researchers have used nomograms to predict ESCC, and this study has several advantages over previous

TABLE 4 | Prognostic ability and accuracy of ESCC.

Factor	Primary cohort		External cohort 1		External cohort 2	
	p-value	C-index	*p*-value	C-index	*p*-value	C-index
Nomogram	<0.0001	0.75	0.0093	0.695	<0.0001	0.72
PDZK1ID1	0.00014	0.701	0.071	0.661	0.0086	0.639
TM9SF1	<0.0001	0.659	0.0021	0.52	0.0077	0.656
Age	0.015	0.532	0.4	0.376	0.061	0.596
TNM	0.00019	0.65	0.01	0.6	0.025	0.630

To further evaluate the predictive capacity and accuracy Nomogram. The higher C-index score represent the better prognostic performance of the system.

studies. First, to avoid specificity, we confirmed the gene markers in samples from multiple medical centers. Second, the majority of previous studies did not consider mRNA in ESCC. Marker genes play an important role in the development of ESCC, and their differential expressions are associated with the development of ESCC. Finally, we adopted a combinatorial strategy in our study, which is different from previous studies that used only one algorithm to select markers. The purpose of the combinatorial algorithm is to reduce the possibility of missing or ignoring important marker genes.

The limits of this study includes the following points. First, we used limited sample sizes in the training and test cohorts that results in some discrepancy. Second, all ESCC patients in primary cohort are Chinese origin contrasting with ESCC patients from TCGA in external cohort one across world. Therefore, the predictative power of our nomogram model in ESCC patients from TCGA is less efficient compared with primary and external cohort two. We speculate that the different predicative capability is ascribed to regional difference of ESCC cases. Third, our nomogram predicts overall survival well for subgroup patients with or without adjuvant therapy, indicating that it is not appropriate for decision-making on adjuvant therapy (**Supplementary Figure S1**). The underlying causes are not clear at present and warrants further study to investigate.

In summary, the nomogram presented in this study can accurately and impersonally predict the prognosis of ESCC patients after partial resection of the esophagus. More research is required to determine whether it can be applied to other patient populations. Moreover, we found two marker genes directly related to the prognosis of ESCC, which will provide a basis for future research.

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, The First Affiliated Hospital, Henan University of Science and Technology. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

S-GG and Y-JQ conceived and designed this study. KL and Y-JQ wrote and revised the manuscript, and Z-JL, B-LG, Y-LJ, XY and Q-WL prepared the figures. YZ, PC, M-XL, L-QS, H-JR and F-BX performed the statistical analyses. All authors read and gave final approval of the manuscript.

FUNDING

This study was supported by National Natural Science Foundation Nos. 81972571 and 81872037. This study was supported by Henan provincial medical science and technology project No. LHGJ20190574.

ACKNOWLEDGMENTS

Thanks to Letpub for providing language editing services. (www.letpub.com).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2022.823619/full#supplementary-material

REFERENCES

Cao, J., Yuan, P., Wang, L., Wang, Y., Ma, H., Yuan, X., et al. (2016). Clinical Nomogram for Predicting Survival of Esophageal Cancer Patients after Esophagectomy. *Sci. Rep.* 6, 26684. doi:10.1038/srep26684

Duan, J., Deng, T., Ying, G., Huang, D., Zhang, H., Zhou, L., et al. (2016). Prognostic Nomogram for Previously Untreated Patients with Esophageal Squamous Cell Carcinoma after Esophagectomy Followed by Adjuvant Chemotherapy. *Jpn. J. Clin. Oncol.* 46 (4), 336–343. doi:10.1093/jjco/hyv206

Enzinger, P. C., and Mayer, R. J. (2003). Esophageal Cancer. *N. Engl. J. Med.* 349, 2241–2252. doi:10.1056/nejmra035010

Frank, E., and Harrell, Jr (2014). *Rms: Regression Modeling Strategies*. R Package version 3.4-0. Nashville, USA: Springer. Available at: http://www.rproject.org (accessed APRIL 17, 2014).

Gao, S.-G., Yang, J.-Q., Ma, Z.-K., Yuan, X., Zhao, C., Wang, G.-C., et al. (2018). Preoperative Serum Immunoglobulin G and A Antibodies to Porphyromonas Gingivalis Are Potential Serum Biomarkers for the Diagnosis and Prognosis of Esophageal Squamous Cell Carcinoma. *BMC Cancer* 18 (1), 17. doi:10.1186/s12885-017-3905-1

Heidarpour, M., Taheri, M., Akhavan, A., Goli, P., and Kefayat, A. (2020). Investigation of HER-2 Expression an its Correlation with Clinicopathological Parameters and Overall Survival of Esophageal Squamous Cell Carcinoma Patients. *Iran J. Pathol.* 15 (4), 274–281. doi:10.30699/IJP.2020.113829.2235

Iasonos, A., Iasonos, D., Raj, G. V., and Panageas, K. S. (2008). How to Build and Interpret a Nomogram for Cancer Prognosis. *J. Clin. Oncol.* 26, 1364–1370. doi:10.1200/JCO.2007.12.9791

Ikeno, S., Nakano, N., Sano, K., Minowa, T., Sato, W., Akatsu, R., et al. (2019). PDZK1-interacting Protein 1 (PDZK1IP1) Traps Smad4 Protein and Suppresses Transforming Growth Factor-β (TGF-β) Signaling. *J. Biol. Chem.* 294 (13), 4966–4980. doi:10.1074/jbc.ra118.004153

International Agency for Research on Cancer (2012). *GLOBOCAN 2012, Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012*. Geneva: World Health Organization.

Li, J., Chen, Z., Tian, L., Zhou, C., He, M. Y., Gao, Y., et al. (2014). LncRNA Profile Study Reveals a Three-lncRNA Signature Associated with the Survival of Patients with Oesophageal Squamous Cell Carcinoma. *Gut* 63 (11), 1700–1710. doi:10.1136/gutjnl-2013-305806

Liang, H., Fan, J. H., and Qiao, Y. L. (2017). Epidemiology, Etiology, and Prevention of Esophageal Squamous Cell Carcinoma in China. *Cancer Biol. Med.* 14, 33–41. doi:10.20892/j.issn.2095-3941.2016.0093

Mariani, L., Miceli, R., Kattan, M. W., Brennan, M. F., Colecchia, M., Fiore, M., et al. (2005). Validation and Adaptation of a Nomogram for Predicting the Survival of Patients with Extremity Soft Tissue Sarcoma Using a Three-Grade System. *Cancer* 103, 402–408. doi:10.1002/cncr.20778

Miyazaki, T., Ikeda, K., Sato, W., Horie-Inoue, K., and Inoue, S. (20182). Extracellular Vesicle-Mediated EBAG9 Transfer from Cancer Cells to Tumor Microenvironment Promotes Immune Escape and Tumor Progression. *Oncogenesis* 7 (1), 7. doi:10.1038/s41389-017-0022-6

N. Ando (Editor) (2015). *Esophageal Squamous Cell Carcinoma* (Tokyo, Japan: Springer).

Owen, R. P., White, M. J., Severson, D. T., Braden, B., Bailey, A., Goldin, R., et al. (2018). Single Cell RNA-Seq Reveals Profound Transcriptional Similarity between Barrett's Oesophagus and Oesophageal Submucosal Glands. *Nat. Commun.* 9 (1), 06796. doi:10.1038/s41467-018-06796-9

Qi, Y.-J., Chao, W.-X., and Chiu, J.-F. (2012). An Overview of Esophageal Squamous Cell Carcinoma Proteomics. *J. Proteomics* 75 (11), 3129–3137. doi:10.1016/j.jprot.2012.04.025

Qiu, J., Peng, B., Tang, Y., Qian, Y., Guo, P., Li, M., et al. (2017). CpG Methylation Signature Predicts Recurrence in Early-Stage Hepatocellular Carcinoma: Results from a Multicenter Study. *Jco* 35 (7), 734–742. doi:10.1200/jco.2016.68.2153

Rubenstein, J. H., and Shaheen, N. J. (2015). Epidemiology, Diagnosis, and Management of Esophageal Adenocarcinoma. *Gastroenterology* 149, 302–317. doi:10.1053/j.gastro.2015.04.053

Sternberg, C. N. (2006). Are Nomograms Better Than Currently Available Stage Groupings for Bladder Cancer. *Jco* 24, 3819–3820. doi:10.1200/jco.2006.07.1290

Sun, Y., Wang, J., Li, Y., Pan, S., Yang, T., Sun, X., et al. (2019). Nomograms to Predict Survival Rates for Esophageal Cancer Patients with Malignant Behaviors Based on ICD-0-3. *Future Oncol.* 15, 121–132. doi:10.2217/fon-2018-0493

Wang, L., Hricak, H., Kattan, M. W., Chen, H.-N., Scardino, P. T., and Kuroiwa, K. (2006). Prediction of Organ-Confined Prostate Cancer: Incremental Value of MR Imaging and MR Spectroscopic Imaging to Staging Nomograms. *Radiology* 238, 597–603. doi:10.1148/radiol.2382041905

Wierda, W. G., O'Brien, S., Wang, X., Faderl, S., Ferrajoli, A., Do, K.-A., et al. (2007). Prognostic Nomogram and index for Overall Survival in Previously Untreated Patients with Chronic Lymphocytic Leukemia. *Blood* 109, 4679–4685. doi:10.1182/blood-2005-12-051458

Wu, H., Yu, J., Li, Y., Hou, Q., Zhou, R., Zhang, N., et al. (2018). Single-cell RNA Sequencing Reveals Diverse Intratumoral Heterogeneities and Gene Signatures of Two Types of Esophageal Cancers. *Cancer Lett.* 438, 133–143. doi:10.1016/j.canlet.2018.09.017

Zhang, W.-Y., Chen, X.-X., Chen, W.-H., Zhang, H., and Zou, C.-L. (2018). Nomograms for Predicting Risk of Locoregional Recurrence and Distant Metastases for Esophageal Cancer Patients after Radical Esophagectomy. *BMC Cancer* 18 (1), 879. doi:10.1186/s12885-018-4796-5

Zhao, J., He, Y.-T., Zheng, R.-S., Zhang, S.-W., and Chen, W.-Q. (2012). Analysis of Esophageal Cancer Time Trends in China, 1989-2008. *Asian Pac. J. Cancer Prev.* 13, 4613–4617. doi:10.7314/apjcp.2012.13.9.4613

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

frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 22 March 2022
doi: 10.3389/fbioe.2022.856745

Identification and Validation of miRNA-TF-mRNA Regulatory Networks in Uterine Fibroids

Xiaotong Peng[1], *Yanqun Mo*[1], *Junliang Liu*[1], *Huining Liu*[1]* *and Shuo Wang*[2]*

[1]*Department of Gynaecology and Obstetrics, Xiangya Hospital, Central South University, Changsha, China,* [2]*Department of Orthopaedics, Shanghai Jiaotong University Affiliated Sixth People's Hospital, Shanghai, China*

Uterine fibroids (UF) are the most common benign gynecologic tumors and lead to heavy menstrual bleeding, severe anemia, abdominal pain, and infertility, which seriously harm a women's health. Unfortunately, the regulatory mechanisms of UF have not been elucidated. Recent studies have demonstrated that miRNAs play a vital role in the development of uterine fibroids. As a high-throughput technology, microarray is utilized to identify differentially expressed genes (DEGs) and miRNAs (DEMs) between UF and myometrium. We identified 373 candidate DEGs and the top 100 DEMs. Function enrichment analysis showed that candidate DEGs were mainly enriched in biological adhesion, locomotion and cell migration, and collagen-containing extracellular matrix. Subsequently, protein-protein interaction (PPI) networks are constructed to analyze the functional interaction between DEGs and screen hub DEGs. Subsequently, the expression levels of hub DEGs were validated by real-time PCR of clinical UF samples. The DGIdb database was used to select candidate drugs for hub DEGs. Molecular docking was applied to test the affinity between proteins and drugs. Furthermore, target genes for 100 candidate DEMs were predicted by miRwalk3.0. After overlapping with 373 candidate DEGs, 28 differentially expressed target genes (DEGTs) were obtained. A miRNA-mRNA network was constructed to investigate the interactions between miRNA and mRNA. Additionally, two miRNAs (hsa-miR-381-3p and hsa-miR-181b-5p) were identified as hub DEMs and validated through RT-PCR. In order to better elucidate the pathogenesis of UF and the synergistic effect between miRNA and transcription factor (TF), we constructed a miRNA-TF-mRNA regulatory network. Meanwhile, *in vitro* results suggested that dysregulated hub DEMs were associated with the proliferation, migration, and apoptosis of UF cells. Our findings provided a novel horizon to reveal the internal mechanism and novel targets for the diagnosis and treatment of UF.

Keywords: uterine fibroids, miRNA-TF-mRNA, regulatory network, bioinformatics, biomarkers

OPEN ACCESS

Edited by:
Francesco Pappalardo, University of Catania, Italy

Reviewed by:
Maxim Yuryevich Shkurnikov, National Research University Higher School of Economics, Russia
Vesselin Baev, Plovdiv University "Paisii Hilendarski", Bulgaria

***Correspondence:**
Huining Liu
liuhuining@163.com
Shuo Wang
shuowang97@126.com

Specialty section:
This article was submitted to Preclinical Cell and Gene Therapy, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *17 January 2022*
Accepted: *02 March 2022*
Published: *22 March 2022*

Citation:
Peng X, Mo Y, Liu J, Liu H and Wang S (2022) Identification and Validation of miRNA-TF-mRNA Regulatory Networks in Uterine Fibroids. Front. Bioeng. Biotechnol. 10:856745. doi: 10.3389/fbioe.2022.856745

INTRODUCTION

Uterine fibroids (UF) are recognized as the most common benign tumors of the female internal genitalia (Pitter et al., 2013). The cumulative incidence of UF in women before the age of 50 is as high as 70% (Walker and Stewart, 2005). Despite their benign nature, UF still causes a variety of symptoms, such as pelvic pressure, uterine bleeding, anemia, or infertility (Islam et al., 2013), and is mainly treated through surgery and drugs (Zhang L. et al., 2021), such as hysterectomy, leading to

devastating effects on approximately two 200,000 per year in United States (Cardozo et al., 2012). Unfortunately, their origins and internal mechanism have not been characterized so far (Andersen and Barbieri, 1995). Therefore, it is urgent to reveal the pathogenesis of UF and identify novel biomarkers for early screening, diagnosis, and individualized precision therapy.

At the post-transcriptional level, microRNAs (miRNAs) are single-stranded RNA molecules (approximately 18–25 nucleotides long) and negatively regulate gene expression at the post-transcriptional level through binding to complementary sites in the 3′-untranslated regions (3′UTR) of target genes, resulting in messenger RNAs (mRNAs) translational arrest or degradation (Rupaimoole and Slack, 2017). The above processes are similar to the regulation mechanism of transcription factors (TFs) through binding to a TF binding site in the promoter region, which suggests that they share a mutual binding mechanism. Recently, there has been increasing evidence that miRNA can regulate more than 30% of the encoding genes in mammals and play a vital role in the pathogenesis and development of various human diseases, including UF (Ali et al., 2020). For example, downregulated miR-139-5p contributed to the progression of UF through collagen I deposition and p38 phosphorylation (Ahn et al., 2021). The caspase-3 cleavage and elongation factor-2 phosphorylation were activated by miR-21, leading to UF cell apoptosis and stalled translation (Fitzgerald et al., 2012). Furthermore, transfected leiomyoma cells with miR-150–5p mimics had stronger migrational ability and secreted more collagen through $p27^{Kip1}$ activation and Akt inhibition (Lee et al., 2019). Hence, miRNA–mRNA regulatory networks are strongly associated with tumorigenesis and progression of UF. However, the specific microRNAs in UF and their potential role on target genes or transcription factors (TFs) regulatory networks have remained unrevealed.

Benefiting from bioinformatics and high-throughput sequencing technology's rapid development, network-based methods emerge and expand our knowledge of regulatory relationships among genes (Fu and Dong, 2018). Regulatory networks have been widely applied to screen pathogenic genes, elucidate miRNAs' functions, and provide targeted drugs for disease treatment for subsequent experimental verification (Hu et al., 2017). As shown in **Scheme 1**, differentially expressed genes (DEGs) and miRNAs (DEMs) in UF tissues versus normal myometrium tissues were identified through analyzing two mRNA microarrays (GSE593 and GSE18096) and one miRNA microarray (GSE159959) downloaded from gene expression omnibus (GEO) databases. Subsequently, we performed functional annotation, enrichment analysis, and established protein-protein interactions (PPI) and drug-hub DEGs networks. Real-time PCR (RT-PCR) and molecular docking were used to check the DEG expression based on human tissue samples for diagnosis and screen effective target-drugs respectively. Furthermore, 28 differentially expressed target genes (DETGs) out of the top 100 DEMs were obtained from the intersection between predicted target genes and DEGs. MiRNA-target gene regulatory networks associated with UF were established and provided the interactions that have not yet been reported. Hub DEMs' functional annotation and pathway enrichment analysis were also implemented. We identified the interaction between TFs and hub DEMs to reveal a common regulatory mechanism. Combining with the above bioinformatics results, we designed mimics and inhibitors of hub DEMs and evaluated their functions on the proliferation, migration, and apoptosis of UF cells. Taken together, our study looked forward to providing a novel horizon for UF treatment through unscrambling the miRNA-TF-mRNA regulatory network, resulting in the avoidance of irretrievable harm to female reproductive health.

MATERIALS AND METHODS

Data Extraction

UF-related expression profiles for miRNA (GSE159959) and mRNA (GSE593) and GSE18096) were collected from the gene expression omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). The associated information from the above three datasets was shown in **Figure 1A**. Due to the different process platforms of the included datasets, we normalized these data using R's limma package.

Analysis of DEMs and DEGs

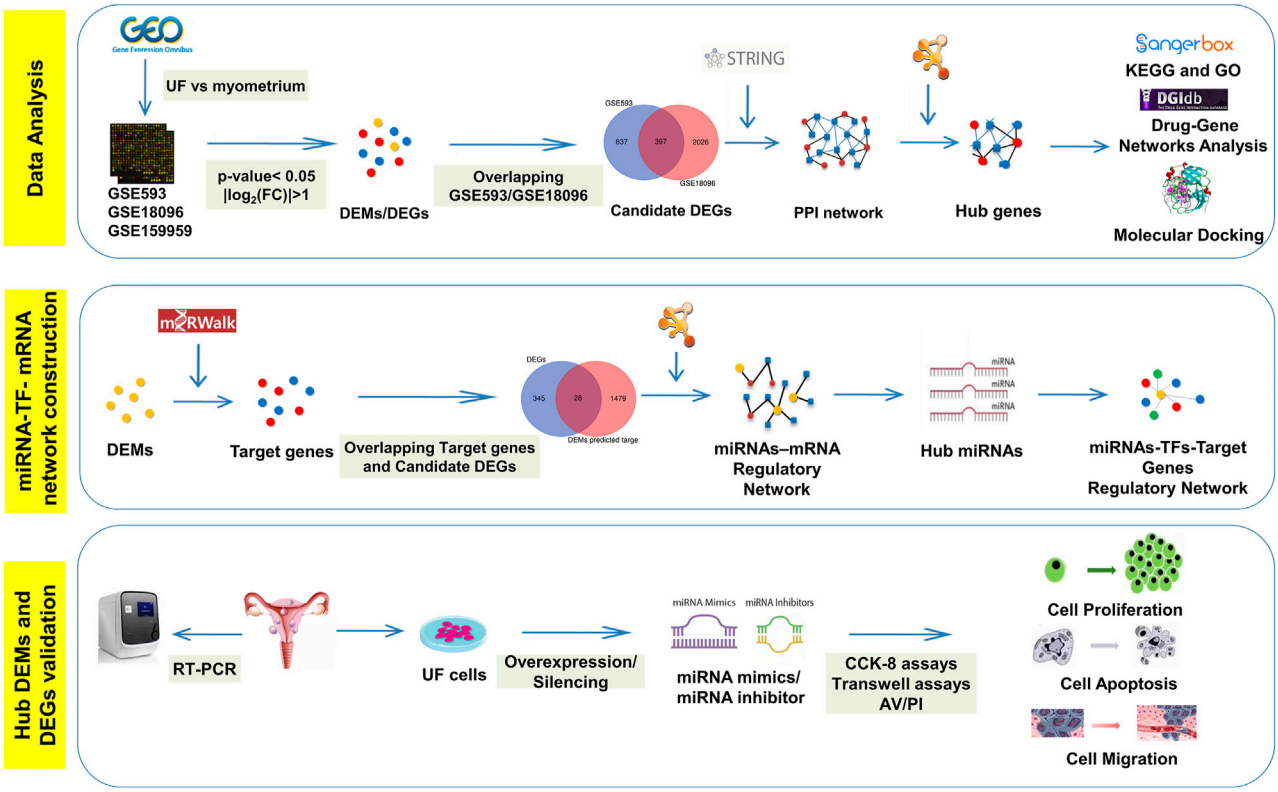
The normalized expression datasets were analyzed using the GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r/) for comparison between UF tissues and matched myometrium tissues. The screening criteria were set as p-value <0.05 and |log2-fold change (FC)| > 1, the top 100 regulated miRNAs were selected as DEMs. With the same threshold, the upregulated and downregulated overlapping DEGs were obtained between GSE593 and GSE18096. The DEMs and DEGs in the respective profiles were visualized in volcano plots. Furthermore, heatmaps of DEMs and DEGs were visualized by TB tools, which is a toolkit that integrates various biological data-handling tools.

Construction of the PPI Network and Analysis of Modules

The online STRING database (https://string-db.org) was used to visualize the positive functional interaction between DEGs and the threshold (combined scores ≥0.4). Cytoscape software vision 3.8.2 was utilized to screen the upregulated and downregulated DEGs and construct the PPI network. Additionally, molecular complex detection (MCODE) v1.5.1.16 (Bader and Hogue, 2003) was used to identify hub genes with the following cut-off criteria: degree cut-off = 2, max. depth = 100, node score cut-off = 0.2, and k-Core = 2.

Functional Enrichment Analysis

To further clarify the functional annotation of DEGs and hub genes, Gene Ontology (GO) functional annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were conducted in the Sangerbox database (http://sangerbox.com/Tool). GO functional analysis was mainly



SCHEME 1 | The schematic workflow of current study.

classified into cellular components (CC), biological processes (BP), and molecular functions (MF).

Drug-Hub Gene Network Analysis

In order to explore the target drugs of hub genes, the drug-gene interaction database (DGIdb, https://dgidb.genome.wustl.edu/) was used, which is the online tool for drug-gene predication. Finally, Cytoscape was used to construct a drug-gene network.

Homology Modeling and Molecular Docking

Recently, artificial intelligence brought revolutionary changes to the field of protein structural biology through convolutional neural networks. As an AI prediction tool, Alphafold (https://alphafold.ebi.ac.uk/), as an AI prediction tool, predicts highly accurate protein structures competitive with experiments based on their amino acid sequence. Therefore, despite the crystal structures of human BIRC5 and TYMS having yet to be solved in the RCSB protein data bank (PDB), we adopted Alphafold to build the structure of STAT6 through in silico modelling. From the amino acid sequence of human BIRC5 and TYMS (accession numbers: O15392 and P04818, respectively) from the UniProtKB database (http://www.uniprot.org/) and predicted protein structures starting from the sequence through AlphaFold v2.0. The stereochemical quality of the predicted model was evaluated by the local distance difference test (LDDT) score in the AlphaFold database.

Molecular docking simulation is a technique to predict the first-rank orientation of small-molecules to macromolecule targets to form stabilized complexes. Associated small-molecules were both downloaded from the Pubchem database (https://pubchem.ncbi.nlm.nih.gov). The AutoDockTools and PyMol were applied to preprocess the input file, including hydrogenation and deletion of crystallographic water and ligands. Molecular docking between small-molecules and BIRC5, TGFBR2, and TYMS binding pockets was performed by Autodock Vina with default parameters. The predicted binding interaction geometries of irinotecan (aka irinotecan hydrochloride) and BIRC5 or TYMS were visualized and the docking affinity between small-molecules and protein targets was scored. The optimal docking conformation and related results were analyzed by PyMol.

Prediction of Target Genes for DEMs and Construction of miRNA-Target Gene Regulatory Network

MiRWalk3.0 (http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk3/) is an online tool to predict the downstream target genes of DEMs. Then, differentially expressed target genes (DETGs) were obtained by overlapping the predicted genes and DEGs for further analysis. In addition, the miRNA-target gene regulatory networks were visualized by Cytoscape. Among

A

Accession	Platform	Sample	Normal	Uterine Fibroids	Gene/microRNA
GSE593	GPL96	Tissues	5	5	Gene
GSE18096	GPL570	Tissues	9	18	Gene
GSE159959	GPL29277	Tissues	6	6	microRNA

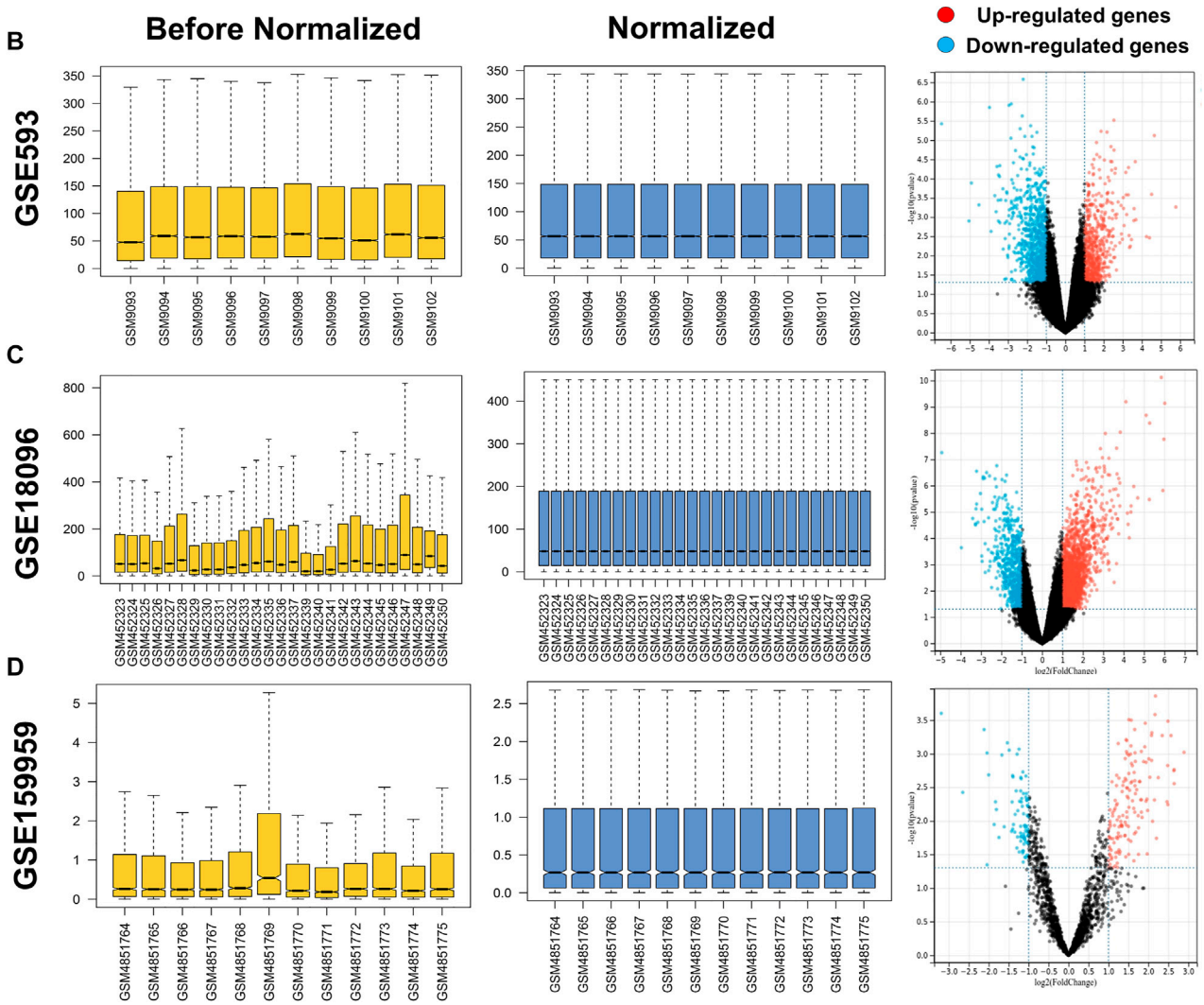


FIGURE 1 | (A) Characteristics of mRNA and miRNA expression profiles of UF. The before and after normalization and volcano plots of **(B)** GSE593, **(C)** GSE159959 and **(D)** GSE18096.

the DEMs in this network, we selected the DEMs with the highest number of regulatory genes ranking top two as the hub DEMs.

Functional and Pathway Enrichment Analysis of Hub DEMs

To further analyze the function of hub DEMs, GO and KEGG pathways were performed using miRPath v3.0 (http://snf-515788.vm.okeanos.grnet.gr/). An adjusted $p < 0.05$ was considered significantly enriched.

Establishment of the miRNA-TFs-Target Gene Networks

It has been confirmed that TF-mediated transcriptional regulation and miRNA-mediated post-transcriptional regulation are tightly coupled, which implies that they share a common regulatory mechanism (Lewis et al., 2005). Transcription factors and miRNAs are known to regulate each other as well as their target genes. Therefore, further research on the regulatory relationship between TFs and miRNAs can help

elucidate the underlying mechanism of tumorigenesis and progress. Based on the hub DEMs, we predicted the potential TFs *via* the iRegulon plug-in (Lewis et al., 2005). Relevant parameters were as follows: enrichment score threshold = 5.0, ROC threshold for AUC calculation = 0.05, rank threshold = 5,000, minimum identity between orthologous genes = 0.05, and maximum FDR on motif similarity = 0.001. The Cytoscape software was subsequently utilized to visualize miRNA-TF-target gene regulatory networks.

Tissue Samples

Ten cases of UF tissue and matched normal myometrial tissue samples were collected from patients who underwent myomectomy in Xiangya Hospital from January 2021 to January 2022, aged 36–49 years old. All patients had regular menstruation without other complications, and they did not take hormone drugs within 3 months before surgery. All samples were taken with the consent of the patients, and informed consent was signed. Consent was granted by all women included in the study, and informed consent was signed. Tissues were stored at −80°C immediately after harvest until further use. The pathological diagnosis of the tissue was confirmed by at least two independent pathologists.

Hematoxylin-Eosin Staining

The tissues were fixed for 24 h and dehydrated with gradient alcohol. Paraffin-embedded tissues were prepared and subsequently cut into 5 μm slices, which were then stained using the H&E staining kit (Beijing Solarbio Science and Technology Co., Ltd., Beijing, China). The pathological changes in UF were observed and captured with five random images under a light microscope (Leica DMi6-M, Leica Microsystems Co., Ltd.).

Isolation and Culture of Human Leiomyoma Cells and Myometrial Smooth Muscle Cells.

The human UF tissue was washed three times with PBS solution and minced into small pieces (3–6 mm^3) with sterilized eye scissors. To obtain single-cell suspensions, the tissue fragments were incubated with Dulbecco's modified eagle media (DMEM, Hyclone, United States) medium supplemented with 4 mg/ml type I collagenase (Sigma-Aldrich, United States) and 1 mg/ml DNase I (Sigma-Aldrich, United States) at 37°C in a humidified incubator with 5% CO2 for 2 h. Subsequently, the suspensions were passed through a 40-μm cell strainer (Corning, United States) to eliminate undigested tissue. The primary cells were cultivated in DMEM medium containing 10% fetal bovine serum (FBS, Gibco, United States), 100 U/ml penicillin, and 100 μg/ml streptomycin at 37°C in a 5% CO2 incubator. Cultured human primary cells were passaged every 4–6 days and analyzed at the third passage for further study.

Cell Transfection

To explore the function of DEMs in leiomyoma cells, the primary cells were seeded in 6-well plates at an appropriate density. When cells were in log-phase growth, cell transfection was performed with negative control (NC) or hub DEM mimics or inhibitors (50 nM) using Lipofectamine 3000 (Invitrogen, United States). The sequences were synthesized by GenePharma (Shanghai, China).

RNA Extraction and RT-PCR Validation

RNA was extracted from tissues using the TRIZOL method (TransGen Biotech, Beijing, China) and the RNA was reverse transcribed into cDNA according to the instructions of the NovoScript® II reverse transcriptase (Novoprotein, Shanghai, China). MiRNA was extracted using the miRNA extraction kit (Tiangen, Beijing, China) and reversely transcribed into cDNA through the polyA tailing method. RT-PCR was carried out by using SYBR® Green MasterMix (Takara, Japan) for hub DEGs and miRNA universal SYBR qPCR Master Mix (Vazyme Biotech, Nanjing, China) for hub DEMs in an ABI PRISM® 7500 Sequence Detection System (Applied Biosystems Inc: Foster City, CA). RT-PCR reaction system: For hub DEGs, 45 s at 95°C, 40 cycles for 30 s at 95°C, 45 s at 60°C, 60 s at 72°C and a final cycle for 10 min at 72°C; For hub DEMs, 5 min at 95°C, 40 cycles for 10 s at 95°C, 30 s at 60°C. Fluorescence values were collected. Replicate wells were set up for each sample. GAPDH and U6 were used as the internal reference for hub DEGs and DEMs detection, respectively, and the experiment was repeated three times independently. The mean CT value was calculated. Relative quantification (ΔΔCT method) was applied for semi-quantitative analysis, and the target gene expression was represented by $2^{-\Delta\Delta CT}$ values. The sequences of the primer were synthesized by Sangon Biotech (Shanghai, China).

CCK-8 Assay

Cells were transfected for 48 h and then inoculated into 96-well plates at a density of 4×10^3/well. The cell proliferation was measured at 1–3 days, respectively. At each time point, 10 μl of CCK-8 solution (New Cell and Molecular Biotech, Jiangsu, China) was added to each well and incubated in the incubator for 2 h. Finally, the absorbance at 450 nm, as the optical density (OD) of the samples, was detected by spectrophotometry at different time points.

Cell Migration Assay

After 24 h of transfection of mimic or inhibitor, we trypsinized and resuspended to obtain a single cell suspension. The upper layer of the Transwell chamber was added with FBS-free suspension at a cell density of 3×10^4/well. The lower layer was added with 20% FBS medium to drive cell migration. After 24 h of culture, we used cotton swabs to remove the nonmigrated cells in the upper chamber, and migrated cells in the lower chamber were fixed and stained with 0.1% crystal violet. Five randomly selected microscopic views were recorded through a light microscope and analysed by ImageJ software to count the number of cell migrations.

Apoptosis Detection

To determine the effect of hub DEMs on the apoptosis of leiomyoma cells, cells were harvested and centrifuged for annexin V-FITC/PI apoptosis detection (Boster Biotech Co.,

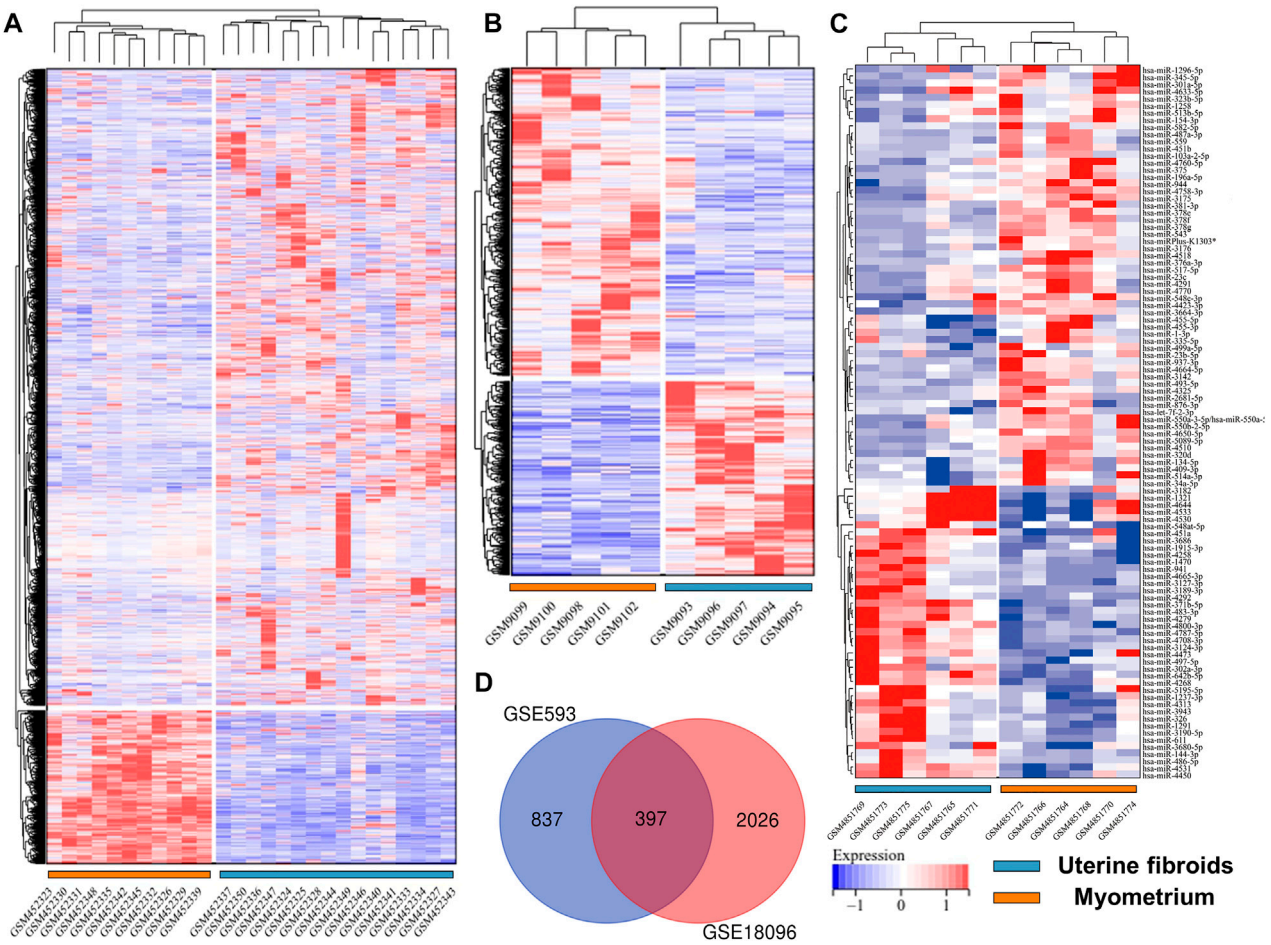


FIGURE 2 | Heatmap of **(A)** GSE18096. **(B)** GSE593, **(C)** GSE159959. **(D)** Venn diagram of screening DEGs form mRNA expression profile.

Ltd., Wuhan, China) after 24 h of transfection. After being washed twice, transfected cells were incubated with 5 μl of 488-annexin V and 5 μl of PI for 15 min at room temperature and protected from light. Subsequently, the flow cytometer (BD, United States) was applied to detect the apoptosis rates in each group. The apoptosis rate = sum of the early and late apoptotic rates.

Statistical Analysis

Statistical analysis was carried out using SPSS 16.0 statistical software. Data was expressed as mean ± SD, and the differences between the two groups were compared using the independent sample *t*-test. The differences between the two groups were compared using the independent sample t-test, and the differences were considered statistically significant at $p < 0.05$.

RESULTS

Identification of DEGs and DEMs in UF

Before screening DEGs and DEMs in UF, the three independent expression datasets processed on different platforms were normalized. Based on the mentioned threshold, differential expression analysis was applied to obtain DEGs and DEMs. 1185 and 2054 DEGs and 229 DEMs were identified in GSE593, GSE18096, and GSE159959, respectively. To shrink the scope of DEM screening, DEMs whose |log2 FC| ranked in the top 100 were chosen as candidate DEMs. As shown in **Figures 1B–D**, **2A–C**, volcano plots and heatmaps exhibited the overall distribution of DEGs and DEMs for UF versus normal myometrium. Furthermore, a total of 397 candidate DEGs were identified by Venn diagram (Chen and Boutros, 2011) analysis through the intersection of these two mRNA datasets (**Figure 2D**). Excluding DEGs with inconsistent expression trends in these two datasets, 373 genes containing 192 upregulated and 181 downregulated genes were obtained as candidate DEGs.

PPI Network Construction and Module Analysis

To identify potential interactions among 373 candidate DEGs, the PPI network was constructed using the STRING database and visualized by Cytoscape (**Figure 3A**). The network included

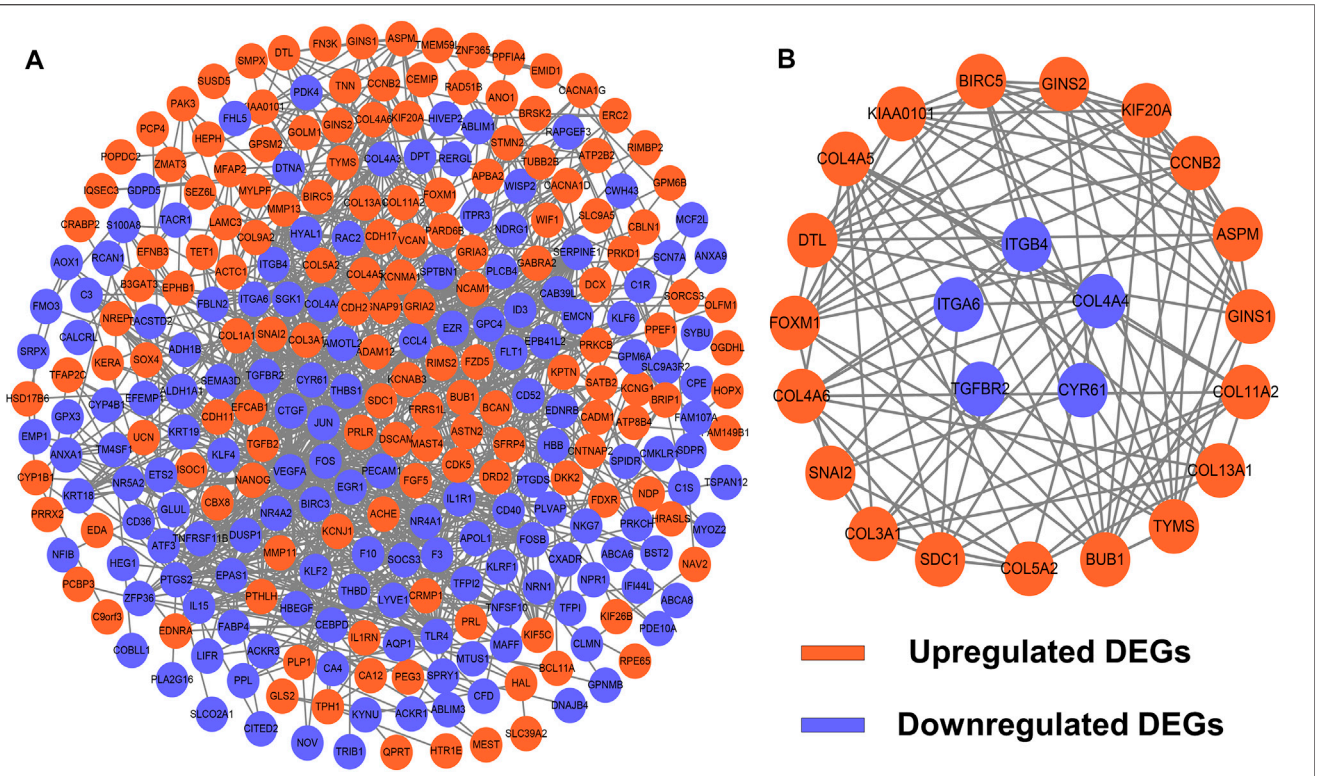


FIGURE 3 | PPI network and module analysis. **(A)** 373 DEGs were visualized in the PPI network. **(B)** Module analysis using MCODE.

297 nodes (genes) and 1,111 edges (interactions). Furthermore, hub DEGs was validated by MCODE, and then we obtained 8 cluster modules. As indicated in **Figure 3B**, the highest-rated module included 24 nodes (genes) and 99 edges (interactions). The 24 hub DEGs included 19 upregulated and 5 downregulated genes.

Enrichment Analysis of DEGs and Hub Genes

To further reveal the vital function and biological pathways of DEGs and hub DEGs. We performed GO and KEGG analysis of 373 candidate DEGs and 24 hub DEGs. As illustrated in **Figures 4A–D**, for the BP-correlated category, DEGs were mainly enriched in biological adhesion, locomotion and cell migration; for CC-correlated category, DEGs were primarily involved in collagen containing extracellular matrix (ECM), an intrinsic component of the plasma membrane and cell surface; and for the MF-correlated category, DEGs were principally concerned with extracellular matrix structural consitituent, growth factor binding and signaling receptor binding. Furthermore, KEGG pathway results demonstrated that DEGs primarily participated in ECM-receptor interaction, focal adhesion, and the PI3K-AKT pathway.

As for hub DEGs, BP-enriched terms included external encapsulating structure organization and collagen fibril organization; CC-enriched terms were involved in supramolecule complexes and collagen trimers; MF-enriched terms referred to extracellular matrix structural consitituent and structural molecule activity. KEGG results contained ECM-receptor interactions and pathways in cancer (**Figures 5A–D**).

Drug-Gene Networks

To seek effective target-drugs for hub DEGs, drug-gene interactions, including 75 potential drugs for UF, were obtained through the DGIdb database. For ease of reading, drug-gene networks were visualized by Cytoscape (**Figure 6**). However, the underlying mechanism between most potential drugs and hub DEGs remains unrevealed.

Molecular Docking

Molecular docking was applied to further screen potential drugs and elucidate possible molecular mechanisms. The crystal structure of human BIRC5 and TYMS was predicated by Alphafold v2.0 based on AI technology. As shown in **Figures 7A,B**, the stereochemical quality of the predicted structure was evaluated by the LDDT score. Autocdock Vina was used for drug-protein molecular docking to screen the optimal potential target drugs. The Affinity score can be used as a standard to judge the merits of docking. The high absolute value of the score indicated stronger binding between small molecules and proteins. The docking scores of potential drugs were presented in **Figures 7C–F** and suggested that irinotecan had the strongest binding affinity towards BIRC5 (7.0–8.6 kcal/mol, |interval score|) and TYMS (7.7–10.0 kcal/mol, |interval score|). The

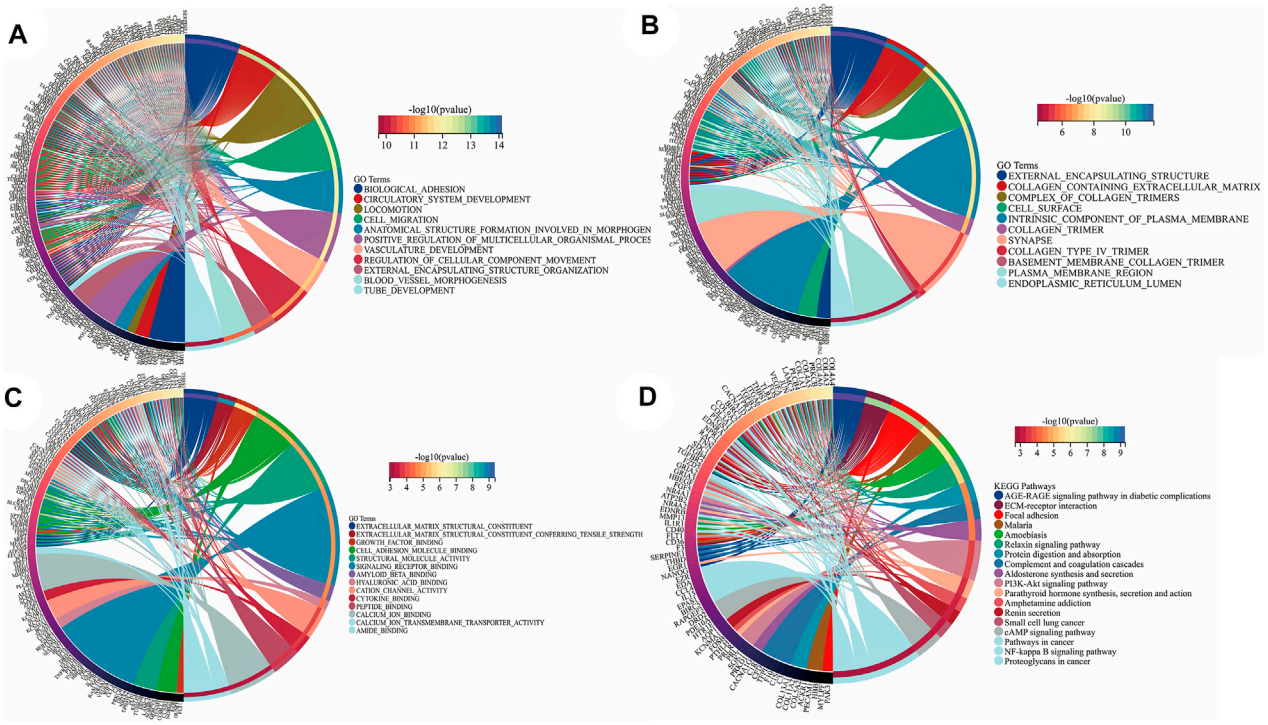


FIGURE 4 | Function annotation and KEGG analysis of DEGs in UF. **(A)** BP-terms; **(B)** CC-terms; **(C)** MF-terms; **(D)** KEGG pathways.

perfect conformation exhibited that irinotecan interacted with residues of BIRC5 (THR21 and HIS45) and TYMS (ARG50, ILE108, ASN226 and TYR258) through hydrogen bonds. After screening, irinotecan was found to be the optimal drug for intervention of both BIRC5 and TYMS and was used to treat UF patients in low doses.

Validation of Hub DEGs by RT-PCR

Based on UF and matched normal myometrial tissues in patients at our hospital, we validated the expression level of hub DEGs through RT-PCR (**Figures 8A, B**). Compared with normal myometrium, the expression levels of ASPM, SDC1, SNAI2, GINS1, CCNB2, KIAA0101, BIRC5, BUB1, TYMS, FOXM1, DTL, KIF20A, COL3A1 were increased, which was consistent with microarray analysis results. Moreover, the expression levels of TGFBR2, ITGA6, CYR61, ITGB4, COL11A2, COL4A5, COL4A6, and COL4A4 were decreased in comparison to normal myometrium. There were no significant differences in COL13A1 and COL5A2 expression.

Prediction of Target Genes for DEMs and miRNA–Target Gene Regulatory Network Establishment

Recent research indicated that tumorigenesis and development were regulated by miRNAs through binding with pairing bases of 3′UTR, resulting in target mRNAs degradation. Therefore, predication of target mRNAs can contribute to revealing the function and internal mechanism of DEMs. MiRWalk 3.0 was used to predict the target genes of DEMs. A total of 1,507 genes were identified as the potential target genes of the DEMs. After intersecting with 373 candidate DEGs, 28 overlapping genes were identified as DETGs for further analysis (**Figure 9A**). The miRNA-target gene regulatory network was visualized using Cytoscape. As indicated in **Figure 9B**, triangle represents DEMs, orange ellipses represent upregulated DETGs, and purple ellipses represent downregulated DETGs. The regulatory network included 21 DEMs, 16 upregulated DETGs, and 12 downregulated DETGs. The DEMs were ranked by the number of regulated DETGs, and the top two were identified as hub DEMs. In vitro and vivo experiments validated the upregulated hsa-miR-181b-5p and hsa-miR-381-3p as well as five screened DETGs.

Functional and Pathway Analysis of Hub DEMs

GO and KEGG pathway analysis were performed to elucidate the function of hub DEMs through miRPath v3.0. As illustrated in **Figures 9C–F**, these hub DEMs were enriched in ion binding, cellular nitrogen compound metabolic process, organelle, biosynthetic process and transcription, DNA-templated in the BP terms; Organelle and cellular component in CC terms; Ion binding, nucleic acid binding transcription factor activity and molecular function in MF terms. Furthermore, KEGG results indicated that hsa-miR-181b-5p and hsa-miR-381-3p involved in amphetamine addiction and TGF-β signaling pathway and proteoglycans in cancer respectively.

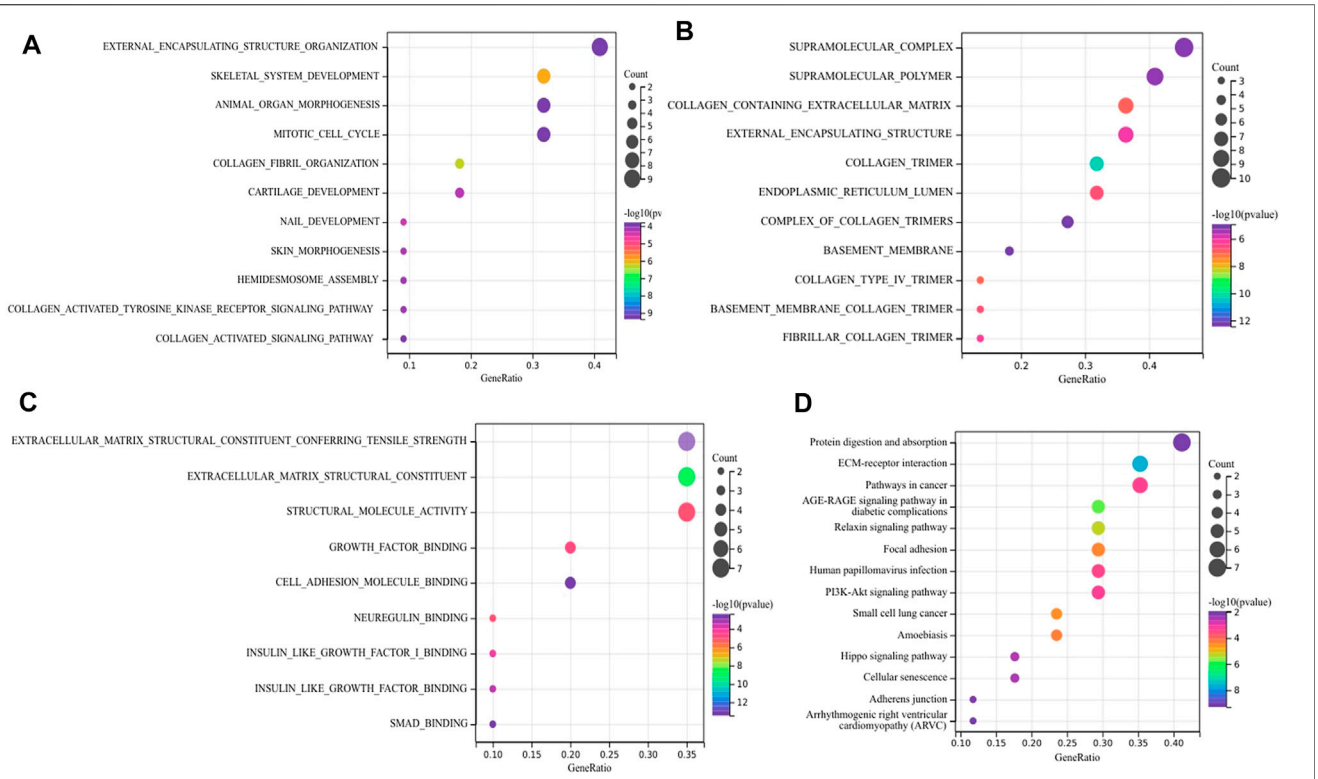


FIGURE 5 | Function annotation and KEGG analysis of hub DEGs. **(A)** BP-terms **(B)** CC-terms; **(C)** MF-terms; **(D)** KEGG pathways.

Construction of the miRNA-TF-Target Gene Regulatory Network

The transcriptional regulatory relationships among these miRNAs, TFs, and target genes were identified through the iRegulon plug-in. Then, miRNA-TF-target gene networks were established through Cytoscape software. In the transcriptional regulatory network, hsa-miR-181b-5p consisted of 2 TFs (STAT3 and ONECUT1) and hsa-miR-381-3p consisted of 2 TFs (FOXA1 and KAT2A) (**Figures 10A, B**).

Validation of Hub DEMs

In order to assess the expression of miR-181b-5p and miR-381b-3p in UF, RT-PCR was performed in UF tissues compared to normal myometrial tissues ($n = 5$). Consistent with our analysis of hub DEMs, we found that miR-181b-5p and miR-381b-3p were markedly upregulated in UF tissues compared to normal myometrium tissues (**Figure 10C**).

The Effect of Hub DEMs on Cell Proliferation

In the functional verification section, the hub DEMs, mimics, and inhibitors were transfected into UF cells. The CCK-8 assay was used to examine the effects of hsa-miR-181b-5p and hsa-miR-381-3p on cell proliferation. As demonstrated in **Figures 11A,B**, overexpression of both hub DEMs enhanced the proliferation of UF cells compared with NC mimics, while hsa-miR-181b-5p and hsa-miR-381-3p inhibitors could reverse the above results.

The Effect of Hub DEMs on Cell Migration

Although UF is a benign tumor with poor metastatic ability, it may locally migrate. Therefore, the evaluation of cell migration still reflects hub DEMs' function. Many studies have also shown that UF cells have the metastatic ability. After transfection with hsa-miR-181b-5p and hsa-miR-381-3p mimics, the number of migrational cells was higher than NC mimic group. On the contrary, the metastatic ability of transfected cells with hsa-miR-181b-5p and hsa-miR-381-3p inhibitors was remarkably reduced, with fewer cells crossing through the membranes. Above, results suggested that there were positive correlations between the metastatic ability of UF cells and hsa-miR-181b-5p and hsa-miR-381-3p expression (**Figure 11C**)

The Effect of Hub DEMs on Cell Apoptosis

Although surgical excision is the undisputed treatment for benign tumors, apoptosis of tumors induced by non-invasive procedures can effectively avoid operative trauma and tends to improve the outcome. MiRNA therapeutics provides a novel horizon for tumor treatment. To verify whether hub DEMs' silence can induce UF cells' apoptosis, we transfected NC, hsa-miR-181b-5p, and hsa-miR-381-3p inhibitors to UF cells. As indicated in **Figure 11D**. Downregulation of hsa-miR-181b-5p and hsa-miR-381-3 could significantly contribute to cell apoptosis through comparison of cell apoptosis rates in each group by flow cytometry.

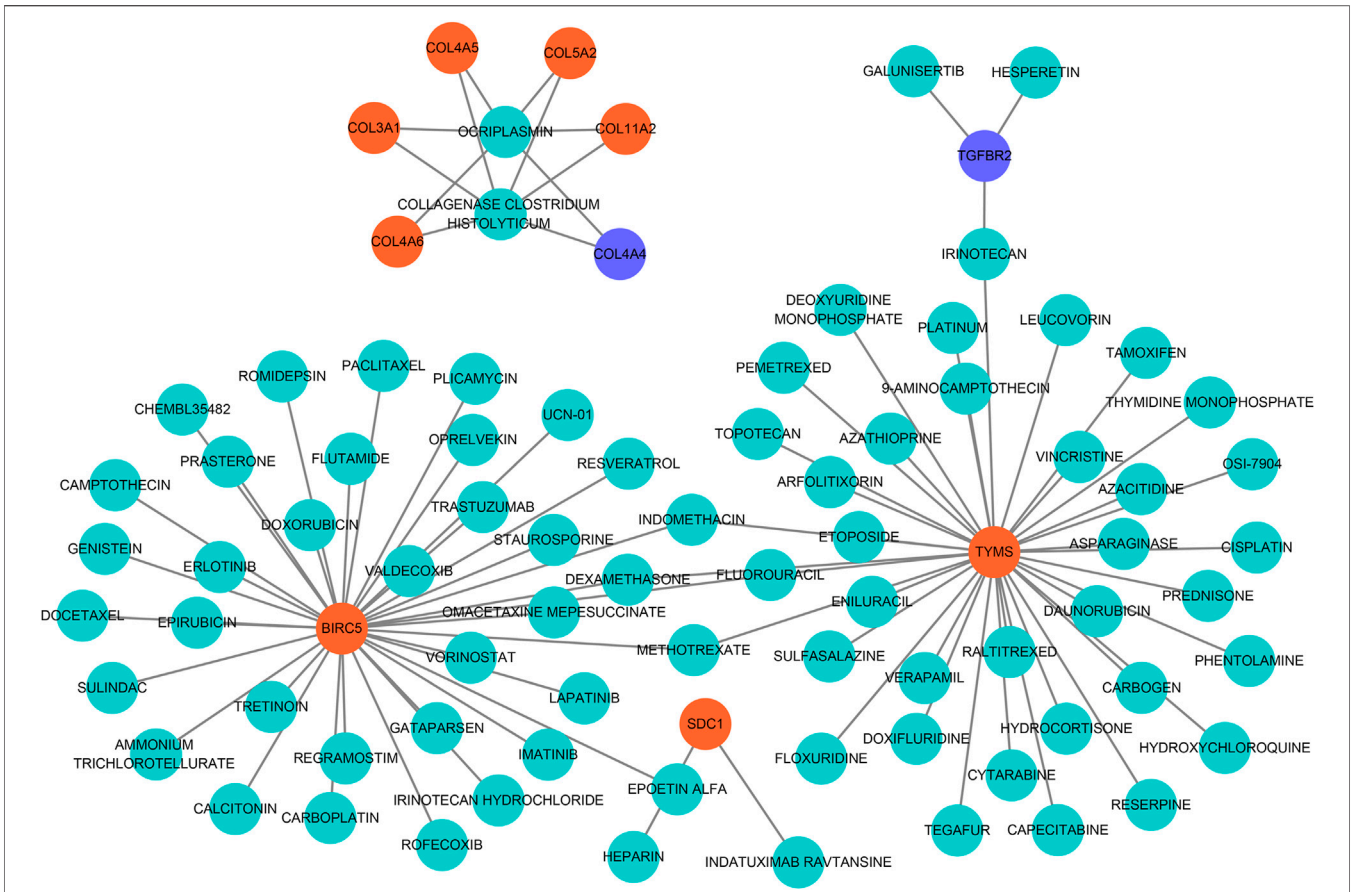


FIGURE 6 | Drug-hub DEGs regulatory networks. Cyan circles, potential target drugs; Orange circles, upregulated DEGs; Purple circles, downregulated DEGs.

DISCUSSION

As the most common benign tumor of the female reproductive system, the worldwide prevalence of UF has been increasing year by year (Pitter et al., 2013). Although UF is rarely malignant, it is often associated with menstrual changes, abdominal pain, abortions, and infertility (Giuliani et al., 2020). In addition, hysterectomy caused by UF seriously reduces the quality of life and places a heavy burden on basic social services such as health care (Cardozo et al., 2012). Unlike other types of cancer that have clear origins, there is a long-standing debate about whether the origin of UF is the immature smooth muscle cell derived from the uterine wall or the walls of uterine vessels (Akhter et al., 2021). Therefore, UF treatment cannot be applied to other tumor targets, indicating that screening of hub miRNAs and mRNAs associated with UF is urgently needed. Besides, the synergistic effects of TFs and miRNAs have become a new hot research topic for uncovering the etiology of diseases (Ciebiera et al., 2020). Unfortunately, to our knowledge, there is no correlated report establishing miRNA-TF-mRNA regulatory networks of UF. As pioneers, we first revealed internal interactions among miRNAs, TFs, and then mRNAs and revisited UF therapy from a novel perspective.

From the mRNA expression profile, 373 candidate DEGs were screened by taking the intersection of the two mRNA datasets. The sangerbox database was utilized to reveal the function and signaling pathway of DEGs. In the results of enrichment analysis, DEGs were closely associated with extracellular matrix structural constituents and biological adhesion, which was consistent with histopatological features of UF. Moreover, KEGG analysis also indicated that pathways enriched by DEGs are involved in ECM-receptor interaction, focal adhesion, and the PI3K-AKT pathway. In these DEGs, cysteine-rich angiogenic inducer 61 (CYR61) was widely reported to be correlated to UF (Sampath et al., 2001) (Arslan et al., 2005). Interesting, as an insulin-like growth factor (IGF)-binding protein, CYR61 degradation could promote IGF2 binding to IGF1R, leading to proliferation acceleration, ECM formation, and apoptosis inhibition (Sarkissyan et al., 2014).

To reduce the scope of DEGs, we established a PPI network and screened 24 hub DEGs, each containing 5 upregulated DEGs and 19 downregulated DEGs. Although most of these genes were subtypes of collagen, these data are also in line with the most striking pathological feature of UF (Bulun, 2013), excessive accumulation of ECM composed of collagen, resulting in irregular bleeding, pelvic pain, and compression. The ECM provides a microenvironment for cell adhesion, proliferation, and migration. Unfortunately, the internal interaction between

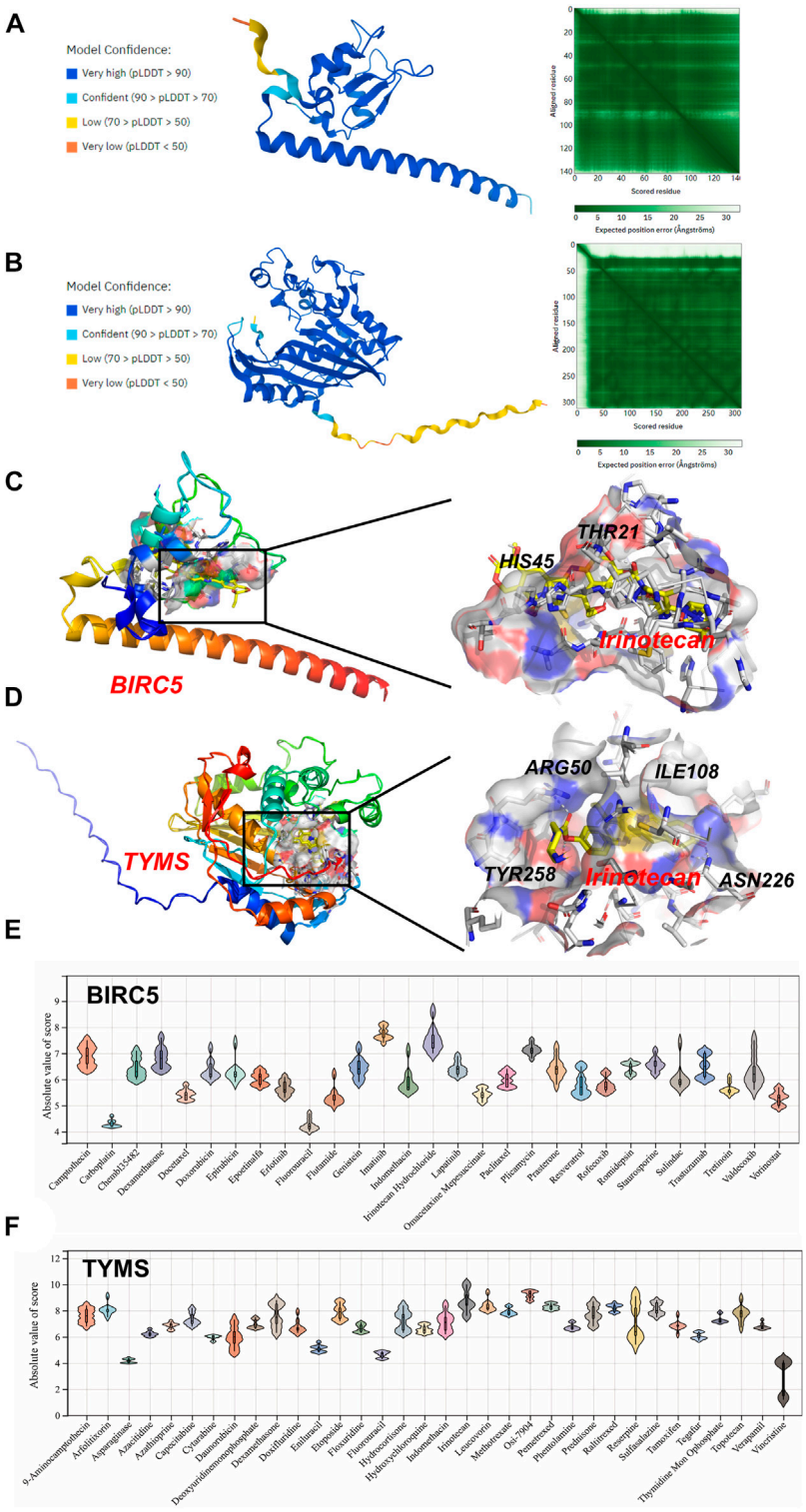


FIGURE 7 | Homology Modeling and Molecular Docking. The crystal structure and evaluation of **(A)** BIRC5 and **(B)** TYMS. The best docking position between irinotecan and **(C)** BIRC5 or **(D)** TYMS was indicated. The absolute value of affinity between predicated small molecules and **(E)** BIRC5 or **(F)** TYMS was exhibited.

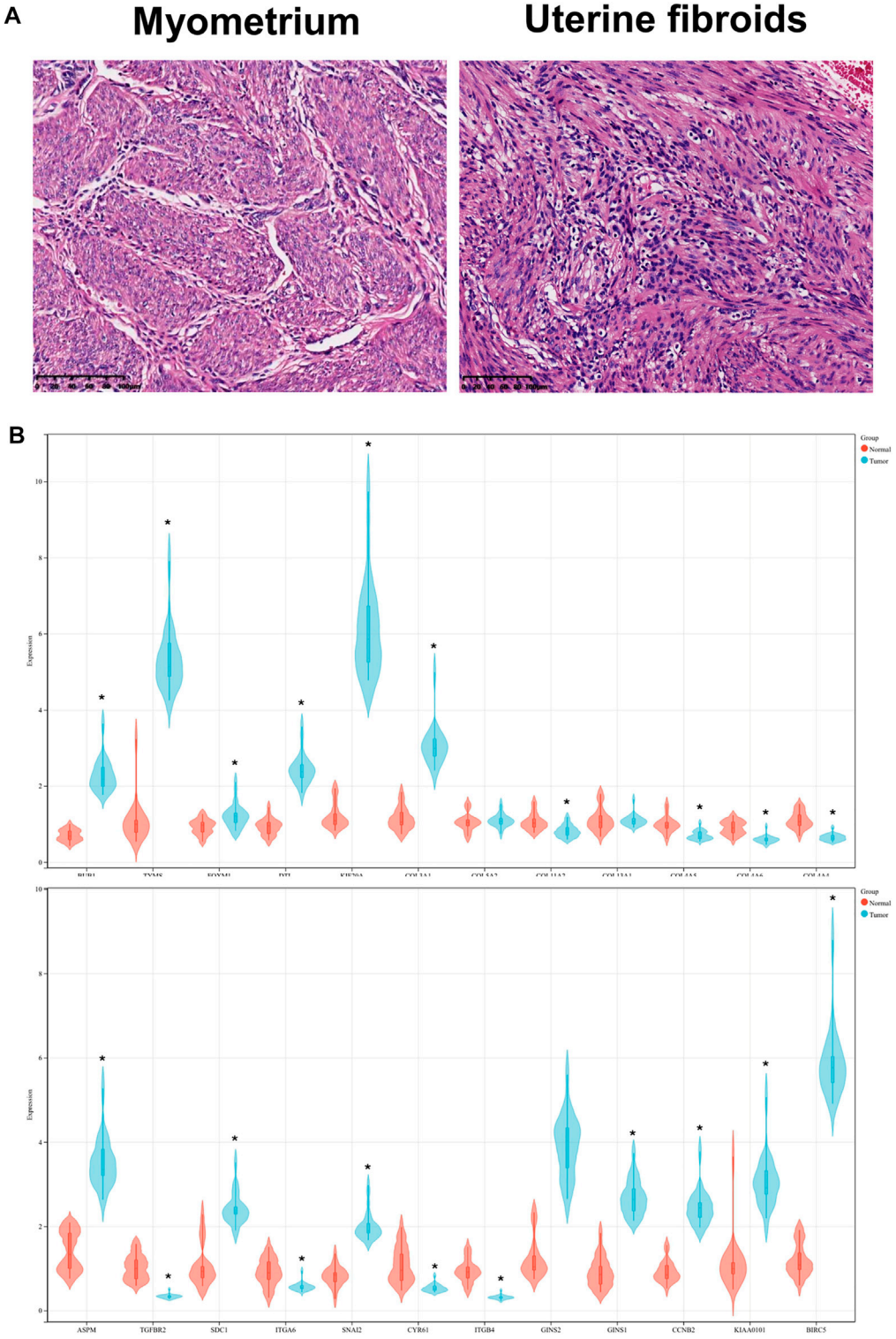


FIGURE 8 | (A) HE staining of UF tissues and myometrium. **(B)** Validation of hub DEGs expression in UF by RT-PCR.

ECM and UF cells remains unknown and is a hot issue to investigate deeply. Furthermore, as an essential transcriptional activator, aberrant SNAI2 can induce cell proliferation, invasion, and the epithelial-mesenchymal transition (EMT) process in a variety of malignancies (Casas et al., 2011). In gastric and lung cancers, abnormal expression of SNAI2 is associated with poor

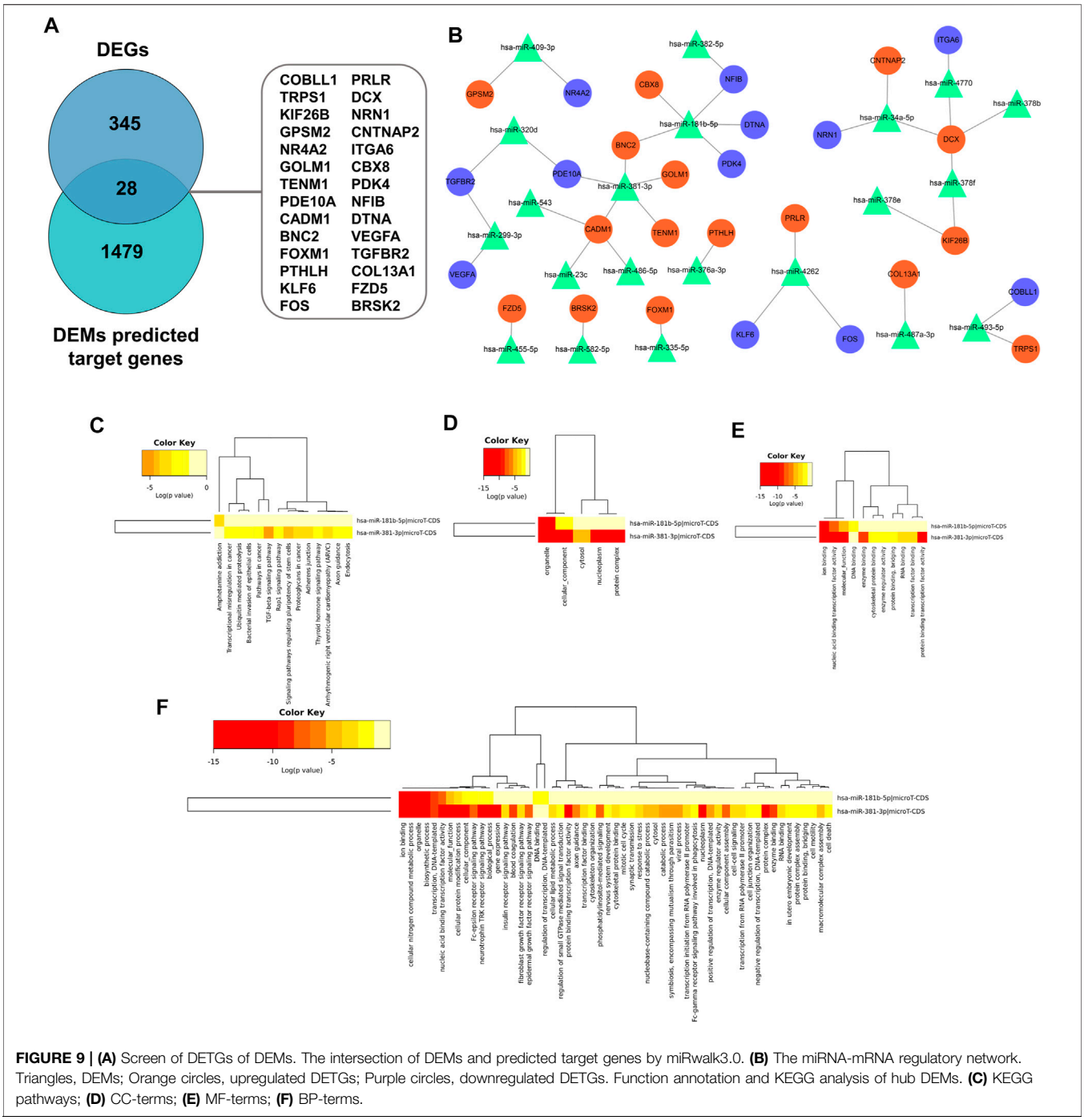


FIGURE 9 | (A) Screen of DETGs of DEMs. The intersection of DEMs and predicted target genes by miRwalk3.0. **(B)** The miRNA-mRNA regulatory network. Triangles, DEMs; Orange circles, upregulated DETGs; Purple circles, downregulated DETGs. Function annotation and KEGG analysis of hub DEMs. **(C)** KEGG pathways; **(D)** CC-terms; **(E)** MF-terms; **(F)** BP-terms.

prognosis (Wang et al., 2021). Studies also indicated that the increased expression level of SDC1 could promote the growth, invasion and metastasis of tumor cells (Cao et al., 2021). FOXM1 plays a critical role in the regulation of cell proliferation, migration, and angiogenesis (Liu et al., 2021). It has been found that FOXM1 plays a vital role in mitosis, and its abnormal expression can cause spindle defects, which can delay mitosis and ultimately participate in the development of cervical (Li et al., 2021), liver (Nandi et al., 2021), lung (Liang et al., 2021) and other cancers. To validate the expression level of hub DEGs, we performed RT-PCR based on collected samples from our hospital. Most of the results on gene expression were in accordance with microarray results and provided corroborative evidence for our analysis.

Although limited drugs, such as GnRH (Dababou et al., 2021), mifepristone (Bi et al., 2021), are widely used in the treatment of UF, they lack a well-defined target and always bring patients side-effects (for example, recidivation after drug withdrawal). Therefore, target drugs for UF urgently need to be discovered to avoid physical injury caused by hysterectomy. Based on

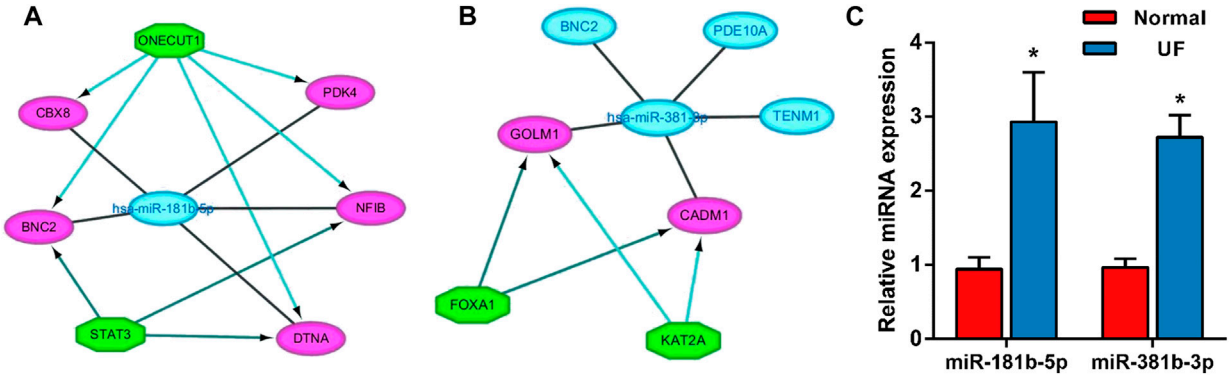


FIGURE 10 | The miRNA-TF-mRNA regulatory networks of **(A)** hsa-miR-181-5p (2 TFs) and **(B)** hsa-miR-381-3p (2 TFs). **(C)** Expression of the miR-181b-5p, miR-381b-3p in UF tissues (n = 5) compared to normal myometrium tissues (n = 5), as assessed by reverse transcription–quantitative polymerase chain reaction.

predicted hub DEGs, 75 target drugs approved by the FDA were discovered through the DGIdb database. As a natural polyphenol with anti-tumor property, resveratrol is a potent inhibitor of the progesterone X receptor (PXR) and may inhibit UF cell proliferation by decreasing estrogen and progesterone levels or antagonizing their receptors (Singh et al., 2011). In addition, ocriplasmin is applied to treat vitreomacular adhesions through the hydrolysis of collagen (Al-Nawaiseh et al., 2021). Considering the large number of target drugs for BIRC5 and TYMS, we utilized molecular docking to select the best candidates. Interestingly, irinotecan had the strongest binding to BIRC5 and TYMS through hydrogen bond formation. We hoped that the above results could give clinicians more novel therapeutic strategies.

Over the last few decades, miRNAs became one of the most discussed issues and was deemed as a biomarker to achieve early screening and clinical treatment. Especially in recent years, the advent of the first miRNA drug (miravirsen) rekindled the topic of morbigenous miRNA screening (Zhang et al., 2021). Certainly, miRNAs are involved in the regulation of multiple activities of UF. We identified DEMs based on the GSE159959 dataset and obtained 28 DEGTs through overlapping between target genes predicted by miRwalk3.0 and DEGs. Additionally, the miRNA-mRNA regulatory network was constructed and miR-181b-5p and miR-381-3p were identified as hub DEMs. The function analysis of hub DEMs indicated that they were enriched in ion binding, cellular nitrogen compound metabolic processes, organelles, biosynthetic processes, and transcription. Furthermore, the KEGG pathway analysis exhibited hsa-miR-181b-5p and hsa-miR-381-3p involved in amphetamine addiction and TGF-β signaling pathway and proteoglycans in cancer respectively.

Dysregulated miR-181b-5p was observed in various cancers and influenced the function of tumor cells. A previous study demonstrated that upregulated miR-181b-5p could promote cell proliferation, migration, and invasion by targeting PARK2 via the PTEN/PI3K/AKT pathway in cholangiocarcinoma (Jiang et al., 2021). Additionally, miR-181b-5p is also closely linked to drug resistance for gallbladder carcinoma (Wu et al., 2019). The potency of ginsenoside Rg3 could be abolished by miR-181b-5p through enhancing autophagy flux *via* the CREBRF/CREB3 pathway (Wu et al., 2019). Wu et al. also reported that miR-181b-5p was verified to be a biomarker for pituitary adenoma (Wu et al., 2017). Similarly, miR-181b-5p upregulation was involved in head and neck cancer pathogenesis (Nurul-Syakima et al., 2011). Nakajima G et al. found that miR-181b-5p was overexpressed in colon cancer tissues and promoted the occurrence and development of tumors by regulating cell signal transduction and cell cycle (Nakajima et al., 2006). Taken together, we have reasons to believe that miR-181b-5p plays a vital role in tumorigenesis and the progression of UF.

MiR-381-3p was reported as an oncogenic regulator involving in a wide range of cancers. Zhao et al. indicated that miR-381-3p overexpression could accelerate cell proliferation and colony formation, which further initiates renal tumor deterioration and poor prognosis (Zhao et al., 2020). A gynecologic cancer-associated study also elucidated that miR-381-3p had a positive correlation with uterine corpus endometrial carcinoma (Zheng et al., 2021). Furthermore, some studies showed that miR-381-3p was upregulated in glioma (Li et al., 2016), osteosarcoma (Li et al., 2016), synovial sarcoma (Hisaoka et al., 2011) and epithelioid sarcoma (Papp et al., 2014). Therefore, we hypothesized that miR-381-3p may become a clinical diagnostic and therapeutic target of UF. Definitely, due to spatial heterogeneity in tumor tissue, miR-381-3p was regarded as a tumor suppressor in gastric cancer (Gao et al., 2022), breast cancer (Yu et al., 2021) and bladder cancer (Li et al., 2019). We performed *in vitro* experiments to examine the function of miR-381-3p on UF, and the results verified our hypothesis.

To investigate the synergistic effect of TFs, miRNA-TF-mRNA regulatory networks of miR-181b-5p and miR-381-3p were established. miR-181b-5p and 2 TFs, STAT3 and ONECUT1, were verified to be coregulators of CBX8, PDK4, NFIB, BNC2, and DTNA. miR-381-3p and 2 TFs, FOXA1 and KAT2A, were verified to be coregulators of GOLM1, TENM1, PDE10A, CADM1, and BNC2. These two networks provided a novel horizon to explain synergistic interactions between miRNAs and TFs and potential combined treatments for UF.

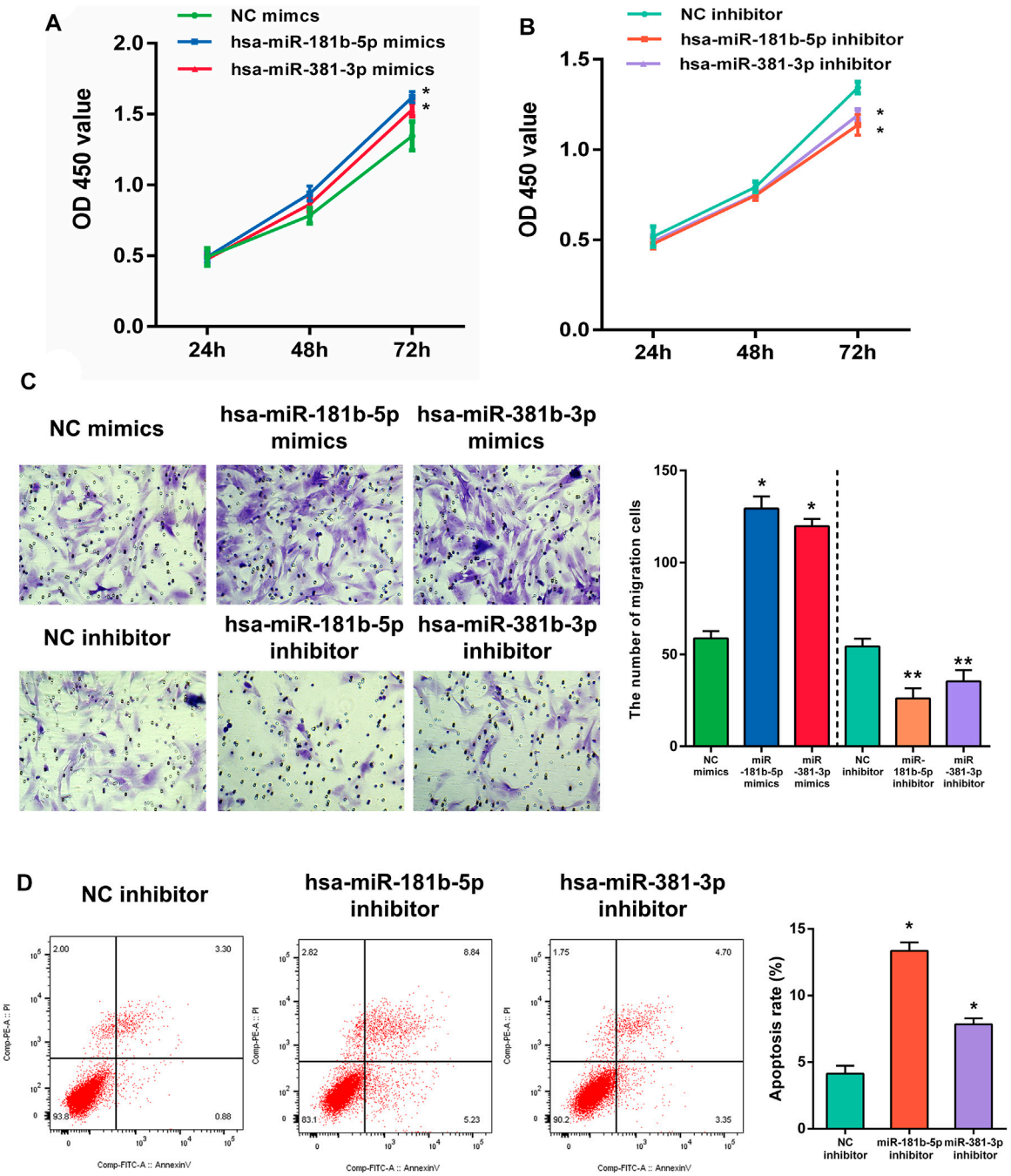


FIGURE 11 | The proliferation of UF cells regulated by **(A)** overexpression or **(B)** slience of hub DEMs was detected by CCK-8 assay. **(C)** The migrational ability of UF cells regulated by overexpression or slience of hub DEMs was detected by Transwell. **(D)** The apoptosis rates of UF cells regulated by overexpression or slience of hub DEMs was detected by AV/PI assay.

