frontiers | Research Topics

Advanced interpretable machine learning methods for clinical NGS big data of complex hereditary diseases – volume II

Edited by
Yudong Cai, Peilin Jia and Tao Huang

Published in
Frontiers in Genetics

ISSN 1664-8714
ISBN 978-2-83251-446-7
DOI 10.3389/978-2-83251-446-7

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

Advanced interpretable machine learning methods for clinical NGS big data of complex hereditary diseases – volume II

Topic editors

Yudong Cai — School of Life Sciences, Shanghai University, China

Peilin Jia — Beijing Institute of Genomics, China

Tao Huang — Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences (CAS), China

Citation

Cai, Y., Jia, P., Huang, T., eds. (2023). *Advanced interpretable machine learning methods for clinical NGS big data of complex hereditary diseases – volume II*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-446-7

Table of contents

frontiers in Genetics

ORIGINAL RESEARCH
published: 03 January 2022
doi: 10.3389/fgene.2021.814798

Identification of Survival-Associated Hub Genes in Pancreatic Adenocarcinoma Based on WGCNA

Liya Huang[1], Ting Ye[1], Jingjing Wang[1], Xiaojing Gu[1], Ruiting Ma[1], Lulu Sheng[2] and Binwu Ma[3]**

[1]Department of Gastroenterology, The General Hospital of NingXia Medical University, Yinchuan, China, [2]Department of Emergency Medicine, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai, China, [3]Department of Neurology, The General Hospital of NingXia Medical University, Yinchuan, China

Pancreatic adenocarcinoma is one of the leading causes of cancer-related death worldwide. Since little clinical symptoms were shown in the early period of pancreatic adenocarcinoma, most patients were found to carry metastases when diagnosis. The lack of effective diagnosis biomarkers and therapeutic targets makes pancreatic adenocarcinoma difficult to screen and cure. The fundamental problem is we know very little about the regulatory mechanisms during carcinogenesis. Here, we employed weighted gene co-expression network analysis (WGCNA) to build gene interaction network using expression profile of pancreatic adenocarcinoma from The Cancer Genome Atlas (TCGA). STRING was used for the construction and visualization of biological networks. A total of 22 modules were detected in the network, among which yellow and pink modules showed the most significant associations with pancreatic adenocarcinoma. Dozens of new genes including PKMYT1, WDHD1, ASF1B, and RAD18 were identified. Further survival analysis yielded their valuable effects on the diagnosis and treatment of pancreatic adenocarcinoma. Our study pioneered network-based algorithm in the application of tumor etiology and discovered several promising regulators for pancreatic adenocarcinoma detection and therapy.

Keywords: WGCNA, gene module, network construction, functional regulator, pancreatic adenocarcinoma

OPEN ACCESS

Edited by:
Tao Huang,
Shanghai Institute of Nutrition and Health (CAS), China

Reviewed by:
Meng Wang,
University of California, San Francisco, United States
Zhihao Xing,
Shenzhen Children's Hospital, China

****Correspondence:***
Lulu Sheng
yl20111211856@sjtu.edu.cn
Binwu Ma
nxmbw@126.com

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *14 November 2021*
Accepted: *06 December 2021*
Published: *03 January 2022*

Citation:
Huang L, Ye T, Wang J, Gu X, Ma R, Sheng L and Ma B (2022) Identification of Survival-Associated Hub Genes in Pancreatic Adenocarcinoma Based on WGCNA. Front. Genet. 12:814798. doi: 10.3389/fgene.2021.814798

INTRODUCTION

Pancreatic cancer ranks the 7th leading cause of cancer mortality worldwide with increasing incidence and poor outcomes (McGuigan et al., 2018). According to the SEER registry, 60,430 new cases and 48,220 deaths of pancreatic cancer have been estimated in the United States in 2021 (Siegel et al., 2021). And pancreatic cancer is predicted to rise from being the 4th to the 2nd most common cause of cancer-related death in the United States by 2030 (Rahib et al., 2014). Pancreatic adenocarcinoma and its variant are the most prevalent subtype of pancreatic cancer and attributed to approximately 90% of all cases (Feldmann et al., 2007). Pancreatic intraepithelial neoplasia, intraductal papillary mucinous neoplasms and mucinous cystic neoplasms are the best characterized precursors of this cancer (Esposito et al., 2014). Due to the insidious nature of pancreatic adenocarcinoma, most patients have already carried metastases such as node upon diagnosis, resulting in 5-years relative survival of about 10% (Luchini et al., 2016; Siegel et al., 2021). Although great efforts including surgical resection, adjuvant chemotherapy and serum biomarker CA19-9 have been made in early detection and treatment of pancreatic adenocarcinoma, medical limitations still exist because of low sensitivity

and high expenses. Recently, several driver genes including KRAS, CDKN2A, TP53 and SMAD4 have been identified from an evolutionary perspective (Vogelstein and Kinzler, 2015; Makohon-Moore and Iacobuzio-Donahue, 2016). Another study has explored novel immune-related gene signatures in pancreatic adenocarcinoma (Chen B. et al., 2021). However, the above evidence is far from adequate to provide therapeutic targets as multiple initiating events embedded with genes were undiscovered in pancreatic adenocarcinoma. Hence, it is imperative to identify new risk genes and their regulatory network in order to elucidate pancreas carcinogenetic mechanisms as well as guide researchers to develop new therapeutic strategies.

As a systematic method used in oncology research that aims at finding co-expressed genes through calculating gene connectivity, weighted gene co-expression network analysis (WGCNA) can analyze the relationship between modules and specific traits followed by clustering genes and forming modules (Langfelder and Horvath, 2008; Chang et al., 2013; Yang et al., 2018). It is widely used in exploring functionality of the whole transcriptome for its particularly powerful computing capability (Zhou et al., 2018). In this study, we employed WGCNA to build a gene interaction network using the expression profile of pancreatic adenocarcinoma from The Cancer Genome Atlas (TCGA). A total of 22 modules were detected in the network, among which yellow and pink modules showed the most significant associations with pancreatic adenocarcinoma. Dozens of new genes including PKMYT1, WDHD1, ASF1B, and RAD18 were identified. Further survival analysis yielded their valuable effects on the progression of pancreatic adenocarcinoma.

METHODS

Pancreatic Adenocarcinoma RNA-Sequencing Datasets

The RNA-sequencing data of 175 pancreatic adenocarcinoma patients was downloaded from the TCGA database (https://portal.gdc.cancer.gov/). As previously described, the gene expression levels were quantified as FPKM (fragments per kilobase per million mapped reads) using TopHat and HTSeq-count (Kim et al., 2013; Anders et al., 2015). The TCGA sample information was listed in **Supplementary Material S1**. Within the 175 tumor samples, 2 samples did not have stage information, 7 samples were in stage I, 24 samples were in stage II, 139 samples were in stage III, and 3 samples were in stage IV. The sample distribution also stressed the importance of identification of new genes for diagnosis and therapy.

Co-Expression Network Construction

R package WGCNA was used for hub genes screening and co-expression of gene pair detection. Elements in the gene co-expression matrix were the weighted values of correlation coefficient between gene pairs. The soft-thresholding function was applied to calculate the power parameter. Dynamic tree cut method was utilized to identify co-expressed gene modules and a dendrogram of genes was produced via a hierarchical clustering approach based on dissimilarity of the unsigned topological overlap matrix (TOM). Finally, genes with similar expression profiles were grouped into network modules.

Enrichment Analysis of Module Genes

R package clusterProfiler was used to perform functional enrichment analysis on clustered genes in pink and yellow modules. A hypergeometric distribution test was applied to detect enrichment terms, and p values were adjusted by false discovery rate (FDR) method with a cutoff FDR <0.05 (Yu et al., 2012).

Visualization of Gene Networks

Construction and analysis of networks were carried out using STRING (11.0) (Szklarczyk et al., 2019).

The Survival Analysis of Hub Genes

There were 175 pancreatic adenocarcinoma patients with the overall, disease free survival time (months) and the survival status. We performed survival analysis using the Cox proportional hazard regression model on these samples (Andersen and Gill, 1982). For each gene, the patients were divided into two groups: the patients with expression levels smaller than the median and the patients with expression levels greater than or equal to the median. The Kaplan-Meier plot was used to describe the survival curves of these two groups of patients. The significance of the survival difference between these two patient groups was evaluated by the log-rank test p value. If the p value was less than 0.05, its survival was considered as significantly different. The R package survival (https://CRAN.R-project.org/package=survival) was used to perform the survival analysis.

RESULTS AND DISCUSSION

Construction of Co-Expression Network

We used the Pearson's correlation coefficient to cluster the samples in TCGA. After removing outliers, we drew a sample clustering tree (**Figure 1A**). The weighted gene co-expression network was constructed from 60,483 genes through WGCNA approach. Here, soft-thresholding power was set to be twelve to satisfy scale-free topology of the network (**Figure 1B**), in which R^2 was used to check how well the network fit the scale freeness. And we detected 22 modules in this network, whose relationship was shown in a cluster dendrogram (**Figure 1C**). The number of members in different modules varied widely. Besides the grey module comprised of many un-classified members, turquoise module contained the maximum 1,508 genes, while the minimum 36 genes were included in darkred module.

Each module represented a group of genes with similar expression profiles across samples. Next, we quantified module-trait associations (**Supplementary Material S2**), among which the pink and yellow modules showed the most significant associations with pancreatic adenocarcinoma. The corresponding correlation coefficients of pink and yellow

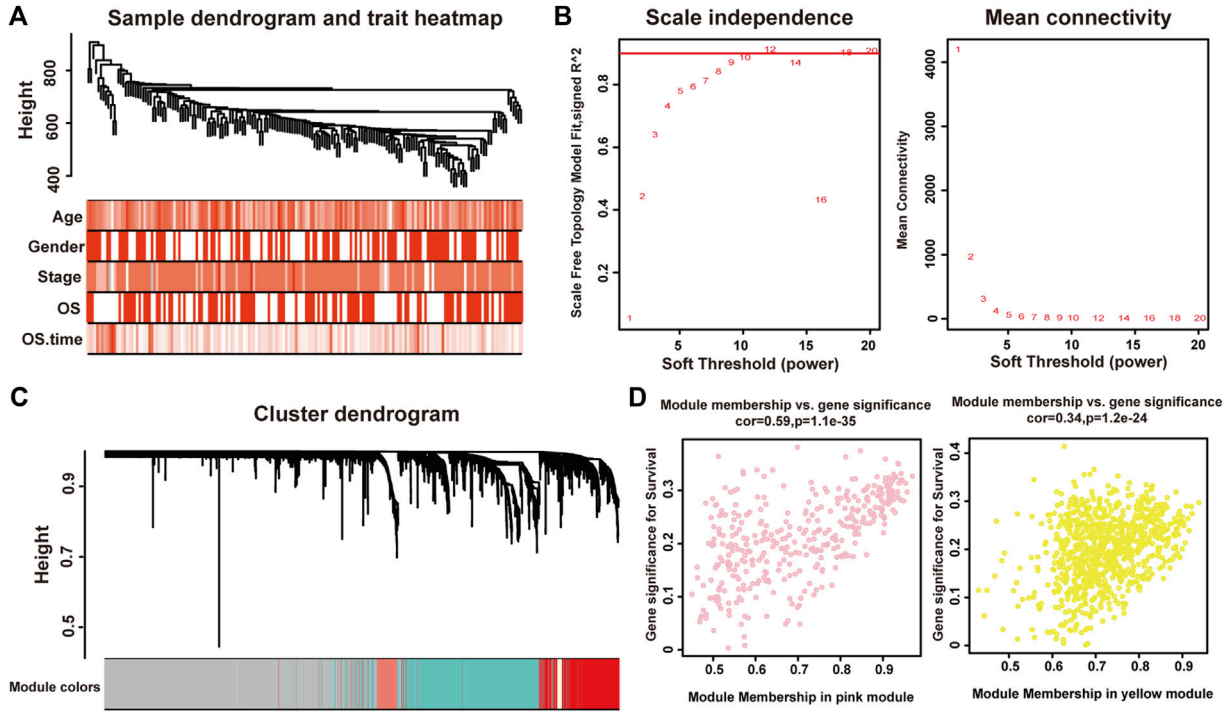


FIGURE 1 | Identification of modules associated with the clinical traits of pancreatic adenocarcinoma. **(A)** Clustering dendrogram of 175 samples. **(B)** The relationship between soft threshold (power) and network properties. Left panel: The relationship between soft-threshold (power) and scale-free topology. Right panel: The relationship between soft threshold (power) and mean connectivity. We set the soft threshold (power) to be twelve to satisfy scale-free topology of the network. **(C)** Total genes were clustered in 22 modules. Each module was marked with one color. **(D)** A scatterplot of Gene Significance (GS) for disease vs. Module Membership (MM) in the pink (left panel) and yellow (right panel) modules.

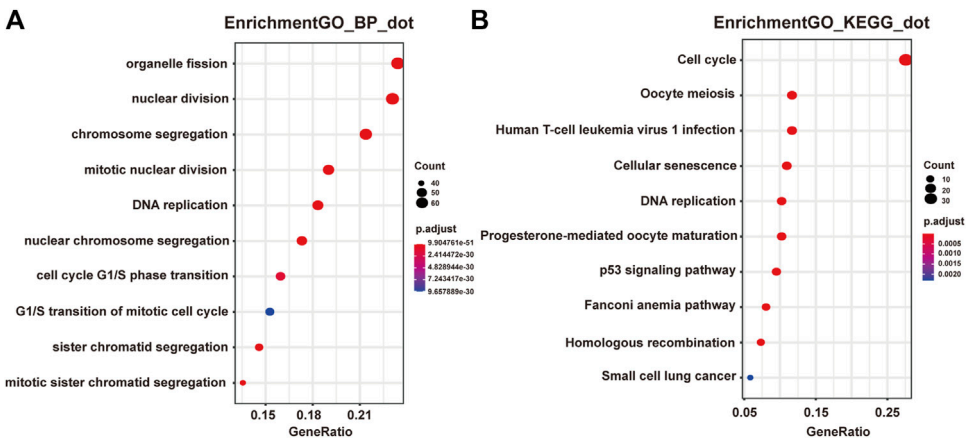


FIGURE 2 | Functional enrichment analysis of genes in the pink module. **(A)** GO analysis showed the top 10 enriched biological processes in the pink modules. **(B)** KEGG analysis showed the top 10 enriched pathways in the pink modules.

modules were 0.3 ($P = 4 \times 10^{-5}$) and −0.28 ($P = 2 \times 10^{-4}$), respectively. Clearly, Gene Significance (GS) and Module Membership (MM) analysis illustrated that genes highly significantly associated with pancreatic adenocarcinoma were also the most important elements of modules associated with pancreatic adenocarcinoma (**Figure 1D**).

Enrichment Analysis of Module Genes

Next we performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of these two modules. As presented in **Figures 2A,B**, genes in the pink module were significantly enriched in cell cycle, DNA replication, nuclear division, and regulation of cell cycle phase transition with adjusted

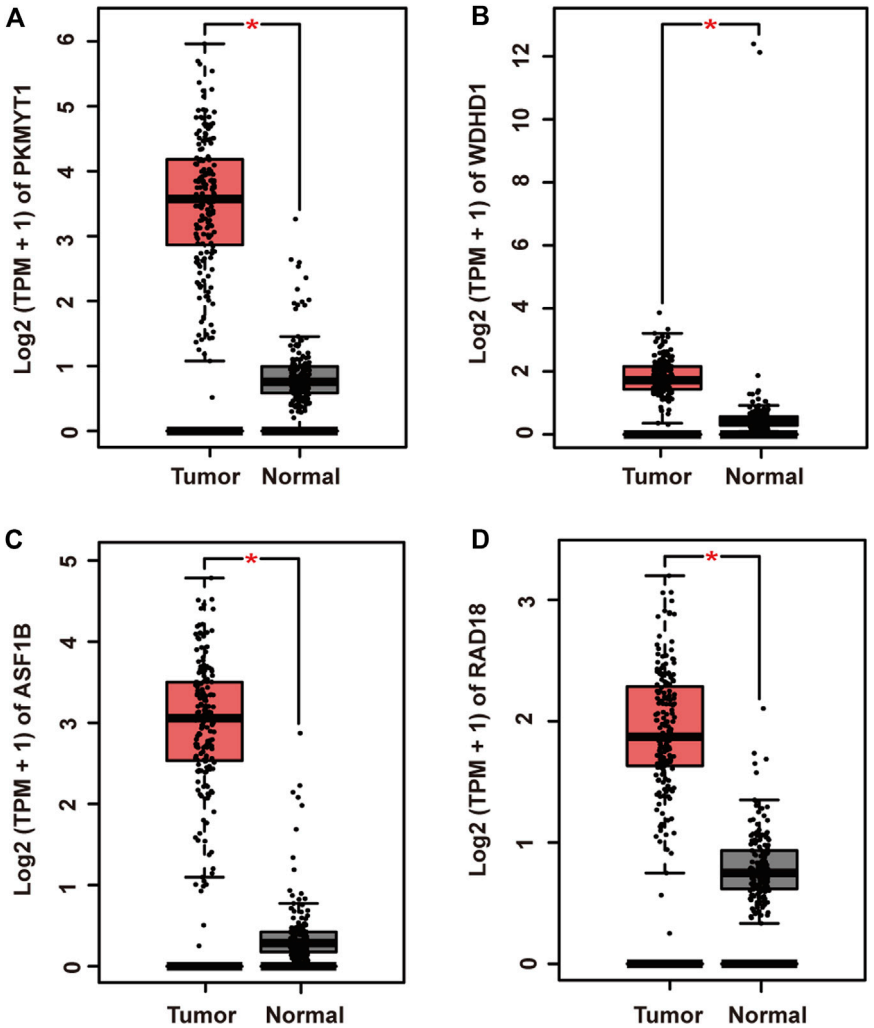


FIGURE 3 | Relative mRNA expression of four hub genes in pancreatic adenocarcinoma and adjacent normal tissues from TCGA. **(A)** PKMYT1. **(B)** WDHD1. **(C)** ASF1B. **(D)** RAD18. The expressions of these genes were significantly up-regulated in tumor samples.

p value smaller than 0.05, while genes in the yellow module were not enriched in any terms or pathways (data not shown), conferring the importance of these biological functions on pancreatic adenocarcinoma development. And genes in the pink module were selected for the following functional analysis.

Identification of Hub Genes

To further elucidate gene regulatory relationship in the module, we constructed co-expressed gene networks in the pink module (**Supplementary Material S3**) and identified master regulators with most connections with others. Finally, PKMYT1, WDHD1, ASF1B, and RAD18 stood out in the network. Next, we investigated their expression patterns in pancreatic adenocarcinoma. As expected, PKMYT1 (**Figure 3A**), WDHD1 (**Figure 3B**), ASF1B (**Figure 3C**), and RAD18 (**Figure 3D**) were significantly up-regulated in tumor tissues compared to adjacent normal tissues from TCGA, emphasizing their promising roles in carcinogenesis.

Functional Analysis of Survival-Associated Key Genes

Protein kinase, membrane associated tyrosine/threonine (PKMYT1), also known as MYT1, is a member of the WEE1 family of protein kinases, exerting key effects on Golgi and endoplasmic reticulum assembly (Chen et al., 2020). PKMYT1 was firstly recognized as a kinase capable of phosphorylating Cdc2 at Thr14 and Tyr15 (Mueller et al., 1995). Increasing studies have revealed its negative roles in cell cycle progression through suppressing cell cycle-associated factors, such as Cyclin A and CDK1 (Varadarajan et al., 2016), leading to its promising relationship with cancer. Previous studies have shown that PKMYT1 promoted cell proliferation and apoptosis resistance in multiple cancers, such as esophageal squamous cell carcinoma (Zhang et al., 2019), non-small cell lung cancer (Sun et al., 2019), prostate cancer (Wang et al., 2020), hepatocellular carcinoma (Liu

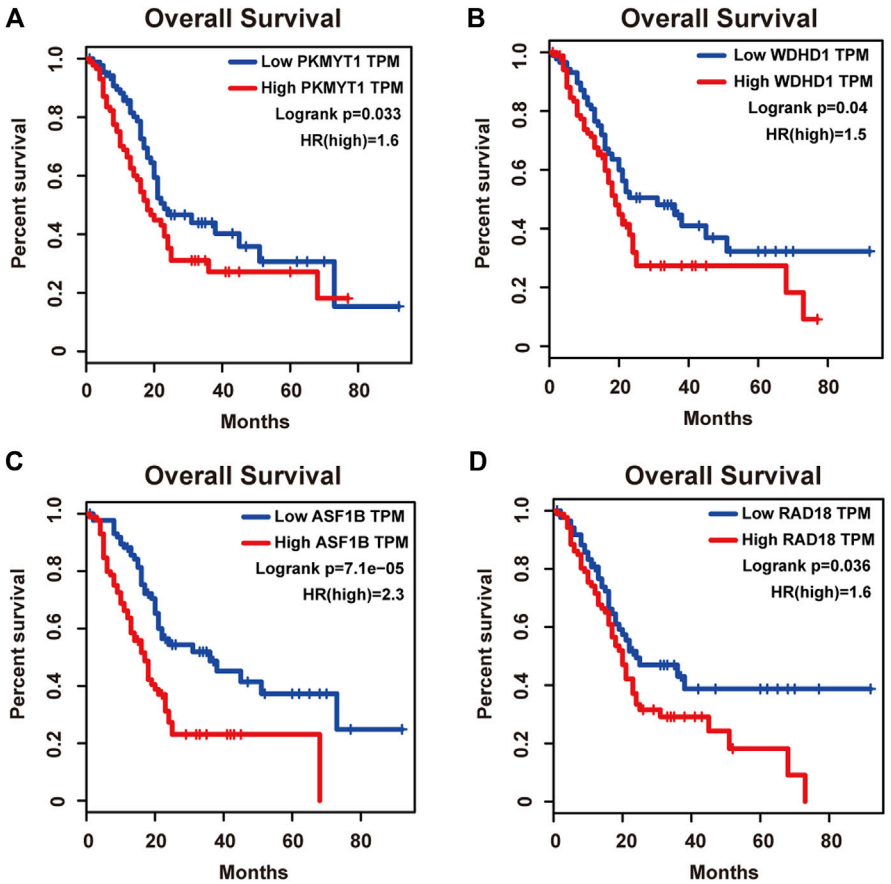


FIGURE 4 | Overall survival analysis of the four hub genes in pancreatic adenocarcinoma based on the Kaplan-Meier plotter. **(A)** PKMYT1. **(B)** WDHD1. **(C)** ASF1B. **(D)** RAD18. The high expressions of these four genes were associated with high risk.

et al., 2017), and gastric cancer (Zhang et al., 2020). Also overexpression of PKMYT1 predicted unfavorable prognosis in breast cancer (Liu Y. et al., 2020) and clear cell renal cell carcinoma (Chen et al., 2020), implicating it as an appealing therapeutic target. Nevertheless, an understanding of the correlation between PKMYT1 and pancreatic adenocarcinoma remains elusive, and our Kaplan-Meier analysis indicated, for the first time, better overall survival in the low transcription group approaching significance (**Figure 4A**), consistent with its tumor promotion roles in other cancers. However, disease free survival time did not show difference between low and high expression groups of PKMYT1 (**Figure 5A**). Through our network-based analysis, we also emphasized its crucial roles in the regulation of cell cycle (**Figure 2A**), which provided experimental clues for further investigations.