RT-PCR was carried out to detect hub DEG expression in myometrium and uterine fibroids, RT-PCR was carried out. Compared with normal myometrium, the expression levels of ASPM, SDC1, SNAI2, GINS1, CCNB2, KIAA0101, BIRC5, BUB1, TYMS, FOXM1, DTL, KIF20A, COL3A1 were increased, which was consistent with microarray analysis results. Moreover, the expression levels of TGFBR2, ITGA6, CYR61, ITGB4, COL11A2, COL4A5, COL4A6, COL4A4 were decreased in comparison to normal myometrium. There were no significant differences in COL13A1 and COL5A2 expression.

Furthermore, in order to explore the effect of hub DEMs on UF cell proliferation, migration, and apoptosis, hsa-miR-181-5p and hsa-miR-381-3p were overexpressed or silenced and delivered into UF cells. CCK-8 demonstrated that hsa-miR-181-5p and hsa-miR-

381-3p overexpression enhanced cell proliferation and migration in uterine fibroids cells compared with NC mimics, while silencing of hsa-miR-181-5p and hsa-miR-381-3p reversed the above results. Furthermore, we examined UF cells transfected with hub DEMs' inhibitor. Downregulation of hsa-miR-181b-5p and hsa-miR-381-3p could significantly contribute to cell apoptosis in comparison of cell apoptosis rates of NC inhibitor by flow cytometry. Therefore, the function of hsa-miR-181b-5p and hsa-miR-381-3p deserves further study to elucidate the internal mechanism in UF.

Certainly, our study inevitably has some limitations. Although we performed *in vitro* experiments through overexpression and silence of hub DEMs, *in vivo* experiments could better simulate tumorigenesis and progression of UF to verify their function. Additionally, we did not assess the expression of target genes upon transfection. Therefore, our further study will conduct RT-PCR to evaluate the expression of target genes and establish an animal model to reveal the underlying mechanism of hub DEMs in UF. Screening chip data from the same platform can effectively avoid batch effects. In addition, miRNA is a type of noncoding RNA that includes lncRNAs, circRNAs etc. Whether other noncoding RNAs also play a vital role in UF is worth discussing. We hope to discover more noncoding RNAs as specific biomarkers to enhance the accuracy of early screening and improve patients' outcomes.

CONCLUSION

In conclusion, we identified some specific biomarkers (DEGs and DEMs) associated with the diagnosis and prognosis of UF based on bioinformatics analysis. Next, hub DEGs were confirmed by RT-PCR of clinical UF samples. Furthermore, we first established miRNA-TF-mRNA regulatory networks of UF and screened hsa-hsa-miR-181-5p and hsa-miR-381-3p as hub DEMs that are associated with regulation of cell proliferation, migration, and apoptosis in UF through *in vitro* validation. Our study provided several novel biomarkers to reveal possible mechanisms and was believed to be helpful for the diagnostic and therapeutic strategy of UF.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/geo/, GSE593; https://www.ncbi.nlm.nih.gov/geo/, GSE18096; https://www.ncbi.nlm.nih.gov/geo/, GSE159959.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Medical Ethics Committee at Xiangya Hospital, Central South University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Conception and design: XP, HL, and SW. Writing, review, and/or revision of the manuscript: XP, HL, SW, and YM. Administrative, technical, or material support: XP, HL, SW, and JL. All authors approved the final version of the manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (Grant Number: 81472434), the Science and Technology Planning Project of Changsha (Grant Number: K1508031-31).

REFERENCES

Ahn, S. H., Kim, H., Lee, I., Lee, J. H., Cho, S., and Choi, Y. S. (2021). MicroRNA-139-5p Regulates Fibrotic Potentials via Modulation of Collagen Type 1 and Phosphorylated P38 MAPK in Uterine Leiomyoma. *Yonsei Med. J.* 62 (8), 726–733. doi:10.3349/ymj.2021.62.8.726

Akhter, S., Jahan, K., Afrose, R., Jahan, R., Fatema, S., and Dey, S. (2021). Risk Association, Clinical Presentation and Management of Uterine Leiomyoma. *Mymensingh Med. J.* 30 (4), 907–912.

Al-Nawaiseh, S., Rickmann, A., Seuthe, A.-M., Al-Rimawi, D., Viktor Stanzel, B., and Szurman, P. (2021). *A Revised Predictive Biomarker for Ocriplasmin Therapy in Vitreomacular Traction Disorders*, 41. Philadelphia, Pa: Retina, 2549–2555. doi:10.1097/iae.0000000000003244

Ali, M., Esfandyari, S., and Al-Hendy, A. (2020). Evolving Role of microRNAs in Uterine Fibroid Pathogenesis: Filling the gap!. *Fertil. Sterility* 113 (6), 1167–1168. doi:10.1016/j.fertnstert.2020.04.011

Andersen, J., and Barbieri, R. L. (1995). Abnormal Gene Expression in Uterine Leiomyomas. *J. Soc. Gynecol. Investig.* 2 (5), 663–672. doi:10.1177/107155769500200501

Arslan, A. A., Gold, L. I., Mittal, K., Suen, T.-C., Belitskaya-Levy, I., Tang, M.-S., et al. (2005). Gene Expression Studies Provide Clues to the Pathogenesis of Uterine Leiomyoma: New Evidence and a Systematic Review. *Hum. Reprod. (Oxford, England)* 20 (4), 852–863. doi:10.1093/humrep/deh698

Bader, G. D., and Hogue, C. W. (2003). An Automated Method for Finding Molecular Complexes in Large Protein Interaction Networks. *BMC Bioinformatics* 4, 2. doi:10.1186/1471-2105-4-2

Bi, C., Qiao, M., Jia, Y., and Wang, H. (2021). Network Meta-Analysis of 6 Kinds of Chinese Patent Medicines Combined with Mifepristone in the Treatment of Uterine Fibroids. *Medicine* 100 (42), e27523. doi:10.1097/md.0000000000027523

Bulun, S. E. (2013). Uterine Fibroids. *N. Engl. J. Med.* 369 (14), 1344–1355. doi:10.1056/NEJMra1209993

Cao, G., Zhang, C., Tian, X., Jing, G., Zhou, X., and Yan, T. (2021). circCEP128 Knockdown Suppresses Bladder Cancer Progression via Regulating microRNA-515-5p/SDC1 Axis. *Cmar* 13, 2885–2896. doi:10.2147/cmar.S288229

Cardozo, E. R., Clark, A. D., Banks, N. K., Henne, M. B., Stegmann, B. J., and Segars, J. H. (2012). The Estimated Annual Cost of Uterine Leiomyomata in the United States. *Am. J. Obstet. Gynecol.* 206 (3), e1e211–211. doi:10.1016/j.ajog.2011.12.002

Casas, E., Kim, J., Bendesky, A., Ohno-Machado, L., Wolfe, C. J., and Yang, J. (2011). Snail2 Is an Essential Mediator of Twist1-Induced Epithelial Mesenchymal Transition and Metastasis. *Cancer Res.* 71 (1), 245–254. doi:10.1158/0008-5472.Can-10-2330

Chen, H., and Boutros, P. C. (2011). VennDiagram: a Package for the Generation of Highly-Customizable Venn and Euler Diagrams in R. *BMC Bioinformatics* 12, 35. doi:10.1186/1471-2105-12-35

Ciebiera, M., Włodarczyk, M., Zgliczyński, S., Łoziński, T., Walczak, K., and Czekierdowski, A. (2020). The Role of miRNA and Related Pathways in

Pathophysiology of Uterine Fibroids-From Bench to Bedside. *Ijms* 21 (8), 3016. doi:10.3390/ijms21083016

Dababou, S., Garzon, S., Laganà, A. S., Ferrero, S., Evangelisti, G., Noventa, M., et al. (2021). Linzagolix: a New GnRH-Antagonist under Investigation for the Treatment of Endometriosis and Uterine Myomas. *Expert Opin. Investig. Drugs* 30, 903–911. doi:10.1080/13543784.2021.1957830

Fitzgerald, J. B., Chennathukuzhi, V., Koohestani, F., Nowak, R. A., and Christenson, L. K. (2012). Role of microRNA-21 and Programmed Cell Death 4 in the Pathogenesis of Human Uterine Leiomyomas. *Fertil. Sterility* 98 (3), 726–734.e722. doi:10.1016/j.fertnstert.2012.05.040

Fu, X., and Dong, D. (2018). Bioinformatic Analysis of MicroRNA Sequencing Data. *Methods Mol. Biol.* 1751, 109–125. doi:10.1007/978-1-4939-7710-9_8

Gao, X., Liu, H., Wu, Q., Wang, R., Huang, M., Ma, Q., et al. (2022). miRNA-381-3p Functions as a Tumor Suppressor to Inhibit Gastric Cancer by Targeting Fibroblast Growth Factor Receptor-2. *Cancer Biother. Radiopharm*. doi:10.1089/cbr.2021.0357

Giuliani, E., As-Sanie, S., and Marsh, E. E. (2020). Epidemiology and Management of Uterine Fibroids. *Intl J. Gynecol. Obste* 149 (1), 3–9. doi:10.1002/ijgo.13102

Hisaoka, M., Matsuyama, A., Nagao, Y., Luan, L., Kuroda, T., Akiyama, H., et al. (2011). Identification of Altered MicroRNA Expression Patterns in Synovial Sarcoma. *Genes Chromosom. Cancer* 50 (3), 137–145. doi:10.1002/gcc.20837

Hu, Y., Lan, W., and Miller, D. (2017). Next-Generation Sequencing for MicroRNA Expression Profile. *Methods Mol. Biol.* 1617, 169–177. doi:10.1007/978-1-4939-7046-9_12

Islam, M. S., Protic, O., Stortoni, P., Grechi, G., Lamanna, P., Petraglia, F., et al. (2013). Complex Networks of Multiple Factors in the Pathogenesis of Uterine Leiomyoma. *Fertil. sterility* 100 (1), 178–193. doi:10.1016/j.fertnstert.2013.03.007

Jiang, Z.-L., Zhang, F.-X., Zhan, H.-L., Yang, H.-J., Zhang, S.-Y., Liu, Z.-H., et al. (2021). miR-181b-5p Promotes the Progression of Cholangiocarcinoma by Targeting PARK2 via PTEN/PI3K/AKT Signaling Pathway. *Biochem. Genet.* 60, 223–240. doi:10.1007/s10528-021-10084-5

Lee, J. H., Choi, Y. S., Park, J. H., Kim, H., Lee, I., Won, Y. B., et al. (2019). MiR-150-5p May Contribute to Pathogenesis of Human Leiomyoma via Regulation of the Akt/p27Kip1 Pathway *In Vitro*. *Ijms* 20 (11), 2684. doi:10.3390/ijms20112684

Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes Are microRNA Targets. *Cell* 120 (1), 15–20. doi:10.1016/j.cell.2004.12.035

Li, J., Ying, Y., Xie, H., Jin, K., Yan, H., Wang, S., et al. (2019). Dual Regulatory Role of CCNA2 in Modulating CDK6 and MET-mediated Cell-cycle Pathway and EMT Progression Is Blocked by miR-381-3p in Bladder Cancer. *FASEB j.* 33 (1), 1374–1388. doi:10.1096/fj.201800667R

Li, Y., Zhao, C., Yu, Z., Chen, J., She, X., Li, P., et al. (2016). Low Expression of miR-381 Is a Favorite Prognosis Factor and Enhances the Chemosensitivity of Osteosarcoma. *Oncotarget* 7 (42), 68585–68596. doi:10.18632/oncotarget.11861

Li, Z., Liu, X., Yu, H., Wang, S., Zhao, S., and Jiang, G. (2021). USP21 Regulates Hippo Signaling to Promote Radioresistance by Deubiquitinating FOXM1 in Cervical Cancer. *Hum. Cel.* 35, 333–347. doi:10.1007/s13577-021-00650-9

Liang, S.-K., Hsu, C.-C., Song, H.-L., Huang, Y.-C., Kuo, C.-W., Yao, X., et al. (2021). Correction: FOXM1 Is Required for Small Cell Lung Cancer Tumorigenesis and Associated with Poor Clinical Prognosis. *Oncogene* 40, 6705. doi:10.1038/s41388-021-02044-5

Liu, C., Barger, C. J., and Karpf, A. R. (2021). FOXM1: A Multifunctional Oncoprotein and Emerging Therapeutic Target in Ovarian Cancer. *Cancers* 13 (12), 3065. doi:10.3390/cancers13123065

Nakajima, G., Hayashi, K., Xi, Y., Kudo, K., Uchida, K., Takasaki, K., et al. (2006). Non-coding MicroRNAs Hsa-Let-7g and Hsa-miR-181b Are Associated with Chemoresponse to S-1 in Colon Cancer. *Cancer Genomics Proteomics* 3 (5), 317–324.

Nandi, D., Cheema, P. S., Singal, A., Bharti, H., and Nag, A. (2021). Artemisinin Mediates its Tumor-Suppressive Activity in Hepatocellular Carcinoma through Targeted Inhibition of FoxM1. *Front. Oncol.* 11, 751271. doi:10.3389/fonc.2021.751271

Papp, G., Krausz, T., Stricker, T. P., Szendrői, M., and Sápi, Z. (2014). SMARCB1expression in Epithelioid Sarcoma Is Regulated by miR-206, miR-381, and miR-671-5p on Both mRNA and Protein Levels. *Genes Chromosomes Cancer* 53 (2), 168–176. doi:10.1002/gcc.22128

Pitter, M. C., Gargiulo, A. R., Bonaventura, L. M., Lehman, J. S., and Srouji, S. S. (2013). Pregnancy Outcomes Following Robot-Assisted Myomectomy. *Hum. Reprod.* 28 (1), 99–108. doi:10.1093/humrep/des365

Rupaimoole, R., and Slack, F. J. (2017). MicroRNA Therapeutics: towards a new era for the Management of Cancer and Other Diseases. *Nat. Rev. Drug Discov.* 16 (3), 203–222. doi:10.1038/nrd.2016.246

Sampath, D., Zhu, Y., Winneker, R. C., and Zhang, Z. (2001). Aberrant Expression of Cyr61, a Member of the CCN (CTGF/Cyr61/Cef10/NOVH) Family, and Dysregulation by 17β-Estradiol and Basic Fibroblast Growth Factor in Human Uterine Leiomyomas. *J. Clin. Endocrinol. Metab.* 86 (4), 1707–1715. doi:10.1210/jcem.86.4.7423

Sarkissyan, S., Sarkissyan, M., Wu, Y., Cardenas, J., Koeffler, H. P., and Vadgama, J. V. (2014). IGF-1 Regulates Cyr61 Induced Breast Cancer Cell Proliferation and Invasion. *PloS one* 9 (7), e103534. doi:10.1371/journal.pone.0103534

Singh, M., Parent, S., Leblanc, V., and Asselin, E. (2011). Resveratrol Modulates the Expression of PTGS2 and Cellular Proliferation in the Normal Rat Endometrium in an AKT-dependent Manner1. *Biol. Reprod.* 84 (5), 1045–1052. doi:10.1095/biolreprod.110.090076

Walker, C. L., and Stewart, E. A. (2005). Uterine Fibroids: the Elephant in the Room. *Science* 308 (5728), 1589–1592. doi:10.1126/science.1112063

Wang, N., Wu, S., Zhao, J., Chen, M., Zeng, J., Lu, G., et al. (2021). Bile Acids Increase Intestinal Marker Expression via the FXR/SNAI2/miR-1 axis in the Stomach. *Cell Oncol.* 44, 1119–1131. doi:10.1007/s13402-021-00622-z

Wu, K., Huang, J., Xu, T., Ye, Z., Jin, F., Li, N., et al. (2019). MicroRNA-181b Blocks Gensenoside Rg3-Mediated Tumor Suppression of Gallbladder Carcinoma by Promoting Autophagy Flux via CREBRF/CREB3 Pathway. *Am. J. Transl Res.* 11 (9), 5776–5787.

Wu, S., Gu, Y., Huang, Y., Wong, T.-C., Ding, H., Liu, T., et al. (2017). Novel Biomarkers for Non-functioning Invasive Pituitary Adenomas Were Identified by Using Analysis of microRNAs Expression Profile. *Biochem. Genet.* 55 (3), 253–267. doi:10.1007/s10528-017-9794-9

Yoke-Kqueen, C., Yoke-Kqueen, C., Sabariah, A., Shiran, M., Singh, A., and Learn-Han, L. (2011). Differential microRNA Expression and Identification of Putative miRNA Targets and Pathways in Head and Neck Cancers. *Int. J. Mol. Med.* 28 (3), 327–336. doi:10.3892/ijmm.2011.714

Yu, Y.-Z., Mu, Q., Ren, Q., Xie, L.-J., Wang, Q.-T., and Wang, C.-P. (2021). miR-381-3p Suppresses Breast Cancer Progression by Inhibition of Epithelial-Mesenchymal Transition. *World J. Surg. Onc* 19 (1), 230. doi:10.1186/s12957-021-02344-w

Zhang, L., Xu, Q., Li, Y., Zhao, H., Shi, X., Peng, F., et al. (2021a). Ameliorative Effects of Component Chinese Medicine from Curcumae Rhizoma and Sparganii Rhizoma, a Traditional Herb Pair, on Uterine Leiomyoma in a Rat Model. *Front. Public Health* 9, 674357. doi:10.3389/fpubh.2021.674357

Zhang, S., Cheng, Z., Wang, Y., and Han, T. (2021). The Risks of miRNA Therapeutics: In a Drug Target Perspective. *Dddt* 15, 721–733. doi:10.2147/dddt.S288859

Zhao, C., Zhou, Y., Ran, Q., Yao, Y., Zhang, H., Ju, J., et al. (2020). MicroRNA-381-3p Functions as a Dual Suppressor of Apoptosis and Necroptosis and Promotes Proliferation of Renal Cancer Cells. *Front. Cel Dev. Biol.* 8, 290. doi:10.3389/fcell.2020.00290

Zheng, X., Xiuyi, L., Zhu, L., Xu, K., Shi, C., Cui, L., et al. (2021). The Circadian Gene NPAS2 Act as a Putative Tumor Stimulative Factor for Uterine Corpus Endometrial Carcinoma. *Cmar* 13, 9329–9343. doi:10.2147/cmar.S343097

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 Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 04 April 2022
doi: 10.3389/fbioe.2022.862619

A Novel Quantification System Combining iTRAQ Technology and Multi-Omics Assessment to Predict Prognosis and Immunotherapy Efficacy in Colon Cancer

Tianyi Xia[1†], *Junnan Guo*[1†], *Bomiao Zhang*[1], *Weinan Xue*[1], *Shenhui Deng*[2], *Yanlong Liu*[1*] *Binbin Cui*[1*]

[1]*Department of Colorectal Surgery, Harbin Medical University Cancer Hospital, Harbin Medial University, Harbin, China,* [2]*Department of Anesthesiology, The Fourth Affiliated Hospital of Harbin Medical University, Harbin, China*

OPEN ACCESS

Edited by:
Francesco Pappalardo, University of Catania, Italy

Reviewed by:
Na Li, Shandong Cancer Hospital, China
Cangang Zhang, Xi'an Jiaotong University, China

***Correspondence:**
Binbin Cui
cuibinbin@hrbmu.edu.cn
Yanlong Liu
liuyanlong@hrbmu.edu.cn

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Preclinical Cell and Gene Therapy, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *26 January 2022*
Accepted: *21 March 2022*
Published: *04 April 2022*

Citation:
Xia T, Guo J, Zhang B, Xue W, Deng S, Liu Y and Cui B (2022) A Novel Quantification System Combining iTRAQ Technology and Multi-Omics Assessment to Predict Prognosis and Immunotherapy Efficacy in Colon Cancer. Front. Bioeng. Biotechnol. 10:862619. doi: 10.3389/fbioe.2022.862619

Background: Colon cancer is one of the most common cancer types, although it has certain unique genetic features. This study aimed to develop a unique score for assessing prognosis and immunotherapy efficacy using integrated multi-omics analysis.

Methods: Isobaric tagging for relative and absolute quantification (iTRAQ) based proteomic analysis was used to screen differentially expressed proteins (DEP) between tumor and normal samples. DEP mRNA obtained from TCGA were clustered into different categories to show landscape-related prognosis and function. Following that, DEG was extracted from DEP mRNA, and the DEP-related score (DEPRS) was constructed to investigate the difference in immunotherapy prognosis and sensitivity. Finally, WCGNA, random forest, and artificial neural networks were used to screen for key genes. The prognostic value and protein level of these genes were validated.

Results: A total of 243 DEPs were identified through iTRAQ analysis, and the corresponding DEP mRNA was clustered into three. Following a series of tests, 1,577 DEGs were identified from overlapped DEP mRNA clusters and were classified into three gene clusters. The two types of clusters described above shared comparable characteristics in terms of prognosis and function. Then, it was established that a high DEPRS indicated a poor prognosis and DEPRS had significant associations with TMB, MSI status, and immunotherapeutic response. Finally, the key genes HART3 and FBLN2 were identified and were found to be implicated in immunotherapy and prognosis.

Conclusion: The development of a DEPRS based on multi-omics analysis will aid in improving our understanding of colon cancer and guiding a more effective immunotherapy strategy. DEPRS and key genes are used as biomarkers in the clinical evaluation of patients.

Keywords: colon cancer, left-sided, right-sided, multi-omics, prognosis, immunotherapy

INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer in the world, and it has a significant impact on both the global economy and patients' lives. Tumor metastasis is the cause of death in half of all patients with CRC (Siegel et al., 2019). Globally, 1,400,000 new cases and 700,000 CRC-related deaths were reported in 2018 (Bray et al., 2018). Colon cancer (CC) accounts for approximately 70% of all CRC (Ahmed, 2020), and CC can be classified into two distinct diseases: left-sided colon cancer (LCC) and right-sided colon cancer (RCC). Due to their unique embryonic origins, LCC and RCC exhibit a variety of clinical characteristics, including drug sensitivity. Additionally, in our previous study, we established the genetic distinction between LCC and RCC in terms of immunotherapy and prognosis (Guo et al., 2021). Based on difference in LCC and RCC, we want to develop a novel prognosis system that isn't effected by tumor site.

Tumorigenesis results from the interplay of multiple factors. With a better understanding of the etiology and pathogenesis of CC, as well as treatment strategies such as surgery and chemoradiotherapy, the survival rate of CC has significantly improved. Several limitations to single therapy and prognostic evaluation of CC, however, have contributed to the high mortality rate associated with advanced CC. At John Hopkins University, Le et al. (2015) discovered that mCRC patients with mismatch repair-deficient (dMMR) or microsatellite instability-high (MSI-H) can benefit from immune checkpoint inhibitors (ICIs). Thus, researchers have evaluated immunotherapy for CC (Le et al., 2015). Numerous studies have established that immunotherapy can benefit a significant number of dMMR/MSI-H patients. Microsatellite stability (MSS) tumors, in theory, have less immune cell infiltration and expression of immune-related genes than MSI tumors. Then, as a result of increased immunocyte infiltration, MSI tumors express neoantigen more easily, making them more sensitive to immunotherapy (Lin et al., 2020). Mismatch repair (MMR), on the other hand, can recognize and fix mutation errors, hence preventing the mutant protein from leading to a tumor. Defects in MMR-related genes result in impaired repair function and the accumulation of numerous altered genes throughout the DNA synthesis process. Tumor mutation burden (TMB) is enhanced by DNA impairment, and genome stability is affected to some extent. Previous research has shown that increased TMB may promote the translation of mutant proteins and stimulate the generation of neoantigens via major histocompatibility complex (MHC) binding (Sha et al., 2020). Meanwhile, elevated TMB can improve the immunogenicity of MSI tumors, resulting in a better immunotherapy outcome than before. Previous research also reports that 97% of MSI-H tumors have TMB ≥10 mutations/Mb (Chalmers et al., 2017). Additionally, a portion of MSS tumors contains a high concentration of TMB, which promotes the enrichment of activated CD4 and CD8 T cells, hence enhancing the tumor's response to ICIs (Ghorani et al., 2020). Numerous studies have demonstrated that TMB can be a reliable predictive index of antitumor response to ICIs (Jiang et al., 2021).

Progress in genomic technology, which began with the completion of the Human Genome Project in 2003, has been accelerated by the advent of transcriptome analysis, biochips (Kozal et al., 1996), and high-throughput sequencing (Reuter et al., 2015). Transcriptome sequencing is currently one of the most widely used high-throughput sequencing technologies, with next-generation sequencing being the most popular (Mosele et al., 2020). Following an integrated analysis of sequencing results, numerous tumor pathological mechanisms have been defined at the molecular level. These findings have aided the development of tumor treatment strategies, with the ultimate goal of translating laboratory findings to the clinic. Currently, the implementation of single transcriptomic analysis has some limitations. For example, there are numerous molecular stages involved in the translation of mRNA to proteins. Certain aberrations in this process may impair protein stability and disrupt the relationship between mRNA and protein levels. The combination of proteomics and transcriptome analysis can accurately uncover the biological mechanisms and clinical transformation of tumors. (Ross et al., 2004) established the concept of quantitative proteomics, a critical component of proteomics research. It is capable of identifying and quantifying all proteins expressed from a single genome or a mixture (Graves and Haystead, 2002). Isotope-based quantitative proteomics has been widely used to analyze specific tumor biomarkers. Numerous isotope-based quantitative proteomics technologies have been developed, including isotope-coded affinity tags (ICATs) (Gygi et al., 1999) and stable isotope labeling of amino acids in cell culture (Ong et al., 2002). Compared with these technologies, isobaric tags for relative and absolute quantification (iTRAQ), developed in 2004 by AB SCIEX, offer significant advantages, including better sensitivity and efficiency (Ross et al., 2004). iTRAQ technology allows for simultaneous labeling of up to eight samples, which may then be analyzed quantitatively using liquid chromatography tandem-mass spectrometry (LC-MS/MS). iTRAQ has been applied to different types of samples, and has yielded significant results, most notably in cancer research. iTRAQ-based studies have identified biomarkers in CRC (Bai et al., 2020), breast cancer (Jézéquel et al., 2019), bladder cancer (Zhang et al., 2017), and other cancers, establishing a solid foundation for further research and analysis of tumor pathogenesis using this technology.

In this study, we first used iTRAQ to analyze six paired samples between the CC and matched samples. After identifying the protein that differed significantly between tumor and normal samples, we obtained mRNA expression data from the TCGA and Gene Expression Omnibus (GEO) databases. To the best of our knowledge, there was rare research that apply differentially expressed protein-related RNA (DEP mRNA) to analyze CC. Surprisingly, there was a significant variation in prognosis and immunocytes infiltration between different DEP mRNA types. The DEP mRNA may then play a unique role in the immunologic microenvironment, influencing whether or not CC patients benefit from immunotherapy. To assess the characteristics of DEP mRNA, we extracted intersecting differentially expressed genes (DEG) from a public database and divided them into three gene clusters. Following that, the DEGs were used to develop a specific system referred to as the DEP-related score (DEPRS), which was used to grade them and compare the high- and low- score groups in terms of prognosis, TMB, MSI status, and immunotherapy sensitivity. This analysis presents a novel and effective strategy for predicting the

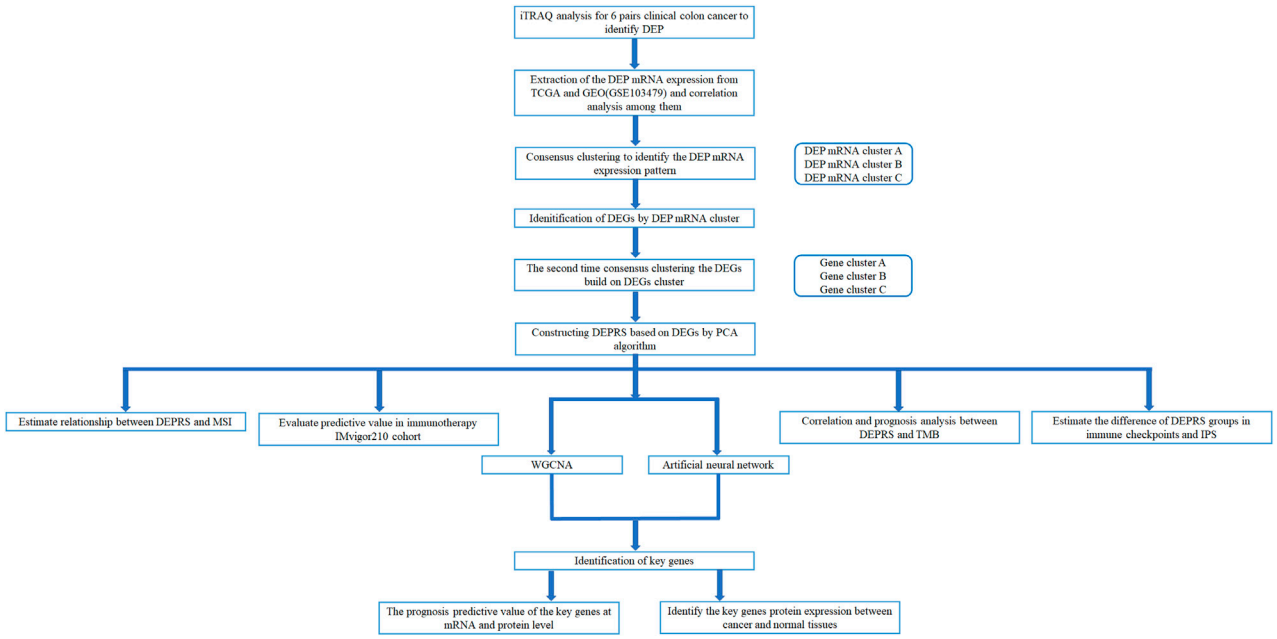


FIGURE 1 | Workflow of whole analysis.

prognosis of CC and evaluating the therapeutic effects of immunotherapy, without affect by LCC and RCC. Simultaneously, we identified important genes based on DEPRS using weighted gene co-expression network analysis (WGCNA), random forest, and artificial neural networks to better direct clinical work.

MATERIALS AND METHODS

The workflow of this whole study was offered in the **Figure 1**.

Collection of Clinical Samples

All human tumor samples and matched adjacent normal samples were collected from patients diagnosed with CC at the Harbin Medical University Cancer Hospital in October 2020. Patients were diagnosed with adenocarcinoma by endoscopy and had not undergone any preoperative chemotherapy or radiotherapy. Tissue samples were immediately stored in liquid nitrogen after resection. This study was approved by the Ethical Committees of Harbin Medical University Cancer Hospital, and patients signed an informed consent form.

Colon Cancer Transcriptome Data Download and Preprocessing

In this study, we downloaded 629 CC RNA-sequencing data from two high-throughput sequencing platforms; 473 cases from TCGA and 156 cases from GEO (GSE103479) (http://www.ncbi.nlm.nih.gov/geo/). The above data included information about each patient's somatic mutation, primary tumor site, clinical stage, and survival status. The tumor was characterized as RCC when it was located in the cecum, ascending colon, and hepatic flexure of the colon. The primary location of the tumor was the splenic flexure of the colon, descending colon, sigmoid and recto-sigmoid junction and were defined as LCC. After excluding patients with incomplete survival data, 444 samples were included in the study; TCGA 322 and GEO 122. We obtained the normalized matrix files from GEO for the microarray data. For data from TCGA, we downloaded the RNA-sequencing data (FPKM value) of gene expression, which was then converted into transcripts per kilobase million (TPM) values for combined analysis. To adjust the batch effect caused by non-biotechnology deviation, the "Combat" function of the R package "SVA" was used.

Unsupervised Clustering Based on Differentially Expressed Protein-Related mRNA

iTRAQ was used to examine six pairs of left- and right- colon cancer and matched normal samples, and 243 DEPs ($|\log2foldchange| > 0.5$, $p\text{-adj} < 0.05$) were identified (Detail iTRAQ method and protein information in **Supplementary Data**). Then, DEPs related mRNA expression levels were extracted from the integrated data set. The DEP mRNA was used to distinguish between tumor types. The hierarchical agglomerative cluster was used to cluster all tumor samples in the R package "ConsensusClusterPlus" (Wilkerson and Hayes, 2010). Cluster count and membership were identified through unsupervised analysis using stability evidence. To ensure the cluster's stability, the analysis was repeated 1,000 times.

Gene Set Variation Analysis

To compare the biological processes of DEP mRNA in distinct tumor types, we utilized the R package "GSVA" to perform enrichment difference analysis. GSVA uses a nonparametric and unsupervised method to estimate the variation pathway and enrichment of

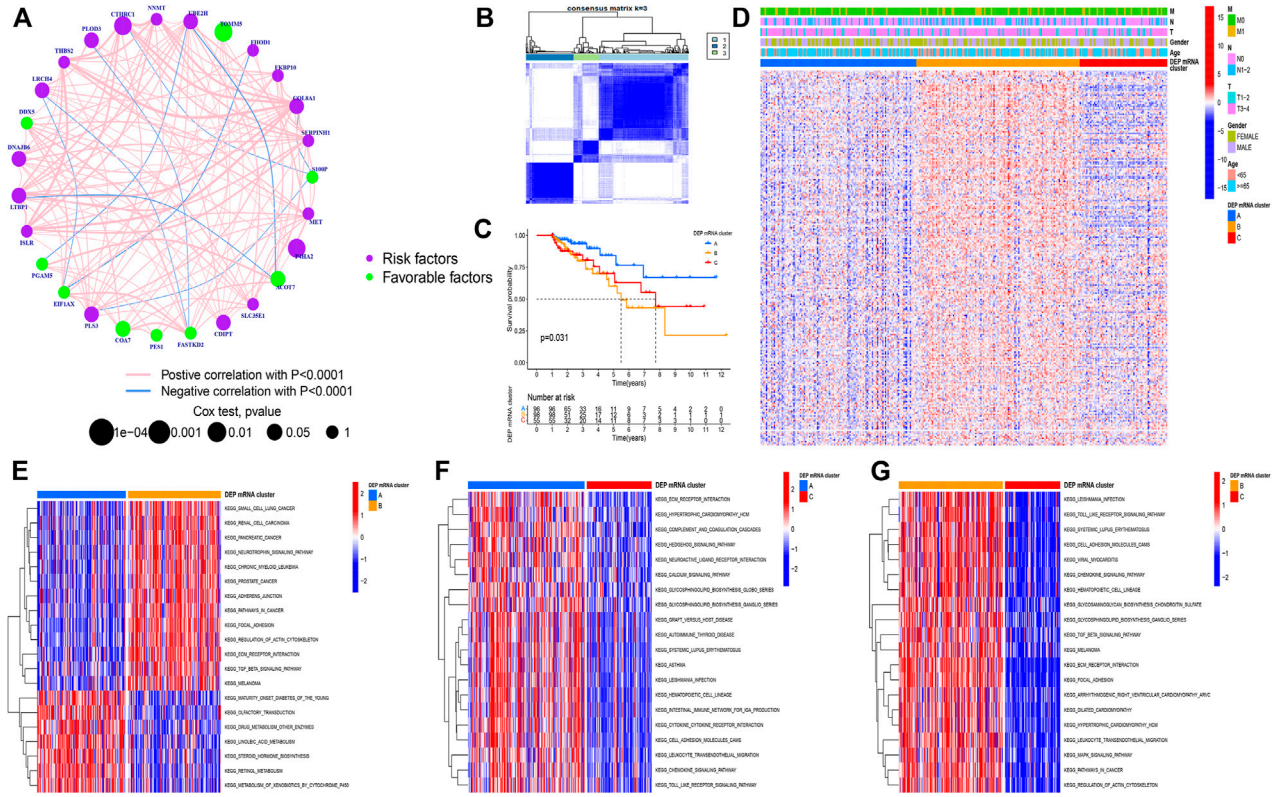


FIGURE 2 | (A) Correlation among DEP mRNA in CC. The circle size represents the effect of each regulator on the prognosis. The range of values calculated by Log-rank test was $p < 0.001$, $p < 0.01$, $p < 0.05$, and $p < 0.1$, respectively. The green dots in the circle represent favorable prognostic factors; Purple dots in the circle represent prognostic risk factors. The lines linking regulators indicate interactions, and their thickness shows correlation strength between genes. Negative correlation is marked in blue and positive correlation in red. **(B)** Consensus matrixes of all CC samples for appropriate k value (k = 3), displaying the clustering stability using 1,000 iterations of hierarchical clustering. All samples were clustered into three subtypes. **(C)** Survival analyses for the three DEP mRNA cluster based on 1,051 patients with CC from GEO cohorts including 96 cases in cluster-A, 98 cases in cluster-B, and 55 cases in cluster-C. Kaplan-Meier curves with Log-rank *p*-value = 0.031 showing significant survival difference among the three DEP mRNA patterns. The DEP mRNA cluster A had significantly better overall survival than the other two clusters. **(D)** A heat map showing the unsupervised clustering of DEP mRNA in all CC samples. Columns represent samples. A heatmap visualizing the clinical parameters. Red represents activation and blue represents inhibition. CC cohorts were used as sample annotations. **(E–G)** Results of GSVA enrichment analysis showing different tumor types in the three DEP mRNA patterns. Red represents activation and blue represents inhibition. **(E)** DEP mRNA cluster A vs DEP mRNA cluster B; **(F)** DEP mRNA cluster A vs DEP mRNA cluster C; **(G)** DEP mRNA cluster B vs DEP mRNA cluster C.

biological processes across several expression datasets. Following that, we download Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG)-related databases from the MSigDB database (http://softwar.broadinstiture.org/gsea/msigdb/) to finish the GSVA analysis. The heat map displays the significantly distinct pathway in the analysis results (p-adj < 0.05).

Evaluation and Difference of Immune Infiltration in Different Types of DEP mRNA

To estimate immunocyte infiltration of samples, the R package "GSVA" was used to analyze single-sample gene-set enrichment analysis (ssGSEA). We acquired information on immune cell marker gene expression from Charoentong's Charoentong et al., 2017 research and calculated the enrichment coefficients, which showed the relative abundance of immune cells in samples. Finally, we compared immune infiltration patterns across various DEP mRNA clusters.

Identification of Differentially Expressed Genes Between Differentially Expressed Proteins mRNA Clusters

R package "Limma" was used to identify DEGs (p-adj < 0.01) in different DEP mRNA clusters (Ritchie et al., 2015). Following that, numerous sets of DEGs were obtained, and the intersection was used to extract their expression in all samples for subsequence processing. The intersecting DEGs were analyzed using the R package "clusterProfiler" (Yu et al., 2012), which included GO and KEGG functional annotations analysis.

Construction of Differentially Expressed Proteins Related Scores

We developed an algorithm and defined it as DEPRS, for quantification of DEP mRNA-related types in LCC and RCC. The entire process was as follows: Firstly, the intersecting DEGs

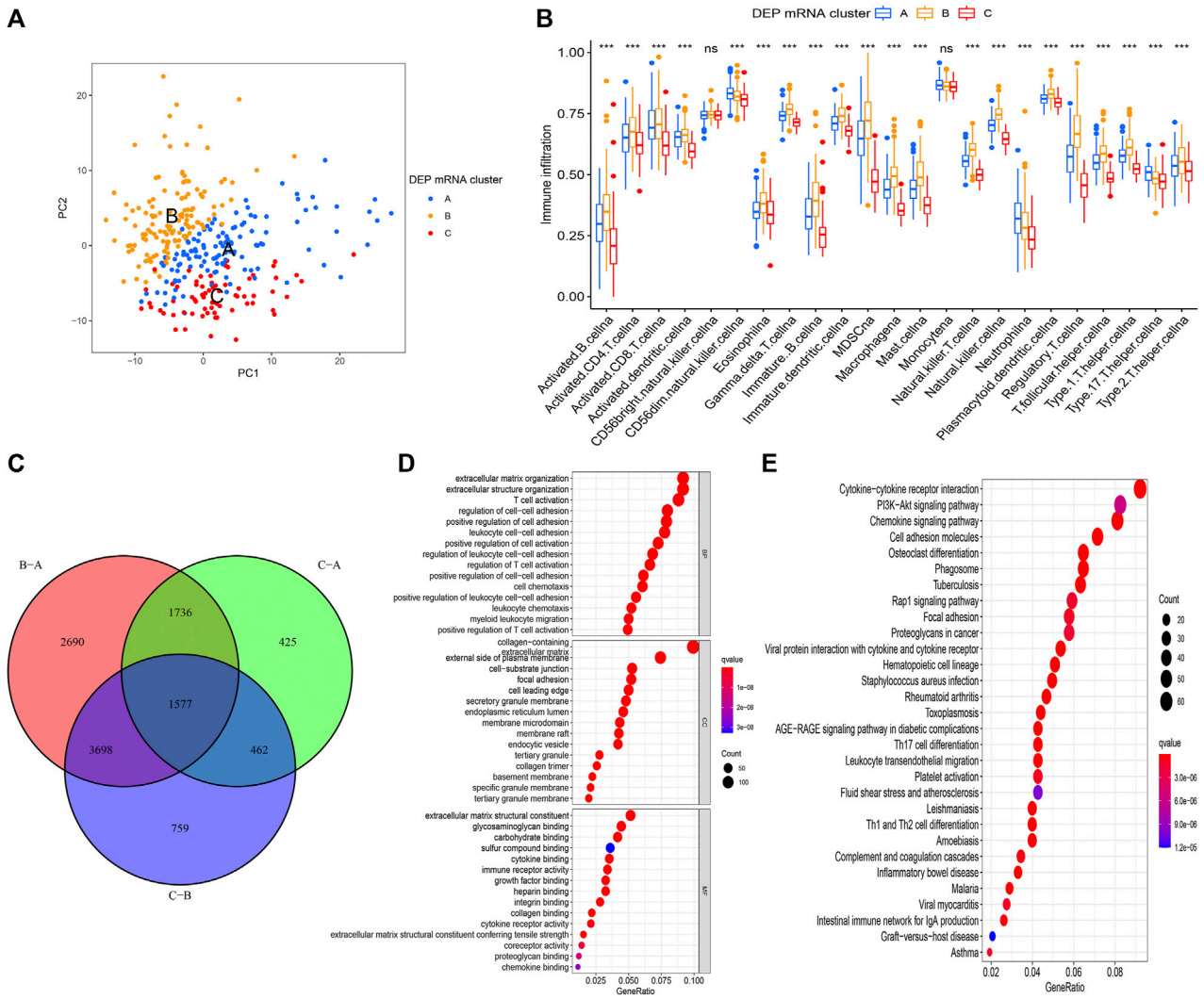


FIGURE 3 | (A) PCA for the transcriptome profiles of three DEP mRNA clusters showing significant differences in transcriptome between different clusters. **(B)** The abundance of each TME infiltrating cell in the three DEP mRNA clusters. The upper and lower ends of the boxes represent interquartile range of values. The lines in the boxes represent the median value, and the colored dots represent outliers. The asterisks represent the statistical *p*-value (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$). **(C)** The venn diagram showing the overlapping genes between the three clusters. **(D)** GO enrichment analysis of the overlapping gene signatures. **(E)** KEGG enrichment analysis of the overlapping gene signatures.

were subjected to univariate Cox proportional hazard regression analysis (COX) using the R package "glmnet" (Friedman et al., 2010). The gene that had a significant effect on prognosis was standardized. Based on these DEP mRNA-associated DEGs, the unsupervised clustering method was used to cluster all patients for further analysis. Meanwhile, we performed principal component analysis (PCA) to identify the key components of these genes to construct a DEP mRNA-related gene signature. Principal component 1 and principal component 2 were chosen as the signature scores. Finally, we used an approach comparable to the gene expression level index to calculate each patients' DEPRS: DEPRS= $\sum$PCA1i+$\sum$PCA2i (i is the expression of DEP mRNA-related DEGs). The best cut-off value for dividing high and low DEPRS groups for prognosis was obtained using the R package "maxstat" (Laska et al., 2012).

Prediction of Immunotherapy Sensitivities

Immunotherapy sensitivities were estimated in the high and low DEPRS groups from two perspectives: immune checkpoint-related genes immunophenoscores (IPS) and open-label immunotherapy cohort. Various immune-related genes, including effector cells, immune suppressor cells, MHC molecules, and immunoregulatory cytokines were used to determine immunogenicity. Machine-learning algorithms can accurately estimate and quantify immunogenicity. The IPS of CC in TCGA was downloaded from the TCIA database (https://tcia.at/). Then, we compared the immunophenoscore differences between high and low DEPRS groups in immunotherapy, to predict immunotherapy sensitivities. Simultaneously, a thorough search for gene expression profiles in publicly available immunotherapy cohort of metastatic urothelial tumors (IMvigor210: http://research-pub.gene.

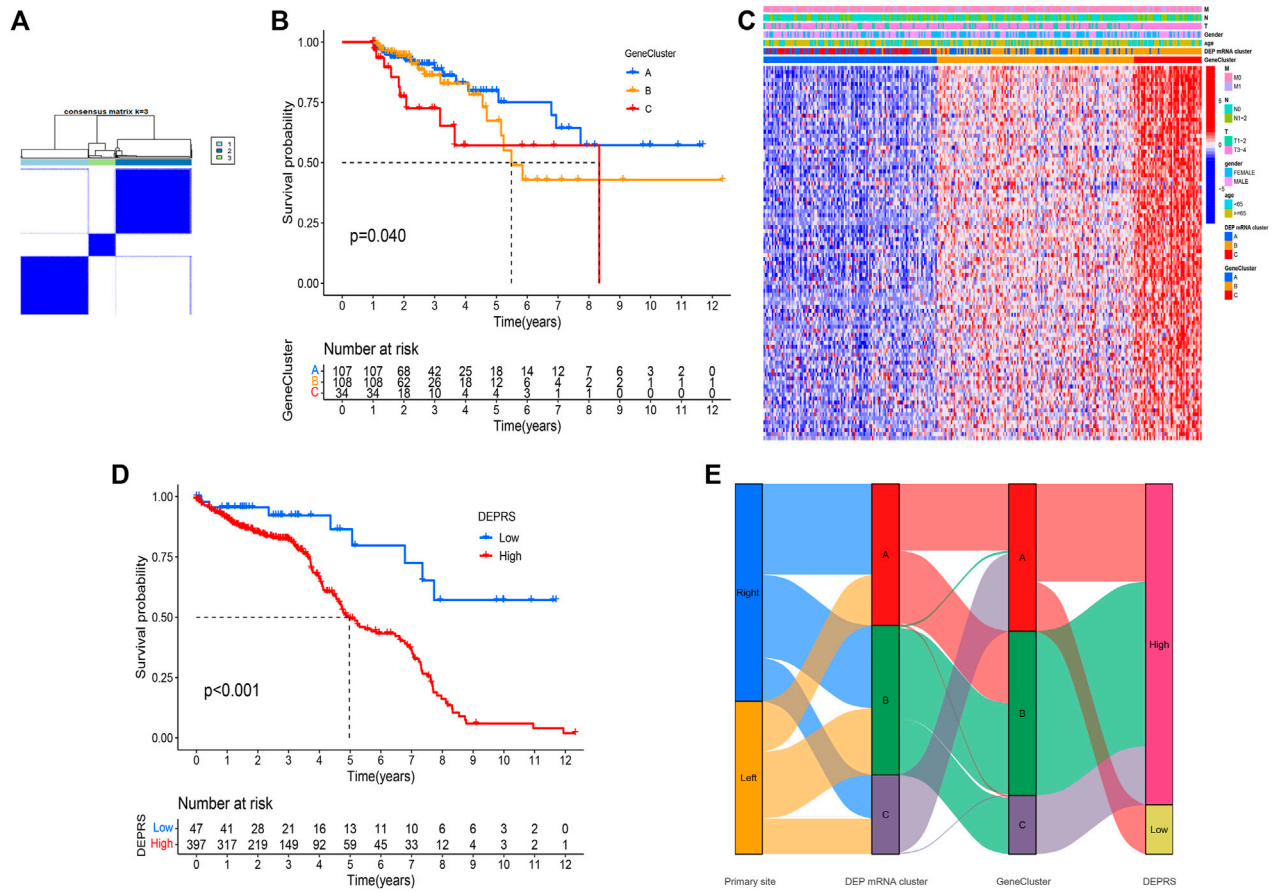


FIGURE 4 | (A) Consensus matrixes of TCGA-COAD cohorts for appropriate k value (k = 3), displaying the clustering stability using 1,000 iterations of hierarchical clustering. TCGA samples were clustered into three subtypes based on the DEGs among three DEP mRNA clusters. **(B)** Kaplan–Meier curves showing the overall survival across the gene clusters. The log rank test showed an overall $p = 0.040$. **(C)** A heat map showing the expression of DEGs in different gene clusters. Heat map colors indicate relative DEGs expression levels. **(D)** Kaplan–Meier curves showing the overall survival in high-DEPRS and low-DEPRS groups. The log rank test showed an overall $p < 0.001$. **(E)** The Sankey diagram displaying the distribution of patients with primary tumor sites, DEP mRNA clusters, gene clusters, and DEPRS.

com/IMvigor210CoreBiologies) Mariathasan et al., 2018 was performed. Data were pretreated using the R package "IMvigor210CireBiologies." The RNA-SEQ data was filtered and normalized using the R package "edgeR," and transformed using voom in the R package "limma." In addition, we downloaded and organized prognostic status and therapeutic effect data. Based on the above computation, the DEPRS for each sample in this cohort was calculated and divided into high and low score groups to compare the difference in therapeutic response between the two groups.

Identification of Key Genes

To screen and identify DEP-related prognostic key genes in LCC and RCC, two methods were used: weighted gene co-expression network analysis (WGCNA), random forest, and artificial neural network. First, we used the combination of DEP mRNA and DEPRS to run the WGCNA. To transform the adjacency matrix (AM) to a topological overlap matrix with DEGs, the appropriate power index was selected. The higher the value of the mean connectivity, the more the network conforms to the scale-free characteristics. According to the relationship between soft threshold (power) and mean connectivity, the minimum index when the R^2 of scale-free network reaches 0.8 was taken as the appropriate index. Then, a correlation between gene consensus modules and DEPRS was established, and gene significance (GS) was defined as the mediated p-value of each gene (GS = lgP) in a linear regression between gene expression and the scores. Subsequently, GS >0.6 genes in the module with the highest positive correlation coefficient with DEPRS were screened. Following that, the random Forest software package was used to classify all DEP mRNA. The parameter mtry (Optimal variable number of binary trees in the nodes) was set to 6, and the optimal number of trees included in the random forest was 49. After constructing the random forest module, the dimensional importance value (IV) was determined using the module's decreasing accuracy method (Gini coefficient method). Genes with a length of more than two were treated as special genes to construct subsequent modules. Unsupervised hierarchical cluster analysis was performed on these specific genes and a heat map was generated to demonstrate their classification effect. Following

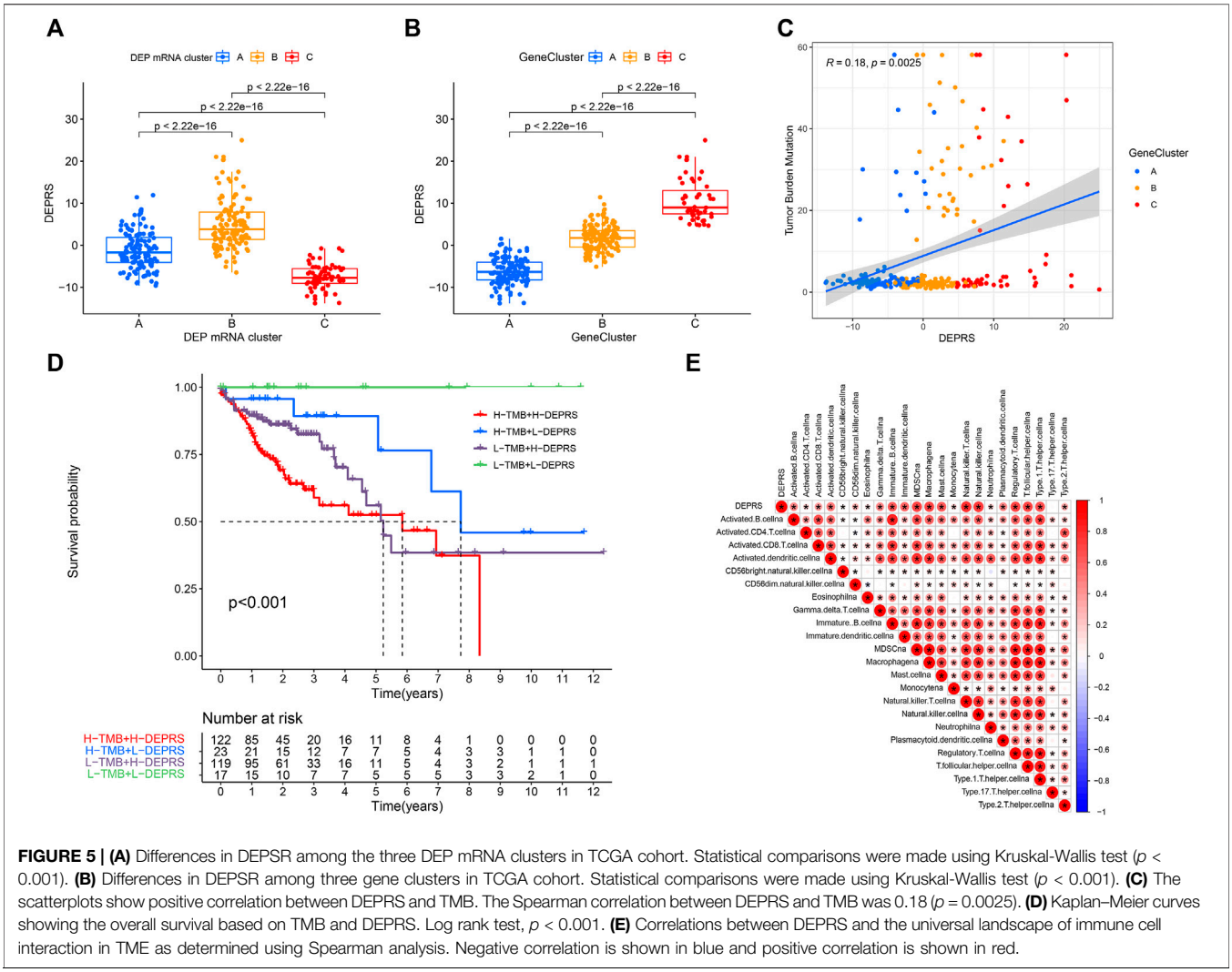


FIGURE 5 | (A) Differences in DEPSR among the three DEP mRNA clusters in TCGA cohort. Statistical comparisons were made using Kruskal-Wallis test (p < 0.001). **(B)** Differences in DEPSR among three gene clusters in TCGA cohort. Statistical comparisons were made using Kruskal-Wallis test (p < 0.001). **(C)** The scatterplots show positive correlation between DEPRS and TMB. The Spearman correlation between DEPRS and TMB was 0.18 (p = 0.0025). **(D)** Kaplan–Meier curves showing the overall survival based on TMB and DEPRS. Log rank test, p < 0.001. **(E)** Correlations between DEPRS and the universal landscape of immune cell interaction in TME as determined using Spearman analysis. Negative correlation is shown in blue and positive correlation is shown in red.

that, we classified these genes into high and low expression groups based on their median expression and categorized them as gene scores (low expression as 0, high expression as 1). On significant variables, the R package "Neuralnet" was used to construct an artificial neural network model. The model's hidden layer parameter was set to 5, and the outcomes of weight score multiplied by gene score were used to construct a classification model for HDEPRSG and LDEPRSG. The R package "pROC" was used to create the receiver operating characteristic curve (ROC), as well as calculate the area under the curve to confirm classification performance. Finally, the specific gene was obtained from an intersection of the particular gene and the screened gene. To repeatedly validate the important gene at the protein level, we downloaded the proteomics cohort in the TCGA COAD sample (including 29 normal samples and 64 tumor samples) from The Clinical Proteomic Tumor Analysis Consortium (CPTAC) (https://proteomics.cancer.gov/programs/cptac). Additionally, we investigated the prognostic significance and expression of key genes.

Statistical Analyses

All the statistical analyses were performed using R-version 4.0.5. The Wilcoxon test and Kruskal-Wallis test were used to compare two groups, and more than two groups respectively. The prognosis curve was constructed using the Kaplan-Meier plotter, and the log-rank test was used to determine whether there was a statistically significant difference in prognosis. Spearman correlation coefficient was used to measure the relationship between variables. The mutation gene status in different groups was demonstrated using the R package "maftool" (Mayakonda et al., 2018). p < 0.05 was considered statistically significant.

RESULTS

Differentially Expressed Proteins mRNA Landscape in Colon Cancer

The CC samples and matched normal samples were subjected to iTRAQ analysis. Then, 243 DEPs (|log2foldchange| > 0.5, p-adj <

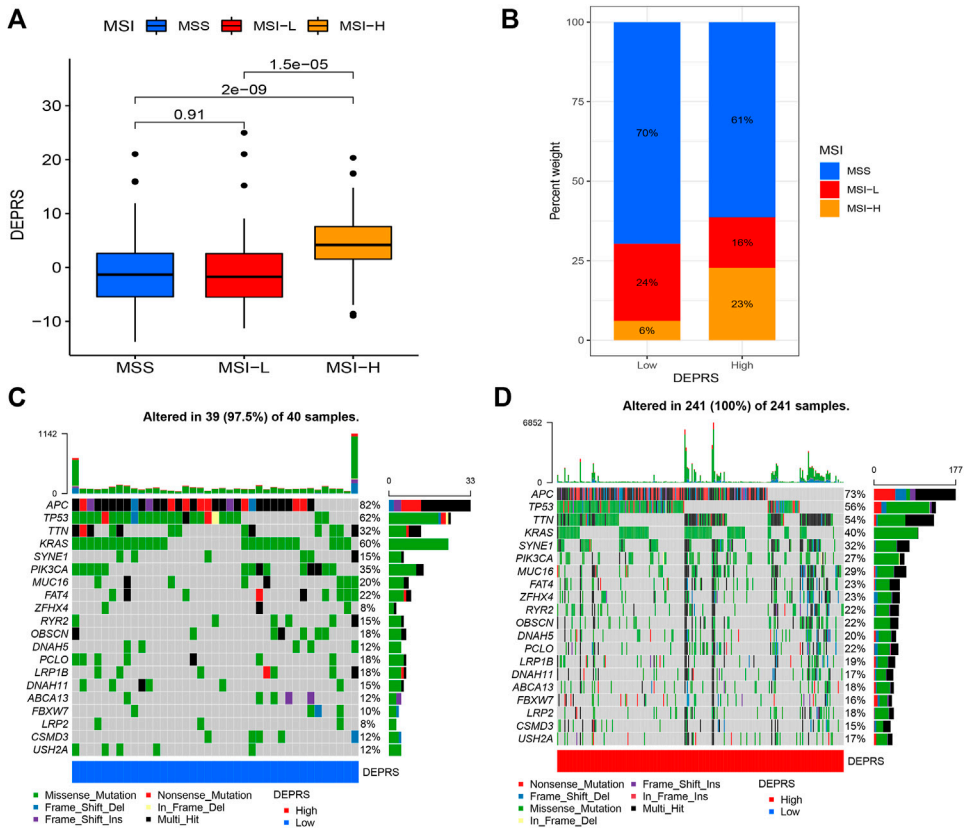


FIGURE 6 | (A) The profile of DEPSR among different MSI status. The Kruskal-Wallis test was adopted to make statistical comparisons between different MSI status ($p < 0.001$). **(B)** Proportion of patients with different MSI status in high DEPES and low DEPRS groups. The proportion of MSS and MSI-L patients in the low DEPRS group was significantly lower than that in the high DEPRS ($p < 0.05$). The groups were compared using the Kruskal-Wallis test. **(C,D)** A waterfall diagram showing the top 20 driver genes with the highest mutation frequency in low DEPRS **(C)** and high DEPRS **(D)** groups.

0.05) were identified and the corresponding mRNA data were extracted from the integrated dataset. The network was used to depict the comprehensive DEP mRNA landscape. Univariate COX analysis was used to investigate all of the DEP mRNAs, and the interaction between them was demonstrated (**Figure 2A**). As shown in the figure, there was an explicit negative association between the favorable factor and risk factor. Next, Following that, all CC samples were clustered using the R package "ConsensusClusterPlus" based on 243 DEP mRNA, and 3 clusters were identified (**Figure 2B**).

A Kaplan-Meier (K-M) curve was used to compare the prognosis of various DEP mRNA clusters. We found that DEP mRNA cluster B had a significantly worse prognosis than the other two clusters ($p = 0.031$) (**Figure 2C**). Then, DEP mRNA expression in the three clusters was shown using a heat map, and the expression of the 3 clusters was significantly different (**Figure 2D**). The KEGG-related GSVA was used to investigate the biological function of these DEP mRNA clusters. When cluster A was compared to cluster B, we observed that DEP mRNA cluster B was significantly more abundant in carcinogenic activation pathways, such as small cell lung cancer, renal cell carcinoma, adherens junction, and TGF-β pathway (**Figure 2E**). Then, GSVA analysis between clusters A and C revealed that cluster A enriched for a variety of pathways not previously associated with it, including ECM receptor interaction, chemokine pathway, and calcium signaling pathway (**Figure 2F**). In comparison to clustering C, DEP mRNA cluster B enriched for a variety of carcinogenic activation pathways including the MAPK signaling pathway, focal adhesion, etc. (**Figure 2G**). As a result of the aforementioned finding, DEP mRNA cluster B was found to have the worst prognosis among three clusters and an enrichment analysis with numerous carcinogenic pathways.

Differentially Expressed Genes Selected from Differentially Expressed Proteins mRNA Clusters

The PAC clustering method was used to confirm the DEP mRNA clustering result previously obtained. As a result of the PCA grouping, three distinct DEP mRNA groups were identified, indicating the consistency and accuracy of the test (**Figure 3A**). The DEP mRNA cluster B showed a poor prognosis and was associated with the cancer signaling pathway. The ssGSEA was used to evaluate each sample to determine the link between the three clusters and immune infiltration. Surprisingly, an investigation of immunological infiltration revealed that DEP mRNA cluster B was significantly abundant in immune cells such as MDSC, activated

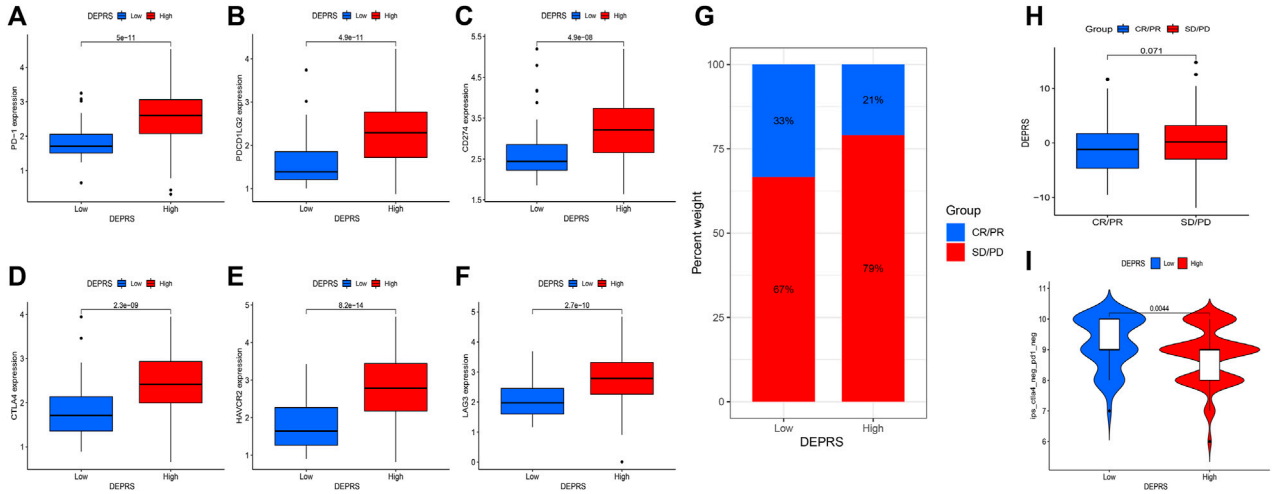


FIGURE 7 | (A–F) Comparison of immunosuppressive checkpoints expression between high DEPRS expression and low DEPRS expression groups. The expression of PD-1 **(A)**, PDCD1LG2 **(B)**, CD274 **(C)**, CTLA4 **(D)**, HAVCR2 **(E)** and LAG3 **(F)** was higher in high DEPRS than in the low DEPRS expression group (all $p < 0.05$). Statistical comparisons were done using the Wilcoxon test. **(G)** Proportion of patients with different treatment outcomes in high DEPRS and low DEPRS. The proportion of CR/PR patients in high DEPRS was lower than that in low DEPRS (21 vs 33%). **(H)** Comparison of DEPRS between different treatment outcome groups ($p = 0.071$). **(I)** The relationship between IPS and DEPRS groups in patients ($p = 0.0044$).

B cell, activated CD4 T cell, and natural killer cell (**Figure 3B**). However, when compared to clusters A and C, cluster B showed no advantage in terms of survival time. To further investigate the differences in the immunological microenvironment of DEP mRNA, the R package "Limma" was used to identify DEG between different DEP mRNA clusters. There were 1,577 overlapping DEGs (**Figure 3C**). Additionally, using the R package "clusterProfiler" the DEGs were subjected to GO and KEGG analysis. In GO enrichment, the overlapped DEGs were enriched in immune response (**Figure 3D**), while in KEGG they were enriched in inflammation-related pathways and carcinogenic pathways (**Figure 3E**).

Construction of Differentially Expressed Proteins Related Score

To conduct in-depth research on the various expression types of DEP that contribute to the differences in the immune microenvironment, the TCGA-COAD sample was re-divided into three gene clusters based on overlapping DEGs (**Figure 4A**). To determine the survival difference between the three gene clusters, a prognostic analysis was performed. Although the results presented on the survival curve weren't perfect, but the novel gene clusters were shown the significant difference of prognosis among different clusters ($p = 0.04$) (**Figure 4B**). Following that, the DEGs expression in distinct gene clusters and DEP mRNA clusters was visualized using the heat map (**Figure 4C**). Taking into account the complexity and variability of individual differential protein expression patterns and the subsequent identification of important genes, we developed a novel algorithm called DEPRS to quantify the DEP mRNA expression in individual patients. The optimal cut-off value was determined using the R package "maxstat," and patients were divided into high DEPRS and low DEPRS groups (HDEPRSG and LDEPRSG). As shown in the prognostic analysis, LDEPRSG had a better prognosis than HDEPRSG ($p < 0.001$) (**Figure 4D**). Following that, a Sankey diagram illustrating the distribution of patients with tumor sites, DEP mRNA cluster, gene clusters, and DEPRS was displayed (**Figure 4E**). As indicated in **Figure 4E**, CC was divided into LCC and RCC, which were further divided into three DEP mRNA clusters. Following that, the DEP mRNA clusters were stratified into three gene clusters. Surprisingly, samples in the DEP mRNA cluster B and gene cluster B were classified as having a high DEPRS score, indicating a poor prognosis. A portion of DEP mRNA cluster B occupied a section of gene cluster B, but partial samples in DEP mRNA cluster A were retained in gene cluster A distributed in low DEPRS, which was associated with a better prognosis. It was indirectly demonstrated that multiple clustering modes produced consistent results. Following the above result, the difference between LCC and RCC wasn't distinct that means the DEPRS model own the special advantage to ignore tumor site.

Differentially Expressed Proteins Related Score and Correlation Analysis of Somatic Mutation

To further illustrate the relationship between DEPRS and the preceding two cluster modes, we examined the correlation between the obtained clusters and DEPRS. There was a significant difference in DEPRS between various DEP mRNA clusters (**Figure 5A**). The DEP mRNA cluster B had had a much higher median DEPRES value than the other two clusters. Then, when compared to other clusters, gene cluster C had significantly increased DEPRS (**Figure 5B**). Because the DEG was shown to be enriched in immune-related pathways, we examined the correlation

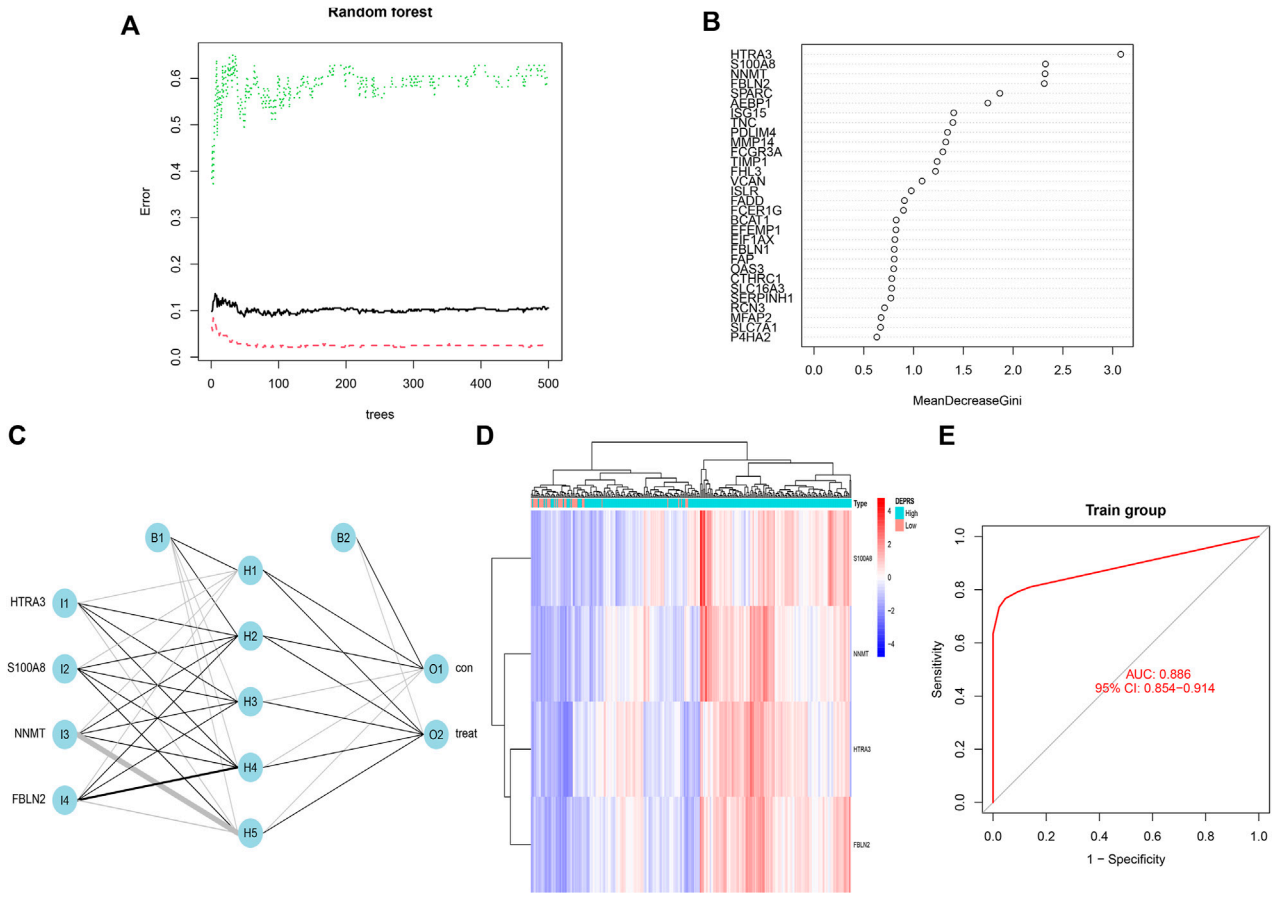


FIGURE 8 | (A) Relationship between the error rate and number of classification trees. **(B)** Gini coefficient method used to screen specific genes (IV >2). There were HTRA3, S100A8, NNMT, FBLN2. **(C)** Results of neural network visualization. **(D)** A heat map showing differences among the four genes between high DEPRS and low DEPES. Red color indicates positive expression and blue indicates negative expression. **(E)** The predictive value of the artificial neural network model based on four key genes in immunotherapy efficacy (AUC, 0.886, 95% CI: 0.854–0.914).

between TMB and DEPES. There was a significant correlation between TMB and DEPRS using association analysis (Coefficient: R = 0.018, $p = 0.0025$) (**Figure 5C**). With an increase in DEPRS, the distribution of gene clusters followed a significant rule. TMB showed an increasing tendency as DEPRS increased, indicating that patients with high TMB have a poor prognosis. According to the preceding data, gene cluster A had a low DEPRS and a low TMB, indicating a better prognosis. Additionally, we examined the effect of TMB and DEPRS integration on prognosis. All samples were stratified into high- and low- TMB subgroups. K-M curves were used to analyze the combined effect of TMB and DEPRS on prognosis. Within the same DEPRS status, the group with a high TMB had a worse outcome than the group with a low TMB. Nonetheless, TMB status did not affect the DEPRS predictive ability of prognosis, such that patients with high DEPRS invariably had a poor prognosis (**Figure 5D**). We also created a correlation map to visualize the DEPRS and immune cell interaction in TME, owing to the relationship between TMB and DEPRS. It is demonstrated unequivocally that the characteristics of DEPRS were highly correlated with high levels of immunocyte infiltration (**Figure 5E**). We aimed to determine the association between MSI status and DEPRS because it was critical for immunotherapy sensitivity. MSI-H differed significantly from MSS and MSI-L in this study, indicating that MSI-H was associated with a high DEPRS (**Figure 6A**). Consistent with this finding, 23% of MSI-H patients had high DEPRS compared with 6% with low DEPRS (**Figure 6B**). In addition, we compared the differences in somatic variation driver genes between individuals with high and low DEPRS. The top 20 driver genes exhibiting the highest mutation frequency were selected (**Figures 6C,D**). The mutation rate of the majority of driver genes was higher in the HDEPRSG than in the LDEPRSG.

Assessment of Differentially Expressed Proteins Related Score in Predicting Immunotherapy Efficacy

The purpose of this study was to compare immunotherapy sensitivity parameters in two DEPRS groups. For immune checkpoints, PD-1, PDCD1LG2, CD274, CTLA4, HAVCR2, and LAG3 were significantly expressed in HDEPRSG than LDEPRSG ($p < 0.05$) (**Figures 7A–F**). Therefore, the parameters indicated that

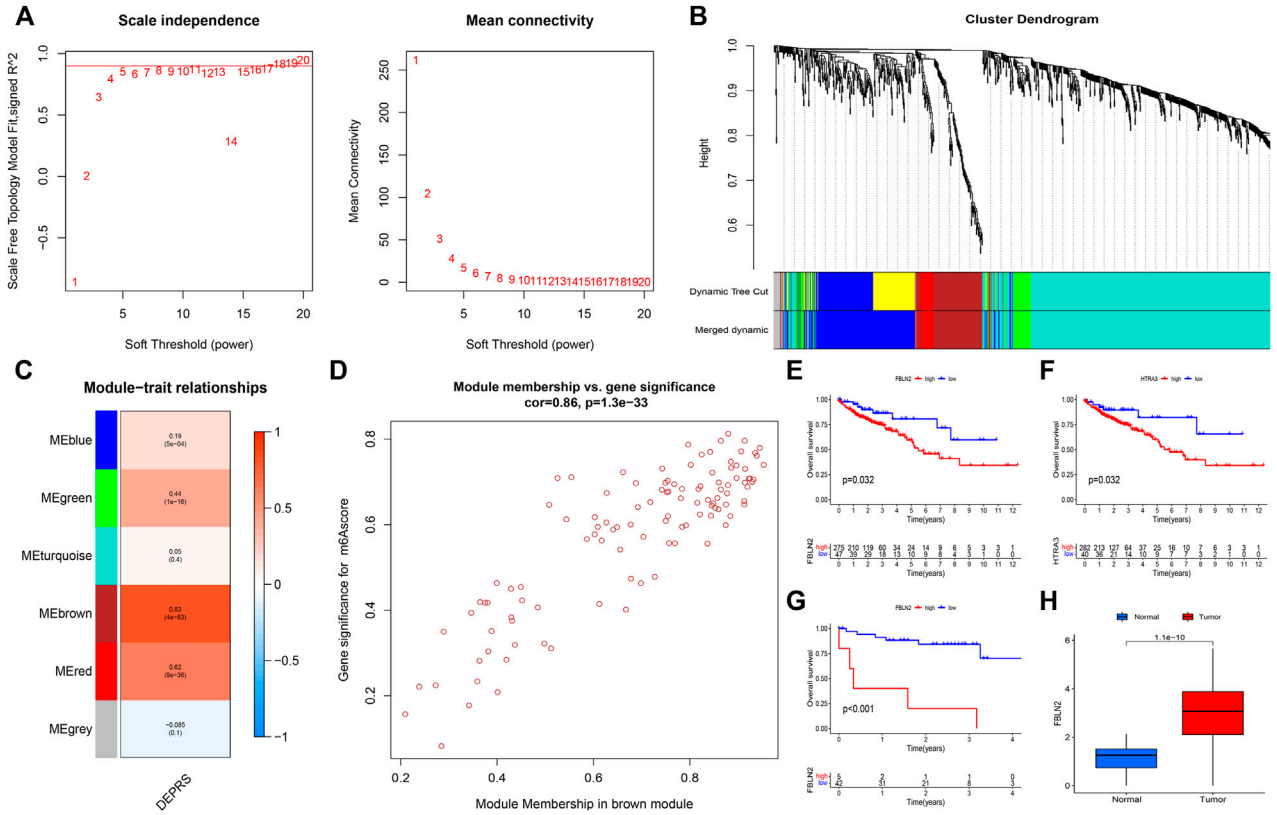


FIGURE 9 | (A) To achieve a scale-free co-expression network, the power index = 5 was chosen as the appropriate soft threshold. **(B)** The branches of the dendrogram correspond to six different gene modules. **(C)** The correlation between gene modules and DEPRS. Each cell contains corresponding correlation coefficient and *p*-value. **(D)** Significant positive correlation between module membership and gene significance (Correlation coefficient = 0.86, $p < 0.001$). **(E,F)** Kaplan-Meier curves showing the overall survival of the two key genes FBLN2 ($p = 0.032$) **(E)** and HTRA3 ($p = 0.032$) **(F)**. Red represents high expression, and blue represents low expression. **(G)** The difference in OS between high and low FBLN2 expression groups ($p < 0.001$). **(H)** The difference in FBLN2 expression between cancer tissues and normal tissues at protein level.

different groups responded differently to immunotherapy. Additionally, we demonstrated the stability of DEPRS as a predictor of immunotherapeutic efficacy in a public immunotherapy cohort. The relationship between DEPRS and treatment outcome was shown in **Figure 7G**. The results showed that the stable disease (SD)/progression disease (PD) proportion was significantly higher in the high DEPRS than in the low DEPRS group (79 vs 67%) and that SD/PD patients had a higher DEPRS than the complete response (CR)/partial response (PR) group (p =0.071) (**Figure 7H**). On the other hand, the relationship between IPS and DEPRS was used to estimate the predictive potential of DEPRS. The IPS, which measures immunogenicity, was a significant difference between DEPRS groups, with the IPS in the high DEPRS group being lower than the IPS in the low DEPRS ($p = 0.0044$) (**Figure 7I**). Based on these findings, we hypothesized that DEPRS may possess the ability to predict prognosis and immunotherapy efficacy.

Identification of Key Genes in Differentially Expressed Proteins Related Score

The random forest module was constructed to identify key genes based on DEPRS, and **Figure 8A** depicts the relationship between reference model error and the number of decision trees (**Figure 8A**). A total of 49 decision trees were selected since the error rate was the lowest and the most relatively stable. The IV > 2 genes (HTRA3, S100A8, NNMT, and FBLN2) were chosen for further analysis using the Gini coefficient method (**Figure 8B**). Unsupervised hierarchical cluster analysis was used to study these specific genes, and the resulting heat map was sued to illustrate the relationship between their expression and DEPRS status (**Figure 8C**). The expression of these four genes was significantly increased in the high DEPRS group. Next, these four genes were used to construct a neural network prediction model as illustrated in **Figure 8D**. The ROC curve was used to determine the sensitivity of the module, and the AUC value was 0.886 (**Figure 8E**). The results suggested that the neural network prediction model based on four specific key genes produced satisfactory results and that HDEPRSG was more accurately classified than LDEPRSG. On the other hand, a gene co-expression network based on DEPRS was constructed to identify the key DEP mRNA. By selecting number 5 as the appropriate soft threshold (**Figure 9A**), a scale-free co-expression network was constructed (**Figure 9B**), yielding six modules. The brown module had the highest correlation with DEPRA (coefficient = 0.83, $p < 0.001$) (**Figure 9C**). After interacting with the four

special genes identified by the neural network, two key genes HTRA3 and FBLN2 were obtained. The K-M curves were used to demonstrate that the high expression of two genes was associated with a poor prognosis (**Figures 9E,F**). At the protein level, high FBLN2 expression was associated with a poor prognosis ($p < 0.001$) (**Figure 9G**), and FBLN2 expression was significantly higher in tumor samples than in normal samples ($p < 0.001$) (**Figure 9H**).

DISCUSSION

Numerous studies have analyzed single transcriptome, resulting in a lack of stability for estimating prognosis and immunotherapy. In comparison, multi-omics integration research has several distinct advantages in the field of tumor research. To the best of our knowledge, integration of proteomics and transcriptomics research is rare in CC. In this study, iTRAQ analysis was used to identify DEPs, and then extract DEP mRNA expression from a publicly available database. The construction of the DEP mRNA clusters and gene clusters confirmed that various clusters had significant differences. Then, DEPRS were further established to confirm the stability of prognosis prediction and immunotherapy sensitivity for DEP mRNA-related DEGs and its evaluation ability didn't be affected by LCC and RCC. Finally, key genes that may offer potential clinical value for immunotherapy were identified.

Currently, immunotherapy is considered a novel treatment for cancer. Excepted PD-1, more and more immunotherapy markers were discovered and TMB is an emerging one. TMB refers to the total number of somatic mutations in the tumor genome that it can be used to show the ability of nonsynonymous mutation. It also indirectly reflect the function of tumor producing neoantigen. High levels of neoantigen can easily be recognized by the autoimmune system, and lead to stimulation of $CD8^+$ T cells to trigger immune response (Schumacher and Schreiber, 2015). Studies have shown that tumors with high TMB have high number of natural killer (NK) cells and T cells in their tumor microenvironment which indirectly show patients may receive good efficacy of immunotherapy. However, using TMB alone to evaluate the curative effect of immunotherapy is not highly effective. Therefore, researchers have attempted to find other indicators that can be integrated with TMB for the assessment of immunotherapy efficacy. In clear cell renal cell carcinoma, TNFSF14 was highly expressed in the high-TMB group, and the copy number of TNFSF14 was significantly correlated with classical immunocyte infiltration (Xu et al., 2020). A previous study showed that the mutation of ZFHX3 was significantly associated with high TMB and neoantigen load. In addition, the mutation of ZFHX3 showed a strong relationship with high-level T cell infiltration and immune-related genes (Zhang et al., 2021). In cutaneous melanoma, high TMB reflects good prognosis and low grade pathology, increased macrophage M1 and M2, and decreased ratio of Treg cell to memory B cells (Kang et al., 2020). In head and neck squamous cell carcinoma, low TMB level indicates better prognosis than high TMB level, and high immunocyte infiltration (Zhang et al., 2020). In a clinical trial in which patients with advanced melanoma received ipilimumab or tremilimumab treatment, patients in high TMB group (>100 nonsynonymous coding mutations) had longer overall survival (OS) (Snyder et al., 2014). A similar finding was obtained in a Chinese clinical trial on advanced gastric cancer. They found that TMB-H patients showed strong response to toripalimab, and patients who showed double positive TMB-H and PD-had better immune response and improved survival rate (Wang et al., 2019). The integration of TMB and a single gene may be an effective strategy for evaluating immune response to immunotherapy. In our study, we integrated integration proteome and transcriptome data to develop DERPS which can combined with TMB to facilitate evaluation of the prognosis of patients. When patients kept the same level of DEPRS, the high TMB level may remind patients own a bad outcome (**Figure 5D**).

As shown, immunotherapy effectively controlled MSI-H tumors, and TMB served as an important indicator of immunotherapy efficacy in MSI-H patients. The best cut-off point range of TMB was 37–41 mutations/Mb in MSI-H CRC patients (Schrock et al., 2019). TMB not only can be used as an auxiliary reference index for MSI tumor response to immunotherapy, but also as an indicator of MSI status. Study indicates that tumors lacking the mismatch repair protein duo MLH1/PMS2 always have a lower TMB than those tumor lacking a different protein heterodimer, MLH2/MSH6. Then, even tumor loss the same mismatch repair protein, the different origin of tumor may affect the TMB level (Salem et al., 2020). In our study, MSI-H tumors showed high DEPRS expression (**Figures 6A,B**), and expression of PD-1, CD174, CTLA4 and other parameter also shown high level in high DEPRS (**Figures 7A–F**), but tumors with high DEPRS expression had poor response to immunotherapy (**Figure 7G**). Borrowing this novel score system, we innovatively combined the two parameters to estimate response to immunotherapy. Previous studies mainly emphasize the heterogeneity between LCC and RCC, but the score system own its stability to assess prognosis and efficacy of immunotherapy which didn't affected by tumor site (**Figure 4E**). On one hand, further studies are needed to explore the detailed mechanism of TMB and MSI status in CC. On the other hand, our next step try to look for similar gene characters in significant different LCC and RCC, in order to offer accurate evaluation for patients.

To identify the key genes influencing the prognosis and response to immunotherapy, multiple algorithms were applied to screen genes based on DEPRS and DEP mRNA. Two key genes, Fibulin 2 (FBLN2) and HtrA serine peptidase 3 (HTRA3) were identified. HTRA3 was found to be a trimeric protein belonging to the homo-oligomeric serine proteases family. Functionally, HTRA3 was found to play an important role in mitochondrial homeostasis, cell death, and signal transduction (Clausen et al., 2002). A previous study reported that HTRA3 is a pro-apoptotic protein which also suppresses tumor formation. For example, in non-small cell lung cancer, over-expression of HTRA3 inhibited TGF-β1 to suppress tumor metastasis (Zhao et al., 2019). However, persistent expression of HTRA3 results in poor prognosis of CRC. Indeed, high expression of HTRA3 in CRC tumor stroma was associated with adverse outcomes such as high tumor budding (Forse et al., 2017). Evidence from studies has shown that high expression of HTRA3 was correlated with poor prognosis in oral squamous cell carcinoma (Moriya et al., 2015). In this study, we identified another key gene, *FBLN2*. This gene encodes fibulins which is a protein belonging to the extracellular matrix (ECM) glycoprotein family. All FBLN family protein contain epidermal growth factor (EGF)-like domains and a

C-terminal structure. Bases on this protein structure, they interact with other proteins to execute their functions (Gallagher et al., 2005). Many studies have shown that FBLN2 can bind to many ligands and function as a scaffold protein in the ECM (Yi et al., 2007). Given that FBLN2 functions in the ECM, downregulation of FBLN2 can promote the migration and invasion of tumor cells thereby causing damage to the basement membrane (Klingen et al., 2021). Furthermore, being a secretory metalloproteinase, ADAMTS-12 participates in tissue remodeling and cell migration. It interacts with FBLN2 to suppress the invasiveness of breast cancer cells. Interestingly, ADAMTS-12 was found to promote tumor development in breast cancer cells lacking FBLN2 by regulating metalloproteinase (Fontanil et al., 2014). In contrast, another study found that FBLN2 promoted tumor growth by interacting with activated β integrin receptor in CRC (Vaes et al., 2021). Consistent with the above finding, high expression level of these two genes (*FBLN2* and *HTRA3*) was linked to worse prognosis in colon cancer (**Figures 9E,F**), and FBLN2 was found to be significantly differentially expressed in CC (**Figure 9H**). Further research is needed to clarify their roles during the development of CC.

In summary, this study integrated omics tools ranging from proteomics to transcriptomics to estimate the prognosis and response to immunotherapy. HART3 and FBLN2 were found to be the key genes that can offer predictive role for immunotherapy in CC. Our study provides a reliable method for establishing a quantitative model that can be adopted to explore the pathogenesis of CC. This model may offer its own value to help clinical practice and not affected by tumor site. Then, the application of the model in clinical practice requires further investigation.

DATA AVAILABILITY STATEMENT

The authors acknowledge that the data presented in this study must be deposited and made publicly available in an acceptable repository, prior to publication. Frontiers cannot accept an article that does not adhere to our open data policies.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Ethical Committees of Harbin Medical University Cancer Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

BC and YL designed the study. TX and JG drafted the manuscript. TX, JG, and SD collected, analyzed and interpreted the data. BZ and WX drew the figures. BC, YL, BZ, and WX helped with the final revision of the article. All authors have read and approved the final manuscript.

FUNDING

This work was supported by Nn10 Program of Harbin Medical University Cancer Hospital (Nn102017-02), the Post-doctoral Scientific Research Developmental Fund of Heilongjiang (LBH-Q18085), the Youth Science Foundation of Heilongjiang (JJ2018QN0724), the Natural Science Foundation of Heilongjiang Province (QC2018111) and Haiyan Foundation of Harbin Medical University Cancer Hospital (JJZD 2022-17).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2022.862619/full#supplementary-material

REFERENCES

Ahmed, M. (2020). Colon Cancer: A Clinician's Perspective in 2019. *Gastroenterol. Res.* 13, 1–10. doi:10.14740/gr1239

Bai, Y., Wang, J., Gao, Z., and Dai, E. (2020). Identification and Verification of Two Novel Differentially Expressed Proteins from Non-Neoplastic Mucosa and Colorectal Carcinoma via iTRAQ Combined with Liquid Chromatography-Mass Spectrometry. *Pathol. Oncol. Res.* 26, 967–976. doi:10.1007/s12253-019-00651-y

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* 68, 394–424. doi:10.3322/caac.21492

Chalmers, Z. R., Connelly, C. F., Fabrizio, D., Gay, L., Ali, S. M., Ennis, R., et al. (2017). Analysis of 100,000 Human Cancer Genomes Reveals the Landscape of Tumor Mutational burden. *Genome Med.* 9, 34. doi:10.1186/s13073-017-0424-2

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

Clausen, T., Southan, C., and Ehrmann, M. (2002). The HtrA Family of Proteases: Implications for Protein Composition and Cell Fate. *Mol. Cell* 10, 443–455. doi:10.1016/s1097-2765(02)00658-5

Fontanil, T., Rúa, S., Llamazares, M., Moncada-Pazos, A., Quirós, P. M., García-Suárez, O., et al. (2014). Interaction Between the ADAMTS-12 Metalloprotease and Fibulin-2 Induces Tumor-Suppressive Effects in Breast Cancer Cells. *Oncotarget* 5, 1253–1264. doi:10.18632/oncotarget.1690

Forse, C. L., Rahimi, M., Diamandis, E. P., Assarzadegan, N., Dawson, H., Grin, A., et al. (2017). HtrA3 Stromal Expression is Correlated with Tumor Budding in Stage II Colorectal Cancer. *Exp. Mol. Pathol.* 103, 94–100. doi:10.1016/j.yexmp.2017.07.002

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

Gallagher, W. M., Currid, C. A., and Whelan, L. C. (2005). Fibulins and Cancer: Friend or Foe? *Trends Mol. Med.* 11, 336–340. doi:10.1016/j.molmed.2005.06.001

Ghorani, E., Reading, J. L., Reading, J. L., Henry, J. Y., Massy, M. R. d., Rosenthal, R., et al. (2020). The T Cell Differentiation Landscape Is Shaped by Tumour Mutations in Lung Cancer. *Nat. Cancer* 1, 546–561. doi:10.1038/s43018-020-0066-y

Graves, P. R., and Haystead, T. A. J. (2002). Molecular Biologist's Guide to Proteomics. *Microbiol. Mol. Biol. Rev.* 66, 39–63. doi:10.1128/mmbr.66.1.39-63.2002

Guo, J.-N., Xia, T.-Y., Deng, S.-H., Xue, W.-N., Cui, B.-B., and Liu, Y.-L. (2021). Prognostic Immunity and Therapeutic Sensitivity Analyses Based on Differential Genomic Instability-Associated LncRNAs in Left- and Right-Sided Colon Adenocarcinoma. *Front. Mol. Biosci.* 8, 668888. doi:10.3389/fmolb.2021.668888

Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999). Quantitative Analysis of Complex Protein Mixtures Using Isotope-Coded Affinity Tags. *Nat. Biotechnol.* 17, 994–999. doi:10.1038/13690

Jézéquel, P., Guette, C., Lasla, H., Gouraud, W., Boissard, A., Guérin-Charbonnel, C., et al. (2019). iTRAQ-Based Quantitative Proteomic Analysis Strengthens Transcriptomic Subtyping of Triple-Negative Breast Cancer Tumors. *Proteomics* 19, e1800484. doi:10.1002/pmic.201800484

Jiang, A.-M., Ren, M.-D., Liu, N., Gao, H., Wang, J.-J., Zheng, X.-Q., et al. (2021). Tumor Mutation Burden, Immune Cell Infiltration, and Construction of Immune-Related Genes Prognostic Model in Head and Neck Cancer. *Int. J. Med. Sci.* 18, 226–238. doi:10.7150/ijms.51064

Kang, K., Xie, F., Mao, J., Bai, Y., and Wang, X. (2020). Significance of Tumor Mutation Burden in Immune Infiltration and Prognosis in Cutaneous Melanoma. *Front. Oncol.* 10, 573141. doi:10.3389/fonc.2020.573141

Klingen, T. A., Chen, Y., Aas, H., Wik, E., and Akslen, L. A. (2021). Fibulin-2 Expression Associates with Vascular Invasion and Patient Survival in Breast Cancer. *PLoS One* 16, e0249767. doi:10.1371/journal.pone.0249767

Kozal, M. J., Shah, N., Shen, N., Yang, R., Fucini, R., Merigan, T. C., et al. (1996). Extensive Polymorphisms Observed in HIV-1 Clade B Protease Gene Using High-Density Oligonucleotide Arrays. *Nat. Med.* 2, 753–759. doi:10.1038/nm0796-753

Laska, E., Meisner, M., and Wanderling, J. (2012). A Maximally Selected Test of Symmetry About Zero. *Statist. Med.* 31, 3178–3191. doi:10.1002/sim.5384

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

Lin, A., Zhang, J., and Luo, P. (2020). Crosstalk Between the MSI Status and Tumor Microenvironment in Colorectal Cancer. *Front. Immunol.* 11, 2039. doi:10.3389/fimmu.2020.02039

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

Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: Efficient and Comprehensive Analysis of Somatic Variants in Cancer. *Genome Res.* 28, 1747–1756. doi:10.1101/gr.239244.118

Moriya, Y., Uzawa, N., Morita, T., Mogushi, K., Miyaguchi, K., Takahashi, K.-I., et al. (2015). The High-Temperature Requirement Factor A3 (HtrA3) is Associated with Acquisition of the Invasive Phenotype in Oral Squamous Cell Carcinoma Cells. *Oral Oncol.* 51, 84–89. doi:10.1016/j.oraloncology.2014.10.001

Mosele, F., Remon, J., Mateo, J., Westphalen, C. B., Barlesi, F., Lolkema, M. P., et al. (2020). Recommendations for the Use of Next-Generation Sequencing (NGS) for Patients with Metastatic Cancers: A Report from the ESMO Precision Medicine Working Group. *Ann. Oncol.* 31, 1491–1505. doi:10.1016/j.annonc.2020.07.014

Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., et al. (2002). Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell Proteom.* 1, 376–386. doi:10.1074/mcp.m200025-mcp200

Reuter, J. A., Spacek, D. V., and Snyder, M. P. (2015). High-Throughput Sequencing Technologies. *Mol. Cell* 58, 586–597. doi:10.1016/j.molcel.2015.05.004

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

Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., et al. (2004). Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-Reactive Isobaric Tagging Reagents. *Mol. Cell Proteom.* 3, 1154–1169. doi:10.1074/mcp.m400129-mcp200

Salem, M. E., Bodor, J. N., Puccini, A., Xiu, J., Goldberg, R. M., Grothey, A., et al. (2020). Relationship Between MLH1, PMS2, MSH2 and MSH6 Gene-Specific Alterations and Tumor Mutational Burden in 1057 Microsatellite Instability-High Solid Tumors. *Int. J. Cancer* 147, 2948–2956. doi:10.1002/ijc.33115

Schrock, A. B., Ouyang, C., Sandhu, J., Sokol, E., Jin, D., Ross, J. S., et al. (2019). Tumor Mutational Burden is Predictive of Response to Immune Checkpoint Inhibitors in MSI-High Metastatic Colorectal Cancer. *Ann. Oncol.* 30, 1096–1103. doi:10.1093/annonc/mdz134

Schumacher, T. N., and Schreiber, R. D. (2015). Neoantigens in Cancer Immunotherapy. *Science* 348, 69–74. doi:10.1126/science.aaa4971

Sha, D., Jin, Z., Budczies, J., Kluck, K., Stenzinger, A., and Sinicrope, F. A. (2020). Tumor Mutational Burden as a Predictive Biomarker in Solid Tumors. *Cancer Discov.* 10, 1808–1825. doi:10.1158/2159-8290.cd-20-0522

Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer Statistics, 2019. *CA Cancer J. Clin.* 69, 7–34. doi:10.3322/caac.21551

Snyder, A., Makarov, V., Merghoub, T., Yuan, J., Zaretsky, J. M., Desrichard, A., et al. (2014). Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma. *N. Engl. J. Med.* 371, 2189–2199. doi:10.1056/nejmoa1406498

Vaes, N., Schonkeren, S. L., Rademakers, G., Holland, A. M., Koch, A., Gijbels, M. J., et al. (2021). Loss of Enteric Neuronal Ndrg4 Promotes Colorectal Cancer via Increased Release of Nid1 and Fbln2. *EMBO Rep.* 22, e51913. doi:10.15252/embr.202051913

Wang, F., Wei, X. L., Wang, F. H., Xu, N., Shen, L., Dai, G. H., et al. (2019). Safety, Efficacy and Tumor Mutational burden as a Biomarker of Overall Survival Benefit in Chemo-Refractory Gastric Cancer Treated with Toripalimab, a PD-1 Antibody in Phase Ib/II Clinical Trial NCT02915432. *Ann. Oncol.* 30, 1479–1486. doi:10.1093/annonc/mdz197

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

Xu, F., Guan, Y., Zhang, P., Xue, L., Yang, X., Gao, K., et al. (2020). The Impact of TNFSF14 on Prognosis and Immune Microenvironment in clear Cell Renal Cell Carcinoma. *Genes Genom.* 42, 1055–1066. doi:10.1007/s13258-020-00974-0

Yi, C.-H., Smith, D. J., West, W. W., and Hollingsworth, M. A. (2007). Loss of Fibulin-2 Expression is Associated with Breast Cancer Progression. *Am. J. Pathol.* 170, 1535–1545. doi:10.2353/ajpath.2007.060478

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

Zhang, J., Zhou, N., Lin, A., Luo, P., Chen, X., Deng, H., et al. (2021). ZFHX3 Mutation as a Protective Biomarker for Immune Checkpoint Blockade in Non-Small Cell Lung Cancer. *Cancer Immunol. Immunother.* 70, 137–151. doi:10.1007/s00262-020-02668-8

Zhang, L., Li, B., Peng, Y., Wu, F., Li, Q., Lin, Z., et al. (2020). The Prognostic Value of TMB and the Relationship between TMB and Immune Infiltration in Head and Neck Squamous Cell Carcinoma: A Gene Expression-Based Study. *Oral Oncol.* 110, 104943. doi:10.1016/j.oraloncology.2020.104943

Zhang, Q., Huang, S., Luo, H., Zhao, X., Wu, G., and Wu, D. (2017). Eight-Plex iTRAQ Labeling and Quantitative Proteomic Analysis for Human Bladder Cancer. *Am. J. Cancer Res.* 7, 935–945.

Zhao, J., Feng, M., Liu, D., Liu, H., Shi, M., Zhang, J., et al. (2019). Antagonism Between HTRA3 and TGFβ1 Contributes to Metastasis in Non-Small Cell Lung Cancer. *Cancer Res.* 79, 2853–2864. doi:10.1158/0008-5472.can-18-2507

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 Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 13 April 2022
doi: 10.3389/fbioe.2022.850669

Identification and Validation of a Prognostic Immune-Related Gene Signature in Esophageal Squamous Cell Carcinoma

Kai Xiong[1†], *Ziyou Tao*[2†], *Zeyang Zhang*[1†], *Jianyao Wang*[1] *and Peng Zhang*[1*]

[1]*Department of Cardiovascular Thoracic Surgery, Tianjin Medical University General Hospital, Tianjin, China,* [2]*Department of Thoracic Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China*

Esophageal carcinoma (EC) is a common malignant cancer worldwide. Esophageal squamous cell carcinoma (ESCC), the main type of EC, is difficult to treat because of the widespread morbidity, high fatality rates, and low quality of life caused by postoperative complications and no specific molecular target. In this study, we screened genes to establish a prognostic model for ESCC. The transcriptome expression profiles of 81 ESCC tissues and 340 normal esophageal mucosal epithelium tissues were obtained from The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) cohorts. The transcriptome expression datasets of 19 esophageal squamous carcinoma cell lines were downloaded from Cancer Cell Line Encyclopedia (CCLE). The R software Limma package was used to identify 6,231 differentially expressed genes and 647 differentially expressed immune-related genes between normal and ESCC tissues. Gene functional analysis was performed using Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). Weighted gene co-expression network analysis (WGCNA) was used to screen out 18 immune-related prognostic genes. We then established the prognostic and risk signature using these genes, and the patients were divided into low-risk and high-risk groups. Compared with high-risk group patients, the low-risk group patients had longer overall survival. M1 macrophages and resting dendritic cells were differentially distributed between the low-risk and high-risk groups and were related to patient survival. We also examined the functional immune cell and immune molecule levels in low-risk and high-risk group patients, with significant differences in the tumor microenvironment between the two groups. To further verify the accuracy of the prognostic risk model, we performed area under the ROC curve (AUC) analysis. The AUC value was 0.931 for the prognostic risk, which was better than the microsatellite instability (MSI) and Tumor Immune Dysfunction and Exclusion (TIDE) scores. In conclusion, we found 18 immune-related prognostic genes related to the occurrence of ESCC and established a prognostic model for predicting disease severity.

Keywords: esophageal squamous cell carcinoma, TCGA, prognostic model, WGCNA, gene signature

OPEN ACCESS

Edited by:
Ping Zhang,
Griffith University, Australia

Reviewed by:
Ting-Ting Liu,
Beijing Red Cross Blood Center, China
Xibo Gao,
Tianjin Children's Hospital, China

***Correspondence:**
Peng Zhang
pengzhang01@tmu.edu.cn

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

Specialty section:
This article was submitted to Preclinical Cell and Gene Therapy, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *08 January 2022*
Accepted: *24 March 2022*
Published: *13 April 2022*

Citation:
Xiong K, Tao Z, Zhang Z, Wang J and Zhang P (2022) Identification and Validation of a Prognostic Immune-Related Gene Signature in Esophageal Squamous Cell Carcinoma. Front. Bioeng. Biotechnol. 10:850669. doi: 10.3389/fbioe.2022.850669

INTRODUCTION

Esophageal carcinoma (EC) is a common malignant cancer worldwide. In 2018, about half of all new or fatal global cases of EC originated in China. Esophageal squamous cell carcinoma (ESCC), the main type of EC, is more difficult to treat because it is frequently diagnosed at a late stage (Sung et al., 2021; Wanqing Chen et al., 2021). At present, radiotherapy and chemotherapy are still the main treatments for unresectable advanced EC. Neoadjuvant and molecular targeted therapies are also commonly combined with traditional treatments to extend patient survival (Kakeji et al., 2021). In addition, with the great promise of immune checkpoint inhibitors, such as PD-1, PD-L1, CTLA-4, researchers have started using immunotherapy for advanced EC (Sadanand, 2021). In the KEYNOTE-181 clinical trial study, pembrolizumab had a higher total effective rate and better safety profile compared with chemotherapy. The above results indicated that pembrolizumab should be considered as the new standard of second-line treatment for patients with Combined Positive Score (CPS) ≥10 (Kojima et al., 2020). However, the response rate to EC immunotherapy treatments must urgently be improved to reduce drug resistance and increase patient survival time. The immunotherapy for EC patients is obstructed and long.

For EC patients, especially those with advanced disease, it is of practical significance to better distinguish patients who will potentially respond to immunotherapy. The immune-related gene prognostic index is a recently emerging algorithm that calculates the score of each patient based on the expression levels of related genes. It can be used to predict the patient's immunotherapy response and prognostic risk index (Yue Chen et al., 2021). In this study, we obtained tumor-related differentially expressed genes by comparing gene expression levels in ESCC tissues and normal esophageal mucosa tissues. We obtained immune-related differential genes by comparing them with immune genes databases. The hub immune-related genes, obtained by weighted gene co-expression network analysis (WGCNA), were used to establish the prognostic and risk signature by immune-related gene prognostic index. Finally, we verified the changes in the immune status of the low-risk and high-risk patient groups and the accuracy of predicting the prognosis and immunotherapy response of ESCC patients.

MATERIALS AND METHODS

Data From the Cancer Genome Atlas, Genotype-Tissue Expression Cohorts and Cancer Cell Line Encyclopedia

The transcriptome expression profiles of 81 ESCC tissues were downloaded from TCGA and the corresponding clinical information (https://portal.gdc.cancer.gov/). The 340 normal esophageal mucosal epithelium samples were obtained from GTEx cohorts (https://www.gtexportal.org/home/index.html). The transcriptome expression datasets of 19 esophageal squamous carcinoma cell lines (TE1, TE4, TE5, TE6, TE8, TE9, TE10, TE11, TE14, TE15, KYSE30, KYSE70, KYSE140, KYSE150, KYSE180, KYSE270, KYSE410, KYSE510, and KYSE520) were downloaded from CCLE (http://www.broadinstitute.org/ccle/home). Using R software to integrate TCGA and GTEx data, 274 normal esophageal mucosa samples and 81 ESCC samples were selected for further analysis. The immune-related gene list was downloaded from the ImmPort dataset (https://immport.org/home).

Differentially Expressed Genes in TCGA and GTEx Cohorts

The Limma package in R was used to screen the differentially expressed genes in ESCC and normal esophageal mucosa samples. A fold change >2 and p-value < 0.05 were the screening conditions (Law et al., 2014).

Enrichment Analysis

The ClusterProfiler, org. Hs.eg.db, enrichplot, ggplot2, and GOplot packages in R were used to perform functional enrichment analyses of the differentially expressed genes. The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis results are shown in circular maps (Bindea et al., 2009). The Gene Set Enrichment Analysis (GSEA) was performed to analyze the distribution trend in gene table sorted by phenotypic correlation.

Screening of Immune-Related Prognostic Genes

The WGCNA package in R was used for the cluster analysis of immune-related differentially expressed genes and screening of prognostic differentially expressed genes. Kaplan-Meier (KM) survival curve analysis was used to further verify the accuracy of the impact of these genes on prognosis (Langfelder and Horvath, 2008).

CIBERSORT

CIBERSORT is a useful tool for evaluating the abundance of various immune cell types in the tumor microenvironment. A p-value < 0.05 was a selective condition, and 73 ESCC samples were obtained for further analysis (Newman et al., 2015).

Receiver Operating Characteristic Analysis and COX Regression Analysis

The area under the ROC curve (AUC) was used to calculate the predictive ability of the ESCC prognosis model and the COX regression analysis was used to calculate the risk coefficients of ESCC patients. The effect was compared with conventional indexes (Kamarudin et al., 2017).

Cell Culture

Human ESCC cell lines (TE13, EC109) and normal human esophageal epithelial cells (HEEC) were provided by cardiovascular thoracic surgery department of Tianjin Medical University General Hospital. Cells were cultured with RPMI-1640

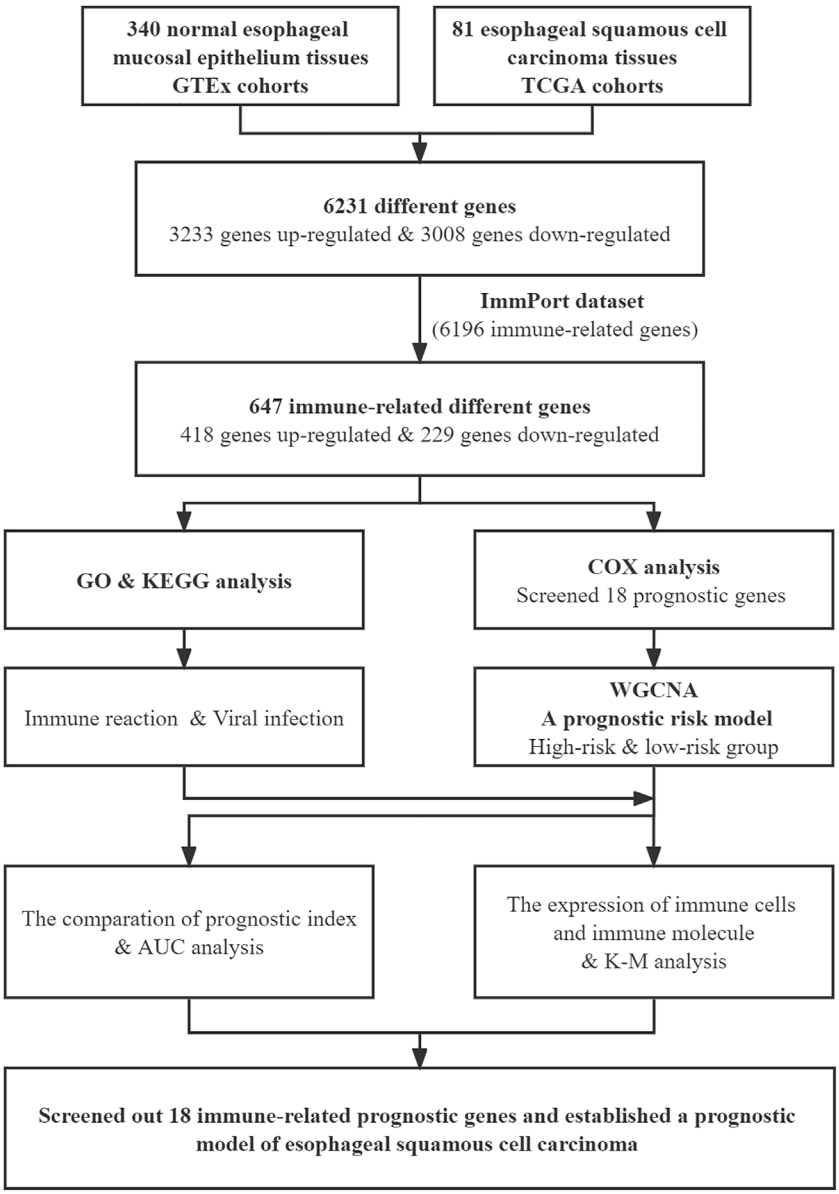


FIGURE 1 | Flow chart of this study.

containing 10% Fetal Bovine Serum (FBS). All cells were maintained in a humidified chamber containing 5% CO_2 at 37°C.

RNA Isolation and Quantitative Real-Time Polymerase Chain Reaction

Total RNA of cultured cells was isolated with RNAiso Plus (TaKaRa) and reverse-transcribed to cDNA with PrimeScript Strand cDNA Synthesis Kit (TaKaRa). Quantitative real-time polymerase chain reaction (qRT-PCR) was performed with TB Green Premix Ex Taq II (TaKaRa). The primer sequences are shown in **Supplementary Table S1**. Relative gene expression was determined by the comparative $2^{-\Delta\Delta CT}$ method.

RESULTS

Differentially Expressed Genes and Immune-Related Genes in Normal and ESCC Tissues

The flow chart of the study is shown in **Figure 1**. We downloaded the transcriptome expression profiles of 95 ESCC tissues and 340 normal esophageal mucosal epithelium samples from TCGA and GTEx cohorts, then performed batch integration analysis on the data. Finally, 274 normal esophageal mucosa samples and 81 ESCC samples with 34,350 genes were selected for further analysis (**Supplementary Data S1**). Differentially expressed

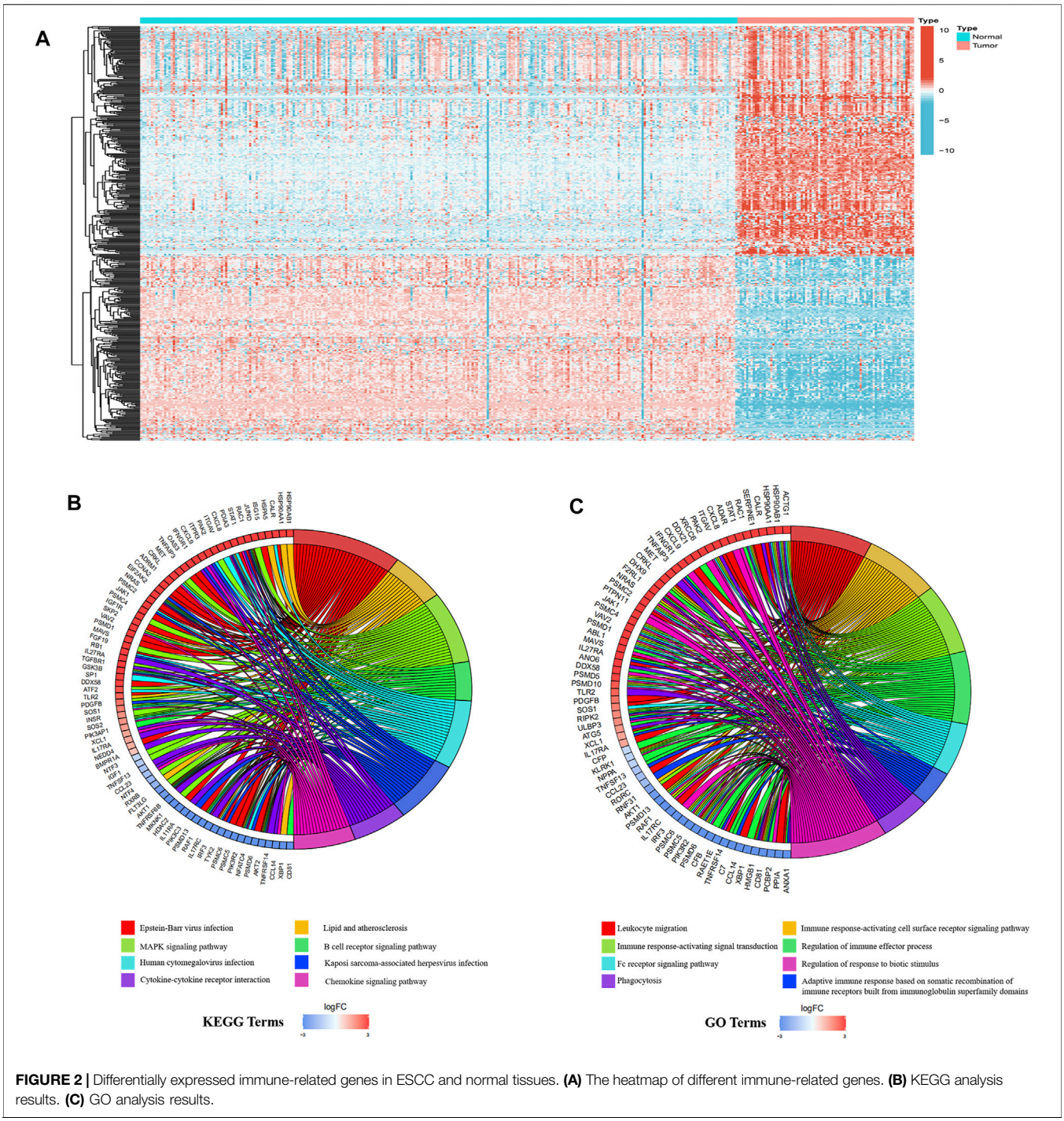


FIGURE 2 | Differentially expressed immune-related genes in ESCC and normal tissues. **(A)** The heatmap of different immune-related genes. **(B)** KEGG analysis results. **(C)** GO analysis results.

genes were screened by Limma. Overall, 6,231 genes were found to be differentially expressed between ESCC and normal tissues. Of these, 3,008 genes were downregulated and 3,223 genes were upregulated in ESCC tissues relative to normal tissues.

Next, we downloaded the list of 6,196 immune-related genes from the ImmPort dataset and screened for any differentially expressed genes among them. There were 647 differentially expressed immune-related genes, with 418 upregulated and 229 downregulated in ESCC tissues relative to normal tissues. The differentially expressed genes are shown in **Supplementary Table S2**. The heatmaps of differentially expressed immune-related genes and other differentially expressed genes are shown in **Figure 2A**; **Supplementary Figure S1**, respectively.

Immune reactions and viral infections may be the main factors that maintain the malignant state of ESCC. GO and KEGG analyses were performed on the differentially expressed immune-related genes, with the results showing that the first eight most significant GO terms enriched were all related to immune

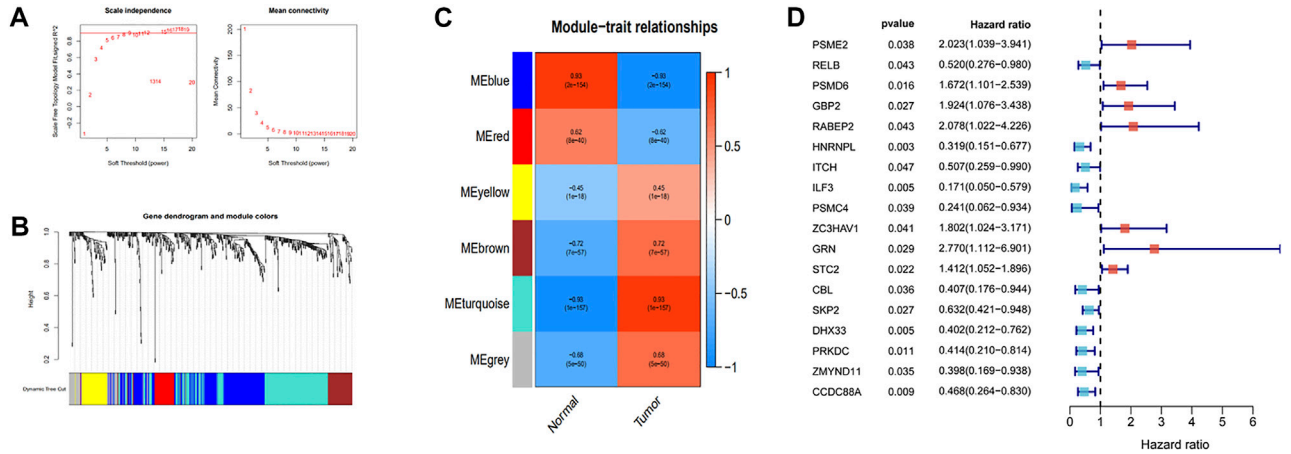


FIGURE 3 | Identification of the independent prognostic genes by WGCNA. **(A)** Analysis of the scale-free index and mean connectivity for various soft-threshold powers. **(B)** Dendrogram of all different genes clustered based on the measurement of dissimilarity. **(C)** The heatmap of the relationship between module eigengenes and tissue types. **(D)** Forest plot of the relationship between gene expression and the risk of survival time.

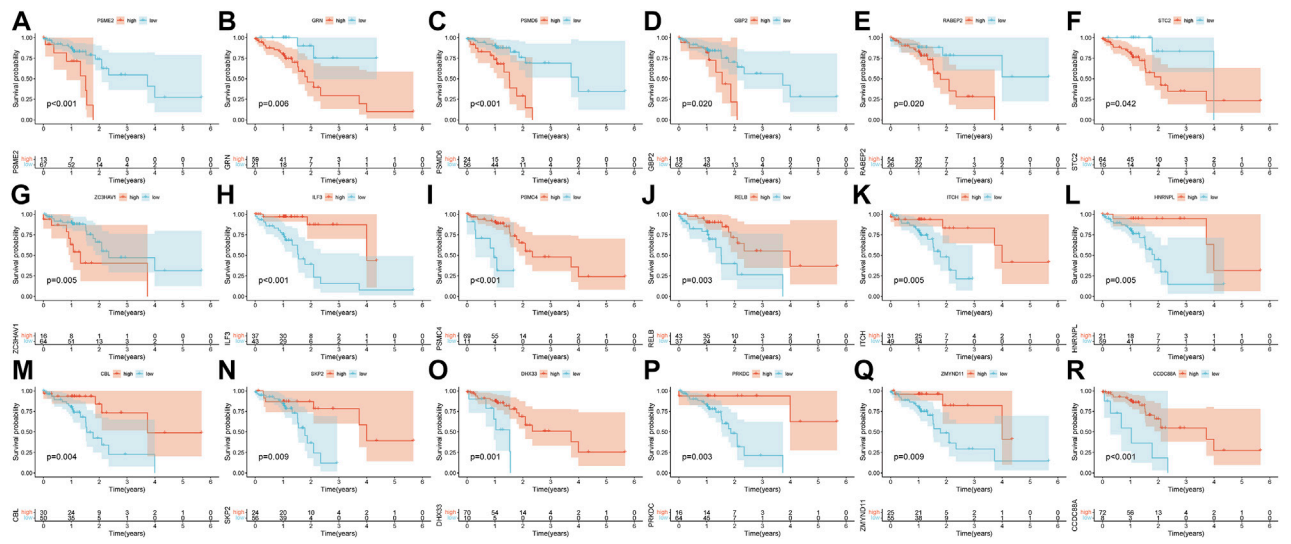


FIGURE 4 | Kaplan-Meier survival curves of the 18 prognostic genes **(A–R)**.

response. Additionally, three of the first eight KEGG terms enriched were related to immune response and three were related to viral infection. The enriched KEGG terms, B cell receptor signaling pathway, cytokine-cytokine receptor interaction, and chemokine signaling pathway, were consistent with the enriched GO terms, such as immune response-activating cell surface receptor signaling pathway, immune response-activating signal transduction, and regulation of immune effector process (**Figures 2B,C**).

Independent Prognostic Genes Screened by WGCNA in ESCC Patients

Pearson's correlation coefficient was used to cluster the immune-related differentially expressed genes in ESCC and normal tissues. The outliers were all removed, and the clustering tree was built (**Figures 3A,B**). Next, the adjacency matrix and topological overlap matrix were constructed. The six most important modules reflecting the combined function of multiple genes were identified based on the clustering tree and the adjacency and topological overlap matrices (**Figure 3C**). The blue module with 202 genes and the turquoise module with 222 genes were highly related to the development of ESCC, so these modules were selected for further analysis. Then, we combined gene expression data and patient survival information to calculate the relationship between gene expression levels and days of survival. We identified 18 genes that were related to ESCC patient prognosis, 11 of which were related to good prognosis (*RELB*, *HNRNPL*, *ITCH*, *ILF3*, *PSMC4*, *CBL*, *SKP2*, *DHX33*, *PRKDC*, *ZMYND11*, and *CCDC88A*) and seven that were related to poor

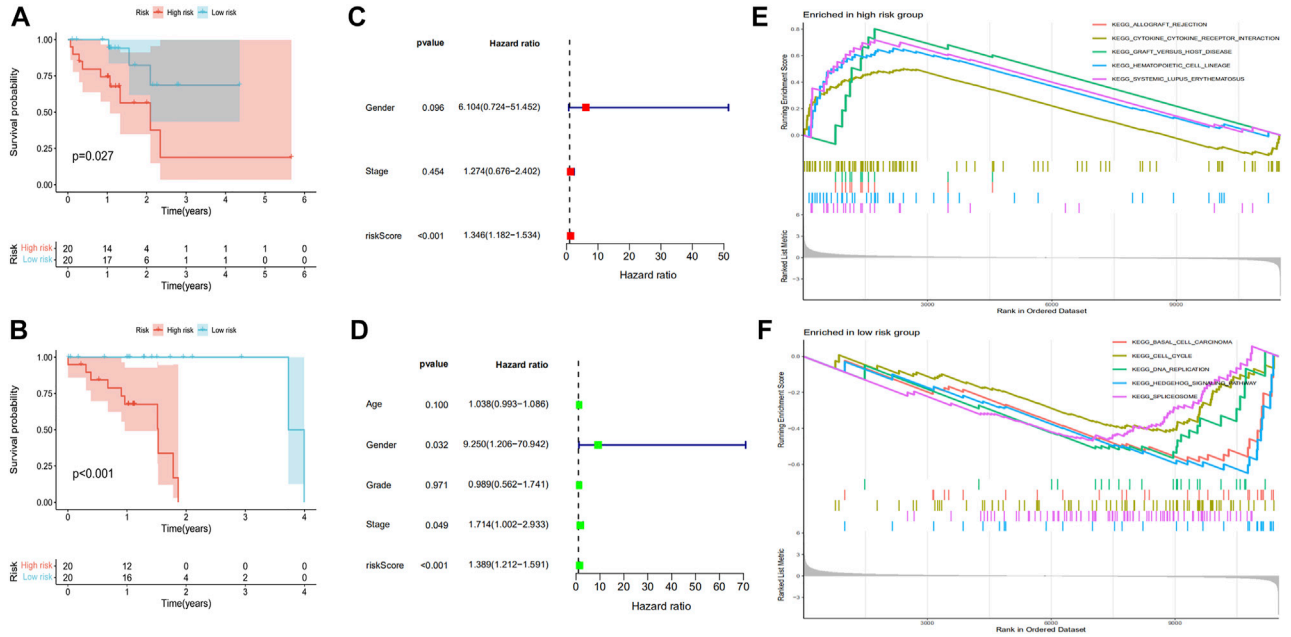


FIGURE 5 | Establishment of a prognostic risk model for ESCC. The Kaplan-Meier survival time was analyzed in the **(A)** training group and **(B)** test group. The independent prognostic analysis was analyzed in the **(C)** training group and **(D)** test group. **(E,F)** show the GSEA enrichment between the high-risk and low-risk groups.

prognosis (*PSME2*, *PSMD6*, *GBP2*, *RABEP2*, *ZC3HAV1*, *GRN*, and *STC2*) (**Figures 3D**, **4A–R**).

A Prognostic Risk Model for ESCC

The 81 ESCC selected samples were randomly divided into two groups, training and test. Then, the risk coefficients of 18 prognostic genes were calculated by COX regression analysis. The sum of the product of risk coefficient and expression level for each gene was used to determine the risk degree of ESCC. The R survival package was then used to establish the prognostic risk model. Finally, we found that the prognostic risk model composed of seven of the 18 prognostic genes had high accuracy (*PSMD6*, *RABEP2*, *GRN*, *STC2*, *ITCH*, *ILF3*, and *PSMC4*). The expression levels of *PSMD6*, *RABEP2*, *GRN*, and *STC2* were related to a poor ESCC prognosis, while the expression levels of *ITCH*, *ILF3*, and *PSMC4* were related to a good ESCC prognosis. According to the prognostic risk model, we divided the training and test groups into high-risk and low-risk subgroups. For the training group, patients in the low-risk group had better survival rates, with similar results in the test group (**Figures 5A,B**). Independent prognostic analyses were then performed on the training and test groups. The results suggested that the risk score was related to ESCC prognosis in both groups (**Figures 5C,D**). We performed GSEA to verify the accuracy of the established prognostic risk model in predicting ESCC prognosis. There were significant differences in GSEA enrichment between the high-risk and low-risk groups. The GSEA enrichment in the high-risk group was mainly related to immune disease and immune response, while this enrichment in the low-risk group was mainly related to tumor disease and tumorigenesis (**Figures 5E,F**). The gene mutations between the high-risk and low-risk groups were also different (**Supplementary Figure S2**).

The Distribution of Immune Cells and Functional Molecules in ESCC Patients

To examine the immune status of low-risk and high-risk group patients, CIBERSORT was used to evaluate the distribution of 23 immune cell types in the ESCC tumor microenvironment. The proportions of immune cells in the low-risk and high-risk groups are shown in **Figure 6A**. There was abundant immune cell infiltration in both the high-risk and low-risk groups. We then compared the immune cell composition between these groups, finding increased levels of M1 macrophages and resting dendritic cells in the high-risk groups (**Figure 6B**). Accordingly, we analyzed the relationship between the immune cell composition and survival time, observing that high levels of resting memory CD4 T cells, M0 macrophages, and regulatory T cells were related to a better survival time (**Figures 6C–E**). High infiltration of M1 macrophages, M2 macrophages, CD8 T cells, and resting dendritic cells showed opposite effects, suggesting that M1 macrophages and resting dendritic cells were increased in malignant ESCC patients and related to a poor prognosis (**Figures 6F–I**). Interestingly, high levels of M1 and M2 macrophages were related to poor prognosis, while high amounts of M0 macrophages showed the opposite. These data imply that the different states of macrophages are related to ESCC prognosis.

The presence of functional immune cells and functional molecules in the high-risk and low-risk groups were also compared. The results showed that a large number of functional immune cells and functional immune molecules were increased in the high-risk group (**Figure 7A**). Consistent with these data, high levels of functional

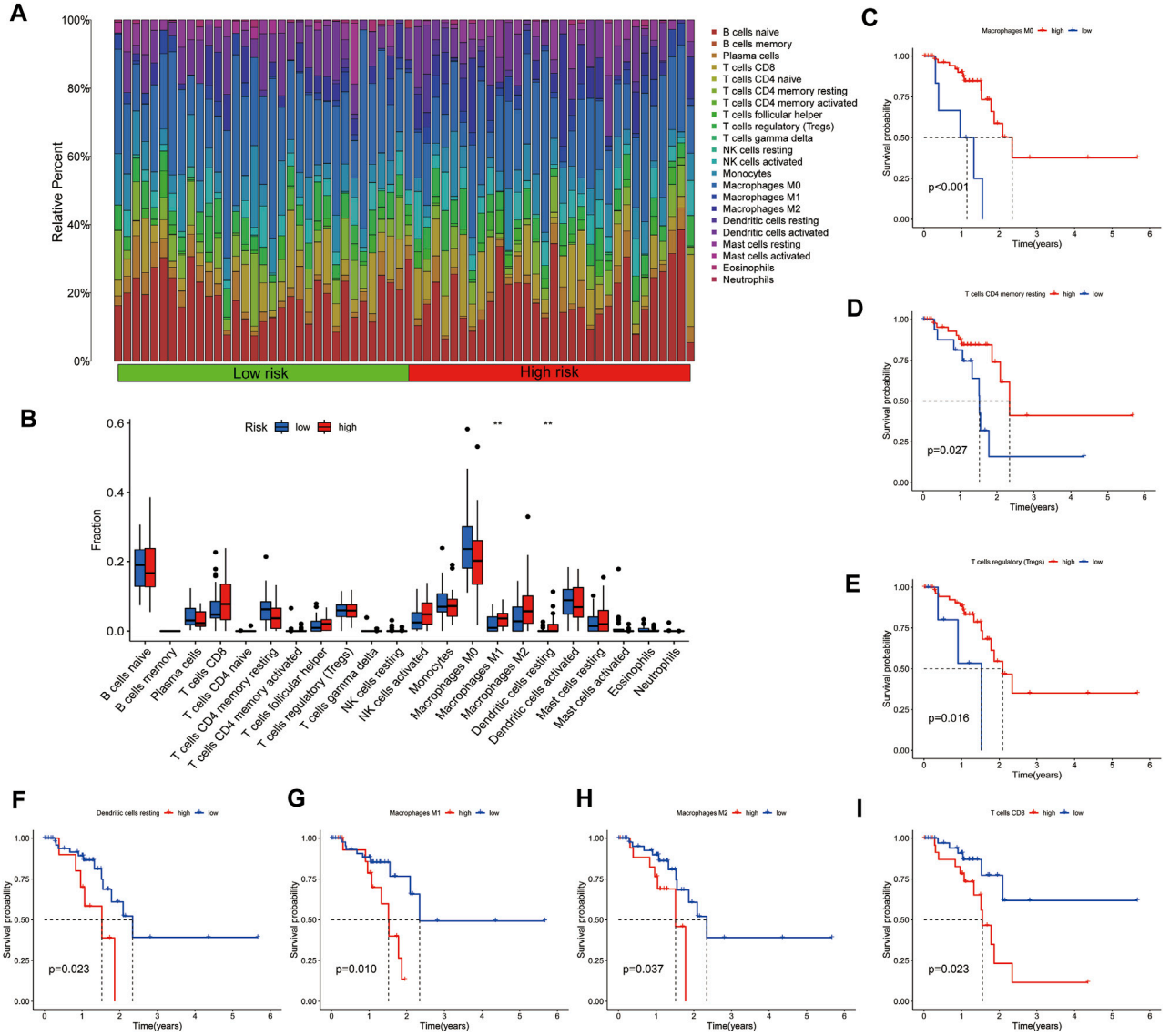


FIGURE 6 | (A) Heatmap and **(B)** distribution of immune cells in the low-risk and high-risk groups of ESCC patients. The immune cell types related to good prognosis are shown in **(C–E)**. The immune cell types related to poor prognosis are shown in **(F–I)**.

immune cells and functional immune molecules were related to poor survival. In particular, patients with low levels of Chemokine receptors (CCR), immune checkpoint molecules, plasmacytoid DC (pDCs), dendritic cells (DCs), and T cell co-stimulation had better survival than those with high levels (**Figure 7B**).

The Immune-Related Gene Prognostic Index Could Predict the Prognostic Status and Immunotherapy Response of ESCC Patients

We examined the Tumor Immune Dysfunction and Exclusion (TIDE) scores of the high-risk and low-risk groups to further analyze the difference of immune status between the two groups. The results showed that the TIDE scores were higher in the high-risk group than in the low-risk group (**Figure 8A**). The dysfunction score of high-risk patients was higher than that of low-risk patients, while the exclusion score showed the opposite trend (**Figures 8B,D**). These data indicate that the immune escape risk and poor prognosis were higher in the high-risk group. The high degree of microsatellite instability (MSI) in the high-risk group also confirmed this phenomenon (**Figure 8C**). After comparing the area under the ROC curve (AUC) for the risk score, TIDE score, and Tumor Inflammation Signature (TIS) score, we found the TIDE score to have the best ability of predicting survival time in ESCC patients (**Figure 8E**). The risk score had

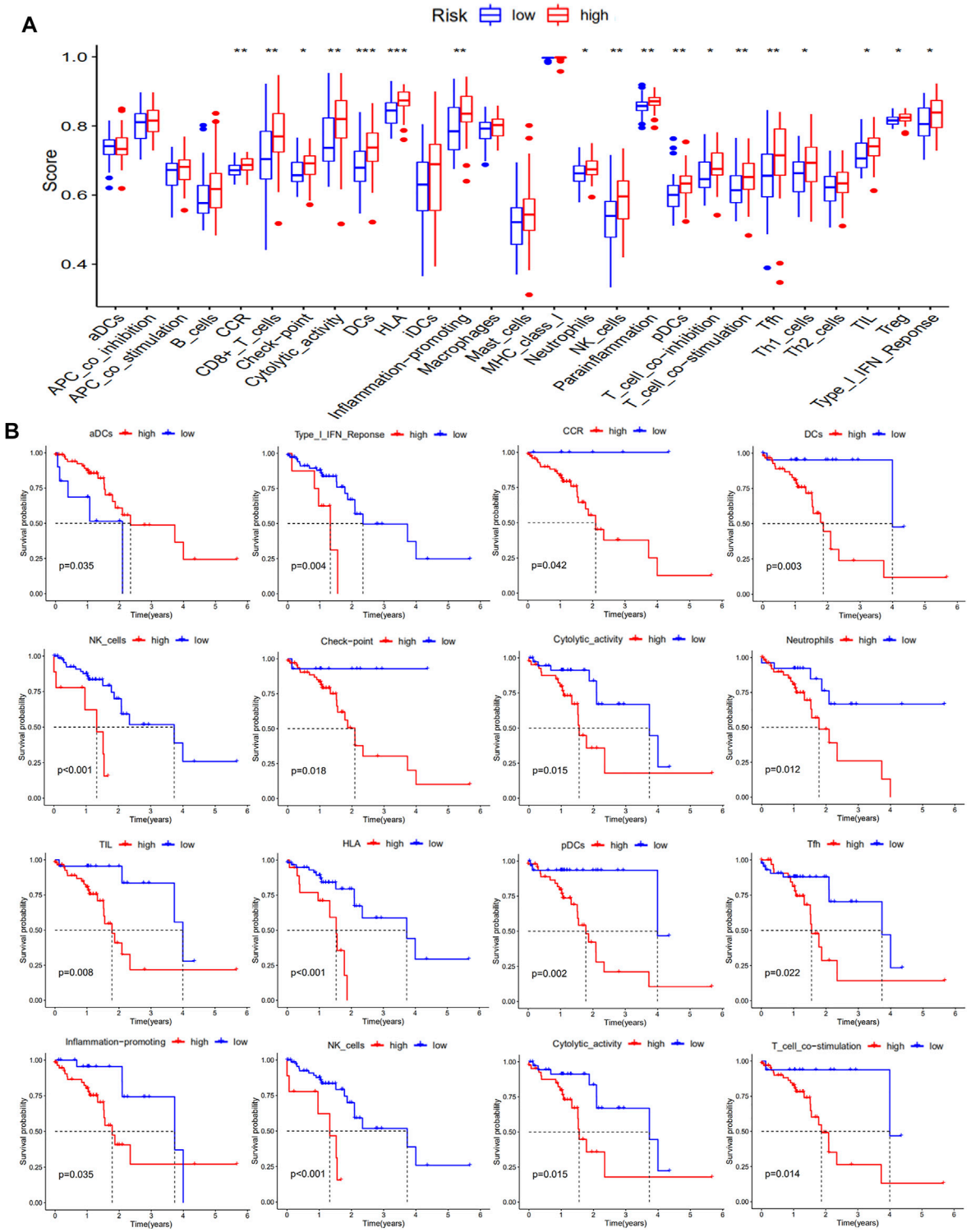


FIGURE 7 | Levels of functional immune cells and functional molecules in ESCC patients. The levels of functional immune cells and functional molecules are shown in **(A)**, and the Kaplan-Meier survival analysis is shown in **(B)**.

a good predictive effect on the 1, 2, and 3-years survival of ESCC patients, and the AUCs were 0.852, 0.842, and 0.931, respectively (**Figure 8F**). In conclusion, we found 18 immune-related prognostic genes related to the occurrence of ESCC and used several of them to establish a prognostic model for predicting disease severity.

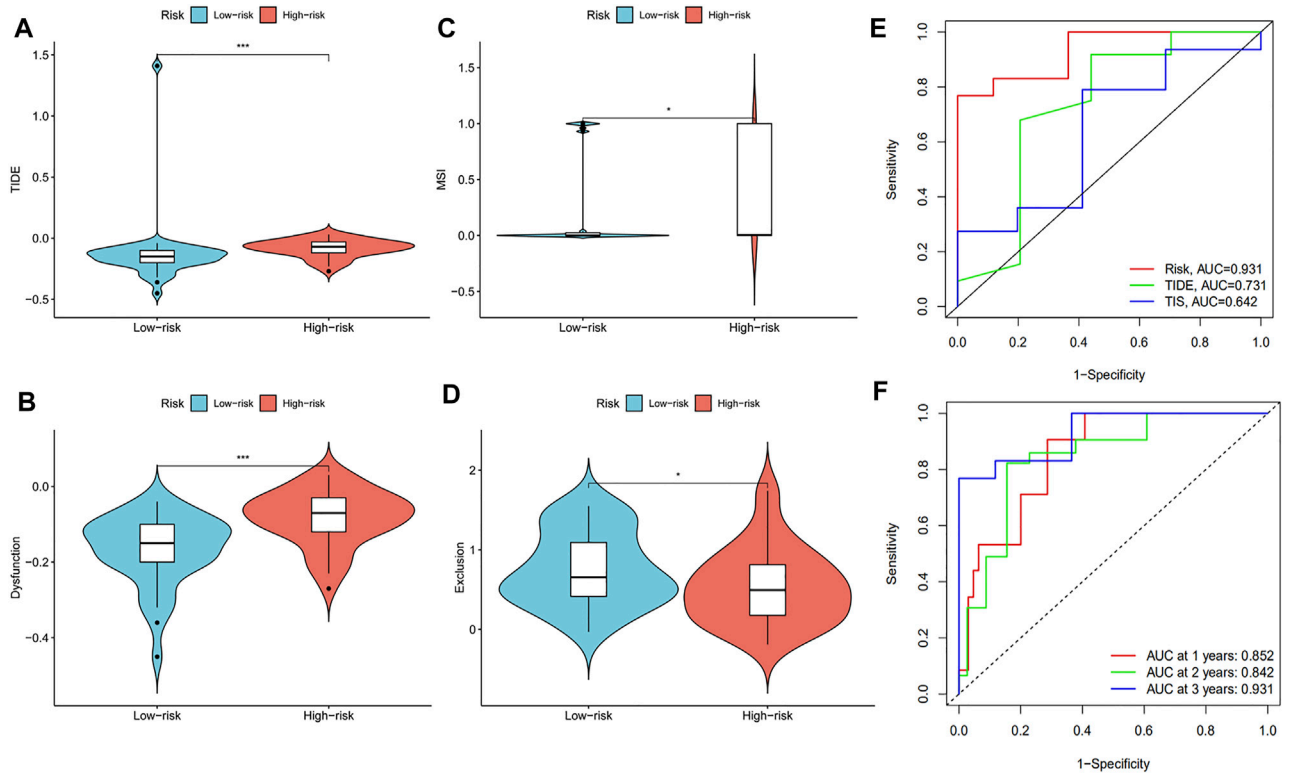


FIGURE 8 | The evaluation of prognostic risk model. The **(A)** TIDE scores, **(B)** dysfunction scores, **(C)** MSI, and **(D)** exclusion scores were analyzed in the low-risk and high-risk groups of ESCC patients. **(E)** ROC analysis of the risk model, TIDE scores, and TIS scores. **(F)** ROC analysis of the 1, 2, and 3-years survival of ESCC patients.

The Expression Levels of seven Prognostic Genes

We analyzed the relative mRNA expression levels of seven prognostic genes in the normal human esophageal epithelial cell (HEEC) and two ESCC cell lines (TE13, EC109) by qRT-PCR. As shown in **Figure 9A**, there were obvious differences in the expression of these genes. Meanwhile, we analyzed the expression of seven prognostic genes in 19 ESCC cell lines in the CCLE database. We found that the expression trends of these seven genes in poorly differentiated (TE5, TE9, KYSE70, KYSE150, KYSE410) and well differentiated (TE1, TE4, TE6, TE10, TE15, KYSE30, KYSE180, KYSE270, and KYSE510) ESCC cell lines were nearly consistent with our previous analysis (**Figures 9B,C**).

DISCUSSION

With the emergence of immune checkpoint inhibitor therapies, the treatment options for cancer patients have significantly increased. For ESCC patients, the results of the KEYNOTE-181 clinical trial study highlighted the outstanding achievements of immunotherapy (Kojima et al., 2020). For patients with advanced ESCC, distinguishing the risk degree of individual patients and their response to immunotherapy could better improve the treatment effects. However, there is still no clear molecular targeting index to indicate ESCC severity, nor is there a corresponding prediction model to infer the response of patients to immunotherapy. In this study, we aimed to address this problem by screening prognostic genes and establishing a prognostic immune response model for ESCC. The complex tumor microenvironment is affected by many factors, and the synergistic effect of many genes can affect tumor progression. Through module analysis, WGCNA eliminates the offset of a single factor and obtains important modules related to disease to further analyze important modules and obtain key genes related to prognosis (Langfelder and Horvath, 2008). Using this method, we obtained 18 differentially expressed immune-related genes, 11 of which were associated with a good ESCC prognosis. The other seven genes showed the opposite. Multiple prognostic hub genes still cannot easily prompt the prognosis and risk of ESCC patients. To solve this problem, a prognostic risk model was established using the immune-related gene prognostic index, which could effectively predict the prognostic risk and immune response of ESCC patients (Huang et al., 2021).

The prognostic risk model contained seven genes (*PSMD6*, *RABEP2*, *GRN*, *STC2*, *ITCH*, *ILF3*, and *PSMC4*). PSMD6 encodes a member of the protease subunit 26S family that is part of a multicatalytic proteinase complex with a highly ordered structure composed of two complexes, a 20S core, and a 19S

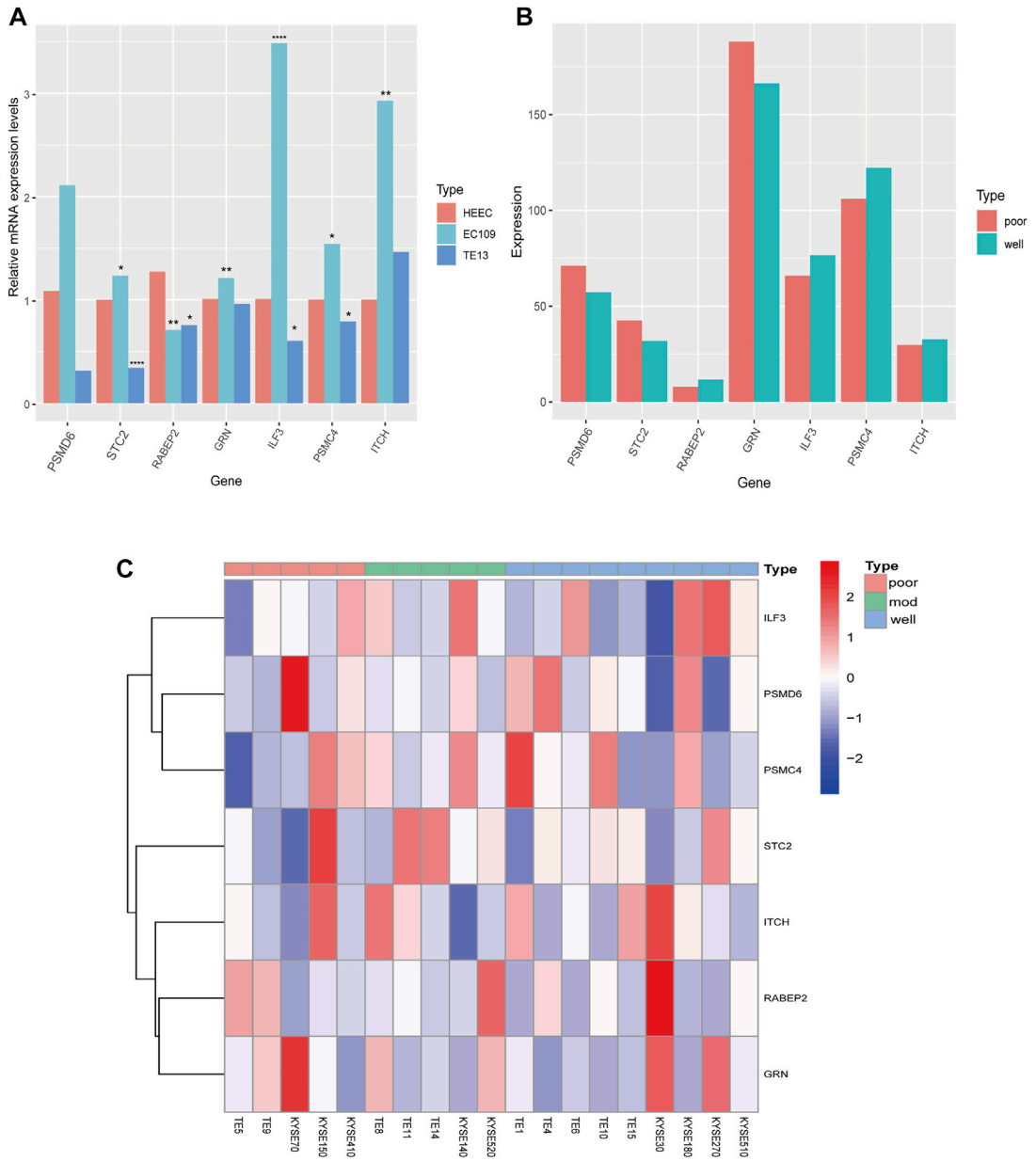


FIGURE 9 | The expression of seven prognostic genes in HEEC and ESCC cell lines. **(A)** The relative mRNA expression levels of seven prognostic genes in HEEC and two ESCC cell lines (EC109, TE13) by qRT-PCR. **(B)** The expression levels of seven prognostic genes in poorly differentiated and well differentiated ESCC cell lines. **(C)** The heatmap of seven prognostic genes expression in 19 ESCC cell lines.

regulator. PSMD6 colocalizes with DNA damage foci and is involved in the ATP-dependent degradation of ubiquitinated proteins, and PSMC4 is involved in the non-ATPase subunits of the 19S regulator lid (Jacquemont and Taniguchi, 2007). The effect of PSMC4 and PSMD6 expression in tumor progression is still unclear. RABEP2 is a recently identified protein required for the formation of collateral vessels during development, which is possibly why it can promote the development of an ESCC tumor (Aghajanian et al., 2021). GRNs are a family of secreted, glycosylated peptides that are cleaved from a single precursor protein with 7.5 repeats of a highly conserved 12-cysteine granulin/epithelin motif. Research has suggested that GRN can promote ESCC progression *via* the autocrine-dependent FAM135B/AKT/mTOR signaling pathway (Dong et al., 2021). STC2 encodes a secreted homodimeric glycoprotein that is expressed in a wide variety of tissues and may have autocrine or paracrine functions. A previous study regarded STC2 as a predictive marker for lymph node metastasis in ESCC (Kita et al., 2011). Moreover, it was reportedly involved in breast, gynecologic, gastric, colorectal, liver, and respiratory cancers (Li et al., 2021). ITCH is a regulator of lymphocyte differentiation and the immune response. Mutations in this gene are involved in the development of multi-system autoimmune diseases (Kleine-Eggebrecht et al., 2019).

Studies have reported that circulating ITCH may have an inhibitory effect on ESCC by regulating the Wnt pathway (Yan et al., 2020). ILF3 has been reported to contribute to the development of various cancers (Hu et al., 2013; Cheng et al., 2018; Liu et al., 2019). High expression of ILF3 was shown to be involved in the metabolic alterations in ESCC patients, especially for the intermediate metabolites of the glycolysis pathway (Zang et al., 2021). Through survival analysis, we found that the expression of *PSMD6*, *RABEP2*, *GRN*, and *STC2* were related to a poor ESCC prognosis, while the expression of *ITCH*, *ILF3*, and *PSMC4* were related to a good prognosis. The immune prognosis analysis model established with these genes was reliable.

According to this immune prognostic risk model, we divided ESCC patients into low-risk and high-risk groups. We then analyzed the relationship between immune cells and the degree of malignancy of ESCC patients. We found the level of resting DCs to be elevated in malignant ESCC patients and related to poor prognosis. DCs are considered to be key antigen-presenting cells (APCs) that are mainly used to control the initiation of T cell-dependent immune responses (Saiz et al., 2018). Resting DCs are present in a variety of tissues, with their main role being to bind and internalize antigens. However, their role in antigen presentation and activation of T cells is limited. The transformation of resting DCs into activated DCs is affected by a variety of factors that regulate the expression of costimulatory and adhesion molecules. Through this process, the capacity for antigen uptake is lost and the ability of potent T cell stimulation is acquired (Montoya et al., 2002; Probst et al., 2005; Steptoe et al., 2007). Studies have shown that resting DCs are abnormally distributed in a variety of tumors, including clear cell renal cell carcinoma (Pan et al., 2020), colorectal cancer (Li et al., 2020), lung adenocarcinoma (Zhang et al., 2021), and colon cancer (Cui et al., 2021). This also illustrates the important role of resting DCs in tumorigenesis and development.

As a plastic and pluripotent cell population, macrophages exhibit significant functional differences under the influence of different microenvironments *in vivo* and *in vitro* (Artyomov et al., 2016). According to the activation state and function, macrophages can be divided into M1 and M2 (Yunna et al., 2020). M1 type macrophages secrete pro-inflammatory cytokines and chemokines and participate in the positive immune response. M2 type macrophages have only weak antigen presentation ability and downregulate the immune response by secreting inhibitory cytokines such as IL-10 and TGF-β (Orecchioni et al., 2019; Anderson et al., 2021). Our results show that the various macrophage stages have different effects on the prognosis of ESCC patients, which is consistent with the previously described findings in the field. Therefore, the distribution of immune cell types can affect the disease malignancy of ESCC.

The TIDE score is an advanced method to predict the sensitivity of tumors to immune checkpoint treatment by using gene expression information. The TIDE score calculated by the software consists of two parts: dysfunction score and exclusion score. The calculation principle of the dysfunction score is that genes with immune disorders have a higher weight, and it can be obtained by multiplying by the expression level. The exclusion score is obtained by multiplying the weight of immune rejection genes by the expression level (Wang et al., 2020). The TIDE score has been confirmed to have higher accuracy in evaluating the efficacy of PD1 and CTLA4 monoclonal antibody treatment (Kim et al., 2021). The TIS score was an important index to evaluate the reactivity to PD1 monoclonal antibody treatment by calculating the levels of inhibited T cells in the tumor microenvironment (Danaher et al., 2018).

MSI refers to the phenomenon that new microsatellite alleles appear in the tumor from a change to the microsatellite length caused by the insertion or deletion of a repeat unit (Vilar and Gruber, 2010). MSI has been widely confirmed in colon cancer (Boland and Goel, 2010), and its mechanism mainly relies on the lack of a DNA mismatch repair system. MSI has also been detected in other cancer types, including gastric cancer (Ratti et al., 2018), endometrial cancer (Latham et al., 2019), and ovarian cancer (Evrard and Alexandre, 2021). TIDE, TIS, and MSI are classic indicators of tumor microenvironmental immunity and tumor evaluation, with TIDE and TIS focusing on T cell function and MSI focusing on genetic changes. However, these indicators could not fully reflect the complex microenvironment of the tumors. In our study, compared with the TIDE and TIS scores, the immune prognostic and risk model was a better predictor of survival time. This model is possibly a better prediction method for the prognosis and immune response of ESCC patients.

CONCLUSION

Here, we found 18 immune-related prognostic genes related to the occurrence of ESCC and used several of them to establish a prognostic model for predicting disease severity.

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

Conception and design: KX and PZ; Collection and assembly of data: KX and ZZ; Data analysis and interpretation: KX, ZT, and JW; Manuscript writing: ZT and ZZ; The final manuscript was approved by all authors.

FUNDING

This work was supported by the Beijing-Tianjin-Hebei Basic Research Cooperation Special Project (20JCZXJC00190).

ACKNOWLEDGMENTS

Thanks to all the peer reviewers and editors for their opinions and suggestions. We thank J. Iacona, Ph.D., from Liwen Bianji (Edanz) for editing the English text of a draft of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2022.850669/full#supplementary-material

REFERENCES

Aghajanian, A., Zhang, H., Buckley, B. K., Wittchen, E. S., Ma, W. Y., and Faber, J. E. (2021). Decreased Inspired Oxygen Stimulates De Novo Formation of Coronary Collaterals in Adult Heart. *J. Mol. Cel. Cardiol.* 150, 1–11. doi:10.1016/j.yjmcc.2020.09.015

Anderson, N. R., Minutolo, N. G., Gill, S., and Klichinsky, M. (2021). Macrophage-Based Approaches for Cancer Immunotherapy. *Cancer Res.* 81, 1201–1208. doi:10.1158/0008-5472.CAN-20-2990

Artyomov, M. N., Sergushichev, A., and Schilling, J. D. (2016). Integrating Immunometabolism and Macrophage Diversity. *Semin. Immunol.* 28, 417–424. doi:10.1016/j.smim.2016.10.004

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., et al. (2009). ClueGO: a Cytoscape Plug-In to Decipher Functionally Grouped Gene Ontology and Pathway Annotation Networks. *Bioinformatics* 25, 1091–1093. doi:10.1093/bioinformatics/btp101

Boland, C. R., and Goel, A. (2010). Microsatellite Instability in Colorectal Cancer. *Gastroenterology* 138, 2073–2087.e3. doi:10.1053/j.gastro.2009.12.064

Cheng, C. C., Chou, K. F., Wu, C. W., Su, N. W., Peng, C. L., Su, Y. W., et al. (2018). EGFR-mediated Interleukin Enhancer-Binding Factor 3 Contributes to Formation and Survival of Cancer Stem-like Tumorspheres as a Therapeutic Target against EGFR-Positive Non-small Cell Lung Cancer. *Lung Cancer* 116, 80–89. doi:10.1016/j.lungcan.2017.12.017

Chen, W., Li, H., Ren, J., Zheng, R., Shi, J., Li, J., et al. (2021). Selection of High-Risk Individuals for Esophageal Cancer Screening: A Prediction Model of Esophageal Squamous Cell Carcinoma Based on a Multicenter Screening Cohort in Rural China. *Int. J. Cancer* 148, 329–339. doi:10.1002/ijc.33208

Chen, Y., Li, Z. Y., Zhou, G. Q., and Sun, Y. (2021). An Immune-Related Gene Prognostic Index for Head and Neck Squamous Cell Carcinoma. *Clin. Cancer Res.* 27, 330–341. doi:10.1158/1078-0432.CCR-20-2166

Cui, Z., Sun, G., Bhandari, R., Lu, J., Zhang, M., Bhandari, R., et al. (2021). Comprehensive Analysis of Glycolysis-Related Genes for Prognosis, Immune Features, and Candidate Drug Development in Colon Cancer. *Front Cel. Dev. Biol.* 9, 684322. doi:10.3389/fcell.2021.684322

Danaher, P., Warren, S., Lu, R., Samayoa, J., Sullivan, A., Pekker, I., et al. (2018). Pan-cancer Adaptive Immune Resistance as Defined by the Tumor Inflammation Signature (TIS): Results from the Cancer Genome Atlas (TCGA). *J. Immunother. Cancer* 6, 63. doi:10.1186/s40425-018-0367-1

Dong, D., Zhang, W., Xiao, W., Wu, Q., Cao, Y., Gao, X., et al. (2021). A GRN Autocrine-dependent FAM135B/AKT/mTOR Feedforward Loop Promotes Esophageal Squamous Cell Carcinoma Progression. *Cancer Res.* 81, 910–922. doi:10.1158/0008-5472.CAN-20-0912

Evrard, C., and Alexandre, J. (2021). Predictive and Prognostic Value of Microsatellite Instability in Gynecologic Cancer (Endometrial and Ovarian). *Cancers (Basel)* 13, 2434. doi:10.3390/cancers13102434

Hu, Q., Lu, Y. Y., Noh, H., Hong, S., Dong, Z., Ding, H. F., et al. (2013). Interleukin Enhancer-Binding Factor 3 Promotes Breast Tumor Progression by Regulating Sustained Urokinase-type Plasminogen Activator Expression. *Oncogene* 32, 3933–3943. doi:10.1038/onc.2012.414

Huang, R. Z., Mao, M., Zheng, J., Liang, H. Q., Liu, F. L., Zhou, G. Y., et al. (2021). Development of an Immune-Related Gene Pairs index for the Prognosis Analysis of Metastatic Melanoma. *Sci. Rep.* 11, 1253. doi:10.1038/s41598-020-80858-1

Jacquemont, C., and Taniguchi, T. (2007). Proteasome Function Is Required for DNA Damage Response and Fanconi Anemia Pathway Activation. *Cancer Res.* 67, 7395–7405. doi:10.1158/0008-5472.CAN-07-1015

Kakeji, Y., Oshikiri, T., Takiguchi, G., Kanaji, S., Matsuda, T., Nakamura, T., et al. (2021). Multimodality Approaches to Control Esophageal Cancer: Development of Chemoradiotherapy, Chemotherapy, and Immunotherapy. *Esophagus* 18, 25–32. doi:10.1007/s10388-020-00782-1

Kamarudin, A. N., Cox, T., and Kolamunnage-Dona, R. (2017). Time-dependent ROC Curve Analysis in Medical Research: Current Methods and Applications. *BMC Med. Res. Methodol.* 17, 53. doi:10.1186/s12874-017-0332-6

Kim, K. H., Chang, J. S., and Kim, Y. B. (2021). A Novel Gene Signature Associated with Poor Response to Chemoradiotherapy in Patients with Locally Advanced Cervical Cancer. *Int. J. Radiat. Oncol. Biol. Phys.* 111, e614. doi:10.1016/j.ijrobp.2021.07.1636

Kita, Y., Mimori, K., Iwatsuki, M., Yokobori, T., Ieta, K., Tanaka, F., et al. (2011). STC2: a Predictive Marker for Lymph Node Metastasis in Esophageal Squamous-Cell Carcinoma. *Ann. Surg. Oncol.* 18, 261–272. doi:10.1245/s10434-010-1271-1

Kleine-Eggebrecht, N., Staufner, C., Kathemann, S., Elgizouli, M., Kopajtich, R., Prokisch, H., et al. (2019). Mutation in ITCH Gene Can Cause Syndromic Multisystem Autoimmune Disease with Acute Liver Failure. *Pediatrics* 143, e20181554. doi:10.1542/peds.2018-1554

Kojima, T., Shah, M. A., Muro, K., Francois, E., Adenis, A., Hsu, C. H., et al. (2020). Randomized Phase III KEYNOTE-181 Study of Pembrolizumab versus Chemotherapy in Advanced Esophageal Cancer. *J. Clin. Oncol.* 38, 4138–4148. doi:10.1200/JCO.20.01888

Langfelder, P., and Horvath, S. (2008). WGCNA: an R Package for Weighted Correlation Network Analysis. *BMC Bioinformatics* 9, 559. doi:10.1186/1471-2105-9-559

Latham, A., Srinivasan, P., Kemel, Y., Shia, J., Bandlamudi, C., Mandelker, D., et al. (2019). Microsatellite Instability Is Associated with the Presence of Lynch Syndrome Pan-Cancer. *J. Clin. Oncol.* 37, 286–295. doi:10.1200/JCO.18.00283

Law, C. W., Chen, Y., Shi, W., and Smyth, G. K. (2014). Voom: Precision Weights Unlock Linear Model Analysis Tools for RNA-Seq Read Counts. *Genome Biol.* 15, R29. doi:10.1186/gb-2014-15-2-r29

Li, M., Wang, H., Li, W., Peng, Y., Xu, F., Shang, J., et al. (2020). Identification and Validation of an Immune Prognostic Signature in Colorectal Cancer. *Int. Immunopharmacol.* 88, 106868. doi:10.1016/j.intimp.2020.106868

Li, S., Huang, Q., Li, D., Lv, L., Li, Y., and Wu, Z. (2021). The Significance of Stanniocalcin 2 in Malignancies and Mechanisms. *Bioengineered* 12, 7276–7285. doi:10.1080/21655979.2021.1977551

Liu, Y., Li, Y., Zhao, Y., Liu, Y., Fan, L., Jia, N., et al. (2019). ILF3 Promotes Gastric Cancer Proliferation and May Be Used as a Prognostic Marker. *Mol. Med. Rep.* 20, 125–134. doi:10.3892/mmr.2019.10229

Montoya, M., Schiavoni, G., Mattei, F., Gresser, I., Belardelli, F., Borrow, P., et al. (2002). Type I Interferons Produced by Dendritic Cells Promote Their Phenotypic and Functional Activation. *Blood* 99, 3263–3271. doi:10.1182/blood.v99.9.3263

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

Orecchioni, M., Ghosheh, Y., Pramod, A. B., and Ley, K. (2019). Macrophage Polarization: Different Gene Signatures in M1(LPS+) vs. Classically and M2(LPS-) vs. Alternatively Activated Macrophages. *Front. Immunol.* 10, 1084. doi:10.3389/fimmu.2019.01084

Pan, Q., Wang, L., Chai, S., Zhang, H., and Li, B. (2020). The Immune Infiltration in clear Cell Renal Cell Carcinoma and Their Clinical Implications: A Study Based on TCGA and GEO Databases. *J. Cancer* 11, 3207–3215. doi:10.7150/jca.37285

Probst, H. C., McCoy, K., Okazaki, T., Honjo, T., and van den Broek, M. (2005). Resting Dendritic Cells Induce Peripheral CD8+ T Cell Tolerance through PD-1 and CTLA-4. *Nat. Immunol.* 6, 280–286. doi:10.1038/ni1165

Ratti, M., Lampis, A., Hahne, J. C., Passalacqua, R., and Vale, . (2018). Microsatellite Instability in Gastric Cancer: Molecular Bases, Clinical Perspectives, and New Treatment Approaches. *Cell. Mol. Life Sci.* 75, 4151–4162. doi:10.1007/s00018-018-2906-9

Sadanand, S. (2021). Immunotherapy for Esophageal Cancer. *Nat. Med.* [Online ahead of print]. doi:10.1038/d41591-021-00022-8

Saiz, M. L., Rocha-Perugini, V., and Sánchez-Madrid, F. (2018). Tetraspanins as Organizers of Antigen-Presenting Cell Function. *Front. Immunol.* 9, 1074. doi:10.3389/fimmu.2018.01074

Steptoe, R. J., Ritchie, J. M., Wilson, N. S., Villadangos, J. A., Lew, A. M., and Harrison, L. C. (2007). Cognate CD4+ Help Elicited by Resting Dendritic Cells Does Not Impair the Induction of Peripheral Tolerance in CD8+ T Cells. *J. Immunol.* 178, 2094–2103. doi:10.4049/jimmunol.178.4.2094

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

Vilar, E., and Gruber, S. B. (2010). Microsatellite Instability in Colorectal Cancer-The Stable Evidence. *Nat. Rev. Clin. Oncol.* 7, 153–162. doi:10.1038/nrclinonc.2009.237

Wang, Q., Li, M., Yang, M., Yang, Y., Song, F., Zhang, W., et al. (2020). Analysis of Immune-Related Signatures of Lung Adenocarcinoma Identified Two Distinct Subtypes: Implications for Immune Checkpoint Blockade Therapy. *Aging (Albany NY)* 12, 3312–3339. doi:10.18632/aging.102814

Yan, H., Xiang, H., Sun, B., Feng, F., and Chen, P. (2020). Circular RNA-ITCH Inhibits the Proliferation of Ovarian Carcinoma by Downregulating lncRNA HULC. *Reprod. Sci.* 27, 375–379. doi:10.1007/s43032-019-00049-w

Yunna, C., Mengru, H., Lei, W., and Weidong, C. (2020). Macrophage M1/M2 Polarization. *Eur. J. Pharmacol.* 877, 173090. doi:10.1016/j.ejphar.2020.173090

Zang, B., Wang, W., Wang, Y., Li, P., Xia, T., Liu, X., et al. (2021). Metabolomic Characterization Reveals ILF2 and ILF3 Affected Metabolic Adaptions in Esophageal Squamous Cell Carcinoma. *Front. Mol. Biosci.* 8, 721990. doi:10.3389/fmolb.2021.721990

Zhang, A., Yang, J., Ma, C., Li, F., and Luo, H. (2021). Development and Validation of a Robust Ferroptosis-Related Prognostic Signature in Lung Adenocarcinoma. *Front. Cel. Dev. Biol.* 9, 616271. doi:10.3389/fcell.2021.616271

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 Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 19 April 2022
doi: 10.3389/fbioe.2022.846812

Molecular Characteristics of m6A Regulators and Tumor Microenvironment Infiltration in Soft Tissue Sarcoma: A Gene-Based Study

Kang-Wen Xiao[1†], *Zhi-Qiang Yang*[1†], *Xin Yan*[2], *Zhi-Bo Liu*[1], *Min Yang*[1], *Liang-Yu Guo*[1] *and Lin Cai*[1*]

[1]Department of Orthopedics, Zhongnan Hospital of Wuhan University, Wuhan, China, [2]Department of Urology, Zhongnan Hospital of Wuhan University, Wuhan, China

Background: N6-methyladenosine (m6A) methylation played a key role in tumor growth. However, the relationship between m6A and soft tissue sarcoma (STS) was still unclear.

Methods: The characterization and patterns of m6A modification in STS (TCGA-SARC and GSE17674) were analyzed comprehensively through bioinformatics and real-time polymerase chain reaction (RT-PCR). The effects of different m6A modification patterns on prognosis and immune infiltration of STS were further explored. Differentially expressed gene (DEG) analysis was performed. Moreover, an m6Ascore was constructed by principal component analysis (PCA). In addition, two immunotherapy datasets (IMvigor210 and GSE78220) and a sarcoma dataset (GSE17618) were used to evaluate the m6Ascore.

Results: Huge differences were found in somatic mutation, CNV, and expression of 25 m6A regulators in STS. Two modification patterns (A and B) in STS were further identified and the m6A cluster A showed a better clinical outcome with a lower immune/stromal score compared with the m6A cluster B ($p < 0.050$).In addition to , most STS samples from m6A cluster A showed a high m6Ascore, which was related to mismatch repair and a better prognosis of STS ($p < 0.001$). In contrast, the m6A cluster B, characterized by a low m6Ascore, was related to the MYC signaling pathway, which led to a poor prognosis of STS. A high m6Ascore also contributed to a better outcome of PD-1/PD-L1 blockade immunotherapy.

Conclusion: The modification patterns of 25 m6A regulators in the STS microenvironment were explored comprehensively. The novel m6Ascore effectively predicted the characteristics of the tumor microenvironment (TME) and outcome in STS and provided novel insights for future immunotherapy.

OPEN ACCESS

Edited by:
Ping Zhang,
Griffith University, Australia

Reviewed by:
Fengbiao Mao,
Peking University Third Hospital, China
Guoying Ni,
University of the Sunshine Coast, Australia

***Correspondence:**
Lin Cai
orthopedics@whu.edu.cn

[†]*These authors share first authorship*

Specialty section:
This article was submitted to Preclinical Cell and Gene Therapy, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *31 December 2021*
Accepted: *22 March 2022*
Published: *19 April 2022*

Citation:
Xiao K-W, Yang Z-Q, Yan X, Liu Z-B, Yang M, Guo L-Y and Cai L (2022) Molecular Characteristics of m6A Regulators and Tumor Microenvironment Infiltration in Soft Tissue Sarcoma: A Gene-Based Study. Front. Bioeng. Biotechnol. 10:846812. doi: 10.3389/fbioe.2022.846812

Abbreviations: AUC, area under the curve; cMap, Connectivity Map; CNV, copy number variation; DEGs, differentially expressed genes; ES, enrichment score; FDR, false discovery rate; FPKM, Fragments Per Kilobase Million; GEO, Gene Expression Omnibus, GO, Gene Ontology; GSEA, Gene Set Enrichment Analysis; HR, hazard ratio; HSMC, human skeletal muscle cell; KEGG, Kyoto Encyclopedia of Genes and Genomes; m6A, N6-methyladenosine; PCA, principal component analysis; PPI, protein–protein interaction; ROC, receiver operating characteristic; RT-PCR, real-time polymerase chain reaction; STRING, Search Tool for Recurring Instances of Neighbouring Genes; STS, soft tissue sarcoma; TCGA, The Cancer Genome Atlas; TME, tumor microenvironment.

Keywords: m6A, soft tissue sarcoma, immunotherapy, tumor microenvironment, survival

INTRODUCTION

Since m6A was first detected in poly (A) RNA in 1974, this RNA modification has been found to regulate huge numbers of biological processes in many fields (Alarcón et al., 2015). Currently, m6A regulators could be divided into three types: adenosine methyltransferases (writers), demethylases (erasers), and binding proteins (readers) (Cao et al., 2016). Dysfunction of these regulators would contribute to incomplete m6A modification, which further regulated tumor development and progression (Delaunay and Frye, 2019). For instance, increased expression of ALKBH5 caused by hypoxia could accelerate mammosphere growth in cancer stem cells (Zhang et al., 2016). Furthermore, high expression of ALKBH5 also contributed to the inhibition of the progression of bladder cancer (Yu et al., 2021). A recent study demonstrated that overexpression of YTHDF1 could promote the growth of hepatocellular carcinoma through autophagy (Li et al., 2021a). Moreover, METTL3 was also found to accelerate tumorigenesis in glioblastoma cells (Cui et al., 2017). Although the role of m6A regulators in other tumors was widely studied recently, the relationship between m6A and STS remained unclear. Hence, it was of great importance to evaluate the role of m6A in STS.

Being rare mesenchymal malignancies with heterogeneity, STSs have been paid more attention in the last few decades (Stiller et al., 2013). The latest research calculated that nearly 13,130 new STS cases and 5,350 deaths would be detected in America (Siegel et al., 2020). However, the current treatment of STS is still surgery combined with radiotherapy, which led to distant metastasis in 25% of STS patients, and the metastasis rate rose to about 50% in high-grade STS (Brennan, 2005). Therefore, it was important to explore novel strategies against STS.

Immunotherapy, mainly consisting of PD-1/L1 and CTLA-4, has emerged as a promising treatment for cancer. However, the responsiveness to those immune checkpoint blockade (ICB) was low and few tumors were reported to effectively respond to ICB, which disappointed clinicians and patients (Topalian et al., 2012). The TME has been reported to be correlated with many tumor activities including tumor angiogenesis and growth (Hanahan and Coussens, 2012). As a key part of the TME, several immune cells have been observed to affect the progression and the response to cancer immunotherapy. Recent research revealed that excluded T cells could inhibit the tumor response to ICB (Mariathasan et al., 2018). Moreover, a recent study reported that decreased YTHDF1 expression was found to enhance the antitumor ability of CD8 (+) T cells in the mouse model (Wilkerson and Hayes, 2010). Although many components of the TME have been reported to be associated with tumor progression and immunotherapeutic effects, these studies were usually based on individual immune components and were not related to the m6A modification. Therefore, integrative analysis of m6A modification in the STS microenvironment and exploring effective markers to predict the therapeutic effect of ICB were urgently needed.

In this study, clinical and transcriptome data of STS from TCGA (The Cancer Genome Atlas) and GEO (Gene Expression Omnibus) databases were collected. Genetic variation and expression of m6A regulators in STS were further analyzed. The Search Tool for Recurring Instances of Neighboring Genes (STRING) database was used to detect connections among m6A regulators (Han et al., 2019). Then, two different m6A modification patterns were identified using consensus cluster analysis (Wilkerson and Hayes, 2010), a method that has been widely used in bioinformatics. Moreover, significantly different prognoses, immune infiltration, and pathways of STS were detected between these two m6A modification patterns. In addition, the m6Ascore accurately evaluated the prognosis and immunotherapy response of the tumor, which brought novel insights into the immunotherapy of STS.

METHODS

Sample and Data Collection

TCGA-SARC with 265 STS samples and the corresponding clinical information, somatic mutation, and CNV were collected from UCSC-XENA (http://xena.ucsc.edu/). Here, we chose somatic mutation data to explore the somatic mutation of m6A regulators in STS, while CNV data were to explore the difference in CNV of m6A regulators in STS. Transcriptome data were used to explore the expression of m6A regulators between STS and adipose tissue and for further bioinformatics analysis. GSE17674 with 62 samples (Hugo et al., 2016) and GSE17618 (Savola et al., 2011) including 44 STS samples were collected from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). The human genome annotation GTF file was collected from the gencode platform (https://www.gencodegenes.org/). GPL570 was used for GSE17618 and GSE17674. Robust multi-array average (RMA) normalization was performed in GSE17674 and GSE17618, while transcripts per kilobase million (TPM) normalization was performed in TCGA-SARC. Basic information on these three datasets were shown in **Supplementary Table S1**. Three pairs of STS samples and the adjacent normal tissue were collected from Zhongnan Hospital of Wuhan University. This study was also approved by the institutional ethics board of Zhongnan Hospital of Wuhan University.

Immunotherapy datasets IMvigor210 (anti-PD-L1) including 298 samples with complete clinical information and GSE78220 (anti-PD-1) (Hugo et al., 2016) including 27 samples with complete clinical information were collected from a previous study (Mariathasan et al., 2018) and GEO database, respectively. IMvigor210 was normalized by the trimmed mean of M-values, and GSE78220 was normalized by FPKM (Fragments Per Kilobase Million). A total of 25 m6A regulators were selected based on recent studies (Zhang et al., 2020; Chen et al., 2021). The flowchart of this study is displayed in **Supplementary Figure S1**.