WD repeat and high-mobility group box DNA-binding protein 1 (WDHD1), an orthologue of Ctf4 in budding yeast and Mcl1 in fission yeast, is a DNA-binding protein involved in DNA replication and cell cycle (Abe et al., 2018). Recent studies have observed the overexpression of WDHD1 in the great majority of lung cancers and esophageal squamous cell carcinomas (Sato et al., 2010). WDHD1 has also been reported to facilitate the abrogation of G1 checkpoint upon DNA damage, leading to genomic instability and eventually tumorigenesis (Zhou et al., 2020). Moreover, WDHD1 could accelerate cell proliferation, cell viability, and metastasis in several cancers including cholangiocarcinoma and breast cancer (Sato et al., 2010; Liu et al., 2019; Ertay et al., 2020; Zhou and Chen, 2021). In accordance with the above research, both our overall and disease free survival analysis exhibited that high levels of WDHD1 correlated with poor patient outcome (**Figures 4B**, **5B**), confirming the oncogenic function of WDHD1 in pancreatic adenocarcinoma and expanding it roles in cancer biology which need further validations.

As one of histone H3-H4 chaperone anti-silencing function 1 (ASF1) isoforms, ASF1B plays important roles in chromatin-based progression of cellular DNA replication and transcription regulation, especially in cell proliferation (Paul et al., 2016). Accumulating evidence has shown that up-regulation of ASF1B stimulated cancer cell proliferation, DNA replication and migration, accompanied by restrained cell cycle arrest and apoptosis (Misiewicz-Krzeminska et al., 2013; Han et al., 2018; Zhou et al., 2019; Liu X. et al., 2020; Zhang et al., 2021), which was also consistent with our

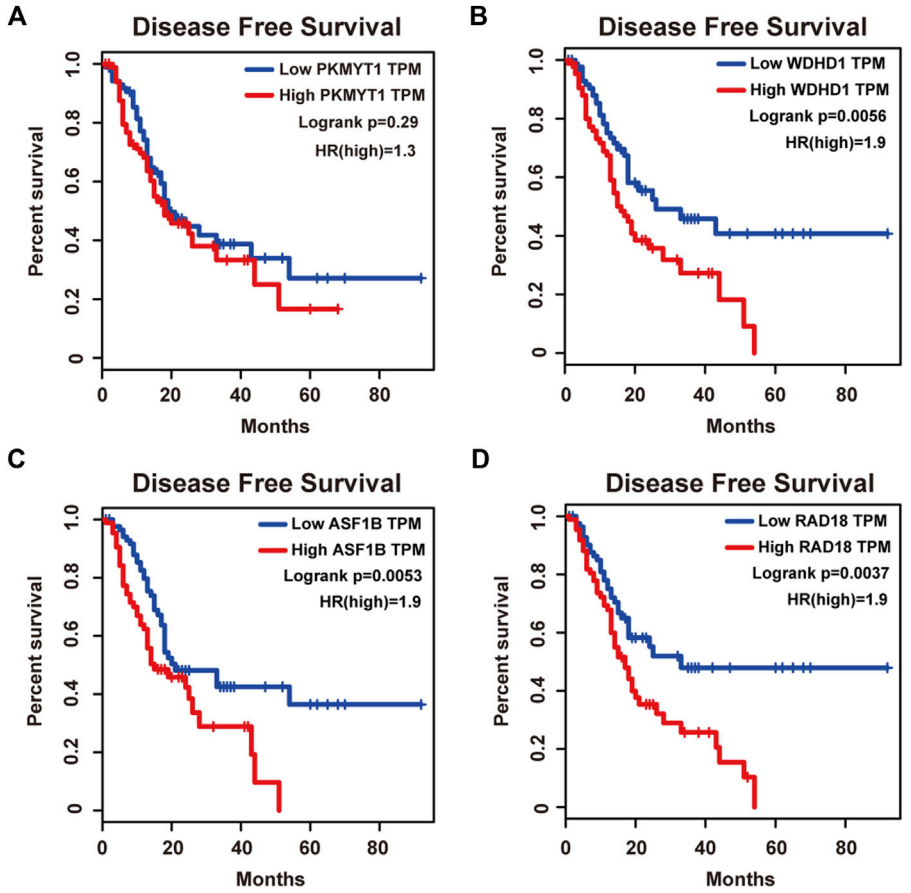


FIGURE 5 | Disease free survival analysis of the four hub genes in pancreatic adenocarcinoma based on the Kaplan-Meier plotter. **(A)** PKMYT1. **(B)** WDHD1. **(C)** ASF1B. **(D)** RAD18. The high expressions of WDHD1, ASF1B and RAD18 were associated with high risk.

enrichment analysis of module genes (**Figure 2A**). In addition, several studies have underlined the important prognostic impact of ASF1B on a variety of cancers (Corpet et al., 2011; Chen Z. et al., 2021; Feng et al., 2021). Another exciting finding was that ASF1B could regulate immune infiltration through affecting immune-related genes and pathways such as antigen processing and presentation and natural killer cell-mediated cytotoxicity (Hu et al., 2021; Zhan et al., 2021), conferring its strong hints on potential immunotherapeutic target for several malignancies. The both overall and disease free survival analysis also demonstrated that the high expression of ASF1B was associated with high risk (**Figures 4C**, **5C**), suggesting it serving as a tumor promoter in pancreatic adenocarcinoma.

RAD18 is an E3 ubiquitin ligase best known for its key roles in the monoubiquitylation of proliferating cell nuclear antigen (PCNA) in response to stalled replication forks, thus initiating DNA damage repair signaling (Williams et al., 2011; Kanu et al., 2016; Yang et al., 2017). Increasing reports have shown that RAD18 enhanced motility and invasiveness of cancerous cells, evidenced by the positive correlations between RAD18 and vital mediators of cell invasion and proliferation such as MMP-1 and MMP-9 (Zou et al., 2018; Xie et al., 2019). Meanwhile, a recent study has found elevated RAD18 was associated with gastric cancer progression and poor prognosis (Baatar et al., 2020), considering it as a novel prognostic biomarker. Accordingly, both overall and disease free survival time was significantly higher in patients with low RAD18 expression, compared with the high RAD18 expression group (**Figures 4D**, **5D**), highlighting its potential values in the treatment and prognosis of pancreatic adenocarcinoma.

CONCLUSION

Gene correlation approaches provide preliminary steps toward genetic interaction networks and offer clues about the function of unknown genes. Here, we employed WGCNA to identify novel hub genes including PKMYT1, WDHD1, ASF1B, and RAD18, and proposed for the first time their oncogenic roles during pancreatic adenocarcinoma progression. Further survival analysis verified their effective roles in predicting prognosis. In-depth mechanisms explaining their ability to allow neoplastic cells to breach tumorigenic barriers are needed. Meanwhile, we should not ignore the limitations of WGCNA as it is based on transcriptomic data and insufficient to reflect cell status globally, in which

multilayer data including mutations, copy number variations and proteomic data is also needed to be taken into account for bettering understanding mechanisms triggering cancer. Also, stronger computing power and more reasonable statistical methods should be stressed to improve gene correlation analysis.

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

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

LH conceived the study. TY and JW conducted the database search and analysis. XG and RM designed the algorithm and network construction. LH, LS, and BM wrote the manuscript. All authors contributed to biological analysis, interpretation of the results, read, and approved the final version of the manuscript.

FUNDING

The study was supported by National Natural Science Foundation of China (81860434).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.814798/full#supplementary-material

REFERENCES

Abe, T., Kawasumi, R., Giannattasio, M., Dusi, S., Yoshimoto, Y., Miyata, K., et al. (2018). AND-1 fork protection Function Prevents fork Resection and Is Essential for Proliferation. *Nat. Commun.* 9 (1), 3091. doi:10.1038/S41467-018-05586-7

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq--a Python Framework to Work with High-Throughput Sequencing Data. *Bioinformatics* 31 (2), 166–169. doi:10.1093/bioinformatics/btu638

Andersen, P. K., and Gill, R. D. (1982). Cox's Regression Model for Counting Processes: A Large Sample Study. *Ann. Statist.* 10 (4), 1100–1120. doi:10.1214/aos/1176345976

Baatar, S., Bai, T., Yokobori, T., Gombodorj, N., Nakazawa, N., Ubukata, Y., et al. (2020). High RAD18 Expression Is Associated with Disease Progression and Poor Prognosis in Patients with Gastric Cancer. *Ann. Surg. Oncol.* 27 (11), 4360–4368. doi:10.1245/s10434-020-08518-2

Chang, X., Shi, L., Gao, F., Russin, J., Zeng, L., He, S., et al. (2013). Genomic and Transcriptome Analysis Revealing an Oncogenic Functional Module in Meningiomas. *Foc* 35 (6), E3. doi:10.3171/2013.10.FOCUS13326

Chen, B., Hu, C., Jiang, L., Xiang, Z., Zuo, Z., Lin, Y., et al. (2021). Exploring the Significance of Novel Immune-Related Gene Signatures in the Prognosis and Immune Features of Pancreatic Adenocarcinoma. *Int. Immunopharmacol.* 92, 107359. doi:10.1016/J.Intimp.2020.107359

Chen, P., Zhang, Z., and Chen, X. (2020). Overexpression of PKMYT1 Facilitates Tumor Development and Is Correlated with Poor Prognosis in Clear Cell Renal Cell Carcinoma. *Med. Sci. Monit.* 26, e926755. doi:10.12659/MSM.926755

Chen, Z., Ou, D., Huang, Z., and Shen, P. (2021). Identification of Hsa_circ_0002024 as a Prognostic Competing Endogenous RNA (ceRNA) through the hsa_miR_129-5p/Anti-Silencing Function 1B Histone Chaperone (ASF1B) axis in Renal Cell Carcinoma. *Bioengineered* 12 (1), 6579–6593. doi:10.1080/21655979.2021.1974650

Corpet, A., De Koning, L., Toedling, J., Savignoni, A., Berger, F., Lemaître, C., et al. (2011). Asf1b, the Necessary Asf1 Isoform for Proliferation, Is Predictive of Outcome in Breast Cancer. *Embo J.* 30 (3), 480–493. doi:10.1038/emboj.2010.335

Ertay, A., Liu, H., Liu, D., Peng, P., Hill, C., Xiong, H., et al. (2020). WDHD1 Is Essential for the Survival of PTEN-Inactive Triple-Negative Breast Cancer. *Cell Death Dis* 11 (11), 1001. doi:10.1038/s41419-020-03210-5

Esposito, I., Konukiewitz, B., Schlitter, A. M., and Kloppel, G. (2014). Pathology of Pancreatic Ductal Adenocarcinoma: Facts, Challenges and Future Developments. *World J. Gastroenterol.* 20 (38), 13833–13841. doi:10.3748/wjg.v20.i38.13833

Feldmann, G., Beaty, R., Hruban, R. H., and Maitra, A. (2007). Molecular Genetics of Pancreatic Intraepithelial Neoplasia. *J. Hepatobiliary Pancreat. Surg.* 14 (3), 224–232. doi:10.1007/s00534-006-1166-5

Feng, Z., Zhang, J., Zheng, Y., Wang, Q., Min, X., and Tian, T. (2021). Elevated Expression of ASF1B Correlates with Poor Prognosis in Human Lung Adenocarcinoma. *Personalized Med.* 18 (2), 115–127. doi:10.2217/pme-2020-0112

Han, G., Zhang, X., Liu, P., Yu, Q., Li, Z., Yu, Q., et al. (2018). Knockdown of Anti-silencing Function 1B Histone Chaperone Induces Cell Apoptosis via Repressing PI3K/Akt Pathway in Prostate Cancer. *Int. J. Oncol.* 53 (5), 2056–2066. doi:10.3892/ijo.2018.4526

Hu, X., Zhu, H., Zhang, X., He, X., and Xu, X. (2021). Comprehensive Analysis of pan-cancer Reveals Potential of ASF1B as a Prognostic and Immunological Biomarker. *Cancer Med.* 10 (19), 6897–6916. doi:10.1002/cam4.4203

Jiangqiao, Z., Tao, Q., Zhongbao, C., Xiaoxiong, M., Long, Z., Jilin, Z., et al. (2019). Anti-silencing Function 1B Histone Chaperone Promotes Cell Proliferation and Migration via Activation of the AKT Pathway in clear Cell Renal Cell Carcinoma. *Biochem. Biophysical Res. Commun.* 511 (1), 165–172. doi:10.1016/j.bbrc.2019.02.060

Kanu, N., Zhang, T., Burrell, R. A., Chakraborty, A., Cronshaw, J., Dacosta, C., et al. (2016). RAD18, WRNIP1 and ATMIN Promote ATM Signalling in Response to Replication Stress. *Oncogene* 35 (30), 4009–4019. doi:10.1038/onc.2015.427

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: Accurate Alignment of Transcriptomes in the Presence of Insertions, Deletions and Gene Fusions. *Genome Biol.* 14 (4), R36. doi:10.1186/gb-2013-14-4-r36

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

Liu, B., Hu, Y., Qin, L., Peng, X.-B., and Huang, Y.-X. (2019). MicroRNA-494-dependent WDHDI Inhibition Suppresses Epithelial-Mesenchymal Transition, Tumor Growth and Metastasis in Cholangiocarcinoma. *Dig. Liver Dis.* 51 (3), 397–411. doi:10.1016/j.dld.2018.08.021

Liu, L., Wu, J., Wang, S., Luo, X., Du, Y., Huang, D., et al. (2017). PKMYT1 Promoted the Growth and Motility of Hepatocellular Carcinoma Cells by Activating Beta-Catenin/TCF Signaling. *Exp. Cel Res.* 358 (2), 209–216. doi:10.1016/j.yexcr.2017.06.014

Liu, X., Song, J., Zhang, Y., Wang, H., Sun, H., Feng, X., et al. (2020). ASF1B Promotes Cervical Cancer Progression through Stabilization of CDK9. *Cel Death Dis* 11 (8), 705. doi:10.1038/s41419-020-02872-5

Liu, Y., Qi, J., Dou, Z., Hu, J., Lu, L., Dai, H., et al. (2020). Systematic Expression Analysis of WEE Family Kinases Reveals the Importance of PKMYT1 in Breast Carcinogenesis. *Cell Prolif* 53 (2), e12741. doi:10.1111/cpr.12741

Luchini, C., Capelli, P., and Scarpa, A. (2016). Pancreatic Ductal Adenocarcinoma and its Variants. *Surg. Pathol. Clin.* 9 (4), 547–560. doi:10.1016/j.path.2016.05.003

Makohon-Moore, A., and Iacobuzio-Donahue, C. A. (2016). Pancreatic Cancer Biology and Genetics from an Evolutionary Perspective. *Nat. Rev. Cancer* 16 (9), 553–565. doi:10.1038/nrc.2016.66

Mcguigan, A., Kelly, P., Turkington, R. C., Jones, C., Coleman, H. G., and Mccain, R. S. (2018). Pancreatic Cancer: A Review of Clinical Diagnosis, Epidemiology, Treatment and Outcomes. *World J. Gastroenterol.* 24 (43), 4846–4861. doi:10.3748/wjg.v24.i43.4846

Misiewicz-Krzeminska, I., Sarasquete, M. E., Quwaider, D., Krzeminski, P., Ticona, F. V., Paino, T., et al. (2013). Restoration of microRNA-214 Expression Reduces Growth of Myeloma Cells through Positive Regulation of P53 and Inhibition of DNA Replication. *Haematologica* 98 (4), 640–648. doi:10.3324/haematol.2012.070011

Mueller, P. R., Coleman, T. R., Kumagai, A., and Dunphy, W. G. (1995). Myt1: A Membrane-Associated Inhibitory Kinase that Phosphorylates Cdc2 on Both Threonine-14 and Tyrosine-15. *Science* 270 (5233), 86–90. doi:10.1126/science.270.5233.86

Paul, P. K., Rabaglia, M. E., Wang, C.-Y., Stapleton, D. S., Leng, N., Kendziorski, C., et al. (2016). Histone Chaperone ASF1B Promotes Human β-cell Proliferation via Recruitment of Histone H3.3. *Cell Cycle* 15 (23), 3191–3202. doi:10.1080/15384101.2016.1241914

Rahib, L., Smith, B. D., Aizenberg, R., Rosenzweig, A. B., Fleshman, J. M., and Matrisian, L. M. (2014). Projecting Cancer Incidence and Deaths to 2030: the Unexpected burden of Thyroid, Liver, and Pancreas Cancers in the United States. *Cancer Res.* 74 (11), 2913–2921. doi:10.1158/0008-5472.CAN-14-0155

Sato, N., Koinuma, J., Fujita, M., Hosokawa, M., Ito, T., Tsuchiya, E., et al. (2010). Activation of WD Repeat and High-Mobility Group Box DNA Binding Protein 1 in Pulmonary and Esophageal Carcinogenesis. *Clin. Cancer Res.* 16 (1), 226–239. doi:10.1158/1078-0432.CCR-09-1405

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

Sun, Q. S., Luo, M., Zhao, H. M., and Sun, H. (2019). Overexpression of PKMYT1 Indicates the Poor Prognosis and Enhances Proliferation and Tumorigenesis in Non-small Cell Lung Cancer via Activation of Notch Signal Pathway. *Eur. Rev. Med. Pharmacol. Sci.* 23 (10), 4210–4219. doi:10.26355/eurrev_201905_17925

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING V11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-wide Experimental Datasets. *Nucleic Acids Res.* 47 (D1), D607–D613. doi:10.1093/nar/gky1131

Varadarajan, R., Ayeni, J., Jin, Z., Homola, E., and Campbell, S. D. (2016). Myt1 Inhibition of Cyclin A/Cdk1 Is Essential for Fusome Integrity and Premeiotic Centriole Engagement in Drosophila Spermatocytes. *MBoC* 27 (13), 2051–2063. doi:10.1091/mbc.E16-02-0104

Vogelstein, B., and Kinzler, K. W. (2015). The Path to Cancer - Three Strikes and You're Out. *N. Engl. J. Med.* 373 (20), 1895–1898. doi:10.1056/Nejmp1508811

Wang, J., Wang, L., Chen, S., Peng, H., Xiao, L., E Du, E., et al. (2020). PKMYT1 Is Associated with Prostate Cancer Malignancy and May Serve as a Therapeutic Target. *Gene* 744, 144608. doi:10.1016/j.gene.2020.144608

Williams, S. A., Longerich, S., Sung, P., Vaziri, C., and Kupfer, G. M. (2011). The E3 Ubiquitin Ligase RAD18 Regulates Ubiquitylation and Chromatin Loading of FANCD2 and FANCI. *Blood* 117 (19), 5078–5087. doi:10.1182/blood-2010-10-311761

Xie, C., Lu, D., Xu, M., Qu, Z., Zhang, W., and Wang, H. (2019). Knockdown of RAD18 Inhibits Glioblastoma Development. *J. Cel Physiol* 234 (11), 21100–21112. doi:10.1002/jcp.28713

Yang, L., Li, Y., Wei, Z., and Chang, X. (2018). Coexpression Network Analysis Identifies Transcriptional Modules Associated with Genomic Alterations in Neuroblastoma. *Biochim. Biophys. Acta (Bba) - Mol. Basis Dis.* 1864 (6 Pt B), 2341–2348. doi:10.1016/j.bbadis.2017.12.020

Yang, Y., Gao, Y., Mutter-Rottmayer, L., Zlatanou, A., Durando, M., Ding, W., et al. (2017). DNA Repair Factor RAD18 and DNA Polymerase Polκ Confer Tolerance of Oncogenic DNA Replication Stress. *J. Cel Biol* 216 (10), 3097–3115. doi:10.1083/jcb.201702006

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Zhan, T., Gao, X., Wang, G., Li, F., Shen, J., Lu, C., et al. (2021). Construction of Novel lncRNA-miRNA-mRNA Network Associated with Recurrence and Identification of Immune-Related Potential Regulatory Axis in Hepatocellular Carcinoma. *Front. Oncol.* 11, 626663. doi:10.3389/fonc.2021.626663

Zhang, Q.-Y., Chen, X.-Q., Liu, X.-C., and Wu, D.-M. (2020). PKMYT1 Promotes Gastric Cancer Cell Proliferation and Apoptosis Resistance. *Onco Targets Ther.* 13, 7747–7757. doi:10.2147/OTT.S255746

Zhang, Q., Zhao, X., Zhang, C., Wang, W., Li, F., Liu, D., et al. (2019). Overexpressed PKMYT1 Promotes Tumor Progression and Associates with Poor Survival in Esophageal Squamous Cell Carcinoma. *Cancer Manag. Res.* 11, 7813–7824. doi:10.2147/CMAR.S214243

Zhang, W., Gao, Z., Guan, M., Liu, N., Meng, F., and Wang, G. (2021). ASF1B Promotes Oncogenesis in Lung Adenocarcinoma and Other Cancer Types. *Front. Oncol.* 11, 731547. doi:10.3389/fonc.2021.731547

Zhou, X.-g., Huang, X.-l., Liang, S.-y., Tang, S.-m., Wu, S.-k., Huang, T.-t., et al. (2018). Identifying miRNA and Gene Modules of colon Cancer Associated with Pathological Stage by Weighted Gene Co-expression Network Analysis. *Onco Targets Ther.* 11, 2815–2830. doi:10.2147/Ott.S163891

Zhou, Y., and Chen, J. J. (2021). STAT3 Plays an Important Role in DNA Replication by Turning on WDHD1. *Cell Biosci* 11 (1), 10. doi:10.1186/s13578-020-00524-x

Zhou, Y., Pei, F., Ji, M., Zhang, F., Sun, Y., Zhao, Q., et al. (2020). WDHD1 Facilitates G1 Checkpoint Abrogation in HPV E7 Expressing Cells by Modulating GCN5. *BMC Cancer* 20 (1), 840. doi:10.1186/s12885-020-07287-1

Zou, S., Yang, J., Guo, J., Su, Y., He, C., Wu, J., et al. (2018). RAD18 Promotes the Migration and Invasion of Esophageal Squamous Cell Cancer via the JNK-MMPs Pathway. *Cancer Lett.* 417, 65–74. doi:10.1016/j.canlet.2017.12.034

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 03 March 2022
doi: 10.3389/fgene.2022.819699

Extracellular Vesicle-Derived miR-105-5p Promotes Malignant Phenotypes of Esophageal Squamous Cell Carcinoma by Targeting SPARCL1 *via* FAK/AKT Signaling Pathway

Binjun He[1,2], Kang Zhang[2], Xiaoliang Han[2], Chao Su[2], Jiaming Zhao[2], Guoxia Wang[2], Guzong Wang[2], Liuya Zhang[2] and Wenbin Hu[2]**

[1]Department of Thoracic Surgery, Shaoxing People's Hospital (Zhejiang University School of Medicine), Shaoxing, China, [2]Department of Thoracic Surgery, Affiliated Hospital of Shaoxing University /Shaoxing Municipal Hospital, Shaoxing, China

OPEN ACCESS

Edited by:
Peilin Jia,
University of Texas Health Science Center, United States

Reviewed by:
Zhengliang Tu,
Zhejiang University, China
Zhijun Chen,
Zhoushan Hospital, China

***Correspondence:**
Wenbin Hu
kyhuwenbin@163.com
Liuya Zhang
zhangly956888@126.com

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: 22 November 2021
Accepted: 12 January 2022
Published: 03 March 2022

Citation:
He B, Zhang K, Han X, Su C, Zhao J, Wang G, Wang G, Zhang L and Hu W (2022) Extracellular Vesicle-Derived miR-105-5p Promotes Malignant Phenotypes of Esophageal Squamous Cell Carcinoma by Targeting SPARCL1 via FAK/AKT Signaling Pathway. Front. Genet. 13:819699. doi: 10.3389/fgene.2022.819699

Objective: Esophageal squamous cell carcinoma (ESCC) presents high morbidity and mortality. It was demonstrated that blood-derived vesicles can facilitate ESCC development and transmit regulating signals. However, the molecular mechanism of vesicle miRNA secreted by tumor cells affecting ESCC progression has not been explored.