Cell Culture

Human skeletal muscle cell line (HSMC) and sarcoma cell line (A673) were collected from the American Type Culture Collection. A673 was cultured in RPMI 1640 medium (Hyclone) with 10% fetal bovine serum (Gibco) and 1% antibiotics (100 U/ml penicillin and 100 μg/ml streptomycin). HSMC was cultured in DMEM (Hyclone) medium with 10% fetal bovine serum and 1% antibiotics (100 U/ml penicillin and 100 μg/ml streptomycin). The cells were maintained in an incubator set to 37°C with 5% CO_2 and passaged regularly.

Real-Time Polymerase Chain Reaction

The total RNA of cell lines and tissue was extracted by the Trizol method (Invitrogen), and then, the RNA was reverse transcribed by using a reverse transcription kit (Roche) to obtain cDNA; the experimental operation was carried out according to the instructions of Trizol and the reverse transcription kit. RT-PCR was performed according to the instructions. The primer sequence of each m6A regulator is shown in **Supplementary Table S2**.

Identification of Different N6-Methyladenosine Modification Patterns in Soft Tissue Sarcoma Through Consensus Cluster Analysis

The 25 m6A regulators included 15 readers (IGF2BP1, IGF2BP2, IGF2BP3, YTHDF1, YTHDF2, YTHDF3, YTHDC1, YTHDC2, FMR1, HNRNPA2B1, HNRNPC, RBMX, LRPPRC, ELAVL1, and EIF3A), eight writers (METTL3, METTL14, WTAP, VIRMA, RBM15, RBM15B, ZC3H13, andCBLL1) and two erasers (FTO and ALKBH5). Different m6A modification patterns in STS were determined by consensus cluster analysis in R software according to the expression of these m6A in TCGA-SARC. Furthermore, the aforementioned process was repeated 1,000 times to obtain a stable clustering effect by using the ConsensusClusterPlus R package (Wilkerson and Hayes, 2010).

Differentially Expressed Gene Analysis, Protein–Protein Interaction Analysis, and Connectivity Map Analysis

Gene signatures of different m6A modification patterns were identified based on DEG analysis using the limma package (Ritchie et al., 2015) in R software. DEGs were also analyzed between normal samples and STSs in GSE17674. The Benjamini–Hochberg method (Storey, 2002) was used here to adjust multiple hypotheses. Adjust $p < 0.050$ and $\log^{FC} > 1$ or $\log^{FC} < -1$ were considered significant. The Connectivity Map used a genome-wide transcriptome system to comprehensively describe the biological status of the disease, physiology, and drug induction and further linked genes, drugs, and pathways (Lamb et al., 2006). The DEGs were further uploaded to the cMap database for drug prediction. $p < 0.05$ indicated statistical significance. In addition, 25 m6A regulators were used for PPI analysis and further visualized by Cytoscape (Shannon et al., 2003). The confidence level was 0.4.

Characteristics of the Soft Tissue Sarcoma Microenvironment Based on Different N6-Methyladenosine Modification Patterns

CIBERSORT algorithm was used to evaluate the immune infiltration of TCGA-SARC (Chen et al., 2018). The permutations of the signature matrix were 1,000. The immune and stromal scores of STSs were evaluated by the ESTIMATE package (Yoshihara et al., 2013).

Functional Enrichment Analysis

To further explore the differences in enrichment pathways among different m6A modifications, all genes from TCGA-SARC were used for gene set enrichment analysis (GSEA) based on different m6A modification patterns (A and B). Moreover, the clusterProfiler package (Yu et al., 2012) was used to screen significant pathways using Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) based on 25 m6A regulators and m6A modification-related DEGs, respectively. False discovery rate <0.05 and $p < 0.05$ were considered significant.

Construction of the m6A Score

In order to quantify the modification patterns of m6A in STS, we constructed an m6A score learning from a previous study (Sotiriou et al., 2006). The specific procedures were as follows: first, the consensus cluster analysis was used to divide patients into several clusters according to DEGs between different m6A modification patterns; second, the prognostic DEGs were screened based on univariate Cox regression analysis; and finally, after z-score normalization, principal component analysis (PCA) was used to construct m6A score based on prognostic DEGs using principal component 1 as the signature score. The formula of the m6A score is shown as follows:

$$m6Ascore = \sum pc1_m - \sum pc1_n,$$

where m represents the expression of prognostic DEGs with hazard ratio (HR) < 1, while n represents the expression of prognostic DEGs with HR > 1.

Statistical Analysis

Statistical Product and Service Solutions software (SPSS 22.0) and R 3.6.2 were used for data analysis. The Maftool package (Mayakonda et al., 2018) was used to display the mutation landscape in TCGA-SARC, while the RCircos package (Krzywinski et al., 2009) was used to show the variation of 25 m6A regulators on human chromosomes. Pearson correlation analysis was performed using the corrplot package (https://cran.r-project.org/web/packages/corrplot/index.html) to assess the relationship among different m6A regulators and different immune cells, respectively. Cox regression analysis (Harrell et al., 1996) was performed along with Kaplan–Meier curve analysis to identify the prognostic m6A regulators and DEGs, respectively. For the survival analysis, a survival package was used and a cut-off point was set using the survminer package (Ranstam and Cook, 2017). Furthermore, different datasets were separately divided into different groups based on low and high m6A score, and prognostic differences were explored. All heatmaps were

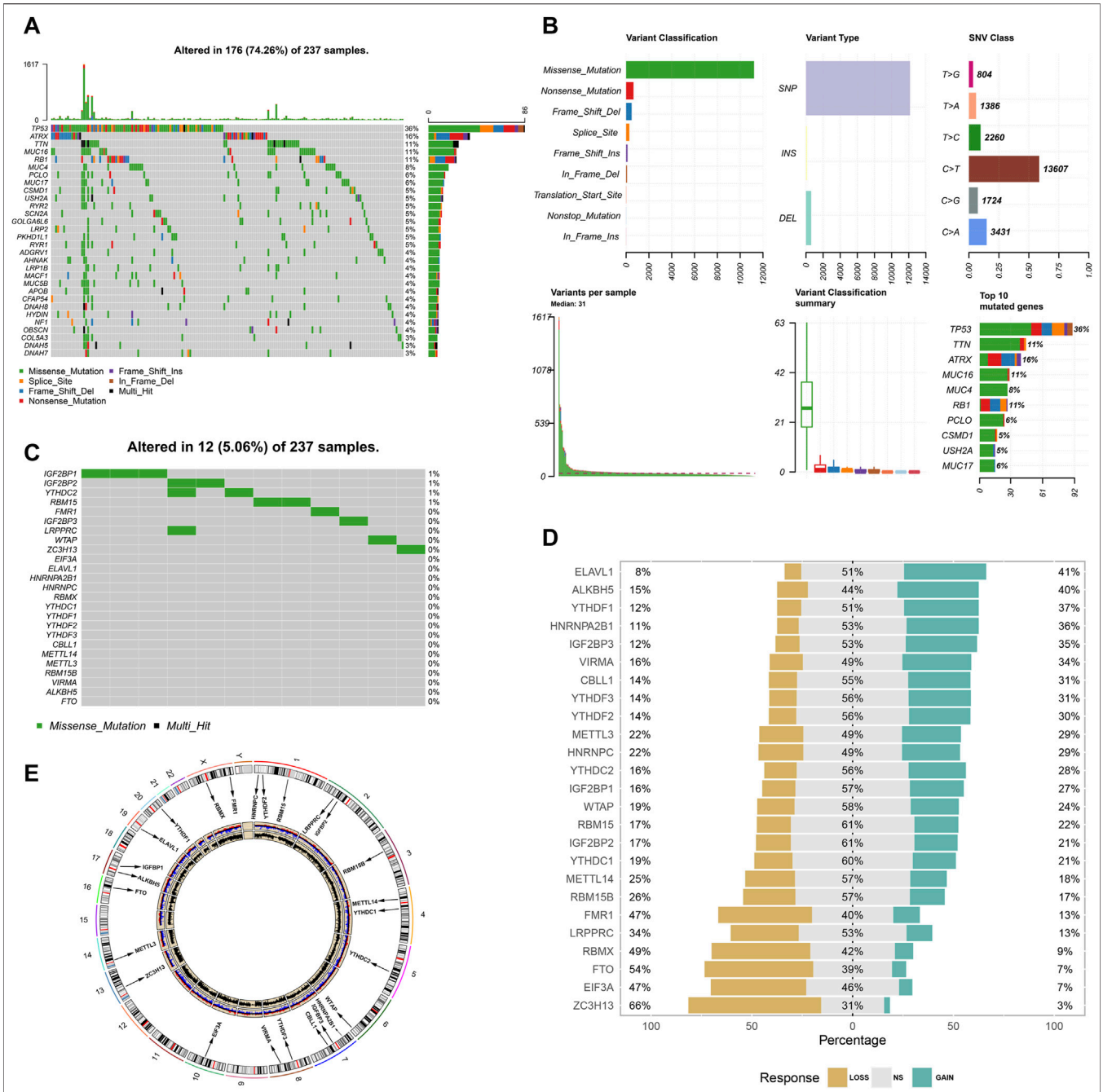


FIGURE 1 | Landscape of somatic mutations and CNV of m6A regulators in STS. **(A)** Summary of somatic mutations in STS; **(B)** variant classifications of mutations in STS; **(C)** summary of somatic mutations of m6A regulators in STS; **(D)** CNV of 25 m6A regulators in STS; and **(E)** locations of different m6A regulators in human chromosomes.

shown using the Pheatmap package (Wang et al., 2014) in R software. The survival rate was compared by the logrank test. Meanwhile, the receiver operating characteristic (ROC) curve for predicting the prognosis of TCGA-SARC and GSE17618 was performed by using the timeROC package (Blanche et al., 2013). PCA was performed using the FactoMineR package (Lê et al., 2008). The Kruskal–Wallis test was performed to compare differences between groups. All significance levels were $p < 0.05$.

RESULTS

Landscape of N6-Methyladenosine Variation in Soft Tissue Sarcoma

Somatic mutations in TCGA-SARC are shown in **Figure 1A**. Of the 237 samples, 176 were detected to have somatic mutations, accounting for 76.3% of the total. Furthermore, the mutation frequency of TP53, ATRX, and TTN was 36, 16, and 11%,

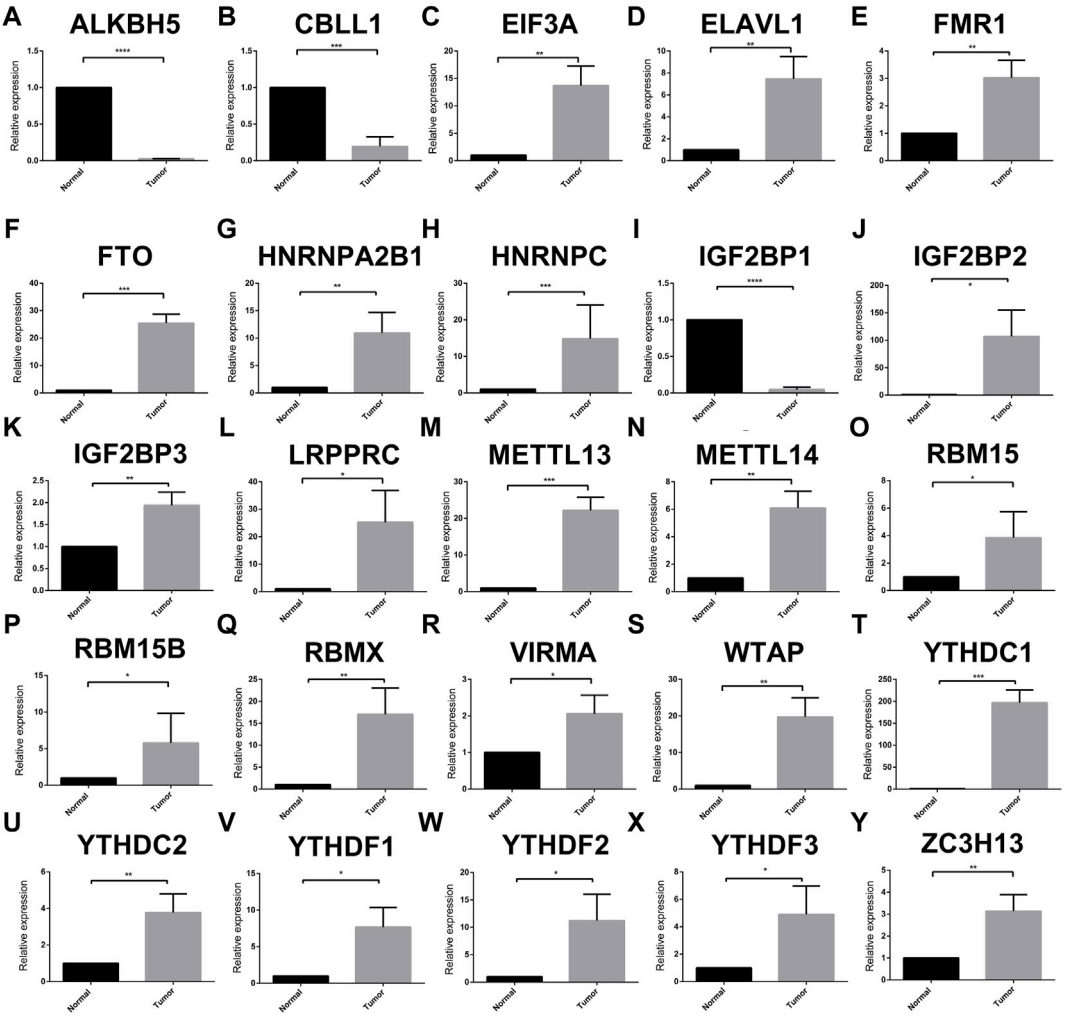


FIGURE 2 | Expression of 25 m6A regulators between STS and normal adjacent tissue. **(A–Y)** Expression of different m6A regulators between normal adjacent tissue and STS samples. $p < 0.05$*, $p < 0.01$**, $p < 0.001$***, and $p = 0$****.

respectively. The summary of variant classification and type is also shown in **Figure 1B**. The major variant classification, type, and single-nucleotide variant type were missense mutation, single-nucleotide polymorphism, and C-T transition, respectively. Considering the widespread somatic variation that existed in STS, the somatic mutations of m6A regulators are also shown in **Figure 1C**. Among 237 samples, only 12 samples had m6A mutations with low mutation frequency. The CNV alteration of 25 m6A regulators is also shown in **Figure 1D**. CNV generally occurred in every m6A regulator. Among them, ZC3H13 (66%), EIF3A (47%), FTO (54%), RBMX (49%), and FMR1 (47%) were found to have higher frequency of CNV gain. ELAVL1 (41%), ALKBH5 (40%), YTHDF1 (37%), HNRNPA2B1 (36%), and IGF2BP3 (35%) had higher frequency of CNV loss. **Figure 1E** also displays the location of different m6A regulators. We also explored the expression of these regulators at the cellular and tissue levels (**Figure 2** and **Supplementary Figure S2**). The result turned out that a wide difference in m6A expression existed in STS. Compared with normal samples, ALKBH5, CBLL1, and IGFBP1 showed lower expression while most of the m6A regulators showed high expression. Considering huge differences in expression among the m6A regulators, we further compared the difference in m6A expression between normal samples and STSs in GSE17674, and the results were consistent with the abovementioned trend (**Figure 3A**). PPI was used to show interactions between m6A regulators, and the result isshown in **Figure 3B**. These m6A regulators were well connected to each other. The correlation plot of each m6A regulator is also displayed in **Figure 3C**. Most of the m6A regulators were correlated with each other, which was consistent with the result of PPI. In addition, univariate Cox analysis indicated that IGF2BP1 ($p < 0.001$), IGF2BP2 ($p = 0.001$), IGF2BP3 ($p = 0.003$), YTHDF2 ($p < 0.001$), HNRNPA2B1 ($p = 0.002$), HNRNPC ($p = 0.002$), RBMX ($p = 0.002$), and VIRMA ($p = 0.044$) were significantly correlated with the prognosis of STS. The abovementioned result is also shown in **Figure 3D**.In addition, , GO and KEGG pathways based on 25 m6A regulators were analyzed, and the results are shown in

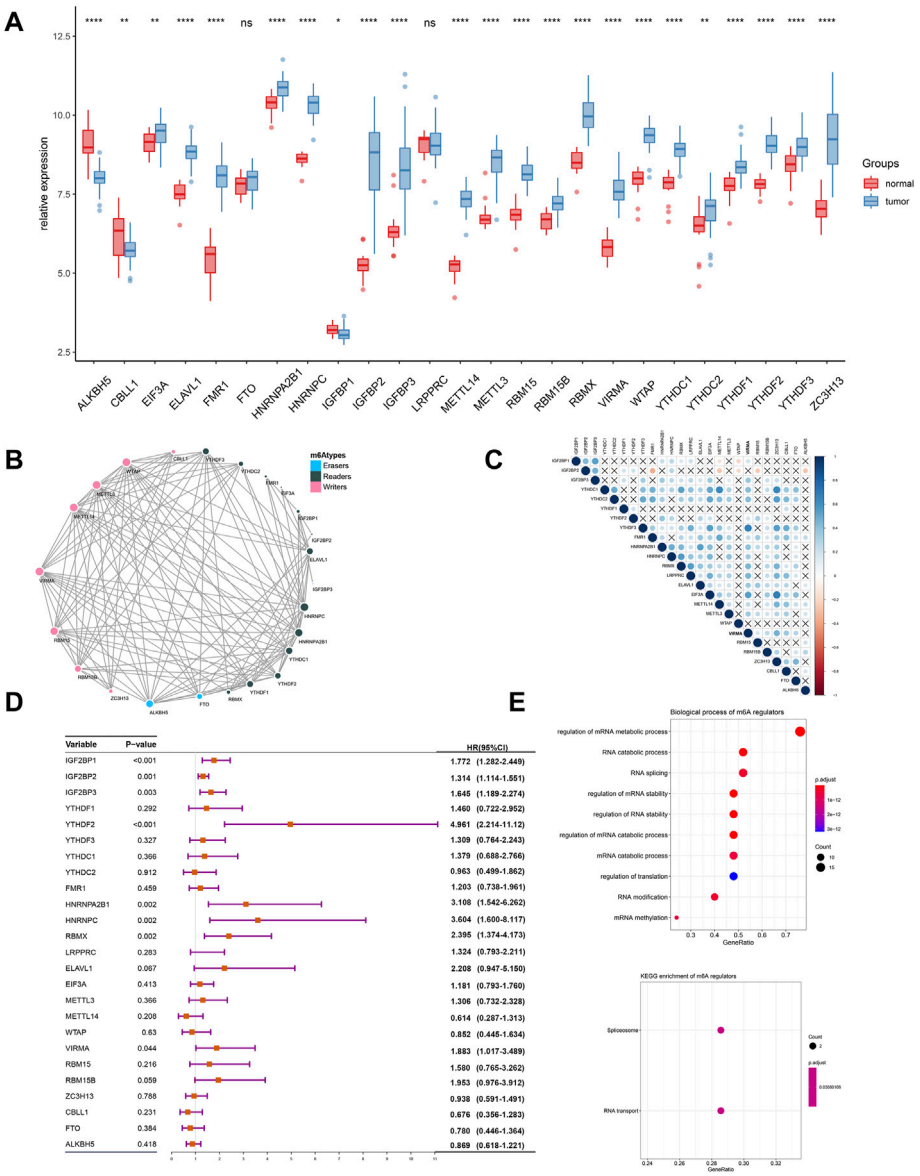


FIGURE 3 | Expression, interactions, prognosis, and functional annotations of 25 m6A regulators in STS. **(A)** Expression of different m6A regulators between normal samples and STS samples using the GSE17674 dataset; **(B)** PPI analysis of 25 m6A regulators; and **(C)** correlation plot among 25 regulators using Pearson correlation analysis. p <0.010 indicated statistical significance. **(D)** Univariate Cox regression analysis for 25 m6A regulators in STS samples; **(E)** functional annotations for 25 m6A regulators. p <0.05*, p <0.01**, p <0.001***, and p = 0****; ns, no significance.

Figure 3E. These genes were enriched in RNA modification-associated pathways. Since widespread variation and expression of m6A regulators existed in STS, m6A regulators might play an important role in the progression and prognosis of STS.

Identification of Two N6-Methyladenosine Modification Patterns in Soft Tissue Sarcoma

To further analyze the effect of m6A modification on STS, we performed a consensus cluster analysis in TCGA-SARC. Two m6A modification patterns (A and B) were identified, and the result of clustering is shown in **Figure 4A**. The corresponding cumulative distribution function plot and delta area plot for clustering are also shown in **Supplementary Figures S3A, B**. m6A cluster A had 140 STS samples, while m6A cluster B had 125 samples. The heatmap of m6A expression among different modification patterns is also shown in **Figure 4B**. Among them, FMR1 (p < 0.001), HNRNPC (p = 0.001), IGF2BP1 (p < 0.001), IGF2BP2 (p < 0.001), IGF2BP3 (p < 0.001), YTHDC1 (p < 0.001), YTHDC2 (p < 0.001), YTHDF1 (p = 0.006), YTHDF1 (p = 0.013), METTL14 (p < 0.001), METTL3 (p = 0.001), RBM15 (p < 0.001), RBM15B (p = 0.010), WTAP (p = 0.010), and ALKBH5 (p = 0.010) were significantly differentially expressed in m6A cluster A and B. As in **Figure 4C**, the result of

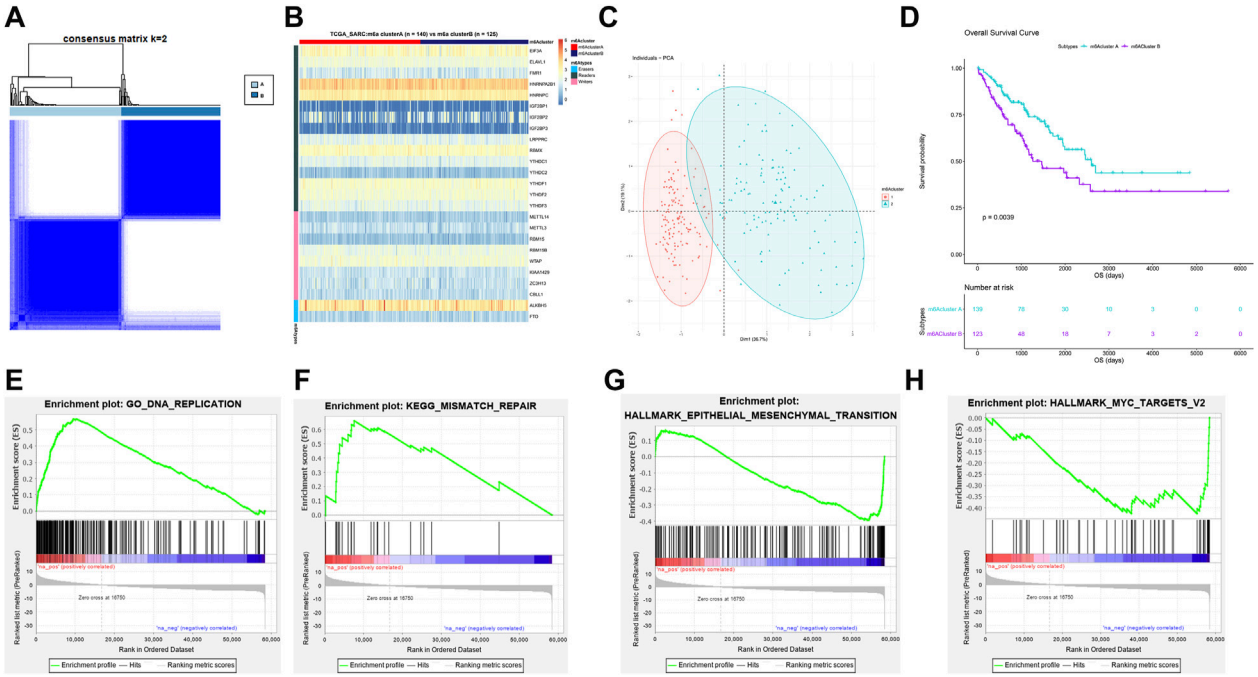


FIGURE 4 | Identification of two m6A modification patterns in STS. **(A)** The result of consensus clustering analysis in STS; **(B)** heatmap of expression of 25 m6A regulators in m6A clusters A and B; **(C)** the result of PCA, indicating two distinct clusters were identified; **(D)** survival plot of two clusters in TCGA-SARC ($p = 0.0039$); **(E)** enriched pathways in m6A cluster A: DNA replication; **(F)** enriched pathways in m6A cluster A: mismatch repair; **(G)** enriched pathways in m6A cluster B: epithelial–mesenchymal transition; and **(H)** enriched pathways in m6A cluster B: MYC target.

PCA also indicated that the consensus cluster well differentiated m6A cluster A and m6A cluster B. The result of survival analysis based on m6A clusters A and B is subsequently shown in **Figure 4D**. m6A cluster A had a significantly improved prognosis of STS than m6A cluster B ($p = 0.004$). The Cox regression analysis also indicated that different m6A clusters were correlated with the prognosis of STS ($p = 0.004$). The 5-year survival rate of m6A cluster A (63.5%) was significantly better than that of m6A cluster B (46.1%) with $p = 0.038$. The GSEA was also performed, and the result is shown in **Figures 4E–H**. Pathways associated with better prognosis were significantly enriched in m6A cluster A including DNA replication [false discovery rate (FDR) = 0; enrichment score (ES) = 0.721] and mismatch repair (FDR = 0; ES = 0.662). In contrast, pathways including epithelial–mesenchymal transition (EMT, $p = 0$, FDR = 0.001, ES = −0.390) and MYC signaling pathway ($p = 0$, FDR = 0.003, ES = −0.360) were enriched in m6A cluster B, which often led to poorer outcome of STS. Full lists of enriched pathways are shown in **Supplementary Tables S3, S4**. The aforementioned analysis revealed that different m6A modifications were associated with different pathways, which further affected the prognosis of STS.

Characteristics of N6-Methyladenosine Modification Patterns in the Soft Tissue Sarcoma Microenvironment

In order to understand the influence of m6A modification patterns on the STS microenvironment, the immune cell infiltration in TCGA-SARC is shown in **Figure 5A**. The correlation plot of each immune cell is also displayed in **Figure 5B**. CD8 T cells and follicular helper T cells had higher correlation with other cells. The relationships between 22 kinds of immune cells and different m6A modification patterns were analyzed by the Kruskal–Wallis test, respectively (**Figure 5C**). m6A cluster A showed higher infiltration of M1 macrophage, CD8 T cell, and NK cell, while the M2 macrophage showed higher infiltration in m6A cluster B. M2 macrophage ($p < 0.050$), mast cell activated ($p < 0.050$), mast cell resting ($p < 0.0001$), neutrophils ($p < 0.050$), and T cell CD4 memory activated ($p < 0.010$) were significantly correlated with different m6A modification patterns in STS. In addition, the immune and stromal scores of m6A cluster A/B were separately calculated, and the results are shown in **Figures 6A,B**. Compared with m6A cluster A, m6A cluster B was characterized by a significantly higher immune score ($p < 0.050$) and stromal score ($p < 0.010$). Survival analysis of different m6A modification patterns in different immune scores and stromal scores was subsequently performed, and the results are shown in **Figures 6C,D**. Significant prognostic differences were found in different m6A modification patterns and immune/stromal score ($p < 0.0001$). The Cox regression analysis also showed that different m6A modification patterns were significantly correlated with the prognosis of STS in different immune scores ($p = 0.001$) and stromal scores ($p < 0.001$), respectively. In addition, differences in the microenvironment between m6A cluster A/B and normal adjacent tissue were also explored in GSE17674. As shown in **Supplementary Figure S4A**, consensus clustering analysis also

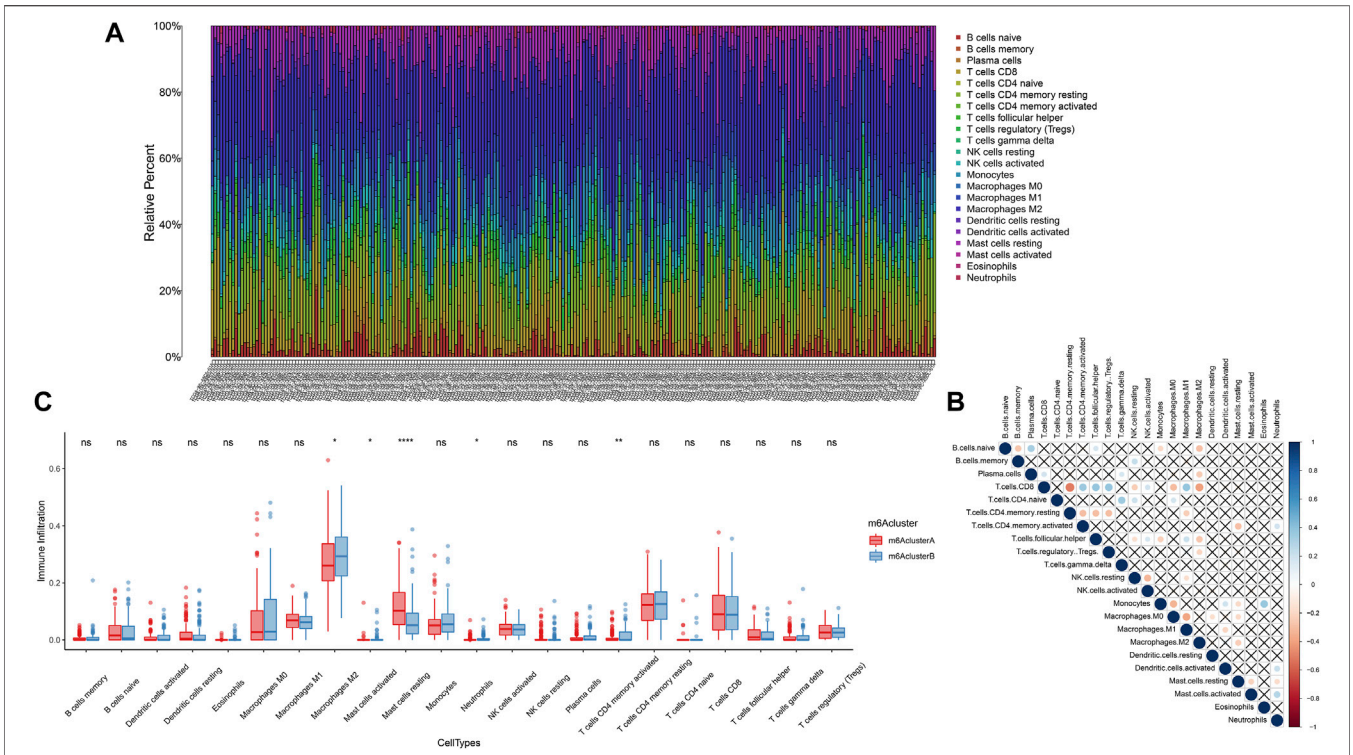


FIGURE 5 | Effects of different m6A modification patterns on the immune infiltration of STS. **(A)** Immune infiltration of each sample in TCGA-SARC. **(B)** Correlation plot of each immune cell in TCGA-SARC. **(C)** Comparison of immune cell infiltration among m6A clusters A and B.

identified two distinct m6A modification patterns. The immune score of both m6A clusters A and B was higher than that of normal tissue with $p = 0.159$ and $p = 0.034$, respectively (**Supplementary Figures S4B,C**). The aforementioned results suggested that different m6A modification patterns affected the immune infiltration in the STS microenvironment, which further affected the prognosis of STS.

Construction of m6A Score

A total of 328 DEGs between different m6A modification patterns were determined. The full list of 328 DEGs is shown in **Supplementary Table S5**. Among them, 227 genes were upregulated in m6A cluster A while 101 genes were upregulated in m6A cluster B (**Figure 7A**). The GO and KEGG enrichment analysis was also performed, and the results are shown in **Supplementary Tables S6, S7**. Similar to the result of m6A regulators, these DEGs were enriched in the p53 signaling pathway ($p = 0.002$), ECM–receptor interaction ($p = 0.005$), and PI3K-Akt signaling pathway ($p = 0.023$), which were closely related to the progression of STS. The heatmap of these DEGs in STS is also shown in **Figure 7B**. Cmap was analyzed according to these 328 DEGs, and the results are shown in **Supplementary Table S8**. Imatinib ($p = 0.001$) and furazolidone ($p = 0.025$) were regarded as important targets treating STS. Univariate Cox regression analysis was further performed to identify prognostic genes among these 328 DEGs. Of the 328 DEGs, 90 of them were identified as prognostic genes ($p < 0.050$). Furthermore, 58 genes were identified as protective genes due to HR < 1, while 32 genes were regarded as risk genes based on HR > 1. The full list of prognostic DEGs is displayed in **Supplementary Table S9**. The relationships between DEGs and these prognostic genes are visualized in a Sankey diagram in **Figure 7C**. To our surprise, all protective DEGs, upregulated in m6A cluster A, were from m6A cluster A. Contrary to m6A cluster A, all risk DEGs, upregulated in m6A cluster B, belonged to m6A cluster B. This was also consistent with the result of the survival analysis. To further evaluate the stability of m6A modification patterns, consensus cluster analysis was performed based on 328 DEGs, and two distinct gene clusters A and B were identified (**Figure 7D**). The corresponding cumulative distribution function plot and delta area plot for clustering are also shown in **Supplementary Figures S5A, B**. Among them, 92.9% samples of gene cluster A were from m6A cluster A while 66.7% samples of gene cluster B were from m6A cluster B, which also implied that the identification of m6A modification patterns was relatively stable. The survival plot between different gene clusters is also shown in **Figure 7E**. Gene cluster A had a better prognosis of STS than gene cluster B ($p = 0.027$), which was also consistent with the survival result of m6A clusters. The Cox regression analysis indicated that different gene clusters were significantly correlated with the prognosis of STS ($p = 0.029$). The 5-year survival rates of gene clusters A and B were 66.0 and 50.0%, respectively. Due to the significant difference in immune infiltration and prognosis between m6A modification patterns, m6A score was constructed to quantify the modification pattern

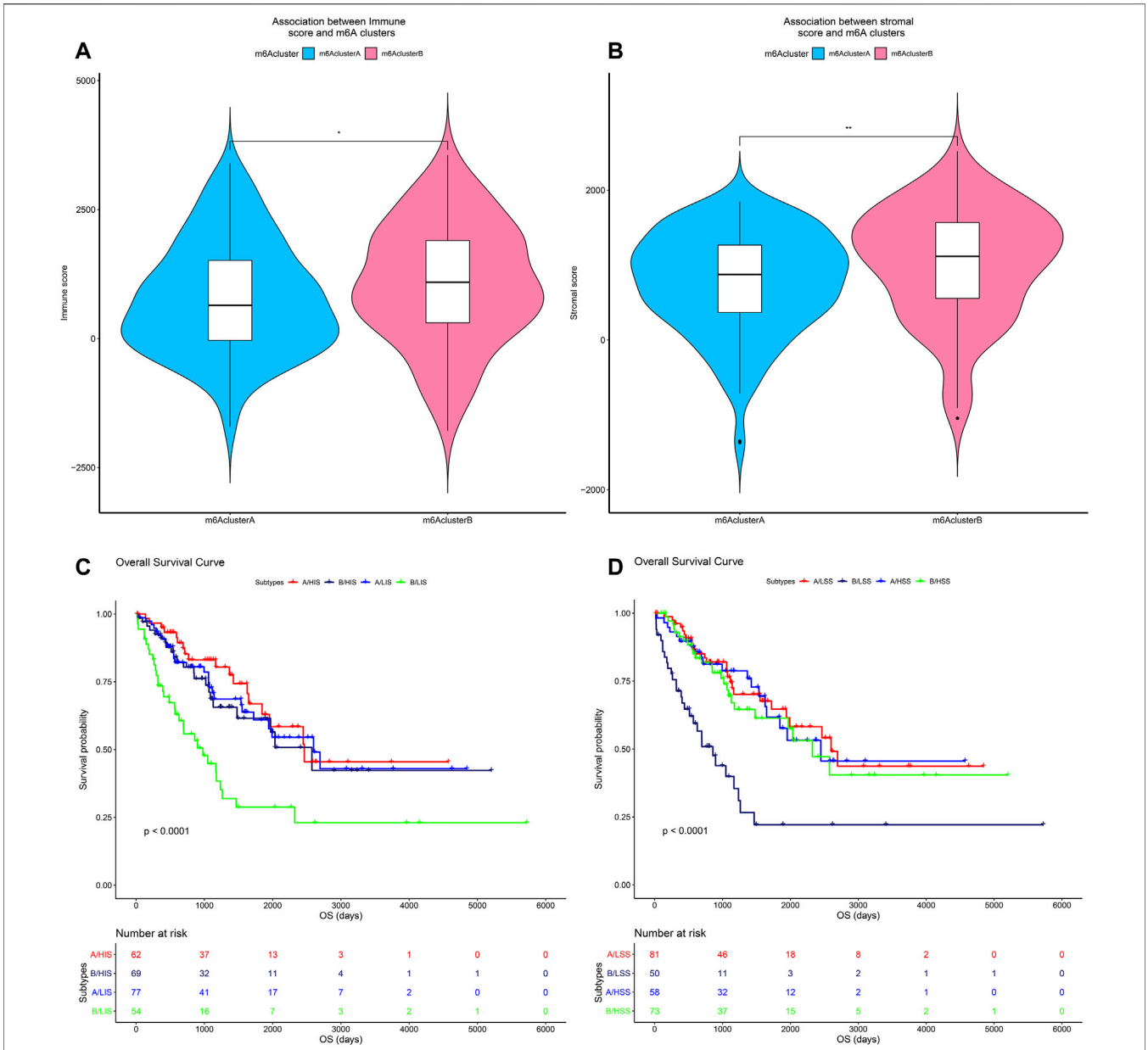


FIGURE 6 | Effects of different m6A modification patterns on the immune infiltration of STS. **(A)** Comparison of immune score among m6A clusters A and B; **(B)** Comparison of stromal score among m6A clusters A and B; **(C)** survival analysis of different immune scores among m6A clusters A and B (A: m6A cluster A, B: m6A cluster B, LSS, low immune score, and HSS, high immune score); and **(D)** survival analysis of different stromal scores among m6A clusters A and B. (A: m6A cluster A, B: m6A cluster B, LIS: low stromal score, and HIS, high stromal score); p <0.05* and p <0.01**.

of each STS sample. Then, STS samples were divided into high- and low-m6A score groups. Similar to what we found in gene clusters, 75.7% samples of high m6A score were from m6A cluster A while 78.4% samples of low m6A score were from m6A cluster B. The survival plot between high and low m6A score is also shown in **Figure 7F**. A high m6A score had a better prognosis of STS than a low m6A score ($p < 0.0001$). The Cox regression analysis indicated that different m6A scores were significantly correlated with the prognosis of STS ($p < 0.001$). The m6A score was also compared in different m6A clusters and gene clusters, and the result is shown in **Figures 7G,H**. The high-m6A score group showed a significantly better outcome than the low-m6A score group ($p < 0.0001$), and the result was the same in gene cluster A ($p < 0.001$). In addition, in **Supplementary Figure S6**, the area under the curve (AUC) of m6A score for prediction of the 1-, 3-, and 5-year survival of STS was 0.77, 0.71, and 0.68, respectively. Finally, a Sankey diagram was performed to summarize the correlation between m6A clusters, gene clusters, m6A score, and prognosis of STS (**Figure 7I**). It could be clearly seen that m6A cluster A, gene cluster A, and

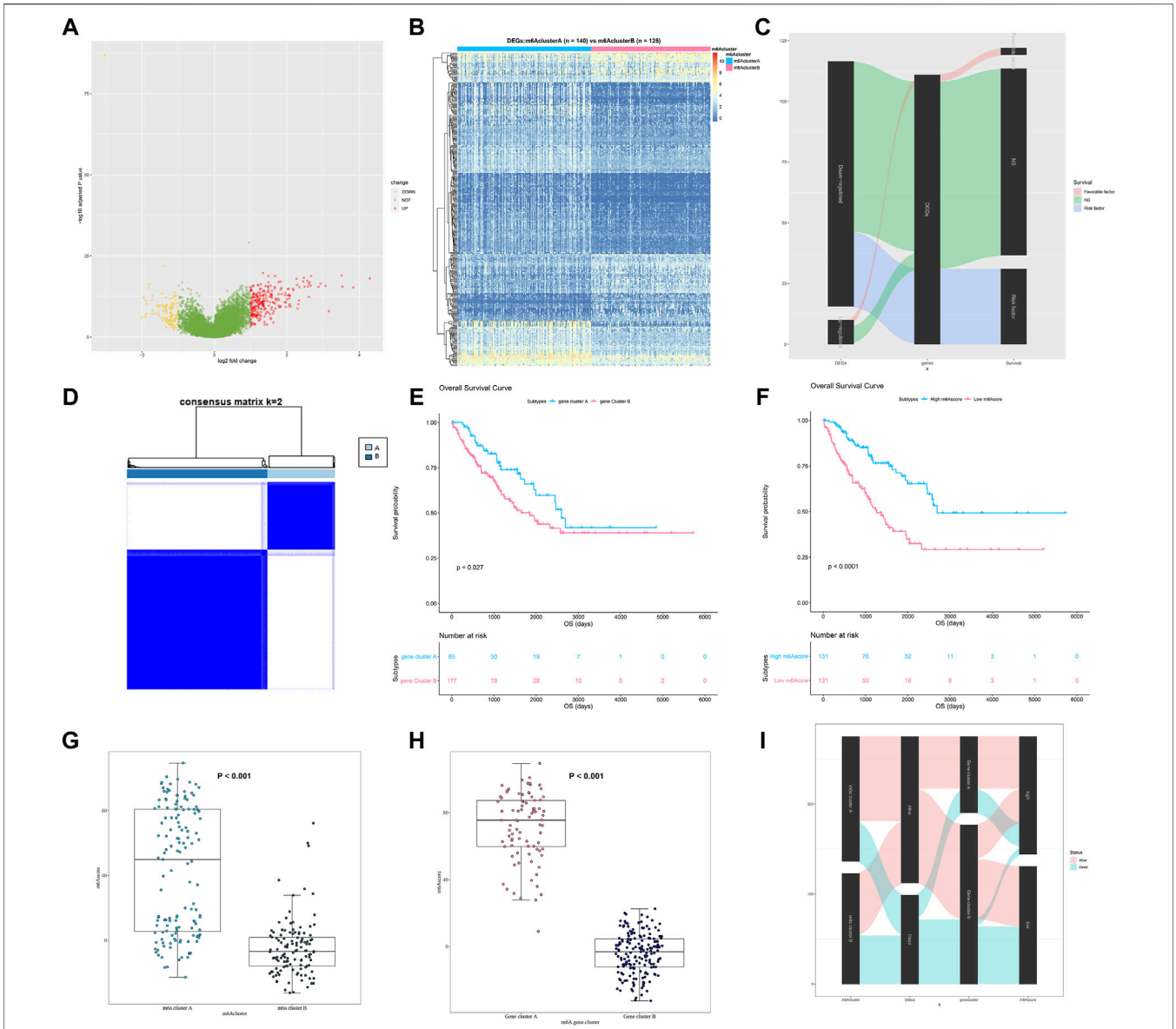


FIGURE 7 | Construction of m6Ascore. **(A)** Volcano plot of DEGs between m6A cluster A and B; **(B)** the heatmap of the expression of DEGs in m6A cluster A and B; **(C)** the relationship between DEGs and these prognostic genes visualized as a Sankey diagram; **(D)** the result of consensus clustering analysis in STS based on 328 DEGs; **(E)** survival plot of gene clusters A and B in TCGA-SARC ($p = 0.027$); **(F)** survival plot of high and low m6A score in TCGA-SARC ($p < 0.0001$); **(G)** comparison of m6A score among m6A clusters A and B; **(H)** comparison of m6Ascore among gene clusters A and B; and **(I)** the relationship between m6A clusters, gene cluster survival status, and m6A score visualized as a Sankey diagram.

high m6A score were correlated with a better prognosis of STS while m6A cluster B, gene cluster B, and low m6A score were related to a poorer prognosis of STS. The aforementioned results indicated that the m6A score could accurately distinguish the m6A modification patterns and predict the prognosis of STS.

m6A score Could Predict Immunotherapy Response

Considering that m6A score could well predict the prognosis of STS, we further explore whether m6A score could effectively predict the response to ICB in IMvigor210 and GSE78220 datasets. For the IMvigor210 dataset, 298 samples were divided into high-m6A score ($n = 243$) and low-m6A score groups ($n = 55$). The result of the survival analysis turned out that a high m6A score showed a better prognosis than a low m6A score (**Figure 8A**, $p = 0.036$). The corresponding Cox regression analysis also revealed that different m6A scores were significantly correlated with the prognosis of each sample in IMvigor210 ($p = 0.037$). Furthermore, the relative percent of complete response (CR), partial response (PR), progression disease (PD), and stable disease (SD) in high- and low-m6A score groups were compared, and the results are shown in

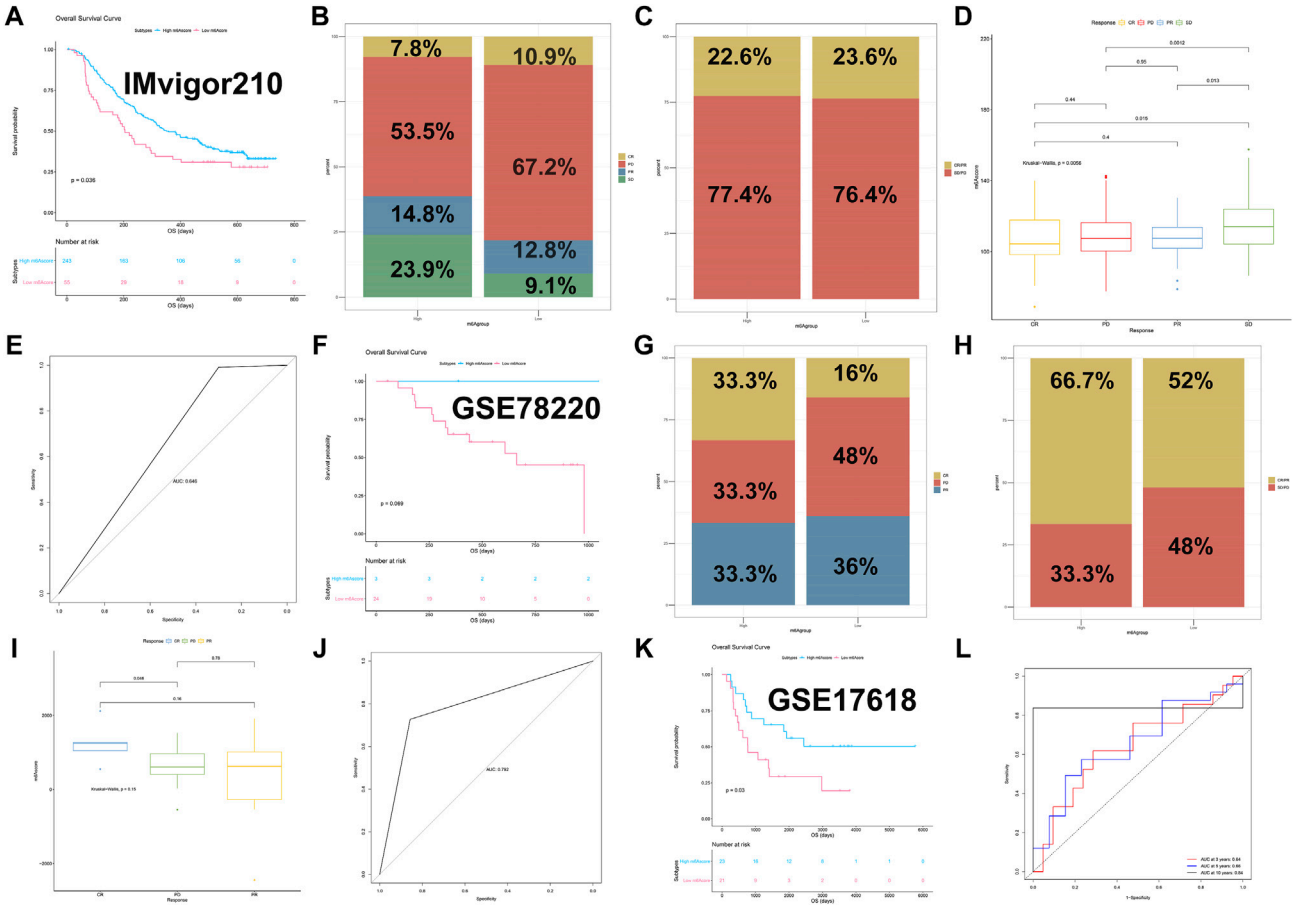


FIGURE 8 | The m6Ascore predicted immune response in immunotherapy datasets and validation of m6Ascore. **(A)** Survival plot of high and low m6A score in the IMvigor210 dataset (p = 0.036); **(B)** relative percent of immune responses in high and low m6A score in the IMvigor210 dataset; CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease; **(C)** relative percent of CR/PR and SD/PD in high and low m6A score in the IMvigor210 dataset; CR, complete response, PR, partial response, SD, stable disease, PD, progressive disease; **(D)** box plots of m6Ascore in different immune responses in the IMvigor210 dataset; **(E)** ROC curve for prediction of immune response in the IMvigor210 dataset; **(F)** survival plot of high and low m6A score in the GSE78220 dataset (p = 0.069); **(G)** relative percent of immune responses in high and low m6A score in the GSE78220 dataset; CR, complete response; PR, partial response; PD, progressive disease; **(H)** relative percent of CR/PR and SD/PD in high and low m6A score in the GSE78220 dataset; CR, complete response; PR, partial response; PD, progressive disease; **(I)** box plots of m6A score in different immune responses in the GSE78220 dataset; **(J)** ROC curve for prediction of immune response in the GSE78220 dataset; **(K)** survival plot of high and low m6A score in the GSE17618 dataset (p = 0.030); and **(L)** ROC curve for prediction of the 3-, 5-, and 10-year survival of STS in the GSE17618 dataset.

Figures 8B–D. 53.4% of the high-m6A score group had PD, while 67.2% of the low-m6A score group had PD, and patients with SD had higher m6A score than patients with PD (p = 0.001). The ROC analysis for the prediction of the response to anti-PD-L1 is shown in **Figure 8E**, and the AUC was 0.646. For the GSE78220 dataset, 27 samples were divided into high-m6A score (n = 3) and low-m6A score groups (n = 24). Survival analysis was further performed between different m6A score groups. Similar to IMvigor210, the high-m6A score group showed a better prognosis than the low-m6A score group (p = 0.069, **Figure 8F**). The corresponding Cox regression analysis also revealed that different m6A scores were associated with the prognosis of each sample in GSE78220 (p = 0.018). The relative percent of CR, PR, and PD in the high-m6A score group were all 33.3%, while in the low-m6Ascore group, the proportion was 16.0, 48.0,, and 36%, respectively (**Figures 8G,H**). The m6A score did not show a significant difference among samples with different immunotherapy responses (**Figure 8I**), which could be due to the small sample size of GSE78220. The ROC curve for prediction of the response to anti-PD-1 is shown in **Figure 8J**, and the AUC was 0.792.In addition, the independent sarcoma dataset GSE17618 was used for the validation of the m6Ascore. GSE171618 dataset was divided into high- (n = 23) and low-m6A score groups (n = 21). As in **Figure 8K**, the high-m6A score group showed a better prognosis than the low-m6A score group (p = 0.030). The corresponding Cox regression analysis also indicated that different m6A scores were significantly correlated with the prognosis of STS in GSE17618 (p = 0.034). The 5-year survival rates of the high- and low-m6A score groups were 55.9 and 29.3%, respectively. The AUC for prediction of the 3-, 5-, and 10-year survival of STS was 0.64, 0.66, and 0.84, respectively (**Figure 8L**). The event-free survival curve between high and low m6A score is also displayed in **Supplementary Figure S7**. Similar to the result of overall

survival, the high-m6A score group had a better prognosis (p = 0.074). The corresponding Cox regression analysis also indicated that different m6A scores were significantly correlated with the event-free survival of STS in GSE17618 (p = 0.070). In general, our study revealed a non-negligible role of m6A modification in the STS microenvironment, which could well predict the response to PD-1/PD-L1 immunotherapy.

DISCUSSION

Due to diverse histopathological classification, STS has not been fully understood so far, which received widespread attention (Trojani et al., 1984). A previous study constructed an m6A-related risk model to predict the prognosis of STS (Hou et al., 2020). However, different m6A modification patterns were not analyzed comprehensively. Recent studies also revealed the significant role of m6A modification in gastric cancer and pancreatic cancer (Zhou et al., 2021). However, systematic analysis of the m6A modification in STS was still rare.

The CNV analysis indicated that ZC3H13 and ELAVL1 had the highest variant frequency. A recent study also indicated that ZC3H13 could inhibit colorectal cancer through Ras-ERK pathways (Zhu et al., 2019). Moreover, ELVAL1 knockout mice were also found to have lower tumor growth in a previous report (Chang et al., 2014). Therefore, our findings suggested that these regulators might also play a key role in STS progression. Besides, 96% of m6A regulators were found to have significantly different expressions between STS and normal samples, which was consistent with a former study that m6A regulators expressed differently in colorectal cancer based on bioinformatics analysis (Zhang et al., 2021a). These results also suggested that m6A regulation might exist in various types of cancer. Univariate Cox regression analysis showed that IGF2BP1 and YTHDF2 were top significantly correlated with the poor prognosis of STS. Previous studies also found that IGFBP2 could accelerate the migration of tumor cells by regulating LEF1 and SNAI2 (Zirkel et al., 2013). Moreover, IGF2BP1 was also regarded as a risk factor in neuroblastoma (Bell et al., 2015). YTHDF2 was also related to the poor prognosis of glioma in Lin et al.'s (2020) research, and the latest research implied that stabilized YTHDF2 could enhance the growth capacity of glioblastoma (Fang et al., 2021). These studies were consistent with our results, which also revealed the potential prognostic value of IGFBP1 and YTHDF2 in STS.

Furthermore, two m6A modification patterns in STS were identified. m6A cluster A was correlated with DNA replication, mismatch repair, and a better prognosis, while m6A cluster B was related to MYC, EMT signaling pathways, and a poor prognosis. As an important part of the cell cycle, stable DNA replication was key to the normal activities of cells. Dysfunction of DNA replication might result in the occurrence of diseases. For example, abnormal replication of MCM10 would lead to NK cell deficiency (Mace et al., 2020). A recent study also reported that DNA replication stress could be used to enhance the antitumor ability in squamous cell carcinoma (Zhang et al., 2021b). In addition, mismatch repair (MMR) was also known for its tumor suppressor function. For instance, dysfunction of MMR genes would lead to Lynch syndrome, which was susceptible to cancer including ovarian cancer (Zhang et al., 2021b). MMR deficiency was also related to the occurrence of endometrial cancer (McDougal et al., 2021). The MYC gene was a proto-oncogene, which has been studied for many decades (Cole and McMahon, 1999). A recent study showed that MYC could be regulated by USP16, which further inhibited prostate cancer progression (Ge et al., 2021). MYC was also found to participate in angiogenesis (Meškytė et al., 2020). EMT was considered to be related to tumor metastasis. A previous study reported that EMT induced by HOXA10 contributed to gastric cancer metastasis (Song and Zhou, 2021). These studies supported our results, which provided novel insights into the role of m6A modification in STS. The immune cell infiltration analysis showed that M1 macrophage, CD8 T cell, and NK cell were enriched in m6A cluster A while M2 macrophage was enriched in m6A cluster B. M1 macrophages were generally thought to inhibit tumor growth while M2 macrophages were found to promote tumor progression, which was also proved in many studies. For example, M1 macrophage could inhibit colon cancer growth (Engström et al., 2014). In Jackute et al.'s (2018) study, low infiltration of M2 macrophages was correlated with an improved prognosis of lung cancer. Improved CD8 T cell infiltration was also considered to be correlated with a better prognosis (Li et al., 2021b). As a part of innate immunity, NK cells also played an anti-tumor role. Recent research reported that NK cells could inhibit lung tumor growth in mice models (Yamamoto et al., 2018).In addition, both m6A clusters A/B showed higher immune scores than normal tissue. This may be due to the local inflammatory and immune response of tumor tissue, which recruits more immune cells, while there is no tumor-mediated immune cell aggregation in normal tissue. For example, a recent study revealed that macrophages could be recruited for breast cancer by increasing CCL2 (Wolfsberger et al., 2021). A former study also indicated that many neutrophils were recruited in non-small-cell lung cancer (Mollaoglu et al., 2018). Our findings were consistent with these studies, indicating that the m6A cluster A was related to a strong anti-tumor immune response while m6A cluster B was related to a suppressed anti-tumor immune response.

The result of Cmap analysis showed imatinib and furazolidone as important drugs for treating STS. A recent study illustrated that furazolidone could induce apoptosis in lung cancer by downregulating NF-kappa B (Yu et al., 2020). Furazolidone was also found to prevent the growth of hepatoma cells by enhancing reactive oxygen species (Sun et al., 2015). Furazolidone was also found to be a potential drug for the treatment of acute myeloid leukemia by enhancing the expression of p53 (Jiang et al., 2013). Imatinib was found to improve the prognosis of gastrointestinal stromal tumors (Kurtovic-Kozaric et al., 2017). A recent study also implied that imatinib might slow down the progression of leukemia (Druker et al., 2001). Given the important role these drugs played in other tumors, they might also become promising drugs for STS treatment. We further identified two gene clusters based on DEGs between m6A clusters A and B. Similar to our m6A cluster modification patterns, gene cluster A showed a better prognosis of STS, which also validated that our previous m6A modification patterns were reliable. The m6Ascore was significantly higher in m6A cluster A and gene cluster A, respectively. Moreover, the AUC of m6Ascore for predicting the 1-, 3-, and 5-year survival of STS

was 0.77, 0.71, and 0.68. In addition, the survival results of validation dataset GSE17618 also showed that a high m6Ascore was related to a better prognosis, and the AUC for predicting the 3-, 5-, and 10-year survival was 0.64, 0.66, and 0.84, respectively. These results demonstrated that the m6A score could identify different m6A modification patterns and be used as a potential prognostic indicator in STS.

Our results also showed that m6A score could predict the response to immunotherapy. A high m6A score was correlated with a better immunotherapy response. Previous studies focused on suppressing macrophages to obtain a better immunotherapy response (Dong et al., 2021) and adding nanoparticles into the TME to further strengthen the anti-tumor ability (Yang et al., 2021). These studies mainly changed the immunotherapy response by affecting the components of the TME. However, biomarkers to directly predict the response of immunotherapy were still rare. Here, we reported the AUC of m6A score for prediction of immune response to PD-1/L1 was 0.646 and 0.794, respectively. Therefore, our m6A score could also be a promising predictor for tumor immunotherapy. In addition, the m6A score was also validated by an independent sarcoma dataset, which also proved that the m6A score was reliable.

Our research also had some limitations. More immunotherapy datasets were needed to validate the m6A score. In addition, more clinical trials were also needed to further validate the drugs and cancer-related pathways we screened in this study.

CONCLUSION

In general, the m6A modification patterns in the STS microenvironment were comprehensively analyzed. An m6A score to evaluate different m6A modification patterns was established through integrative analysis. A high m6A score showed a better prognosis of STS, while a low m6Ascore led to a poor prognosis of STS. In addition, the m6A score was validated by an independent dataset successfully and accurately predicted the prognosis of STS, which could be a promising predictor for cancer immunotherapy.

DATA AVAILABILITY STATEMENT

The original contributions presented in this 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 board of Zhongnan Hospital of Wuhan University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LC, K-WX, L-YG and Z-BL conceived and designed the study. Z-QY, MY, Z-BL and XY searched databases and analyzed data. Z-QY, K-WX, and XY prepared tables and figures. Z-QY, LC and K-WX wrote the manuscript. LC supervised the study.

FUNDING

This study was funded by the Health Care of Yellow Crane Talent Plan (Project No. 17).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2022.846812/full#supplementary-material

Supplementary Figure S1 | The flowchart of this study.

Supplementary Figure S2 | The expression of 25 m6A regulators between HSMC and A673 cells; (A–Y) the expression of different m6A regulators between HSMC and A673 cells. $p < 0.05$*, $p < 0.01$**, $p < 0.001$***, and $p = 0$****.

Supplementary Figure S3 | The cumulative distribution function plot and delta area plot for clustering in TCGA-SARC. **(A)** The cumulative distribution function plot in TCGA-SARC; **(B)** the delta area plot for clustering in TCGA-SARC.

Supplementary Figure S4 | Exploration of differences in the microenvironment between different m6A modification patterns and normal tissue in GSE17674. **(A)** The result of consensus clustering analysis in GSE17674; **(B)** comparison of immune score between m6A cluster A and normal tissue in GSE17674; and **(C)** comparison of immune score between m6A cluster B and normal tissue in GSE17674.

Supplementary Figure S5 | The cumulative distribution function plot and delta area plot for clustering in TCGA-SARC based on 328 DEGs. **(A)** The cumulative distribution function plot in TCGA-SARC based on 328 DEGs; **(B)** the delta area plot for clustering in TCGA-SARC based on 328 DEGs.

Supplementary Figure S6 | The ROC curve of m6A score for prediction of the 1-, 3-, and 5-years survival of STS in TCGA-SARC.

Supplementary Figure S7 | The event-free survival plot of high and low m6A score in the GSE17618 dataset ($p = 0.074$)

REFERENCES

Alarcón, C. R., Lee, H., Goodarzi, H., Halberg, N., and Tavazoie, S. F. (2015). N6-methyladenosine marks Primary microRNAs for Processing. *NATURE* 519, 482–485. doi:10.1038/nature14281

Bell, J. L., Turlapati, R., Liu, T., Schulte, J. H., and Hüttelmaier, S. (2015). IGF2BP1 Harbors Prognostic Significance by Gene Gain and Diverse Expression in Neuroblastoma. *Jco* 33, 1285–1293. doi:10.1200/JCO.2014.55.9880

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

Brennan, M. F. (2005). Soft Tissue Sarcoma: Advances in Understanding and Management. *The Surgeon* 3, 216–223. doi:10.1016/s1479-666x(05)80044-7

Cao, G., Li, H.-B., Yin, Z., and Flavell, R. A. (2016). Recent Advances in Dynamic M 6 A RNA Modification. *Open Biol.* 6, 160003. doi:10.1098/rsob.160003

Chang, S.-H., Elemento, O., Zhang, J., Zhuang, Z. W., Simons, M., and Hla, T. (2014). ELAVL1 Regulates Alternative Splicing of eIF4E Transporter to Promote Postnatal Angiogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 111, 18309–18314. doi:10.1073/pnas.1412172111

Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., and Alizadeh, A. A. (2018). Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol. Biol. (Clifton, N.J.)*, 243–259. doi:10.1007/978-1-4939-7493-1_12

Chen, H., Yao, J., Bao, R., Dong, Y., Zhang, T., Du, Y., et al. (2021). Cross-talk of Four Types of RNA Modification Writers Defines Tumor Microenvironment and Pharmacogenomic Landscape in Colorectal Cancer. *MOL. CANCER* 20. doi:10.1186/s12943-021-01322-w

Cole, M. D., and McMahon, S. B. (1999). The Myc Oncoprotein: a Critical Evaluation of Transactivation and Target Gene Regulation. *ONCOGENE* 18, 2916–2924. doi:10.1038/sj.onc.1202748

Cui, Q., Shi, H., Ye, P., Li, L., Qu, Q., Sun, G., et al. (2017). m 6 A RNA Methylation Regulates the Self-Renewal and Tumorigenesis of Glioblastoma Stem CellsA RNA Methylation Regulates the Self-Renewal and Tumorigenesis of Glioblastoma Stem Cells. *Cel Rep.* 18, 2622–2634. doi:10.1016/j.celrep.2017.02.059

Delaunay, S., and Frye, M. (2019). RNA Modifications Regulating Cell Fate in Cancer. *NAT. CELL BIOL.* 21, 552–559. doi:10.1038/s41556-019-0319-0

Dong, Y., Zhang, S., Gao, X., Yin, D., Wang, T., Li, Z., et al. (2021). HIF1α Epigenetically Repressed Macrophages via CRISPR/Cas9-EZH2 System for Enhanced Cancer Immunotherapy. *Bioactive Mater.* 6, 2870–2880. doi:10.1016/j.bioactmat.2021.02.008

Druker, B. J., Talpaz, M., Resta, D. J., Peng, B., Buchdunger, E., Ford, J. M., et al. (2001). Efficacy and Safety of a Specific Inhibitor of the BCR-ABL Tyrosine Kinase in Chronic Myeloid Leukemia. *N. Engl. J. Med.* 344, 1031–1037. doi:10.1056/NEJM200104053441401

Engström, A., Erlandsson, A., Delbro, D., and Wijkander, J. (2014). Conditioned media from Macrophages of M1, but Not M2 Phenotype, Inhibit the Proliferation of the colon Cancer Cell Lines HT-29 and CACO-2. *INT. J. ONCOL.* 44, 385–392. doi:10.3892/ijo.2013.2203

Fang, R., Chen, X., Zhang, S., Shi, H., Ye, Y., Shi, H., et al. (2021). EGFR/SRC/ERK-stabilized YTHDF2 Promotes Cholesterol Dysregulation and Invasive Growth of Glioblastoma. *NAT. COMMUN.* 12. doi:10.1038/s41467-020-20379-7

Ge, J., Yu, W., Li, J., Ma, H., Wang, P., Zhou, Y., et al. (2021). USP16 Regulates Castration-Resistant Prostate Cancer Cell Proliferation by Deubiquitinating and Stablizing C-Myc. *J. Exp. Clin. Cancer Res.* 40. doi:10.1186/s13046-021-01843-8

Han, D., Liu, J., Chen, C., Dong, L., Liu, Y., Chang, R., et al. (2019). Anti-tumour Immunity Controlled through mRNA m6A Methylation and YTHDF1 in Dendritic Cells. *NATURE* 566, 270–274. doi:10.1038/s41586-019-0916-x

Hanahan, D., and Coussens, L. M. (2012). Accessories to the Crime: Functions of Cells Recruited to the Tumor Microenvironment. *CANCER CELL* 21, 309–322. doi:10.1016/j.ccr.2012.02.022

Harrell, F. E., Lee, K. L., and Mark, D. B. (1996). Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors. *Statist. Med.* 15, 3612–3874. doi:10.1002/(sici)1097-0258(19960229)15:4<361::aid-sim168>3.0.co;2-4

Hou, M., Guo, X., Chen, Y., Cong, L., and Pan, C. (2020). A Prognostic Molecular Signature of N^6-Methyladenosine Methylation Regulators for Soft-Tissue Sarcoma from the Cancer Genome Atlas Database. *Med. Sci. Monit.* 26. doi:10.12659/MSM.928400

Hugo, W., Zaretsky, J. M., Sun, L., Song, C., Moreno, B. H., Hu-Lieskovan, S., et al. (2016). Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *CELL* 165, 35–44. doi:10.1016/j.cell.2016.02.065

Jackute, J., Zemaitis, M., Pranys, D., Sitkauskiene, B., Miliauskas, S., Vaitkiene, S., et al. (2018). Distribution of M1 and M2 Macrophages in Tumor Islets and Stroma in Relation to Prognosis of Non-small Cell Lung Cancer. *BMC IMMUNOL.* 19. doi:10.1186/s12865-018-0241-4

Jiang, X., Sun, L., Qiu, J. J., Sun, X., Li, S., Wang, X., et al. (2013). A Novel Application of Furazolidone: Anti-leukemic Activity in Acute Myeloid Leukemia. *PLOS ONE* 8, e72335. doi:10.1371/journal.pone.0072335

Krzywinski, M., Schein, J., Birol, İ., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: An Information Aesthetic for Comparative Genomics. *Genome Res.* 19, 1639–1645. doi:10.1101/gr.092759.109

Kurtovic-Kozaric, A., Kugic, A., Hasic, A., Beslija, S., Ceric, T., Pasic, A., et al. (2017). Long-term Outcome of GIST Patients Treated with Delayed Imatinib Therapy. *Eur. J. Cancer* 78, 118–121. doi:10.1016/j.ejca.2017.03.024

Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., et al. (2006). The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease. *SCIENCE* 313, 1929–1935. doi:10.1126/science.1132939

Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: AnRPackage for Multivariate Analysis. *J. Stat. Soft.* 25, 1–18. doi:10.18637/jss.v025.i01

Li, J.-Y., Chen, Y.-P., Li, Y.-Q., Liu, N., and Ma, J. (2021). Chemotherapeutic and Targeted Agents Can Modulate the Tumor Microenvironment and Increase the Efficacy of Immune Checkpoint Blockades. *MOL. CANCER* 20. doi:10.1186/s12943-021-01317-7

Li, Q., Ni, Y., Zhang, L., Jiang, R., Xu, J., Yang, H., et al. (2021). HIF-1α-induced Expression of m6A Reader YTHDF1 Drives Hypoxia-Induced Autophagy and Malignancy of Hepatocellular Carcinoma by Promoting ATG2A and ATG14 Translation. *Sig Transduct Target. Ther.* 6, 453. doi:10.1038/s41392-020-00453-8

Lin, X., Wang, Z., Yang, G., Wen, G., and Zhang, H. (2020). YTHDF2 Correlates with Tumor Immune Infiltrates in Lower-Grade Glioma. *Aging* 12, 18476–18500. doi:10.18632/aging.103812

Mace, E. M., Paust, S., Conte, M. I., Baxley, R. M., Schmit, M. M., Patil, S. L., et al. (2020). Human NK Cell Deficiency as a Result of Biallelic Mutations in MCM10. *J. CLIN. INVEST.* 130, 5272–5286. doi:10.1172/JCI134966

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

Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: Efficient and Comprehensive Analysis of Somatic Variants in Cancer. *Genome Res.* 28, 1747–1756. doi:10.1101/gr.239244.118

McDougal, M., Nair-Fairless, P., Weiss, T., Dao, E., Chapple, A. G., and Jernigan, A. (2021). Practice Patterns and Results of Tumor and Germline Genetic Evaluation of Women with Endometrial Cancer in South Louisiana. *Gynecol. Oncol. Rep.* 36, 100717. doi:10.1016/j.gore.2021.100717

Meškytė, E. M., Keskas, S., and Ciribilli, Y. (2020). MYC as a Multifaceted Regulator of Tumor Microenvironment Leading to Metastasis. *Ijms* 21, 7710. doi:10.3390/ijms21207710

Mollaoglu, G., Jones, A., Wait, S. J., Mukhopadhyay, A., Jeong, S., Arya, R., et al. (2018). The Lineage-Defining Transcription Factors SOX2 and NKX2-1 Determine Lung Cancer Cell Fate and Shape the Tumor Immune Microenvironment. *IMMUNITY* 49, 764–779. doi:10.1016/j.immuni.2018.09.020

Ranstam, J., and Cook, J. A. (2017). Kaplan-meier Curve. *BRIT J. SURG.* 104, 442. doi:10.1002/bjs.10238

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

Savola, S., Klami, A., Myllykangas, S., Manara, C., Scotlandi, K., Picci, P., et al. (2011). High Expression of Complement Component 5 (C5) at Tumor Site Associates with Superior Survival in Ewing's Sarcoma Family of Tumour Patients. *ISRN Oncol.* 2011, 1–10. doi:10.5402/2011/168712

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

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

Song, C., and Zhou, C. (2021). HOXA10 Mediates Epithelial-Mesenchymal Transition to Promote Gastric Cancer Metastasis Partly via Modulation of TGFB2/Smad/METTL3 Signaling axis. *J. Exp. Clin. Cancer Res.* 40. doi:10.1186/s13046-021-01859-0

Sotiriou, C., Wirapati, P., Loi, S., Harris, A., Fox, S., Smeds, J., et al. (2006). Gene Expression Profiling in Breast Cancer: Understanding the Molecular Basis of Histologic Grade to Improve Prognosis. *Jnci-j NATL. CANCER* 98, 262–272. doi:10.1093/jnci/djj052

Stiller, C. A., Trama, A., Serraino, D., Rossi, S., Navarro, C., Chirlaque, M. D., et al. (2013). Descriptive Epidemiology of Sarcomas in Europe: Report from the RARECARE Project. *Eur. J. Cancer* 49, 684–695. doi:10.1016/j.ejca.2012.09.011

Storey, J. D. (2002). A Direct Approach to False Discovery Rates. *J. R. STAT. SOC. B* 64, 479–498. doi:10.1111/1467-9868.00346

Sun, Y., Tang, S., and Xiao, X. (2015). The Effect of GADD45a on Furazolidone-Induced S-phase Cell-Cycle Arrest in Human Hepatoma G2 Cells. *J. Biochem. Mol. Toxicol.* 29, 489–495. doi:10.1002/jbt.21719

Topalian, S. L., Hodi, F. S., Brahmer, J. R., Gettinger, S. N., Smith, D. C., McDermott, D. F., et al. (2012). Safety, Activity, and Immune Correlates of Anti-PD-1 Antibody in Cancer. *N. Engl. J. Med.* 366, 2443–2454. doi:10.1056/NEJMoa1200690

Trojani, M., Contesso, G., Coindre, J. M., Rouesse, J., Bui, N. B., De Mascarel, A., et al. (1984). Soft-tissue Sarcomas of Adults; Study of Pathological Prognostic Variables and Definition of a Histopathological Grading System. *Int. J. Cancer* 33, 37–42. doi:10.1002/ijc.2910330108

Wang, L., Cao, C., Ma, Q., Zeng, Q., Wang, H., Cheng, Z., et al. (2014). RNA-seq Analyses of Multiple Meristems of Soybean: Novel and Alternative Transcripts, Evolutionary and Functional Implications. *BMC PLANT BIOL.* 14, 169. doi:10.1186/1471-2229-14-169

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

Wolfsberger, J., Sakil, H. A. M., Zhou, L., van Bree, N., Baldisseri, E., de Souza Ferreira, S., et al. (2021). TAp73 Represses NF-Kb-Mediated Recruitment of Tumor-Associated Macrophages in Breast Cancer. *Proc. Natl. Acad. Sci. U.S.A.* 118. doi:10.1073/pnas.2017089118

Yamamoto, Y., Miyazato, K., Takahashi, K., Yoshimura, N., Tahara, H., and Hayakawa, Y. (2018). Lung-resident Natural Killer Cells Control Pulmonary Tumor Growth in Mice. *CANCER SCI.* 109, 2670–2676. doi:10.1111/cas.13703

Yang, M., Li, J., Gu, P., and Fan, X. (2021). The Application of Nanoparticles in Cancer Immunotherapy: Targeting Tumor Microenvironment. *Bioactive Mater.* 6, 1973–1987. doi:10.1016/j.bioactmat.2020.12.010

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring Tumour Purity and Stromal and Immune Cell Admixture from Expression Data. *NAT. COMMUN.* 4. 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: A J. Integr. Biol.* 16, 284–287. doi:10.1089/omi.2011.0118

Yu, H., Yang, X., Tang, J., Si, S., Zhou, Z., Lu, J., et al. (2021). ALKBH5 Inhibited Cell Proliferation and Sensitized Bladder Cancer Cells to Cisplatin by m6A-Ck2α-Mediated Glycolysis. *Mol. Ther. - Nucleic Acids* 23, 27–41. doi:10.1016/j.omtn.2020.10.031

Yu, J. G., Ji, C. H., and Shi, M. H. (2020). The Anti-infection Drug Furazolidone Inhibits NF-κB Signaling and Induces Cell Apoptosis in Small Cell Lung Cancer. *Kaohsiung J. Med. Sci.* 36, 998–1003. doi:10.1002/kjm2.12281

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

Zhang, C., Samanta, D., Lu, H., Bullen, J. W., Zhang, H., Chen, I., et al. (2016). Hypoxia Induces the Breast Cancer Stem Cell Phenotype by HIF-dependent and ALKBH5-Mediated M 6 A-Demethylation of NANOG mRNA. *Proc. Natl. Acad. Sci. U.S.A.* 113, E2047–E2056. doi:10.1073/pnas.1602883113

Zhang, W., Liu, W., Jia, L., Chen, D., Chang, I., Lake, M., et al. (2021). Targeting KDM4A Epigenetically Activates Tumor-Cell-Intrinsic Immunity by Inducing DNA Replication Stress. *Mol. Cel* 81 (10), 2148–2165. doi:10.1016/j.molcel.2021.02.038

Zhang, Z., Wang, Q., Zhang, M., Zhang, W., Zhao, L., Yang, C., et al. (2021). Comprehensive Analysis of the Transcriptome-wide m6A Methylome in Colorectal Cancer by MeRIP Sequencing. *Epigenetics* 16, 425–435. doi:10.1080/15592294.2020.1805684

Zhou, Z., Zhang, J., Xu, C., Yang, J., Zhang, Y., Liu, M., et al. (2021). An Integrated Model of N6-Methyladenosine Regulators to Predict Tumor Aggressiveness and Immune Evasion in Pancreatic Cancer. *EBIOMEDICINE* 65, 103271. doi:10.1016/j.ebiom.2021.103271

Zhu, D., Zhou, J., Zhao, J., Jiang, G., Zhang, X., Zhang, Y., et al. (2019). ZC3H13 Suppresses Colorectal Cancer Proliferation and Invasion via Inactivating Ras-ERK Signaling. *J. Cell Physiol.* 234, 8899–8907. doi:10.1002/jcp.27551

Zirkel, A., Lederer, M., Stöhr, N., Pazaitis, N., and Hüttelmaier, S. (2013). IGF2BP1 Promotes Mesenchymal Cell Properties and Migration of Tumor-Derived Cells by Enhancing the Expression of LEF1 and SNAI2 (SLUG). *NUCLEIC ACIDS RES.* 41, 6618–6636. doi:10.1093/nar/gkt410

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

frontiers | Frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 12 May 2022
doi: 10.3389/fbioe.2022.852734

Identification of a Pyroptosis-Related Gene Signature for Predicting the Immune Status and Prognosis in Lung Adenocarcinoma

Zetian Gong[1†], Qifan Li[2†], Jian Yang[2†], Pengpeng Zhang[1], Wei Sun[1], Qianhe Ren[1], Junjie Tang[1], Wei Wang[1], Hui Gong[3]* and Jun Li[1]**

[1]Department of Thoracic Surgery, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China, [2]Department of Thoracic Surgery, The First Affiliated Hospital of Soochow University, Suzhou, China, [3]Department of Trauma Center, Affiliated Hospital of Nantong University, Nantong, China

OPEN ACCESS

Edited by:
Ping Zhang,
Griffith University, Australia

Reviewed by:
Pin-Yen Chen,
Monash University, Australia
Jiyuan An,
Queensland University of Technology, Australia

***Correspondence:**
Wei Wang
wangwei15261883958@163.com
Hui Gong
gongh8899@163.com
Jun Li
dr_lijun1990@163.com

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

Specialty section:
This article was submitted to Preclinical Cell and Gene Therapy, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *11 January 2022*
Accepted: *14 April 2022*
Published: *12 May 2022*

Citation:
Gong Z, Li Q, Yang J, Zhang P, Sun W, Ren Q, Tang J, Wang W, Gong H and Li J (2022) Identification of a Pyroptosis-Related Gene Signature for Predicting the Immune Status and Prognosis in Lung Adenocarcinoma. Front. Bioeng. Biotechnol. 10:852734. doi: 10.3389/fbioe.2022.852734

Background: Pyroptosis is a form of programmed cell death triggered by the rupture of cell membranes and the release of inflammatory substances; it is essential in the occurrence and development of cancer. A considerable number of studies have revealed that pyroptosis is closely associated to the biological process of several cancers. However, the role of pyroptosis in lung adenocarcinoma (LUAD) remains elusive. The purpose of this study was to explore the prognostic role of pyroptosis-related genes (PRGs) and their relationship with the tumor immune microenvironment (TIME) in LUAD.

Methods: Gene expression profiles and clinical information were downloaded from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases. A prognostic PRG signature was established in the training set and verified in the validation sets. Functional enrichment and immune microenvironment analyses related to PRGs were performed and a nomogram based on the risk score and clinical characteristics was established. What is more, quantitative real-time PCR (qRT-PCR) analysis was applied in order to verify the potential biomarkers for LUAD.

Results: A prognostic signature based on five PRGs was constructed to separate LUAD patients into two risk groups. Patients in the high-risk group had worse prognoses than those in the low-risk group. The signature was identified as independent via Cox regression analyses and obtained the largest area under the curve (AUC = 0.677) in the receiver operating characteristic (ROC). Functional enrichment and immune microenvironment analyses demonstrated that the immune status was significantly different in the two subgroups and that immunotherapy may be effective for the high-risk group. Furthermore, qRT-PCR analysis verified that serum PRKACA and GPX4 could serve as diagnostic biomarkers for LUAD.

Conclusion: Overall, a risk signature based on five PRGs was generated, providing a novel perspective on the determinants of prognosis and survival in LUAD, as well as a basis for the development of individualized regimes.

Keywords: lung adenocarcinoma, pyroptosis, bioinformatics analysis, TCGA, PRKACA, GPX4, immune checkpoint genes

1 INTRODUCTION

According to the latest global cancer statistics, lung cancers remain one of the most commonly diagnosed cancers, and they have the highest incidence of deaths worldwide (Sung et al., 2021). Lung adenocarcinoma (LUAD) is the most common histological subtype of non-small cell lung cancer (NSCLC), which also occupies almost 80% of lung cancer cases (Gridelli et al., 2015). Despite progress in surgery, targeted therapy, chemotherapy, and radiotherapy, the 5-year overall survival (OS) rate for lung cancer remains only around 21% (Siegel et al., 2021). Currently, therapeutic regimens for individual LUAD patients are based mainly on specific factors such as radiomic features, tumor-node-metastasis (TNM) staging, tumor subtypes, and the differentiation grade. With the rapid rise in precision medicines, novel therapeutic schedules, especially immunotherapies and targeted therapies, have been proposed to prolong the lives of LUAD patients (Bronte et al., 2010; Saito et al., 2018). However, only a portion of patients have received benefits from them, leaving an urgent need to explore potential biomarkers for efficient and prognostic predictions.

Pyroptosis, also known as inflammatory "necrosis," is an inflammatory caspase-dependent cell death type triggered by the cell rupture and the release of many proinflammatory factors, including IL-1β, IL-18, ATP, and HMGB1 (Fang et al., 2020; Tang et al., 2020). It has been demonstrated that the process of pyroptotic cell death is mediated mainly through GSDMD (gasdermin D)-dependent activation regulated by caspase 1/4/5/11 (Shi et al., 2015). Activated caspases cleave the hinge region between the N- and C-terminal domains of GSDMD, releasing the segment with lethal activity and leading to pyroptosis (Ding et al., 2016). Several studies have indicated that pyroptosis was both a friend and a foe of cancers (Nagarajan et al., 2019; Xia et al., 2019; Fang et al., 2020). On the one hand, the inflammatory mediators released and several signaling pathways are bound up with the tumorigenesis and their chemotherapeutic drugs resistance. On the other hand, as a type of programmed cell death, pyroptosis can suppress the emergence and progression of tumors. In NSCLC, a high level of GSDMD expression was shown to be linked with invasive features, including more advanced TNM stages and larger tumor sizes (Burdette et al., 2021). Recent studies have identified pyroptosis-related gene (PRG) signatures for the prognosis of ovarian cancer and gastric cancer (Shao et al., 2021; Ye et al., 2021), while the performance of PRGs in LUAD has not yet been clarified.

Given the existing findings, we know that pyroptosis is critical to the development of tumors and to antitumor processes; however, its precise functions in LUAD have not been explored as extensively. In the present work, we aimed to construct a scoring model based on PRGs to predict the prognosis of LUAD and explore the latter's relationship with immune checkpoint genes (ICGs), hoping to find additional therapeutic targets.

2 MATERIALS AND METHODS

2.1 Data Acquisition and Processing

The lung adenocarcinoma RNA-seq (FPKM) data and the corresponding clinical information were obtained from the TCGA database (https://portal.gdc.cancer.gov/). The cohort consisted of 497 tumor tissues and 54 normal tissues, with the complete clinical information of 486 patients (tumor = 439, normal = 47) extracted as a training set. The Ensemble IDs were transformed into gene symbols via the use of the "rtracklayer" and "dplyr" R packages, and the pieces of clinical information were merged into a single matrix for further analysis. To increase the reliability of the study, two Gene Expression Omnibus (GEO) datasets, i.e., GSE31210 and GSE50081 (both using the GPL570 platform), which contained the microarray-based expression data of LUAD patients and the relevant clinical information (n = 536), were extracted for validation from the GEO website (https://www.ncbi.nlm.nih.gov/geo/). In this study, we also identified 79 ICGs from a review of the literature (Pardoll, 2012; Hu et al., 2021), most of which were ligands, receptors or important molecules in immune checkpoint pathways (**Supplementary Table S1**).

2.2 Identification of Differentially Expressed Pyroptosis-Related mRNAs

A total of 33 PRGs were extracted from prior reviews (Ye et al., 2021) and are presented in **Supplementary Table S2**. Differentially expressed PRGs (DE-PRGs) were identified in the training cohort between normal and tumor tissues, using the "limma" R package with thresholds of $p < 0.05$.

2.3 Establishment and Validation of the Pyroptosis-Related Prognostic Gene Signature

To identify the prognostic genes among all PRGs, we further employed Cox regression analysis with the "survival" R package to assess the links between each gene and survival status in the training cohort. To avoid omissions, we set 0.2 as the cut-off *p*-value, and seven survival-related genes were screened for further analysis. Subsequently, multivariate Cox regression analysis was conducted to narrow down the candidate genes based on the lowest Akaike information criterion (AIC). Ultimately, five PRGs and their coefficients were retained. A prognostic risk score was created for each patient via the following formula: Risk score = ∑Coef (PRGs) * Exp (PRGs), where Exp (PRGs) is the relative expression of the candidate PRGs, and Coef (PRGs) is the regression coefficient. Based on the median value of the risk score, patients in the training set were divided into the high-risk group and the low-risk group. The OS between the two groups was compared by means of Kaplan–Meier analysis with the "survival" and "survminer" R packages. The predictive performance of the model was further validated in two GEO datasets (GSE31210 and GSE50081). Samples in the validation cohort were separated into high-risk and low-risk groups based on the formula for the risk score derived from the training dataset, respectively. The receiver operating characteristic (ROC) curve was used to assess the prognostic performance through the "timeROC" R package. The area under the curve (AUC) of each cohort was calculated for detailed evaluations.

2.4 Independent Prognostic Analysis of the Risk Score

We extracted the clinical information (age, gender, AJCC stage, TNM stage, tobacco history, and anatomical location) of patients in the training cohort. These elements were analyzed in company with the risk score in our regression model, employing univariate and multivariable Cox regression models. Furthermore, a time-dependent ROC curve was used to evaluate the predictive accuracy for OS by different clinicopathological factors and risk scores by means of the "survivalROC" package.

2.5 Development of a Predictive Nomogram

A nomogram incorporating the signature and clinical parameters was developed via the "rms" R package to predict the overall survival of LUAD patients. Then, the calibration curves and ROC curves were plotted to assess the predictive accuracy of the nomogram.

2.6 Functional Enrichment Analysis of PRGs

Depending on the median risk score, patients in the training cohort were stratified into two subgroups. The differentially expressed genes (DEGs) between the low- and high-risk groups were filtered at the specific threshold ($|\log2FC| \geq 1$ and FDR < 0.05). To clarify the biological functions of the prediction model, gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed, based on the DEGs, by applying the "clusterProfiler" R package with the criteria of $p < 0.05$ and FDR < 0.05. The "gsva" R package was also employed to conduct the ssGSEA to calculate the scores of infiltrating immune cells and to assess the activity of immune-related pathways.

2.7 Evaluation of the Immune Status Between the Two Subgroups

To further explore the link between the prediction model and the immune system, the single-sample gene set enrichment analysis (ssGSEA) method was utilized to quantify the overall immune status of the two subgroups by analyzing the expression profiles of the 29 immune signature gene sets. Subsequently, the ESTIMATE algorithm was performed to calculate stromal and immune scores, determining the levels of stromal and immune cell tumor infiltration. Thereafter, correlations between the risk score and several key ICGs, such as PD-L1, CTLA4, LAG-3, and so on, were evaluated. Spearman correlation analyses were used to examine the relationship among the risk score, the stromal and immune scores, and the expression of ICGs.

2.8 Protein Levels of NRGs in the HPA Database

The Human Protein Atlas (HPA) is a database containing all of the human proteins in cells, tissues and organs, where all images of tissues are stained via immunohistochemistry. To compare the protein expression levels related to the prognostic signature, we extracted the immunohistochemical images of the candidate PRGs from the HPA database (https://www.proteinatlas.org/).

2.9 Cell Lines and Cell Culture

Three LUAD cells—namely A549, H1299, and H1650—and one normal epithelial cell line (HBE) were purchased from the Cell Bank of the Chinese Academy of Sciences (Shanghai, China) and cultured in an RPMI-1640 medium (HyClone, Logan, UT, United States) with 10% fetal bovine serum at 37°C in a humidified atmosphere with 5% CO2.

2.10 Real-Time Quantitative Reverse-Transcriptase Polymerase Chain Reaction Analysis

Total RNA was extracted by TRIzol reagent (Thermo Fisher Scientific, Carlsbad, CA, United States) according to the protocol and reverse-transcribed to cDNA through the use of random primer amplification. Real-time qRT-PCR analysis was carried out using Platinum SYBR Green qPCR SuperMix-UDG kits (Life Technologies, Gaithersburg, MD, United States). Primers used for the qRT-PCR analysis were performed as follows. Glyceraldehyde 3-phosphate dehydrogenase (GADPH) levels were used to normalize PRKACA and GPX4 expression. Relative expression was calculated using the ΔΔCt method.

Statistical Analysis

All statistical analyses were performed using R language (Version 4.1.0). The Kaplan–Meier method with a two-sided log-rank test was performed to compare the OS of patients between the two subgroups. To determine the independent risk characteristics, univariate and multivariate Cox analyses were applied. Correlation coefficients between two non-bivariate normally distributed variables were computed via Spearman analyses. The hazard ratios (HRs) and the 95% confidence intervals of the aforementioned elements were estimated in order to quantify the strength of these associations. All statistical tests were two-tailed. The overall flowchart is shown in **Figure 1**.

3 RESULTS

3.1 Defining the Differentially Expressed PRGs in LUAD

The expression of the 33 PRGs in LUAD and normal lung tissues was first obtained by means of the TCGA dataset. Following differential expression analysis of the training set, 27 PRGs were either upregulated or downregulated in LUAD. More definitely, the expression of IL6, NLRC4, CASP5, IL1B, CASP1, NLRP3, NLRP1, PYCARD, IL18, PRKACA, TNF, and NOD1 was raised, while the expression of AIM2, TIRAP, PLCG1, GSDMD, CASP4, GPX4, CASP8, GSDME, PJVK, CASP3, CASP6, GSDMA, GSDMB, NLRP7, and GSDMC was declined in LUAD in comparison with normal tissues (**Figure 2A**, FDR < 0.05).

Gene	Forward primer	Reverse primer
GAPDH	ACAACTTTGGTATCGTGGAAGG	GCCATCACGCCACAGTTTC
PRKACA	CAAGGAGACCGGGAACCACTA	CATTCAGGGTGTGTTCGATCTG
GPX4	GAGGCAAGACCGAAGTAAACTAC	CCGAACTGGTTACACGGGAA

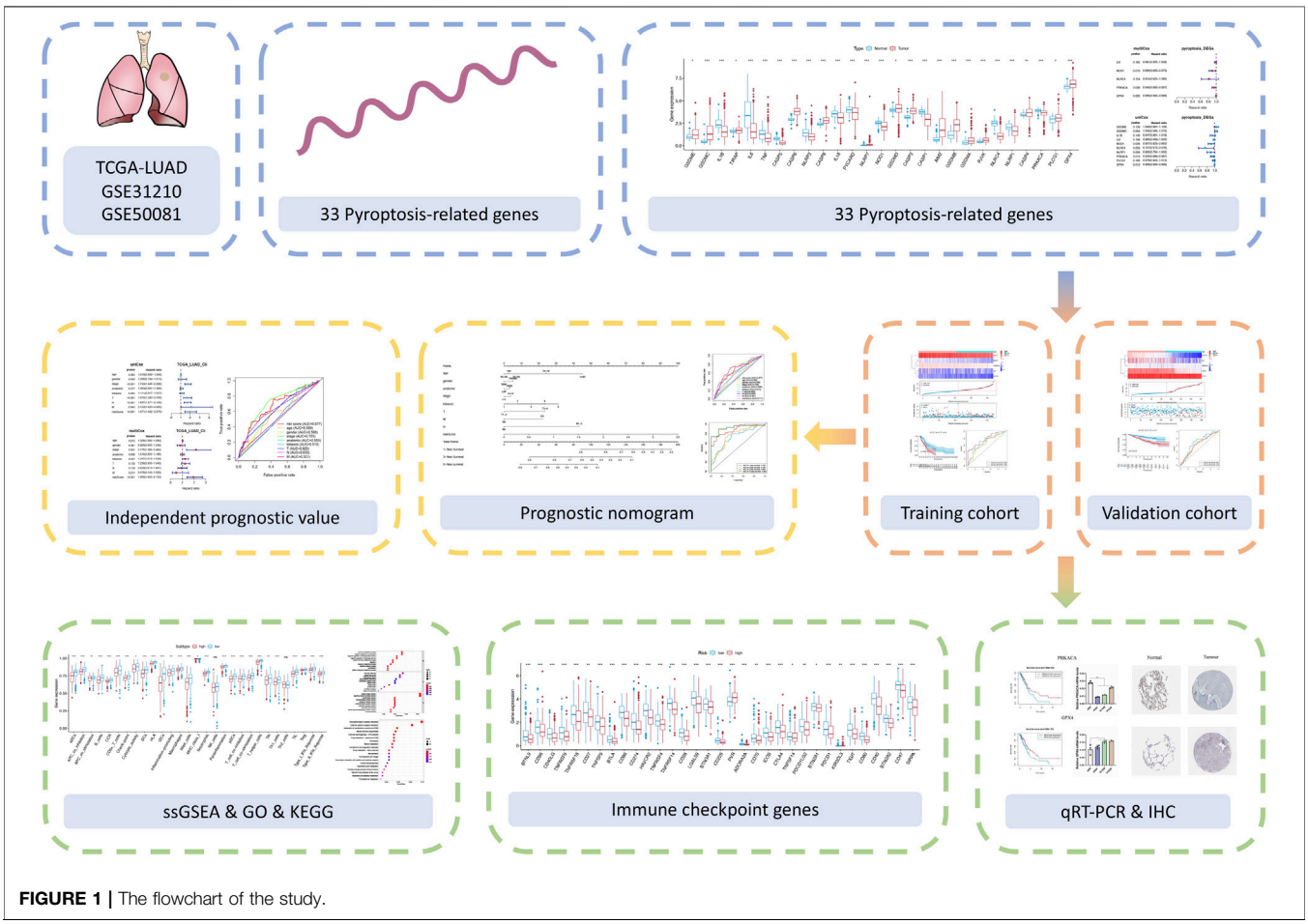


FIGURE 1 | The flowchart of the study.

3.2 The Establishment and Verification of a Pyroptosis-Related Prognostic Model

Univariate Cox regression analysis was first applied in order to verify the candidate PRGs associated with prognosis ($p < 0.2$) (**Figure 2B**), and multivariate Cox regression analysis further identified five PRGs, namely IL6, NOD1, NLRC4, PRKACA, and GPX4, based on the lowest AIC (1,059.83) (**Figure 2C**). The formula is shown as follows: risk score = (−0.019 * expression level of IL6) + (−0.110 * expression level of NOD1) + (−0.205 * expression level of NLRC4) + (−0.059 * expression level of PRKACA) + (−0.005 * expression level of GPX4). We classified the LUAD cases into low-risk (n = 220) and high-risk (n = 219) groups depending on the median risk score. The risk score distribution, survival status, and gene expression pattern of the two groups are presented in **Figure 3A**. As the risk score raised, the patients' risk of death also increased and the survival time reduced. The Kaplan–Meier analysis revealed that LUAD patients in the high-risk group had shorter OS (**Figure 3C**, p = 1.496e-05), with AUCs of 0.683, 0.659, and 0.776 in the 1-year, 3-year, and 5-year ROC curves, respectively (**Figure 3D**).

3.3 Validation of the Signature in Two GEO Datasets

To evaluate the accuracy and stability of the prognostic signature, two GEO datasets (GSE31210 and GSE50081, both based on GPL570, **Supplementary Table S3**) were performed as external validations. Patients in the validation cohort were classified into low-risk and high-risk groups depending on the formula for the risk score derived from the training cohort, respectively (**Figure 3B**). Similarly, in the validation cohort, better OS belonged to the patients with low-risk scores (**Figure 3E**, p = 4.737e-05), with AUCs of 0.68, 0.619, and 0.625 in the 1-year, 3-year, and 5-year ROC curves, respectively (**Figure 3F**).

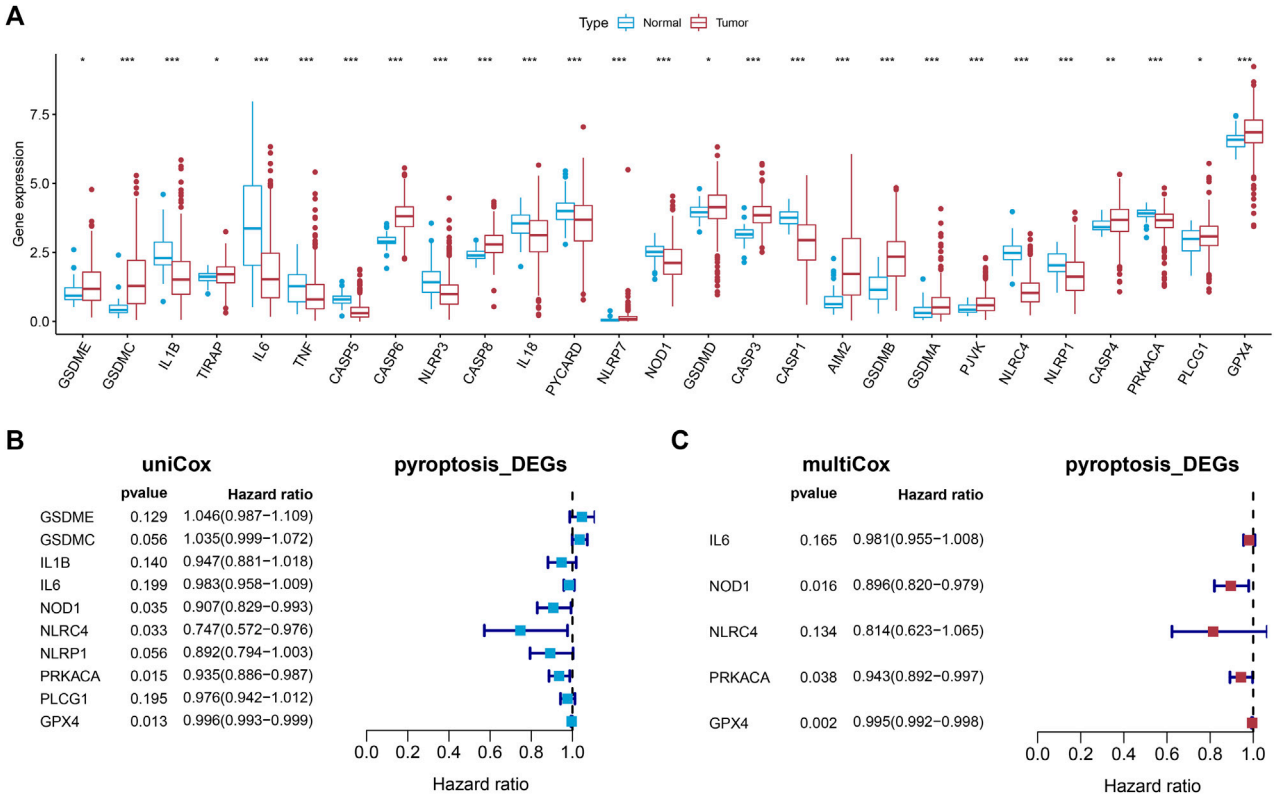


FIGURE 2 | Establishment of the prognostic model. **(A)** The boxplot showed the differentially expressed PRGs between normal and tumor tissues of LUAD. **(B,C)** Univariate and multivariate Cox regression identified the PRGs associated with prognosis.

3.4 Risk Factors Predictive of Survival in LUAD

Univariate and multivariate Cox regression analyses were applied to evaluate whether the risk score derived from the signature could function as an independent prognostic factor, using the "survival" package. As shown in **Figure 4A**, the AJCC stage ($p < 0.001$), T stage ($p < 0.001$), N stage ($p < 0.001$), M stage ($p = 0.043$), and risk score ($p < 0.001$) were significantly related to OS in the univariate Cox regression analysis, with only the AJCC stage ($p = 0.001$) and risk score ($p < 0.001$) also being significantly related to OS in the multivariate Cox regression analysis (**Figure 4B**). Furthermore, a time-dependent ROC curve was performed to testify as to the predictive accuracy. According to the results, the AUC of the risk score was 0.677, which was higher than the AUC of the T stage, N stage, and M stage and similar to the AUC of the AJCC stage (**Figure 4C**), indicating that the prognostic risk model was relatively reliable. To sum up, the prediction model could be regarded as an independent prognostic factor for LUAD patients.

3.5 Construction of a Predictive Nomogram

To predict the patients' survival time accurately, nomograms are usually applied by calculating the nomogram score based on each prognostic elements included in the nomogram (Balachandran et al., 2015). In this study, we established a nomogram to evaluate the probabilities of 1-year, 3-year, 5-year, and 10-year survival by using the risk score and other clinicopathological elements, like gender, AJCC stage, TNM stage, tobacco history, and anatomical location (**Figure 5A**). Calibration curves were also plotted and showed a high degree of consistency between the actual and the predicted 1-year, 3-year, 5-year, and 10-year survival when compared to the reference line (**Figure 5B**). Then, we observed that the AUC of the nomscore calculated from the nomogram was 0.711, which was greater than the riskscore (AUC = 0.677) (**Figure 5C**). Furthermore, the AUC of the nomscore in the 1-year, 3-year, 5-year, and 10-year ROC curves reached 0.726, 0.759, 0.885, and 0.923 (**Figure 5D**). These results suggested that the prediction efficiency would be more accurate and reliable when the risk score was jointed with other clinicopathological parameters.

3.6 Functional Enrichment Analysis

To further elucidate the biological functions and pathways of DE-PRGs in pyroptosis, the "limma" R package was applied in order to extract the DEGs between the two groups. FDR < 0.05 and $|\log2FC| \geq 1$ were statistically significant. Altogether, 820 DEGs were identified in the TCGA cohort. Among them, 343 genes were highly expressed in the high-risk group, while the other 477 genes were low expressed. Subsequently, based on the DEGs, GO enrichment analysis and KEGG pathway analysis were performed. The top GO terms were "hormone metabolic

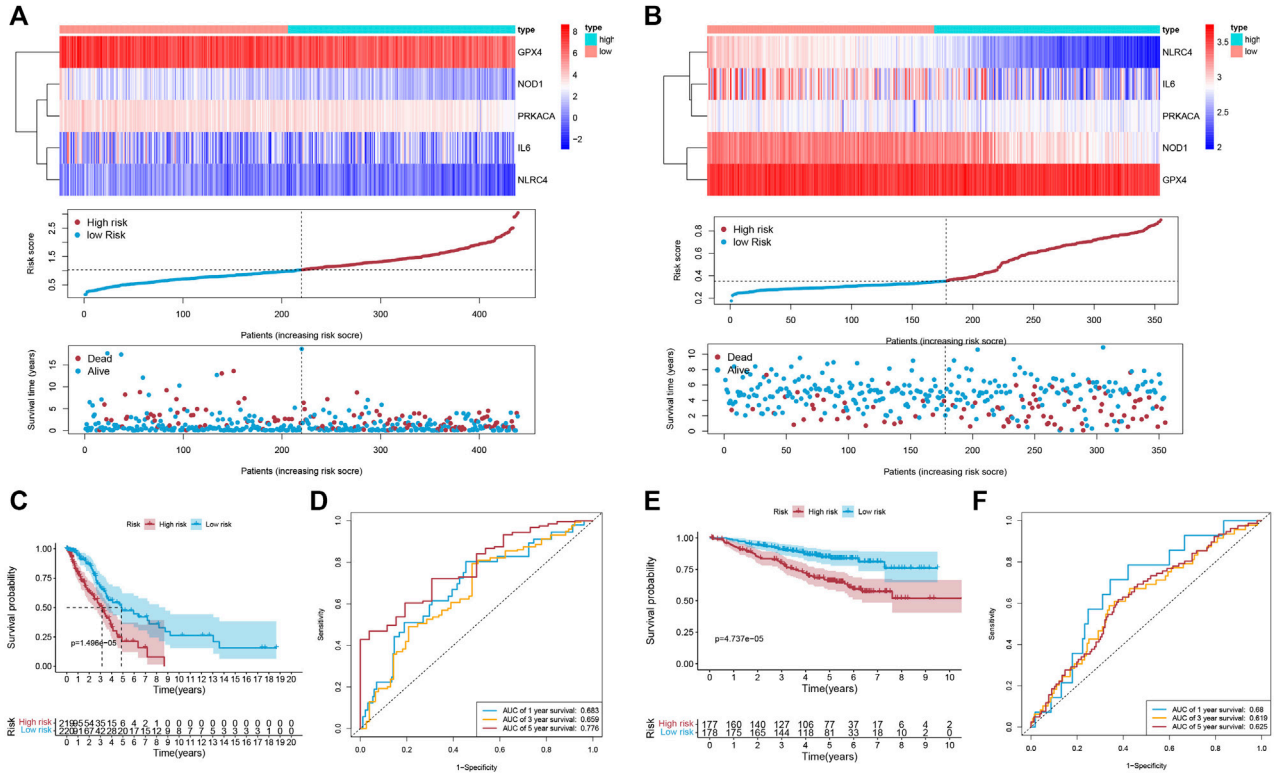


FIGURE 3 | The accuracy of the prognostic model. **(A,B)** The gene expression pattern; distribution of risk scores between the two groups; distribution of survival status in the training cohort **(A)** and validation cohort **(B)**. **(C,E)** K-M survival curves of OS in the training cohort (C; p = 1.496e-05) and validation cohort (E; p = 4.737e-05). **(D,F)** The ROC curve representing patients' survival for different numbers of years in the two cohorts, AUC of 1-year, 3-year, and 5-year OS were 0.683, 0.659, and 0.776 (respectively) in the training cohort **(D)**, and were 0.68, 0.619, and 0.625 (respectively) in the validation cohort **(F)**.

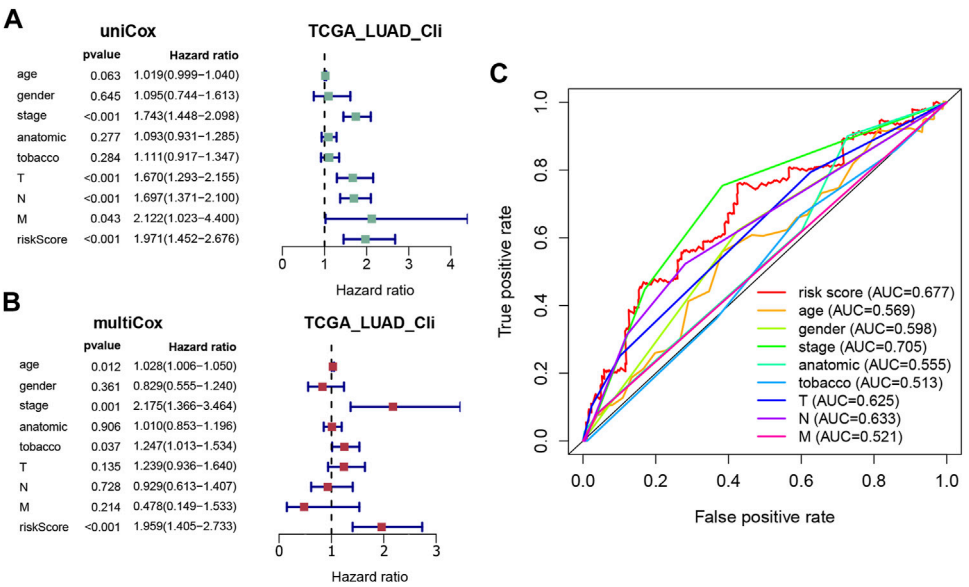


FIGURE 4 | Evaluation of the prognostic accuracy of the model and other clinicopathological characteristics. **(A,B)** Univariate and multivariate Cox regression analyses of the risk score and other clinical parameters (age, gender, AJCC stage, anatomic location, tobacco history, T, N, and M). **(C)** ROC curves for the risk score (AUC = 0.677) and other clinical features.

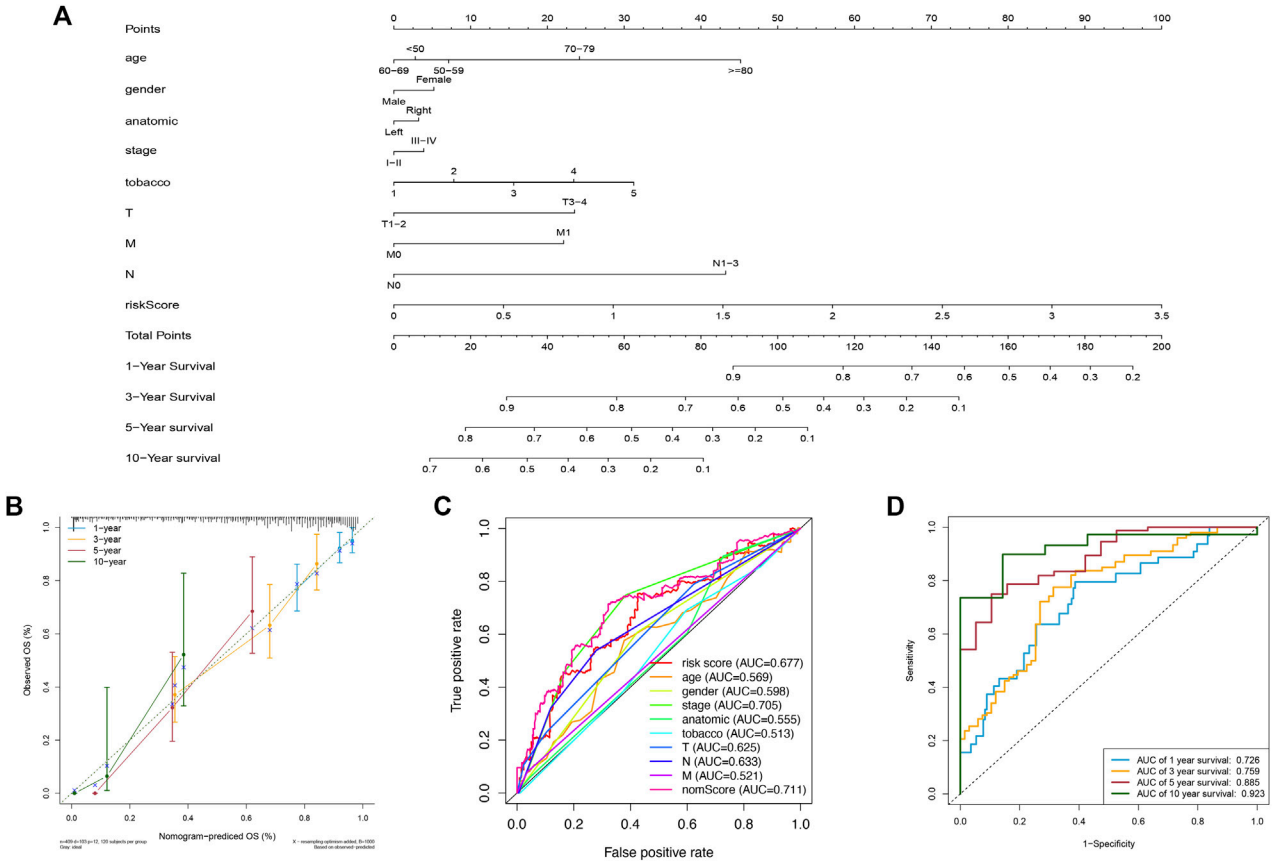


FIGURE 5 | Construction and assessment of a nomogram. **(A)** The nomogram of 1-year, 3-year, 5-year, and 10-year OS on basis of the risk model and other clinical features. **(B)** Calibration plots used for evaluating the consistency between the actual and the predicted 1-year, 3-year, 5-year, and 10-year OS. **(C)** ROC curves for the nomscore (AUC = 0.711) and other elements. **(D)** The ROC curves of nomscore for predicting OS. The AUC of 1-year, 3-year, 5-year, and 10-year OS were 0.726, 0.759, 0.885, and 0.923 respectively.

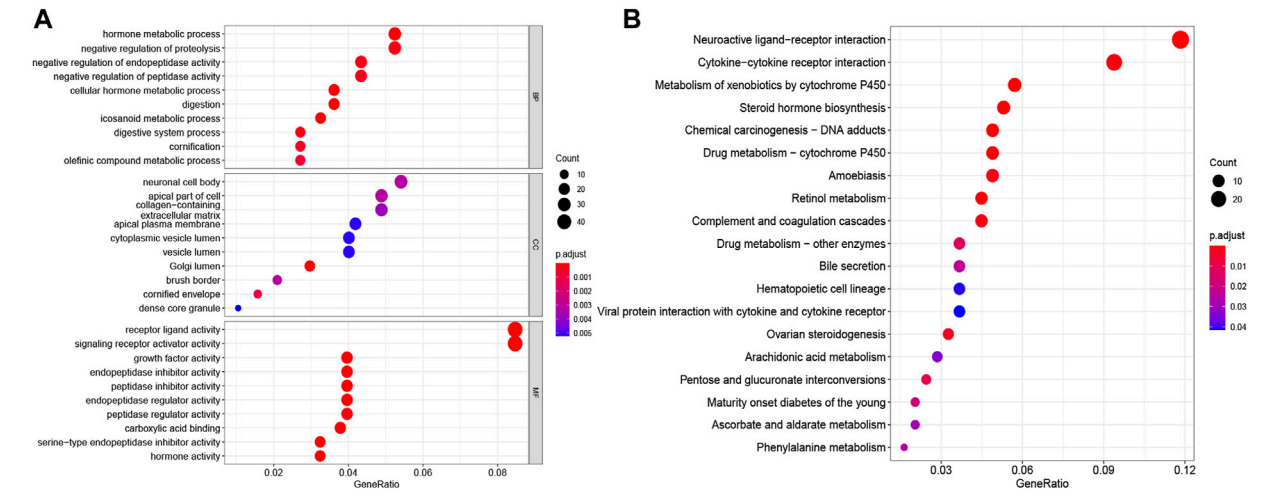


FIGURE 6 | Functional enrichment analyses depending on the signature. The bubble graphs for GO enrichment analysis **(A)** and KEGG pathway analysis **(B)** were displayed. The bigger the bubble, the richer the genes, and the darker the red, the more pronounced the difference.

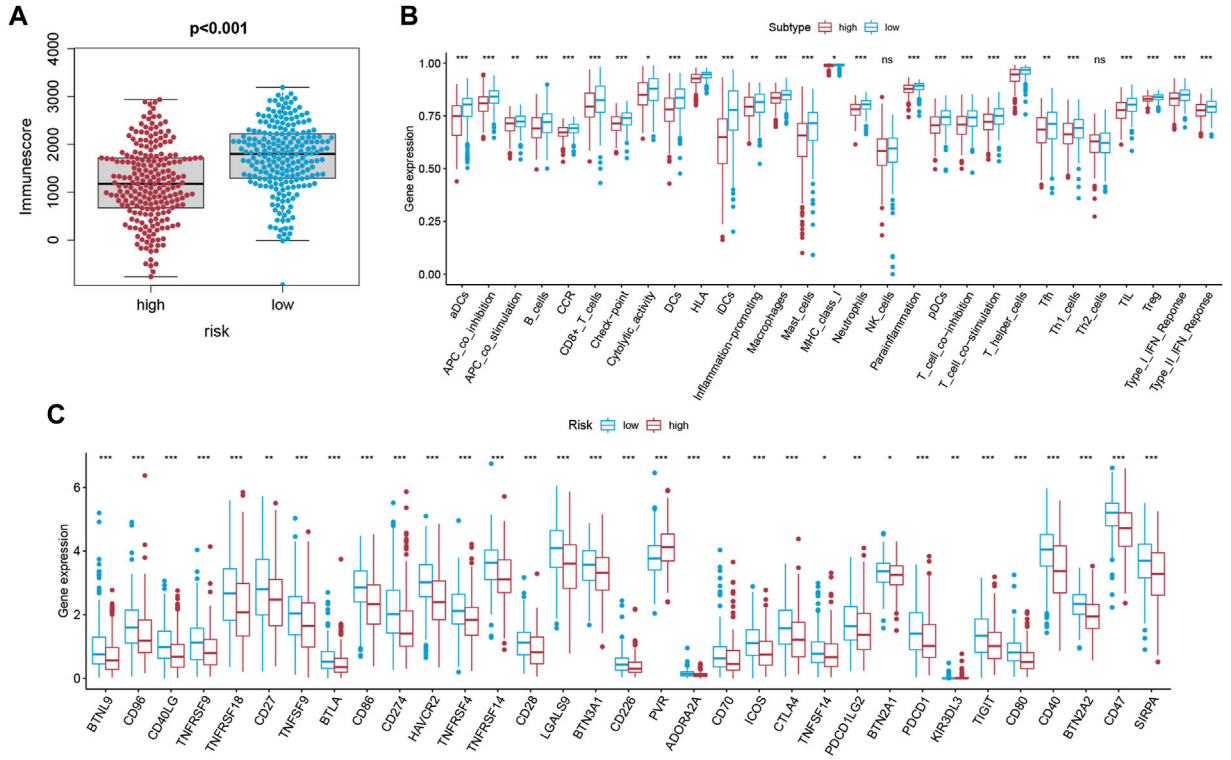


FIGURE 7 | Analysis of the immune status of LUAD. **(A)** The correlation between the risk score and the immune score was displayed ($p < 0.001$). **(B)** Differential distribution of enrichment scores of 16 immune cells types and 13 immune-related pathways between the low-risk (blue box) and high-risk (red box) groups via ssGSEA. **(C)** Boxplots of the expression level of ICGs between the two subgroups (low-risk: blue; high-risk: red). Only the expression level of PVR was positively correlated to the risk score.

process" in the biological process (BP), "neuronal cell body" in the cellular component (CC), and "receptor ligand activity" in the molecular function (MF), respectively (**Figure 6A**). According to KEGG analysis, "neuroactive ligand–receptor interaction," "cytokine–cytokine receptor interaction," "steroid hormone biosynthesis," and "chemical carcinogenesis—DNA adducts" were the main pathways (**Figure 6B**). On the whole, the results implied that the DEGs were mainly correlated with the immune response, receptor interaction, and chemical metabolism.

3.7 Comparison of the Tumor Immune Microenvironment Between Groups

Based on the ESTIMATE algorithm, we successfully obtained the immune scores, stromal scores, and estimate scores. The immune scores were distributed between −934.47 and 3,190.06 and represented a significant difference between the subgroups (**Figure 7A**, $p < 0.001$). Thereafter, we applied ssGSEA to quantify the immune activation level between the two subgroups by analyzing the expression profiles of the 29 immune signature gene sets. As shown in **Figure 7B**, the levels of immune cell infiltration, especially of B cells, $CD8^+$ T cells, neutrophils, T helper 1 (Th1) cells, tumor-infiltrating lymphocytes (TILs), and regulatory T (Treg) cells, were generally lower in the high-risk group than in the low-risk group. Moreover, the levels of 13 immune pathways displayed a similar distribution between the two groups. Furthermore, from the previous literature, 79 ICGs were extracted (**Supplementary Table S1**). After removing the HLA related genes, 60 genes remained, in which 51 out of 60 ICGs have expression values. The results of the Spearman correlation analyses between the risk score and the ICGs revealed that the signature may be closely related to immunotherapy. As shown in **Figure 7C**, most of the ICGs were highly expressed in the low-risk group, except for PVR, which may be a promising therapeutic target. We then validated our results through the use of clinical specimens from the HPA. The OS and the histological expressions of PRKACA and GPX4 in normal and tumor tissues were exhibited, in accordance with the results front (**Figures 8A,C–E,G,H**).

3.8 qRT-PCR Analysis

We validated this result by applying the qRT-PCR analysis to three LUAD cell lines and a normal lung cell line. We found that the mRNA expression level of PRKACA was higher in the normal cell line than in the cancer cell lines, and for GPX4 it was the opposite, consistent with our results above (**Figures 8B,F**).

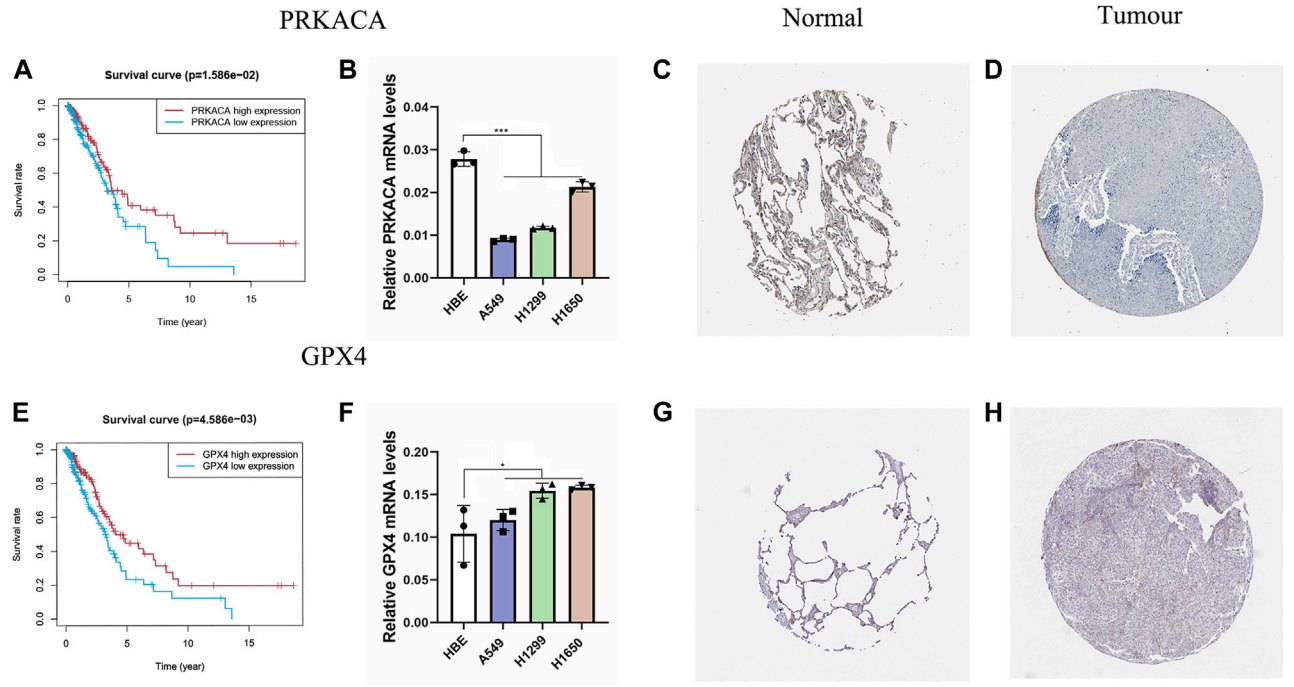


FIGURE 8 | Validation of PRKACA and GPX4. **(A,E)** The prognostic differences of PRKACA (p = 1.586e-02) and GPX4 (p = 4.586e-03) between the high-risk and low-risk groups were investigated. **(B,F)** Relative mRNA levels of PRKACA and GPX4 were displayed via qRT-PCR. PRKACA was highly expressed in normal tissues, and GPX4 showed the opposite. **(C–D,G–H)** Protein expression of the two candidate genes in the HPA database. **(C–D)** Protein expression of PRKACA. **(G–H)** Protein expression of GPX4.

4 DISCUSSION

Lung cancer is one of the leading causes of cancer morbidities and the most common cause of cancer-related deaths worldwide (Sung et al., 2021). For early-stage LUAD patients, surgery is the recommended treatment (Vansteenkiste et al., 2014), whereas chemotherapy, radiotherapy, immunotherapy, and targeted therapy are recommended for advanced LUAD patients (Hirsch et al., 2017). Several studies have demonstrated a strong association between TMB and the clinical benefits of receiving immunotherapy (Proto et al., 2019; Samstein et al., 2019). Moreover, TIME is closely correlated with the efficacy of immunotherapy (Frankel et al., 2017; Lei et al., 2020), suggesting that immunotherapy may be more effective for high-risk patients based on the risk model.

Pyroptosis is a type of programmed cell death triggered by a family of inflammatory caspases that plays as a double-edged sword in the tumorigenesis and therapeutic mechanisms, which has been newly recognized in recent years. Pyroptosis was found to release inflammatory factors and stimulate normal cells, leading to transformation into tumor cells (Karki and Kanneganti, 2019). Having said that, pyroptosis can also promote tumor cell death and restrain proliferation and migration of cancer cells, making pyrolysis itself a hopeful prognostic and therapeutic target for cancer (Zheng and Li, 2020). A pyroptosis-related signature has been established to predict prognosis in ovarian cancer (Ye et al., 2021). While, the role of PRGs in LUAD has not yet been identified; thus, our study aimed to explain it.

In the present study, 33 PRGs were systematically analyzed to identify those associated with OS. Following the differential expression analysis, univariate Cox regression analysis, and multivariate Cox regression analysis, five optimal mRNAs, namely IL6, NOD1, NLRC4, PRKACA, and GPX4, were screened out for the pyroptosis-related prognostic signature. In our study, GPX4 and PRKACA were found to be elements of the prediction model. Previous literature indicated that GPX4, an antioxidant enzyme that participate in repairing oxidative damage to lipids, was an important negative regulator of the pyroptotic cell death pathway (Kang et al., 2018; Russo and Rathinam, 2018; Kajarabille and Latunde -Dada, 2019). PAKACA is a catalytic subunit alpha of protein kinase A activated by cAMP that closely related to the progression of tumor (Chao et al., 2021). Increased transcription of PRKACA has been detected in patients of breast cancer that resistant to trastuzumab, which becomes a routine treatment for HER2-positive breast cancer (Moody et al., 2015). This suggests that PRKACA may perform as a biomarker for cancer and a prognostic indicator. The five PRGs' signature was testified to be an independent indicator for LUAD prognosis. Following this, a prediction model depending on the PRGs' signature was constructed. The AUC of the signature could reach 0.776 in the training set when predicting 5-year survival.

Until now, the mechanism of pyroptosis has not been fully discovered. What we know is that as tumors progress, multiple cell death modes may be concurrent and interact with one another (Fritsch et al., 2019). Generally, pyroptosis features the release of many proinflammatory factors and the rupture of cell plasma membranes (Zheng and Li, 2020). We analyzed the DEGs between two subgroups and discovered they were mainly involved in receptor ligand activity and metabolic processes, implying that dying cells may induce complex metabolic processes. According to the GO and KEGG analyses results, we have reason to suppose that pyroptosis plays a vital role in the regulation of the tumor microenvironment.

Cancer immunotherapies targeting immune checkpoints have been proven to improve OS in various cancers (Dyck and Mills, 2017; Li et al., 2019). Thus, exploring novel targets and developing new schemes for antitumor therapy are always main tasks within current medicine. In this study, we hoped to discover the association between the signature and TIME. Based on the ESTIMATE algorithm and ssGSEA, we could speculate that in the high-risk group there is an overall lesion of immune functions. Previous researches revealed that Treg cells were recruited into the human tumor microenvironment and inhibited T cell immunity to abolish the therapeutic efficacy of PD-L1, CTLA-4, and the TGF-β blockade, regardless of whether they were live or apoptotic (Zou, 2006; Maj et al., 2017; Principe et al., 2021). Surprisingly, the level of Treg cells was higher in the low-risk group than in the high-risk group, which indicated that immunotherapy may be effective for the high-risk group. Furthermore, we detected the correlation between the risk score and the expression levels of ICGs. The results demonstrated that the level of poliovirus receptor (PVR, CD155) was higher in the high-risk group. PVR, a member of the nectin-like family of adhesion molecules, has been proved to decrease the expansion and function of tumor antigen-specific $CD8^+$ T cells. PVR also has a high affinity to TIGHT, which is a promising new target for cancer immunotherapy (Shibuya et al., 1996; Kucan Brlic et al., 2019; Chauvin and Zarour, 2020). According to these findings, immunotherapy based on PVR may be promising for LUAD patients applicable to the model.

Despite the prognostic value of the signature, this study still encountered several limitations which must be considered. First, our report was retrospective and based on public databases, devoid of certain crucial clinicopathological information. Second, the way in which pyroptosis modulates the precise process of LUAD is unclear. Moreover, further biochemical experiments, such as immunohistochemistry, cell function experiments, and so on, need to be conducted to confirm the findings.

5 CONCLUSION

In conclusion, a prognostic signature of five PRGs was established in LUAD and validated in the GEO datasets to explore the role of pyroptosis in tumor malignancy. These PRGs were also associated with TIME, as well as helping to predict potential therapeutic regimens for LUAD. Further studies are necessary in order to verify these results in our study.

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

ZG, QL, and JL performed study concept and design; ZG, QL, JY, PZ, WS, and HG performed development of methodology and writing, review and revision of the paper; ZG, JY, QR, JT, and JL provided acquisition, analysis and interpretation of data, and statistical analysis; ZG, JY, and JL provided technical and material support. All authors read and approved the final paper.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2022.852734/full#supplementary-material

REFERENCES

Balachandran, V. P., Gonen, M., Smith, J. J., and DeMatteo, R. P. (2015). Nomograms in Oncology: More Than Meets the Eye. *Lancet Oncol.* 16 (4), e173–e180. doi:10.1016/s1470-2045(14)71116-7

Bronte, G., Rizzo, S., La Paglia, L., Adamo, V., Siragusa, S., Ficorella, C., et al. (2010). Driver Mutations and Differential Sensitivity to Targeted Therapies: a New Approach to the Treatment of Lung Adenocarcinoma. *Cancer Treat. Rev.* 36 (Suppl. 3), S21–S29. doi:10.1016/s0305-7372(10)70016-5

Burdette, B. E., Esparza, A. N., Zhu, H., and Wang, S. (2021). Gasdermin D in Pyroptosis. *Acta Pharm. Sin. B* 11 (9), 2768–2782. doi:10.1016/j.apsb.2021.02.006

Chao, B., Jiang, F., Bai, H., Meng, P., Wang, L., and Wang, F. (2021). Predicting the Prognosis of Glioma by Pyroptosis-Related Signature. *J. Cell Mol. Med.* 26(1). 133-143. doi:10.1111/jcmm.17061

Chauvin, J. M., and Zarour, H. M. (2020). TIGIT in Cancer Immunotherapy. *J. Immunother. Cancer* 8 (2), e000957. doi:10.1136/jitc-2020-000957

Ding, J., Wang, K., Liu, W., She, Y., Sun, Q., Shi, J., et al. (2016). Erratum: Pore-Forming Activity and Structural Autoinhibition of the Gasdermin Family. *Nature* 540 (7631), 150. doi:10.1038/nature20106

Dyck, L., and Mills, K. H. G. (2017). Immune Checkpoints and Their Inhibition in Cancer and Infectious Diseases. *Eur. J. Immunol.* 47 (5), 765–779. doi:10.1002/eji.201646875

Fang, Y., Tian, S., Pan, Y., Li, W., Wang, Q., Tang, Y., et al. (2020). Pyroptosis: A New Frontier in Cancer. *Biomed. Pharmacother.* 121, 109595. doi:10.1016/j.biopha.2019.109595

Frankel, T., Lanfranca, M. P., and Zou, W. (2017). The Role of Tumor Microenvironment in Cancer Immunotherapy. *Adv. Exp. Med. Biol.* 1036, 51–64. doi:10.1007/978-3-319-67577-0_4

Fritsch, M., Günther, S. D., Schwarzer, R., Albert, M.-C., Schorn, F., Werthenbach, J. P., et al. (2019). Caspase-8 Is the Molecular Switch for Apoptosis, Necroptosis and Pyroptosis. *Nature* 575 (7784), 683–687. doi:10.1038/s41586-019-1770-6

Gridelli, C., Rossi, A., Carbone, D. P., Guarize, J., Karachaliou, N., Mok, T., et al. (2015). Non-small-cell Lung Cancer. *Nat. Rev. Dis. Prim.* 1, 15009. doi:10.1038/nrdp.2015.9

Hirsch, F. R., Scagliotti, G. V., Mulshine, J. L., Kwon, R., Curran, W. J., Jr., 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

Hu, F. F., Liu, C. J., Liu, L. L., Zhang, Q., and Guo, A. Y. (2021). Expression Profile of Immune Checkpoint Genes and Their Roles in Predicting Immunotherapy Response. *Brief. Bioinform* 22 (3), bbaa176. doi:10.1093/bib/bbaa176

Kajarabille, N., and Latunde-Dada, G. O. (2019). Programmed Cell-Death by Ferroptosis: Antioxidants as Mitigators. *Int. J. Mol. Sci.* 20 (19), 4968. doi:10.3390/ijms20194968

Kang, R., Zeng, L., Zhu, S., Xie, Y., Liu, J., Wen, Q., et al. (2018). Lipid Peroxidation Drives Gasdermin D-Mediated Pyroptosis in Lethal Polymicrobial Sepsis. *Cell Host Microbe* 24 (1), 97–108. doi:10.1016/j.chom.2018.05.009

Karki, R., and Kanneganti, T.-D. (2019). Diverging Inflammasome Signals in Tumorigenesis and Potential Targeting. *Nat. Rev. Cancer* 19 (4), 197–214. doi:10.1038/s41568-019-0123-y

Kucan Brlic, P., Lenac Rovis, T., Cinamon, G., Tsukerman, P., Mandelboim, O., and Jonjic, S. (2019). Targeting PVR (CD155) and its Receptors in Anti-tumor Therapy. *Cell Mol. Immunol.* 16 (1), 40–52. doi:10.1038/s41423-018-0168-y

Lei, X., Lei, Y., Li, J.-K., Du, W.-X., Li, R.-G., Yang, J., et al. (2020). Immune Cells within the Tumor Microenvironment: Biological Functions and Roles in Cancer Immunotherapy. *Cancer Lett.* 470, 126–133. doi:10.1016/j.canlet.2019.11.009

Li, B., Chan, H. L., and Chen, P. (2019). Immune Checkpoint Inhibitors: Basics and Challenges. *Cmc* 26 (17), 3009–3025. doi:10.2174/0929867324666170804143706

Maj, T., Wang, W., Crespo, J., Zhang, H., Wang, W., Wei, S., et al. (2017). Oxidative Stress Controls Regulatory T Cell Apoptosis and Suppressor Activity and PD-L1-Blockade Resistance in Tumor. *Nat. Immunol.* 18 (12), 1332–1341. doi:10.1038/ni.3868

Moody, S. E., Schinzel, A. C., Singh, S., Izzo, F., Strickland, M. R., Luo, L., et al. (2015). PRKACA Mediates Resistance to HER2-Targeted Therapy in Breast Cancer Cells and Restores Anti-apoptotic Signaling. *Oncogene* 34 (16), 2061–2071. doi:10.1038/onc.2014.153

Nagarajan, K., Soundarapandian, K., Thorne, R. F., Li, D., and Li, D. (2019). Activation of Pyroptotic Cell Death Pathways in Cancer: An Alternative Therapeutic Approach. *Transl. Oncol.* 12 (7), 925–931. doi:10.1016/j.tranon.2019.04.010

Pardoll, D. M. (2012). The Blockade of Immune Checkpoints in Cancer Immunotherapy. *Nat. Rev. Cancer* 12 (4), 252–264. doi:10.1038/nrc3239

Principe, D. R., Chiec, L., Mohindra, N. A., and Munshi, H. G. (2021). Regulatory T-Cells as an Emerging Barrier to Immune Checkpoint Inhibition in Lung Cancer. *Front. Oncol.* 11, 684098. doi:10.3389/fonc.2021.684098

Proto, C., Ferrara, R., Signorelli, D., Lo Russo, G., Galli, G., Imbimbo, M., et al. (2019). Choosing Wisely First Line Immunotherapy in Non-small Cell Lung Cancer (NSCLC): what to Add and what to Leave Out. *Cancer Treat. Rev.* 75, 39–51. doi:10.1016/j.ctrv.2019.03.004

Russo, A. J., and Rathinam, V. A. K. (2018). Lipid Peroxidation Adds Fuel to Pyr(optosis). *Cell Host Microbe* 24 (1), 8–9. doi:10.1016/j.chom.2018.06.010

Saito, M., Suzuki, H., Kono, K., Takenoshita, S., and Kohno, T. (2018). Treatment of Lung Adenocarcinoma by Molecular-Targeted Therapy and Immunotherapy. *Surg. Today* 48 (1), 1–8. doi:10.1007/s00595-017-1497-7

Samstein, R. M., Lee, C.-H., Shoushtari, A. N., Hellmann, M. D., Shen, R., Janjigian, Y. Y., et al. (2019). Tumor Mutational Load Predicts Survival after Immunotherapy across Multiple Cancer Types. *Nat. Genet.* 51 (2), 202–206. doi:10.1038/s41588-018-0312-8

Shao, W., Yang, Z., Fu, Y., Zheng, L., Liu, F., Chai, L., et al. (2021). The Pyroptosis-Related Signature Predicts Prognosis and Indicates Immune Microenvironment Infiltration in Gastric Cancer. *Front. Cell Dev. Biol.* 9, 676485. doi:10.3389/fcell.2021.676485

Shi, J., Zhao, Y., Wang, K., Shi, X., Wang, Y., Huang, H., et al. (2015). Cleavage of GSDMD by Inflammatory Caspases Determines Pyroptotic Cell Death. *Nature* 526 (7575), 660–665. doi:10.1038/nature15514

Shibuya, A., Campbell, D., Hannum, C., Yssel, H., Franz-Bacon, K., McClanahan, T., et al. (1996). DNAM-1, a Novel Adhesion Molecule Involved in the Cytolytic Function of T Lymphocytes. *Immunity* 4 (6), 573–581. doi:10.1016/s1074-7613(00)70060-4

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer Statistics, 2021. *CA A Cancer J. Clin.* 71 (1), 7–33. doi:10.3322/caac.21654

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

Tang, R., Xu, J., Zhang, B., Liu, J., Liang, C., Hua, J., et al. (2020). Ferroptosis, Necroptosis, and Pyroptosis in Anticancer Immunity. *J. Hematol. Oncol.* 13 (1), 110. doi:10.1186/s13045-020-00946-7

Vansteenkiste, J., Crinò, L., Dooms, C., Douillard, J. Y., Faivre-Finn, C., Lim, E., et al. (2014). 2nd ESMO Consensus Conference on Lung Cancer: Early-Stage Non-small-cell Lung Cancer Consensus on Diagnosis, Treatment and Follow-Up. *Ann. Oncol.* 25 (8), 1462–1474. doi:10.1093/annonc/mdu089

Xia, X., Wang, X., Cheng, Z., Qin, W., Lei, L., Jiang, J., et al. (2019). The Role of Pyroptosis in Cancer: Pro-cancer or Pro-"host"? *Cell Death Dis.* 10 (9), 650. doi:10.1038/s41419-019-1883-8

Ye, Y., Dai, Q., and Qi, H. (2021). A Novel Defined Pyroptosis-Related Gene Signature for Predicting the Prognosis of Ovarian Cancer. *Cell Death Discov.* 7 (1), 71. doi:10.1038/s41420-021-00451-x

Zheng, Z., and Li, G. (2020). Mechanisms and Therapeutic Regulation of Pyroptosis in Inflammatory Diseases and Cancer. *Int. J. Mol. Sci.* 21 (4), 1456. doi:10.3390/ijms21041456

Zou, W. (2006). Regulatory T Cells, Tumour Immunity and Immunotherapy. *Nat. Rev. Immunol.* 6 (4), 295–307. doi:10.1038/nri1806

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 Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 14 July 2022
doi: 10.3389/fbioe.2022.806851

Spectrum of BRAF Aberrations and Its Potential Clinical Implications: Insights From Integrative Pan-Cancer Analysis

Qiaoli Yi[1†], Jinwu Peng[2,3†], Zhijie Xu[2,3,4], Qiuju Liang[1], Yuan Cai[2], Bi Peng[2], Qingchun He[5,6] and Yuanliang Yan[1]**

[1]Department of Pharmacy, Xiangya Hospital, Central South University, Changsha, China, [2]Department of Pathology, Xiangya Hospital, Central South University, Changsha, China, [3]Department of Pathology, Xiangya Changde Hospital, Changde, China, [4]National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China, [5]Department of Emergency, Xiangya Hospital, Central South University, Changsha, China, [6]Department of Emergency, Xiangya Changde Hospital, Changde, China

OPEN ACCESS

Edited by:
Andrea Banfi,
University of Basel, Switzerland

Reviewed by:
Monica Neagu,
Victor Babes National Institute of Pathology (INCDVB), Romania
Cheng Zhan,
Fudan University, China

***Correspondence:**
Yuanliang Yan
yanyuanliang@csu.edu.cn
Qingchun He
404113@csu.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Preclinical Cell and Gene Therapy, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: 01 November 2021
Accepted: 30 May 2022
Published: 14 July 2022

Citation:
Yi Q, Peng J, Xu Z, Liang Q, Cai Y, Peng B, He Q and Yan Y (2022) Spectrum of BRAF Aberrations and Its Potential Clinical Implications: Insights From Integrative Pan-Cancer Analysis. Front. Bioeng. Biotechnol. 10:806851. doi: 10.3389/fbioe.2022.806851

B-Raf proto-oncogene serine/threonine-protein kinase (BRAF) is frequently altered in multiple cancer types, and BRAF V600 mutations act as a prime target for precision therapy. Although emerging evidence has investigated the role of BRAF, the comprehensive profiling of BRAF expression, alteration and clinical implications across various cancer types has not been reported. In this study, we used the TCGA dataset, covering 10,967 tumor samples across 32 cancer types, to analyze BRAF abnormal expression, DNA methylation, alterations (mutations and amplification/deletion), and their associations with patient survival. The results showed that BRAF expression, alteration frequency, mutation site distribution, and DNA methylation patterns varied tremendously among different cancer types. The expression of BRAF was found higher in PCPG and CHOL, and lower in TGCT and UCS compared to normal tissues. In terms of pathological stages, BRAF expression was significantly differentially expressed in COAD, KIRC, LUSC, and OV. The methylation levels of BRAF were significantly lower in LUSC, HNSC, and UCEC compared to normal tissue. The expression of BRAF and downstream gene (ETS2) was negatively correlated with methylation levels in various cancers. The overall somatic mutation frequency of BRAF was 7.7% for all cancer samples. Most fusion transcripts were found in THCA and SKCM with distinct fusion patterns. The majority of BRAF mutations were oncogenic and mainly distributed in the Pkinase_Tyr domain of THCA, SKCM, COADREAD, and LUAD. The BRAF mutations were divided into five levels according to the clinical targeted therapy implication. The results showed level 1 was mainly distributed in SKCM, COADREAD, and LUAD, while level 3B in THCA. The overall BRAF CNV frequency was about 42.7%, most of which was gain (75.9%), common in GBM, TGCT, and KIRP. In addition, the forest plot showed that increased BRAF expression was associated with poor patient overall survival in LIHC, OV, and UCEC. Taken together, this study provided a novel insight into the full alteration spectrum of BRAF and its implications for treatment and prognosis.

Keywords: BRAF, gene fusion, alteration, prognosis, pan-cancer

INTRODUCTION

The B-Raf proto-oncogene serine/threonine-protein kinase (BRAF) is located on chromosome 7q34 and encodes a protein which belongs to the RAF family (ARAF, BRAF, and CRAF) of serine/threonine protein kinases. As a direct downstream effector of RAS, BRAF protein plays an important role in regulating the mitogen-activated protein kinase (MAPK)/extracellular signal-regulated kinase (ERK) signaling pathway, which mediates a variety of essential cellular processes, including cell growth, proliferation, differentiation and survival (Ritterhouse and Barletta, 2015; Song et al., 2020; Xu et al., 2020).

Mutations in BRAF, most frequently the valine (V) to glutamate (E) substitution at residue 600 (V600E), are identified as cancer-causing mutations in thyroid carcinoma (THCA) and skin cutaneous melanoma (SKCM). As an oncogenic driver, BRAF V600 mutations account for approximately 60% of all BRAF mutations in cancer patients. In contrast to wild-type BRAF, the constitutively active BRAF V600E mutation dramatically enhances kinase activity in an RAS-independent manner and is independent of protein homodimerization to switch to the highly active state (Samatar and Poulikakos, 2014; Alos et al., 2020). Up to now, the second-generation BRAF inhibitors vemurafenib (PLX4032) and dabrafenib (GSK2118436) were approved by the United States Food and Drug Administration (FDA) for the treatment of patients with metastatic melanomas harboring BRAF V600 mutations (Karoulia et al., 2017). Point mutations are not the only alterations found in BRAF. Fusion transcripts arising from translocations have been identified in melanoma, prostate cancer, gastric cancer, etc. (Palanisamy et al., 2010). The oncogenic potential of BRAF fusions has been attributed to the lack of a crucial N-terminal domain that mediates BRAF autoinhibition. Owing to the deletion of the N-terminal inhibitory domain, gene fusions lead to constitutive dimerization of BRAF protein aberrantly activating the downstream MAPK signaling pathway (Cremolini et al., 2019; Kratz and Deming, 2019).

Since previous studies of BRAF gene alterations in cancers are limited to a single cancer type and/or insufficient sample sizes, an integrative analysis across a variety of tumor types to investigate its function is of particular importance. In this article, we comprehensively analyzed the large dataset from The Cancer Genome Atlas (TCGA) to fill in the gaps. We first systematically profiled BRAF expression, methylation, gene alterations, and its clinical and therapeutic implications across 32 TCGA cancer types covering 10,967 tumor samples. In addition, the survival associations between BRAF expression and prognosis in distinct cancer types were conducted to explore its potential therapeutic implication. In general, our study provided a novel insight into the full alteration spectrum of BRAF and its implications for treatment and prognosis in diverse tumor types.

MATERIALS AND METHODS

Data Acquisition and Reanalysis Using Bioinformatics Tools

The essential bioinformatics tools used in this research could be found in Supplementary Table S1. Tumor Immune Estimation Resource (TIMER2.0) is a comprehensive resource for the systematical analysis of immune infiltrates across diverse cancer types (Li et al., 2020). We studied the differential expression of the BRAF gene between tumor samples and adjacent normal tissues across all TCGA tumors by using the "Gene_DE" module of the TIMER2.0 database. The transcripts per million (TPM) values of transcription factors were log2-converted. For certain cancer types without adjacent normal tissues, we further explored Gene Expression Profiling Interactive Analysis 2 (GEPIA2) portal to investigate the BRAF mRNA expression difference between tumor samples and matched TCGA normal and Genotype-Tissue Expression (GTEx) data. Additionally, the GEPIA2 data portal was also used to generate violin plots of BRAF expression across pathological stages for all TCGA cancer types. The log2 (TPM +1) transformed expression data were applied for the violin plots here. GEPIA2 is an interactive web server for analyzing the RNA sequencing expression data from the TCGA and the GTEx projects and provides customizable functions such as tumor/normal differential expression analysis, survival analysis, and so on (Tang et al., 2019). Next, we analyzed the methylation difference of BRAF and its downstream genes of the MAPK signaling pathway between tumor samples and adjacent normal tissues in various TCGA cancer types by using the "TCGA Cancer-Methylation" module in the Gene Set Cancer Analysis (GSCALite) platform. Furthermore, the correlation between methylation and gene expression of BRAF and downstream genes was also visualized by the GSCALite platform. GSCALite is a web-based analysis platform that integrates cancer genomics data to analyze gene methylation, drug sensitivity, and so on (Liu et al., 2018).

The cBioPortal is an open-access portal that provides an interactive investigation of multidimensional cancer genomics and clinical data (Gao et al., 2013). In this study, we selected the "TCGA PanCancer Atlas Studies" covering 10,967 samples across 32 cancer types to further explore BRAF alterations. Data files including copy-number alterations, mutations, mRNA Expression, Log2 copy-number values (CNV), and clinical data were downloaded from cBioportal. The BRAF mRNA expression data were performed based on RSEM (batch normalized from Illumina HiSeq_RNASeqV2) and then log10 transformed. For the BRAF CNV data, the log-ratio value represents: 2 = amplification; 1 = gain; 0 = diploid; -1 = shallow deletion; and −2 = deep deletion. In addition, BRAF fusion data were collected from the TCGA Fusion Gene Database, which enables researchers to query cancer-associated transcript fusions in an interactive manner (Hu et al., 2018).

The Kaplan-Meier plotter is an open-access online database that enables researchers to assess the effect of a candidate gene on survival analysis in pan-cancer (Gyorffy, 2021). The correlations between BRAF mRNA expression and clinical prognosis across various cancer types were analyzed with the "Start KM Plotter for pan-cancer" module. Data including hazard ratio (HR), *p*-value, and 95% confidence interval (CI) were collected to draw the forest plots to summarize survival analysis.

Statistical Analyses

The statistical analysis was performed with Graphpad PRISM software Version 8.0. Student's t-test, Cox regression analysis, and linear regression analysis were conducted when appropriate. $p < 0.05$ was defined as a statistically significant difference.

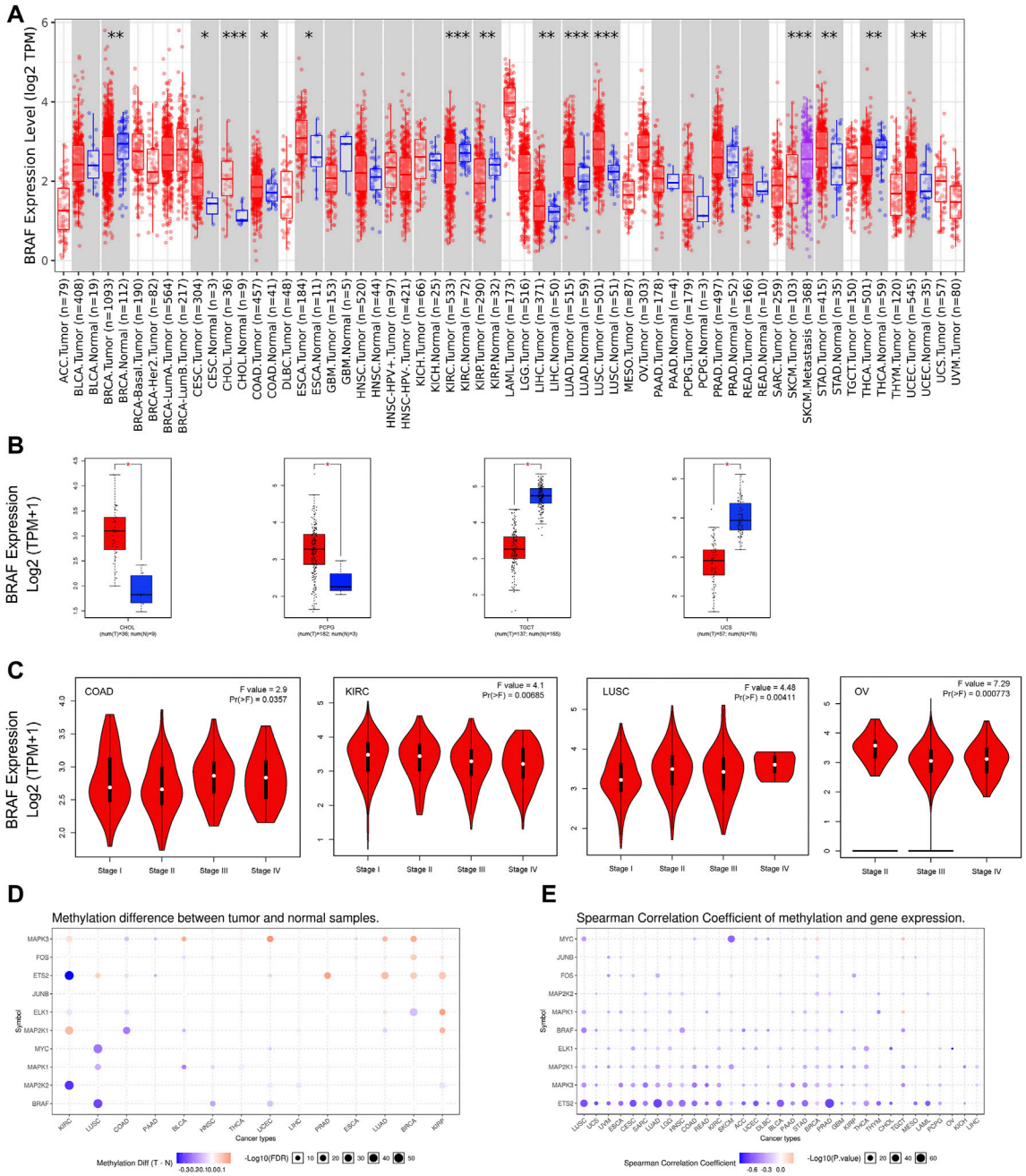


FIGURE 1 | BRAF mRNA expression and DNA methylation in TCGA tumor tissues. **(A)** BRAF mRNA expression across different cancer types by TIMER2. The log2 [TPM (Transcripts per million)] was applied for the log-scale. **(B)** Differential expression of BRAF between tumors samples and normal tissues using combined data from TCGA and GTEx datasets based on the GEPIA2 portal. BRAF expression was up-regulated in CHOL and PCPG, but down-regulated in TGCT and UCS. The log2 (TPM + 1) was applied for log-scale. **(C)** Differential expression of BRAF in different pathological stages of COAD, KIRC, LUSC, and OV. The log2 (TPM + 1) was applied for log-scale. **(D)** Bubble map depicting the methylation difference of BRAF and its downstream genes between tumors and normal samples. Blue dots indicate down-regulated methylation in tumors. Red dots indicate up-regulated methylation. **(E)** Bubble map exhibiting correlations between methylation and gene expression of BRAF and its downstream genes. Blue dots denote down-regulated methylation in tumors. Blue dots represent the negative Spearman correlation coefficient, and red dots represent the positive. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

RESULTS

Expression and Methylation Level of BRAF in Pan-Cancer

Aberrant expression of BRAF gene has been demonstrated in various cancer types (Wang et al., 2021; Zhao et al., 2021). Previous studies on BRAF expression in cancer have used inconsistent research methods and have been limited to small sample sizes and/or to single or limited cancer types. In this study, we conducted a more comprehensive analysis of BRAF expression in pan-cancer. At the outset, we explored the mRNA expression pattern of BRAF between tumor samples and adjacent normal tissues in pan-cancer by TIMER2.0 (**Figure 1A**). Compared with the corresponding adjacent normal tissues or metastatic lesions, significantly differential expression of BRAF was found in 14 cancer types, with 9 tumor types up-regulated (cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), stomach adenocarcinoma (STAD) and uterine corpus endometrial carcinoma (UCEC)) and 5 tumor types down-regulated (breast invasive carcinoma (BRCA), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), SKCM, THCA). After adding GTEx normal tissue samples as a control, we further investigated BRAF differential expression by GEPIA2. As shown in **Figure 1B**, BRAF expression was up-regulated in pheochromocytoma and paraganglioma (PCPG) and CHOL, down-regulated in testicular germ cell tumors (TGCT) and uterine carcinosarcoma (UCS). We further analyzed the correlation between BRAF expression and pathological stages in pan-cancer. As shown in **Figure 1C**, we found that BRAF expression was correlated with pathological stages in several tumor types, including COAD, KIRC, LUSC, and ovarian serous cystadenocarcinoma (OV) ($p < 0.05$).

A growing body of evidence has suggested that DNA methylation is strongly correlated with gene alteration in cancers (Wang et al., 2020; Zhang et al., 2020). Therefore, we searched the methylation profiles of BRAF and its downstream genes in TCGA cancers by using the GSCALite database (**Figures 1D,E**). The results indicated that the methylation of BRAF was down-regulated in LUSC, head and neck squamous cell carcinoma (HNSC), and UCEC. Then, we evaluated the correlation between methylation and BRAF expression in pan-cancer. The results revealed that the expression profiles of BRAF and downstream genes were generally negatively correlated with methylation in various cancers.

BRAF Somatic Mutation Patterns in Pan-Cancer

The overall somatic mutation frequency of BRAF was 7.7% for all cancer samples (848/10,976) and 7.0% for all patients (767/10,953) across the 32 TCGA cancer types. And the detailed information on 848 BRAF somatic mutations was shown in **Supplementary Table S2**. The sample size of each tumor type varied from 36 (CHOL) to 1,084 (BRCA), and the cancer types with a small sample size might not reflect the general spectrum of BRAF mutation status (**Supplementary Table S3**). As shown in **Figure 2A**, the most frequent cancer types with BRAF mutations were THCA (59.3%), SKCM (53.6%), colon adenocarcinoma/rectum adenocarcinoma (COADREAD) (10.6%), LUAD (7.2%) and UCEC (4.7%). Instead, almost no BRAF mutations were observed in kidney chromophobe (KICH), acute myeloid leukemia (LAML), LIHC, TGCT, thymoma (THYM), and uveal melanoma (UVM).

Based on the Pfam database (http://pfam.xfam.org/protein/braf_human), BRAF harbors 3 functional domains, including the RBD domain (156-227 aa), C1_1 domain (235-280 aa), and Pkinase_Tyr domain (458-712 aa). The 848 BRAF somatic mutations were observed in various cancer types and widely distributed across different functional domains of the BRAF gene. The most common one was the Pkinase_Tyr domain (722 samples), followed by the other domains whose functions were barely known (75 samples), the C1_1 domain (11 samples), and the RBD domain (8 samples). Fusions (32 samples) were also observed in BRAF somatic mutations across all cancer types. The location distribution of BRAF mutations was dramatically different among numerous cancers. Mutations in THCA, SKCM, COADREAD, and LUAD were most frequently distributed in the Pkinase_Tyr domain. However, mutations in UCEC were predominantly located in the other domains amounting to half of the total mutations. Furthermore, fusions were mainly distributed in THCA and SKCM (**Figure 2B** and **Supplementary Table S4**).

Fusion genes generated by cleavage and re-splicing at the genome level are often the targets for tumor diagnosis and treatment. We analyzed fusion transcripts of BRAF across various cancer types by using the TCGA Fusion Gene Database (**Figure 3**). BRAF fusion transcripts were detected in THCA (17), SKCM (9), prostate adenocarcinoma (PRAD) (3), pancreatic adenocarcinoma (PAAD) (2), READ (2), LIHC (1), LUSC (1), STAD (1), KIRP (1), brain lower-grade glioma (LGG) (1) and bladder urothelial carcinoma (BLCA) (1). The highest number of fusion transcripts was found in THCA (three SND1_BRAF, one BRAF_SND1, one AGK_BRAF, one BRAF_AGK, one MACF1_BRAF, one BRAF_MACF1, one FAM114A2_BRAF, one BRAF_ FAM114A2, one CCNY_BRAF, one MKRN1_BRAF, etc.). AGK_BRAF and BRAF_AGK were also detected in SKCM. The vast majority of these BRAF fusion transcripts were classified as in-frame, while three BRAF fusion transcripts (one BRAF_HIBADH and one HIBADH_BRAF in SKCM, one TMPRSS2_BRAF in PRAD) were classified as out-of-frame and three (one BRAF_MRPS33 in BLCA and one in STAD, one BRAF_CUL1 in KIRP) were classified as CDS-5UTR.

According to functional impact on protein coding, these 848 BRAF somatic mutations could be classified into four categories: missense (778 mutations), truncating (32 mutations), fusion (32 samples), and in-frame (6 mutations) (**Supplementary Figure S1A**). The 600aa in the Pkinase_Tyr domain was the most mutated position, which was observed in 590 samples (545 samples with V600E, 39 with V600M, 4 with V600G, 1 with

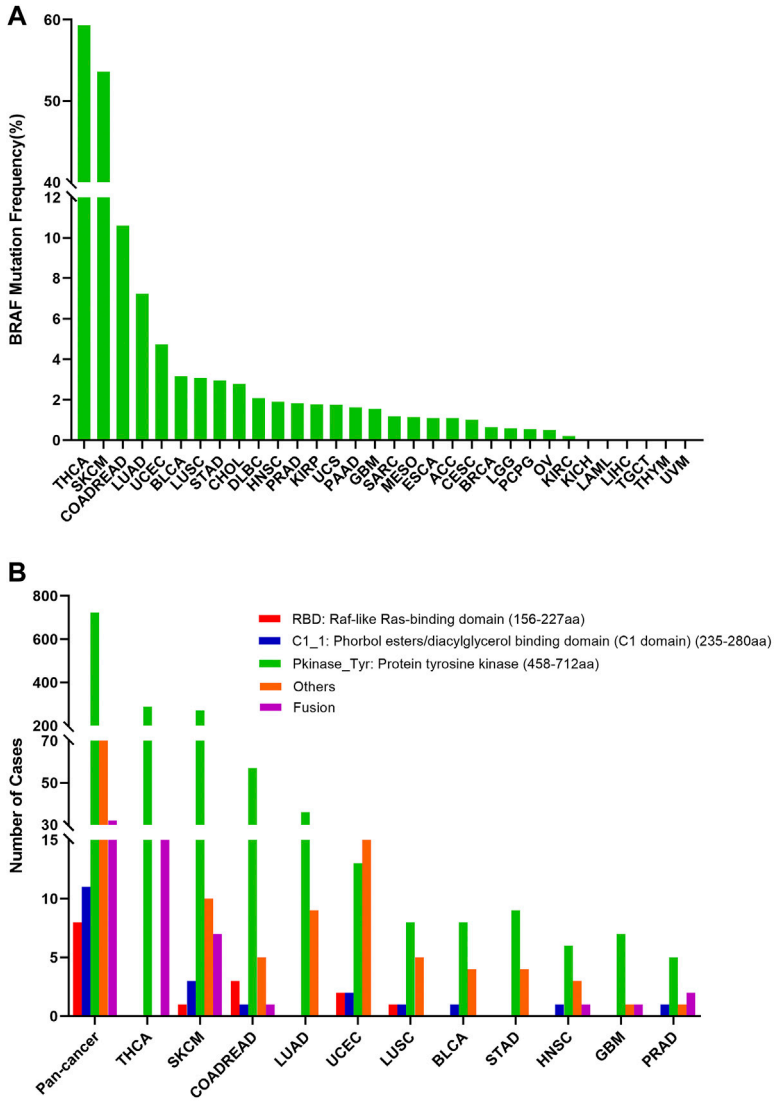


FIGURE 2 | BRAF mutation distribution in various cancer types and protein functional domains. **(A)** BRAF mutation frequency across 32 TCGA cancer types. **(B)** BRAF mutation distribution in different protein functional domains in all and top 10 tumor types. Abbreviation: aa: amino acid.

V600_K601delinsE and 1 with V600_R603del). Mutations at V600E were primarily distributed in THCA samples (284/545) and SKCM samples (193/545) (**Supplementary Figure S1B–C**). V600E is known to be oncogenic and serves as target for BRAF inhibitors approved by FDA, while other mutation types (V600M/G) are likely oncogenic. Other tumors with mutations at this position were COADREAD (48 samples), LUAD (9 samples), glioblastoma multiforme (GBM) (5 samples), KIRP (2 samples), BLCA (1 sample), LGG (1 sample), CHOL (1 sample) and HNSC (1 sample). Studies on their role in COADREAD and LUAD are underway (Planchard et al., 2017; Kopetz et al., 2019) and its function remains little known in other cancer types.

Based on the oncogenic effect and predictive significance, the 848 BRAF somatic mutations could be classified into four categories. As shown in **Figure 4A**, 616 (72.6%) BRAF mutations were oncogenic, 98 (11.6%) likely oncogenic, 1 (0.1%) inconclusive and 133 (15.7%) unknown. Although a major portion of BRAF somatic mutations was distributed in the functional categories, there were still some mutations in the unknown class deserving further study to characterize the potential functional significances of these mutations. As displayed in **Figure 4B**, mutations distributed in the functional categories comprised the majority of BRAF mutations in several cancers such as THCA, SKCM, COADREAD, and LUAD. However, more than two-thirds of mutations belonged to the unknown class in UCEC, LUSC, BRCA, and STAD.

The 848 BRAF mutations could be divided into five levels by the clinical targeted therapy implication, containing level NA (193 mutations), level 4 (22 mutations), level 3B (326 mutations), and level 3A (14 mutations), and level 1 (293 mutations). Only

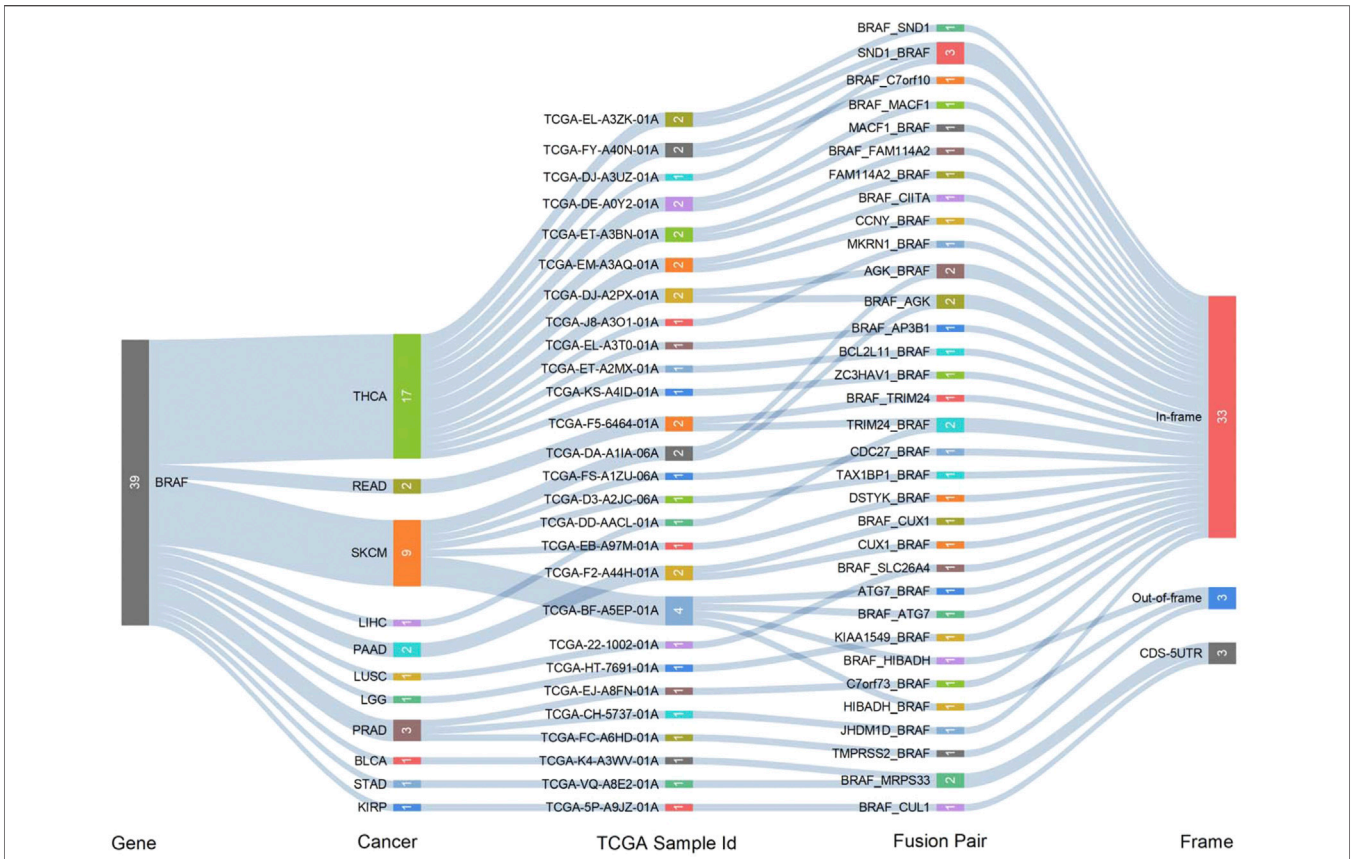


FIGURE 3 | Gene fusions of BRAF across different cancer types.

level 1 mutations are indicated for targeted therapy with FDA-approved drugs. All level 1 mutations were found in SKCM (236), COADREAD (48), and LUAD (9). These mutations (including V600E, V600M, and V600G) were concentrated in 600aa of the Pkinase_Tyr domain. Although all mutations in THCA belonged to oncogenic (289/302)/likely oncogenic (13/302), all of them were in level 3B (301/302) and level NA (1/302) without treatment implications (**Figures 5A,B**).

BRAF CNVs in Different Tumor Types

The overall BRAF CNV frequency was about 42.7% (4,684/10,967 samples). The majority of BRAF CNV types were gain (3,555 samples), followed by shallow deletion (972 samples), application (133 samples), and deep deletion (24 samples). The most common tumor types with BRAF CNVs were GBM (81.3%), TGCT (69.1%), ESCA (63.7%), OV (62.9%), adrenocortical carcinoma (ACC) (62.0%), KIRP (61.1%), UCS (59.6%) and LUSC (59.5%). By contrast, THYM (16.3%), PCPG (15.7%), LAML (12.0%), UVM (8.8%) and THCA (5.2%) harbored a relatively low BRAF CNV frequency (**Figure 6A**). Among the 848 samples with BRAF mutations described above, 276 also had BRAF CNV changes, of which 236 with gain, 23 with amplification, 14 with shallow deletion, and 3 with deep deletion. SKCM had the highest numbers of amplification or gain among different cancer types. Meanwhile, SKCM and LUSC were the two cancer types with the highest numbers of shallow deletion, and deep deletion only occurred in THCA (**Figure 6B**, **Supplementary Table S2**).

BRAF mRNA expression was compared across 32 TCGA cancer types and exhibited a relatively consistent trend, suggesting that there may be a common mechanism to promote BRAF expression. Based on the interquartile range, BRAF expression was widely distributed in COADREAD and SKCM, and narrowly distributed in UVM and mesothelioma (MESO), which may be attributed to the fact that some cancer types had more than one subtype and therefore more genetic diversity (**Supplementary Figure 2A**). In addition, we explored the correlation between BRAF mRNA expression and CNVs. The results showed that there was a positive correlation between BRAF mRNA expression and CNVs in pan-cancer (r = 0.1716, p < 0.0001) (**Supplementary Figure 2B**).

Combined BRAF Alterations (CNVs and Mutation) Across Various Cancer Types

The combined BRAF alteration (CNV and mutation) frequency in all cancers was about 8.3% (905 of 10,967 samples). As shown in **Figure 7A**, BRAF alteration frequency among various cancer types was dramatically different. While KICH, TGCT, THYM, and UVM had neither BRAF mutation nor BRAF CNVs, BRAF

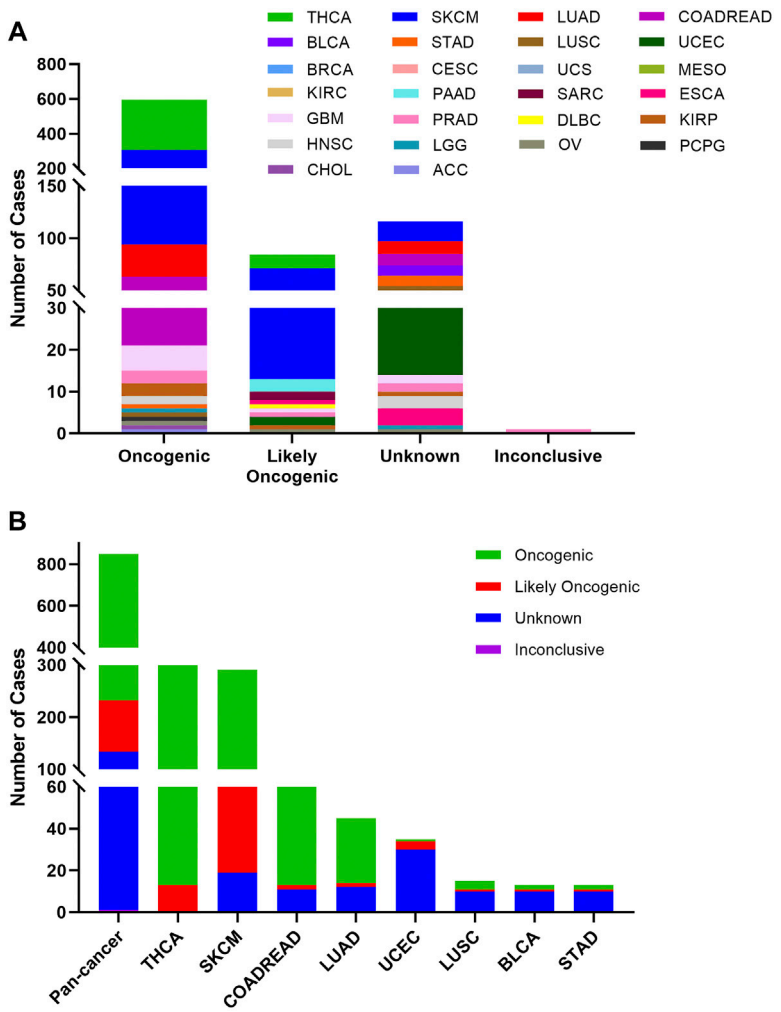


FIGURE 4 | BRAF mutation classification by functional impacts. **(A)** BRAF mutations based on functional impacts on all tumors together. **(B)** Functional impact class distribution of BRAF mutations in pan-cancer and the top 8 cancer types.

alterations were most frequently observed in THCA (59.6%) and SKCM (53.8%), in which mutation was more common than CNV, with BRAF mutation rates of 57.4 and 49.3%, respectively. Other cancers with dominant BRAF mutation but at much lower mutation rate included COADREAD (10.4%), LUAD (7.2%), UCEC (4.7%), BLCA (3.2%), STAD (3.0%), LUSC (2.9%), CHOL (2.8%) and DLBC (2.1%). Amplification was more common in OV (7.7 vs. 0.3%), SARC (1.6 vs. 0.4%) and LGG (1.4 vs. 0.6%). Deep deletion was mainly distributed in LAML (1.5%), SARC (1.2%) and ESCA (1.1%).

BRAF mutation location and its CNVs occurrence appeared to be associated. 201 of 664 (30.3%) mutations in the Pkinase_Tyr domain and 21 of 75 (28%) mutations in the other function-unknown domain were accompanied by copy gain. Amplification was mainly distributed in the Pkinase_Tyr domain. Mutations in the RBD domain and C1_1 domain rarely had concurrent CNVs (**Figure 7B**).

BRAF Expression and the Prognosis of Cancer Patients

In order to assess the clinical significance of BRAF expression, we analyzed patient survival in pan-cancer and showed that increased BRAF expression was associated with poor patient overall survival (OS) in LIHC, OV, and UCEC. Interestingly, increased BRAF expression was correlated with better prognosis in BRCA, HNSC, and KIRC (**Figure 8A**). In addition, survival analysis between BRAF expression and patient relapse-free survival (RFS) across various cancer types exhibited that increased BRAF expression was associated with poor RFS in LIHC, LUSC, and UCEC, while high BRAF expression was correlated with better RFS in BRCA and OV (**Figure 8B**). The contradictory results in OV may be attributed to insufficient sample size and diverse genetic backgrounds.