Methods: The mRNA-related signaling pathways and differentially expressed genes were screened out in TCGA dataset. The levels of miRNA-105-5p and SPARCL1 were determined by qRT-PCR. Protein level determination was processed using Western blot. The interaction between the two genes was verified with the dual-luciferase method. A transmission electron microscope was utilized to further identify extracellular vesicles (EVs), and co-culture assay was performed to validate the intake of EVs. *In vitro* experiments were conducted to evaluate cell function changes in ESCC. A mice tumor formation experiment was carried out to observe tumor growth *in vivo*.

Results: MiRNA-105-5p expression was increased in ESCC, while SPARCL1 was less expressed. MiRNA-105-5p facilitated cell behaviors in ESCC through targeting SPARCL1 and regulating the focal adhesion kinase (FAK)/Akt signaling pathway. Blood-derived external vesicles containing miRNA-105-5p and EVs could be internalized by ESCC cells. Then, miRNA-105-5p could be transferred to ESCC cells to foster tumorigenesis as well as cell behaviors.

Conclusion: EV-carried miRNA-105-5p entered ESCC cells and promoted tumor-relevant functions by mediating SPARCL1 and the FAK/Akt signaling pathway, which indicated that the treatment of ESCC *via* serum EVs might be a novel therapy and that miRNA-105-5p can be a molecular target for ESCC therapy.

Keywords: EVs, miR-105-5p, SPARCL1, ESCC, FAK/AKT, proliferation, migration, invasion

INTRODUCTION

Esophageal carcinoma (EC) ranks sixth in mortality (544,000 deaths) and seventh in incidence (604,000 new cases) in 2020 (Sung et al., 2021), which means that EC is a serious threat to human health. Esophageal squamous cell carcinoma (ESCC) makes up the majority of EC (Lin et al., 2013). ESCC is formed by the abnormal proliferation of esophagus squamous epithelium, which is a kind of cancer with unobvious early clinical features. Besides, most ESCC patients are diagnosed at advanced stages accompanied by distant metastasis, and their 5-year survival rates are lower than 20% after receiving conventional therapies (Siegel et al., 20182018; Sawada et al., 2019). Accordingly, finding effective therapeutic approaches for ESCC patients is of substantial importance to the enhancement of advanced ESCC patients' survival.

With deeper investigations of miRNAs in recent years, it has been confirmed that miRNAs can mediate the growth and metastasis of multiple cancers (Xu et al., 2018; Sheng et al., 2019; Wu et al., 2021). For instance, miRNA-10b-3p is capable of fostering ESCC growth and metastasis (Priniski et al., 2019). MiRNA-133b can inhibit ESCC cell processes (Zeng et al., 2019). MiRNA-105-5p is a member of miRNAs, yet there have been relatively few studies about its role in cancers. Currently, it has only been reported that miRNA-105-5p modulates PES1 in liver cancer stem cells to facilitate cell growth (Wei et al., 2019). Nonetheless, there has been no research on the relationship between miRNA-105-5p and ESCC. To this end, the present study aimed to gain deeper insight into the regulatory mechanism.

Extracellular vesicles (EVs) are extracellular membrane particles' component with diameters of 40–1,000 nm, whose outer layer is a bilayer lipid membrane while the inside encapsulates different proportions of DNA, RNA, and protein components (Jin et al., 2021). Numerous reports indicate that EVs are usually existing in body fluids, including cerebrospinal fluid, blood, urine, saliva, pleural effusion, and ascites (Caivano et al., 2015). EVs have many biological functions, such as intercellular information exchange, protecting and repairing damaged cells and tissue, participating in immune response, and promoting angiogenesis (Abels and Breakefield, 2016; Xu et al., 2016). Recently, research has suggested that serum EVs can transfer miRNAs to tumor cells to regulate tumorigenesis. For example, exosomal miRNA-660-5p is sent to foster the metastasis of non-small cell lung cancer (NSCLC) cells (Qi et al., 2019). MiRNA-21 carried on Evs can be transferred to ESCC cells to induce ESCC cell proliferation (Tanaka et al., 2013). Nevertheless, whether miRNA-105-5p in Evs could modulate ESCC remains unknown.

To conclude, we confirmed that serum EV-derived miRNA-105-5p could be transferred to ESCC cells to foster the progression of ESCC cells by regulating the focal adhesion kinase (FAK)/Akt signaling pathway, indicating that miRNA-105-5p might be a novel signal facilitating ESCC development.

MATERIALS AND METHODS

Bioinformatics Methods

ESCC-related miRNA chip data GSE55856 (normal: n = 108, tumor: n = 108) were accessed from the GEO database, and differential analysis was performed to screen differentially expressed miRNAs (DEmiRNAs) using the "limma" package (|logFC|>2, $p$adj < 0.01). Meanwhile, the ESCC-related miRNA dataset (normal: n = 13, tumor: n = 96) was downloaded from the TCGA database and differential analysis was conducted to screen DEmiRNAs using the "edgeR" package (|logFC|>2, $p$adj < 0.01). The target miRNA was confirmed, and the expression position of the target miRNA was localized *via* the EVmiRNA database. Bioinformatics databases including miRDB (http://mirdb.org/), mirDIP (http://ophid.utoronto.ca/mirDIP/index.jsp#r), and TargetScan (http://www.targetscan.org/vert_71/) were employed to predict the target mRNAs of the target miRNA. Besides, ESCC-related mRNA expression data were obtained from TCGA database and differential analysis was carried out to screen differentially expressed mRNAs (DEmRNAs) (|logFC|>1.5, $p$adj < 0.01). Then, the differential mRNA and the predicted target gene were taken to intersect. The ultimate target mRNA that had the binding sites of the target miRNA was obtained. GSEA software was utilized to conduct pathway enrichment analysis on target genes, so as to further investigate the mechanism of target miRNA and its target genes affecting ESCC.

Patient Samples

This paper included 64 ESCC patients admitted to the Shaoxing People's Hospital from June 2017 to December 2018. ESCC tissue and adjacent normal tissue samples (within at least 2 cm away from the edge of a tumor) were obtained by surgical resection. All specimens had detailed clinical information, and their tissue samples were confirmed by experts. All patients did not receive radiotherapy, chemotherapy, biotherapy, or traditional Chinese medicine treatment before operation. Peripheral blood was collected from all patients before the operation, and the blood samples of 20 healthy people were also obtained at the same time. This study was authorized by the Hospital Ethics Committee, and written informed consent was acquired from all participants.

Cell Culture

Normal human esophageal epithelial cell line HEEC and ESCC cell line TE-1 were purchased from BeNa Culture Collection (BNCC, Beijing, China). ESCC cell lines Eca109, EC9706, and NEC were ordered from American Type Culture Collection (ATCC, Manassas, VA, USA). All the above cell lines were cultured in Dulbecco's Modified Eagle's Medium (Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 10% fetal bovine serum (Thermo Fisher Scientific, Waltham, MA, USA) routinely. The culture conditions were 37°C and 5% CO_2 atmosphere. All cells were utilized for subsequent assays after four passages.

Cell Transfection

The MiRNA-105-5p-mimic, miRNA-105-5p-inhibitor, and their corresponding negative controls (NCs) ordered from GenePharma (Shanghai, China) were transfected into ESCC cells using Lipofectamine 2000 (Thermo Fisher Scientific, Waltham, MA, USA).

The SPARCL1 overexpression vector (oe-SPARCL1) and overexpression negative control (oe-NC) were bought from GenePharma (Shanghai, China). Oe-SPARCL1 and oe-NC were transfected into ESCC cells by the Lipofectamine 2000 reagent kit. After 48 h of cell incubation, the transfection efficiency was tested.

Separation of EVs in Serum

Serum-derived EVs were purified using the EVs isolation kit (Invitrogen, Carlsbad, CA, United States). In short, reagent and serum samples were subjected to incubation (30 min) and proceeded to centrifugation (10,000 rpm, 5 min, room temperature). Thereafter, precipitated EVs were gathered for resuspending in phosphate-buffered saline (PBS).

Transmission Electron Microscope Analysis

After EVs were suspended in PBS, a single drop of suspension was dripped in a sample-loaded copperplate. After 1 min of standing at room temperature, the redundant liquid was removed using a piece of filter paper. Subsequently, 2% uranyl acetate (5 μl) was dripped in the sample-loaded copperplate and the samples were negatively stained for 1 min at general temperature. Then, the redundant negative staining reagent was removed. After being dried under an incandescent lamp, the electron microscope H-7600 (Hitachi High-Technologies, Japan) was employed to observe the result, and images were photographed for further analysis.

Intake of EVs by ESCC Cells

Serum EVs were labeled using the PKH67 Green Fluorescent Cell Linker Mini Kit (Sigma, USA). ESCC cells [10 cells/ml (Sawada et al., 2019)] were plated and cultured for 4 h to foster cell attachment. Subsequently, 2 μl PKH67-labeled EVs were added into the plate and maintained to culture for 48 h. Next, the medium was discarded and the cells were rinsed with PBS and fixed in 4% paraformaldehyde for 10 min to enhance cell permeability. Then, the paraformaldehyde was abandoned and cells were washed with PBS, after which 1 ml 1% bovine serum albumin (BSA) was added and cells were incubated for 20–30 min at room temperature to reduce the staining background. The F-actin and nucleus in the cytoskeleton were stained through adding Alexa Fluor® 594 Phalloidin and the DAPI of appropriate concentration. At last, the intake of EVs in blood-adherent cells by ESCC cells was observed *via* a fluorescence microscope.

Co-culture of ESCC Cells and Serum-Derived EVs

After being filtered by a 0.22 μm sterilizing filter, serum-derived EVs (20 μg/well) were added into a plate laid with ESCC cells for routine culture in an incubator according to the quantitative results of a bicinchoninic acid kit (Beyotime, China). ESCC cells were isolated for subsequent experiments after 48 h.

TABLE 1 | Primer sequences for qRT-PCR.

Genes	Sequences
miRNA-105-5p	Forward: 5′-TCGGCAGGTCAAATGCTCAGAC-3′ Reverse: 5′-CTCAACTGGTGTCGTGGA -3′
U6	Forward: 5′-CTCGCTTCGGCAGCACA-3′ Reverse: 5′-AACGCTTCACGAATTTGCGT-3′
SPARCL1	Forward: 5′- GCCTGGAGAGCACCAAGAGGCC -3′ Reverse: 5′- ATGGTCCCCAGCCAAAAGCCTC -3′
GAPDH	Forward: 5′- GACCTGACCTGCCGTCTA-3′ Reverse: 5′-AGGAGTGGGTGTCGCTGT-3′

RNA Extraction and qRT-qPCR

For the quantification of miRNA-105-5p, total RNA was extracted from cells, tissue samples, and EVs using RNAiso Plus (TaKaRa, Japan) and QIAzol Lysis Reagent (Qiagen, Hilden, Germany). Then, the total RNA was reverse-transcribed into complementary DNA (cDNA) by using the Superscript II Reverse Transcription Assay Kit (Invitrogen, USA). For detection of SPARCL1 in cells and tissue samples, total RNA was extracted from tissue samples and cells by using TRIzol reagent (TaKaRa, Japan) and transcribed into cDNA by using the M-MLV Reverse transcriptase Assay Kit (TaKaRa, Japan). The concentration of RNA was measured *via* NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, USA). qRT-PCR was performed on the ABI 7500 Real-Time PCR system (Thermo Fisher Scientific, USA) using SYBR Prime Script TM RT-qPCR Kit (Takara, Japan). GAPDH and U6 were used as internal controls. The primers were displayed in **Table 1**. Relative expression levels of miRNA-105-5p and SPARCL1 were analyzed by the $2^{-\Delta\Delta Ct}$ method.

Western Blot

Total proteins were isolated from ESCC cells and EVs using radio-immunoprecipitation assay (Beyotime, China) lysis buffer containing 1% proteinase inhibitor (Beyotime, China). The concentration of the proteins was measured by the BCA protein assay kit (Thermo Fisher Scientific, United States). After denatured at high temperature, protein samples were separated on 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (30 μg/lane) and transferred onto polyvinylidene fluoride membranes (Millipore, USA). After being blocked with 5% BSA/Tris-buffered saline with Tween-20 (TBST) for 60 min, the membranes were incubated with primary antibodies overnight at 4°C. On the following day, the membranes were washed with 1 × TBST (Solarbio, China) at room temperature three times with 5 min for each time. Thereafter, the membranes were reacted with horseradish peroxidase-conjugated secondary antibody goat anti-rabbit at room temperature for 120 min, after which the membranes were washed with 1 × TBST in triplicate with

TABLE 2 | Antibodies for Western blot.

Antibodies	Sources	Dilution multiple	Co. No
TSG101	Rabbit antibody	1:2,000	abcam (ab125011)
CD9	Rabbit antibody	1:2,000	abcam (ab92726)
CD63	Rabbit antibody	1:1,000	abcam (ab217345)
SPARCL1	Rabbit antibody	1:500	abcam (ab107533)
FAK	Rabbit antibody	1:2,000	abcam (ab40794)
p-FAK	Rabbit antibody	1:1,000	abcam (ab4792)
Akt	Rabbit antibody	1:500	abcam (ab8805)
p-Akt	Rabbit antibody	1:1,000	abcam (ab38449)
GAPDH	Rabbit antibody	1:10,000	abcam (ab181602)
IgG H&L	Goat anti-rabbit	1:10,000	abcam (ab6721)

20 min for each time. Finally, the electrochemiluminescence (ECL) assay kit (Solarbio, China) was utilized for protein bands visualization and images were captured for further observation. The antibodies for Western blot were listed in **Table 2**.

Cell Counting Kit-8 Assay

ESCC cells at the logarithmic phase were digested and then 200 μl cells (1×10^4 cells/ml) were seeded in plates for routine incubation. After the cells were cultured for 0, 24, 48, and 72 h, each well was supplemented with 20 μl Cell Counting Kit-8 (CCK-8) solution (Yeasen, Shanghai, China) and the cells were incubated in a constant-temperature incubator for an additional 1 h. At last, the absorbance of each well at 450 nm was identified by SpectraMax M5 (Molecular Devices, USA).

Wound Healing Assay

ESCC cells at the logarithmic phase were digested, and then 5×10^5 cells were seeded into plates. After single cell layers were formed, a scratch was made on cells using the tip of a sterile pipette. After being washed with PBS, cells were routinely cultured in FBS-free medium. The cell migration area was photographed at 0 and 48 h under an inverted microscope, and the cell migration rate was calculated as: cell migration rate = (width at 0 h − width at 48 h)/ width at 0 h.

Transwell Invasion Assay

ESCC cells at the logarithmic phase were digested, and 200 μl of cells in an FBS-free medium were then seeded into upper Transwell chambers laid with matrix at a density of 2×10^5 cells/ml. The upper chambers were placed in plates, and the lower chambers were added with a fresh medium containing 10% FBS. After being cultivated at 37°C for 24 h, a soaked cotton swab was used to remove the cells in the upper chambers, after which the cells in the lower chambers were fixed in 4% paraformaldehyde for 10 min and stained with 0.1% crystal violet for 15 min. Five fields were randomly selected, photographed, and counted under an inverted microscope.

Tumor Formation in Nude Mice

A total of 10 female BALB/C nude mice (4–6 weeks old, weighing 15–20 g) were ordered from the Shanghai Institute of Materia Medica, Chinese Academy of Sciences (Shanghai, China). The mice were divided into two groups with five mice in each group. Approximately 1×10^7 ESCC cells that were diluted by 200 μl PBS were subcutaneously injected into the left hindlimb of nude mice. Tumor volume was monitored every 3 days. Tumor volume was calculated with a formula of: volume= [length × width (Lin et al., 2013)]/2 (mm^3). When the tumor volume of nude mice reached 100 mm^3, the serum-derived EVs (20 μg) of ESCC patients were injected into nude mice, with an equivalent amount of PBS being blank control. The injection was performed every 3 days with three consecutive weeks in total. At the end of the third week, the mice were given euthanasia. Then, the tumors were isolated and weighed. The isolated tumors were frozen and stored in liquid nitrogen for further analysis.

Immunohistochemistry

The tumor tissue of nude mice was fixed in 4% paraformaldehyde and then put into a refrigerator at −4°C. Then, tumor tissue samples were dehydrated in a graded ethanol solution, transparently disposed by xylene, embedded in paraffin, and sliced to about 5 μm slices. Thereafter, the slices were stained with corresponding primary and secondary antibodies, counterstained with hematoxylin, dehydrated, and fixed for further observation. The antibodies for immunohistochemistry were listed in **Table 3**.

Dual-Luciferase Reporter Gene Assay

The amplified wild-type (WT) and mutant (MUT) 3′UTR of SPARCL1 were cloned into the multiple cloning sites of the pmirGLO luciferase vector (Promega, USA) to generate SPARCL1-WT and SPARCL1-MUT vectors. Using the Lipofectamine 2000 reagent kit, SPARCL1-WT/SPARCL1-MUT vectors and NC-mimic/miRNA-105-5p mimic were co-transfected into ESCC cells. After 48 h of transfection, Firefly and Renilla luciferase activities were evaluated using the Dual-Luciferase Reporter Assay kit (Promega, United States). Each transfection was repeated three times.

Statistical Analysis

Each assay underwent at least three repetitions. Statistics were processed using GraphPad Prism seven software (La Jolla, CA, USA). Measurement data were exhibited as mean ± standard deviation. The differences between two groups were analyzed by Student's *t*-test, and one-way analysis of variance was adopted to compare more than two groups. $p < 0.05$ was regarded as statistically significant.

TABLE 3 | Antibodies for immunohistochemistry.