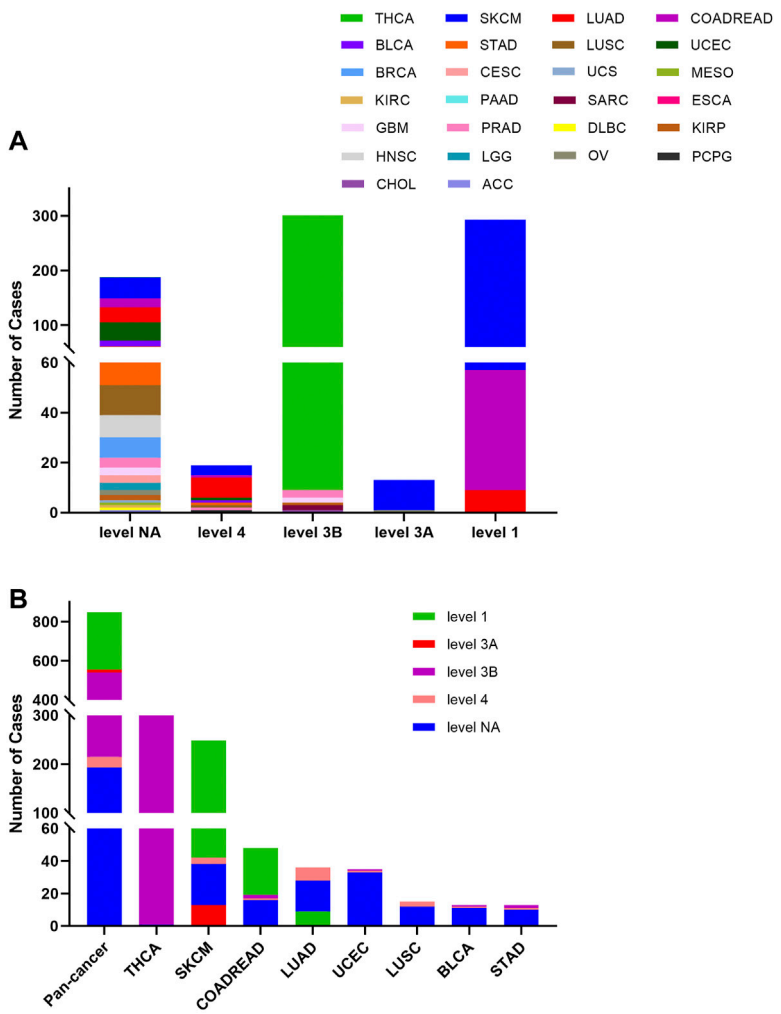


FIGURE 5 | BRAF mutation distribution according to clinical therapeutic implications. **(A)** BRAF mutation distribution based on the OncoKB therapeutic evidence levels among diverse cancer types. **(B)** OncoKB therapeutic evidence levels distribution of BRAF mutation in pan-cancer and the top 8 tumor types.

DISCUSSION

In this study, we profiled the characteristics of BRAF in 32 TCGA cancer types by using the cBioPortal tool and the results showed that BRAF expression, methylation, mutations, locations, and CNVs dramatically differed among diverse cancer types, which had significant clinical implications.

DNA methylation, characterized as a methyl group added to cytosines in cytosine-guanine (CpG) dinucleotides, is one of the key epigenetic modifications involved in the regulation of gene expression (Müller and Győrffy, 2022). An existing body of evidence indicates that aberrant DNA methylation leads to activation of oncogenes and silencing of tumor suppressor genes, contributing to tumorigenesis and progression (Klutstein et al., 2016). In a recent study by Noreen et al. (2019), the silencing of Tet methylcytosine dioxygenase 1 (TET1) oxidative DNA demethylase mediated by BRAF V600E mutation was responsible for the initiation of colon cancers with CpG-island methylator phenotype (CIMP). Furthermore, Weisenberger's group (Weisenberger et al., 2006) also identified the tight association between BRAF mutation and CIMP in colorectal cancers. Our results indicated that the methylation of BRAF and downstream genes were correlated with tumor occurrence. Thus, the potential roles of BRAF mutation in the regulation of DNA methylation and tumor initiation deserve further investigation.

Gene fusions originating from the concatenation of two separate genes caused by trans-splicing events or chromosomal translocations may provide fundamental insights into tumorigenesis and progression (Mertens et al., 2015). In this analysis, the distribution patterns of gene fusions involving BRAF varied in pan-cancer. THCA harbored the highest number of fusion transcripts, of which SND1_BRAF was the most common one, while other cancer types such as LIHC, LUSC, STAD, KIRP, LGG, and BLCA had equally few fusion transcripts. The ectopic expression of SND1_BRAF could increase the phosphorylation levels of MEK/ERK and cell proliferation (Jang et al., 2015). However, MEK inhibitors display expected response rates of up to

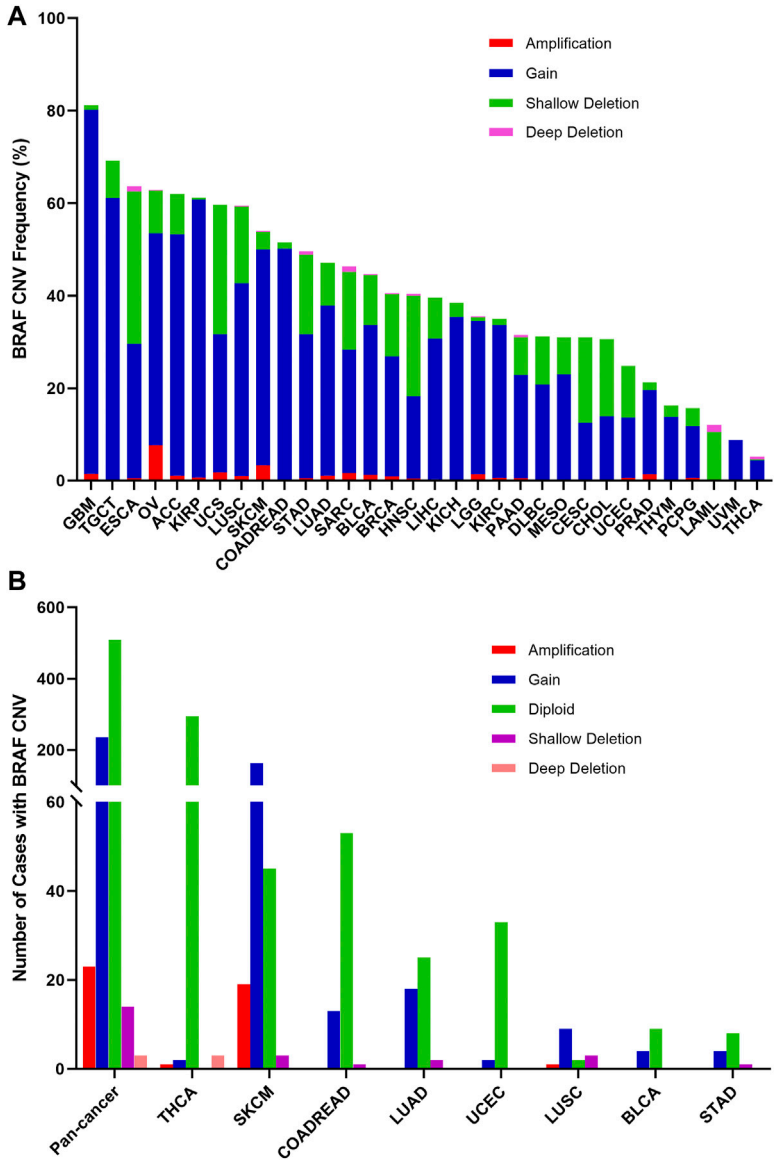


FIGURE 6 | BRAF Copy Number Variant (CNV) distribution across all and selected top 8 cancer types. **(A)** BRAF CNV frequency in 32 TCGA cancer types. **(B)** BRAF CNV distribution in pan-cancer and the top 8 cancer types for the cases with EGFR mutations.

70% of patients with higher BRAFV600, while a randomized phase 2 trial showed no differences in overall survival (Algazi et al., 2020). Combination treatments with BRAF fusions and MEK inhibitors may propose a novel insight to evaluate the effectiveness of chemotherapy in cancers.

Mutations in the Pkinase_Tyr domain accounted for most of BRAF single nucleotide or insertion/deletion (indel) mutations. However, the Pkinase_Tyr domain was far more critical in terms of targeted therapy with BRAF inhibitors as approximately 90% of BRAF mutations in SKCM occurred in this region, particularly the BRAF V600E mutation in exon 15. Mutations in this region have been shown to be predictive markers for effective BRAF inhibitors therapy for SKCM in clinical practice (Chapman et al., 2011; McArthur et al., 2014), with significantly longer survival compared to traditional combination chemotherapy. Vemurafenib (PLX4032) and dabrafenib (GSK2118436) were approved by the FDA for the treatment of melanoma patients with BRAF V600E mutation in August 2011 and May 2013, respectively, (Bollag et al., 2012; Hauschild et al., 2012), marking a significant milestone in precision medicine for advanced melanoma. However, due to the complicated pathogenesis of cancer, most patients develop acquired resistance after several months of monotherapy (Shi et al., 2014), combination therapy holds promise as an effective therapeutic strategy. Compared with BRAF inhibitors alone, combining BRAF and MEK inhibitors have been demonstrated to enhance antitumor activity and delay the emergence of drug resistance in patients who have melanoma harboring BRAF V600E mutations, without increased overall

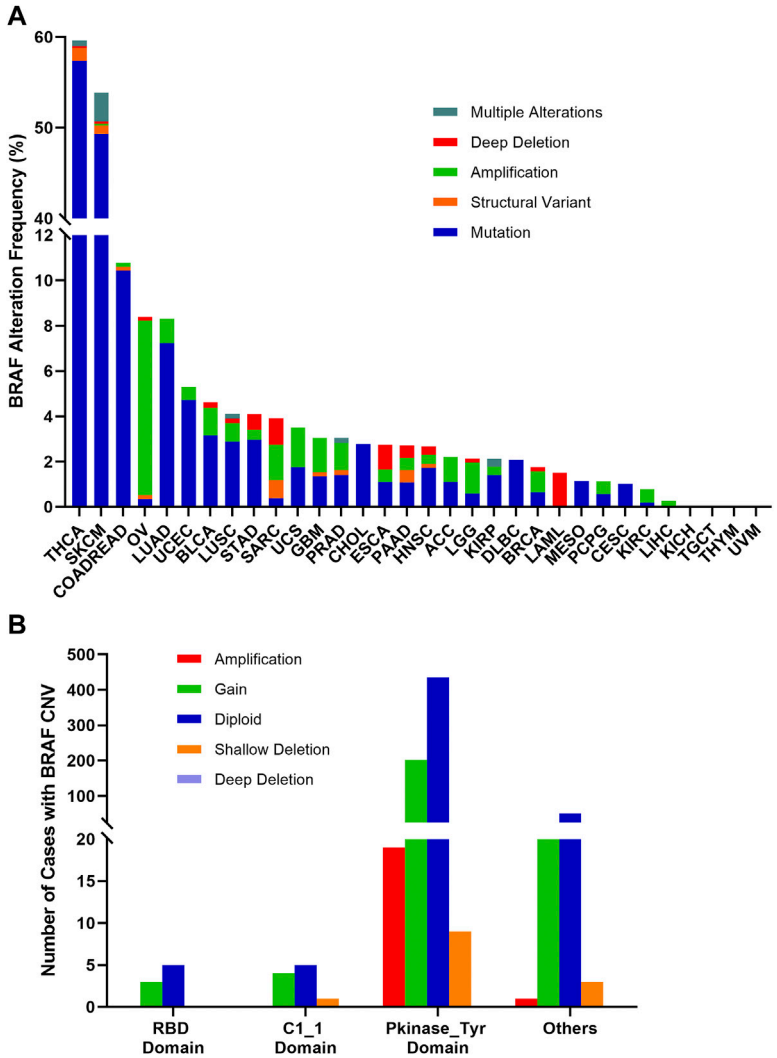


FIGURE 7 | BRAF alterations and distribution in pan-cancer. **(A)** BRAF alteration (mutations and CNVs) frequency across 32 TCGA tumor types. **(B)** BRAF CNVs distribution along with mutations located in different protein functional domains.

toxicity (Robert et al., 2019). In addition, studies assessing triple therapy, BRAF, and MEK inhibitors in combination with immunotherapies, are ongoing (Gutzmer et al., 2020). The BRAF mutation rate of SKCM in this TCGA cohort appeared to be consistent with previous reports and V600E mutations accounted for approximately 95% of all V600 mutated SKCM tumors (Cancer Genome Atlas Network, 2015). For other less frequent BRAF mutations in SKCM, except for the V600K mutation, targeted therapy generated inconsistent results (Menzer et al., 2019). It is clear that different BRAF mutations have different implications, and only those resulting in hyperactivated RAF in MAPK signaling pathway may benefit from BRAF targeted therapy (Yao et al., 2015).

The BRAF V600E mutation in THCA was mainly distributed in papillary thyroid cancer (PTC) and anaplastic thyroid carcinoma (ATC). As previously reported (Chou et al., 2022; Fallahi et al., 2022), PTC and ATC harbored drastically different clinical outcomes while the 5-year survival rate approached 100% in PTC and only 7% in ATC. So, the researches to discover the function of BRAF V600E mutation in ATC are important. Interestingly, the combination therapy of dabrafenib and trametinib was already approved by the FDA in May 2018 for ATC patients with BRAF V600E mutation based on phase II clinical trial (Haddad et al., 2018; Subbiah et al., 2018; Park et al., 2021). Furthermore, the correlation between the BRAF V600E mutation and high-risk clinicopathological features of PTC has been identified in some studies, especially when coexisting with a TERT promoter mutation could remarkably increase transcriptional activities (Liu et al., 2017). As for the PTC patients, studies on BRAF inhibitors have mainly focused on radioactive iodine refractory differentiated thyroid cancer, and relevant clinical trials are in progress (NCT02145143; NCT04462471; NCT04940052; NCT01947023; NCT04554680; NCT05182931).

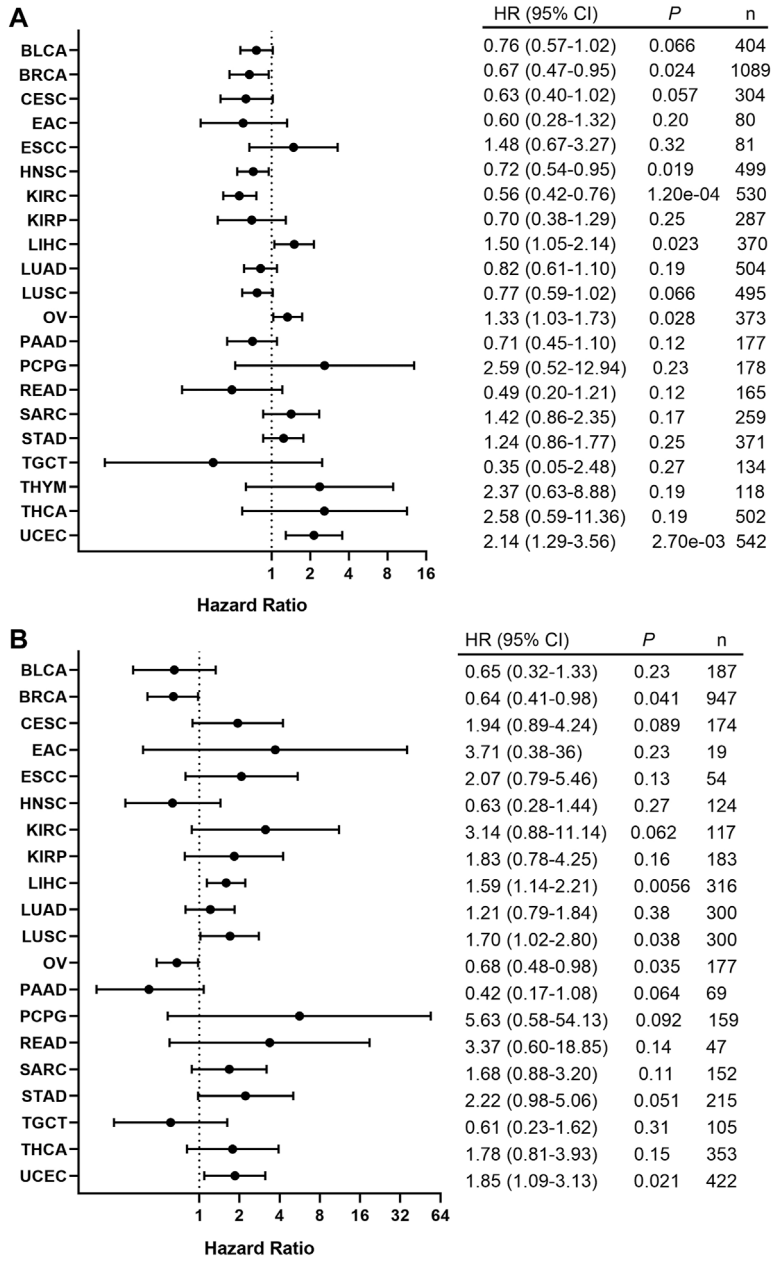


FIGURE 8 | Correlation between the BRAF expression and patient survival. **(A)** Forest plot of the association between BRAF expression and overall survival (OS) based on Kaplan-Meier Plotter. **(B)** Forest plot of the association between BRAF expression and relapse-free survival (RFS) based on Kaplan-Meier Plotter.

Although BRAF mutation is an oncogenic driver in multiple cancers, a single-agent BRAF inhibitor has limited clinical efficacy in BRAF V600E-mutated COADREAD patients (Kopetz et al., 2015). This has been attributed predominantly to the rapid reactivation of the MAPK pathway through the epidermal growth factor receptor (EGFR) (Prahallad et al., 2012). Interestingly, combination therapy of BRAF inhibition (encorafenib) and anti-EGFR monoclonal antibodies (cetuximab) has been well validated and approved by FDA in BRAF V600E-mutated metastatic COADREAD patients (BEACON CRC) (Kopetz et al., 2019). Compared with the cetuximab and traditional irinotecan-based chemotherapy, doublet-therapy with encorafenib and cetuximab showed a significant survival benefit, as well as triplet-therapy group with binimetinib (MEK inhibitor). Nowadays, single or combination treatments of encorafenib have been explored in various BRAF mutant cancers, such as SKCM, PAAD, LUAD, COADREAD, THCA, and other advanced solid tumors in a clinical trial (NCT05003622; NCT04390243; NCT05195632; NCT04673955; NCT04061980; NCT03973918). Regarding the

expected efficiency of immune checkpoint inhibitors, patients treated with BRAF inhibitors and pembrolizumab or nivolumab are validated in ongoing clinical trials (NCT05217446; NCT04044430; NCT04017650). Although OV and LUAD harbored similar alteration frequencies, the functional profiles of BRAF in these two cancers were quite different, with different amplification and mutation patterns. Studies have shown that BRAF inhibitors alone or in combination with MEK inhibitors are effective as a second-line treatment in patients with BRAFV600-mutated LUAD (Mazieres et al., 2020). BRAF mutation may be a good prognostic factor in OV (Kaldawy et al., 2016), and additional studies will be required to further characterize the clinical significance.

In this study, we profiled BRAF expression, fusion transcript, alteration, and the prognostic and clinical implications across 32 TCGA cancer types. However, some limitations needed to be mentioned. Firstly, some rare tumor types did not have sufficient sample sizes to capture the full BRAF expression and alteration spectrum to establish moderate associations. The low frequency of BRAF mutation or amplification also made this analysis challenging. Moreover, it was mainly a pan-cancer investigation without in-depth dive into each cancer type.

CONCLUSION

Our study provides a comprehensive view of BRAF expression, alteration, and clinical prognostic implications across 32 cancer types covering more than ten thousand tumor samples. While some BRAF alternations are involved more in carcinogenesis, others are more therapeutic. Some cancer types have a higher BRAF alternation frequency and its abnormal expression is associated with prognosis. Genomic profiling of BRAF may guide its use in targeted therapy.

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

Conception and Design: YY and JP. Writing, review, and/or revision of the manuscript: QY, ZX, QL, YC, and BP. Administrative, technical, or material support: JP and QH. All authors approved the final version of the manuscript.

FUNDING

This study is supported by grants from the Natural Science Foundation of Hunan Province (2020JJ5934, 2021JJ30904), the Hunan Science and Technology Department Project (2013SK3011), and the Horizontal Project (1 43010100).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2022.806851/full#supplementary-material

REFERENCES

Algazi, A. P., Othus, M., Daud, A. I., Lo, R. S., Mehnert, J. M., Truong, T.-G., et al. (2020). Continuous versus Intermittent BRAF and MEK Inhibition in Patients with BRAF-Mutated Melanoma: a Randomized Phase 2 Trial. *Nat. Med.* 26 (10), 1564–1568. doi:10.1038/s41591-020-1060-8

Alos, L., Fuster, C., Castillo, P., Jares, P., Garcia-Herrera, A., Marginet, M., et al. (2020). TP53 Mutation and Tumoral PD-L1 Expression Are Associated with Depth of Invasion in Desmoplastic Melanomas. *Ann. Transl. Med.* 8 (19), 1218. doi:10.21037/atm-20-1846

Bollag, G., Tsai, J., Zhang, J., Zhang, C., Ibrahim, P., Nolop, K., et al. (2012). Vemurafenib: the First Drug Approved for BRAF-Mutant Cancer. *Nat. Rev. Drug Discov.* 11 (11), 873–886. doi:10.1038/nrd3847

Cancer Genome Atlas Network (2015). Genomic Classification of Cutaneous Melanoma. *Cell* 161 (7), 1681–1696. doi:10.1016/j.cell.2015.05.044

Chapman, P. B., Hauschild, A., Robert, C., Haanen, J. B., Ascierto, P., Larkin, J., et al. (2011). Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation. *N. Engl. J. Med.* 364 (26), 2507–2516. doi:10.1056/NEJMoa1103782

Chou, R., Dana, T., Haymart, M., Leung, A. M., Tufano, R. P., Sosa, J. A., et al. (2022). Active Surveillance versus Thyroid Surgery for Differentiated Thyroid Cancer: A Systematic Review. *Thyroid* 32, 351–367. doi:10.1089/thy.2021.0539

Cremolini, C., Moretto, R., Zucchelli, G., and Falcone, A. (2019). BRAF Mutant Metastatic Colorectal Cancers: New Arrows in Our Quiver. *Ann. Transl. Med.* 7 (Suppl. 8), S367. doi:10.21037/atm.2019.08.118

Fallahi, P., Ferrari, S. M., Galdiero, M. R., Varricchi, G., Elia, G., Ragusa, F., et al. (2022). Molecular Targets of Tyrosine Kinase Inhibitors in Thyroid Cancer. *Seminars Cancer Biol.* 79, 180–196. doi:10.1016/j.semcancer.2020.11.013

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

Gutzmer, R., Stroyakovskiy, D., Gogas, H., Robert, C., Lewis, K., Protsenko, S., et al. (2020). Atezolizumab, Vemurafenib, and Cobimetinib as First-Line Treatment for Unresectable Advanced BRAF(V600) Mutation-Positive Melanoma (IMspire150): Primary Analysis of the Randomised, Double-Blind, Placebo-Controlled, Phase 3 Trial. *Lancet* 395 (10240), 1835–1844. doi:10.1016/S0140-6736(20)30934-X

Gyorffy, B. (2021). Survival Analysis across the Entire Transcriptome Identifies Biomarkers with the Highest Prognostic Power in Breast Cancer. *Comput. Struct. Biotechnol. J.* 19, 4101–4109. doi:10.1016/j.csbj.2021.07.014

Haddad, R. I., Nasr, C., Bischoff, L., Busaidy, N. L., Byrd, D., Callender, G., et al. (2018). NCCN Guidelines Insights: Thyroid Carcinoma, Version 2.2018. *J. Natl. Compr. Canc Netw.* 16 (12), 1429–1440. doi:10.6004/jnccn.2018.0089

Hauschild, A., Grob, J.-J., Demidov, L. V., Jouary, T., Gutzmer, R., Millward, M., et al. (2012). Dabrafenib in BRAF-Mutated Metastatic Melanoma: a Multicentre, Open-Label, Phase 3 Randomised Controlled Trial. *Lancet* 380 (9839), 358–365. doi:10.1016/S0140-6736(12)60868-X

Hu, X., Wang, Q., Tang, M., Barthel, F., Amin, S., Yoshihara, K., et al. (2018). TumorFusions: an Integrative Resource for Cancer-Associated Transcript Fusions. *Nucleic Acids Res.* 46 (D1), D1144–D1149. doi:10.1093/nar/gkx1018

Jang, J. S., Lee, A., Li, J., Liyanage, H., Yang, Y., Guo, L., et al. (2015). Common Oncogene Mutations and Novel SND1-BRAF Transcript Fusion in Lung Adenocarcinoma from Never Smokers. *Sci. Rep.* 5, 9755. doi:10.1038/srep09755

Kaldawy, A., Segev, Y., Lavie, O., Auslender, R., Sopik, V., and Narod, S. A. (2016). Low-grade Serous Ovarian Cancer: A Review. *Gynecol. Oncol.* 143 (2), 433–438. doi:10.1016/j.ygyno.2016.08.320

Karoulia, Z., Gavathiotis, E., and Poulikakos, P. I. (2017). New Perspectives for Targeting RAF Kinase in Human Cancer. *Nat. Rev. Cancer* 17 (11), 676–691. doi:10.1038/nrc.2017.79

Klutstein, M., Nejman, D., Greenfield, R., and Cedar, H. (2016). DNA Methylation in Cancer and Aging. *Cancer Res.* 76 (12), 3446–3450. doi:10.1158/0008-5472.CAN-15-3278

Kopetz, S., Desai, J., Chan, E., Hecht, J. R., O'Dwyer, P. J., Maru, D., et al. (2015). Phase II Pilot Study of Vemurafenib in Patients with Metastatic BRAF-Mutated Colorectal Cancer. *J. Clin. Oncol.* 33 (34), 4032–4038. doi:10.1200/JCO.2015.63.2497

Kopetz, S., Grothey, A., Yaeger, R., Van Cutsem, E., Desai, J., Yoshino, T., et al. (2019). Encorafenib, Binimetinib, and Cetuximab in BRAF V600E-Mutated Colorectal Cancer. *N. Engl. J. Med.* 381 (17), 1632–1643. doi:10.1056/NEJMoa1908075

Kratz, J. D., and Deming, D. A. (2019). The Evolving Treatment Paradigm for BRAF V600 Mutant Colorectal Cancer. *Ann. Transl. Med.* 7 (Suppl. 8), S257. doi:10.21037/atm.2019.12.61

Li, T., Fu, J., Zeng, Z., Cohen, D., Li, J., Chen, Q., et al. (2020). TIMER2.0 for Analysis of Tumor-Infiltrating Immune Cells. *Nucleic Acids Res.* 48 (W1), W509–W514. doi:10.1093/nar/gkaa407

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., Bishop, J., Zhu, G., Zhang, T., Ladenson, P. W., and Xing, M. (2017). Mortality Risk Stratification by Combining BRAF V600E and TERT Promoter Mutations in Papillary Thyroid Cancer: Genetic Duet of BRAF and TERT Promoter Mutations in Thyroid Cancer Mortality. *JAMA Oncol.* 3 (2), 202–208. doi:10.1001/jamaoncol.2016.3288

Mazieres, J., Cropet, C., Montané, L., Barlesi, F., Souquet, P. J., Quantin, X., et al. (2020). Vemurafenib in Non-small-cell Lung Cancer Patients with BRAF(V600) and BRAF(nonV600) Mutations. *Ann. Oncol.* 31 (2), 289–294. doi:10.1016/j.annonc.2019.10.022

McArthur, G. A., Chapman, P. B., Robert, C., Larkin, J., Haanen, J. B., Dummer, R., et al. (2014). Safety and Efficacy of Vemurafenib in BRAF(V600E) and BRAF(V600K) Mutation-Positive Melanoma (BRIM-3): Extended Follow-Up of a Phase 3, Randomised, Open-Label Study. *Lancet Oncol.* 15 (3), 323–332. doi:10.1016/S1470-2045(14)70012-9

Menzer, C., Menzies, A. M., Carlino, M. S., Reijers, I., Groen, E. J., Eigentler, T., et al. (2019). Targeted Therapy in Advanced Melanoma with Rare BRAF Mutations. *J. Clin. Oncol.* 37 (33), 3142–3151. doi:10.1200/JCO.19.00489

Mertens, F., Johansson, B., Fioretos, T., and Mitelman, F. (2015). The Emerging Complexity of Gene Fusions in Cancer. *Nat. Rev. Cancer* 15 (6), 371–381. doi:10.1038/nrc3947

Müller, D., and Győrffy, B. (2022). DNA Methylation-Based Diagnostic, Prognostic, and Predictive Biomarkers in Colorectal Cancer. *Biochimica Biophysica Acta (BBA) - Rev. Cancer* 1877 (3), 188722. doi:10.1016/j.bbcan.2022.188722

Noreen, F., Küng, T., Tornillo, L., Parker, H., Silva, M., Weis, S., et al. (2019). DNA Methylation Instability by BRAF-Mediated TET Silencing and Lifestyle-Exposure Divides Colon Cancer Pathways. *Clin. Epigenet* 11 (1), 196. doi:10.1186/s13148-019-0791-1

Palanisamy, N., Ateeq, B., Kalyana-Sundaram, S., Pflueger, D., Ramnarayanan, K., Shankar, S., et al. (2010). Rearrangements of the RAF Kinase Pathway in Prostate Cancer, Gastric Cancer and Melanoma. *Nat. Med.* 16 (7), 793–798. doi:10.1038/nm.2166

Park, J., Jung, H. A., Shim, J. H., Park, W.-Y., Kim, T. H., Lee, S.-H., et al. (2021). Multimodal Treatments and Outcomes for Anaplastic Thyroid Cancer before and after Tyrosine Kinase Inhibitor Therapy: a Real-World Experience. *Eur. J. Endocrinol.* 184 (6), 837–845. doi:10.1530/EJE-20-1482

Planchard, D., Smit, E. F., Groen, H. J. M., Mazieres, J., Besse, B., Helland, Å., et al. (2017). Dabrafenib Plus Trametinib in Patients with Previously Untreated BRAF(V600E)-mutant Metastatic Non-small-cell Lung Cancer: an Open-Label, Phase 2 Trial. *Lancet Oncol.* 18 (10), 1307–1316. doi:10.1016/S1470-2045(17)30679-4

Prahallad, A., Sun, C., Huang, S., Di Nicolantonio, F., Salazar, R., Zecchin, D., et al. (2012). Unresponsiveness of Colon Cancer to BRAF(V600E) Inhibition through Feedback Activation of EGFR. *Nature* 483 (7387), 100–103. doi:10.1038/nature10868

Ritterhouse, L. L., and Barletta, J. A. (2015). BRAF V600E Mutation-specific Antibody: A Review. *Seminars Diagnostic Pathology* 32 (5), 400–408. doi:10.1053/j.semdp.2015.02.010

Robert, C., Grob, J. J., Stroyakovskiy, D., Karaszewska, B., Hauschild, A., Levchenko, E., et al. (2019). Five-Year Outcomes with Dabrafenib Plus Trametinib in Metastatic Melanoma. *N. Engl. J. Med.* 381 (7), 626–636. doi:10.1056/NEJMoa1904059

Samatar, A. A., and Poulikakos, P. I. (2014). Targeting RAS-ERK Signalling in Cancer: Promises and Challenges. *Nat. Rev. Drug Discov.* 13 (12), 928–942. doi:10.1038/nrd4281

Shi, H., Hugo, W., Kong, X., Hong, A., Koya, R. C., Moriceau, G., et al. (2014). Acquired Resistance and Clonal Evolution in Melanoma during BRAF Inhibitor Therapy. *Cancer Discov.* 4 (1), 80–93. doi:10.1158/2159-8290.CD-13-0642

Song, L.-B., Zhang, Q.-J., Hou, X.-Y., Xiu, Y.-Y., Chen, L., Song, N.-H., et al. (2020). A Twelve-Gene Signature for Survival Prediction in Malignant Melanoma Patients. *Ann. Transl. Med.* 8 (6), 312. doi:10.21037/atm.2020.02.132

Subbiah, V., Kreitman, R. J., Wainberg, Z. A., Cho, J. Y., Schellens, J. H. M., Soria, J. C., et al. (2018). Dabrafenib and Trametinib Treatment in Patients with Locally Advanced or Metastatic BRAF V600-Mutant Anaplastic Thyroid Cancer. *J. Clin. Oncol.* 36 (1), 7–13. doi:10.1200/JCO.2017.73.6785

Tang, Z., Kang, B., Li, C., Chen, T., and Zhang, Z. (2019). GEPIA2: an Enhanced Web Server for Large-Scale Expression Profiling and Interactive Analysis. *Nucleic Acids Res.* 47 (W1), W556–W560. doi:10.1093/nar/gkz430

Wang, Q., Zhao, N., and Zhang, J. (2021). Gene Mutation Analysis in Papillary Thyroid Carcinoma Using a Multi-Gene Panel in China. *Int. J. Gen. Med.* 14, 5139–5148. doi:10.2147/IJGM.S327409

Wang, T.-X., Tan, W.-L., Huang, J.-C., Cui, Z.-F., Liang, R.-D., Li, Q.-C., et al. (2020). Identification of Aberrantly Methylated Differentially Expressed Genes Targeted by Differentially Expressed miRNA in Osteosarcoma. *Ann. Transl. Med.* 8 (6), 373. doi:10.21037/atm.2020.02.74

Weisenberger, D. J., Siegmund, K. D., Campan, M., Young, J., Long, T. I., Faasse, M. A., et al. (2006). CpG Island Methylator Phenotype Underlies Sporadic Microsatellite Instability and Is Tightly Associated with BRAF Mutation in Colorectal Cancer. *Nat. Genet.* 38 (7), 787–793. doi:10.1038/ng1834

Xu, M., Zhou, J., Zhang, Q., Le, K., Xi, Z., Yi, P., et al. (2020). MiR-3121-3p Promotes Tumor Invasion and Metastasis by Suppressing Rap1GAP in Papillary Thyroid Cancer *In Vitro*. *Ann. Transl. Med.* 8 (19), 1229. doi:10.21037/atm-20-4469

Yao, Z., Torres, N. M., Tao, A., Gao, Y., Luo, L., Li, Q., et al. (2015). BRAF Mutants Evade ERK-dependent Feedback by Different Mechanisms that Determine Their Sensitivity to Pharmacologic Inhibition. *Cancer Cell* 28 (3), 370–383. doi:10.1016/j.ccell.2015.08.001

Zhang, R., Li, Y., Yu, H., Liu, L., Zhu, C., Zuo, S., et al. (2020). An Aberrant DNA Methylation Signature for Predicting Hepatocellular Carcinoma. *Ann. Transl. Med.* 8 (24), 1667. doi:10.21037/atm-20-7804

Zhao, R., Gao, S., He, H., Zhang, J., Zhang, G., and Wen, X. (2021). Evaluation on the Distribution of EGFR, KRAS and BRAF Genes and the Expression of PD-L1 in Different Types of Lung Cancer. *Int. J. Gen. Med.* 14, 5615–5620. doi:10.2147/IJGM.S316151

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 Bioengineering and Biotechnology

TYPE Original Research
PUBLISHED 18 July 2022
DOI 10.3389/fbioe.2022.841034

OPEN ACCESS

EDITED BY
Ping Zhang,
Griffith University, Australia

REVIEWED BY
Pin-Yen Chen,
Monash University, Australia
Yujun Shi,
Sichuan University, China
Fei Wu Long,
Sichuan University, China

*CORRESPONDENCE
Yao Chen,
chenyao@cqmu.edu.cn
Lei Zhao,
zhaolei@cqmu.edu.cn
Yue Xiang,
xiangy0906@126.com

†These authors have contributed equally to this work
‡These authors share senior authorship

SPECIALTY SECTION
This article was submitted to Preclinical Cell and Gene Therapy, a section of the journal Frontiers in Bioengineering and Biotechnology

RECEIVED 21 December 2021
ACCEPTED 27 June 2022
PUBLISHED 18 July 2022

CITATION
Ding Q, Hou Z, Zhao Z, Chen Y, Zhao L and Xiang Y (2022), Identification of the prognostic signature based on genomic instability-related alternative splicing in colorectal cancer and its regulatory network.
Front. Bioeng. Biotechnol. 10:841034.
doi: 10.3389/fbioe.2022.841034

Identification of the prognostic signature based on genomic instability-related alternative splicing in colorectal cancer and its regulatory network

Qiuying Ding[1†], Zhengping Hou[1†], Zhibo Zhao[2], Yao Chen[1*‡], Lei Zhao[1*‡] and Yue Xiang[1*‡]

[1]Centre for Lipid Research, Key Laboratory of Molecular Biology for Infectious Diseases, Ministry of Education, Department of Infectious Diseases, Institute for Viral Hepatitis, The Second Affiliated Hospital, Chongqing Medical University, Chongqing, China, [2]The Department of Hepatobiliary Surgery of the Second Affiliated Hospital, Chongqing Medical University, Chongqing, China

Background: Colorectal cancer (CRC) is a heterogeneous disease with many somatic mutations defining its genomic instability. Alternative Splicing (AS) events, are essential for maintaining genomic instability. However, the role of genomic instability-related AS events in CRC has not been investigated.

Methods: From The Cancer Genome Atlas (TCGA) program, we obtained the splicing profiles, the single nucleotide polymorphism, transcriptomics, and clinical information of CRC. Combining somatic mutation and AS events data, a genomic instability-related AS signature was constructed for CRC. Mutations analyses, clinical stratification analyses, and multivariate Cox regression analyses evaluated this signature in training set. Subsequently, we validated the sensitivity and specificity of this prognostic signature using a test set and the entire TCGA dataset. We constructed a nomogram for the prognosis prediction of CRC patients. Differentially infiltrating immune cells were screened by using CIBERSORT. Inmmunophenoscore (IPS) analysis was used to evaluate the response of immunotherapy. The AS events-related splicing factors (SF) were analyzed by Pearson's correlation. The effects of SF regulating the prognostic AS events in proliferation and migration were validated in Caco2 cells.

Results: A prognostic signature consisting of seven AS events (PDHA1-88633-ES, KIAA1522-1632-AP, TATDN1-85088-ES, PRMT1-51042-ES, VEZT-23786-ES, AIG1-77972-AT, and PHF11-25891-AP) was constructed. Patients in the high-risk score group showed a higher somatic mutation. The genomic instability risk score was an independent variable associated with overall survival (OS), with a hazard ratio of a risk score of 1.537. The area under the curve of receiver operator characteristic curve of the genomic instability risk score in predicting the OS of CRC patients was 0.733. Furthermore, a nomogram was established and could be used clinically to stratify patients to predict prognosis. Patients defined as high-risk by this signature showed a lower proportion of eosinophils than the low-risk group. Patients with low risk were more sensitive to anti-CTLA4 immunotherapy. Additionally, HSPA1A and FAM50B were two SF regulating the OS-related AS. Downregulation of HSPA1A and FAM50B inhibited the proliferation and migration of Caco2 cells.

Conclusion: We constructed an ideal prognostic signature reflecting the genomic instability and OS of CRC patients. HSPA1A and FAM50B were verified as two important SF regulating the OS-related AS.

KEYWORDS

colorectal cancer, genomic instability, alternative splicing, splicing factor, overall survival

Introduction

Colorectal cancer (CRC) is the second most common cancer diagnosed in women and the third most common in men (Dekker et al., 2019). It is the second leading cause of cancer deaths (Keum and Giovannucci, 2019). In 2020, 1,148,515 new cases of CRC were diagnosed, and about 576,858 individuals died from malignancy (Sung et al., 2021). CRC is a multifactorial disease characterized by molecular and clinical heterogeneity (Nguyen et al., 2020). Therefore, it is urgently necessary to explore novel biomarkers for improving the clinical outcome of CRC patients.

Alternative splicing (AS) is one of the essential post-transcriptional regulatory mechanisms and contributes to enriching the protein diversity from a limited number of genes (Baralle and Giudice, 2017; Sciarrillo et al., 2020). Increasing evidence has suggested that aberrant alternative splicing (AS) events regulate cell proliferation, invasion, apoptosis, angiogenesis, and drug resistance, resulting in the progression of CRC (Chen et al., 2021). Alternatively spliced CD44 variants have been identified to promote intestinal tumorigenesis induced by the activation of Wnt signaling (Guo and Frenette, 2014). In chemoradiation-resistant colon cancer cells, exon skipping is significantly increased (Xiong et al., 2016). Alternative splicing isoforms of vascular endothelial growth factor A (VEGFA), UDP glucuronosyltransferase family 1 member A complex locus (UGT1A), pregnane X receptor (PXR), and KRAS are potential therapeutic targets for CRC (Audet-Delage et al., 2017; Canavese et al., 2017; Eilertsen et al., 2019).

Genomic instability, an important prognostic factor of cancer, has been reported to be a hallmark of cancer (Duijf et al., 2019). The instability is multifaceted at several different levels, ranging from simple deoxyribonucleic acid sequence changes to chromosomal aberrations. The molecular mechanisms underlying genomic instability implicate numerous levels of gene regulation, such as transcriptional and post-transcriptional regulation (Chen et al., 2022). Studies indicated that 92–94% of human genes undergo AS (Wang et al., 2008). In addition, multiple AS events have been identified to be associated with genomic instability (Liu et al., 2020; Sebastian et al., 2020a). The dysregulated AS disturbs genome integrity resulting in tumorigenesis (Öther-Gee Pohl and Myant, 2022). However, whether the AS events could reflect the genomic instability and overall survival (OS) of CRC is currently unknown.

In this study, we compared the differential AS events between genomic stable and unstable patients. We developed a new prognostic model combining AS profiles and somatic mutation profiles in CRC tumor tissues. In addition, we explored the related splicing factors and infiltrating immune cells in this prognostic model. Our studies identified the potential molecular signature as genomic instability-associated CRC biomarkers, which may be helpful to assess the clinical outcomes of CRC patients accurately.

Materials and methods

Data collection

Clinical information, RNA-seq data, and somatic mutation data of CRC cohorts were obtained from The Cancer Genome Atlas (TCGA) database (https://tcga-data.nci.nih.gov/). AS events from CRC patient samples ($n = 442$) were collected from TCGAspliceSeq database (http://bioinformatics.mdanderson.org). The SpliceSeq tool was used to analyze AS profiles and assess the splicing patterns of mRNA in CRC. The Percent Splicing index (PSI), ranging from zero to one, was used to quantify AS events. AS events with PSI values >75% were selected. The AS events were visualized by using the R package: UpSetR (v1.4.0).

Screening of the genomic instability-related alternative splicing events

A computational framework was performed to analyze genomic instability-related AS events by combining the PSI values of the AS events and somatic mutation profiles (Figure 1). The somatic mutation quantity of each patient in TCGA database were calculated with "varscan". We then ranked each patient's somatic mutations number in descending order, then defined the top 25% ($n = 98$) and the last 25% (n = 97) of the patients as genomic unstable (GU) group and genomic stable (GS) group, respectively. To filter differentially expressed AS events which were defined as genomic instability-related AS events, the significance analysis of microarrays (SAM) method was used to compare the PSI values between these two groups ($p < 0.05$).

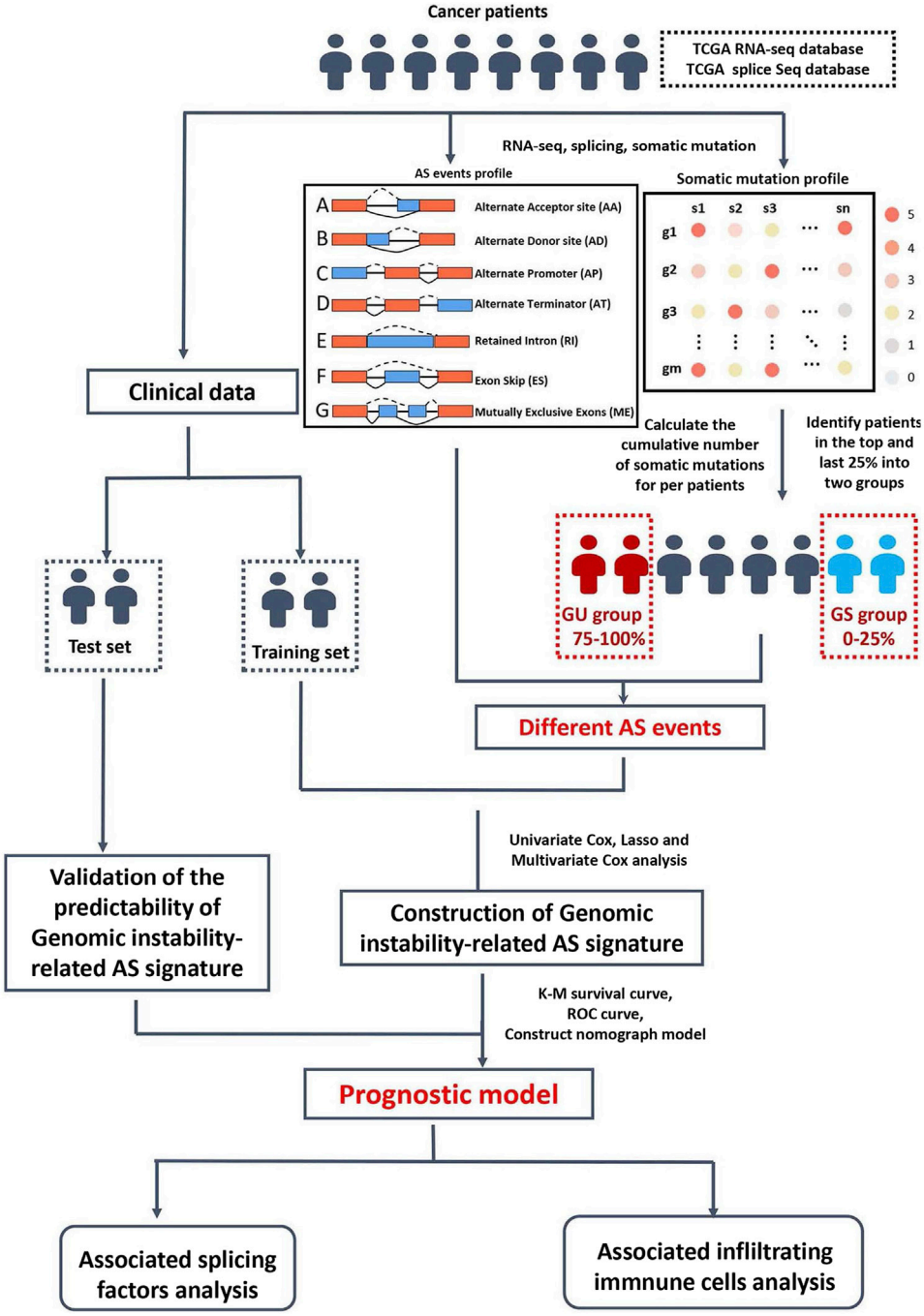


FIGURE 1
Study flowchart.

Identification of survival-associated alternative splicing events and the prognostic signature construction

Univariate Cox analysis was performed to determine the survival-associated genomic instability-related AS events. To remove high correlations among the seven AS events, we used the LASSO (Least absolute shrinkage and selection operator) regression. Multivariate Cox regression analyses were performed to estimate the independent predictor function of each AS event. Finally, we calculate the risk score using the following formula: Risk score = $\sum_{i}^{n} PSIi \times \beta i$, where β denotes the regression

TABLE 1 Clinicopathological information of the patients with CRC in TCGA.

Covariates	Type	Total (n = 433)	Test set (n = 216)	Training set (n = 217)	p-value
Age (%)	≤65	178 (41.11%)	88 (40.74%)	90 (41.47%)	0.9541
	>65	255 (58.89%)	128 (59.26%)	127 (58.53%)	
Sex (%)	Female	204 (47.11%)	101 (46.76%)	103 (47.47%)	0.9594
	Male	229 (52.89%)	115 (53.24%)	114 (52.53%)	
Tumor stage (%)	Stage I-II	244 (56.35%)	123 (56.94%)	121 (55.76%)	0.8317
	Stage III-IV	178 (41.11%)	87 (40.28%)	91 (41.94%)	
	Unknow	11 (2.54%)	6 (2.78%)	5 (2.3%)	
T stage (%)	T1-2	86 (19.86%)	44 (20.37%)	42 (19.35%)	0.8662
	T3-4	346 (79.91%)	171 (79.17%)	175 (80.65%)	
	Unknow	1 (0.23%)	1 (0.46%)	0 (0%)	
M stage (%)	M0	318 (73.44%)	163 (75.46%)	155 (71.43%)	0.6089
	M1	60 (13.86%)	28 (12.96%)	32 (14.75%)	
	Unknow	55 (12.7%)	25 (11.57%)	30 (13.82%)	
N stage (%)	N0	259 (59.82%)	129 (59.72%)	130 (59.91%)	1
	N1-2	174 (40.18%)	87 (40.28%)	87 (40.09%)	

coefficient of each event. Then the median of risk score was defined as a cut-off value distinguishing the high-and low-risk groups. The LASSO regression was performed with "glmnet" and "survival" R package. The multivariate Cox regression analyses were performed with "survival" R package.

Validation of the prognostic signature

To provide additional support to our findings, the 433 CRC patients were randomly divided into a training set (n = 217) and a test set (n = 216) using the R package "caret" (Table1 summarizes the clinicopathological characteristics of the CRC patients). First, each set of CRC patients was divided into high-risk and low-risk groups using the same formula to compute the risk score. Kaplan–Meier survival curve and Log-Rank test were used to compare overall survival between the high-risk and low-risk groups in the training set. The survival receiver operating characteristic (ROC) package (R 4.0.3) was then used to assess the ability of the prediction model. Univariate and Multivariate Cox regression were conducted to calculate the high-risk score's hazard ratio (HR). Subsequently, we validated the prognostic signature in the test set and the entire TCGA dataset (n = 433). Kaplan–Meier, Cox, and ROC analyses were carried out as described above.

Immune cell analysis

We used the CIBERSORT algorithm (http://cibersort.stanford.edu/) to estimate the 22 kinds of infiltrating immune cells in CRC tissue. A total of 163 cases were included for further analysis with p-values <0.05. The median risk scores classified these cases into high-risk (n = 81) and low-risk (n = 82) groups. The R package "vioplot" was used to draw the differential immune cells types between these two groups. The survival curve was generated using the R package "survival". The P- value is calculated based on the log-rank. Immunophenoscore (IPS) data were obtained from The Cancer Immunome Atlas (TCIA) database (https://tcia.at/), predicting the response to cytotoxic T lymphocyte antigen 4 (CTLA4) and programmed cell death protein 1 (PD-1) blockers (n = 433). According to the median risk scores classified these cases into high-risk (n = 216) and low-risk (n = 217) groups. The differential effective immunotherapy responses between high-risk and low-risk groups were analyzed by the chi-square test and visualized by R package "vioplot".

Correlation network of survival associated alternative splicing events and splicing factors construction

A total of 404 splicing factor genes were identified in a previous study(Seiler et al., 2018). The mRNA profile data of the splicing factors (SF) in CRC were obtained from the TCGA database. Correlations between the splicing factor expression and prognosis-related AS events were visualized and analyzed by Cytoscape 3.7.2. In Univariate Cox regression, a p-value of <0.05 and correlation coefficient >0.1 were identified as statistically significant.

TABLE 2 Primers used in this study.

Name	Primer (5′–3′)
HSPA1A	Forward: CATCATCAGCGGACTGTACCA
	Reverse: TGCAAACACAGGAAATTGAGAAC
FAM50B	Forward: AAGAGGTTCTCGGCGCATTAC
	Reverse: CGGGCCTTCATGTCGTTCA
36B4	Forward: CAGCAAGTGGGAAGGTGTAATCC
	Reverse: CCCATTCTATCATCAACGGGTACAA

Cell culture and transfection

Caco2 cells were purchased from ATCC (Manassas, VA, United States) and cultured with a DMEM-High glucose medium (HyClone, Logan, UT, United States) supplied with 10% fetal bovine serum (FBS, Lonsera, UY). and 50 U/ml Penicillin-G, 50 μg Streptomycin (Thermo Scientific, Cambridge, MA, United States). Cell cultures were maintained at 37°C in a humidified atmosphere of 5% CO2.

The small interfering RNA of HSPA1A and FAM50B and negative control were purchased from TsingkeBiotechnology (Beijing, CHN). For the siRNA experiments, cells were seeded in 6-well plates at a density of 5 x 10^4 cells/well and medium was replaced with serum-free medium once confluence reached ~80%. Subsequently, 160 pmol siRNA (80 pmol of each siRNA when two siRNAs were co-transfected) was mixed with 200 μl serum-free medium containing 8 μl Lipofectamine RNAiMAX Reagent (Invitrogen, United States) and added to the cells. In the following experiments, the cells were divided into four groups: control siRNA (NCi) group, HSPA1A siRNA (HSPA1Ai) group, FAM50B siRNA (FAM50Bi) group, HSPA1A siRNA + FAM50B siRNA (HSPA1Ai + FAM50Bi) group. The oligonucleotide sequences were as follows: siRNA-HSPA1A sense strand, 5′- GCCAUGACG AAAGACAACATT-3′ and antisense strand, 5′-UGUUGU CUUUCGUCAUGGCTT -3’; siRNA-FAM50B sense strand, 5′-GCUGGUACGAGAACAATT -3′ and antisense strand, 5′-UUGUUCUUCUCGUACCAGCTT -3’. The primer sequence used were listed in Table2.

RNA isolation and real-time qPCR

Total RNA of siRNA infected cells was extracted using RNAiso Plus (Takara), according to the manufacture’s protocols. Subsequently, the isolated RNA was reverse transcribed into cDNA with PrimeScriptTM Reagent Kit (Takara). The reaction mixture for qPCR containing SYBR (BAOGUAGN, China) was prepared according to the manufacture’s protocols. RT-PCR was performed in a PCR system with HSPA1A, FAM50B, and 36B4 primers. Relative gene expression was calculated using the $2^{-\triangle\triangle CT}$ method, using 36B4 mRNA expression as reference gene. Each sample was analyzed at least three times.

Cell proliferation experiments

Caco2 cells were seeded in 96-well plates (Corning, NY, United States) (3*10^3 cells/well) and transfected with siRNA for 24 h. Then these cells were cultured for 4 days, and the cell proliferation was measured using a CCK8 reagent (Beyotime, Beijing, CHN).

Wound healing assay

The Caco2 cells were seeded in 6-well plates at a density of 5 × 10^4 cells/well. After 24 h of siRNA transfection, wounds were made in center of the well using a sterile 10ul pipette tip. Images of five randomly-selected scratched fields were captured on an inverted light microscope (ZEISS, Oberkochen, BE, GER) at 0 and 24 h. Magnification, X200. Wound areas were measured by Image J, and calculated the wound healing percentage.

Transwell assay

Cells were added to the upper chambers of the Transwells (Cornning, NY, United States) at a density of 5 × 10^5 cells/well and transfected with siRNA. After 24 h, medium in the lower chamber of a transwell was exchanged for complete medium and medium in the upper chamber of a transwell was exchange for serum-free medium. After the cells were cultured in transwell for 24 h, cells on the lower layer of the membrane were fixed by 4% paraformaldehyde (Sangon, Shanghai, CHN) for 10 min, then stained using Crystal Violet Staining Solution (Beyotime) for 30min. The cells number was counted by using an inverted microscope (ZEISS) and five fields were randomly selected to count the cells. Magnification, ×200.

Statistical analysis

Statistical analyses were performed using the statistical package GraphPad Prism, version 8.0 (California, United States). All results are expressed as mean ± SEM. Student’s t-tests were used to compare results between two groups, and One-way ANOVA was used to compare differences among multiple groups. Results were considered statistically significant at *, #, $p < 0.05$.

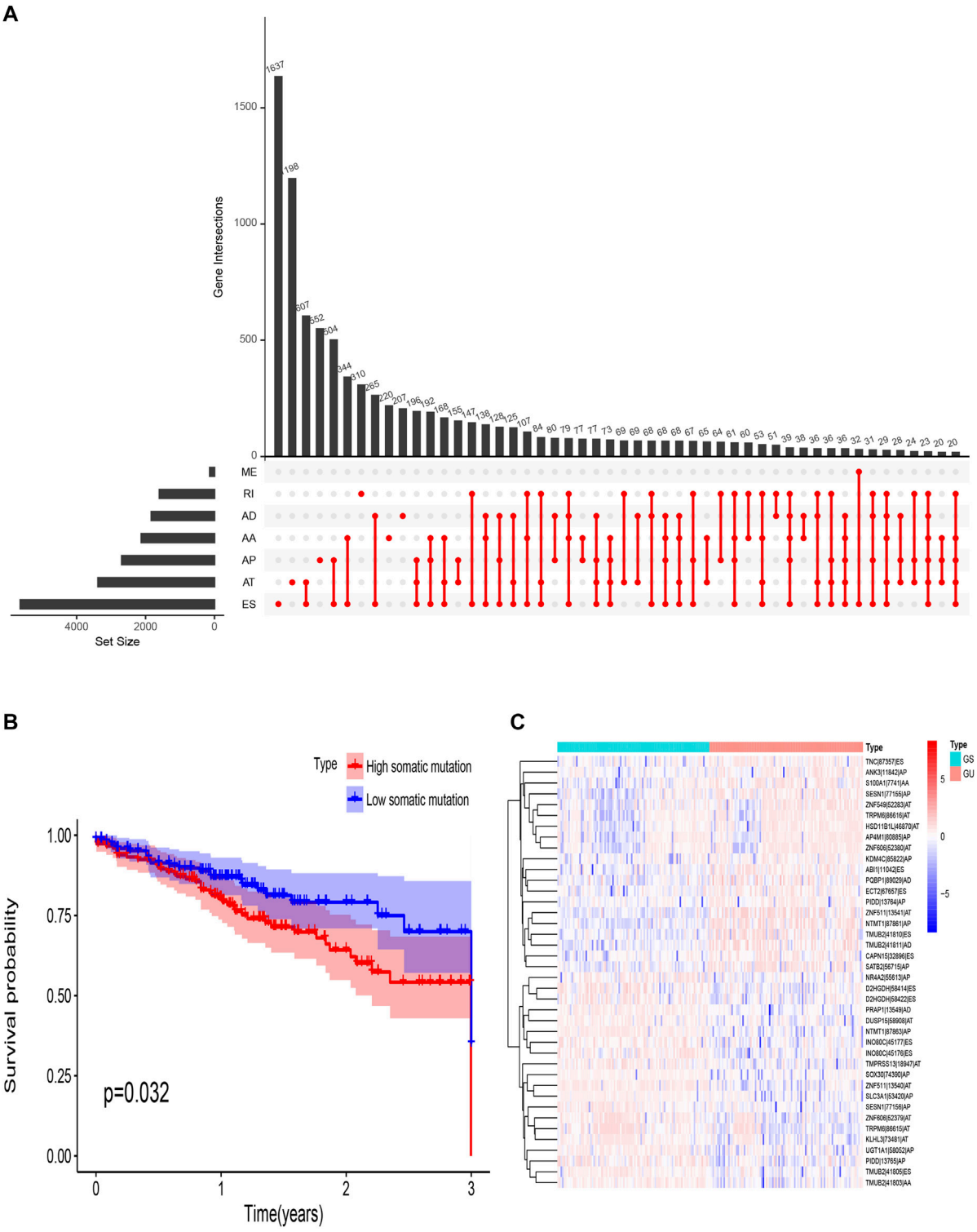


FIGURE 2
Overview of AS events in TCGA CRC cohort. **(A)** Upset plot for all AS events. AS, alternative splicing; RI, retained intron; ME, mutually exclusive exons; ES, exon skipping; AT, alternative terminator; AP, alternative promoter; AD, alternative donor site; AA, alternative acceptor site. **(B)** Survival probability of different somatic mutation group. **(C)** Heat map of genomic instability-related AS events. TCGA, The Cancer Genome Atlas; CRC, Colorectal cancer.

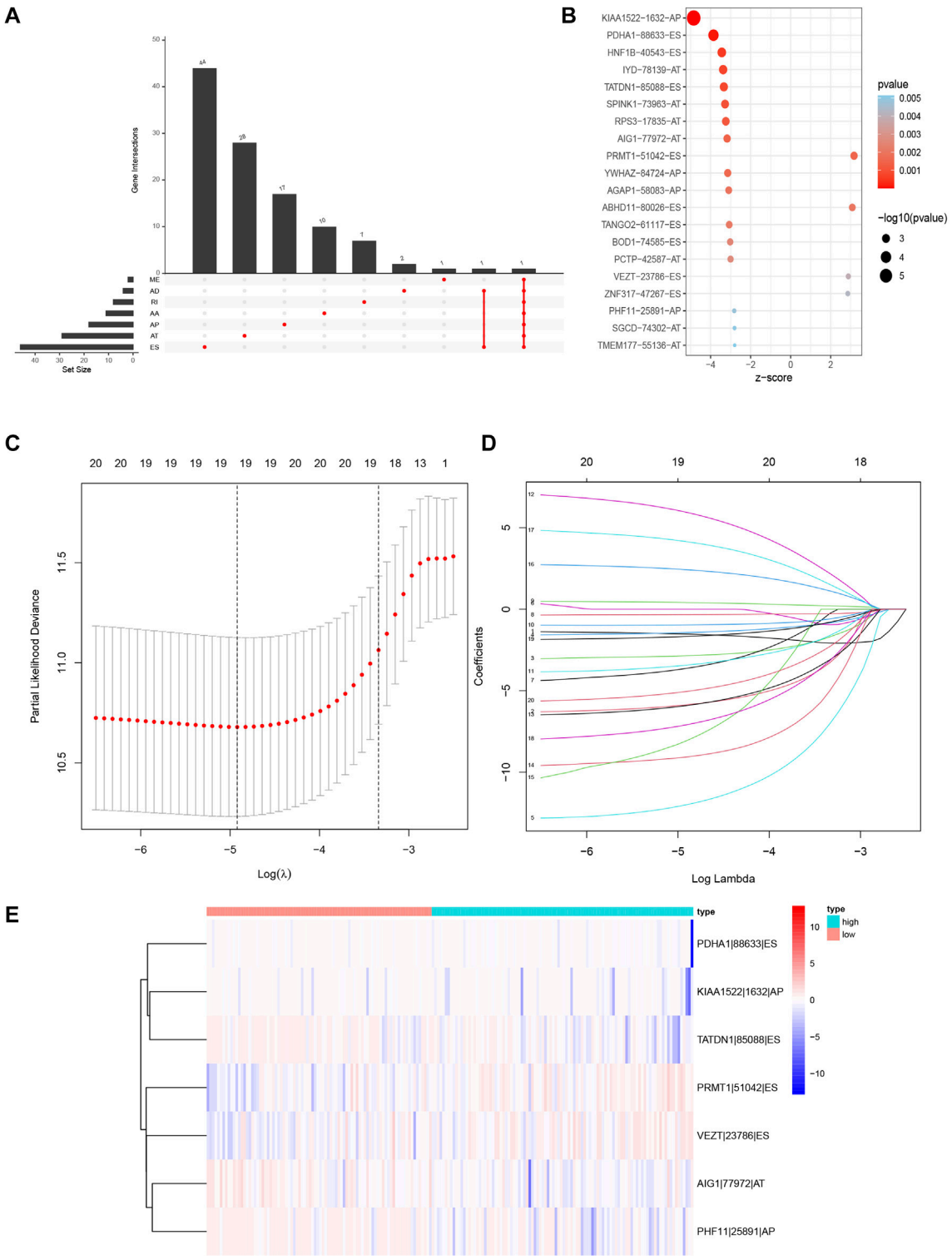


FIGURE 3
Prognosis-related AS events in this study. **(A)** Upset plot for survival-related genes. **(B)** The Bubble plots of survival-associated AS events in CRC. **(C,D)** Optimal survival-related AS events selection in the LASSO regression model. **(E)** Heat map of the seven optimal survival-related AS events. LASSO, least absolute shrinkage and selection operator.

TABLE 3 Details of the selected AS events based on multivariate Cox analysis.

AS event	Coefficient	HR	95% CI	*p*-value
KIAA1522-1632-AP	−2.957	0.052	0.003–1.048	0.054
PDHA1-88633-ES	−6.779	0.001	1.77E-05–0.073	0.001
TATDN1-85088-ES	−36.891	9.52E-17	2.58E-27–3.51E-06	0.003
AIG1-77972-AT	−4.116	0.016	0.0004–0.670	0.030
PRMT1-51042-ES	3.553	34.913	3.040–400.990	0.004
VEZT-23786-ES	5.388	218.699	9.810–4875.360	0.001
PHF11-25891-AP	−9.906	4.99E-05	3.24E-08–0.077	0.008

Results

Screening of genomic instability-related alternative splicing events in colorectal cancer patients

There are seven types of AS events: Mutually Exclusive Exons (ME), Retained Intron (RI), Alternate Donor site (AD), Alternate Acceptor site (AA), Alternate Terminator (AT), Alternate Promoter (AP), and Exon Skip (ES). In this study, 442 CRC patients were included. Seventy-four ME events in 74 genes, 1655 RI in 1184 genes, 1691 AD events in 1316 genes, 2006 AA events in 1576 genes, 3973 AP events in 2330 genes, 4710 AT events in 2783 genes, and 8416 ES events in 4436 genes. In TCGA-CRC, ES events were the most frequent type of spliced signatures, followed by AT and AP events, and ME was the least frequent (Figure 2A). Next, we divided TCGA CRC patients into high-somatic mutation and low- somatic mutation groups according to the median of the somatic mutation counts of each patient. As shown in Figure 2B, the three-year OS was significantly lower in patients with higher level of somatic mutation ($p = 0.032$), demonstrating a key role of genomic instability in the OS of CRC patients. Furthermore, we defined patients with the top 25% ($n = 98$) and the last 25% ($n = 97$) of somatic mutation counts as genomic unstable (GU) group and genomic stable (GS) group, respectively. The mean somatic mutation numbers of GS and GU groups were 88 and 2058, respectively. A heat map of the top 40 different AS events is showed in Figure 2C.

Construct a genomic instability-related alternative splicing signature for overall survival in the training set

We screened the AS events related to the OS of CRC patients. The results showed that 114 AS events were significantly associated with OS in CRC patients (Figures 3A,B). Then Lasso Cox regression analysis was used to further select the AS events related to the OS and prognosis of CRC patients (Figures 3C,D). Then, the risk score was calculated for each AS event (Table3). Seven AS events: PDHA1-88633-ES, KIAA1522-1632-AP, TATDN1-85088-ES, PRMT1-51042-ES, VEZT-23786-ES, AIG1-77972-AT, and PHF11-25891-AP were identified as independent risk factors for OS in CRC using multivariate Cox regression. Moreover, two AS events (PRMT1-51042-ES, VEZT-23786-ES) had positive coefficients suggesting that high expression of these two AS events were associated with poorer prognosis as risk factors. In contrast, the remaining AS events (PDHA1-88633-ES, KIAA1522-1632-AP, TATDN1-85088-ES, AIG1-77972-AT, and PHF11-25891-AP) had negative coefficients suggesting that their upregulated expression were associated with better survival as protective factors. Next, a genomic instability-related AS events prognostic signature was established by quantifying the PSI of the seven AS events and their coefficients from the multivariate Cox regression analysis. Risk Score = (−6.779 × PSI of PDHA1-88633-ES) + (−2.957 × PSI of KIAA1522-1632-AP) + (−36.891 × PSI of TATDN1-85088-ES) + (3.553 × PSI of PRMT1-51042-ES) + (5.388 × PSI of VEZT-23786-ES) + (−4.116 × PSI of AIG1-77972-AT) + (−9.906 × PSI of PHF11-25891-AP). The heatmap of the seven AS events is shown in Figure 3E.

The alternative splicing signature was associated with genomic instability in colorectal cancer patients

To verify whether the AS signature was associated with the somatic mutation pattern, we compared the expression of UBQLN4, a biomarker for driving genomic instability (Jachimowicz et al., 2019), between the two different risk groups in the three sets. As shown in Figures 4A–C, the expression of UBQLN4 and somatic mutation count was higher in the high-risk groups than in the low-risk group, in which the *p* values were 0.02 in the training set, 0.063 in the test set, and 0.0024 in the entire TCGA set. These results

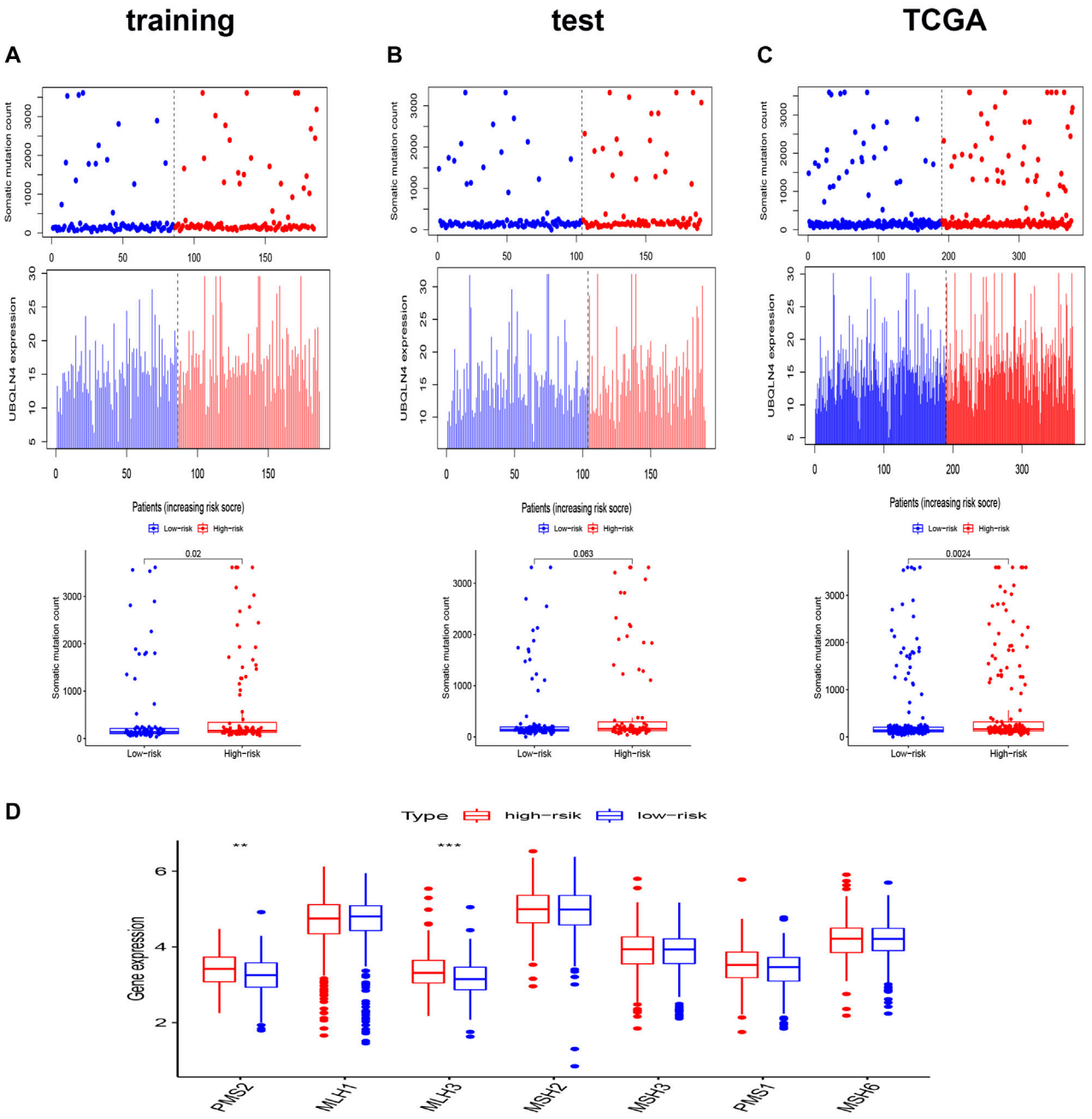


FIGURE 4
Relationship between the genomic instability-related AS signature and somatic mutation patterns of CRC patients. **(A–C)** The distribution of somatic mutation count and UBQLN4 expression of the training set **(A)**, the test set **(B)**, and the entire TCGA set **(C)**. **(D)** Boxplots comparing the seven DNA mismatch repair genes expression between high- and low-risk groups. *$p < 0.05$ high-risk group vs. low-risk group. Statistical analysis was performed using the Mann-Whitney U test.

implied that the AS signature score was associated with genomic instability. Furthermore, DNA somatic mismatch repair (MMR) genes associated with genomic instability (Baross-Francis et al., 1998). We then analyzed the genomic alterations of seven DNA mismatch repair genes (MLH1, MLH3, MSH2, MSH3, MSH6, PMS1, PMS2) in the two different risk score groups. As shown in Figure 4D, the gene expression of PMS2 and MLH3 were significantly higher in the high-risk group. These results reconfirm that our prognostic signature was associated with genomic instability.

Predictability evaluation of genomic instability-related alternative splicing signature in colorectal cancer patients

We performed Kaplan-Meier analysis and constructed a ROC curve to verify the prognostic efficiency of the AS signature in the training set (Figures 5A,B). The risk score distribution curves showed that higher risk score patients in CRC had a shorter survival time. Kaplan-Meier survival curve analysis verified that patients with higher risk scores had poorer OS, $p < 0.001$. Moreover, the AUC value of the predictive accuracy of the model was 0.773. To further validate the prognostic significance of the genomic instability AS signature, we calculated the genomic instability AS signature scores of the test set and the entire TCGA set and constructed the respective ROC curves. In the test set, the survival of the low-risk group was significantly longer than that of the high-risk group, with an AUC value of 0.710. (Figures 5C,D). Similar results were also obtained in the entire TCGA set, where the AUC of the ROC curves was 0.733 (Figures 5E,F). These results suggested that genomic instability-related AS signature had a good survival prediction efficacy.

The genomic instability-related alternative splicing signature was independent of other clinical factors

To evaluate whether the genomic instability-related AS signature could act as an independent prognostic factor of clinicopathological features, univariate and multivariate Cox

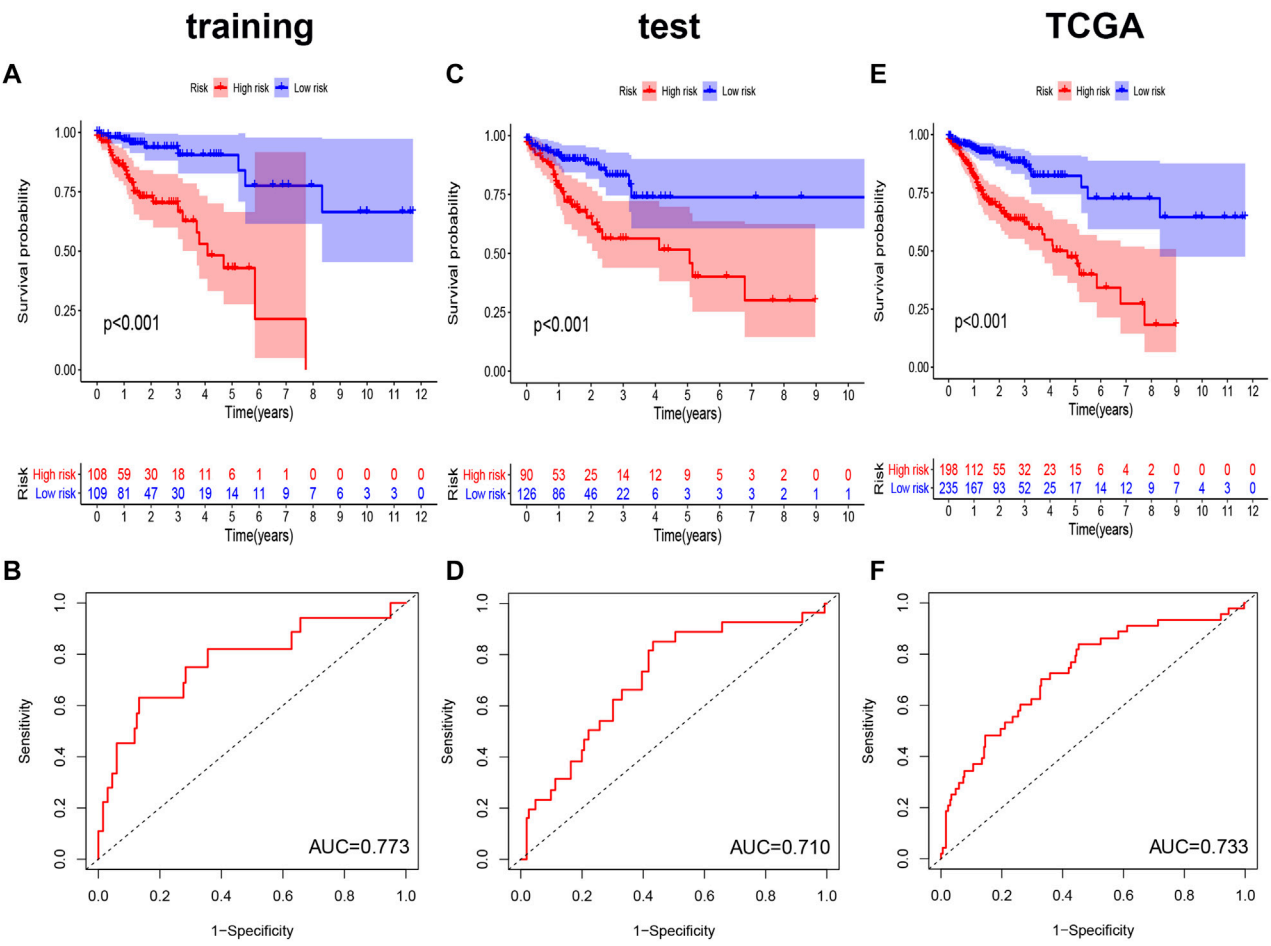


FIGURE 5
Kaplan-Meier curves and ROC curves of the prognostic AS models. **(A,C,E)** Kaplan-Meier plots of the genomic instability-related AS signature in the training set **(A)**, the test set **(C)**, and the entire TCGA set **(E)**. **(B,D,F)** The ROC curves for overall survival of the genomic instability-related AS signature in three datasets, respectively. ROC, Receiver operating characteristic.

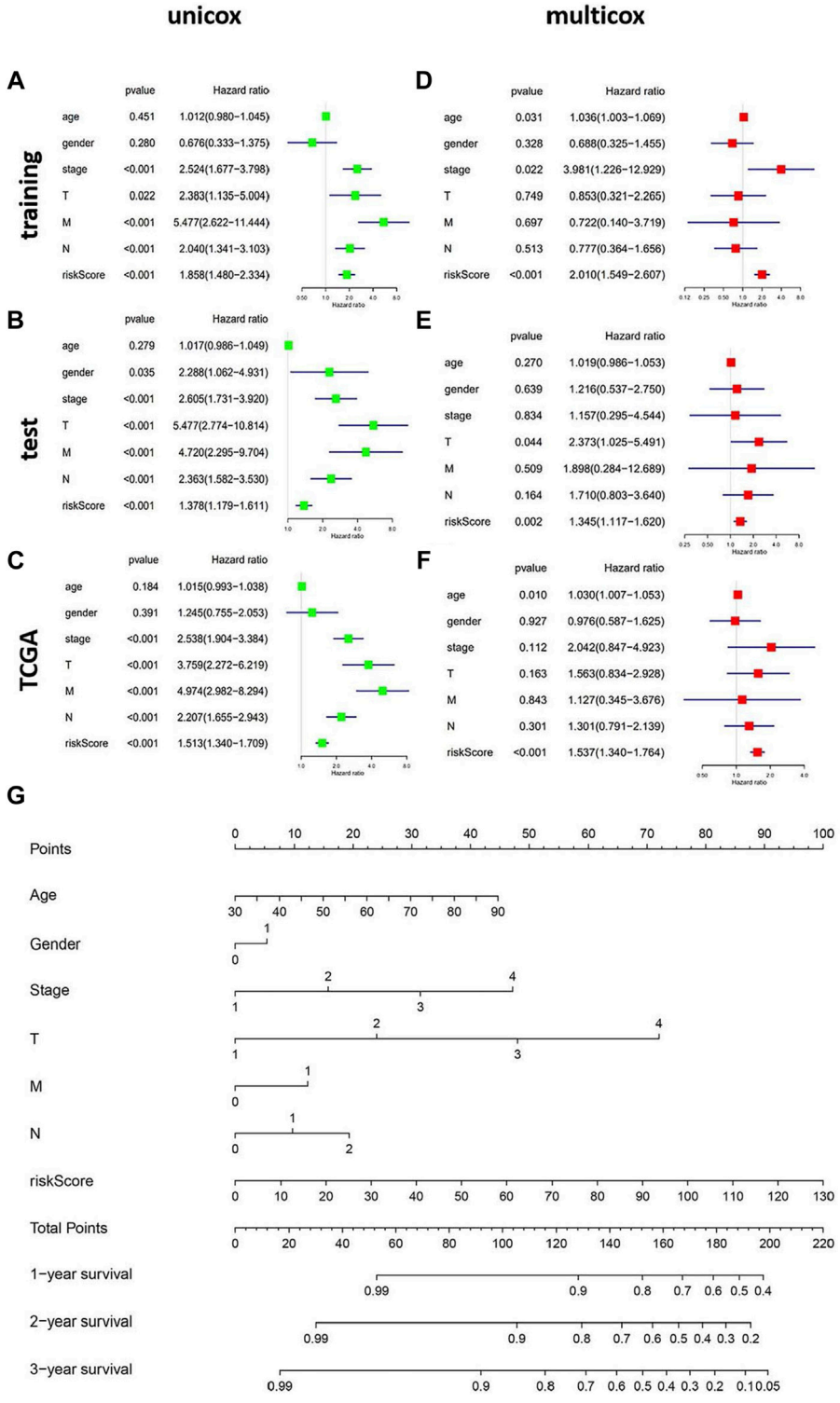


FIGURE 6
The independent prognostic analysis of the genomic instability-related AS signature. Construction of the nomograph model in patients with CRC. (**A–C**) Forest plots of univariate cox regression in the training set, test set, and the entire TCGA set. (**D–F**) Forest plots of multivariate cox regression in the three datasets. (**G**) The nomograph model predicting 1-, 2-, and 3-year survival in patients with CRC based on age, sex, TMN stage, and risk score.

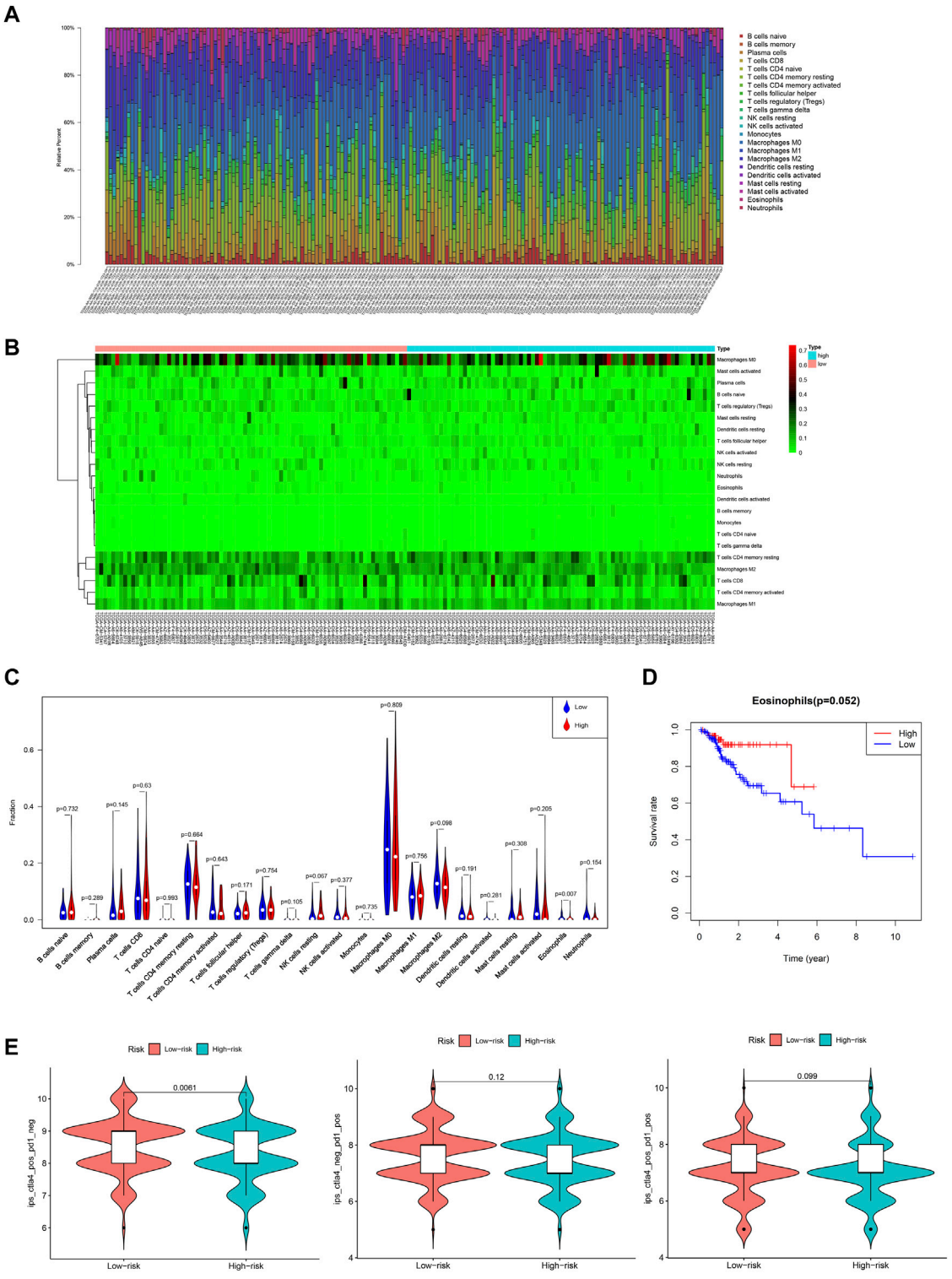


FIGURE 7
Overview of the infiltrating immune cells in CRC. **(A)** Bar plot showing the proportion of the 22 types of immune cells. **(B)** Heat map of the immune cells proportion between the high- and low-risk groups. **(C)** Comparison of each immune cell type in the two risk groups. **(D)** Kaplan-Meier estimates of overall survival of patients with low or high eosinophils expression. **(E)** Violin plots of the IPS in two risk groups. IPS, immunophenoscore.