Antibodies	Sources	Dilution multiple	Co. No
SPARCL1	Rabbit antibody	1:50	abcam (ab125011)
FAK	Rabbit antibody	1:250	abcam (ab40794)
p-FAK	Rabbit antibody	1:200	abcam (ab4792)
Akt	Rabbit antibody	1:1,000	abcam (ab8805)
p-Akt	Rabbit antibody	1:200	abcam (ab38449)
IgG H and L	Goat anti-rabbit	1:1,000	abcam (ab6721)

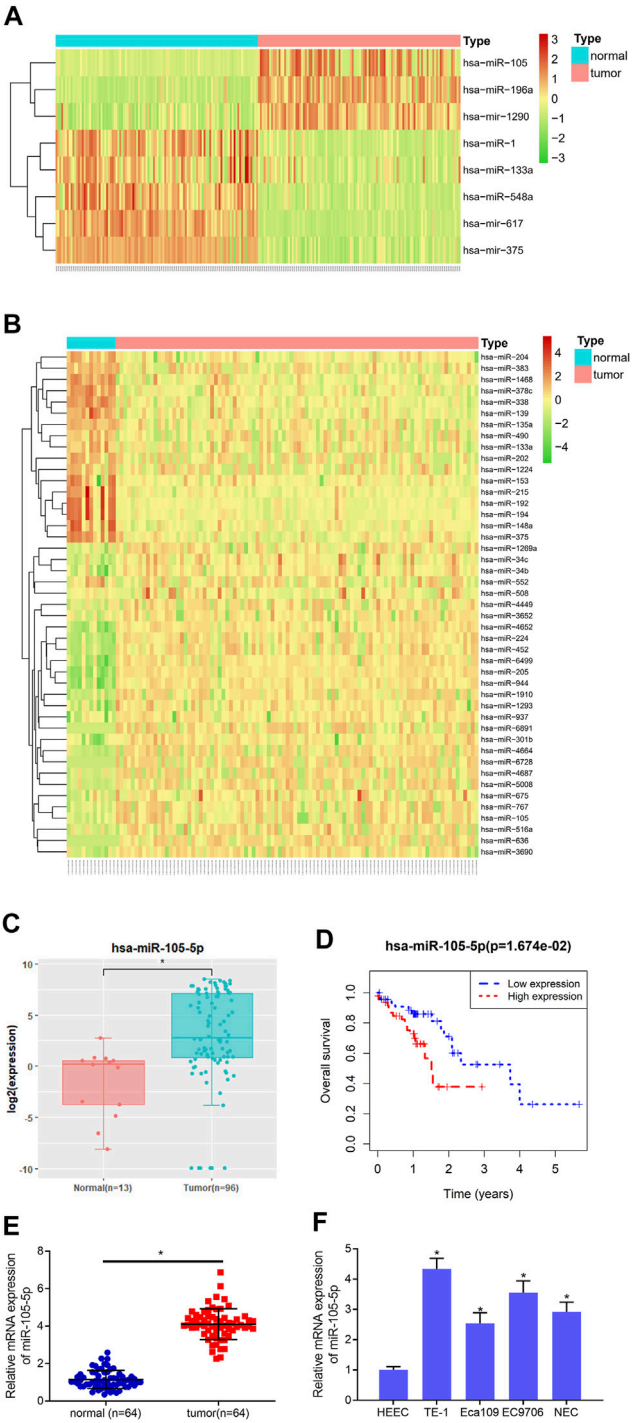


FIGURE 1 | MiRNA-105-5p is activated in ESCC. **(A)** Heatmap for DEmiRNAs in the GSE55856 miRNA chip; **(B)** Heatmap for DEmiRNAs in TCGA-ESCC miRNA dataset; **(C)** MiRNA-105-5p expression in TCGA-ESCC; **(D)** Survival analysis of miRNA-105-5p expression in TCGA-ESCC dataset; **(E)** MiRNA-105-5p expression in cancer and adjacent normal tissue; **(F)** MiRNA-105-5p expression in normal human esophageal epithelial cell line and ESCC cell lines; *$p < 0.05$.

RESULTS

MiRNA-105-5p Is Increased in ESCC Tissue and Cells

8 and 45 DEmiRNAs were obtained from the GSE55856 chip and TCGA-ESCC miRNA dataset, respectively, after differential analysis (**Figures 1A,B**), among which miRNA-105-5p was expressed highly in tumor tissue. Meanwhile, miRNA-105-5p was markedly elevated in tumor tissue in TCGA-ESCC (**Figure 1C**), and high miRNA-105-5p expression level indicated poor prognosis (**Figure 1D**). Hence, miRNA-105-5p was chosen. Then, qRT-PCR result suggested that the miRNA-105-5p level was stimulated in ESCC tissue than in normal tissue (**Figure 1E**). Similarly, cell lines implicated a similar result to tissue: miRNA-105-5p was noticeably increased in ESCC cell lines (**Figure 1F**) and most highly expressed in the TE-1 cell line. As a result, the TE-1 cell line was chosen for subsequent experiments. Taken together, we confirmed that miRNA-105-5p was highly expressed in ESCC and might be a factor indicating the unfavorable prognosis of ESCC patients.

MiRNA-105-5p Facilitates ESCC Cell Progression

Since it was discovered that patients with high expression of miRNA-105-5p had poor prognosis, we speculated that miRNA-105-5p was a predictor for the unfavorable prognosis of ESCC. In order to verify this speculation, firstly, miRNA-105-5p-inhibitor/miRNA-105-5p-mimic was transfected into TE-1 cells and the transfection efficiency was confirmed by qRT-PCR (**Figure 2A**). Then, CCK-8 assay results suggested that ESCC cell viability was reduced/improved upon miRNA-105-5p silencing/overexpression (**Figures 2B,C**). The results of wound healing assay and Transwell assay uncovered that silencing/overexpressing miRNA-105-5p inhibited/enhanced the cell migration and invasion of ESCC (**Figures 2D,E**). Collectively, the above experimental results unveiled that miRNA-105-5p facilitated ESCC cell processes and it was likely to be an oncogene of ESCC.

SPARCL1 is the Direct Target of miRNA-105-5p

To gain further insight into the mechanism by which miRNA-105-5p modulates ESCC, the downstream mRNA of miRNA-105-5p was explored. Bioinformatics databases including miRDB, mirDIP, and TargetScan were utilized to predict the target mRNAs of miRNA-105-5p, which were overlapped with 1,712 downregulated DEmRNAs from the TCGA-ESCC dataset (**Figure 3A**). About 10 DEmRNAs that had binding sites with miRNA-105-5p were obtained. Pearson correlation analysis uncovered that SPARCL1 was the strongest negatively correlated with miRNA-105-5p (**Figure 3B**). Meanwhile, SAPARCL1 was notably decreased in tumor tissue when compared with that in normal tissue (**Figure 3C**). A qRT-PCR result showed that SPARCL1 was poorly expressed in ESCC cancer tissue and cells, and the expression trend of miRNA-105-5p was opposite

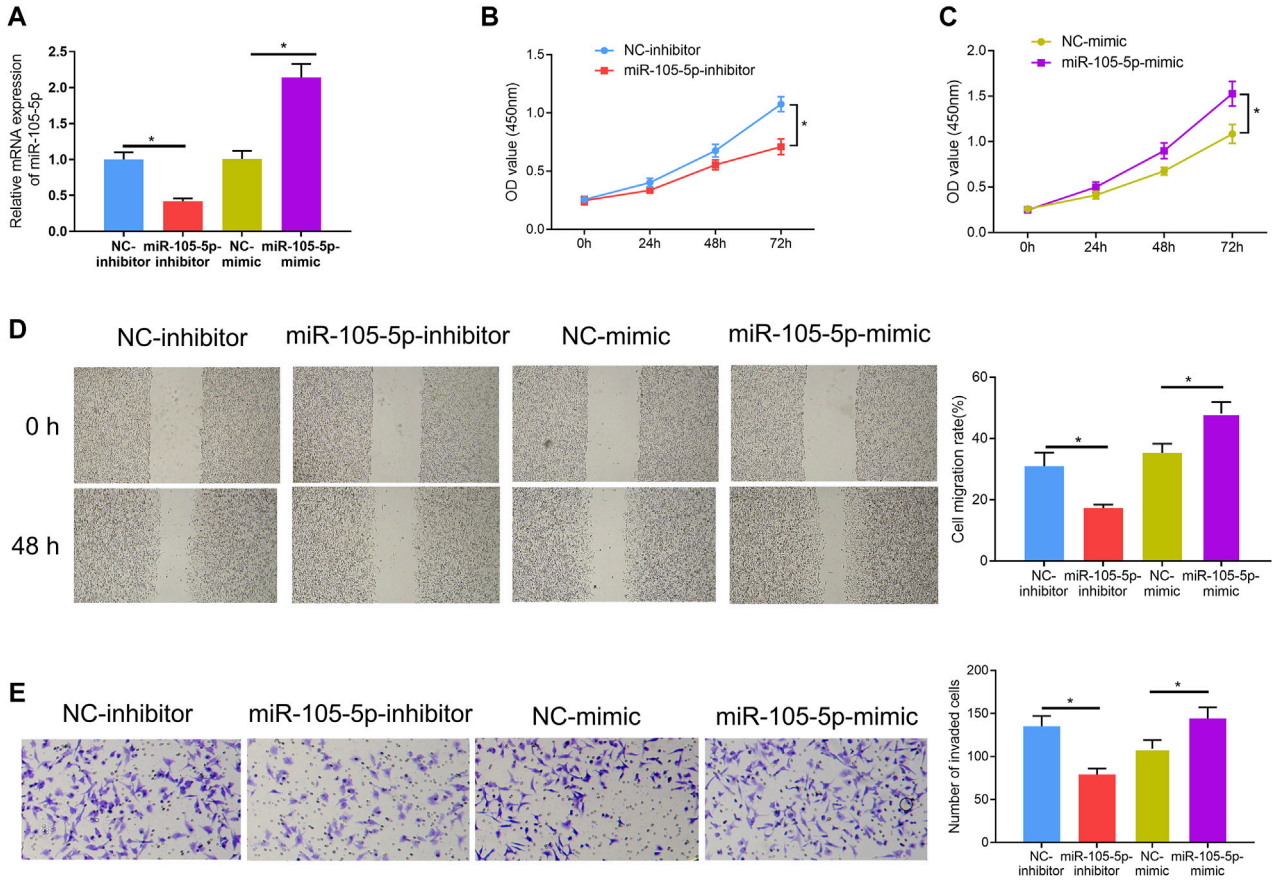


FIGURE 2 | MiRNA-105-5p fosters ESCC cell progression. **(A)** The efficiency of silencing or overexpressing miRNA-105-5p in ESCC cells; **(B,C)** Effect of silencing/overexpressing miRNA-105-5p on cell viability; **(D,E)** The effect of silenced/overexpressed miRNA-105-5p on cell migratory and invasive capacities of ESCC were assessed *via* wound healing assay (40×) **(D)** and Transwell invasion assay (100×) **(E)**, respectively; $^{*}p < 0.05$.

(**Figures 3D,E**). Therefore, we identified SPARCL1 as a potential target of miRNA-105-5p. Firstly, we predicted the binding sequence of miRNA-105-5p and SPARCL1 3′UTR through a bioinformatics database (**Figure 3F**). WT-SPARCL1 and MUT-SPARCL1 were co-transfected into TE-1 cells with miRNA-105-5p mimic and NC mimic. A Luciferase activity result showed that miRNA-105-5p overexpression decreased the luciferase activity of the WT SPARCL1 reporter but not the MUT reporter, indicating that SPARCL1 was the direct target of miRNA-105-5p (**Figure 3G**). To further verify the regulatory relationship between miRNA-105-5p and SPARCL1, we transfected miRNA-105-5p inhibitor/mimic into the TE-1 cell line. The result of qRT-PCR showed that SPARCL1 expression was significantly decreased/increased after the overexpression/silence of miRNA-105-5p in TE-1 cells (**Figure 3H**). Overall, it could be concluded that SPARCL1 was the direct target of miRNA-105-5p.

MiRNA-105-5p Modulates ESCC Cell Functions by Targeting SPARCL1

Rescue experiments were carried out to verify whether miRNA-105-5p could modulate ESCC cell processes *via* mediating SPARCL1. Firstly, miRNA-105-5p and SPARCL1 were overexpressed in TE-1 cells meanwhile, and the qRT-PCR result suggested that SPARCL1 expression in the miRNA-105-5p-mimic + oe-SPARCL1 group was higher than that in the miRNA-105-5p-mimic + oe-NC group (**Figure 4A**). Subsequently, CCK-8 assay revealed that SPARCL1 suppressed the proliferative ability of ESCC cells and counteracted the promoting effect of miRNA-105-5p on ESCC cell proliferation (**Figure 4B**). Similarly, wound healing and Transwell invasion assays uncovered that the facilitating effect of miRNA-105-5p on ESCC cell migration and invasion was attenuated by overexpressing SPARCL1 (**Figures 4C,D**). Collectively, these experiments unveiled that miRNA-105-5p could regulate ESCC cell processes through targeting SPARCL1.

SPARCL1 Inhibits ESCC Cell Behaviors by Modulating FAK/Akt Signaling Pathway

To observe the SPARCL1-related signaling pathway, we performed GSEA and further discovered that SPARCL1 was markedly enriched in FAK signaling (**Figure 5A**). Research unveiled that the FAK/Akt signaling pathway is closely

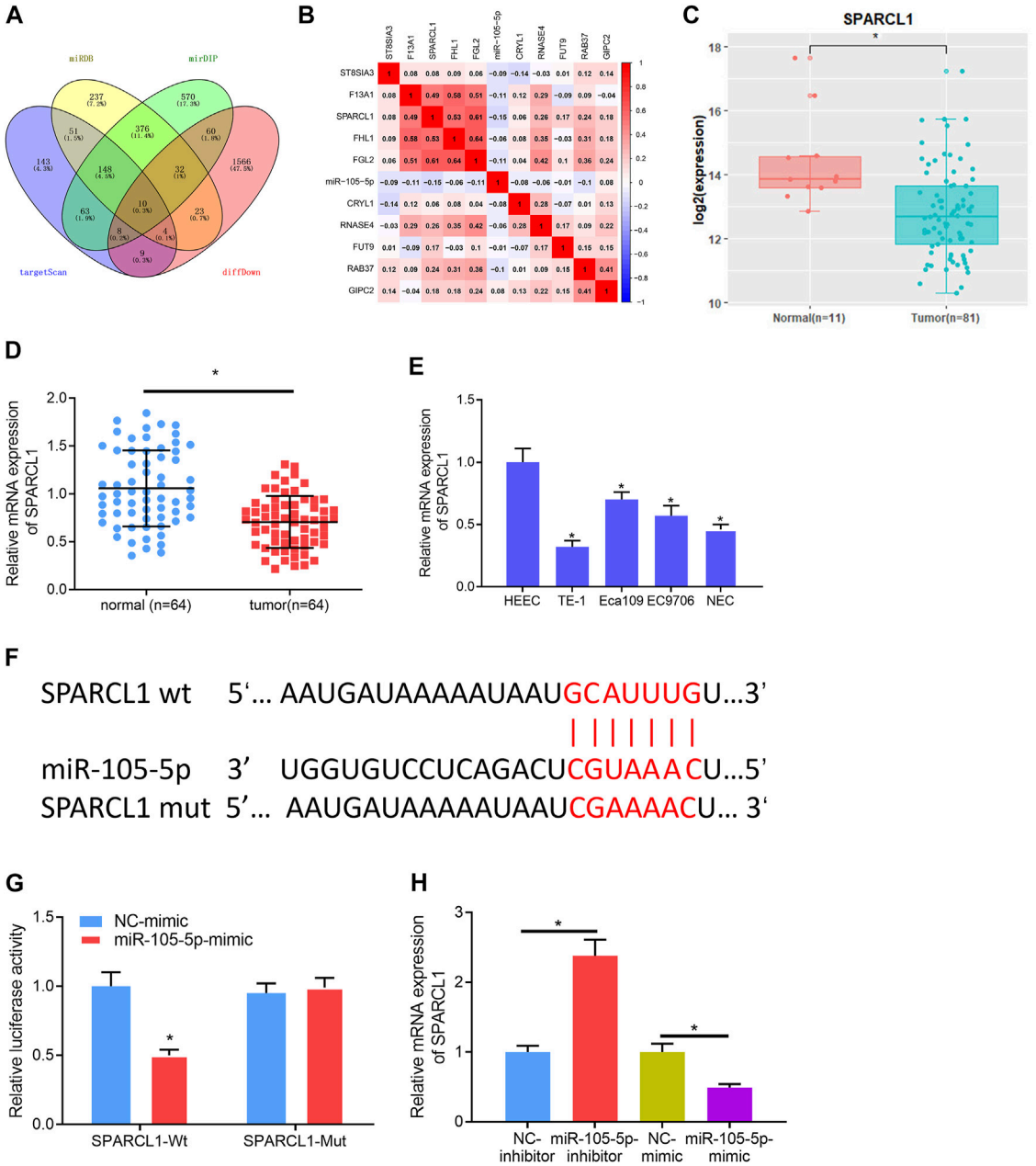


FIGURE 3 | SPARCL1 is the direct target of miRNA-105-5p. **(A)** Venn diagram of downregulated DEmRNAs and the predicted target mRNAs of miRNA-105-5p; **(B)** Correlation of miRNA-105-5p and the 10 overlapping mRNAs; **(C)** SPARCL1 expression in TCGA-ESCC dataset; **(D)** SPARCL1 expression in cancer and adjacent normal tissue; **(E)** SPARCL1 level in normal human esophageal epithelial cell line and ESCC cell lines; **(F)** Putative binding sites between miRNA-105-5p and SPARCL1; **(G)** The targeting relationship between miRNA-105-5p and SPARCL1; **(H)** The effect of silencing/overexpressing miRNA-105-5p in ESCC cells on SPARCL1 level; $^{*}p < 0.05$, $^{**}p < 0.01$.

associated with ESCC malignant progression (Meng et al., 2016; Zhu et al., 2018). Hence, we examined if there was a connection between SPARCL1 and FAK/Akt signaling. We observed that the levels of phosphorylated FAK and phosphorylated Akt were decreased in SPARCL1-overexpression cell lines, while total FAK and Akt levels remained the same (**Figure 5B**), which uncovered that SPARCL1 could suppress the activation of the FAK/Akt signaling pathway. Thereafter, ESCC cells with overexpressed SPARCL1 were processed by using *p*-Akt activator SC-79 (10 μM) (Wang et al., 2017) with PBS as control, after which the Western blot result suggested that the activity of the FAK/Akt signaling pathway inhibited by SPARCL1 was reversed by *p*-Akt activator SC-79 (**Figure 5C**). Finally, tumor-relevant behaviors of ESCC cells were observed in different treatment groups, and finding that adding SC-79 promoted ESCC cell proliferation, migration, and invasion and

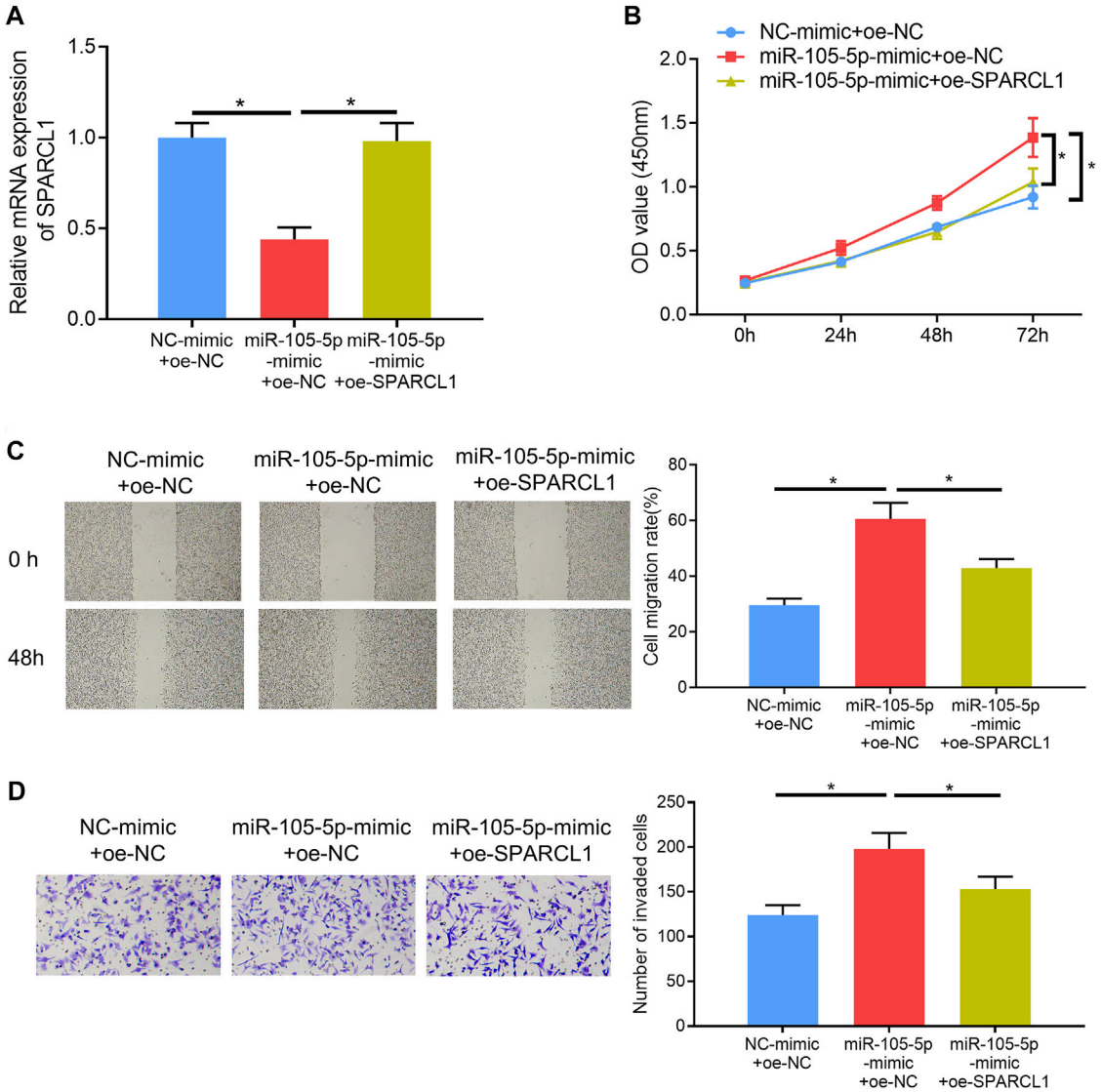


FIGURE 4 | MiRNA-105-5p regulates ESCC cell proliferation, migration, and invasion by targeting SPARCL1. **(A)** SPARCL1 expression in TE-1 cells transfected with NC-mimic + oe-NC, miRNA-105-5p-mimic + oe-SPARCL1, and miRNA-105-5p-mimic + oe-NC; **(B–D)** TE-1 cell proliferation, migration, and invasion in each group were tested *via* CCK-8 assay **(B)**, wound healing assay (×40) **(C),** and Transwell invasion assay (×100) **(D)**, respectively; $^*p < 0.05$.

counteracted the inhibitory effect of SPARCL1 on ESCC cells (**Figures 5D–F**). Collectively, we demonstrated that SPARCL1 was capable of repressing ESCC cell processes *via* inhibiting the FAK/Akt signaling pathway.

MiRNA-105-5p in EVs can Be Transferred to ESCC Cells

Studies indicated that exosomes and EVs could transfer miRNAs that they carry to cancer cells (Zheng et al., 2018; Wan et al., 2020). Accordingly, miRNA-105-5p expression was observed in different exosomes or EVs *via* the EVmiRNA database, finding that miRNA-105-5p mainly existed in serum exosomes and microvesicles, while EVs contained exosomes and microvesicles (**Figure 6A**). Hence, it was speculated that miRNA-105-5p in EVs could be transferred to ESCC cells. In order to verify this speculation, firstly, transmission electron microscope and Western blot were employed to assess the expression of EV-related proteins (TSG101, CD9, and CD63) to identify that EVs were successfully extracted from blood (**Figures 6B,C**). Then, the qRT-PCR result revealed that miRNA-105-5p existed in both the serum EVs of ESCC patients (case-EVs) and normal persons (control-EVs), and miRNA-105-5p expression in the EVs of ESCC patients was higher relatively (**Figure 6D**). Subsequently, Evs of ESCC patients were labeled using PKH67, and ESCC cells were labeled by Alexa Fluor® 594 Phalloidin and DAPI, after which the Evs and ESCC cells were co-cultured with PBS as the blank control. After co-

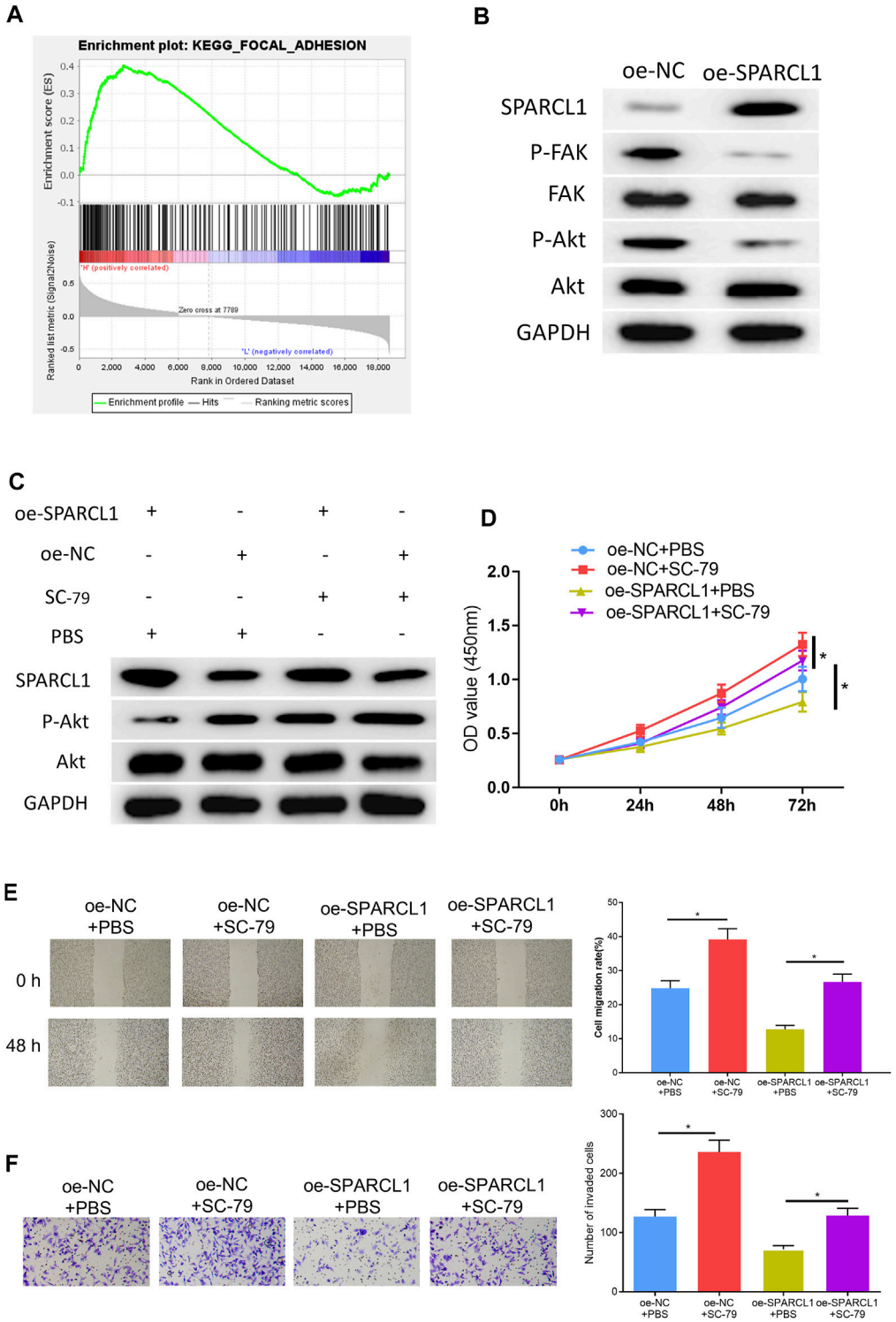


FIGURE 5 | SPARCL1 inhibits ESCC cell progression by modulating FAK/Akt signaling pathway. **(A)** GSEA result of SPARCL1; **(B)** The protein expression levels of SPARCL1 and FAK/Akt signaling pathway-related proteins in ESCC cells upon SPARCL1 overexpression; **(C)** The protein expression levels of SPARCL1 and FAK/Akt signaling pathway-related proteins in ESCC cells with an addition of *p*-Akt activator SC-79; **(D–F)** CCK-8 assay (D), wound healing assay (40×), and **(E)** Transwell invasion assay (×100) **(F)** were carried out to evaluate whether SPARCL1 suppresses ESCC cell proliferation, migration, and invasion *via* FAK/Akt signaling pathway; *$p < 0.05$.

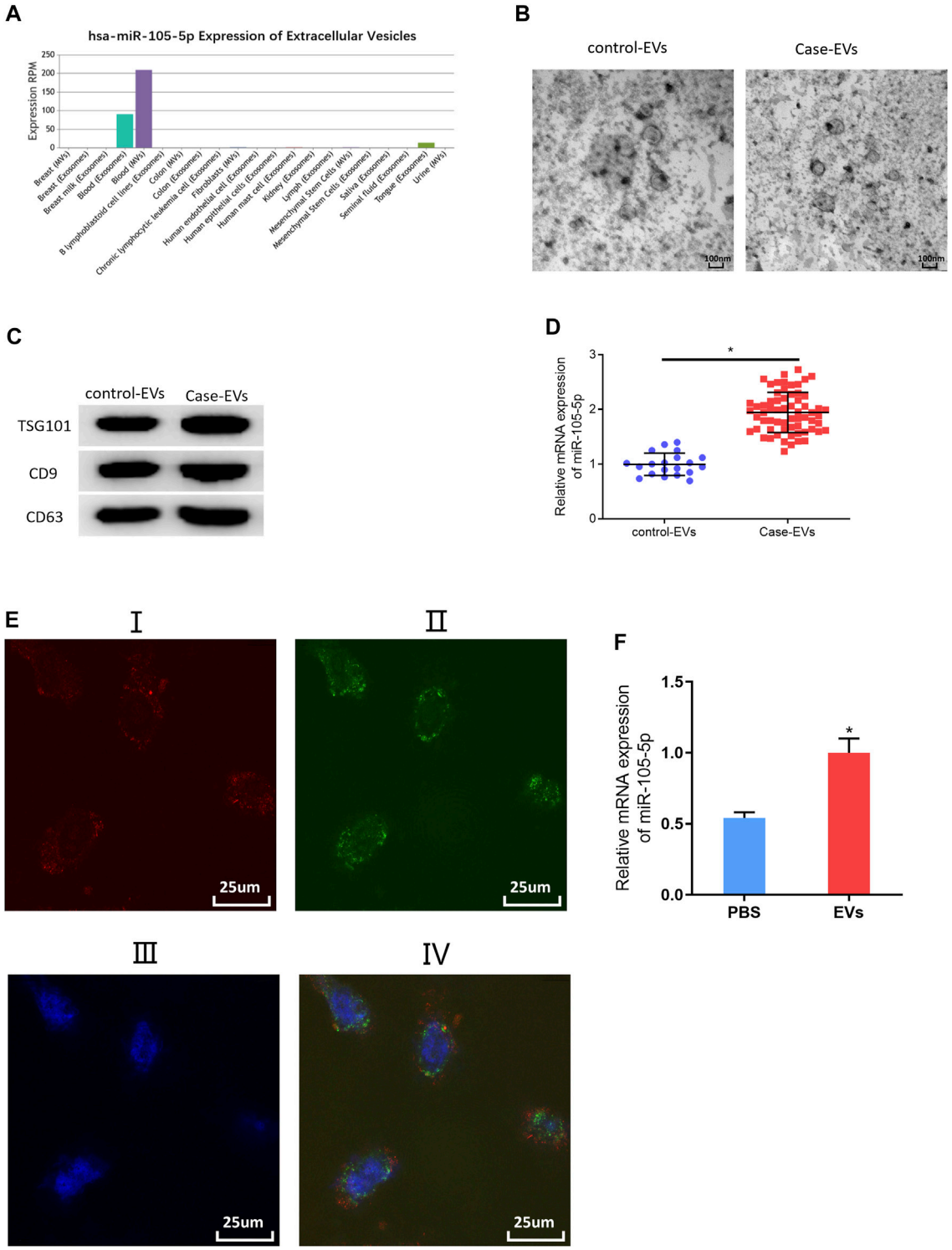


FIGURE 6 | MiRNA-105-5p in EVs can be transferred to ESCC cells. **(A)** Bar chart of miRNA-105-5p expression in different exosomes or microvesicles detected by the EVmiRNA database; **(B)** Transmission electron microscope was used to observe the forms of serum EVs of ESCC patients and normal persons (ratio: 100 nm); **(C)** Protein expression of EVs markers TSG101, CD9, and CD63 in serum EVs of ESCC patients and normal persons; **(D)** MiRNA-105-5p expression in serum EVs of ESCC patients and normal persons; **(E)** A fluorescence microscope was employed to detect whether serum EVs of ESCC patients could be internalized by ESCC cells: I. Alex Fluro594 phalloidin-labeled F-actin (red fluorescence); II. PKH67-labeled microvesicles (green fluorescence); III. DAPI-labeled cell nuclei (blue fluorescence); IV: Merge; Scale bar = 50 nm; **(F)** MiRNA-105-5p expression in ESCC cells after ESCC cells were co-cultured with serum EVs of ESCC patients. *$p < 0.05$.

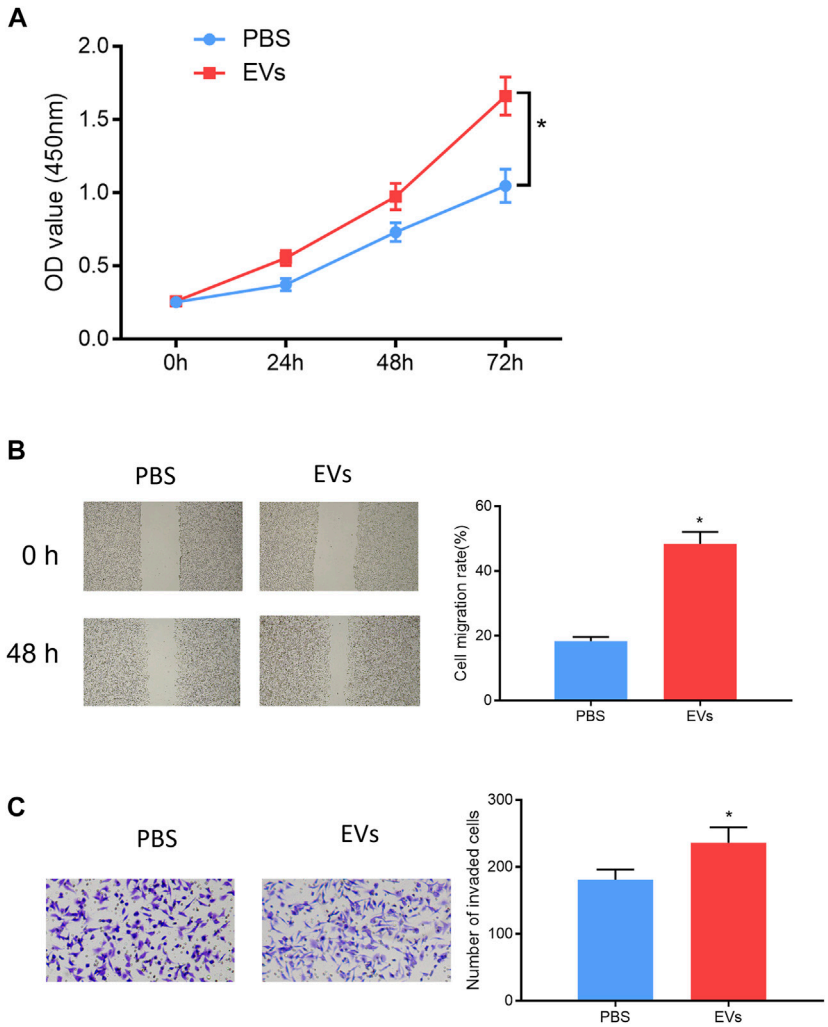


FIGURE 7 | Serum EV-derived miRNA-105-5p of ESCC patients accelerates cell processes. **(A–C)** CCK-8 assay **(A)**, wound healing assay (×40), /**(B)** and Transwell invasion assay (×100) **(C)** were performed to evaluate ESCC cell proliferation, migration, and invasion after ESCC cells were co-cultured with PBS or serum EVs of ESCC patients, respectively; *$p < 0.05$.

culture was finished, it was confirmed *via* a fluorescence microscope that the EVs of ESCC patients could be internalized by ESCC cells (**Figure 6E**). At last, qRT-PCR result indicated that miRNA-105-5p was conspicuously increased in ESCC cells (**Figure 6F**). Taken together, we confirmed that miRNA-105-5p in EVs could be transferred to ESCC cells.

Serum EVs-Derived miRNA-105-5p of ESCC Patients Promotes ESCC Cell Functions

In order to validate that EV-carried miRNA-105-5p could facilitate ESCC cell processes after entering ESCC cells, the ESCC cells that were co-cultured with PBS or serum EVs of ESCC patients were screened, and then cell biological experiments indicated that ESCC cell proliferation, migration, and invasion were noticeably fostered after ESCC cells were co-cultured with the serum EVs of ESCC patients (**Figures 7A–C**). Collectively, these experiments proved that EVs in the blood could promote ESCC cell progression after being internalized by ESCC cells.

Serum EVs-Derived miRNA-105-5p Fosters ESCC Cell Growth *In Vivo* Through Targeting SPARCL1 and Regulating FAK/Akt Signaling Pathway

ESCC cells were subcutaneously injected into the left hindlimb of nude mice. When the tumor volume of nude mice reached 100 mm^3, the serum EVs of ESCC patients were injected into the tumor of nude mice, with an equal amount of PBS as blank control. The injection was performed every 3 days. Three weeks later, the growth curves of the tumors of nude mice were drawn, and the tumors were weighed. The tumors formed by the serum

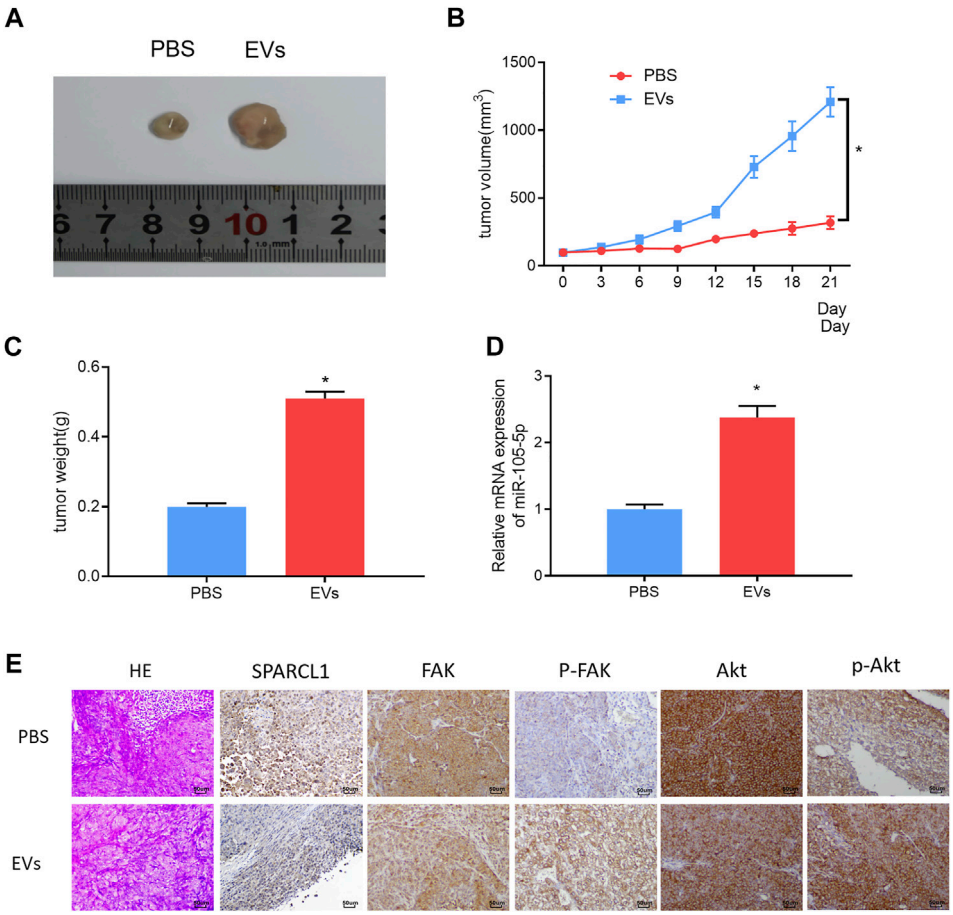


FIGURE 8 | Serum EV-derived miRNA-105-5p fosters ESCC cell growth *in vivo* by targeting SPARCL1 and regulating FAK/Akt signaling pathway. **(A)** Tumors of nude mice that were injected with PBS or serum EVs of ESCC patients; **(B)** Growth curves of tumors of nude mice; **(C)** Tumor weight of nude mice; **(D)** MiRNA-105-5p expression in tumors of nude mice; and **(E)** The protein expression of SPARCL1 and proteins related to FAK/Akt signaling pathway in tumors of nude mice (×200); $^{*}p < 0.05$.

EVs of ESCC patients grew more rapidly and were larger in size than those of the control group (**Figures 8A–C**). The qRT-PCR result revealed that miRNA-105-5p upregulation after tumors were injected with the serum EVs of ESCC patients (**Figure 8D**). Thereafter, immunohistochemistry was carried out to detect the protein levels of SPARCL1, FAK, *p*-FAK, Akt, and *p*-Akt in tumors; the result indicated that the protein expression of SPARCL1 in tumors injected with the serum EVs of ESCC patients was downregulated, whereas those of *p*-FAK and *p*-Akt were upregulated (**Figure 8E**). Collectively, from these experimental results, we identified that serum EV-derived miRNA-105-5p could foster ESCC cell growth *in vivo* by targeting SPARCL1 and modulating the FAK/Akt signaling pathway.

CONCLUSION

Studies have uncovered that the dysregulation of miRNAs can affect the progression of multiple cancers (Chen et al., 2020; Tan et al., 2020; Yehia et al., 2020). Here, it was discovered that miRNA-105-5p was prominently increased in ESCC and was likely to be an unfavorable factor for ESCC prognosis through bioinformatics analysis. A qRT-PCR result implicated that miRNA-105-5p was highly expressed in ESCC. Subsequently, a series of cell biological experiments further proved that miRNA-105-5p could promote the progression of ESCC cells. To sum up, these experimental results validate that miRNA-105-5p is a hopeful molecular target for ESCC treatment.

SPARCL1, also known as Hevin, MAST9, and SCI, is a family member of the SPARC proteins in matricellular proteins (Isler et al., 2001; Bertucci et al., 2004). Research revealed that SPARCL1 was an adhesion molecule mediating the cell–matrix interactions and got involved in physiological processes such as cell proliferation, cell adhesion, muscle differentiation, and B lymphocyte maturation (Bolshakov and Siegelbaum, 1995; Oritani and Kincade, 1998; Claeskens et al., 2000; Maak et al., 2001). With the research on SPARCL1 going deeper in recent years, it has been found to play a vital role in affecting progression of cancers through being regulated by miRNAs. For instance, targeting SPARCL1 by miRNA-539-3p facilitates the progression of epithelial ovarian cancer cells (Gong and Fan, 2019). Nevertheless, the role of SPARCL1 in ESCC has not

been reported yet. Herein, it was discovered that miRNA-105-5p had binding sites with SPARCL1 through bioinformatics methods. Given this finding, firstly, SPARCL1 expression in ESCC tissue and cell lines was detected, finding that SPARCL1 was poorly expressed. Meanwhile, dual-luciferase reporter assay and qRT-PCR confirmed that miRNA-105-5p could target SPARCL1 and regulate its expression. Additionally, a series of *in vitro* experiments validated that miRNA-105-5p was capable of fostering ESCC cell progression *via* mediating SPARCL1.

FAK is located in the focal adhesion that forms between cells growing with extracellular matrix constituents. FAK can activate the FAK/Akt signaling pathway *via* receiving signals from integrin, fibrin, etc. Activated FAK can combine with Scr to form a complex, and the Tyr397 of activated FAK can directly combine with the SH2 domain of P13K to activate P13K. Activated P13K is capable of activating Akt so as to regulate cell growth (Gan et al., 2006). Currently, loads of studies have reported that mRNA can regulate ESCC cell proliferation and metastasis *via* the FAK/Akt signaling pathway. For example, LOXL2 fosters the tumorigenesis of head and neck squamous cell carcinoma through FAK/Akt signaling (Liu et al., 2020). Here, the GSEA result implicated that SPARCL1 was enriched in the FAK/Akt signaling pathway. Based on this, the Western blot result suggested that SPARCL1 was able to suppress the FAK/Akt signaling pathway. Thereafter, ESCC cells with SPARCL1 overexpression were processed using *p*-Akt activator, and cell biological experiments were then conducted to evaluate ESCC cell behaviors. Taken together, we proved that SPARCL1 was capable of repressing the progression of ESCC cells through deactivating the FAK/Akt signaling pathway.