10.3389/fbioe.2022.841034

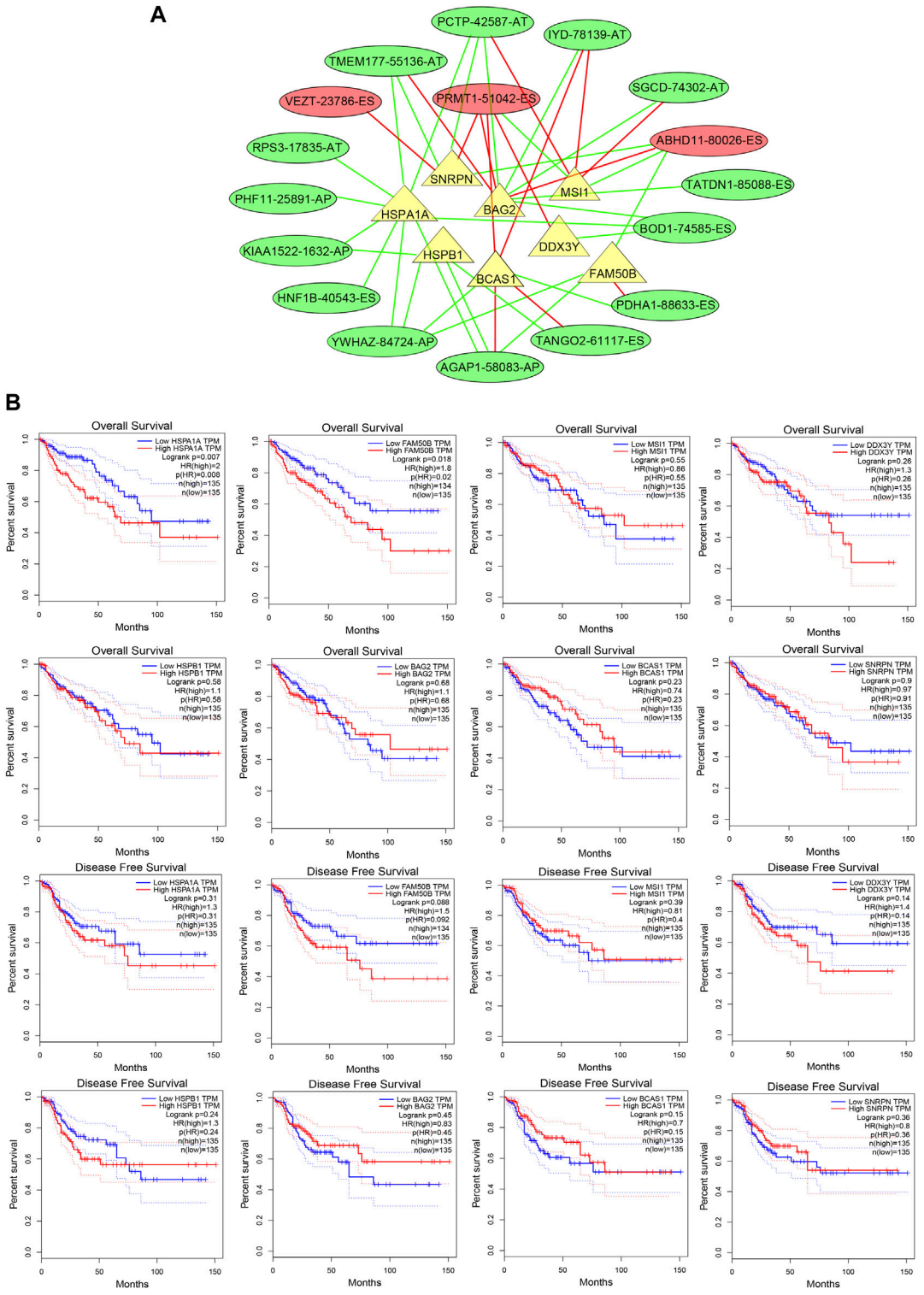


FIGURE 8
The splicing factors are associated with prognostic AS signature. **(A)** Splicing correlation network in CRC. The triangles represent the survival-related SF. The red and green ovals represent SREs that increase and decrease risk, respectively. Red and green lines represent the positive and negative correlations of connected triangles, respectively. SRE, alternative splicing events; SF, splicing factor. **(B)** ROC analysis of overall survival and disease-free survival for the AS signature-related splicing factors in patients with CRC. SF, splicing factor.

regression analyses were performed, adjusting for age, sex, and pathologic stage in the three data sets (training, test, and TCGA data set) (Figures 6A–C). The univariate Cox regression results showed that with the OS, genomic instability AS signature and tumor stage were significantly correlated in these three data sets ($p < 0.001$). The prognostic significance of each data set was also retained in multivariate Cox regression analyses. For other clinical features, only sex had a significant correlation with the genomic instability-related AS signature in the test set using univariate Cox regression analyses ($p = 0.035$) (Figures 6D–F). Moreover, to make this model more practicable in the clinic, we constructed a nomograph model based on the risk score, tumor stage, age, and sex to predict the 1-, 2-, and 3- year survival of patients with CRC (Figure 6G).

Tumor-infiltrating immune cells were associated with the prognostic alternative splicing signature

To investigate the relationship between tumor-infiltrating immune cells and the prognostic AS signature, we used CIBERSORT to identify tumor-infiltrating immune cells in 163 CRC patients. Detailed tumor-infiltrating immune cell information on each patient was illustrated in Figure 7A. Furthermore, according to the risk score, we divided these patients into high-risk and low-risk groups (Figure 7B). As shown in Figure 7C, compared with the low-risk group, the high-risk group had a lower proportion of eosinophils ($p = 0.007$). Additionally, we explored the relationship between the proportion of eosinophils and clinicopathological information of CRC patients. Next, we used a ROC curve to verify the prognostic efficiency of eosinophils expression, and the results showed that higher eosinophils expression had higher OS rates ($p = 0.052$) (Figure 7D). We also explored the association between immunotherapy efficiency between the high-risk and low-risk groups. As shown in Figure 7E, patients with low risk were more sensitive to targeting CTLA4 treatment. For PD-1 alone or in combination with CTLA4 treatment, there were no significant differences between the two risk groups. Therefore, these data implied that our prognostic AS signature might facilitate immunotherapy results prediction.

Exploring the regulatory network of regulating the prognostic alternative splicing signature

Due to the unavailability of inhibiting AS specifically, we explored the regulatory network for regulating the prognostic AS signature and tried to find a target regulating OS-related AS. We constructed a splicing-regulatory network to further determine whether the prognostic AS events were regulated by specific splicing factors in CRC. As shown in Figure 8A, eight splicing factors (SNRPN, HSPA1A, HSPB1, BAG2, BCAS1, DDX3Y, MSI1, and FAM50B) were significantly correlated with survival-associated AS events, and more than half of the survival-related AS events were regulated by more than one splicing factor. Furthermore, we assessed the function of these splicing factors in the prognosis of CRC patients (Figure 8B). The result showed that patients with lower HSPA1A expression levels and FAM50B expression levels had longer OS rates; the p-values were 0.008 and 0.02, respectively. Although the p-value approached insignificance, lower FAM50B expression also showed longer disease-free survival rates ($p = 0.092$).

Validation the effect of HSPA1A and FAM50B in Caco2 cells proliferation and migration

We continued to validate the effect of SF regulating the prognostic AS events in the proliferation and migration of Caco2 cells. The mRNA levels of HSPA1A and FAM50B were decreased after specific siRNA transfection in Caco2 cells (Figure 9A). As detected by the CCK8 proliferation assay, gnomically inhibiting HSPA1A or FAM50B decreased the proliferation of Caco2 significantly (Figure 9B). Wound healing and transwell assays were used to examine the effect of HSPA1A and FAM50B on the migration of Caco2 cells. It was found that both in the HSPA1Ai group, and FAM50Bi group, the Caco2 cells exhibited decreased proliferation and migration (Figures 9C,D). The results suggest that downregulation of HSPA1A and FAM50B inhibited Caco2 proliferation and migration.

Discussion

CRC remains a leading cause of cancer-related death in the world (Siegel et al., 2020). The diagnosis and treatment of CRC have been improved dramatically, but the mortality rate continues to be high, especially in advanced patients (He et al., 2021). With sequencing technology development, new genomic markers have been proposed to guide CRC patients' "personalized" treatment. For example, a multitude of genomic instability-associated events, including SNPs, circular RNAs, long non-coding RNA, and microRNAs, have been reported as predictors of clinical outcomes in CRC patients (Weigl et al., 2018; Ghafouri-Fard et al., 2021; Long et al., 2021).

Recently, the critical roles of AS in maintaining genomic instability have been revealed. Metastasis-associated antigen 1, an oncogenic chromatin modifier, may affect chromosomal instability by regulating related RNA splicing (Liu et al., 2020). MarcoH2A1, a replication stress-protective histone, and its alternative splicing was associated with X chromosome

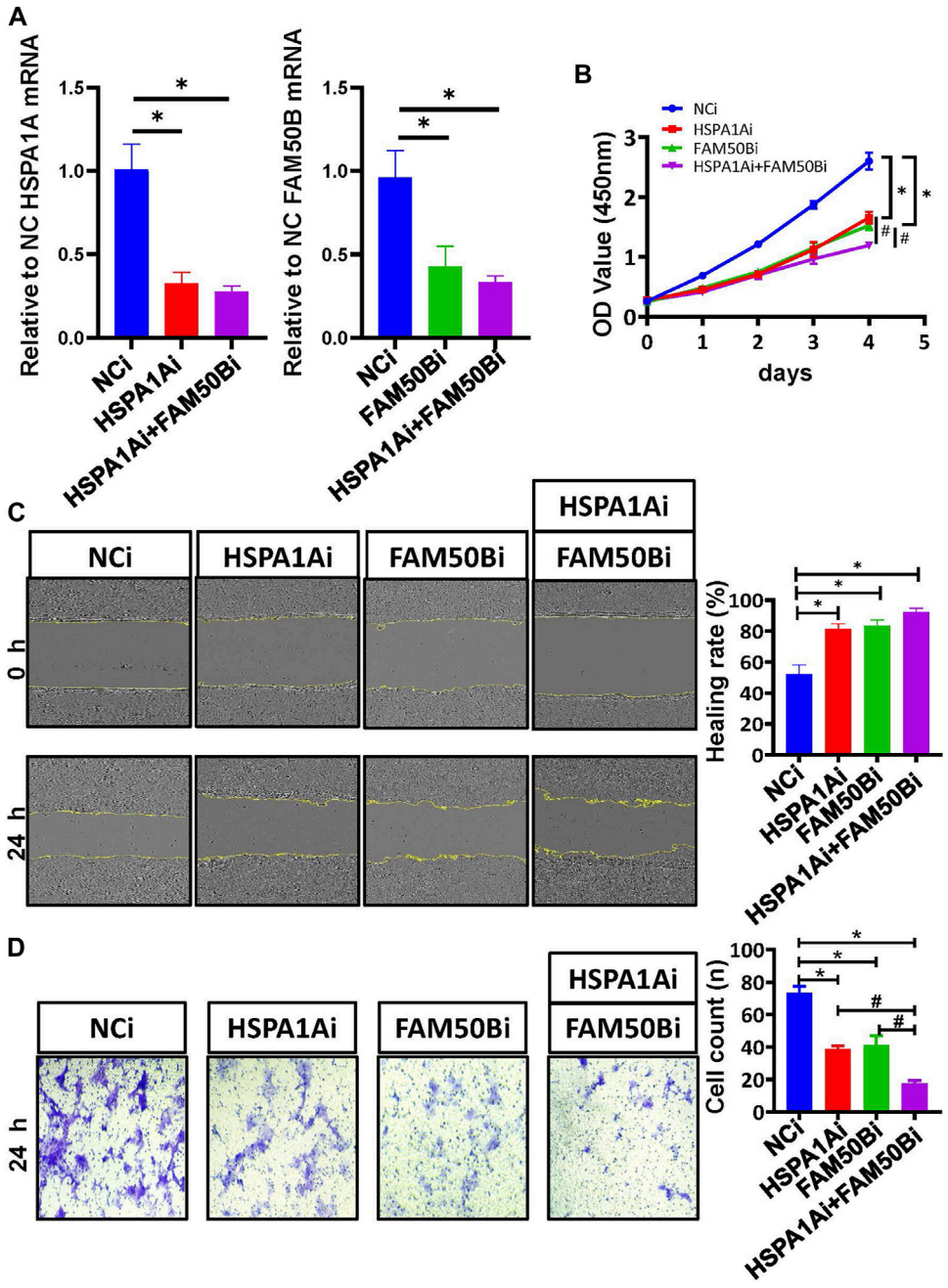


FIGURE 9
Downregulation of HSPA1A and FAM50B inhibits the proliferation and migration of Caco2 cells. Caco2 cells were transfected with control siRNA, HSPA1A siRNA, and FAM50B siRNA, respectively, or co-transfected with HSPA1A siRNA and FAM50B siRNA. **(A)** The mRNA levels of HSPA1A and FAM50B in Caco2 cells after transfection. **(B)** Cell proliferation were assessed by CCK8 assays (n ≥ 4). **(C and D)** The cell migration was detected by scratch **(C)** and transwell assays **(D)** (n = 3). * $p < 0.05$ vs. NCi group. #$p < 0.05$ vs. HSPA1Ai + FAM50Bi. Data are mean ± SEM.

genomic stability (Sebastian et al., 2020b). Despite reports of recent advances in the identification of aberrant splicing events in CRC (Liu J. et al., 2018; Lian et al., 2020), there have been no relevant studies on AS signatures relative to genomic instability in CRC. Herein, we identified a group of genomic instability-related AS events in CRC and revealed their significance in predicting patient survival.

Following this line, we derived a prognostic signature based on the differential AS events. We combined single nucleotide polymorphism profiles with somatic mutation profiles of CRC

and identified seven genomic instability-related AS events (PDHA1-88633-ES, KIAA1522-1632-AP, TATDN1-85088-ES, PRMT1-51042-ES, VEZT-23786-ES, AIG1-77972-AT, and PHF11-25891-AP) to construct our prognostic signature in the training set. The high-risk groups had higher UBQLN4 expression and somatic mutation count in the training set, which indicated that our prognostic signature was available to estimate genomic instability. Furthermore, two MMR genes, PMS2 and MLH3, were also higher in the high-risk group, which revalidated that our prognostic signature was associated with genomic instability. For the PMS2, Kasela, Mariann et al. found that lower expression with higher repair efficiencies (Kasela et al., 2019) indicated that the low-risk group might have better repair efficiencies. Narayanan, Sumana et al. had also found that low expression of PMS2 and MLH3 had significantly improved 5-year OS in CRC patients (Narayanan et al., 2019), which indicated that our low-risk group might have a better OS.

We proposed an ideal prognostic model and verified this model using both the test and the entire TCGA sets. This model showed a great performance in the risk stratification of CRC patients and a good potential in predicting the prognosis of CRC (AUC of the ROC curve was >0.7). Subsequently, we evaluated whether this genomic instability-related AS signature could be an independent prognostic factor. The multivariate Cox regression analysis revealed that patients with a higher risk score had poorer outcome, which was also validated in the test and the entire TCGA dataset. Finally, we established a nomogram combining the genomic instability-related AS signature with the tumor stage to enhance the convenience and accuracy of the prediction model.

Among the seven genomic instability-related AS events, PRMT1-51042-ES and VEZT-23786-ES were two positively coefficient AS events. Protein arginine methyltransferase 1 (PRMT1), the founding member of the PRMT family, has been reported to be associated with histone methylarginine and transcription activation (Yang and Bedford, 2013). Consistent with our results, previous studies have shown that a PRMT1 spice isoform could serve as a biomarker of poor prognosis in CRC (Mathioudaki et al., 2008; Yao et al., 2021). Vezatin, adherens junctions transmembrane protein (VEZT) has been identified as a tumor suppressor gene in gastric cancer (Li et al., 2015). Nevertheless, the function of VEZT variants in CRC is still unclear.

Conversely, PDHA1-88633-ES, KIAA1522-1632-AP, TATDN1-85088-ES, AIG1-77972-AT, and PHF11-25891-AP were AS events inversely correlated with the OS. Downregulation of PDHA1, a gatekeeper enzyme-linked between glycolysis and the mitochondrial citric acid cycle, has been associated with poor survival in gastric cancer and esophageal squamous cancer (Liu Z. et al., 2018; Liu et al., 2019). The KIAA1522 gene was discovered via a sequencing project of human cDNA encoding large proteins (Nagase et al., 2000). Recently, studies have indicated that KIAA1522 may act as an oncogene for non-small cell lung cell cancer (NSCLC) and breast cancer (Liu et al., 2016; Li et al., 2018). As a highly conserved nuclease, TatD DNase domain containing 1 (TATDN1), a member of the TATD family, has been found upregulated in hepatocellular carcinoma (HCC) and cisplatin-resistant NSCLC (Shen et al., 2019; Wang et al., 2019). Androgen-induced gene 1 (AIG1) is a transmembrane protein that regulates cytosolic calcium concentrations (Nickel et al., 2016). Previous studies have determined AIG1 may serve as a new biomarker for the diagnosis and prognostic evaluation of HCC and is associated with the thiopurine treatment of acute lymphoblastic leukemia (Choi et al., 2019). The deletion and methylation of PHD finger protein 11 (PHF11) were associated with chronic lymphocytic leukemia and Ewing sarcoma, respectively (Parker et al., 2011; Alholle et al., 2013). While the roles of these inversely correlated AS events in CRC remain largely unknown and require further research.

Tumor-infiltrating immune cells play essential roles in cancer development and progression. Ye et al. found that CD66b+ tumor-associated neutrophils, Tregs, and CD163+tumor-associated macrophages were significantly correlated with prognosis in CRC patients (Ye et al., 2019). In this study, we identified the differential infiltrating immune cells in CRC patients with the prognostic AS signature and found that the high-risk group showed a lower proportion of eosinophils than the low-risk group. In addition, patients with higher eosinophils had higher OS rates. Consistent with our findings, several studies have reported that eosinophil accumulation was associated with better survival in CRC patients (Pretlow et al., 1983; Harbaum et al., 2015; Prizment et al., 2016). A study of 381 primary CRC patients by Harbaum et al. found that the number of peritumoral eosinophils significantly impacted on the prognosis of CRC patients by assessing peritumoral eosinophils and intratumoral eosinophil counts (Harbaum et al., 2015). Similarly, another study involving 441 CRC patients in the United States observed that the tumor-stromal eosinophil count was an important favorable prognostic factor in CRC (Prizment et al., 2016). The anti-tumorigenic mechanisms of eosinophils in CRC include direct and indirect effects. The direct killing is achieved via degranulation and release of eosinophil-specific proteins, such as major basic protein, eosinophil cationic protein, eosinophil-derived neurotoxin, and granzymes (Legrand et al., 2010; Varricchi et al., 2018). Instead, indirect killing refers to a combination of cytokine-mediated effects, including interleukin (IL)-2, IL-5, IL-4, IL-8, and IL-17E (Benatar et al., 2010; Gatault et al., 2015). Whether our genomic instability-related AS signature could affect the anti-tumor function of eosinophil in CRC also requires further validation *in vitro* and *in vivo*. Recently, tumor immunotherapy has become a new paradigm. Inhibition of CTLA or PD-1 monoclonal antibodies is the most promising treatment approach for many cancers, including the microsatellite instability (MSI) -high advanced CRC (Messersmith, 2019). We then explored the predictive value of the prognostic AS signature in immunotherapy. The results showed that the IPS-CLTA4 score was significantly increased in the low-risk. Indeed, AS may play a 'double-edged sword' role in immunotherapy (Öther-Gee Pohl and Myant, 2022). Some AS variants can produce

neoantigens to increase $CD8^+$ T-cell immunogenicity (Smart et al., 2018). On the other hand, an alternatively spliced variant of CD19 could consult the resistance of CAR-T treatment (Sotillo et al., 2015). In this study, our prognostic AS signature had the potential ability to predict the efficacy of IPS-CLTA4 in CRC patients. Although we have no direct evidence to show that these seven screened AS events could affect the immunotherapy sensitivity, two AS events related genes, PRMT1 and PHF11, had revealed that they were associated with immune cells. PRMT1 is highly expressed in T helper cells, and the inhibition of PRMT1 could attenuate the suppressive functions of regulatory T cells (Kagoya et al., 2019). PHF11 plays an essential role in producing IgE by activated B cells (Ikari et al., 2014). These suggested that maybe AS events could affect the immunotherapy sensitivity. Further studies are required to elucidate these comprehensively.

The splicing factors (SFs) play a critical role in regulating AS events (El Marabti and Abdel-Wahab, 2021). Previous studies have demonstrated that serine and arginine-rich splicing factor 6 regulates AS to mediate CRC progression (Wan et al., 2019), and SET domain containing 2, histone lysine methyltransferase (SETD2) modulated AS events during intestinal tumorigenesis (Yuan et al., 2017). Thus, we further explored the relationship between survival-related AS events and the expression of splicing factors in CRC. Eight SFs, including SNRPN, HSPA1A, HSPB1, BAG2, BCAS1, DDX3Y, MSI1, and FAM50B, were associated with survival-related AS events. Among these SFs, HSPA1A and FAM50B were associated with OS in CRC. We found that patients with lower HSPA1A expression levels had higher OS rates, which was consistent with previous studies (Xing et al., 2021). As a member of the heat shock proteins (HSPs) family, heat shock protein family A (Hsp70) member 1A (HSPA1A) exerted cytoprotective and immunological functions (Wang et al., 2021). Recently, Huang et al. (2021) found that HSPA1A could regulate two types of AS events (SNX5-58744-AT and SNX5-587745-AT), which were correlated with distant metastasis, through the "Class I MHC mediated antigen processing and presentation" pathway in mesothelioma. Our study also found that patients with lower FAM50B expression levels also had higher OS. Loss of FAM50B (also known as Family with sequence similarity 50, member B) expression has also been identified in almost 4% of cancers in the TCGA database. Silencing FAM50B can reduce cellular fitness and cause apoptosis and dysregulation of transcription (Thompson et al., 2021). It has been validated that as a splicing factor, FAM50B serves an independent prognostic factor in glioblastoma (Qiu et al., 2021). In order to assess the impact of HSPA1A and FAM50B in the progression of CRC, we performed functional analysis on Caco2 cells. The results showed that both individual and simultaneous HSPA1A and FAM50B knockdown showed proliferation and migration inhibition function.

Though our study provides significant insights to explore the relationship between genome instability and CRC patients' prognosis, some limitations should also be considered. First, as the TCGA database is the only available database providing the alternative splicing events data, an external examination is unpracticable. In the future, we hope an external examination could be conducted. Second, additional studies will be necessary to unravel the biological roles of these AS events *in vivo* and *in vitro*.

In conclusion, we developed and validated a risk prognostic signature comprising seven genomic instability-related AS events, which could serve as an independent prognostic biomarker for the survival of CRC patients and reflect the change in the microenvironment of CRC. HSPA1A and FAM50B play an important role in the proliferation and migration of Caco2 cells. Our data suggest that this genomic instability-related AS signature and its regulatory network may have important implications for developing new therapeutic targets and individualized therapy in patients with CRC.

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

QD and ZH performed the data analysis and wrote the manuscript. ZZ provide the support for the R scripts. YC, LZ, and YX designed the study and revised the manuscript. All authors approved the submitted version.

Funding

This study was funded by the National Natural Science Foundation of China (Grant No. 81970510 and No. 31971084), Chongqing Natural Science Foundation (cstc2021ycjh-bgzxm0146, cstc2019jcyj-msxmX0336), Talent Project of Chongqing (CQYC2019050790) and the Science and Technology Research Program of Chongqing Municipal Education Commission (KJQN201900438).

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.

Acknowledgments

We acknowledge the use of TCGA and TCGAspliceSeq for the simple nucleotide variation, transcriptomics, clinical information, and splicing data.

References

Alholle, A., Brini, A. T., Gharanei, S., Vaiyapuri, S., Arrigoni, E., Dallol, A., et al. (2013). Functional epigenetic approach identifies frequently methylated genes in Ewing sarcoma. *Epigenetics* 8, 1198–1204. doi:10.4161/epi.26266

Audet-Delage, Y., Rouleau, M., Rouleau, M., Roberge, J., Miard, S., Picard, F., et al. (2017). Cross-talk between alternatively spliced UGT1A isoforms and colon cancer cell metabolism. *Mol. Pharmacol.* 91, 167–177. doi:10.1124/mol.116.106161

Baralle, F. E., and Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* 18, 437–451. doi:10.1038/nrm.2017.27

Baross-Francis, A., Andrew, S. E., Penney, J. E., and Jirik, F. R. (1998). Tumors of DNA mismatch repair-deficient hosts exhibit dramatic increases in genomic instability. *Proc. Natl. Acad. Sci. U. S. A.* 95, 8739–8743. doi:10.1073/pnas.95.15.8739

Benatar, T., Cao, M. Y., Lee, Y., Lightfoot, J., Feng, N., Gu, X., et al. (2010). IL-17E, a proinflammatory cytokine, has antitumor efficacy against several tumor types *in vivo*. *Cancer Immunol. Immunother.* 59, 805–817. doi:10.1007/s00262-009-0802-8

Canavese, M., Ngo, D. T. M., Maddern, G. J., Hardingham, J. E., Price, T. J., Hauben, E., et al. (2017). Biology and therapeutic implications of VEGF-A splice isoforms and single-nucleotide polymorphisms in colorectal cancer. *Int. J. Cancer* 140, 2183–2191. doi:10.1002/ijc.30567

Chen, Y., Huang, M., Liu, X., Huang, Y., Liu, C., Zhu, J., et al. (2021). Alternative splicing of mRNA in colorectal cancer: new strategies for tumor diagnosis and treatment. *Cell Death Dis.* 12, 752. doi:10.1038/s41419-021-04031-w

Chen, M., Linstra, R., and van Vugt, M. A. T. M. (2022). Genomic instability, inflammatory signaling and response to cancer immunotherapy. *Biochim. Biophys. Acta - Rev. Cancer* 1877, 188661. doi:10.1016/j.bbcan.2021.188661

Choi, R., Sohn, I., Kim, M.-J., Woo, H. I., Lee, J. W., Ma, Y., et al. (2019). Pathway genes and metabolites in thiopurine therapy in Korean children with acute lymphoblastic leukaemia. *Br. J. Clin. Pharmacol.* 85, 1585–1597. doi:10.1111/bcp.13943

Dekker, E., Tanis, P. J., Vleugels, J. L. A., Kasi, P. M., and Wallace, M. B. (2019). Colorectal cancer. *Lancet* 394, 1467–1480. doi:10.1016/S0140-6736(19)32319-0

Duijf, P. H. G., Nanayakkara, D., Nones, K., Srihari, S., Kalimutho, M., Khanna, K. K., et al. (2019). Mechanisms of genomic instability in breast cancer. *Trends Mol. Med.* 25, 595–611. doi:10.1016/j.molmed.2019.04.004

Eilertsen, I. A., Sveen, A., Strømme, J. M., Skotheim, R. I., Nesbakken, A., Lothe, R. A., et al. (2019). Alternative splicing expands the prognostic impact of KRAS in microsatellite stable primary colorectal cancer. *Int. J. Cancer* 144, 841–847. doi:10.1002/ijc.31809

El Marabti, E., and Abdel-Wahab, O. (2021). Therapeutic modulation of RNA splicing in malignant and non-malignant disease. *Trends Mol. Med.* 27, 643–659. doi:10.1016/j.molmed.2021.04.005

Gatault, S., Delbeke, M., Driss, V., Sarazin, A., Dendooven, A., Kahn, J.-E., et al. (2015). IL-18 is involved in eosinophil-mediated tumoricidal activity against a colon carcinoma cell line by upregulating LFA-1 and ICAM-1. *J. Immunol. (Baltimore, Md, 1950)* 195, 2483–2492. doi:10.4049/jimmunol.1402914

Ghafouri-Fard, S., Taheri, M., Hussen, B. M., Vafaeimanesh, J., Abak, A., Vafaee, R., et al. (2021). Function of circular RNAs in the pathogenesis of colorectal cancer. *Biomed. Pharmacother.* 140, 111721. doi:10.1016/j.biopha.2021.111721

Guo, W., and Frenette, P. S. (2014). Alternative CD44 splicing in intestinal stem cells and tumorigenesis. *Oncogene* 33, 537–538. doi:10.1038/onc.2013.260

Harbaum, L., Pollheimer, M. J., Kornprat, P., Lindtner, R. A., Bokemeyer, C., Langner, C., et al. (2015). Peritumoral eosinophils predict recurrence in colorectal cancer. *Mod. Pathol.* 28, 403–413. doi:10.1038/modpathol.2014.104

He, J., Wu, F., Han, Z., Hu, M., Lin, W., Li, Y., et al. (2021). Biomarkers (mRNAs and non-coding RNAs) for the diagnosis and prognosis of colorectal cancer - from the body fluid to tissue level. *Front. Oncol.* 11, 632834. doi:10.3389/fonc.2021.632834

Huang, R., Zheng, Z., Liu, S., Yan, P., Song, D., Yin, H., et al. (2021). Identification of prognostic and bone metastasis-related alternative splicing signatures in mesothelioma. *Cancer Med.* 10, 4478–4492. doi:10.1002/cam4.3977

Ikari, J., Inamine, A., Yamamoto, T., Watanabe-Takano, H., Yoshida, N., Fujimura, L., et al. (2014). Plant homeodomain finger protein 11 promotes class switch recombination to IgE in murine activated B cells. *Allergy* 69, 223–230. doi:10.1111/all.12328

Jachimowicz, R. D., Beleggia, F., Isensee, J., Velpula, B. B., Goergens, J., Bustos, M. A., et al. (2019). UBQLN4 represses homologous recombination and is overexpressed in aggressive tumors. *Cell* 176, 505–519. e522. doi:10.1016/j.cell.2018.11.024

Kagoya, Y., Saijo, H., Matsunaga, Y., Guo, T., Saso, K., Anczurowski, M., et al. (2019). Arginine methylation of FOXP3 is crucial for the suppressive function of regulatory T cells. *J. Autoimmun.* 97, 10–21. doi:10.1016/j.jaut.2018.09.011

Kasela, M., Nyström, M., and Kansikas, M. (2019). PMS2 expression decrease causes severe problems in mismatch repair. *Hum. Mutat.* 40, 904–907. doi:10.1002/humu.23756

Keum, N., and Giovannucci, E. (2019). Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat. Rev. Gastroenterol. Hepatol.* 16, 713–732. doi:10.1038/s41575-019-0189-8

Legrand, F., Driss, V., Delbeke, M., Loiseau, S., Hermann, E., Dombrowicz, D., et al. (2010). Human eosinophils exert TNF-α and granzyme A-mediated tumoricidal activity toward colon carcinoma cells. *J. Immunol. (Baltimore, Md, 1950)* 185, 7443–7451. doi:10.4049/jimmunol.1000446

Li, Y.-S., Chen, Y.-Z., Guo, X.-B., Liu, X., and Li, L.-P. (2015). VEZT as a novel independent prognostic factor in gastric cancer. *Cancer Biomark.* 15, 375–380. doi:10.3233/CBM-150476

Li, Y., Wang, Y., Fan, H., Zhang, Z., and Li, N. (2018). miR-125b-5p inhibits breast cancer cell proliferation, migration and invasion by targeting KIAA1522. *Biochem. Biophys. Res. Commun.* 504, 277–282. doi:10.1016/j.bbrc.2018.08.172

Lian, H., Wang, A., Shen, Y., Wang, Q., Zhou, Z., Zhang, R., et al. (2020). Identification of novel alternative splicing isoform biomarkers and their association with overall survival in colorectal cancer. *BMC Gastroenterol.* 20, 171. doi:10.1186/s12876-020-01288-x

Liu, Y.-Z., Yang, H., Cao, J., Jiang, Y.-Y., Hao, J.-J., Xu, X., et al. (2016). KIAA1522 is a novel prognostic biomarker in patients with non-small cell lung cancer. *Sci. Rep.* 6, 24786. doi:10.1038/srep24786

Liu, J., Li, H., Shen, S., Sun, L., Yuan, Y., Xing, C., et al. (2018). Alternative splicing events implicated in carcinogenesis and prognosis of colorectal cancer. *J. Cancer* 9, 1754–1764. doi:10.7150/jca.24569

Liu, Z., Yu, M., Fei, B., Fang, X., Ma, T., Wang, D., et al. (2018). miR-21-5p targets PDHA1 to regulate glycolysis and cancer progression in gastric cancer. *Oncol. Rep.* 40, 2955–2963. doi:10.3892/or.2018.6695

Liu, L., Cao, J., Zhao, J., Li, X., Suo, Z., and Li, H. (2019). PDHA1 gene knockout in human esophageal squamous cancer cells resulted in greater warburg effect and aggressive features *in vitro* and *in vivo*. *Onco. Targets. Ther.* 12, 9899–9913. doi:10.2147/OTT.S226851

Liu, J., Li, C., Wang, J., Xu, D., Wang, H., Wang, T., et al. (2020). Chromatin modifier MTA1 regulates mitotic transition and tumorigenesis by orchestrating mitotic mRNA processing. *Nat. Commun.* 11, 4455. doi:10.1038/s41467-020-18259-1

Long, F., Lin, Z., Li, L., Ma, M., Lu, Z., Jing, L., et al. (2021). Comprehensive landscape and future perspectives of circular RNAs in colorectal cancer. *Mol. Cancer* 20, 26. doi:10.1186/s12943-021-01318-6

Mathioudaki, K., Papadokostopoulou, A., Scorilas, A., Xynopoulos, D., Agnanti, N., Talieri, M., et al. (2008). The PRMT1 gene expression pattern in colon cancer. *Br. J. Cancer* 99, 2094–2099. doi:10.1038/sj.bjc.6604807

Messersmith, W. A. (2019). NCCN guidelines updates: Management of metastatic colorectal cancer. *J. Natl. Compr. Canc. Netw.* 17, 599–601. doi:10.6004/jnccn.2019.5014

Nagase, T., Kikuno, R., Ishikawa, K., Hirosawa, M., and Ohara, O. (2000). Prediction of the coding sequences of unidentified human genes. XVII. The

complete sequences of 100 new cDNA clones from brain which code for large proteins *in vitro*. *DNA Res.* 7, 143–150. doi:10.1093/dnares/7.2.143

Narayanan, S., Kawaguchi, T., Peng, X., Qi, Q., Liu, S., Yan, L., et al. (2019). Tumor infiltrating lymphocytes and macrophages improve survival in microsatellite unstable colorectal cancer. *Sci. Rep.* 9, 13455. doi:10.1038/s41598-019-49878-4

Nguyen, L. H., Goel, A., and Chung, D. C. (2020). Pathways of colorectal carcinogenesis. *Gastroenterology* 158, 291–302. doi:10.1053/j.gastro.2019.08.059

Nickel, N., Cleven, A., Enders, V., Lisak, D., Schneider, L., and Methner, A. (2016). Androgen-inducible gene 1 increases the ER Ca(2+) content and cell death susceptibility against oxidative stress. *Gene* 586, 62–68. doi:10.1016/j.gene.2016.03.055

Öther-Gee Pohl, S., and Myant, K. B. (2022). Alternative RNA splicing in tumour heterogeneity, plasticity and therapy. *Dis. Model. Mech.* 15, dmm049233. doi:10.1242/dmm.049233

Parker, H., Rose-Zerilli, M. J. J., Parker, A., Chaplin, T., Wade, R., Gardiner, A., et al. (2011). 13q deletion anatomy and disease progression in patients with chronic lymphocytic leukemia. *Leukemia* 25, 489–497. doi:10.1038/leu.2010.288

Pretlow, T. P., Keith, E. F., Cryar, A. K., Bartolucci, A. A., Pitts, A. M., Pretlow, T. G., et al. (1983). Eosinophil infiltration of human colonic carcinomas as a prognostic indicator. *Cancer Res.* 43, 2997–3000.

Prizment, A. E., Vierkant, R. A., Smyrk, T. C., Tillmans, L. S., Lee, J. J., Sriramarao, P., et al. (2016). Tumor eosinophil infiltration and improved survival of colorectal cancer patients: Iowa women's health study. *Mod. Pathol.* 29, 516–527. doi:10.1038/modpathol.2016.42

Qiu, J., Wang, C., Hu, H., Chen, S., Ding, X., and Cai, Y. (2021). Transcriptome analysis and prognostic model construction based on splicing profiling in glioblastoma. *Oncol. Lett.* 21, 138. doi:10.3892/ol.2020.12399

Sciarrillo, R., Wojtuszkiewicz, A., Assaraf, Y. G., Jansen, G., Kaspers, G. J. L., Giovannetti, E., et al. (2020). The role of alternative splicing in cancer: from oncogenesis to drug resistance. *Drug resist. Updat.* 53, 100728. doi:10.1016/j.drup.2020.100728

Sebastian, R., Hosogane, E. K., Sun, E. G., Tran, A. D., Reinhold, W. C., Burkett, S., et al. (2020). Epigenetic regulation of DNA repair pathway choice by MacroH2A1 splice variants ensures genome stability. *Mol. Cell* 79, 836–845. doi:10.1016/j.molcel.2020.06.028

Sebastian, R., Hosogane, E. K., Sun, E. G., Tran, A. D., Reinhold, W. C., Burkett, S., et al. (2020). Epigenetic regulation of DNA repair pathway choice by MacroH2A1 splice variants ensures genome stability. *Mol. Cell* 79, 836–845. e837. doi:10.1016/j.molcel.2020.06.028

Seiler, M., Peng, S., Agrawal, A. A., Palacino, J., Teng, T., Zhu, P., et al. (2018). Somatic mutational landscape of splicing factor genes and their functional consequences across 33 cancer types. *Cell Rep.* 23, 282–296. e284. doi:10.1016/j.celrep.2018.01.088

Shen, C., Xu, Y., Lu, T. F., Zhang, J. J., Qian, Y. B., Xu, N., et al. (2019). LncRNA TATDN1 induces the progression of hepatocellular carcinoma via targeting miRNA-6089. *Eur. Rev. Med. Pharmacol. Sci.* 23, 6459–6466. doi:10.26355/eurrev_201908_18529

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

Smart, A. C., Margolis, C. A., Pimentel, H., He, M. X., Miao, D., Adeegbe, D., et al. (2018). Intron retention is a source of neoepitopes in cancer. *Nat. Biotechnol.* 36, 1056–1058. doi:10.1038/nbt.4239

Sotillo, E., Barrett, D. M., Black, K. L., Bagashev, A., Oldridge, D., Wu, G., et al. (2015). Convergence of acquired mutations and alternative splicing of CD19 enables resistance to CART-19 immunotherapy. *Cancer Discov.* 5, 1282–1295. doi:10.1158/2159-8290.CD-15-1020

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

Thompson, N. A., Ranzani, M., van der Weyden, L., Iyer, V., Offord, V., Droop, A., et al. (2021). Combinatorial CRISPR screen identifies fitness effects of gene paralogues. *Nat. Commun.* 12, 1302. doi:10.1038/s41467-021-21478-9

Varricchi, G., Galdiero, M. R., Loffredo, S., Lucarini, V., Marone, G., Mattei, F., et al. (2018). Eosinophils: The unsung heroes in cancer? *Oncoimmunology* 7, e1393134. doi:10.1080/2162402X.2017.1393134

Wan, L., Yu, W., Shen, E., Sun, W., Liu, Y., Kong, J., et al. (2019). SRSF6-regulated alternative splicing that promotes tumour progression offers a therapy target for colorectal cancer. *Gut* 68, 118–129. doi:10.1136/gutjnl-2017-314983

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., et al. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476. doi:10.1038/nature07509

Wang, L., Shang, X., and Feng, Q. (2019). LncRNA TATDN1 contributes to the cisplatin resistance of non-small cell lung cancer through TATDN1/miR-451/TRIM66 axis. *Cancer Biol. Ther.* 20, 261–271. doi:10.1080/15384047.2018.1529091

Wang, X., Wang, Y., Fang, Z., Wang, H., Zhang, J., Zhang, L., et al. (2021). Targeting HSPA1A in ARID2-deficient lung adenocarcinoma. *Natl. Sci. Rev.* 8, nwab014. doi:10.1093/nsr/nwab014

Weigl, K., Chang-Claude, J., Knebel, P., Hsu, L., Hoffmeister, M., Brenner, H., et al. (2018). Strongly enhanced colorectal cancer risk stratification by combining family history and genetic risk score. *Clin. Epidemiol.* 10, 143–152. doi:10.2147/CLEP.S145636

Xing, X.-L., Yao, Z.-Y., Xing, C., Huang, Z., Peng, J., Liu, Y.-W., et al. (2021). Gene expression and DNA methylation analyses suggest that two immune related genes are prognostic factors of colorectal cancer. *BMC Med. Genomics* 14, 116. doi:10.1186/s12920-021-00966-3

Xiong, W., Gao, D., Li, Y., Liu, X., Dai, P., Qin, J., et al. (2016). Genome-wide profiling of chemoradiation-induced changes in alternative splicing in colon cancer cells. *Oncol. Rep.* 36, 2142–2150. doi:10.3892/or.2016.5022

Yang, Y., and Bedford, M. T. (2013). Protein arginine methyltransferases and cancer. *Nat. Rev. Cancer* 13, 37–50. doi:10.1038/nrc3409

Yao, B., Gui, T., Zeng, X., Deng, Y., Wang, Z., Wang, Y., et al. (2021). PRMT1-mediated H4R3me2a recruits SMARCA4 to promote colorectal cancer progression by enhancing EGFR signaling. *Genome Med.* 13, 58. doi:10.1186/s13073-021-00871-5

Ye, L., Zhang, T., Kang, Z., Guo, G., Sun, Y., Lin, K., et al. (2019). Tumor-infiltrating immune cells act as a marker for prognosis in colorectal cancer. *Front. Immunol.* 10, 2368. doi:10.3389/fimmu.2019.02368

Yuan, H., Li, N., Fu, D., Ren, J., Hui, J., Peng, J., et al. (2017). Histone methyltransferase SETD2 modulates alternative splicing to inhibit intestinal tumorigenesis. *J. Clin. Invest.* 127, 3375–3391. doi:10.1172/JCI94292

Glossary

AS Alternative splicing

AD Alternate Donor site

AA Alternate Acceptor site

AT Alternate Terminator

AP Alternate Promoter

AUC Area under curve

AIG1 Androgen-induced gene 1

CRC Colorectal cancer;

ES Exon Skip

GU Genomic unstable

GS Genomic stable

HCC Hepatocellular carcinoma

HR Hazard ratio

HSPA1A Heat shock protein family A (Hsp 70) member 1A

LASSO Least absolute shrinkage and selection operator

MMR Mismatch repair

ME Mutually Exclusive Exons

NSCLC Non-small cell lung cell cancer

OS Overall survival

PXR Pregnane X receptor

PSI Percent Splicing index

POLE Polymerase ε

PRMT1 Protein arginine methyltransferase 1

PHF11 PHD finger protein 11

RI Retained Intron

ROC Receiver operating characteristic

SAM Significance analysis of microarrays

SNP Single nucleotide polymorphism

SF Splicing factors

SETD2 SET domain containing 2, histone lysine methyltransferase

TATDN1 TatD DNase domain containing 1

TCGA The Cancer Genome Altas

UGT1A UDP glucuronosyltransferase family 1 member A complex locus

VEZT Vezatin, adherens junctions transmembrane protein.

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Original Research
PUBLISHED 05 August 2022
DOI 10.3389/fbioe.2022.822835

OPEN ACCESS

EDITED BY
Ping Zhang,
Griffith University, Australia

REVIEWED BY
Nguyen Quoc Khanh Le,
Taipei Medical University, Taiwan
Giulia Russo,
University of Catania, Italy

*CORRESPONDENCE
Ye Feng,
fengye@jlu.edu.cn

†These authors share first authorship

SPECIALTY SECTION
This article was submitted to Preclinical Cell and Gene Therapy, a section of the journal Frontiers in Bioengineering and Biotechnology

RECEIVED 26 November 2021
ACCEPTED 12 July 2022
PUBLISHED 05 August 2022

CITATION
Liao L, Gao Y, Su J and Feng Y (2022), By characterizing metabolic and immune microenvironment reveal potential prognostic markers in the development of colorectal cancer.
Front. Bioeng. Biotechnol. 10:822835.
doi: 10.3389/fbioe.2022.822835

By characterizing metabolic and immune microenvironment reveal potential prognostic markers in the development of colorectal cancer

Liangliang Liao[1†], Yongjian Gao[1†], Jie Su[2] and Ye Feng[1*]

[1]China-Japan Union Hospital of Jilin University, Changchun, China, [2]The First Hospital of Jilin University, Changchun, China

Colon adenocarcinoma (COAD) is one of the deadliest cancers in the world and survival rates vary significantly between early and advanced stage patients. Therefore, the identification of the pathogenesis in the development of COAD and prognostic markers is urgently demanded. Herein, we collected RNA-seq and somatic mutation data of COAD for statistical analysis. Clinical stage-specific differentially expressed genes (DEGs) and tumor development-dependent DEGs were identified. By characterizing the metabolic and immune features of COAD between stages, we found that the energy supply and inflammatory response of advanced tumors were suppressed. Next, the *ETS1, AR, GATA1, GATA2, SREBF1, FOXP3, STAT4,* and *NFKB1* were identified to drive the metabolic and immune-related pathways in the development of COAD. The three potential prognostic markers (*HOXC8, IRF7,* and *CXCL13*) were identified based on Cox regression analysis. Additionally, immune infiltration analysis revealed that the resting $CD4^+$ T cell was significantly related to the overall survival (OS) of COAD patients. Collectively, the specific metabolic and immune characteristics of advanced patients and the identified prognostic biomarkers will contribute to the development of precision medicine.

KEYWORDS

clinical stages, somatic mutation, metabolic and immune, prognostic markers, transcriptional regulation

Introduction

Colorectal cancer is one of the deadliest cancers in the world, killing nearly a million people every year (Labianca et al., 2010; Dekker et al., 2019). Although advances in diagnosis and treatment methods improve the prognosis of early-stage patients, it is still an important cause of cancer-related deaths (Lech et al., 2016). The locations where tumors often occur are divided into proximal colon, distal colon, and rectum, and the ratio of patients reaches 4:2:3 (Cheng et al., 2011). With the development of diagnostic technology, the number of young patients diagnosed with colon cancer increase.

Recent study data suggest that the overall five-year relative survival rate for COAD patients exceeds 60%, it varies depending on the clinical stages. Therefore, the dynamic changes in the physiological mechanisms involved in the development of COAD need to be urgently determined.

In the process of tumor development, it is accompanied by metabolic reprogramming to support tumor cells' demand for proliferation and metastasis (DeBerardinis and Chandel, 2016). Genetics and environmental are important driving factors of cell metabolism (Boroughs and DeBerardinis, 2015; Pavlova and Thompson, 2016). Among them, different tumor stages have specific physiological environments (La Vecchia and Sebastian, 2020). To explore metabolic reprogramming during tumor development, it is necessary to use gene expression to measure metabolic pathway activity (Puram et al., 2017).

So far, the diagnosis of colorectal cancer still relies on colonoscopy, but the treatment methods have developed significantly. In recent years, immune checkpoint therapy have become hot spots in cancer treatment. For example, a novel treatment for colorectal cancer was proposed based on the immune checkpoint PD-1/PD-L1 (Yaghoubi et al., 2019). However, not all COAD patients show complete response to PD-1, and there are some adverse reactions (Wu et al., 2019). Hence, in this work, we explored the immune cell landscape and the activity of immune-related pathways in the development of COAD.

In this study, we collected RNA-seq and mutation data of colorectal tumors and normal samples from The Cancer Genome Atlas (TCGA) database to identify metabolic and immune characteristics in the development of COAD. Transcriptional regulatory networks were constructed to identify drive factors that play important roles in immune and metabolic pathways. Potential prognostic markers identified by Cox regression analysis were used to construct survival risk models for COAD. Moreover, immune infiltration analysis revealed the immune landscape of COAD.

Materials and methods

Data collection

First, we downloaded the gene expression data (including tumor samples and normal samples), somatic mutation data and clinical information of COAD patients from the TCGA database (Tomczak et al., 2015). The hallmark, KEGG, GO Biological Process and metabolic pathway gene sets were collected from The Molecular Signatures Database (MSigDB (Liberzon et al., 2015), http://www.gsea-msigdb.org/gsea/msigdb/) database. Further, the transcription factors (TF)-TG data for human were downloaded from the TRRUST (Han et al., 2018) (https://www.grnpedia.org/trrust/) and ORTI (Vafaee et al., 2016) databases (http://orti.sydney.edu.au/about.html). The signature profiles of leukocyte were collected from the CIBERSORTx (Newman et al., 2015) (https://cibersortx.stanford.edu/) database.

Differential expression analysis

Here, in addition to considering the difference in gene expression between all the tumor sample and the normal sample collected from TCGA, the difference in gene expression between the tumor sample and the corresponding normal sample in each clinical stage was also considered. Differentially expressed genes were identified using the R package Limma (Ritchie et al., 2015). We considered the genes with $|log_2FC| > 1.5$ and $p\text{-}value < 0.01$ as the differentially expressed RNAs (DEGs).

Statistical analysis of mutation data

The somatic mutation data of COAD collected from TCGA was used to describe the mutation of signature genes. Then, the R package maftools (Mayakonda et al., 2018) was used for the statistical and visualization of mutation location, mutation form, mutation frequency and other information for these signature genes.

Gene set level analysis

For the metabolic gene sets collected from the MSigDB database, gene set variation analysis (Hanzelmann et al., 2013) (GSVA) was used to calculated the enrichment score of each stage of COAD in metabolic pathway by R package GSVA (v1.36.3). GSVA is a non-parametric, unsupervised algorithm. Further, the ten pathways and biological pathways related to immune checkpoints, antigen presentation, and immune activation or suppression were extracted from hallmark, KEGG and GO Biological Process gene sets. Using the appealed GSVA algorithm, the activity score of each stage of COAD in these immune-related pathways was calculated.

Construction of transcriptional regulatory network

First, we used the analysis of variance (ANOVA) algorithm to calculate the metabolic genes and immune-related genes specifically expressed between samples grouped in stages. We defined the metabolic and immune-related DEGs specifically expressed between samples grouped by stage as stage-MDEGs and stage-IDEGs. Next, stage-MDEGs and stage-IDEGs were extracted for the construction of transcriptional regulatory

network based on the TF-target gene data collected from TRRUST and ORTI databases. The TF-target gene relationship pairs related to stage-MDEGs and stage-IDEGs were extracted. Further, the Pearson correlation coefficient (*R*) between the genes of each pair was calculated and the cutoff of the *p*-value and *R* were set to 0.05 and 0.2. Moreover, the TFs-target genes network was constructed using Cytoscape (Shannon et al., 2003) (3.7.0) tool. The topological properties of the network then were calculated and the top genes of degree were identified as key drive factors.

Functional enrichment analysis

First, the genes whose expression was affected through the transcriptional regulatory mechanism was collected. The R package clusterprofiler (Yu et al., 2012) was used to perform GO functional enrichment on these genes. We set *p*-value<0.05 to screen for significantly enriched biological pathway. The relationship between these biological pathways and the corresponding genes was depicted using the R package circlize (Gu et al., 2014) (v0.4.10).

Identification of prognosis markers

Based on the appealed transcriptional regulatory network research, the TF and target genes were used as candidate factors for the identification of prognostic markers of COAD. First, the univariate COX regression was used to screen for the prognostic related genes using patient's survival data including survival state and overall survival (OS) of COAD (the cutoff of *p-value* was 0.05). The patients of COAD were randomly sub-grouped into the training set and test set in accordance with the ratio of 7:3. Further, the train set were used to construct multivariate COX regression model (Fisher and Lin, 1999). The reliability of the survival prediction model was described by the receiver operating characteristic curve (ROC), and the area under the curve (AUC) was calculated. The PH hypothesis test was also used to calibrate the model. The gene that *p*-value of the Schoenfeld Individual test greater than 0.05 was reserved for the reconstruction of the multivariate cox regression model. Moreover, we used the nomogram algorithm to build a COAD survival risk prediction model.

Calculation of risk score

First, we calculate the risk score of each patient for COAD based on the linear combination of the expression values weighted by the coefficients of the multivariate Cox regression analysis:

$$Risk\ score\ (i) = \sum_{k=1}^{n} \beta_k * e_{ki} \quad (1)$$

Where n represents the number of prognostic-related genes, i represents the order of genes, and k represents the order of patients. The regression coefficient and gene expression value are represented by β and e respectively. Then, we calculated the risk scores of the samples and divided the samples into high-risk and low-risk categories based on the median risk score. The Kaplan-Meier survival curve (Ranstam and Cook, 2017) was used to describe the patient's survival probability of high- and low-risk group, and calculated the statistical difference with the bilateral log-rank test (Guyot et al., 2012). Besides, the above survival analysis process was also carried out in an independent data set (GSE38832 (Tripathi et al., 2014)) to confirm the robustness and stability of prognostic markers.

Results

Stage-specific transcriptional and mutational landscape in Colon adenocarcinoma development

The key to the treatment of colorectal cancer is early detection and timely diagnosis (The, 2018). Therefore, exploring the dynamic changes of molecules in the development of COAD was beneficial to reveal the driving mechanism of the physiological state of patients in different stages. We developed a pipeline to explore the dynamic molecular mechanisms in COAD development (Supplementary Figure S1). First, differential expression analysis revealed that 9,859 DEGs (3,047 up-regulation and 6,812 down-regulation) were identified between tumor samples and para-cancerous samples of COAD (Figure 1A). The top 10 up-regulated genes were marked. Among them, *WNT2* is an important component in the WNT signaling pathway and promotes tumor angiogenesis in colon cancer (Unterleuthner et al., 2020). With the development of tumors, the prognosis of advanced patients will be severely disrupted (Huang et al., 2020), which was also effective for COAD. We found that the prognosis of COAD patients is consistent with its clinical stage and that advanced patients were associated with the worst OS (Figure 1B). Further, to explore the specific expression of biomolecules in patients of four clinical stages, we have statistically tested the DEGs of each clinical stage. The 2326 DEGs (868 up-regulation and 1,461 down-regulation) in stage I, the 7,857 DEGs (1,959 up-regulation and 5,898 down-regulation) in stage II, the 3,976 DEGs (1,147 up-regulation and 2,820 down-regulation) in stage III, and the 4,200 DEGs (1,320 up-regulation and 2,880 down-regulation) in stage IV were identified. By integrating the DEGs identified in the overall tumor sample and the DEGs identified in each clinical stage, 33 up-regulated genes and 117 down-regulated genes were identified in different stages (Figure 1C), which indicated that molecular and functions have been reprogrammed during the development of COAD. For

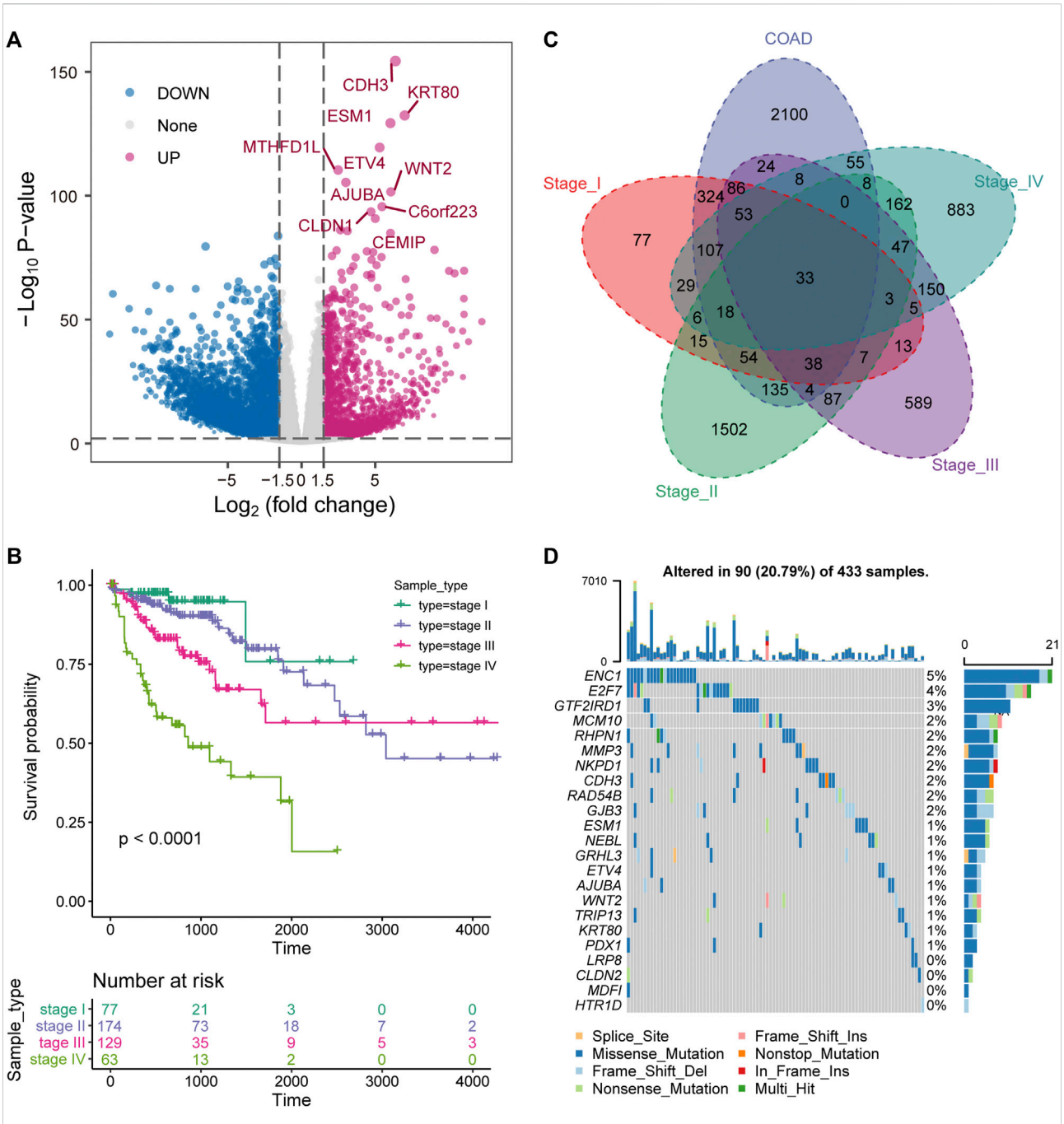


FIGURE 1
Stage-specific transcriptional and mutational landscape in COAD development. **(A)** The results of the DEGs between the whole tumor sample of COAD and the normal sample are displayed by volcano plot. The x-axis represents log2 (Fold Change). The y-axis displays -log10 (*p* value). **(B)** Kaplan-Meier (KM) curves depict the survival of patients in four stages for COAD. Log-rank test is used to calculate statistical significance. **(C)** Venn diagram shows the intersection of up-regulated genes between clinical stages and overall tumor samples. **(D)** The waterfall chart shows the mutation information of genes that are continuously up-regulated in the development of COAD, and the mutation type of each gene in each sample is displayed.

the 33 up-regulated genes, we used COAD mutation data to describe their mutation landscape (Figure 1D). Among them, according to the current mutation data, there are no somatic mutations in the genomic positions of 10 genes. We found that ENC1 has the top mutation frequency (Figure 1D), and it was significant co-occurence with *GRHL3, RHPN1* and *E2F7* at the mutation level (Supplementary Figure S2). Moreover, previous studies have shown that *ENC1* promotes the progression of

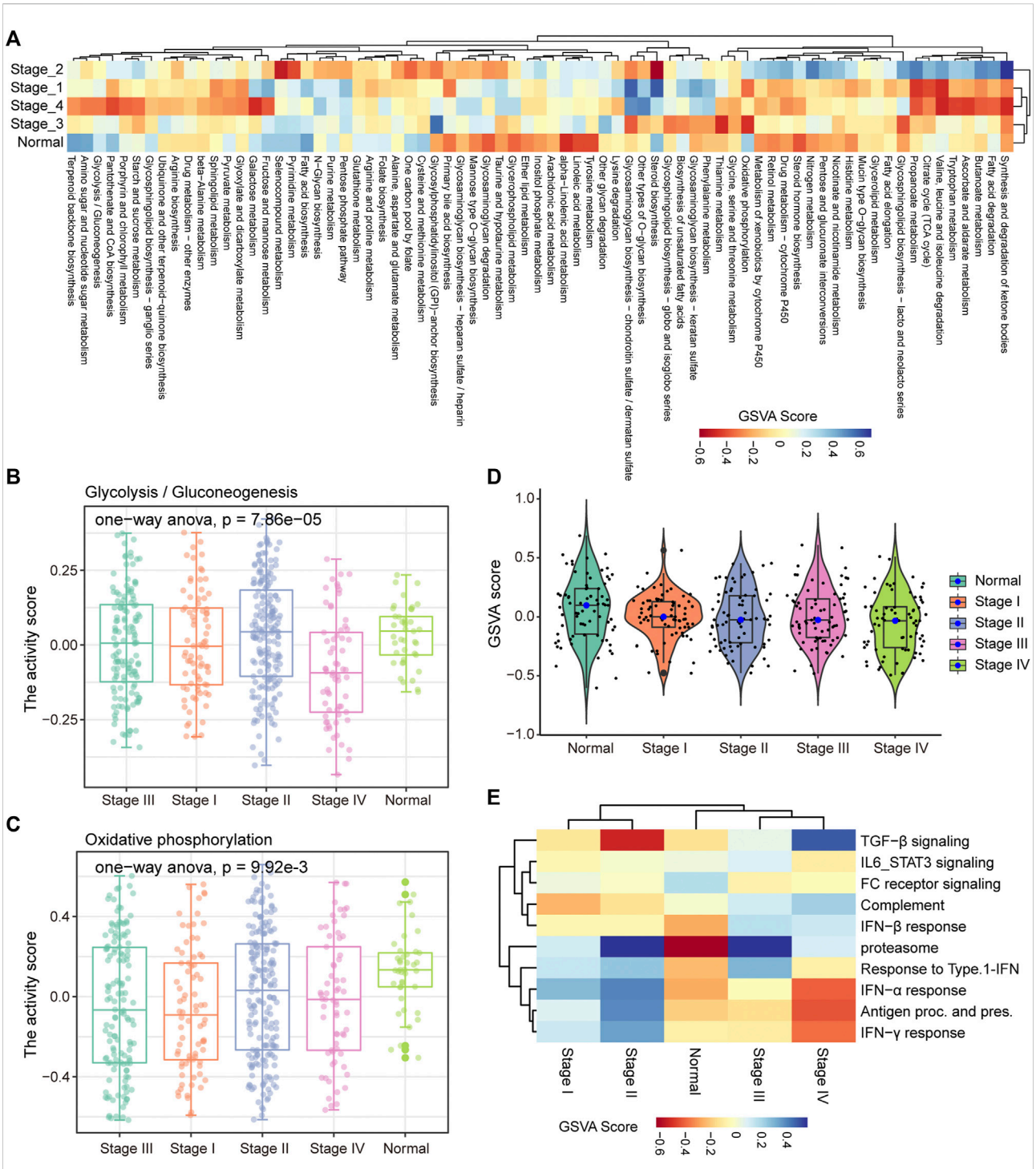


FIGURE 2
Stage-specific metabolic and immune activity. **(A)** The enrichment scores of the 85 metabolic pathways calculated by GSVA in the five categories (stage I, stage II, stage III, stage IV, normal samples) are displayed by the heat map. **(B–C)** Boxplot shows the glycolysis and OXPHOS pathway scores of each sample in the five categories. ANOVA is used to assess statistical differences between groups. **(D)** The distribution of enrichment scores for 85 metabolic pathways in the five categories is shown by violin chart. **(E)** The heat map shows the enrichment scores of each immune-related pathway in the five categories.

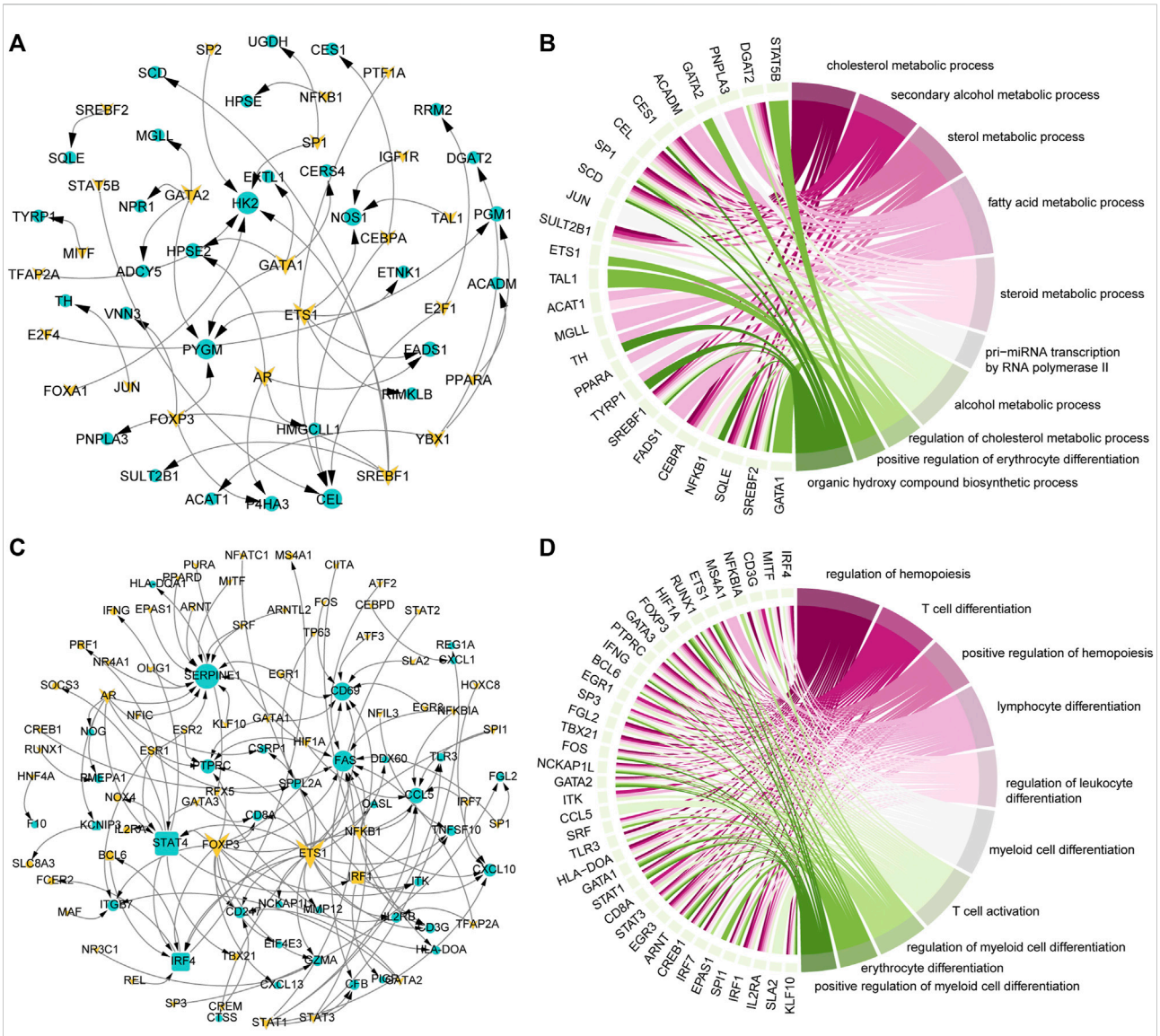


FIGURE 3
Key factors drive the reprogramming of metabolism and immunity. **(A)** Transcriptional regulatory network of metabolism-related genes that continue to be specifically expressed in the development of COAD. The triangle represents TF, the circle represents the target gene. Genes that continue to be specifically expressed in the development of COAD are marked in yellow. **(B)** The top 10 GO items enriched by genes in the **(A)** network. The interaction between genes and the GO items is shown. **(C)** Same as in **(A)** but for immune-related genes that continue to be specifically expressed in the development of COAD. **(D)** Same as in **(B)** but for genes in the **(C)** network.

colorectal cancer through *JAK2/STAT5/AKT* axis-mediated epithelial-mesenchymal transition and stemness (Cui et al., 2021).

Stage-specific metabolic and immune activity

In the process of tumor development, tumor cells undergo metabolic reprogramming to adapt to changes of the environment (Sun et al., 2018). To characterize the dynamic changes of metabolism in the development of COAD, GSVA was used to calculate the activity of 85 metabolic pathways collected from the MsigDB database in four clinical stages. We found that tumor tissue have activated energy supply compared with normal tissues (Figure 2A), which was consistent with previous studies showing that the activated metabolic microenvironment could supply tumor proliferation and metastasis (Wang et al., 2021). In different stages of tumor development, there are obvious differences in the activity of metabolic pathways. The activities of glycolysis and oxidative phosphorylation (OXPHOS) have been

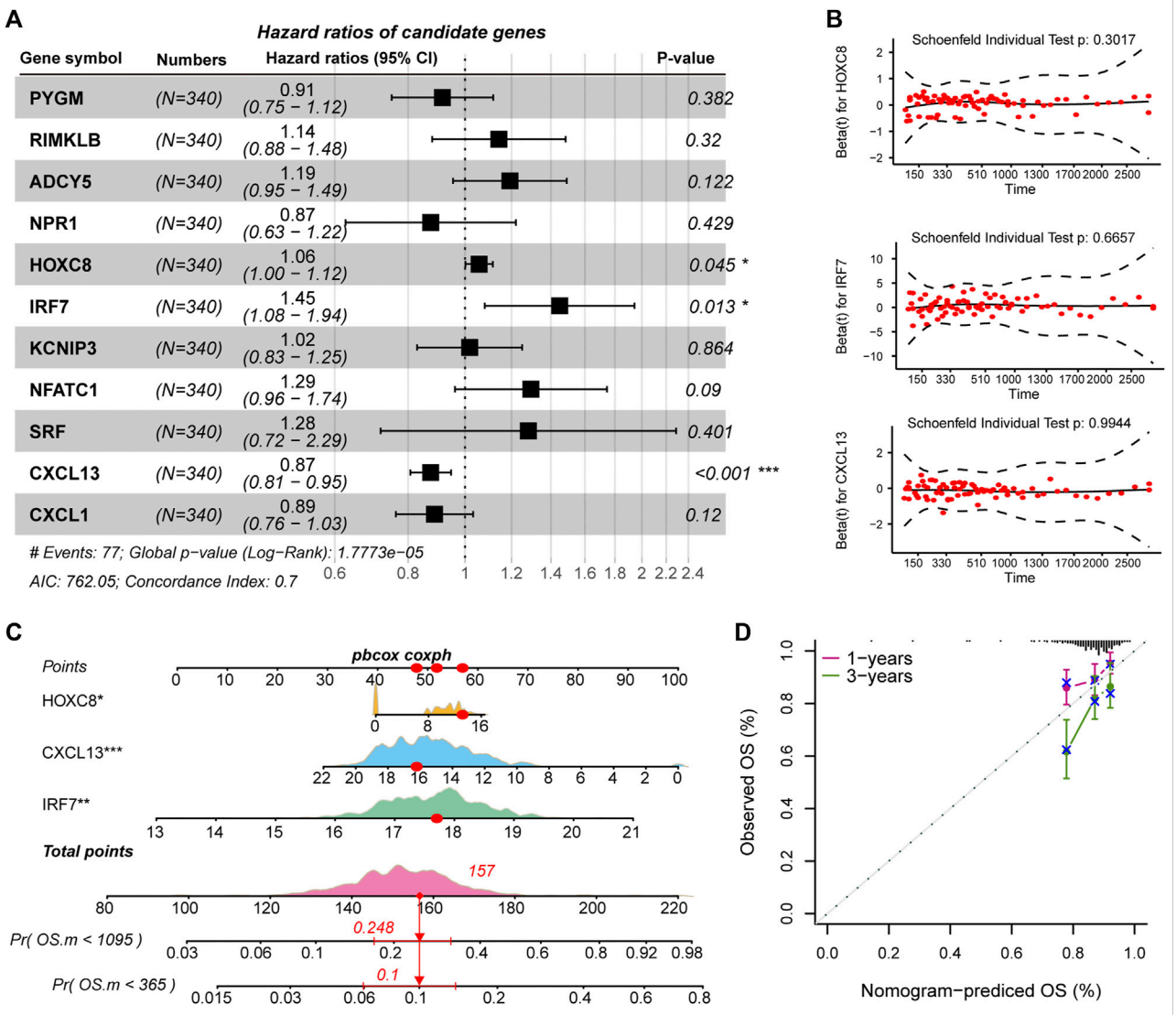


FIGURE 4
Identification of prognostic related genes in COAD. **(A)** Forest plots for multivariate Cox risk regression models. **(B)** PH hypothesis test of *HOXC8, IRF7,* and *CXCL13.* x-axis represents survival time, and y-axis represents Schoenfeld residuals. **(C)** The nomogram shows the prediction of 1-year and 3-years survival risk for patients of COAD. **(D)** Calibration curve of nomogram.

significantly enhanced in stage III (Figures 2B,C), which may be related to the proliferation of tumor cells at this stage. Furthermore, we analyzed the activity distribution of metabolic pathways between normal tissues and tumors of various stages. The global metabolic activity of normal tissues was higher than that of tumor tissues of each stage (Figure 2D), which indicates that tumor cells could selectively activate specific metabolic pathways to adapt to the development of this stage. For example, tumors in stage IV have high purity and metastasis to distant organs (Koo et al., 2020), and the reduction of glycolysis and the enhancement of OXPHOS were important metabolic characteristics of this stage for COAD (Figures 2B,C), which may be related to the increase in oxygen supply caused by the formation of blood vessels in local tissues. With the development of tumors, the immune microenvironment of tumors has also changed. We found that there were significant differences in the activity of the immune signaling pathways between the early and late stages of COAD (Figure 2E). In stage I and II of COAD, IL6/STAT3 signaling pathway, TGF-β signaling and FC receptor response have strong activity, which revealed the inflammatory activation of the immune response in the early stage of the tumor. Taken together, these results suggested that the development of COAD was accompanied by metabolic reprogramming and variation of the immune microenvironment.

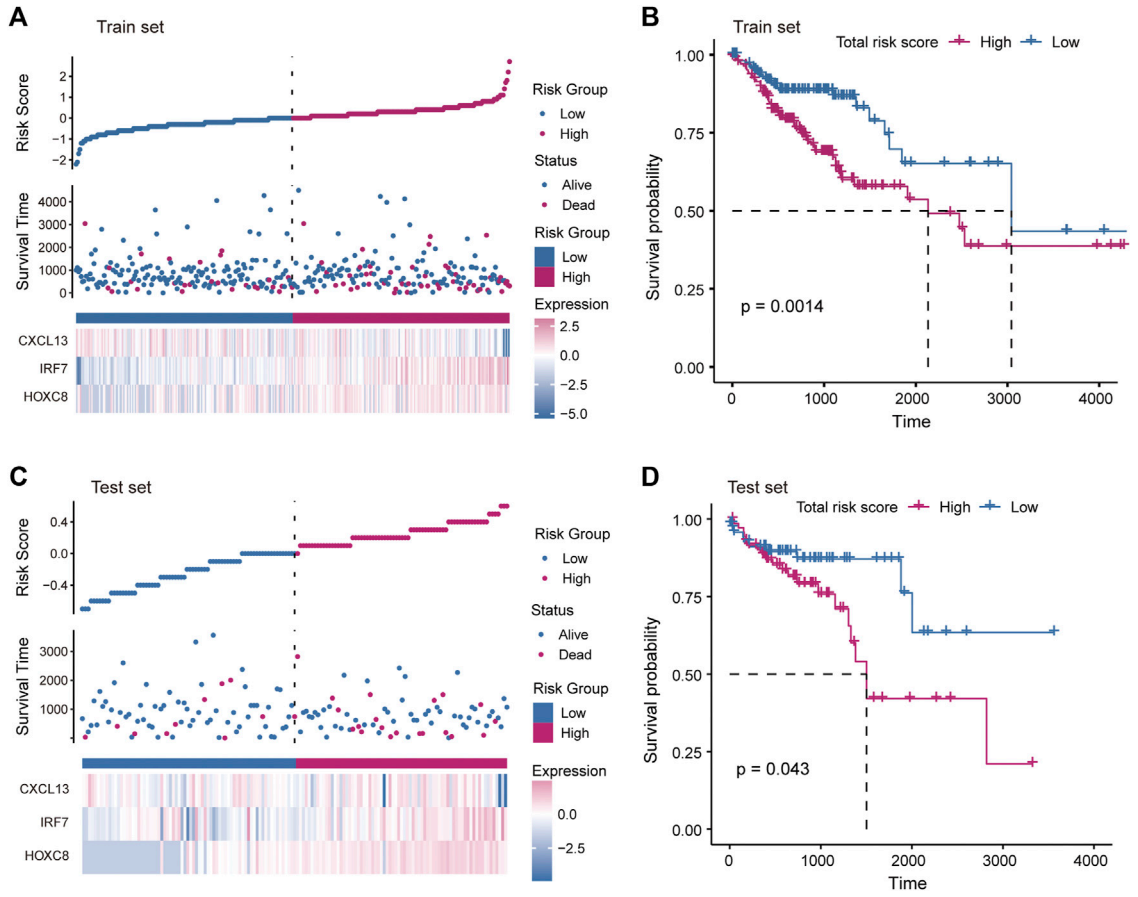


FIGURE 5
Construction of risk scoring model. **(A)** The figure shows the risk scores, survival status, and expression of prognostic markers for the train set samples. **(B)** The Kaplan-Meier curves for the survival of high-risk and low-risk groups in the train set. **(C)** Same as in **(A)** but for the test set samples. **(D)** Same as in **(B)** but for two groups in the test set.

Key factors drive the reprogramming of metabolism and immunity

TFs play an important role in gene expression by regulating the initiation and intensity of gene transcription (Lambert et al., 2018). To identify the driving factors that regulate metabolism and immune activity in the development of COAD, the transcriptional regulatory networks were constructed. Based on metabolism-related genes, 23 TFs and 30 target genes were identified and 46 TF-target gene units were formed (Figure 3A). By analyzing the topological properties of the network, we have identified the top 5 TFs (*ETS1, AR, GATA1, GATA2,* and *SREBF1*) of degree. The *ETS1* has been shown to be a driving factor for the progression of majority cancers (Dittmer, 2015; Chen et al., 2019) and its down-regulation inhibits the progression of colorectal cancer (Gu et al., 2019). These results indicated that *ETS1, AR, GATA1, GATA2,* and *SREBF1* could be used as biomarkers in the development of COAD. In addition to metabolic pathways, the results of functional enrichment analysis showed that genes involved in the transcriptional regulatory network were significantly enriched in the transcription of non-coding RNA and cell differentiation (Figure 3B). In the immune-related transcriptional regulatory network, 55 TFs and 48 target genes constituted the 129 TF-target gene units (Figure 3C). The *ETS1* also the top gene of degree in this network. The remaining four high degree TFs were *FOXP3, STAT4, AR,* and *NFKB1*. Among them, the *FOXP3* was closely related to the differentiation of T cells and was lineage-defining TF for regulatory T cells (Ono, 2020). Moreover, we found that the genes in the immune-related transcriptional regulatory network were significantly enriched in the activation and differentiation of immune cells (Figure 3D). Taken together, these results suggested that the *ETS1* and *AR* were the driving factors of metabolic and immune reprogramming in the development of COAD.

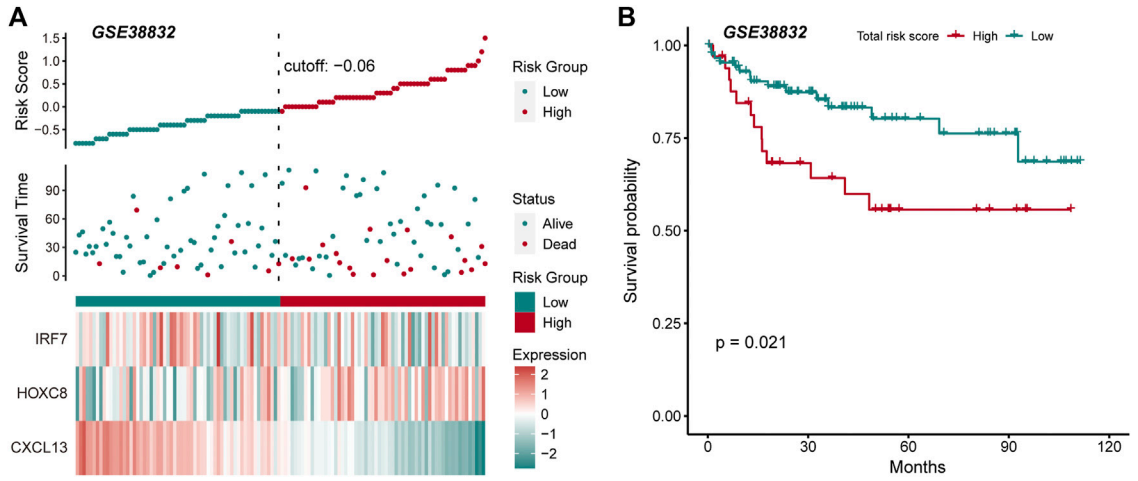


FIGURE 6
External data verify the robustness and stability of prognostic markers. **(A)** The figure shows the risk scores, survival status, and expression of prognostic markers for the GSE38832 set. **(B)** The Kaplan-Meier curves for the survival of high-risk and low-risk groups in the GSE38832 set.

Potential prognostic signature for Colon adenocarcinoma

The driving factors that play an important role in the development of COAD by regulating cell metabolism and immunity may determine the prognosis of patients. To identify potential prognostic markers of COAD, we integrated the statistical analysis pipeline from previous studies in this work (Dang et al., 2021; Le et al., 2021). The univariate Cox regression algorithm was used to fit the relationship between gene expression and patient's survival based on train set (including survival status and survival time). In this step, 11 genes were identified and significantly related to the patient's OS. Further, these genes were used to construct the multivariate Cox regression model. We found that three genes including *HOXC8, IRF7,* and *CXCL13* could be used as a potential prognostic signature for COAD (Figures 4A,B). To identify the best predictive time point for the survival prediction model, we divided the 1–3 years period into four periods and evaluated the prediction results using receiver operating characteristic curve (ROC). We found that the risk prediction result reached the maximum area under curve (AUC) value of 0.69 in the 912.5 days (Supplementary Figure S3). Based on these three prognostic markers, 1-year and 3-years survival risk prediction models of COAD were constructed and visualized through the nomograph (Figure 4C). The results of the calibration curve proved the stability of the risk prediction model (Figure 4D). Moreover, the risk scoring model was constructed as follows: risk score = -0.06 * *HOXC8* + 0.37* *IRF7* -0.13* *CXCL13*. The risk scores of the training and test set samples were calculated and they were divided into high risk and low risk groups based on the median risk score. We found the obvious expression difference of the three prognostic signature between the high/low-risk groups (Figure 5A) and the patients of high-risk score had poor prognosis (Figure 5B). The test set also showed the same prediction results as the train set (Figures 5C,D). Moreover, in the set of GSE38832 series, 122 samples were divided into two groups according to the upper quartile of risk scores (Figure 6A). Similarly, patients with high-risk scores had poorer OS (Figure 6B). All these suggesting that the *HOXC8, IRF7,* and *CXCL13* contributed to the prediction of the patient's prognosis for COAD and could be used for clinical diagnosis.

Immune cell components relate to the patient's survival risk

Previous studies have shown that the immune microenvironment plays an important role in the development of tumors (Hinshaw and Shevde, 2019; Lei et al., 2020). To identify the immune characteristics in the development of COAD, the CIBERSORTx tool was used to calculate the immune cell composition of samples for COAD and normal. For the 22 immune cell fraction matrices obtained, we found that COAD patients had immune infiltration compared with normal samples and there was no significant difference in immune infiltration between the clinical stages of tumors (Figure 7A). Further, the content of major histocompatibility complex (MHC) was calculated. We found that the gene encoding MHC-II molecule has a lower expression level in Stage IV, but a higher expression level in stage I and II (Figure 7B), which may explain the loss of immunogenicity in advanced patients of CAOD. Moreover, we evaluate the correlation of the immune cell fraction and risk score for patients of COAD. We found that the fraction of resting $CD4^+$ T cell, activated M $CD4^+$ T cell, and Treg were significantly related to the patient's survival risk (Figures 7C–E).

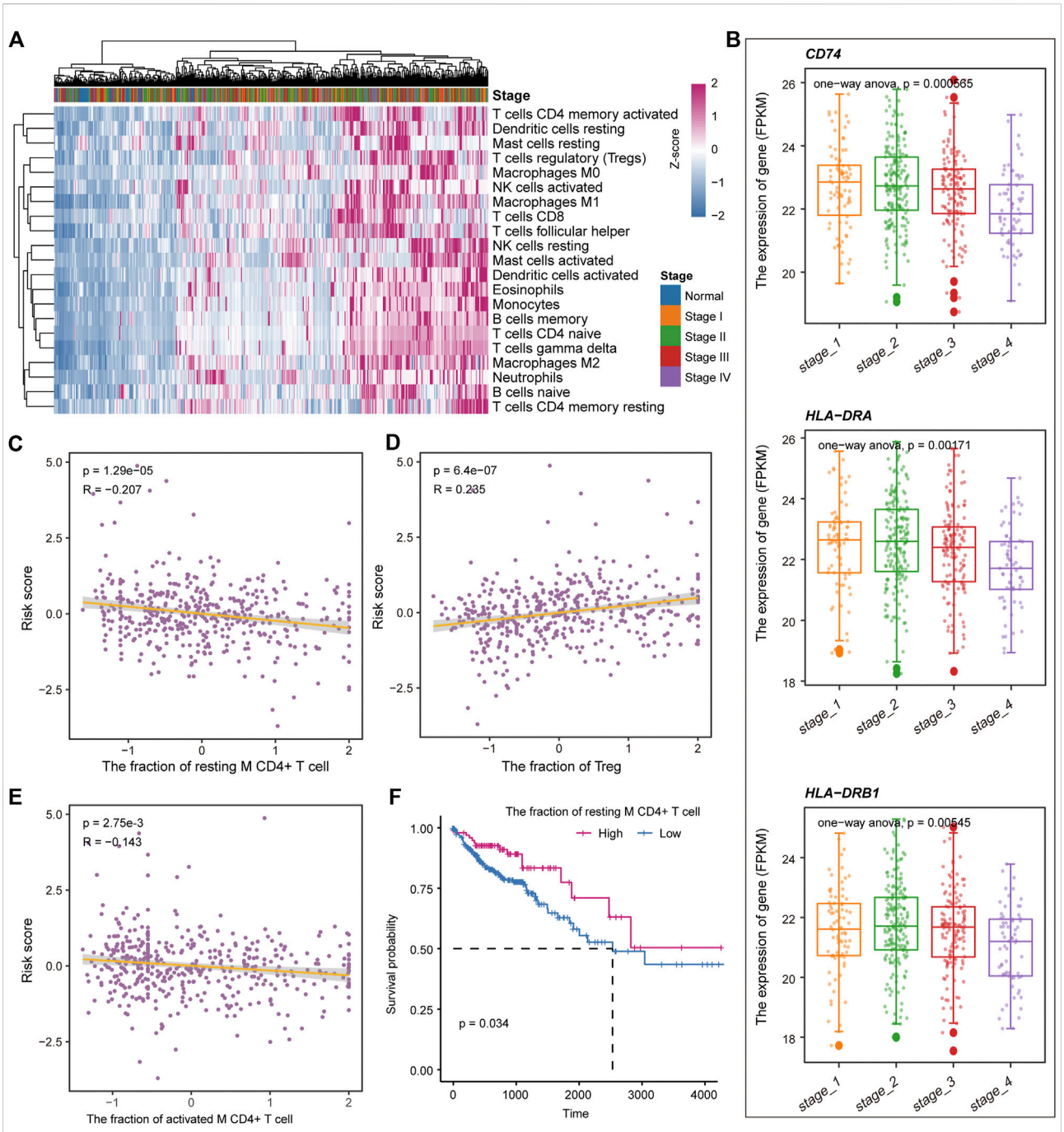


FIGURE 7
Immune cell components relate to the patient's survival risk. **(A)** The immune cell composition of each sample is displayed by heat map. The column label represents the clinical stage of the sample. **(B)** The expression levels of genes encoding MHC II molecules in each clinical stage are shown by boxplot. ANOVA is used to calculate statistical significance. **(C–E)** The correlation between the risk score and the fraction of the resting $CD4^+$ T cell, activated M $CD4^+$ T cell, and Treg. **(F)** Kaplan-Meier curves for survival in high/low-fraction groups of the resting $CD4^+$ T cell. Log-rank test was used to calculate statistical significance.

Based on the median of each immune cell component, the patients of COAD were divided into two groups (high/low-fraction). For the resting $CD4^+$ T cell, the patients with low-fraction of resting $CD4^+$ T cell were related to poor patient's prognosis (Figure 7F), suggesting that resting $CD4^+$ T cell may be a protective factor for COAD. Taken together, all these indicate that immune infiltration and tumor immunogenicity were closely related to the development and patient's survival of COAD.

Discussion

In this work, we have revealed the metabolic and immune characteristics in the development of COAD by integrating multi-omics data analysis. We found that COAD patients with different clinical stages had significant prognostic differences and advanced patients had the worst prognosis. For each clinical stage, stage-specific genes are identified and integrated analysis reveals 33 up-regulated genes and 117 down-regulated genes in all clinical stages. Combined with the somatic mutation data of the patients, the mutation landscape of these genes in COAD was revealed. Furthermore, stage-specific metabolic and immune activity were revealed through functional enrichment analysis. We found that energy metabolism (including glycolysis and OXPHOS) contributed to the development of COAD and is the basis for the changes in the physiological mechanism of each clinical stage. By constructing transcriptional regulatory networks, we have identified the key factors driving the development of COAD by disturbing metabolic and immune pathways. Moreover, we have identified three prognostic markers (*HOXC8, IRF7,* and *CXCL13*) of COAD based on the Cox regression algorithm and constructed a risk score model for the assessment of patient survival risk. By combining the patient's immune infiltration and survival data, we found that the resting $CD4^+$ T cell can be used as a protective factor for the patient.