Quite a few studies have demonstrated that EVs can be used to treat tumors in recent years (Zheng et al., 2017; Wang et al., 2019). Therefore, miRNA-105-5p expression was observed *via* the EVmiRNA database, finding that miRNA-105-5p mainly existed in serum exosomes and Evs. Then, the serum Evs of ESCC patients and normal persons were obtained. qRT-PCR revealed that miRNA-105-5p existed in blood and that the serum Evs of ESCC patients could be internalized by ESCC cells, leading to elevated miRNA-105-5p expression in ESCC cells, ultimately facilitating ESCC cell behaviors. Subcutaneous transplantation xenograft models are widely applied to the study on the treatment of tumors with miRNAs in Evs (Wang et al., 2020; Zhao et al., 2020). In this study, it was also confirmed that serum EV-derived miRNA-105-5p could promote tumor growth by employing subcutaneous transplantation tumor models in nude mice. Collectively, the above experimental results fully verified that serum EV-derived miRNA-105-5p can be transferred to ESCC cells to foster the tumorigenesis of ESCC.

Altogether, we demonstrated that serum EV-derived miRNA-105-5p could be transferred to ESCC cells and foster the progression of ESCC by targeting SPARCL1 and regulating the FAK/Akt signaling pathway. The discovery of this functional mechanism will supply a rationale for ESCC therapy. Furthermore, how miRNA-105-5p regulates the expression of other target genes in ESCC or in a broad scope of other cancers remains to be further studied in the near future. This study also had certain defects, like this study carried out cell function verification only using one cell line, and we will use other ESCC cell lines to further verify the results of this 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.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The studies were reviewed and approved by the Academic Ethics Committee of Shaoxing People's Hospital (2020-32). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

BH contributed to the conception of the study and helped perform the analysis with constructive discussions. KZ performed the experiment. XH contributed significantly to analysis and manuscript preparation. CS and JZ performed the data analyses. GW and GW wrote the manuscript. LZ and WH gave the final approval of the version to be submitted.

FUNDING

This study was funded by the Shaoxing Science and Technology Project (No. 2020A13059).

REFERENCES

Abels, E. R., and Breakefield, X. O. (2016). Introduction to Extracellular Vesicles: Biogenesis, RNA Cargo Selection, Content, Release, and Uptake. *Cell Mol Neurobiol* 36, 301–312. doi:10.1007/s10571-016-0366-z

Bertucci, F., Salas, S., Eysteries, S., Nasser, V., Finetti, P., Ginestier, C., et al. (2004). Gene Expression Profiling of colon Cancer by DNA Microarrays and Correlation with Histoclinical Parameters. *Oncogene* 23, 1377–1391. doi:10.1038/sj.onc.1207262

Bolshakov, V. Y., and Siegelbaum, S. A. (1995). Hippocampal Long-Term Depression: Arachidonic Acid as a Potential Retrograde Messenger. *Neuropharmacology* 34, 1581–1587. doi:10.1016/0028-3908(95)00127-r

Caivano, A., Laurenzana, I., De Luca, L., La Rocca, F., Simeon, V., Trino, S., et al. (2015). High Serum Levels of Extracellular Vesicles Expressing Malignancy-Related Markers Are Released in Patients with Various Types of Hematological Neoplastic Disorders. *Tumor Biol.* 36, 9739–9752. doi:10.1007/s13277-015-3741-3

Chen, J. Y., Xu, L. F., Hu, H. L., Wen, Y. Q., Chen, D., and Liu, W. H. (2020). MiRNA-215-5p Alleviates the Metastasis of Prostate Cancer by Targeting PGK1. *Eur. Rev. Med. Pharmacol. Sci.* 24, 639–646. doi:10.26355/eurrev_202001_20040

Claeskens, A., Ongenae, N., Neefs, J. M., Cheyns, P., Kaijen, P., Cools, M., et al. (2000). Hevin Is Down-Regulated in many Cancers and Is a Negative Regulator of Cell Growth and Proliferation. *Br. J. Cancer* 82, 1123–1130. doi:10.1054/bjoc.1999.1051

Gan, B., Yoo, Y., and Guan, J.-L. (2006). Association of Focal Adhesion Kinase with Tuberous Sclerosis Complex 2 in the Regulation of S6 Kinase Activation and Cell Growth. *J. Biol. Chem.* 281, 37321–37329. doi:10.1074/jbc.M605241200

Gong, Y. B., and Fan, X. H. (2019). MiR-539-3p Promotes the Progression of Epithelial Ovarian Cancer by Targeting SPARCL1. *Eur. Rev. Med. Pharmacol. Sci.* 23, 2366–2373. doi:10.26355/eurrev_201903_17381

Isler, S., Schenk, S., Bendik, I., Schraml, P., Novotna, H., Moch, H., et al. (2001). Genomic Organization and Chromosomal Mapping of SPARC-like 1, a Gene Down Regulated in Cancers. *Int. J. Oncol.* 18, 521–526. doi:10.3892/ijo.18.3.521

Jin, Y., Meng, Q., Zhang, B., Xie, C., Chen, X., Tian, B., et al. (2021). Cancer-associated Fibroblasts-Derived Exosomal miR-3656 Promotes the Development and Progression of Esophageal Squamous Cell Carcinoma via the ACAP2/PI3K-AKT Signaling Pathway. *Int. J. Biol. Sci.* 17, 3689–3701. doi:10.7150/ijbs.62571

Lin, Y., Totsuka, Y., He, Y., Kikuchi, S., Qiao, Y., Ueda, J., et al. (2013). Epidemiology of Esophageal Cancer in Japan and China. *J. Epidemiol.* 23, 233–242. doi:10.2188/jea.je20120162

Liu, C., Guo, T., Sakai, A., Ren, S., Fukusumi, T., Ando, M., et al. (2020). A Novel Splice Variant of LOXL2 Promotes Progression of Human Papillomavirus-Negative Head and Neck Squamous Cell Carcinoma. *Cancer* 126, 737–748. doi:10.1002/cncr.32610

Maak, S., Jaesert, S., Neumann, K., Yerle, M., and von Lengerken, G. (2001). Isolation of Expressed Sequence Tags of Skeletal Muscle of Neonatal Healthy and Splay Leg Piglets and Mapping by Somatic Cell Hybrid Analysis. *Anim. Genet.* 32, 303–307. doi:10.1046/j.1365-2052.2001.00778.x

Meng, X., Chen, X., Lu, P., Ma, W., Yue, D., Song, L., et al. (2016). MicroRNA-202 Inhibits Tumor Progression by Targeting LAMA1 in Esophageal Squamous Cell Carcinoma. *Biochem. Biophysical Res. Commun.* 473, 821–827. doi:10.1016/j.bbrc.2016.03.130

Oritani, K., and Kincade, P. W. (1998). Lymphopoiesis and Matrix Glycoprotein SC1/ECM2. *Leuk. Lymphoma* 32, 1–7. doi:10.3109/10428199809059241

Priniski, S. J., Rosenzweig, E. Q., Canning, E. A., Hecht, C. A., Tibbetts, Y., Hyde, J. S., et al. (2019). The Benefits of Combining Value for the Self and Others in Utility-Value Interventions. *J. Educ. Psychol.* 111, 1478–1497. doi:10.1037/edu0000343

Qi, Y., Zha, W., and Zhang, W. (2019). Exosomal miR-660-5p Promotes Tumor Growth and Metastasis in Non-small Cell Lung Cancer. *J. BUON* 24, 599–607.

Sawada, R., Maehara, R., Oshikiri, T., Nakamura, T., Itoh, T., Kodama, Y., et al. (2019). MDM2 Copy Number Increase: a Poor Prognostic, Molecular Event in Esophageal Squamous Cell Carcinoma. *Hum. Pathol.* 89, 1–9. doi:10.1016/j.humpath.2019.04.002

Sheng, S., Xie, L., Wu, Y., Ding, M., Zhang, T., and Wang, X. (2019). MiR-144 Inhibits Growth and Metastasis in colon Cancer by Down-Regulating SMAD4. *Biosci. Rep.* 39, 2. doi:10.1042/bsr20181895

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

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

Tan, T., Xu, X. H., Lu, X. H., and Wang, X. W. (2020). MiRNA-200a-3p Suppresses the Proliferation, Migration and Invasion of Non-small Cell Lung Cancer through Targeting IRS2. *Eur. Rev. Med. Pharmacol. Sci.* 24, 712–720. doi:10.26355/eurrev_202001_20050

Tanaka, Y., Kamohara, H., Kinoshita, K., Kurashige, J., Ishimoto, T., Iwatsuki, M., et al. (2013). Clinical Impact of Serum Exosomal microRNA-21 as a Clinical Biomarker in Human Esophageal Squamous Cell Carcinoma. *Cancer* 119, 1159–1167. doi:10.1002/cncr.27895

Wan, F.-Z., Chen, K.-H., Sun, Y.-C., Chen, X.-C., Liang, R.-B., Chen, L., et al. (2020). Exosomes Overexpressing miR-34c Inhibit Malignant Behavior and Reverse the Radioresistance of Nasopharyngeal Carcinoma. *J. Transl Med.* 18, 12. doi:10.1186/s12967-019-02203-z

Wang, H., Wei, H., Wang, J., Li, L., Chen, A., and Li, Z. (2020). MicroRNA-181d-5p-Containing Exosomes Derived from CAFs Promote EMT by Regulating CDX2/HOXA5 in Breast Cancer. *Mol. Ther. - Nucleic Acids* 19, 654–667. doi:10.1016/j.omtn.2019.11.024

Wang, R., Wang, W., Ao, L., Wang, Z., Hao, X., and Zhang, H. (2017). Benzo[a]pyrene-7,8-diol-9,10-epoxide Suppresses the Migration and Invasion of Human Extravillous Trophoblast HTR-8/SVneo Cells by Down-Regulating MMP2 through Inhibition of FAK/SRC/PI3K/AKT Pathway. *Toxicology* 386, 72–83. doi:10.1016/j.tox.2017.05.008

Wang, T., Luo, Y., Lv, H., Wang, J., Zhang, Y., and Pei, R. (2019). Aptamer-Based Erythrocyte-Derived Mimic Vesicles Loaded with siRNA and Doxorubicin for the Targeted Treatment of Multidrug-Resistant Tumors. *ACS Appl. Mater. Inter.* 11, 45455–45466. doi:10.1021/acsami.9b16637

Wei, S., Liu, K., He, Q., Gao, Y., and Shen, L. (2019). PES1 Is Regulated byCD44 in Liver Cancer Stem Cells via miR-105-5p. *FEBS Lett.* 593, 1777–1786. doi:10.1002/1873-3468.13459

Wu, M., Kong, C., Cai, M., Huang, W., Chen, Y., Wang, B., et al. (2021). Hsa_circRNA_002144 Promotes Growth and Metastasis of Colorectal Cancer through Regulating miR-615-5p/LARP1/mTOR Pathway. *Carcinogenesis* 42, 601–610. doi:10.1093/carcin/bgaa140

Xu, L.-J., Duan, Y., Wang, P., and Yin, H.-Q. (2018). MiR-199b-5p Promotes Tumor Growth and Metastasis in Cervical Cancer by Down-Regulating KLK10. *Biochem. biophysical Res. Commun.* 503, 556–563. doi:10.1016/j.bbrc.2018.05.165

Xu, R., Greening, D. W., Zhu, H.-J., Takahashi, N., and Simpson, R. J. (2016). Extracellular Vesicle Isolation and Characterization: toward Clinical Application. *J. Clin. Invest.* 126, 1152–1162. doi:10.1172/JCI81129

Yehia, L., Seyfi, M., Niestroj, L.-M., Padmanabhan, R., Ni, Y., Frazier, T. W., et al. (2020). Copy Number Variation and Clinical Outcomes in Patients with GermlinePTENMutations. *JAMA Netw. Open* 3, e1920415. doi:10.1001/jamanetworkopen.2019.20415

Zeng, W., Zhu, J.-F., Liu, J.-Y., Li, Y.-L., Dong, X., Huang, H., et al. (2019). miR-133b Inhibits Cell Proliferation, Migration and Invasion of Esophageal Squamous Cell Carcinoma by Targeting EGFR. *Biomed. Pharmacother.* 111, 476–484. doi:10.1016/j.biopha.2018.12.057

Zhao, G., Li, H., Guo, Q., Zhou, A., Wang, X., Li, P., et al. (2020). Exosomal Sonic Hedgehog Derived from Cancer-associated Fibroblasts Promotes Proliferation and Migration of Esophageal Squamous Cell Carcinoma. *Cancer Med.* 9, 2500–2513. doi:10.1002/cam4.2873

Zheng, R., Du, M., Wang, X., Xu, W., Liang, J., Wang, W., et al. (2018). Exosome-transmitted Long Non-coding RNA PTENP1 Suppresses Bladder Cancer Progression. *Mol. Cancer* 17, 143. doi:10.1186/s12943-018-0880-3

Zheng, Y., Tang, L., Mabardi, L., Kumari, S., and Irvine, D. J. (2017). Enhancing Adoptive Cell Therapy of Cancer through Targeted Delivery of Small-Molecule Immunomodulators to Internalizing or Noninternalizing Receptors. *ACS Nano* 11, 3089–3100. doi:10.1021/acsnano.7b00078

Zhu, X., Wang, J., Li, L., Deng, L., Wang, J., Liu, L., et al. (2018). GPX3 Suppresses Tumor Migration and Invasion via the FAK/AKT Pathway in Esophageal Squamous Cell Carcinoma. *Am. J. Transl Res.* 10, 1908–1920.

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 Genetics

ORIGINAL RESEARCH
published: 13 May 2022
doi: 10.3389/fgene.2022.894865

Identification of Immunological Characteristics and Immune Subtypes Based on Single-Sample Gene Set Enrichment Analysis Algorithm in Lower-Grade Glioma

Yunyang Zhu[1†], *Songwei Feng*[2†], *Zhaoming Song*[1], *Zhong Wang*[1] *and Gang Chen*[3*]

[1]Department of Neurosurgery, The First Affiliated Hospital of Soochow University, Suzhou, China, [2]Department of Obstetrics and Gynaecology, Zhongda Hospital, School of Medicine, Southeast University, Nanjing, China, [3]Department of Neurosurgery and Brain and Nerve Research Laboratory, The First Affiliated Hospital of Soochow University, Suzhou, China

OPEN ACCESS

Edited by:
Tao Huang,
Shanghai Institute of Nutrition and Health (CAS), China

Reviewed by:
Yongbiao Huang,
Huazhong University of Science and Technology, China
Jiaxi Lu,
Chongqing University, China
Chuanzhao Zhang,
Guangdong Provincial People's Hospital, China

***Correspondence:**
Gang Chen
nju_neurosurgery@163.com

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

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: 12 March 2022
Accepted: 08 April 2022
Published: 13 May 2022

Citation:
Zhu Y, Feng S, Song Z, Wang Z and Chen G (2022) Identification of Immunological Characteristics and Immune Subtypes Based on Single-Sample Gene Set Enrichment Analysis Algorithm in Lower-Grade Glioma. Front. Genet. 13:894865. doi: 10.3389/fgene.2022.894865

Few breakthroughs have been achieved in the treatment of lower-grade glioma (LGG) in recent decades. Apart from the conventional pathological and histological classifications, subtypes based on immunogenomics would provide reference for individualized treatment and prognosis prediction. Our study identified four immunotypes of lower-grade glioma (clusters A, B, C, and D) by bioinformatics methods in TCGA-LGG and two CGGA datasets. Cluster A was an "immune-cold" phenotype with the lowest immune infiltration and longest survival expectation, whereas cluster D was an "immune-rich" subtype with the highest immune infiltration and poor survival expectation. The expression of immune checkpoints increased along with immune infiltration degrees among the clusters. It was notable that immune clusters correlated with a variety of clinical and immunogenomic factors such as age, WHO grades, IDH1/2 mutation, PTEN, EGFR, ATRX, and TP53 status. In addition, LGGs in cluster D were sensitive to cisplatin, gemcitabine, and immune checkpoint PD-1 inhibitors. RTK-RAS and TP53 pathways were affected in cluster D. Functional pathways such as cytokine–cytokine receptor interaction, antigen processing and presentation, cell adhesion molecules (CAMs), and ECM–receptor interaction were also enriched in cluster D. Hub genes were selected by the Matthews correlation coefficient (MCC) algorithm in the blue module of a gene co-expression network. Our studies might provide an immunogenomics subtyping reference for immunotherapy in LGG.

Keywords: lower-grade glioma, immunogenomics, immune clusters, glioma, tumor-immune microenvironment

INTRODUCTION

Lower-grade gliomas were previously regarded as World Health Organization (WHO) grade I and grade II gliomas in contrast to high-grade gliomas. Nowadays, the concept of diffuse lower-grade gliomas (LGGs), which refer to WHO grades II and III astrocytomas, oligodendrogliomas, and oligoastrocytomas (Eckel-Passow et al., 2015; Zeng et al., 2018), was applied to better define the slowly invasive and relatively indolent progression features. With a 10-year median survival, nearly 70% of LGG patients tend to gradually transform into high-grade glioma patients in which the tumor

immune microenvironment (TIM) and genetic changes play vital roles (Appolloni et al., 2019; Nejo et al., 2019). The prognosis of LGG could hardly be predicted accurately by conventional pathological and histological classifications; thus, new subtypes based on immunogenomics are urgently required.

Classifications based on molecular and genetic markers had been adapted since the 2016 WHO classification of the central nervous system tumors (Louis et al., 2016) and were emphasized in the 2021 version (Louis et al., 2021). Isocitrate dehydrogenase (IDH) mutation plays an important role in altering the tumor immune microenvironment. The decrease of PD-L1 in IDH mutation gliomas means a strong T-cell activation (Han et al., 2020). In LGG, oligodendroglioma and astrocytoma were further classified based on 1p/19q co-deletion, p53 mutation, alpha-thalassemia/intellectual disability syndrome X-linked (ATRX) mutation, and telomerase reverse transcriptase (TERT) promoter mutation in IDH1/2-mutant gliomas (Ohba et al., 2020). The tumor microenvironment (TME) of glioma could be delineated by infiltrating immunocytes and genetic landscapes. It has been reported that the innate immune cells would be manipulated and reprogrammed in the TME to facilitate the tumorigenesis, progression, and spread and subsequently lead to tumor immune evasion in gliomas (Zindl and Chaplin, 2010; Hinshaw and Shevde, 2019). Tumor mutational burden (TMB), which is closely correlated with immune infiltration, consists of the base substitutions, insertion, or deletion mutations of the whole exome. A study had classified LGG patients into two groups based on TMB and found that the infiltration of B lineage, CD4 T cells, CD8 T cells, neutrophils, macrophages, and dendritic cells would lead to shorter overall survival along with the high expression of immune checkpoints PD-1 and CTLA-4 (Yin et al., 2020).

Immune checkpoint inhibitors had promising therapeutic effects in several tumors (Lipson et al., 2015; Emens et al., 2017). The inhibition of PD-1 and CTLA-4 could notably enhance immunosurveillance and prolong the survival time in mouse glioma models (Wainwright et al., 2014; Xu et al., 2020). However, the clinical application remains challenging because of the "cold phenotype" of glioma (Qi et al., 2020). Our study would seek for the most suitable immunotyping for immune checkpoint blocking therapy.

Here, four immune clusters of LGG based on The Cancer Genome Atlas (TCGA) immune datasets were identified and then verified in two Chinese Glioma Genome Atlas (CGGA) datasets. The immune cell infiltration features, hub genes, potential drugs, and prognosis were studied by bioinformatics methods. This study might offer reference for immunogenomics subtyping for individualized LGG therapy.

METHODS

Data Processing

RNA-seq data (level-3, HTseq-FPKM) and clinical information of lower-grade glioma (LGG, grade II–III) samples were obtained from The Cancer Genome Atlas (TCGA) dataset. A total of 481 samples were finally selected after removing samples with no survival state, no WHO grade, recurrent tumor, reduplicated sequencing, and whose survival time were less than 1 day. In addition, RNA-seq and clinicopathological data were obtained from the Chinese Glioma Genome Atlas (CGGA) website as the validation set. A total of 332 samples with complete survival information were chosen from the CGGA-693 dataset (CGGA-LGG-1) and 162 samples were obtained from the CGGA-325 dataset (CGGA-LGG-2). The batch effect correction was performed by the R package termed "SVA."

Identification of Immune-Related Clusters in Lower-Grade Gliomas

Single-sample Gene Set Enrichment Analysis (ssGSEA) was conducted in the three datasets based on the expression level of 29 immunity-associated signatures by the R package "GSVA." Consensus clustering was then applied to define the immune subgroups based on the ssGSEA scores by the "ConsensusClusterPlus" package in R. The K-means clustering algorithm was performed with 100 resampling iterations by random selection of 80% of the total samples to ensure the clustering stability. The best cluster number was determined by the consensus matrix (CM) heat maps, cumulative distribution function (CDF) curves, and delta area score of CDF curves. A principal component analysis (PCA) was used to illustrate the reliability of optimal methods. Thorsson et al. (2018) had identified six immune function subtypes by an extensive immunogenomic analysis: wound healing (C1), IFN-γ dominant (C2), inflammatory (C3), lymphocyte depleted (C4), immunologically quiet (C5), and TGF-β response (C6). A Sankey plot was applied to visualize the relationships between our four clusters and the six identified immune functional subtypes mentioned previously.

Features of Immune Cell Clusters in the Immune Microenvironment of Lower-Grade Gliomas

The Estimation of Stromal and Immune cells in Tumors using Expression data (ESTIMATE) algorithm was used to evaluate the LGG microenvironment (Yoshihara et al., 2013). Immune scores and stromal scores were calculated to reveal the abundance of infiltrating immune and stromal cells. ESTIMATE scores were calculated for reflecting non-tumor composites. Tumor purity was inferred by the aforementioned scores. The Kruskal–Wallis test was used to compare differences in multiple clusters. Heat maps were drawn by the "pheatmap" package in R.

Estimation of Immune Cell Infiltration

The Microenvironment Cell Populations-counter (MCP counter) algorithm (Becht et al., 2016) was used to quantitate the abundance of immunocytes in heterogeneous tissues by the "MCPcounter" R package. The absolute abundances of two stromal cells and eight immune cells were evaluated by immune cell scores, including T cells, CD8 T cells, cytotoxic lymphocytes, B lineage, NK cells, monocytic lineage, myeloid dendritic cells, neutrophils, endothelial cells, and fibroblasts.

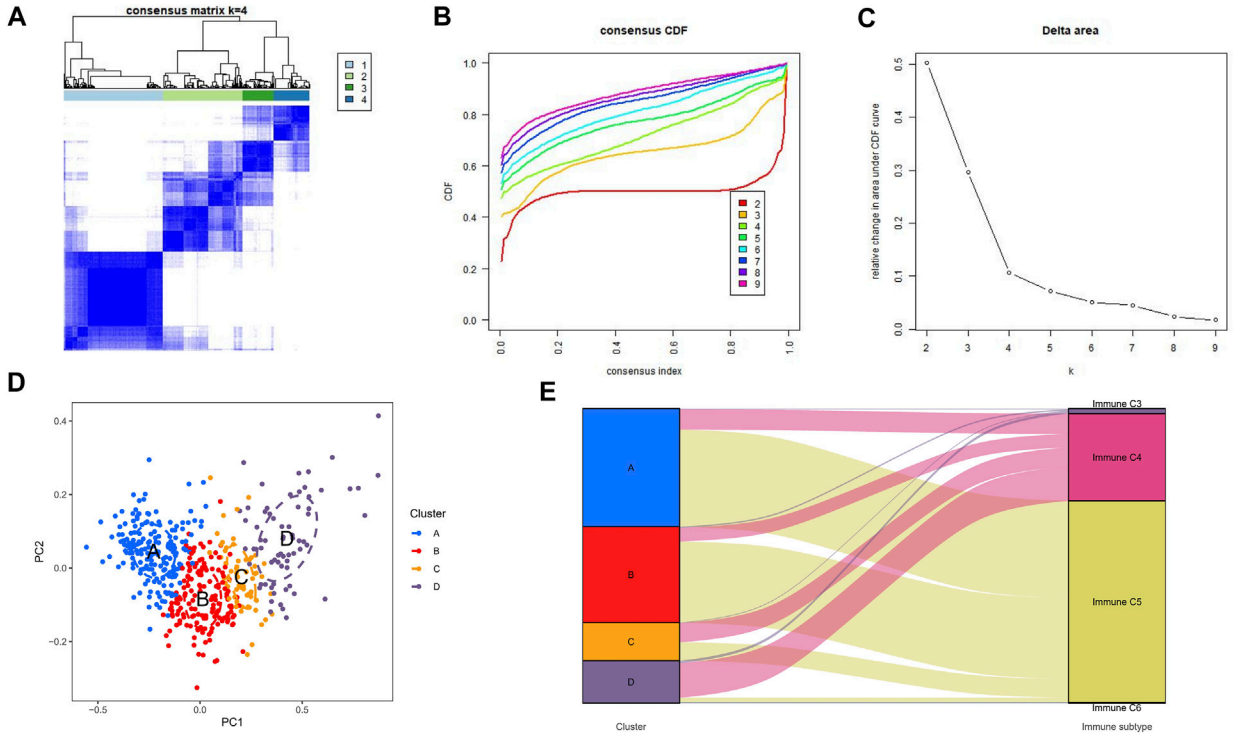


FIGURE 1 | Consensus clustering results in TCGA-LGG cohorts. **(A)** Consensus clustering matrix of $k = 4$ as the optimal cluster number. **(B)** CDF curves of the consensus score from $k = 2$–9. **(C)** Delta area under the CDF curve. **(D)** Principal component analysis for the ssGSEA scores of four clusters. Each subgroup was distinguished by different colors. **(E)** Sankey diagram to visualize the relationships between our four clusters and the six identified immune function subtypes.

Prediction of Potential Drugs

The SubMap analysis (Hoshida et al., 2007; Roh and Chen, 2017) from Gene Pattern (https://www.genepattern.org/) was used to predict the response to immune checkpoint blockade (anti-PD-1 and anti-CTLA-4 immunotherapy). In addition, the chemotherapy response was predicted based on the public pharmacogenomics database termed "Genomics of Drug Sensitivity in Cancer" (GDSC, http://www.cancerrxgene.org). The half-maximum inhibitory concentration (IC50) of each patient was estimated by the R package "pRRophetic" with Ridge's regression, and the accuracy of the prediction was estimated by a 10-fold cross-validation. The IC50 of each sample in TCGA dataset was calculated based on the prediction models of bleomycin and doxorubicin, and cisplatin and gemcitabine.

Gene Set Enrichment Analysis

The GSEA algorithm was used to investigate the biological functions and pathways of clusters A and D, with C2:CP KEGG gene sets from MSigDB as the reference gene sets. False discovery rate (FDR) < 0.05 was the screening threshold.

Weighted Gene Co-Expression Networks Analysis and Protein–Protein Interaction Networks Analysis

A weighted gene co-expression networks analysis (WGCNA) algorithm was used to mine the synergistically expressed gene modules. Immune-related genes from the ImmPort dataset (https://www.immport.org/) were classified into different modules which were significantly correlated with the four immune clusters by the R package "WGCNA." Samples were clustered by a hierarchical clustering algorithm implemented in the R function "hclust." The soft thresholding power $\beta = 3$ was selected by the R function "pickSoftThreshold" (scale free $R^2 = 0.85$). The expression matrix was converted into the adjacent matrix and then into the topological matrix for gene clustering. An average linkage hierarchical cluster approach was used to cluster genes into a dendrogram. The STRING database (Szklarczyk et al., 2011) was explored to construct the protein–protein interaction (PPI) network. In the PPI gene network of the target module, hub genes were the top ten genes ranked by the MCC algorithm of "cytoHubba" plugin in Cytoscape 3.8.0. In addition, the survival curves based on the best cut-off value of hub genes were drawn by the "survminer" package in R.

Statistical Analysis

Student's t-test was applied for normal distributions, and the Mann–Whitney U-test was performed for nonparametric distribution. Chi-square or Fisher's exact tests were used for categorical data. Kaplan–Meier curves and log-rank tests were used to evaluate the survival time of different immune clusters. The nonparametric Kruskal–Wallis (KW) test was used to analyze the difference in IC50 in different clusters. The Benjamini–Hochberg procedure was applied to control the

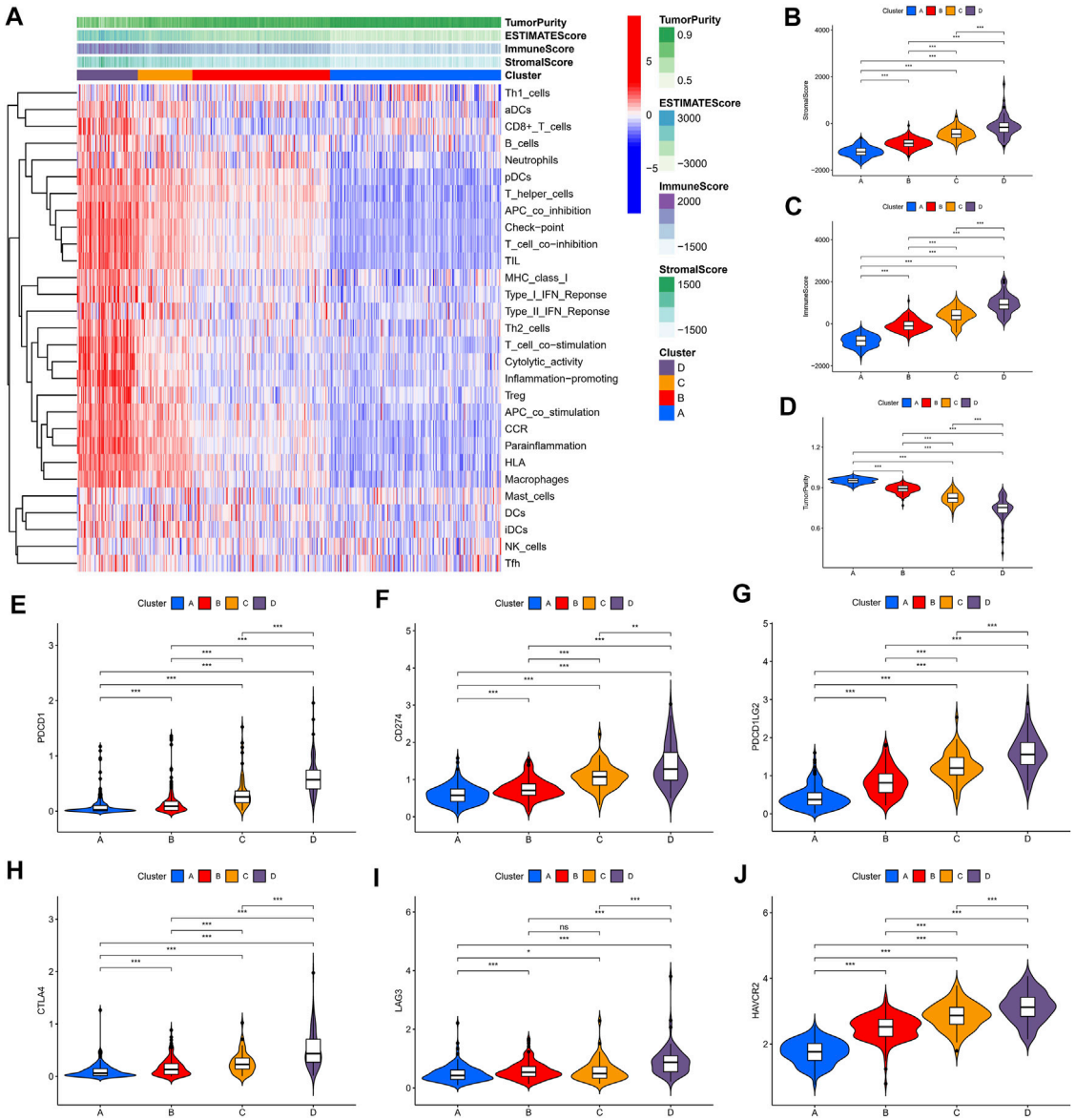


FIGURE 2 | Immune characteristics of the four clusters in the TCGA dataset. **(A)** Heat map of the four immune clusters based on ssGSEA scores. **(B)** Stromal scores, **(C)** immune scores, and **(D)** tumor purity of different clusters. **(E–J)** Violin plots for the immune checkpoint gene expressions of PDCD1, CD274, PDCD1LG2, CTLA-4, LAG3, and HAVCR2 in different clusters.

false discovery rate (FDR) for multiple testing. $p < 0.05$ was considered statistically significant (* represented $p < 0.05$, ** referred $p < 0.01$, and *** referred $p < 0.001$).

RESULTS

Identification of Four Immune-Related Clusters in Lower-Grade Gliomas

Unsupervised consensus clustering was applied to explore a novel immune classification of LGGs based on the ssGSEA scores of TCGA dataset. The optimal clusters number was found to be four with maximal consensus within clusters and minimal ambiguity among clusters (**Figures 1A–C**). PCA verified that the ssGSEA scores could completely be distinguished into four subtypes which were referred to as cluster A, cluster B, cluster C, and cluster D in TCGA dataset (**Figure 1D**). The clustering results were the same in CGGA-1 (**Supplementary Figure S1A**) and CGGA-2 datasets (**Supplementary Figure S1B**). The Sankey diagram revealed the immune function characteristics of the four clusters (**Figure 1E**). The majority of clusters A and B were related to the C5 function of "Immunologically Quiet." Cluster D was related to the C4 function of "Lymphocyte Depleted."

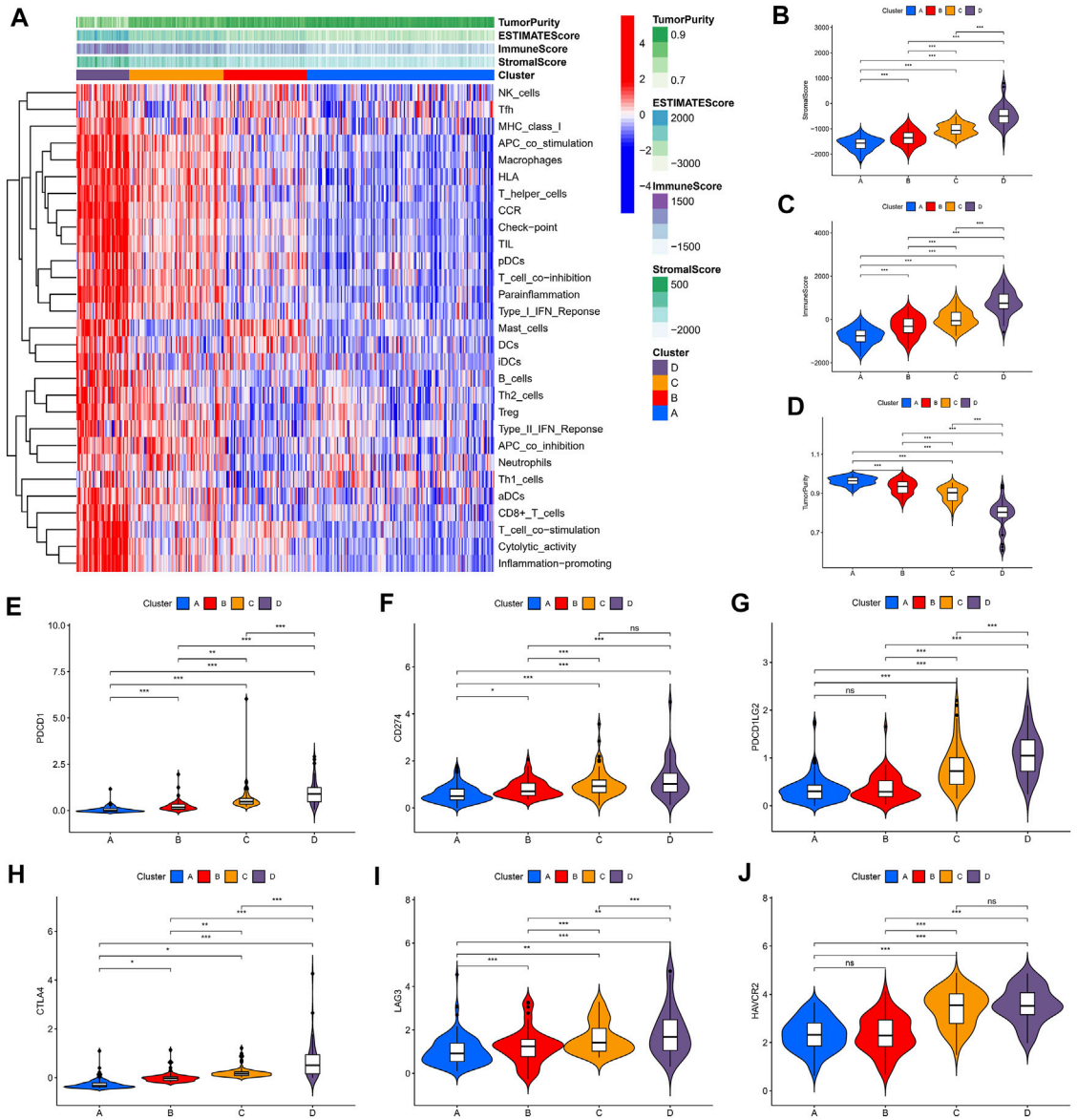


FIGURE 3 | Immune characteristics of the four clusters in the CGGA-1 dataset. **(A)** Heat map of the four immune clusters based on ssGSEA scores. **(B)** Stromal scores, **(C)** immune scores, and **(D)** tumor purity of different clusters. **(E–J)** Violin plots for the immune checkpoint gene expressions of PDCD1, CD274, PDCD1LG2, CTLA-4, LAG3, and HAVCR2 in different clusters.

Landscape of the Four Clusters in the Tumor Immune Environment in Lower-Grade Gliomas

We identified four immune clusters and their immune characteristics in TCGA, CGGA-1, and CGGA-2 datasets (**Figures 2A**, **3A**, **4A**) were shown in the heat maps. Cluster D showed the highest degree of immune infiltration and it was followed by clusters C, B, and A. Cluster D was considered the "immune-rich" phenotype with the highest enrichment scores and the least tumor purity while cluster A was the opposite which was regarded as an "immune cold" phenotype (**Figures 2A–D**, **Figures 3A–D**, **Figures 4A–D**). Apart from that, clusters A and B could be seen as a "low-immune infiltration" subgroup whereas clusters C and D were considered as a "high-immune infiltration" subgroup. The expression of the immune checkpoint genes PDCD1 (PD-1), CD274 (PD-L1), PDCD1LG2 (PD-L-2), CTLA-4, HAVCR2, and LAG3 which played a vital role in the oncogenesis and progression of LGG were expressed in the following order: D > C > B > A (**Figures 2E–J**, **Figures 3E–J**, **Figures 4E–J**).

Clinical Features and Gene Mutation Characteristics of the Four Immune Clusters

To evaluate the clinical features and gene mutation characteristics among the four immune clusters, age, gender,

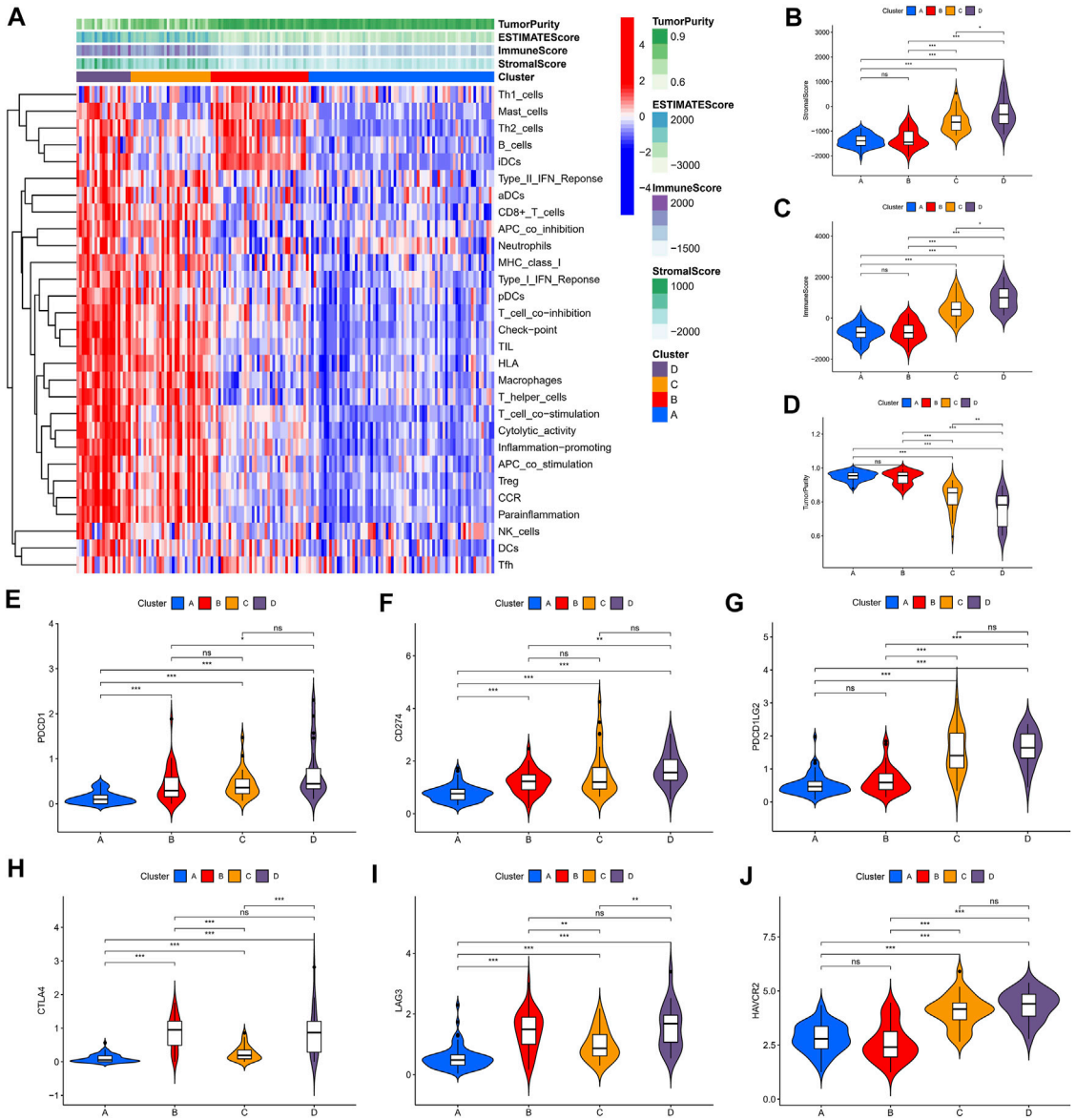


FIGURE 4 | Immune characteristics of the four clusters in the CGGA-2 dataset. **(A)** Heat map of the four immune clusters based on ssGSEA scores. **(B)** Stromal scores, **(C)** immune scores, and **(D)** tumor purity of different clusters. **(E–J)** Violin plots for the immune checkpoint gene expressions of PDCD1, CD274, PDCD1LG2, CTLA-4, LAG3, and HAVCR2 in different clusters.

tumor grade, IDH1 (R132) status, IDH2 R172 status, PTEN status, EGFR status, ATRX status ,and TP53 status in TCGA dataset were counted (**Table 1**; **Figure 5**). In clusters A, C, and D, people aged more than 40 years accounted for more proportion, whereas those younger than 40 years were more common in cluster B. WHO grade II glioma tended to be common in the "low-immune infiltration" subgroup (clusters A and B), whereas the "high-immune infiltration" subgroup (clusters C and D) counted more in WHO grade III glioma. The frequency of IDH1 (R132) mutation was much higher in the "low-immune infiltration" subgroup than the "high-immune infiltration" subgroup. The frequency of PTEN and EGFR mutations was significantly higher in cluster D, which had the highest immune infiltration. In clusters B and C, which had mild immune infiltration changes, ATRX mutation frequencies were higher than those in clusters with extreme immune infiltration changes (clusters A and D). TP53 mutation was common in cluster B. Gender and IDH2 (R172) mutation status were not significant covariates in the immune classification. In addition, the "low-immune infiltration" subgroup (clusters A and B) showed longer overall survival than the "high-immune infiltration" subgroup (clusters C and D) (**Figure 6**), indicating that immune infiltration of LGG played a negative role in the prognosis.

TABLE 1 | Correlations among the four immune clusters and clinical characteristics in the TCGA-LGG dataset.

Covariates	Cluster	Total	A	B	C	D	*p* value
Age	<40	221 (45.95%)	84 (43.3%)	85 (54.49%)	28 (45.16%)	24 (34.78%)	0.035
	≥40	260 (54.05%)	110 (56.7%)	71 (45.51%)	34 (54.84%)	45 (65.22%)	
Gender	FEMALE	214 (44.49%)	95 (48.97%)	60 (38.46%)	28 (45.16%)	31 (44.93%)	0.2738
	MALE	267 (55.51%)	99 (51.03%)	96 (61.54%)	34 (54.84%)	38 (55.07%)	
Grade	G2	230 (47.82%)	107 (55.15%)	83 (53.21%)	28 (45.16%)	12 (17.39%)	0
	G3	251 (52.18%)	87 (44.85%)	73 (46.79%)	34 (54.84%)	57 (82.61%)	
IDH1 R132 status	Mutation	368 (76.51%)	161 (82.99%)	135 (86.54%)	37 (59.68%)	35 (50.72%)	0
	Wild	113 (23.49%)	33 (17.01%)	21 (13.46%)	25 (40.32%)	34 (49.28%)	
IDH2 R172 status	Mutation	20 (4.16%)	12 (6.19%)	6 (3.85%)	1 (1.61%)	1 (1.45%)	0.2292
	Wild	461 (95.84%)	182 (93.81%)	150 (96.15%)	61 (98.39%)	68 (98.55%)	
PTEN status	Mutation	29 (6.03%)	5 (2.58%)	8 (5.13%)	4 (6.45%)	12 (17.39%)	2.00E-04
	Wild	452 (93.97%)	189 (97.42%)	148 (94.87%)	58 (93.55%)	57 (82.61%)	
EGFR status	Mutation	30 (6.24%)	7 (3.61%)	6 (3.85%)	6 (9.68%)	11 (15.94%)	0.001
	Wild	451 (93.76%)	187 (96.39%)	150 (96.15%)	56 (90.32%)	58 (84.06%)	
ATRX status	Mutation	174 (36.17%)	45 (23.2%)	83 (53.21%)	26 (41.94%)	20 (28.99%)	0
	Wild	307 (63.83%)	149 (76.8%)	73 (46.79%)	36 (58.06%)	49 (71.01%)	
TP53 status	Mutation	216 (44.91%)	66 (34.02%)	92 (58.97%)	29 (46.77%)	29 (42.03%)	1.00E-04
	Wild	265 (55.09%)	128 (65.98%)	64 (41.03%)	33 (53.23%)	40 (57.97%)	

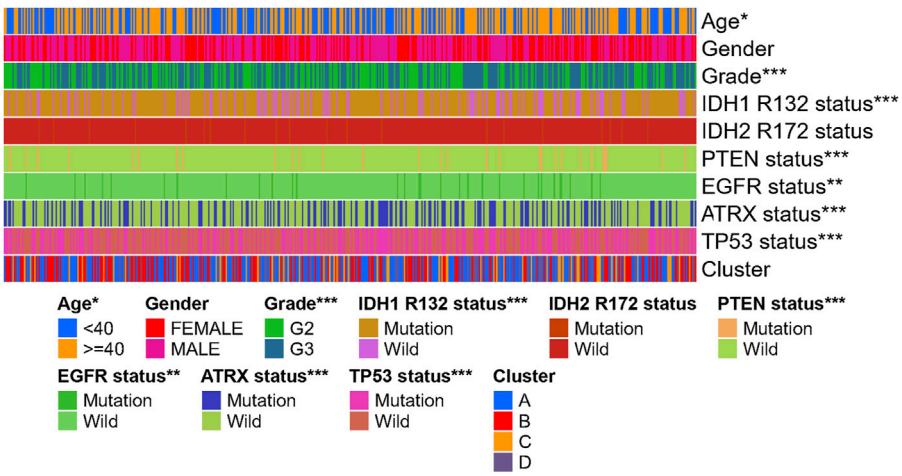


FIGURE 5 | Heat map of clinical features and gene mutation. Characteristics of the four immune subtypes in the TCGA-LGG dataset.

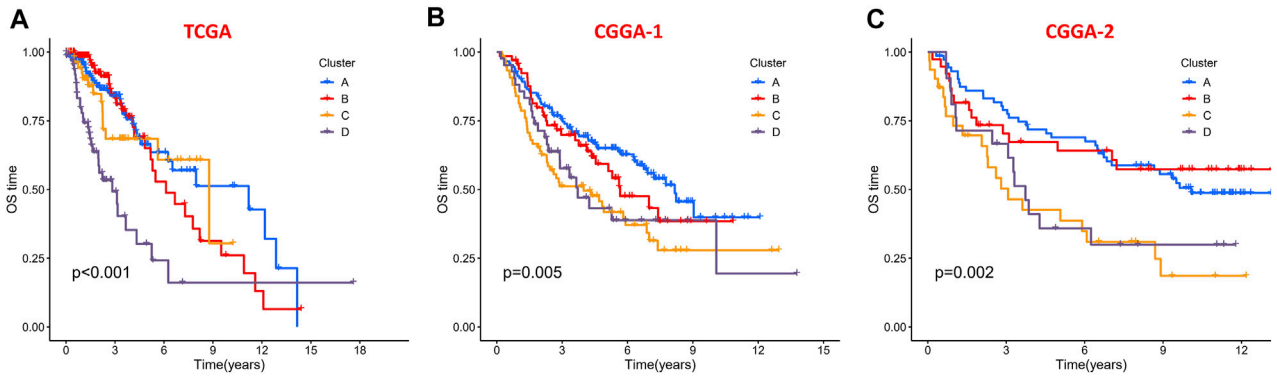


FIGURE 6 | Kaplan–Meier survival curves of the four clusters in the **(A)** TCGA dataset, **(B)** CGGA-1 datasets, and **(C)** CGGA-2 datasets.

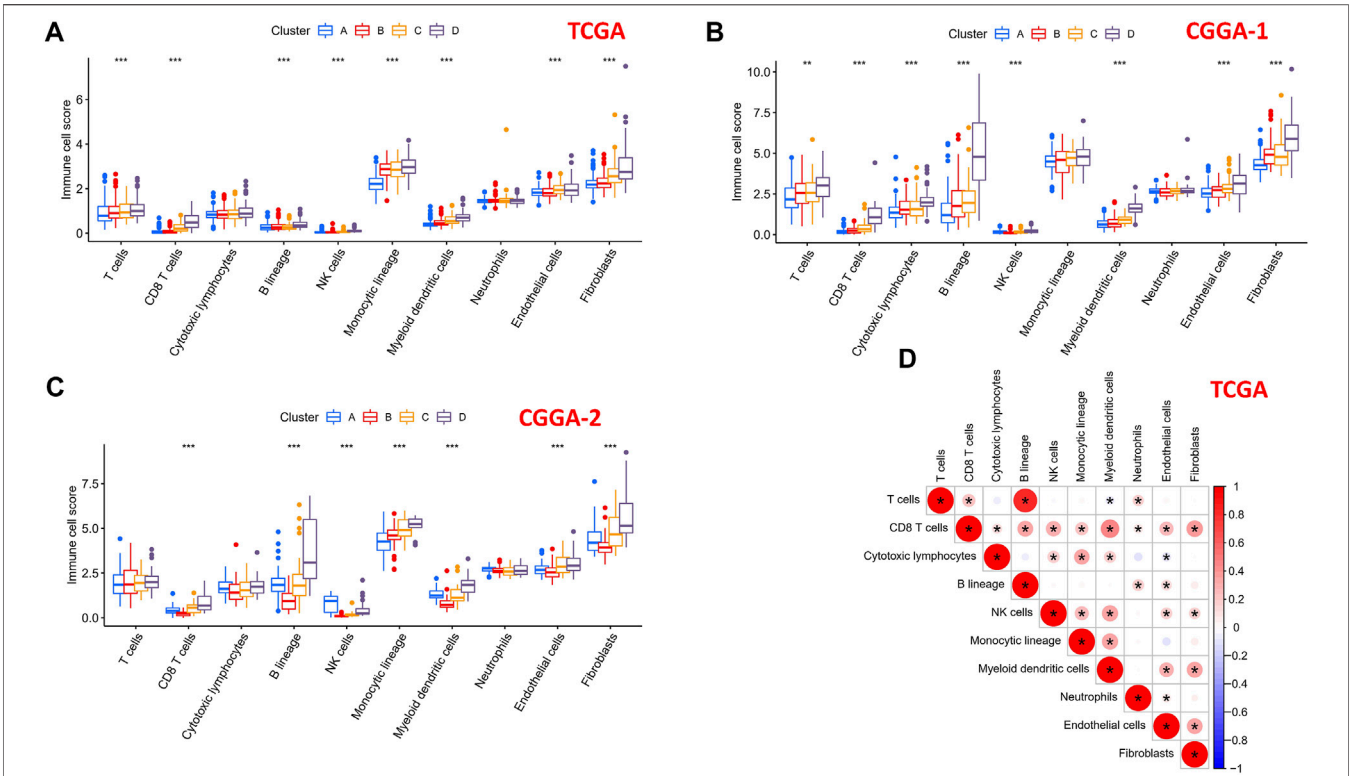


FIGURE 7 | Immune cells scores of the four subtypes in 10 human immunocytes using the MCPcounter algorithm. **(A)** TCGA, **(B)** CGGA-1, and **(C)** CGGA-2 cohorts showed significantly different immune cell populations among the four subtypes. **(D)** Intrinsic correlation between infiltrating immunocytes and immune scores.

Comparison of Immune and Stromal Cells Among the Four Clusters

To explore the differential distribution of immunocytes and stromal cells in tumor immunity clusters, the MCPcounter algorithm was used to calculate the contents of two stromal cells and eight immune cells in the four clusters in TCGA, CGGA-1, and CGGA-2 datasets (**Figures 7A–C**). Immune cell scores of CD8 T cells, B lineage, NK cells, myeloid dendritic cells, endothelial cells, and fibroblasts in the "high-immune infiltration" subgroup (clusters C and D) were significantly higher than those in the "low-immune infiltration" subgroup (clusters A and B). Then, the correlation landscape of immunocytes was characterized to compare the relative subpopulations of infiltrating immune cells and immune scores among the four cluster patterns (**Figure 7D**). Cox regression analysis of the 10 immune cells in TCGA, CGGA-1, and CGGA-2 datasets are shown in **Supplementary Figure S2**, revealing the prognostic risk factors of infiltrating immunocytes.

Therapeutic Potential of Different Immune Clusters

We compared the expression profiles of the four immune clusters in TCGA datasets by the Subclass Mapping algorithm which assessed the response to anti-PD-1 and anti-CTLA-4 therapies. A significant correlation was observed when comparing cluster D with the PD-1 response group (Bonferroni-corrected p = 0.001, **Supplementary Figure S3A**). It revealed that cluster D might have a better response to anti-PD-1 therapy while no significant correlation of anti-CTLA-4 therapy was observed in all the clusters. The "pRRophetic" algorithm was applied to evaluate the sensitivity of four common chemical drugs: cisplatin, bleomycin, doxorubicin, and gemcitabine for the four immune clusters. A lower IC50 value would indicate a better sensitivity in the prediction models. For bleomycin and doxorubicin, the "low-immune infiltration" subgroup (clusters A and B) was more sensitive than the "high-immune infiltration" subgroup (clusters C and D). For cisplatin and gemcitabine, cluster D was the most sensitive and cluster A was the least sensitive compared with the other clusters (**Figures 8A–D**). Moreover, to compare the accuracy of the four immune clusters, prognosis signatures in other references were used to compare the C-index. The results were also exciting: in the C-index for predicting the LGG survival possibility, our immune clusters showed better predictive value than other signatures (Maimaiti, Aierpati et al., 2022; Maimaiti, Aierpati et al., 2021) (0.813 > 0.774 > 0.712 > 0.662, **Supplementary Figure S3B**).

Gene Set Enrichment Analysis

A GSEA analysis was performed to screen the correlated biological pathways in immune clusters A and D. Cluster D

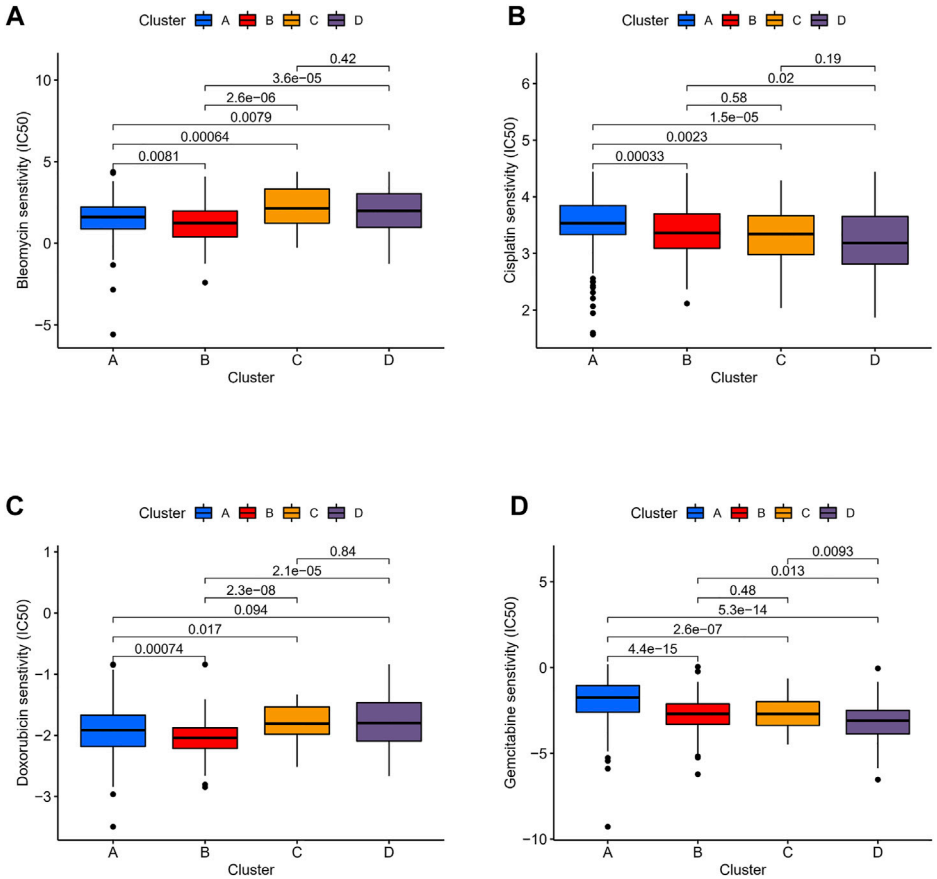


FIGURE 8 | Sensitivity of chemotherapeutics in different immune clusters. Differences in IC50 of **(A)** bleomycin, **(B)** cisplatin, **(C)** doxorubicin, and **(D)** gemcitabine were estimated among the four immune clusters.

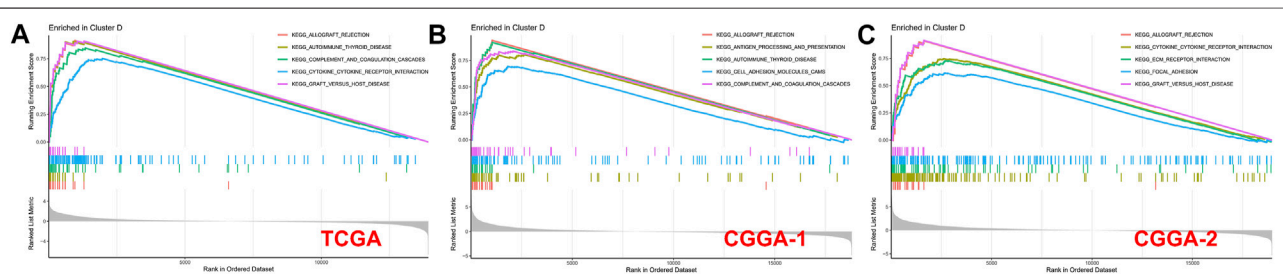


FIGURE 9 | GSEA enrichment analysis for clusters A and D in **(A)** TCGA, **(B)** CGGA-1, and **(C)** CGGA-2 datasets. No significant enrichment was found in cluster A with a threshold of FDR <0.05.

was enriched in Allograft rejection, complement and coagulation cascades, cytokine–cytokine receptor interaction, graft versus host disease, antigen processing and presentation, cell adhesion molecules (CAMs), ECM-receptor interaction, and focal adhesion in TCGA, CGGA-1, and CGGA-2 datasets. Enrichment results of cluster A were not significant under the strict FDR <0.05 threshold in all the three datasets (**Figure 9**).

Mutation Landscapes in Lower-Grade Gliomas

Tumor mutational burden of coding errors is reported to have a certain correlation with the tumor immune microenvironment. We explored this correlation of different immune clusters in TCGA-LGG datasets (**Figures 10A–D**). The frequency of IDH1 missense mutations in the "low-immune infiltration" subgroup was higher than that in

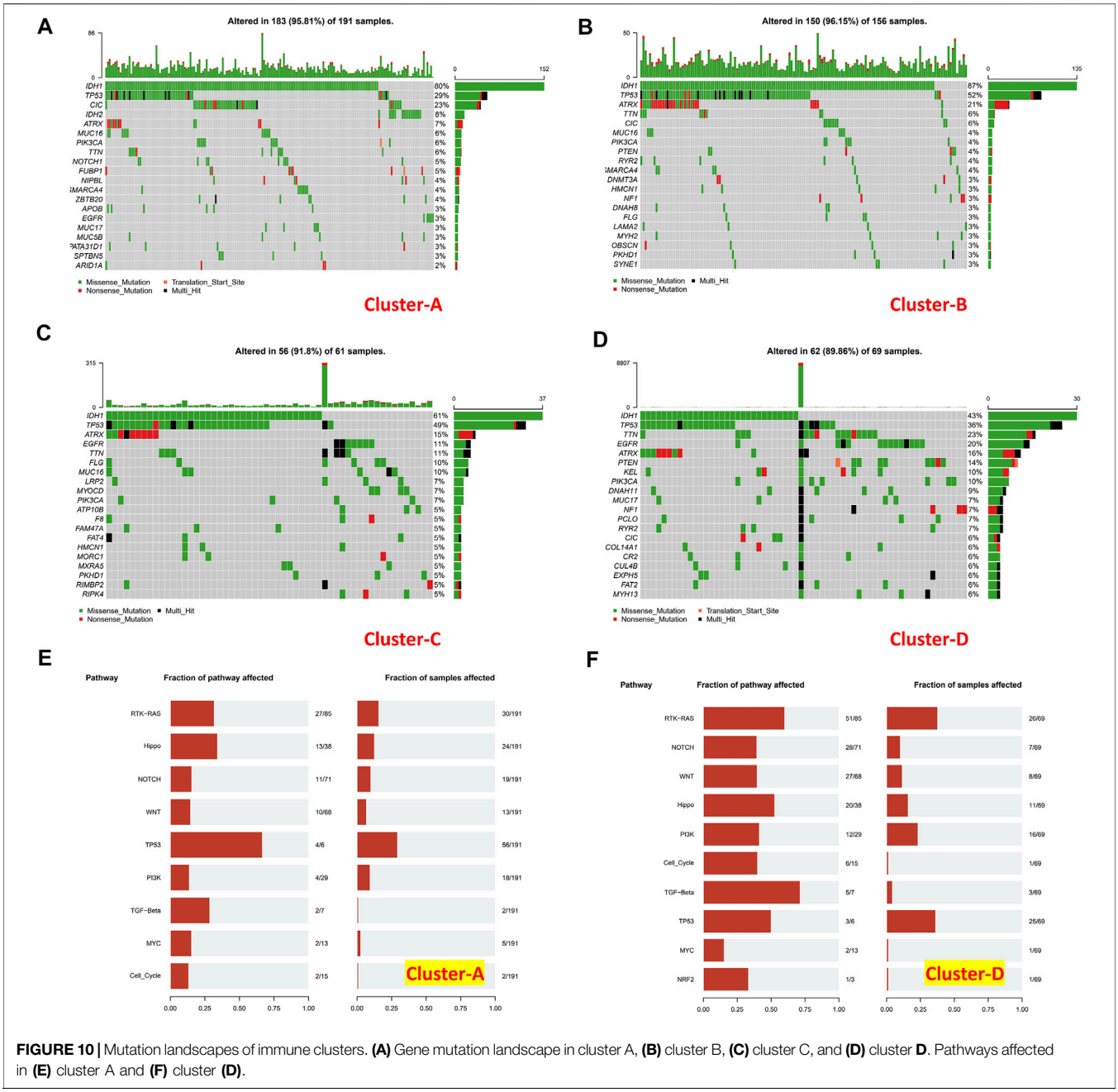


FIGURE 10 | Mutation landscapes of immune clusters. **(A)** Gene mutation landscape in cluster A, **(B)** cluster B, **(C)** cluster C, and **(D)** cluster **D**. Pathways affected in **(E)** cluster A and **(F)** cluster **(D)**.

the "high-immune infiltration" subgroup (80 and 87% in clusters A and B, 61 and 43% in clusters C and D). TP53 mutations were higher in clusters B (52%) and C (49%) than those in clusters A (29%) and D (36%). Meanwhile, most of them were missense mutations. The CIC missense mutation was high in cluster A. ATRX mutations including missense mutations, nonsense mutations, and multi-hit were at a high frequency in clusters B and C. TTN, EGFR, and ATRX mutations were common in the "high-immune infiltration" subgroup. PTEN, KEL, and PLK3CA mutations were higher in cluster D. the TP53 pathway was highly affected in cluster A and RTK-RAS and TP53 pathways were affected in cluster D (**Figures 10E,F**).

Weighted Gene Co-Expression Network Analysis and the Immune Characteristics of Hub Genes

The WGCNA networks of immuno-related genes with immune infiltrating clusters were constructed. The optimal soft thresholding power β was selected (**Figures 11A**, **Figure 9**) and three modules were obtained (**Figures 11B,C**). Red, blue,