Colorectal cancer is the fourth most deadly cancer in the world, causing nearly 900,000 deaths each year (Dekker et al., 2019). Since the disease only has symptoms in the late stages, it is necessary to identify its development mechanism and potential biomarkers. In recent years, there were majority studies on biomarkers and prognostic markers of COAD (Pellino et al., 2018; Patel et al., 2019). For example, Razi et al. revealed DCLK1 as a marker of stem cell regulates tumor progression and invasion from the perspective of ceRNA mechanism (Razi et al., 2021). Pankaj Ahluwalia et al. simply used KM analysis and Cox regression algorithm to identify prognostic markers of COAD (Ahluwalia et al., 2019). We focused on the development of tumors and were committed to revealing its dynamic physiological mechanisms. The clinical stage of COAD patients was revealed to be significantly related to the prognosis, indicating that the clinical stage could partly reflect the development of the tumor.

In the transcriptional regulatory network, we have identified hub TFs for metabolism and immune regulation of COAD. The *ETS1* and *AR* were the driving factors both of metabolic and immune pathway, suggesting that *ETS1* and *AR* could be used as potential biomarkers for COAD. We found that patients with COAD have global immune cell infiltration compared with normal tissues and the wide heterogeneity of immune cells in each clinical stage, which is consistent with previous studies (Ge et al., 2019).

Conclusion

In summary, our research revealed the metabolic and immune characteristics in the development of COAD, and identified potential biomarkers through biological network analysis. Three potential prognostic markers were identified. Through immune infiltration analysis, the immune landscape of COAD was revealed and the resting $CD4^+$ T cell was identified as a protective factor.

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

YF and LL conceived and designed the experiments. LL and YG performed analysis. LL and JS collected the data and wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgments

The authors gratefully thank the TCGA for providing data for this work.

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/fbioe.2022.822835/full#supplementary-material

References

Ahluwalia, P., Mondal, A. K., Bloomer, C., Fulzele, S., Jones, K., Ananth, S., et al. (2019). Identification and clinical validation of a novel 4 gene-signature with prognostic utility in colorectal cancer. *Int. J. Mol. Sci.* 20, 3818. doi:10.3390/ijms20153818

Boroughs, L. K., and DeBerardinis, R. J. (2015). Metabolic pathways promoting cancer cell survival and growth. *Nat. Cell Biol.* 17, 351–359. doi:10.1038/ncb3124

Chen, Y., Peng, C., Chen, J., Chen, D., Yang, B., He, B., et al. (2019). WTAP facilitates progression of hepatocellular carcinoma via m6A-HuR-dependent epigenetic silencing of ETS1. *Mol. Cancer* 18, 127. doi:10.1186/s12943-019-1053-8

Cheng, L., Eng, C., Nieman, L. Z., Kapadia, A. S., and Du, X. L. (2011). Trends in colorectal cancer incidence by anatomic site and disease stage in the United States from 1976 to 2005. *Am. J. Clin. Oncol.* 34, 573–580. doi:10.1097/coc.0b013e3181fe41ed

Cui, Y., Yang, J., Bai, Y., Li, Q., Yao, Y., Liu, C., et al. (2021). Corrigendum: ENC1 facilitates colorectal carcinoma tumorigenesis and metastasis via JAK2/STAT5/AKT axis-mediated epithelial mesenchymal transition and stemness. *Front. Cell Dev. Biol.* 9, 758671. doi:10.3389/fcell.2021.758671

Dang, H. H., Ta, H. D. K., Nguyen, T. T. T., Anuraga, G., Wang, C. Y., Lee, K. H., et al. (2021). Identifying gpsm family members as potential biomarkers in breast cancer: A comprehensive bioinformatics analysis. *Biomedicines* 9, 1144. doi:10.3390/biomedicines9091144

DeBerardinis, R. J., and Chandel, N. S. (2016). Fundamentals of cancer metabolism. *Sci. Adv.* 2, e1600200. doi:10.1126/sciadv.1600200

Dekker, E., Tanis, P. J., Vleugels, J. L. A., Kasi, P. M., and Wallace, M. B. (2019). Colorectal cancer. *Lancet* 394, 1467–1480. doi:10.1016/s0140-6736(19)32319-0

Dittmer, J. (2015). The role of the transcription factor Ets1 in carcinoma. *Semin. Cancer Biol.* 35, 20–38. doi:10.1016/j.semcancer.2015.09.010

Fisher, L. D., and Lin, D. Y. (1999). Time-dependent covariates in the Cox proportional-hazards regression model. *Annu. Rev. Public Health* 20, 145–157. doi:10.1146/annurev.publhealth.20.1.145

Ge, P., Wang, W., Li, L., Zhang, G., Gao, Z., Tang, Z., et al. (2019). Profiles of immune cell infiltration and immune-related genes in the tumor microenvironment of colorectal cancer. *Biomed. Pharmacother.* 118, 109228. doi:10.1016/j.biopha.2019.109228

Gu, C., Cai, J., Xu, Z., Zhou, S., Ye, L., Yan, Q., et al. (2019). MiR-532-3p suppresses colorectal cancer progression by disrupting the ETS1/TGM2 axis-mediated Wnt/β-catenin signaling. *Cell Death Dis.* 10, 739. doi:10.1038/s41419-019-1962-x

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

Guyot, P., Ades, A. E., Ouwens, M. J., and Welton, N. J. (2012). Enhanced secondary analysis of survival data: Reconstructing the data from published kaplan-meier survival curves. *BMC Med. Res. Methodol.* 12, 9. doi:10.1186/1471-2288-12-9

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, D380–D386. doi:10.1093/nar/gkx1013

Hanzelmann, S., Castelo, R., and Guinney, J. (2013). Gsva: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* 14, 7. doi:10.1186/1471-2105-14-7

Hinshaw, D. C., and Shevde, L. A. (2019). The tumor microenvironment innately modulates cancer progression. *Cancer Res.* 79, 4557–4566. doi:10.1158/0008-5472.can-18-3962

Huang, S., Yang, J., Fong, S., and Zhao, Q. (2020). Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges. *Cancer Lett.* 471, 61–71. doi:10.1016/j.canlet.2019.12.007

Koo, M. M., Swann, R., McPhail, S., Abel, G. A., Elliss-Brookes, L., Rubin, G. P., et al. (2020). Presenting symptoms of cancer and stage at diagnosis: Evidence from a cross-sectional, population-based study. *Lancet Oncol.* 21, 73–79. doi:10.1016/s1470-2045(19)30595-9

La Vecchia, S., and Sebastian, C. (2020). Metabolic pathways regulating colorectal cancer initiation and progression. *Semin. Cell Dev. Biol.* 98, 63–70. doi:10.1016/j.semcdb.2019.05.018

Labianca, R., Nordlinger, B., Beretta, G. D., Brouquet, A., Cervantes, A., and Group, E. G. W. (2010). Primary colon cancer: ESMO clinical practice guidelines for diagnosis, adjuvant treatment and follow-up. *Ann. Oncol.* 21 (5), v70–7. doi:10.1093/annonc/mdq168

Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., et al. (2018). The human transcription factors. *Cell* 172, 650–665. doi:10.1016/j.cell.2018.01.029

Le, V. H., Kha, Q. H., Hung, T. N. K., and Le, N. Q. K. (2021). Risk score generated from CT-based radiomics signatures for overall survival prediction in non-small cell lung cancer. *Cancers (Basel)* 13, 3616. doi:10.3390/cancers13143616

Lech, G., Slotwinski, R., Slodkowski, M., and Krasnodebski, I. W. (2016). Colorectal cancer tumour markers and biomarkers: Recent therapeutic advances. *World J. Gastroenterol.* 22, 1745. doi:10.3748/wjg.v22.i5.1745

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., Thorvaldsdottir, H., Ghandi, M., Mesirov, J. P., Tamayo, P., et al. (2015). The molecular signatures database hallmark gene set collection. *Cell Syst.* 1, 417–425. doi:10.1016/j.cels.2015.12.004

Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 28, 1747–1756. doi:10.1101/gr.239244.118

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

Ono, M. (2020). Control of regulatory T-cell differentiation and function by T-cell receptor signalling and Foxp3 transcription factor complexes. *Immunology* 160, 24–37. doi:10.1111/imm.13178

Patel, J. N., Fong, M. K., and Jagosky, M. (2019). Colorectal cancer biomarkers in the era of personalized medicine. *J. Pers. Med.* 9, 3. doi:10.3390/jpm9010003

Pavlova, N. N., and Thompson, C. B. (2016). The emerging hallmarks of cancer metabolism. *Cell Metab.* 23, 27–47. doi:10.1016/j.cmet.2015.12.006

Pellino, G., Gallo, G., Pallante, P., Capasso, R., De Stefano, A., Maretto, I., et al. (2018). Noninvasive biomarkers of colorectal cancer: Role in diagnosis and personalised treatment perspectives. *Gastroenterol. Res. Pract.* 2018, 1–21. doi:10.1155/2018/2397863

Puram, S. V., Tirosh, I., Parikh, A. S., Patel, A. P., Yizhak, K., Gillespie, S., et al. (2017). Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* 171, 1611–1624 e24. doi:10.1016/j.cell.2017.10.044

Ranstam, J., and Cook, J. A. (2017). Kaplan-Meier curve. *Br. J. Surg.* 104, 442. doi:10.1002/bjs.10238

Razi, S., Sadeghi, A., Asadi-Lari, Z., Tam, K. J., Kalantari, E., Madjd, Z., et al. (2021). DCLK1, a promising colorectal cancer stem cell marker, regulates tumor progression and invasion through miR-137 and miR-15a dependent manner. *Clin. Exp. Med.* 21, 139–147. doi:10.1007/s10238-020-00665-w

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

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

Sun, L., Suo, C., Li, S. T., Zhang, H., and Gao, P. (2018). Metabolic reprogramming for cancer cells and their microenvironment: Beyond the Warburg Effect. *Biochimica Biophysica Acta - Rev. Cancer* 1870, 51–66. doi:10.1016/j.bbcan.2018.06.005

The, H. (2018). Lancet gastroenterology, colorectal cancer screening: Is earlier better? *Lancet Gastroenterol. Hepatol.* 3, 519.

Tomczak, K., Czerwinska, P., and Wiznerowicz, M. (2015). Review the cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *wo.* 19, A68–A77. doi:10.5114/wo.2014.47136

Tripathi, M. K., Deane, N. G., Zhu, J., An, H., Mima, S., Wang, X., et al. (2014). Nuclear factor of activated T-cell activity is associated with metastatic capacity in colon cancer. *Cancer Res.* 74, 6947–6957. doi:10.1158/0008-5472.can-14-1592

Unterleuthner, D., Neuhold, P., Schwarz, K., Janker, L., Neuditschko, B., Nivarthi, H., et al. (2020). Cancer-associated fibroblast-derived WNT2 increases tumor angiogenesis in colon cancer. *Angiogenesis* 23, 159–177. doi:10.1007/s10456-019-09688-8

Vafaee, F., Krycer, J. R., Ma, X., Burykin, T., James, D. E., Kuncic, Z., et al. (2016). Orti: An open-access repository of transcriptional interactions for interrogating mammalian gene expression data. *PLoS One* 11, e0164535. doi:10.1371/journal.pone.0164535

Wang, F., Zhang, Y., Hao, Y., Li, X., Qi, Y., Xin, M., et al. (2021). Characterizing the metabolic and immune landscape of non-small cell lung cancer reveals prognostic biomarkers through omics data integration. *Front. Cell Dev. Biol.* 9, 702112. doi:10.3389/fcell.2021.702112

Wu, X., Gu, Z., Chen, Y., Chen, B., Chen, W., Weng, L., et al. (2019). Application of PD-1 blockade in cancer immunotherapy. *Comput. Struct. Biotechnol. J.* 17, 661–674. doi:10.1016/j.csbj.2019.03.006

Yaghoubi, N., Soltani, A., Ghazvini, K., Hassanian, S. M., and Hashemy, S. I. (2019). PD-1/PD-L1 blockade as a novel treatment for colorectal cancer. *Biomed. Pharmacother.* 110, 312–318. doi:10.1016/j.biopha.2018.11.105

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS A J. Integr. Biol.* 16, 284–287. doi:10.1089/omi.2011.0118

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Original Research
PUBLISHED 06 October 2022
DOI 10.3389/fbioe.2022.838129

OPEN ACCESS

EDITED BY
Andrea Banfi,
University of Basel, Switzerland

REVIEWED BY
Andras Szilagyi,
Institute of Enzymology, Hungary
Fátima Alejandra Pardo Avila,
Stanford University, United States

*CORRESPONDENCE
Wolfgang Schreiner,
wolfgang.schreiner@meduniwien.ac.at

SPECIALTY SECTION
This article was submitted to Preclinical Cell and Gene Therapy, a section of the journal Frontiers in Bioengineering and Biotechnology

RECEIVED 17 December 2021
ACCEPTED 14 September 2022
PUBLISHED 06 October 2022

CITATION
Kenn M, Karch R, Tomasiak L, Cibena M, Pfeiler G, Koelbl H and Schreiner W (2022), Molecular dynamics identifies semi-rigid domains in the PD-1 checkpoint receptor bound to its natural ligand PD-L1.
Front. Bioeng. Biotechnol. 10:838129.
doi: 10.3389/fbioe.2022.838129

Molecular dynamics identifies semi-rigid domains in the PD-1 checkpoint receptor bound to its natural ligand PD-L1

Michael Kenn[1], Rudolf Karch[1], Lisa Tomasiak[1], Michael Cibena[1], Georg Pfeiler[2], Heinz Koelbl[2] and Wolfgang Schreiner[1]*

[1]Institute for Biosimulation and Bioinformatics, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria, [2]Division of General Gynecology and Gynecologic Oncology, Department of Obstetrics and Gynecology, Medical University of Vienna, Vienna, Austria

Cells in danger of being erroneously attacked by leucocytes express PD-L1 on their surface. These cells activate PD-1 on attacking leucocytes and send them to death, thus curbing erroneous, autoimmune attack. Unfortunately, cancer cells exploit this mechanism: By expressing PD-L1, they guard themselves against leucocyte attack and thereby evade immune clearance. Checkpoint inhibitors are drugs which re-enable immune clearance of cancer cells by blocking the binding of PD-L1 to PD-1 receptors. It is therefore of utmost interest to investigate these binding mechanisms. We use three 600 ns all-atom molecular dynamics simulations to scrutinize molecular motions of PD-1 with its binding partner, the natural ligand PD-L1. Usually, atomic motion patterns are evaluated against whole molecules as a reference, disregarding that such a reference is a dynamic entity by itself, thus degrading stability of the reference. As a remedy, we identify semi-rigid domains, lending themselves as more stable and reliable reference frames against which even minute differences in molecular motion can be quantified precisely. We propose an unsupervised three-step procedure. In previous work of our group and others, minute differences in motion patterns proved decisive for differences in function. Here, several highly reliable frames of reference are established for future investigations based on molecular motion.

KEYWORDS

molecular dynamics, checkpoint inhibitor, immune therapy, oncology, drug design, cluster analysis

1 Introduction

1.1 Medical background and clinical significance

Immune system T-cells detect cancer cells as they develop, and normally kill them (Smith-Garvin et al., 2009). However, some cancer cells have developed mechanisms to escape this important, immune–mediated clearance (Chen and Mellman, 2013) as follows: T-cells present a suicide tool (PD-1) on their surface. In healthy individuals, this tool is activated (by PD-L1) only if a T-cell should erroneously attack healthy tissue. PD-1 is therefore called an "immune-checkpoint".

However, some cancer cells also express PD-L1 on their surface. They exploit the above checkpoint mechanism, abusively activate the immune checkpoint molecules (Dong et al., 2002) and thereby escape destruction. By increased expression of PD-L1 and/or the release of immunosuppressive factors cancer cells may survive even in a "hot", immune-cell enriched surrounding.

Checkpoint inhibitors are drugs blocking the binding between PD-1 and its natural ligand, PD-L1. Clinical trials have proved their efficacy (Brahmer et al., 2012; Kwa and Adams, 2018). More recently a phase III trial in metastatic triple negative breast cancer patients showed a distinct improvement in progression-free survival and overall survival (Brahmer et al., 2018). This demonstrates the significance of the target (PD-1) being expressed when a PD-L1 antibody is used (Schmid et al., 2018; Cortés et al., 2019).

In order to further improve these promising therapies, a better understanding of the molecular mechanism of the PD-1 receptor is necessary.

1.2 Rationale for multi-level clustering

To evaluate minute movements within molecular dynamics trajectories, all frames need to be fitted to a certain intramolecular region (i.e. domain) at a reference frame (point in time). Such a fitting domain should not significantly deform itself over time (along a trajectory), in order to serve as a stable reference against which very small and intricate movement patterns outside this domain can be detected.

In previous work, domains for fitting were usually selected manually, based on secondary structure, such as beta-strands, beta-sheets or alpha helices. We detect such stable regions in an unsupervised procedure from the computed dynamics itself. In particular for example, if parts of beta-strands participate in the binding mechanics to be evaluated, they should not at the same time be part of the domain to which fitting is performed.

A most direct approach would be clustering according to small changes in distance between pairs of atoms over the whole trajectory. However, it is known that molecular systems tend to switch between metastable states, each of which may pertain over considerable parts of the simulation. During such a metastable state, some pairs of atoms might remain in close vicinity, with little variation of their distances. For example, atoms in some loop, which assumes a certain conformation characteristic for this and only this meta-state. Clustering only during this meta-state would send these pairs into the same cluster. However, as the system switches to another meta-state, the very same pairs of atoms could be detached from each other, become members of different neighborhoods and end up in different clusters if clustering would be performed only over this second meta-state. In consequence, one single pass of clustering over the whole trajectory might particularly conceal minute patterns of motion, being of focal interest. Separate clustering of segments of a trajectory is likely to take account of such minute differences between meta-states and exclude these regions from semi-rigid domains to be obtained.

Deriving rigidity directly and unsupervised from the simulation is considered a promising advantage and basis for future MD-studies.

1.3 Molecular structures

The molecular structure of the immune checkpoint PD-1 is shown in Figure 1, generated with VMD (Humphrey et al., 1996; Hsin et al., 2008; Cross et al., 2009) from PDB (Burley, 2013) entry 4ZQK (Zak et al., 2017). Since 4ZQK does not contain the complete structure of PD-1, we have modelled the missing parts *in silico* already in our previous work (Roither et al., 2021). The immune checkpoint receptor, PD-1, consists of several beta strands in tight mutual binding and respective loops in between, see Table 1. These loops protrude loosely from a rather compact beta core and offer versatile modes of interaction and binding. In particular, the residues 70 to 77, comprising the CC′-loop, are crucial for interaction with the natural ligand PD-L1 (Kundapura and Ramagopal, 2019), see Figure 2. Details of this interaction have been investigated experimentally by Zak (Zak et al., 2015) and in molecular dynamics studies by Liu (Liu et al., 2017) and our group (Roither et al., 2020; Tomasiak et al., 2020; Roither et al., 2021; Tomasiak et al., 2021).

In the present work we draw on previous experience with the same system (Roither et al., 2020; Roither et al., 2021) but focus on unsupervised clustering, using a very efficient algorithm (Kenn et al., 2016) previously developed for MHC-molecules and T-cell receptors (Kenn et al., 2014).

2 Materials and methods

Molecular preparation and technical details of the molecular dynamics (MD) simulation have already been reported (Tomasiak et al., 2021). In Sections 2.1, 2.2, we briefly recapitulate essential

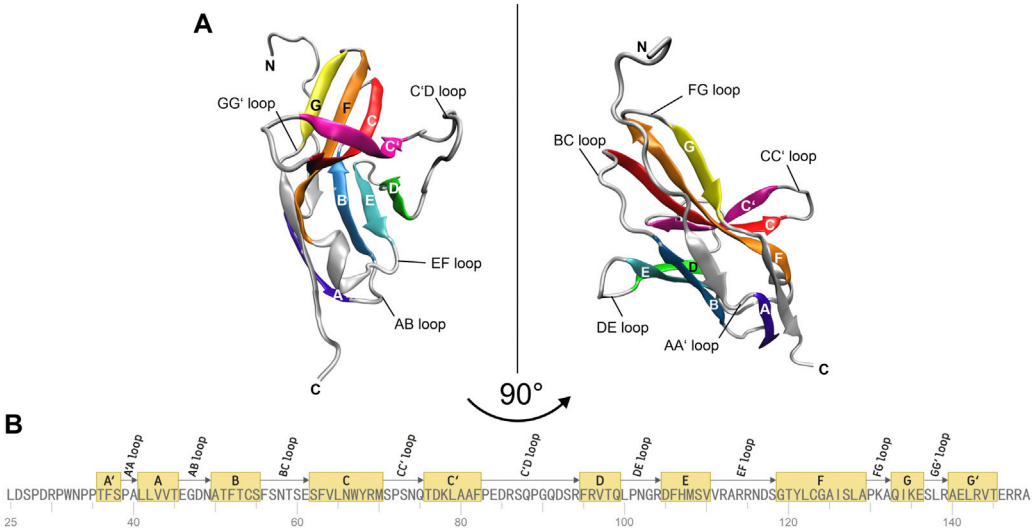


FIGURE 1
Molecular structure of immune checkpoint molecule PD-1. **(A)** Cartoon representation of the extracellular domain of PD-1. A two-layer β sandwich is formed by two β sheets GFCC′ (colored yellow, orange, red, magenta) and ABED (colored violet, blue, cyan, green) with loops connecting the respective β strands (colored silver). **(B)** Sequence of the residues of PD-1. The β strands of the protein are depicted as yellow boxes and the connecting loops as arrows. The figures were prepared using VMD version 1.9.3 (Humphrey et al., 1996).

points for completeness. The remaining subchapters Sections 2.3.1–2.3.5 refer to evaluation methods specific for this work.

2.1 Preparation of molecular complexes

Structural data for MD simulations were downloaded from the protein data bank (PDB, https://www.rcsb.org/) using the following entries: PDB-ID 4ZQK for the PD-1/PD-L1 system (resolution: 2.45 Å) (Zak et al., 2015) and PDB-ID 5GGS for the PD-1/pembrolizumab Fab fragment complex (resolution: 2.0 Å) (Lee et al., 2016). Missing residues in the crystal structure of the endogenous ligand PD-L1 in complex with the extracellular domain of PD-1 (PDB-ID 4ZQK), were added from the PD-1/pembrolizumab system (PDB-ID 5GGS), the N loop was taken from the PD-1/nivolumab system (PDB-ID 5WT9), see Roither et al. (Roither et al., 2020) for further preprocessing details. For determining the protonation states at pH 7.0 the H++ Server was used (http://biophysics.cs.vt.edu/) (Gordon et al., 2005). The assignment of strands, sheets, and loops was made following the classification of the Protein Feature View applet available within the 4ZQK record of the PDB (see Figure 1B).

2.2 All-atom molecular dynamics

As described previously (Tomasiak et al., 2021) all-atom MD simulations were performed with GROMACS 2021.2 (Hess et al., 2008), using the Amber99sb-ildn force field (Lindorff-Larsen et al., 2010) and an explicit water model. For the simulation box a rhombic dodecahedron was chosen with a minimum distance of 2 nm between the respective molecules and the box boundaries. The PD-1/PD-L1 complex consists of 4099 atoms and 240 residues, and the complex was solvated in TIP3P water (Jorgensen et al., 1983). Solute molecules were replaced by sodium and chloride ions to reach a physiological salt concentration of 0.15 mol/L.

For the energy minimization the method of steepest-descent was chosen. Before production runs the systems were equilibrated at NVT and NPT for 100 ps (time step 2 fs) each. In the NVT equilibration run the temperature was set to 310 K using a Berendsen-thermostat (Berendsen et al., 1984) with a time constant of 0.1 ps and position restraint MD. Equilibration in NPT ensembles was performed under the control of a Berendsen-barostat (Berendsen et al., 1984) set to 1 bar with a time constant of 1.0 ps.

All independent production runs had a simulation time of 600 ns with a time step of 2 fs using the LINCS algorithm (Hess, 2008) for constraining bonds to hydrogen atoms. For the van der Waals interactions a single cut-off of 1.47 nm was used and a cut-off distance of 1.4 nm for the short-range neighbor list in the Verlet scheme (Verlet, 1967) for neighborhood search. For electrostatic interactions the particle-mesh Ewald (PME) algorithm (Darden et al., 1993) was applied with a cut-off of 1.4 nm. Temperature coupling was done with the velocity-rescaling algorithm (Bussi et al., 2007) at a temperature of 310 K and for pressure coupling at 1 bar the Parrinello Rahman algorithm (Parrinello and Rahman, 1981) was used with a time constant of 2 ps. 30000 frames for each run were obtained by saving coordinates, velocities, forces, and energies

TABLE 1 Residues and secondary structure of PD-1. The assignment of strands and loops was chosen according to the classification of the Protein Feature View applet available within the 4ZQK record of the PDB. The domains were named following canonical Ig-strand designations (Zak et al., 2015). $ResID_S$ and $ResID_E$ indicate the starting and the ending residue ID of the according domain within chain B of 4ZQK. $Res\#_S$ and $Res\#_E$ indicate the starting and the ending residue number of a domain (continuous numbering for the whole complex in the respective PDB file).

Domain	$ResID_S$	$ResID_E$	Amino acid sequence	4ZQK	
				$Res\#_S$	$Res\#_E$
NtermA′ loop	25	35	LDSPDRPWNPP	116	126
A′ strand	36	38	TFS	127	129
A'A loop	39	40	PA	130	131
A strand	41	45	LLVVT	132	136
AB loop	46	49	EGDN	137	140
B strand	50	55	NATFTCS	141	146
BC loop	56	61	FSNTSE	147	152
C strand	62	70	SFVLNWYRM	153	161
CC′ loop	71	75	SPSNQ	162	166
C′ strand	76	82	TDKLAAF	167	173
C'D loop	83	94	PEDRSQPGQDSR	174	185
D strand	95	99	FRVTQ	186	190
DE loop	100	104	LPNGR	191	195
E strand	105	110	DFHMSV	196	201
EF loop	111	118	VRRRNDS	202	209
F strand	119	129	GTYLCGAISLA	210	220
FG loop	130	132	PKA	221	223
G strand	133	136	QIKE	224	227
GG′ loop	137	139	SLR	228	230
G′ strand	140	145	AELRVT	231	236
G'rest loop	146	149	ERRA	237	240
PD-L1 binding domain	70	77	MSPSNQTD	161	168
Pembrolizumab binding domain	74	99	NQTDKLAAFPEDRSQPGQDCRFRVTQ	165	190
NtermA′ loop	25	35	LDSPDRPWNPP	116	126

every 20 ps to a trajectory file. Three independent 600 ns MD simulations with different initial velocities were carried out for each system, summing up to a total simulation time of 600 ns * 3 = 1.8 μs.

Prior to the evaluation, all frames of each given trajectory were fitted to the first frame of the trajectory, according to minimum root mean square deviation (RMSD) at time t. In mathematical terms, the Cartesian coordinates $\mathbf{x}_i$ of all atoms i were translated and rotated towards minimum RMSD of the backbone within β-strands and α-helices:

$$RMSD(t) = \left[\frac{1}{N_{bb}}\sum_{i=1}^{N_{bb}}\left\|\mathbf{x}_i(t) - \mathbf{x}_i(0)\right\|^2\right]^{1/2} \rightarrow Min \qquad (1)$$

where $\mathbf{x}_i(t)$ is the position of atom i at time t. For the precise regions of secondary structure elements (β-strands and α-helices), see Tables 1, 2. N_{bb} is the total number of backbone atoms (N, C_α, C_O) contained therein. Finally, the first 100 ns of each trajectory were discarded to get rid of initial phase trends, leaving 500 ns with N_t = 25000 frames for each trajectory to be further analyzed.

2.3 Obtaining semi-rigid domains

Semi-rigid domains for a given trajectory were obtained in a two-step process: First, "spatial clustering" was performed by grouping C_α-atoms showing similar movements into each of the clusters. Evidently, such a clustering does not need to (and will not) yield exactly the same clusters if spatial clustering is performed for different subsections of N_f frames each (called "segments" in the following) of a trajectory: Some pairs of C_αs will stay together in a given cluster over many segments, others will not (Kenn et al., 2014). This fact is exploited to perform "time-wise clustering" as a second step, by constructing new clusters from those C_αs which stay together within spatial clusters across successive segments with maximum fidelity. Such groups of atoms form clusters even more stable over time and are hence

TABLE 2 Residues and secondary structure of PD-L1. The assignment of strands, loops and helices was chosen according to the classification of the Protein Feature View applet available within the 4ZQK record of the PDB protein data bank. The domains were named following canonical Ig-strand designations (Zak et al., 2015). Res#$_S$ and Res#$_E$ indicate the starting and the ending residue number of a domain (continuous numbering for the whole complex in the respective PDB file).

Domain	ResID$_S$	ResID$_E$	Amino acid sequence	Res#$_S$	Res#$_E$
NtermA loop	18	26	AFTVTVPKD	1	9
A strand	27	31	LYVVE	10	14
AB loop	32	35	YGSN	15	18
B strand	36	41	MTIECK	19	24
BH1 loop	42	48	FPVEKQL	25	31
Helix1	49	52	DLAA	32	35
H1C loop	53	53	L	36	36
C strand	54	59	IVYWEM	37	42
CC′ loop	60	61	ED	43	44
C′ strand	62	68	KNIIQFV	45	51
C′C″ loop	69	70	HG	52	53
C″ strand	71	72	EE	54	55
C″H2 loop	73	73	D	56	56
Helix2	74	82	LKVQHSSYR	57	65
H2D loop	83	84	QR	66	67
D strand	85	87	ARL	68	70
DH3 loop	88	88	L	71	71
Helix3	89	94	KDQLSL	72	77
H3E loop	95	95	G	78	78
E strand	96	101	NAALQI	79	84
EH4 loop	102	104	TDV	85	87
Helix 4	105	109	KLQDA	88	92
F strand	110	117	GVYRCMIS	93	100
FG loop	118	120	YGG	101	103
G strand	121	131	ADYKRITVKVN	104	114
Grest loop	132	132	A	115	115

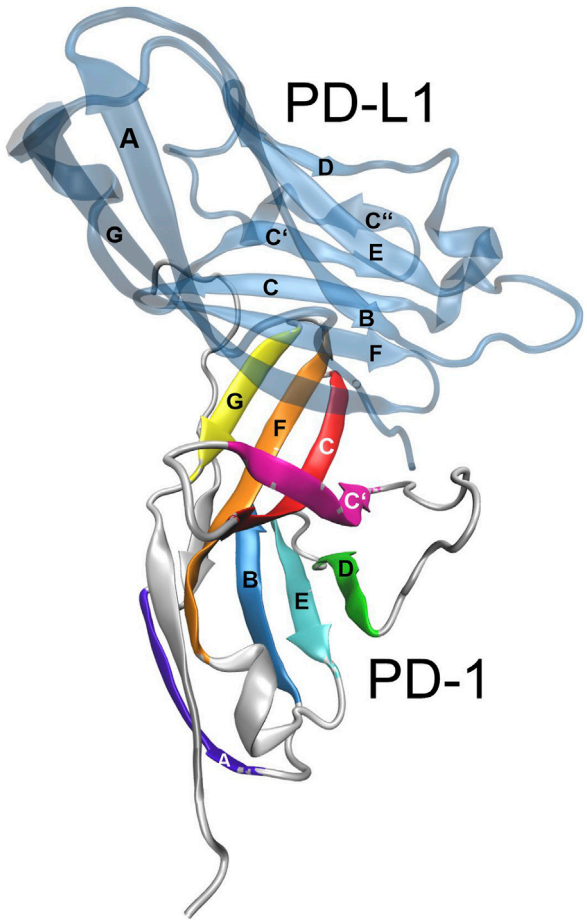


FIGURE 2
Immune checkpoint molecule PD-1 and binding partners. Cartoon representation of the extracellular domain of PD-1 bound to the endogenous ligand PD-L1 (transparent blue). The figure was prepared using VMD version 1.9.3 (Humphrey et al., 1996).

termed “semi-rigid domains” (Kenn et al., 2016). The total number of frames used from a trajectory, N_t, is partitioned into N_s segments, with $N_t = N_s * N_f$. We used $N_s = 500$ and $N_f = 50$, corresponding to 1 ns per segment and a frame length of 20 ps.

Note that time-wise clustering is a special mode of consensus clustering (Monti et al., 2003), since the same clustering algorithm is applied to different parts of a trajectory and a consensus between these results is finally adopted.

2.3.1 Spatial clustering

One crucial aspect of collective motion of atoms is captured by the variability (standard deviation) of mutual distances (Kenn et al., 2014), usually termed STDDV. We use it as an approximation for “motional distance” between two C_αs i and j, and denote it for brevity by D_{ij} defined as

$$D_{ij} = \sqrt{\frac{N_f}{N_f - 1}\left\langle \left(d_{ij} - \langle d_{ij}\rangle\right)^2 \right\rangle} \qquad (2)$$

where $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$ is the Euclidean distance in a given (time-wise) frame and $\langle\rangle$ denotes averaging over all N_f frames for which clustering is intended, see Figure 3. Note that distances are not affected by any fitting of the trajectory to a reference frame.

For actual spatial clustering (over segments or over the whole trajectory) we consider C_α atoms only and follow the concept of Bernhard and Noé (Bernhard and Noé, 2010). Each $C_{\alpha,i}$ is assigned a membership in cluster m, expressed as a real number $0 \le c_{i,m} \le 1$, with zero meaning no membership and 1 standing for full membership. According to Bernhard and

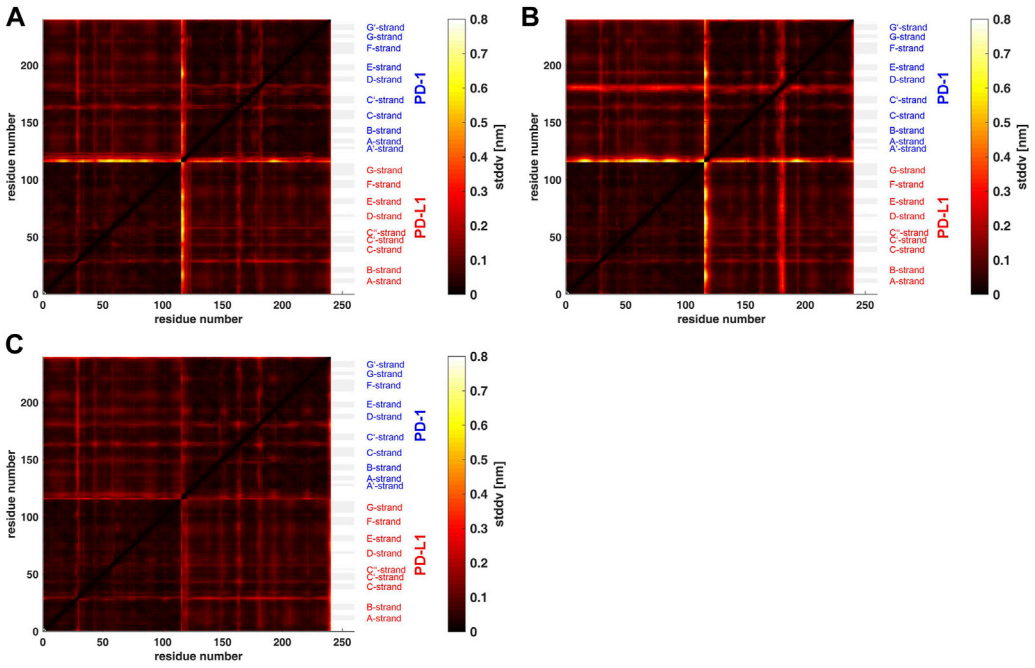


FIGURE 3
Matrix of standard deviations of atom distances over whole trajectories, shown as scaled color image (SCI). **(A)**: Trajectory t_1 for complex 4ZQK, consisting of receptor PD-1 and PD-L1 as ligand, showing enhanced similarity within two large areas (receptor and ligand, respectively). Note that numbering starts with ligand PD-L1 with residue-ID = 18 (lower left corner), corresponding to residue number $i = 1$ in both axes of the SCI shown. PD-L1 extends over $1 \le i \le 115$. The N-terminal end of PD-1 starts with residue-ID = 25 and extends over residues $116 < i < 240$ towards the right upper corner. Elements of secondary structure are denoted right to the SCI (Zak et al., 2017), with their extensions indicated by horizontal grey shaded bars. Standard deviations D_{ij} [nm], computed according to Eq. 2, for values see color bar. **(B)**: trajectory t_2. **(C)**: trajectory t_3.

Noé, the N_α C_α atoms of the backbone are optimally decomposed into k clusters by minimizing the following target function:

$$q(\mathbf{c}) = \sum_{m=1}^{k}\sum_{i=1}^{N_\alpha}\sum_{j=1}^{N_\alpha} c_{im}c_{jm}D_{ij} = \mathrm{tr}\left(\mathbf{c}^{\mathrm{T}}\mathbf{D}\mathbf{c}\right) \rightarrow \min \qquad (3)$$

In the formulation of Bernhard and Noé, memberships were assumed to be real numbers. This works successfully in the end but affords tremendous computational expense. In our previous work (Kenn et al., 2014) we were able to improve Bernhard's and Noé's method by showing mathematically that the membership coefficients, c_{im}, have in fact to be crisp, i.e., $\{0, 1\}$. Knowing this in advance drastically speeds up the minimization specified in Eq. 3. Since a given atom can only fully belong to one and only one cluster (no fragmentary membership), optimization can draw on single atom moves between clusters. We applied such a fast random search with single atom moves, followed by exhaustive searches to obtain a global optimum. Each lap of clustering was performed 100000 times and the result with the best target function retained. For computational details, parameter studies and thorough evaluations of accuracy and performance we refer to our previous work (Kenn et al., 2016).

As a result, spatial clustering yielded crisp memberships, $c_{i,m}^{(s)} = \{0, 1\}$, for C_α-atom i, in cluster m, within segment s, see Figure 4. Note that $1 \le i \le N_\alpha$, $1 \le m \le k$ and $1 \le s \le N_s$.

2.3.2 Time-wise consensus clustering

To arrive at a consensus we start with defining dissimilarity Δ_{ij} between two C_α -atoms i and j as:

$$\Delta_{ij} = \frac{1}{N_s}\sum_{s=1}^{N_s}\Delta_{ij}^{(s)} = \frac{1}{N_s}\sum_{s=1}^{N_s}\left(1 - \sum_{m=1}^{k} c_{im}^{(s)} \cdot c_{jm}^{(s)}\right) \qquad (4)$$

with $\Delta_{ij}^{(s)} = 0$ if atoms i and j belong to the same cluster $C_m^{(s)}$ in segment s and $\Delta_{ij}^{(s)} = 1$ otherwise. Summing up $\Delta_{ij}^{(s)}$ over all segments (s) yields the number of segments within which i and j are not within the same cluster (Monti et al., 2003). Note that the number of segments is an upper bound, e.g. atoms i and j may reside "not in the same cluster" in 30 segments out of 500. The precise choice of segment length has only minor impact on the results. Shorter segment lengths (e.g., 25 frames per segment) yield a similarity matrix of higher resolution, but entails only minute changes in the final results. Naturally, a minimum length of segments is required to obtain a reliable estimate of variances. Division by the number of segments (N_s) finally renders a

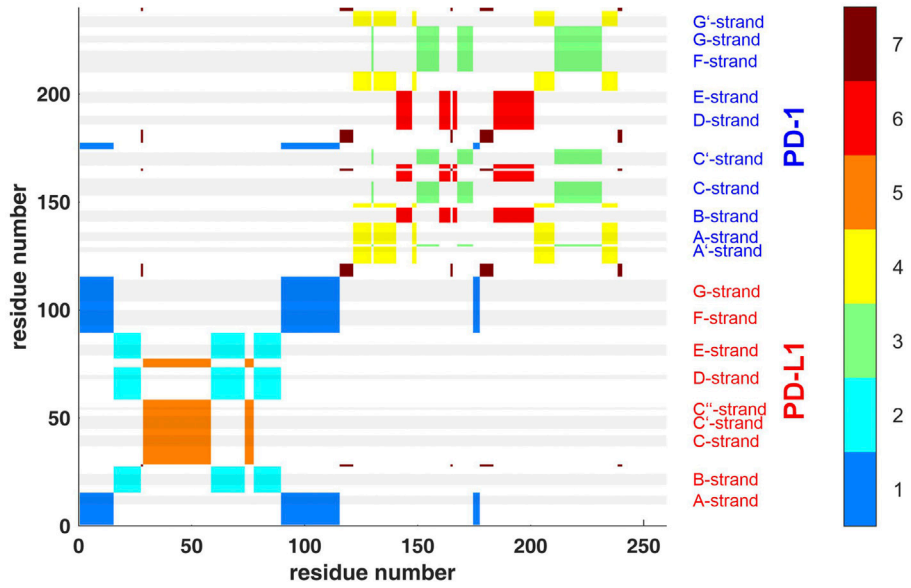


FIGURE 4
Clustering standard deviations of distance variation (STDDV) with k = 7 over the whole trajectory t_1. The best out of 100000 trials in the search for minimum target function is shown. Each cluster (1–7) is shown in a separate color, see the color bar. The sizes of clusters 1 to 7 were 44, 39, 39, 36, 34, 32 and 16. Elements of secondary structure are indicated by grey shaded bars and corresponding labels.

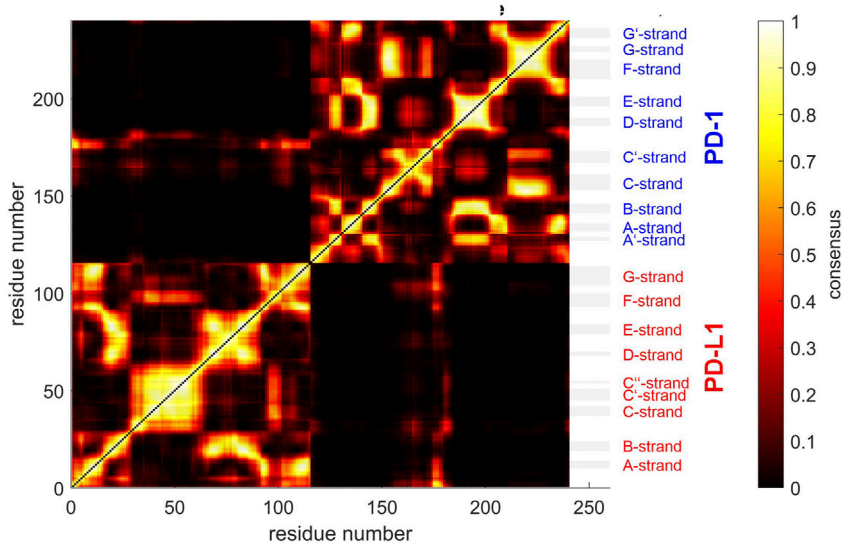


FIGURE 5
Similarity matrix after temporal consensus clustering trajectory t_1, shown as scaled color image (SCI). Complex 4ZQK, consisting of receptor PD-1 and PD-L1 as ligand, showing enhanced similarity within two large areas (receptor and ligand, respectively). Note that numbering starts with ligand PD-L1 with residue-ID = 18 (lower left corner), corresponding to residue number $i = 1$ in both axes of the SCI shown. PD-L1 extends over $1 \le i \le 115$. The N-terminal end of PD-1 starts with residue-ID = 25 and extends over residues $116 < i < 240$ towards the right upper corner. Elements of secondary structure are denoted right to the SCI (Zak et al., 2017), with their extensions indicated by horizontal grey shaded bars. Spatial clusters: 7. Note that the number of spatial clusters influences the similarity matrix and is given as input for computation. Consensus (0–500) indicates in how many (out of 500) timewise segments two C_α atoms belonged to the same spatial cluster (no matter which cluster that was). Consensus shown normalized to 0–1, see Eq. 5 and color bar.

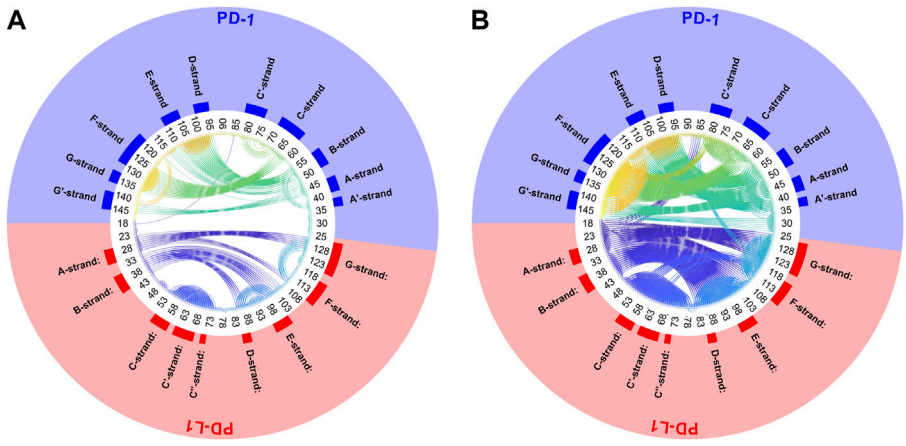


FIGURE 6
Circular plot of small variations of inter-atom distances for PD-1 complexed with PD-L1. Trajectory 2, spatial clusters k = 7. Residues numbered within each chain according to PDB convention. Around the circular plot, elements of secondary structure are indicated. Left: threshold for link to be drawn: Δ_{th} = 0.04 = 20/500. Right: Δ_{th} = 0.1 = 50/500.

normalized measure of dissimilarity between i and j, relating to the whole trajectory (e.g., 30/500 = 0.06). This dissimilarity lends itself as a proxy for "distance" between atoms in this second lap of clustering (consensus clustering). Since cluster memberships are crisp, $c_{im}^{(s)} \in \{0, 1\}$, the concept above can also be expressed more formally (but less intuitively) *via* a product of memberships, see the second part of Eq. 4. Dual to dissimilarity, a *similarity*-matrix can be obtained *via*

$$C_{ij} = 1 - \Delta_{ij} \tag{5}$$

see the example displayed in Figure 5. Note that similarity, as defined above, will be used synonymously with "consensus" in the framework of consensus clustering. Naturally, C_α atoms in close succession along the backbone appear close to the diagonal and show high consensus, see the color bar.

Another very illustrative way to display consensus between atoms is a circular plot, see Figure 6. All C_α-atoms are arranged in a circle and a threshold, Δ_{th}, has to be chosen. Whenever the dissimilarity between two atoms is smaller than the threshold ($\Delta_{ij} \leq \Delta_{th}$), these are connected by a line. Thus, connected atoms show small fluctuation in their distance over time.

2.3.3 Second lap of clustering based on consensus

The dissimilarity matrix Δ_{ij} was then subjected to agglomerative clustering (Ward, 1963; Jain et al., 1999), evaluating two methods for comparison, "average" and "complete" (Mathworks, 2021). They differ in their mode of linkage, i.e., the way, how the distance between two given (intermediate) clusters is computed: Method "average" takes the mean distance between individuals in different clusters to represent the distance between both clusters. Conversely, method "complete" adopts the largest of those between-cluster distances as the distance between the two clusters.

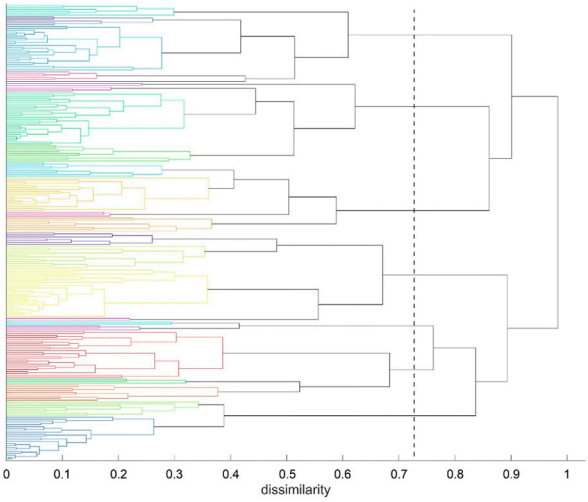


FIGURE 7
Agglomerative clustering according to inter-atomic time-wise consensus. Spatial clustering (k = 7 clusters) within each of 500 time-wise segments. Consensus among these 500 results of clustering was converted into distances and subjected to agglomerative clustering with distance model "average". For reasons of clarity we call the results of agglomerative clustering "groups" in the following—to distinguish from the results of spatial clustering ("clusters"). Agglomerative clustering was terminated at N_G = 24 groups. The dashed line indicates N_G = 7 groups, as an example.

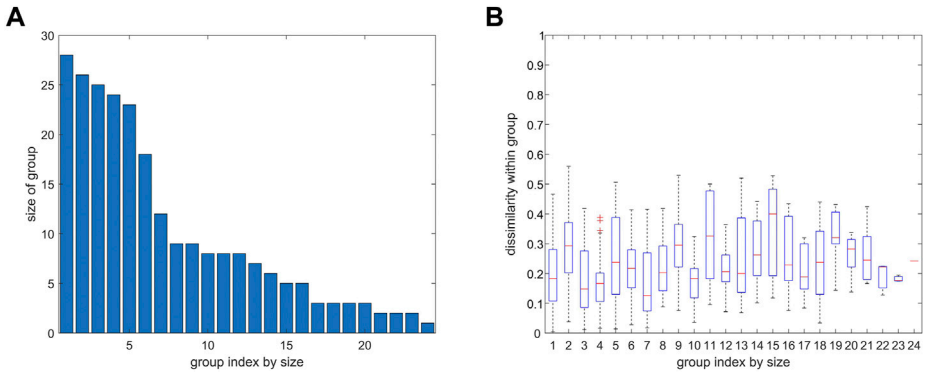


FIGURE 8
Number of atoms and variability of distance variation within groups from agglomerative clustering. 4ZQK, trajectory t_1, parameters k = 7 and N_G = 24, similar to Figure 7. **(A)** Size of group (number of atoms). **(B)** Homogeneity within groups shown by a boxplot of distance variations between pairs of atoms within each group (mean, quantiles, extremes).

The methods "average" and "complete" are both appropriate for Euclidean as well as for non-Euclidean distances, which we worked with, after all. A third method ("single") would also be appropriate for non-Euclidean distances, however it tends to yield a large number of small clusters, what seemed inappropriate for the structure of our molecules. Other methods are restricted to Euclidean distances.

Agglomerative clustering yields a tree-like-structure (dendrogram), an example is shown in Figure 7. At the left vertical axis C_α atoms are arranged and colored according to cluster membership, the residue-index being irrelevant here. The horizontal axis shows dissimilarity, in our case values between 0 (each C_α-atom against itself) and a maximum equal to the number of segments, N_s, into which the trajectory was split (e.g., 500). Note that this maximum applies to the methods chosen in this work but need not apply to other clustering methods, such as "Ward" for example.

Clustering starts at bottom, with each atom representing a cluster of its own (leaves of the tree). Then clustering proceeds upwards (from left to right in Figure 7), in each step joining two clusters, selected among all pairs according to minimum distance. Note that there is no universal definition of "distance" between two clusters but one has to choose among several variants, i.e., "average" or "complete" in this work. Note that "distance" appears on the horizontal axes in Figure 7. As a result, any emerging cluster contains the sum of atoms contained in both of its predecessors. Finally, the algorithm terminates with a cluster containing all atoms, at the root of the tree.

The tree is then retraced from the root towards the leaves (from right towards left in Figure 7), along decreasing dissimilarity. Whenever a bifurcation is crossed, the number of clusters increases, one by one. One may proceed until a preselected number of clusters, N_C, is encountered (e.g. N_C = 7 in Figure 7) and thus obtain a corresponding "cut-point" in terms of dissimilarity, see the dashed line. Quantitatively, the cut-point is computed as the median of those two levels of dissimilarity that have been passed though latest during recovery. In Figure 7, the final cut-point for display was selected at N_C = 24 groups (left, bottom border of tree). This number of clusters was chosen to accommodate several large, compact domains within the molecule (such as beta-sheets) as well as several smaller parts, such as freely moving loops. This intention has been fulfilled as clearly reflected in Figures 4, 8. These clusters represent a partition of all atoms into a given number (N_G) of groups $\{G_1, G_2 ... G_{N_G}\}$, as shown in Figure 9. These groups are shown in different colors.

2.3.4 Estimating the stability of clusters across trajectories

Above we have explained spatial clustering within consecutive segments of a single trajectory and then how to perform agglomerative clustering into domains, based on time-wise stability of these spatial clusters. Resulting clusters were called "semi-rigid". Finally, we evaluate how much clusters differ between independent trajectories of the same molecular system. This comparison yields an estimate of cluster-stability on an upmost level, and was performed as follows.

For a trajectory t, N_G time-wise consensus clusters $\{G_1^{(t)}, G_2^{(t)} ... G_{N_G}^{(t)}\}$ were obtained, with $t = 1, 2, 3$, since three trajectories were generated. Let cluster-memberships of atom i in cluster m within trajectory t be denoted by $G_{im}^{(t)} = \{0, 1\}$, with $1 \le i \le N_\alpha$, $1 \le m \le N_G$ and $1 \le t \le 3$. When comparing results of agglomerative clustering between trajectories, the following problem arises: During agglomeration, emerging labels (identity numbers) of clusters may depend on minute, even somewhat random differences between trajectories. For example, if an existing cluster is to be joined with its "nearest" neighbor cluster, there might be two (or even more) neighbors

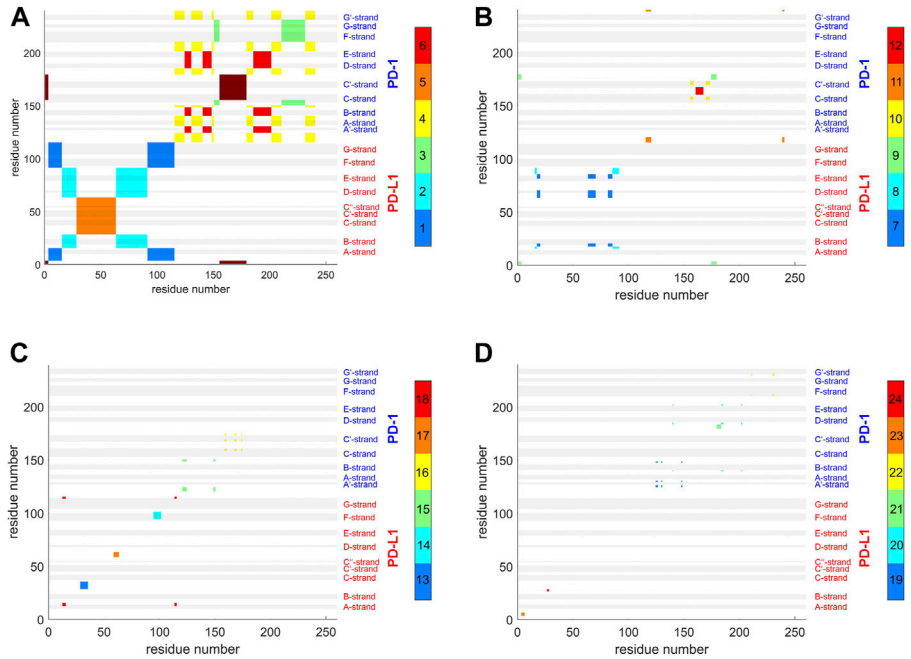


FIGURE 9
Atom groups resulting from agglomerative clustering consensus. 4ZQK, trajectory t_1, parameters k = 7, cutoff N_G = 24 groups. Note that the groups were internally numbered in order of descending size and each cluster is indicated by the color along the color bar to the right. To visually represent as many as 24 groups, 4 panels were generated for groups 1–6 **(A)**, 7–12 **(B)**, 13–18 **(C)** and 19–24 **(D)**. Note also that each cluster does not need to appear as coherent field in the matrix, since remote atoms in the peptide chain may belong to one and the same cluster, as shown in the circular graph, Figure 6. To identify a single cluster, all fields of the same color within one given panel have to be considered together. All in all, the picture reflects the intricate connections of intra-molecular motions. Elements of secondary structure are indicated by grey shaded bars and corresponding labels.

almost equally "near". As a consequence, even minute differences between trajectories in such a case induce different decision paths "which cluster wins", and branch into different joining-operations for each trajectory. Since any new cluster generated (by joining) receives the next available cluster-label in sequence, a certain cluster-label may refer to two physically different groups of atoms in each trajectory. All in all, even though agglomerative clustering may produce nicely compatible physical groups of atoms, the labels of those groups might (and usually will) result totally permuted.

Therefore, after agglomerative clustering two trajectories, a so called "assignment problem" arises (Ramshaw and Tarjan, 2012): How should pairs of corresponding clusters be identified on an algorithmic basis?

In short, we proceeded as follows: We used the "Hungarian Algorithm", drawing on the special target function given in Eq. 7. The value given by this target function represents the metric between trajectories. A vivid display is given in Figure 10, including a description how estimates come about for specific groups of C_α atoms.

In mathematical detail, the following procedure was performed: For each trajectory, NG (e.g., NG = 24) groups are obtained, and out of NG! possible pairings the best matching has to be determined, labels permuted accordingly, and re-assigned. Only on this basis, a comparison—cluster by cluster—is meaningful.

The assignment problem has been mathematically solved (Kuhn, 1955), based on the "Hungarian algorithm", was put in a more general frame by Edmunds and Karp (Edmonds and Karp, 1972), and is now available in the MATLAB routine "matchpairs" (Duff and Koster, 2001): The user has to specify a so called "loss function" which quantifies the "loss" compared to a perfect match between two sets of clusters $\{G_1^{(t_1)}, G_2^{(t_1)} \ldots G_{N_G}^{(t_1)}\}$ and $\{G_1^{(t_2)}, G_2^{(t_2)} \ldots G_{N_G}^{(t_2)}\}$. Note that a comparison is feasible only if both sets contain the same number of clusters, N_G. For example, when evaluating the disparity between a pair of clusters $\{G_i^{(t_1)}, G_j^{(t_2)}\}$, one may use the symmetric difference

$$L_{ij}^{(t_1,t_2)} = \left|\left(G_i^{(t_1)} \cup G_j^{(t_2)}\right) \setminus \left(G_i^{(t_1)} \cap G_j^{(t_2)}\right)\right| \qquad (6)$$

as a proxy for a so called loss function, with $|\ |$ meaning the number of elements in a group (cardinality). If both sets contain the very same atoms, the loss $L = 0$, if they do not share a single atom, the loss $L_{ij}^{(t_1,t_2)} = |G_i^{(t_1)} \cup G_j^{(t_2)}|$, i.e., it equals the total number of atoms in both groups. For intermediate cases, L

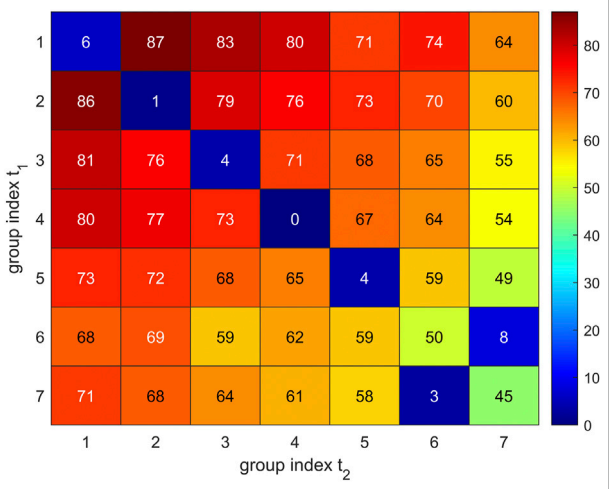


FIGURE 10
Visual representation of group-matching. Comparison of two sets of 7 $\mathbf{C}_\alpha$ groups resulting for trajectory t_2 (horizontal axis) and trajectory t_1 (vertical axis). Group numbers are assigned with decreasing group size (7–1). Values given in matrix elements were evaluated *via* Eq. 6 and represent the loss function $L_{ij}^{(t_1,t_2)}$, i.e., the number of $\mathbf{C}_\alpha$ atoms not contained in both groups. Low losses indicate good matching between groups and are colored blue, see the color bar. Diagonal elements represent a matching according to group size only, e.g., L_{11} = 6 indicates that only 7 $\mathbf{C}_\alpha$ atoms are not members of these both groups (1–1). Elements off the diagonal represent putative losses if group labels were permuted, e.g., L_{12} = 87 indicates that 87 $\mathbf{C}_\alpha$ atoms would mismatch in a putative comparison between group 1 from t_1 and group 2 from t_2. One can see that for groups 1 to 5, the original labelling (according to group size) is already optimum. Conversely, groups 6 and 7 have to be interchanged for optimum match.

represents the number of atoms contained in just one of both sets but not in the other (exclusive or-condition).

Solving the assignment problem allows to re-label clusters in a way that clusters with the same index go in pairs (common index m replaces i, j) and this pairing entails minimum overall loss. For this optimum assignment, losses are added over all clusters to obtain the total clustering disparity between both trajectories:

$$D^{(t_1,t_2)} = \sum_{m=1}^{N_g} \left| \left(\mathrm{G}_m^{(t_1)} \cup \mathrm{G}_m^{(t_2)}\right) \setminus \left(\mathrm{G}_m^{(t_1)} \cap \mathrm{G}_m^{(t_2)}\right) \right| \qquad (7)$$

Note that the solution of the assignment problem is not commutative, i.e. $D^{(t_1,t_2)} \neq D^{(t_2,t_1)}$, i.e. it makes a difference in results which trajectory comes first. We shall call it "reference" in the following, e.g. trajectories t_2 and t_3 may be mapped on reference t_1.

Optimized re-assignment and joint labelling of clusters allows to boil down each cluster to its "stable kernel", K_m, made up by those atoms belonging to the "same" cluster in all three trajectories considered:

$$K_m^{(t_1,t_2,t_3)} = \mathrm{G}_m^{(t_1)} \cap \mathrm{G}_m^{(t_2)} \cap \mathrm{G}_m^{(t_3)},\ m = 1, \ldots, N_\mathrm{G} \qquad (8)$$

Such kernels may be displayed within 3D representations of the molecular complexes.

2.3.5 Relating groups to molecular structures

For each atom i, its kernel-membership k_i is known, with $1 \le k_i \le N_\mathrm{G}$. This allows for visualization of such groups within 3D representations of the molecular complex. From the memberships we generated Tcl-commands (Welch et al., 2003) to color these groups in VMD (Humphrey et al., 1996), see also the figures shown in the results section.

3 Results

3.1 Results for whole trajectories

Applying the methods explained above we obtained results for the complex 4ZQK (PD-1 + PD-L1). First, standard deviations D_{ij} of pair-distances were computed over each whole trajectory, with $N_\mathrm{f} = N_\mathrm{t}$ in Eq. 2. Figure 3 shows the result for trajectories t_1, t_2 and t_3. Considerable differences between trajectories t_1, t_2 and t_3 can be seen.

Second, spatial clustering was performed over whole trajectories, see an example in Figure 4 for t_1 and k = 7. Note that clustering in any case assigns each atom to one of the clusters, even if its STDDV to quite many other atoms are large, see the conspicuous stripes in shiny yellow in Figure 3. As a consequence, clusters obtained this way inevitably also house atoms not intended to be parts of semi-rigid domains.

3.2 Results for segmental clustering

Next, time-wise clustering was performed. Figure 5 shows the similarity matrix with values between 0 and N_s, indicating how often time-wise consensus clustering found two C_α atoms within the same cluster. Note that clusters are neither numbered nor labelled in this step, i.e., they do not have unique identifiers related to their "inhabitants" in terms of physical atoms. For example, the pair of $C_{\alpha,128}$ and $C_{\alpha,237}$ may be together in cluster 4 in time-wise segment 129 and together in cluster 5 in time-wise segment 237. This would yield a consistency count of 2 (out of 500). Naturally, the number of segments, N_s, poses an upper limit of consistency, expressing that these two atoms were in the same cluster in all segments.

As consensus relates to linked mobility, most strong linkages were seen within each molecule (chain) of the complex, i.e., within PD-1 and within PD-L1. This resembles the fact that beta-strands cooperatively fold into beta-sheets, and corresponding atoms move in a more concerted way. However, some weaker linkage is also present between both molecules, see the parts in orange for residues of PD-1 towards multiple parts of PD-L1: these regions show consensus. A few C_α

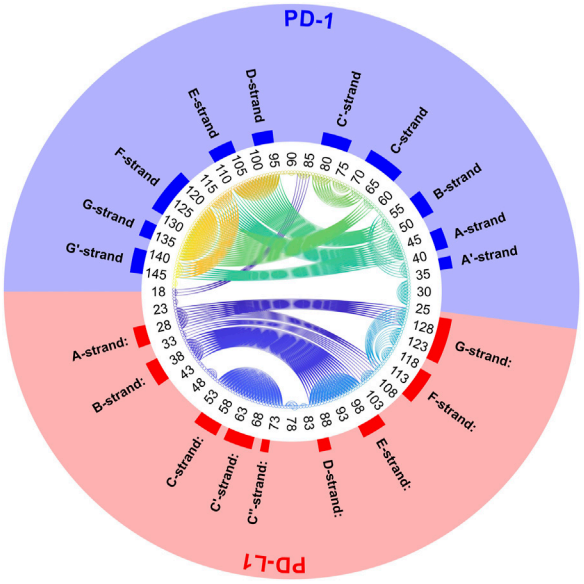


FIGURE 11
Circular graph of semi-rigid domains from agglomerative clustering time-wise consensus. 4ZQK, trajectory t_1, parameters k = 7, cutoff N_G = 24 groups. Connective lines colored according to C_α indices.

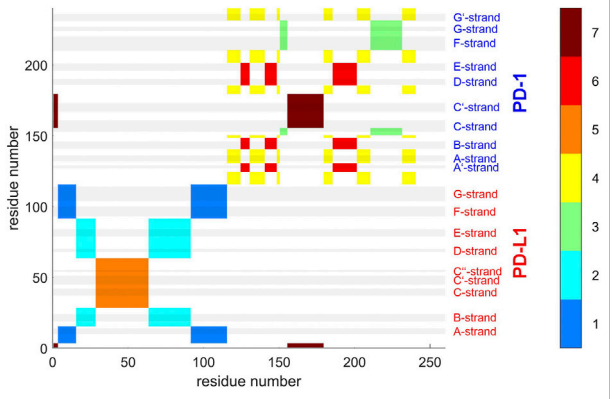


FIGURE 12
Atom groups for agglomerative clustering with cutoff N_G = 7 groups. Complex 4ZQK, trajectory t_1, parameters k = 7, cutoff N_G = 7 groups. For easy comparison with Figure 4, each cluster is shown in a separate color, see color bar. Colors were selected to match those of Figure 4, for direct comparison. Elements of secondary structure are indicated by grey shaded bars and corresponding labels.

atoms at the start (i.e., the N-terminal loop) of the ligand even show close relation to this region of PD-1, with consensus around 0.8 (appearing yellow).

Posing a threshold on dissimilarity, e.g., $\Delta_{ij} \leq \Delta_{th}$, a circular plot can be obtained, see Figure 6. Pairs of atoms are connected by lines to indicate consensus if they appear in different clusters in a fraction of segments smaller than Δ_{th}. For example, selecting Δ_{th} = 0.06 = 30/500, connects atoms only if they end up in different clusters in no more than 30 (time-wise) segments, out of 500. Naturally, the larger the dissimilarity threshold, Δ_{th}, is chosen, the more connection lines populate the circular plot. Moreover, weak similarities, such as those between PD-1 and PD-L1, become visible only if large dissimilarities are tolerated (right panel of Figure 6). They faint away in quite a large percentage of frames.

The above results display concordance (i.e., similarity in movements) between atoms, as it results directly from time-wise consensus clustering, based on pairs of C_α atoms. These pairwise results (consensus matrix) were subjected to a further step of analysis, agglomerative clustering, see Figure 7. Note that choosing a certain number of clusters, e.g. N_G = 24, does not change anything of the algorithm, it just defines the level of cutoff through the tree where splitting into groups is considered as result. Note that dissimilarities between clusters may well exceed the upper limit of dissimilarities Δ_{ij} between single atoms.

Clusters resulting from agglomerative clustering are different in size (number of atoms), see Figure 8. The box plot indicates variability within groups, based in the standard deviations of inter-atom distances used as the key target for spatial clustering. Groups from agglomerative clustering may also be displayed in matrix form, see Figure 9. Like in Figure 5, atoms are numbered consecutively, as they occur in the 4ZQK complex in PDB. Elements of secondary structure have been annotated to hint at possible relations to atomic mobility. In addition, these groups were visualized in circular graphs, see Figure 11.

Agglomerative clustering starts with each atom representing its own cluster and then joins existing clusters. By proceedings upwards level by level, it creates a tree of larger and larger clusters, ending up in one maximum cluster above all others. This tree may be pruned at any level to yield different numbers of clusters. For comparison with clustering STDDV matrices according to Bernhard (Bernhard and Noé, 2010), see Figure 4, we display the agglomerative result pruned at N_G = 7, see Figure 12. Note that colors have been selected to match those of Figure 4, in order to be directly comparable.

3.3 Stability of clusters across trajectories

Note that all visualizations shown so far pertained to one single trajectory and a given set of clustering-parameters (k, N_G). It is interesting, however, to evaluate differences in results between trajectories. To these ends we utilized the disparity $D^{(t_1,t_2)}$ between pairs of trajectories, defined in Eq. 7. Discrepancies in agglomerative clustering between pairs of trajectories were 22, 40 and 28 for (t_1, t_2), (t_1, t_3) and (t_2, t_3),

TABLE 3 Disparity in groups between trajectories. All results refer to $k = 7$ clusters for Bernhard clustering. Agglomerative clustering was performed for 7 and 24 groups, respectively. For each comparison between trajectories, discrepancies in agglomerative clustering are given as numbers of residues within different groups together with corresponding percentages of all residues (240). Note that, for a comparison between three trajectories (right part of table), disparities evaluated according to Eq. 7 depend on the choice of the reference trajectory, listed in position 1—as a coincidence, these results are all equal (44 and 68). Comparing 3 trajectories, means to include differences between 3 pairs of trajectories: For example, an atom counts as disparity if it resides in different clusters for (t_1, t_2) even if it resides in corresponding clusters in (t_1, t_3) and (t_2, t_3). As a consequence, disparities between triples of trajectories appear larger than those between pairs.

Groups	Trajectory comparison					
	(t_1, t_2)	(t_1, t_3)	(t_2, t_3)	(t_1, t_2, t_3)	(t_2, t_1, t_3)	(t_3, t_1, t_2)
7	40 (16.7%)	28 (11.7%)	22 (9.2%)	44 (18.3%)	44 (18.3%)	44 (18.3%)
24	52 (21.7%)	44 (18.3%)	41 (17.1%)	68 (28.3%)	68 (28.3%)	68 (28.3%)

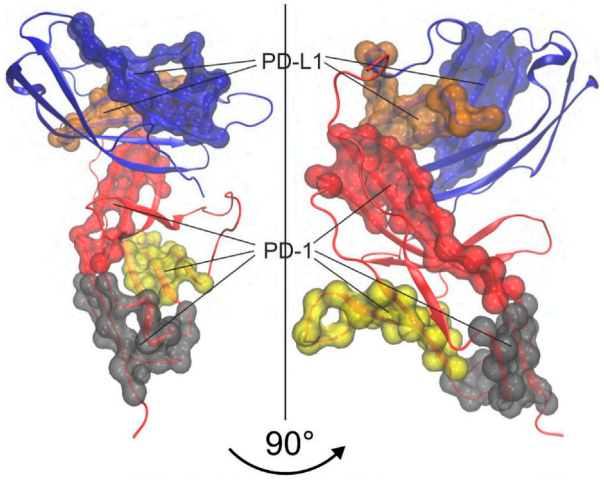


FIGURE 13
3D-Visualization of semi-rigid domains resulting from spatio-temporal consensus clustering and consensus over trajectories. Groups obtained from agglomerative clustering were consolidated over 3 trajectories to obtain "kernels", representing semi-rigid domains. The 5 largest kernels are colored and shown as surface representations. Note that kernels in red, grey and yellow belong to PD-1 while kernels in blue and ochre belong to PD-L1.

respectively (see Table 3). Comparing three trajectories naturally leads larger discrepancies. For comparison, we also added the results for agglomerative clustering in 7 groups, concordant with the preceding Bernhard-clustering. Note that considering more groups increases the chance for residues to switch between groups, and concomitantly discrepancy increases, however, for generating the final consensus, the more discrepant smaller clusters were disregarded, see below.

Finally, however, we also created a consensus between trajectories by estimating "kernels" of atoms belonging to the same cluster in all three trajectories, see Eq. 8. Note that the labelling of agglomerative groups originally varies randomly between trajectories and has to be consolidated as described in the methods section. Such a consolidated numbering—and a corresponding coloring—was used to outline the kernels within a 3D model of the molecular complex, see Figure 13. These kernels are considered as the "semi-rigid domains" aimed at.

4 Discussion

We applied the method of spatio-temporal clustering to the PD-1/PD-L1 complex, aiming at identifying semi-rigid domains within these molecules. Such domains are considered a highly important basis for coming computational research since any detection of minute movement patterns requires to fit molecular configurations to stable kernels. Minute and interesting movement patterns, e.g., of active loops, may then be characterized with reference to such kernels.

During the course of an MD-simulation, larger portions ("domains") of a molecule might collectively move slowly but move broadly back and forth in amplitude. Inside such a domain, however, single amino acids and even more single atoms oscillate at much higher frequencies. The goal is to separate these two types of movement occurring on different spatial and time scales: semi-rigid domains as a whole should go along with the larger but slower movements, while housing those many tiny oscillations of their "inhabitant" atoms. As a result, a single atom performs both motions in superposition—small oscillations at high frequency, superimposed on larger and much slower collective motions of its corresponding domain. Both types of motion in combination influence the distances to its neighbor atoms.

In a non-supervised approach, one can only draw on the variation of distances as such, without knowing their origin (tiny oscillations of single atoms or large-scale movement of whole domains). Clustering atoms with respect to variations of pair-distances will therefore yield different results (clusters), when performed on different (time-wise) parts of a trajectory. Finally, however, a smart clustering algorithm should yield larger clusters "moving" in accordance with those larger domains, each of these holding much the same groups of atoms as inhabitants (members) over time.

For a start, we computed the standard deviations of distance variations (STDDV) matrices of whole trajectories (Figure 3). Since these matrices did not reveal prominent structures which

could be clustered right away, we adopted a refined, three-step procedure. Moreover, distinct differences between trajectories became apparent in these matrices. We have to conclude that the system obviously inhabited different configurational sub-spaces in each trajectory, and simulation time has to be extended in coming studies to closer approach ergodicity and visit all portions of configuration space appropriately.

In the present work, matrices with different properties were studied. The STDDV-matrix does not reflect distances as such but rather variations in distance and therefore in general will not fulfil the triangle inequality. Incidentally, the Bernhard algorithm does not require fulfillment of the triangle inequality. In the second step of our algorithm we computed the dissimilarity matrix, Eq. 4, which fulfills the triangle inequality. This was a main reason for us to adopt this multi-step procedure.

To refine clustering we adopted a three-step procedure: First, clustering according to distance variation, but separately over short segments of the trajectory. Second, these results were consolidated over all segments of the trajectory by characterizing consensus for each pair of atoms: the percentage of time-wise segments in which these two atoms shared (resided in) the same cluster. Note that this second step yielded but pairwise information (consensus matrix), visualized in various forms (Figures 5, 6). Third, we performed agglomerative clustering to derive domain-like regions of coherence, the final result, shown in Figure 13. Note that cluster memberships after agglomerative clustering are in general different from those obtained by spatial clustering in the first lap.

The most intuitive approach would have been agglomerative hierarchical clustering, (Kaufman and Rousseeuw, 1990; Teukolsky et al., 2007). In a preliminary examination of the STDDV matrices (Figure 3) we found that an important precondition of agglomerative clustering is only poorly satisfied by MD data: Atoms may switch between clusters quite freely, without severely changing the target-function (minimum distance variability within clusters). This may easily deteriorate agglomerative clustering, and therefore we refrained from it as a first step. However, in future studies it would be interesting to mend this drawback, possibly by selecting more sophisticated models for linkage between clusters (others than “average” or “complete”). Also, agglomerative clustering allows to optimize the cut-off (i.e. the number of groups, N_G) according to formal criterions such as consistency (Mathworks, 2021). Linkage and cut-off could be systematically evaluated and optimized.

The achievement of the present work is the unsupervised consolidation of quite large domains within the molecular complex, despite considerable movements of its member atoms. Results were additionally consolidated by repeating the entire analysis for three independent trajectories and considering the overlap between these three replicates of a cluster as the final, reliable rigid domain. Based on these semi-rigid domains, subtle movements of active regions may be evaluated in future studies, scrutinizing the molecular basis of receptor activation and action of drugs, including checkpoint blockers in oncology.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://www.rcsb.org/structure/4ZQK, https://www.rcsb.org/structure/5GGS.

Author contributions

Conceptualization, WS, MK, HK, and GP; methodology, MK and RK; MD simulations, LT and RK; software, MK and MC; formal analysis and investigation, MK and WS; resources, HK; writing—original draft preparation, WS; writing—review and editing, WS, RK, and MK; visualization, LT, MC, and MK; supervision, WS. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

A major part of the molecular dynamics computations for this work was performed at the Vienna Scientific Cluster (VSC).

Conflict of interest

GP received grants and honoraria from Pfizer, Roche, Novartis, MSD, Seagen, Daiichi, UCB, Amgen, AstraZeneca, and Gilead.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2022.838129/full#supplementary-material

References

Berendsen, H. J., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81 (8), 3684–3690. doi:10.1063/1.448118

Bernhard, S., and Noé, F. (2010). Optimal identification of semi-rigid domains in macromolecules from molecular dynamics simulation. *PLoS One* 5 (5), e10491. doi:10.1371/journal.pone.0010491

Brahmer, J. R., Lacchetti, C., Schneider, B. J., Atkins, M. B., Brassil, K. J., Caterino, J. M., et al. (2018). Management of immune-related adverse events in patients treated with immune checkpoint inhibitor therapy: American society of clinical oncology clinical practice guideline. *Jco* 36 (17), 1714–1768. doi:10.1200/jco.2017.77.6385

Brahmer, J. R., Tykodi, S. S., Chow, L. Q., Hwu, W. J., Topalian, S. L., Hwu, P., et al. (2012). Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N. Engl. J. Med.* 366 (26), 2455–2465. doi:10.1056/NEJMoa1200694

Burley, S. K. (2013). PDB40: The protein data bank celebrates its 40th birthday. *Biopolymers* 99 (3), 165–169. doi:10.1002/bip.22182

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126 (1), 014101. doi:10.1063/1.2408420

Chen, D. S., and Mellman, I. (2013). Oncology meets immunology: The cancer-immunity cycle. *Immunity* 39 (1), 1–10. doi:10.1016/j.immuni.2013.07.012

Cortés, J., Lipatov, O., Im, S. A., Gonçalves, A., Lee, K. S., Schmid, P., et al. (2019). KEYNOTE-119: Phase III study of pembrolizumab (pembro) versus single-agent chemotherapy (chemo) for metastatic triple negative breast cancer (mTNBC). *Ann. Oncol.* 30, v859–v860. doi:10.1093/annonc/mdz394.010

Cross, S., Kuttel, M. M., Stone, J. E., and Gain, J. E. (2009). Visualisation of cyclic and multi-branched molecules with VMD. *J. Mol. Graph. Model.* 28, 131–139. doi:10.1016/j.jmgm.2009.04.010

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98 (12), 10089–10092. doi:10.1063/1.464397

Dong, H., Strome, S. E., Salomao, D. R., Tamura, H., Hirano, F., Flies, D. B., et al. (2002). Tumor-associated B7-H1 promotes T-cell apoptosis: A potential mechanism of immune evasion. *Nat. Med.* 8 (8), 793–800. doi:10.1038/nm730

Duff, I. S., and Koster, J. (2001). On algorithms for permuting large entries to the diagonal of a sparse matrix. *SIAM J. Matrix Anal. Appl.* 22 (4), 973–996. doi:10.1137/s0895479899358443

Edmonds, J., and Karp, R. M. (1972). Theoretical improvements in algorithmic efficiency for network flow problems. *J. ACM* 19 (2), 248–264. doi:10.1145/321694.321699

Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S., and Onufriev, A. (2005). H++: A server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* 33 (2), W368–W371. doi:10.1093/nar/gki464

Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4 (3), 435–447. doi:10.1021/ct700301q

Hess, B. (2008). P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* 4 (1), 116–122. doi:10.1021/ct700200b

Hsin, J., Arkhipov, A., Yin, Y., Stone, J. E., and Schulten, K. (2008). Using VMD: An introductory tutorial. *Curr. Protoc. Bioinforma.* 24, 51–55. doi:10.1002/0471250953.bi0507s24

Humphrey, W., Dalke, A., and Schulten, K. (1996). Vmd: Visual molecular dynamics. *J. Mol. Graph.* 14 (1), 3327–3838. doi:10.1016/0263-7855(96)00018-5

Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data clustering. *ACM Comput. Surv.* 31 (3), 264–323. doi:10.1145/331499.331504

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79 (2), 926–935. doi:10.1063/1.445869

Kaufman, L., and Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster Analysis*. New York: John Wiley.

Kenn, M., Ribarics, R., Ilieva, N., Cibena, M., Karch, R., and Schreiner, W. (2016). Spatiotemporal multistage consensus clustering in molecular dynamics studies of large proteins. *Mol. Biosyst.* 12 (5), 1600–1614. doi:10.1039/c5mb00879d

Kenn, M., Ribarics, R., Ilieva, N., and Schreiner, W. (2014). Finding semirigid domains in biomolecules by clustering pair-distance variations. *BioMed Res. Int.* 2014, 1–13. doi:10.1155/2014/731325

Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Nav. Res. Logist.* 2 (1-2), 83–97. doi:10.1002/nav.3800020109

Kundapura, S. V., and Ramagopal, U. A. (2019). The CC′ loop of IgV domains of the immune checkpoint receptors, plays a key role in receptor:ligand affinity modulation. *Sci. Rep.* 9 (1), 19191. doi:10.1038/s41598-019-54623-y

Kwa, M. J., and Adams, S. (2018). Checkpoint inhibitors in triple-negative breast cancer (TNBC): Where to go from here. *Cancer* 124 (10), 2086–2103. doi:10.1002/cncr.31272

Lee, J. Y., Lee, H. T., Shin, W., Chae, J., Choi, J., Kim, S. H., et al. (2016). Structural basis of checkpoint blockade by monoclonal antibodies in cancer immunotherapy. *Nat. Commun.* 7, 13354–13364. doi:10.1038/ncomms13354

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., et al. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78 (8), 1950–1958. doi:10.1002/prot.22711

Liu, W., Huang, B., Kuang, Y., and Liu, G. (2017). Molecular dynamics simulations elucidate conformational selection and induced fit mechanisms in the binding of PD-1 and PD-L1. *Mol. Biosyst.* 13 (1742-2051), 892–900. (Electronic)). doi:10.1039/c7mb00036g

Mathworks (2021). *MATLAB function: Linkage*.

Monti, S., Tamayo, P., Mesirov, J., and Golub, T. (2003). Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Mach. Learn.* 52 (1-2), 91–118. doi:10.1023/A:1023949509487

Parrinello, M., and Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* 52 (12), 7182–7190. doi:10.1063/1.328693

Ramshaw, L., and Tarjan, R. E. (2012). *On minimum-cost assignments in unbalanced bipartite graphs*. HP Laboratories.

Roither, B., Oostenbrink, C., Pfeiler, G., Koelbl, H., and Schreiner, W. (2021). Pembrolizumab induces an unexpected conformational change in the CC′-loop of PD-1. *Cancers* 13 (1), 5. doi:10.3390/cancers13010005

Roither, B., Oostenbrink, C., and Schreiner, W. (2020). Molecular dynamics of the immune checkpoint Programmed Cell Death Protein I, PD-1: Conformational changes of the BC-loop upon binding of the ligand PD-L1 and the monoclonal antibody nivolumab. *BMC Bioinformatics* 21 (17). doi:10.1186/s12859-020-03904-9

Schmid, P., Cortes, J., Bergh, J. C. S., Pusztai, L., Denkert, C., Verma, S., et al. (2018). KEYNOTE-522: Phase III study of pembrolizumab (pembro) + chemotherapy (chemo) vs placebo + chemo as neoadjuvant therapy followed by pembro vs placebo as adjuvant therapy for triple-negative breast cancer (TNBC). *Jco* 36 (15), TPS602. doi:10.1200/JCO.2018.36.15_suppl.TPS602

Smith-Garvin, J. E., Koretzky, G. A., and Jordan, M. S. (2009). T cell activation. *Annu. Rev. Immunol.* 27, 591–619. doi:10.1146/annurev.immunol.021908.132706

Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). "Section 16.4. Hierarchical clustering by phylogenetic trees," in *Numerical recipes: The art of scientific computing*. 3 ed (New York: Cambrige University Press), 701–744.

Tomasiak, L., Karch, R., and Schreiner, W. (2020). "Long-term molecular dynamics simulations reveal flexibility properties of a free and TCR-bound pMHC-I system," in *2020 IEEE international conference on bioinformatics and biomedicine* (BIBM), 1295–1302. doi:10.1109/bibm49941.2020.9313545

Tomasiak, L., Karch, R., and Schreiner, W. (2021). "The monoclonal antibody pembrolizumab alters dynamics of the human programmed cell death receptor 1 (PD-1)," in *2021 IEEE international conference on bioinformatics and biomedicine* (BIBM), 3315–3321. doi:10.1109/bibm52615.2021.9669720

Verlet, L. (1967). Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* 159 (1), 98–103. doi:10.1103/PhysRev.159.98

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58 (301), 236–244. doi:10.1080/01621459.1963.10500845

Welch, B. B., Jones, K., and Hobbs, J. (2003). *Practical programming in tcl/tk*. Upper Saddle River, NJ: Prentice Hall Professional.

Zak, K. M., Grudnik, P., Magiera, K., Dömling, A., Dubin, G., and Holak, T. A. (2017). Structural biology of the immune checkpoint receptor PD-1 and its ligands PD-L1/PD-L2. *Structure* 25 (8), 1163–1174. doi:10.1016/j.str.2017.06.011

Zak, K. M., Kitel, R., Przetocka, S., Golik, P., Guzik, K., Musielak, B., et al. (2015). Structure of the complex of human programmed death 1, PD-1, and its ligand PD-L1. *Structure* 23 (12), 2341–2348. doi:10.1016/j.str.2015.09.010

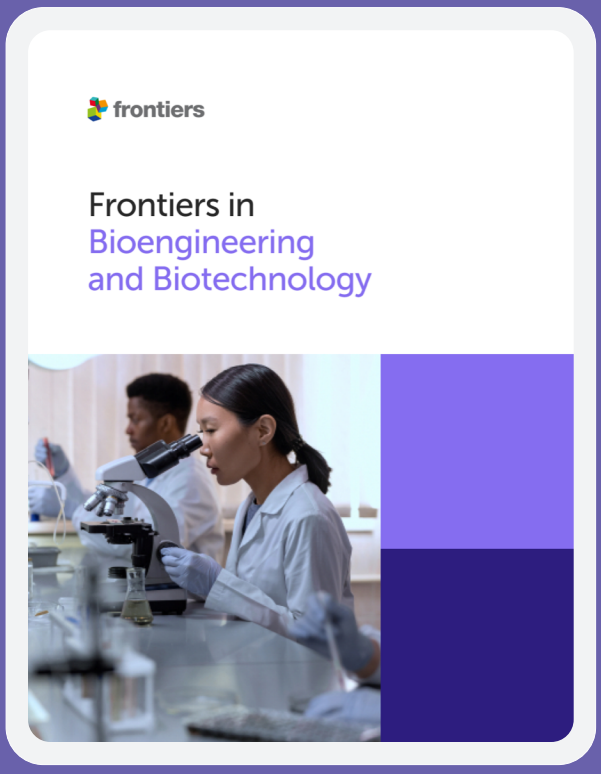
Frontiers in
Bioengineering and Biotechnology

Accelerates the development of therapies, devices, and technologies to improve our lives

A multidisciplinary journal that accelerates the development of biological therapies, devices, processes and technologies to improve our lives by bridging the gap between discoveries and their application.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact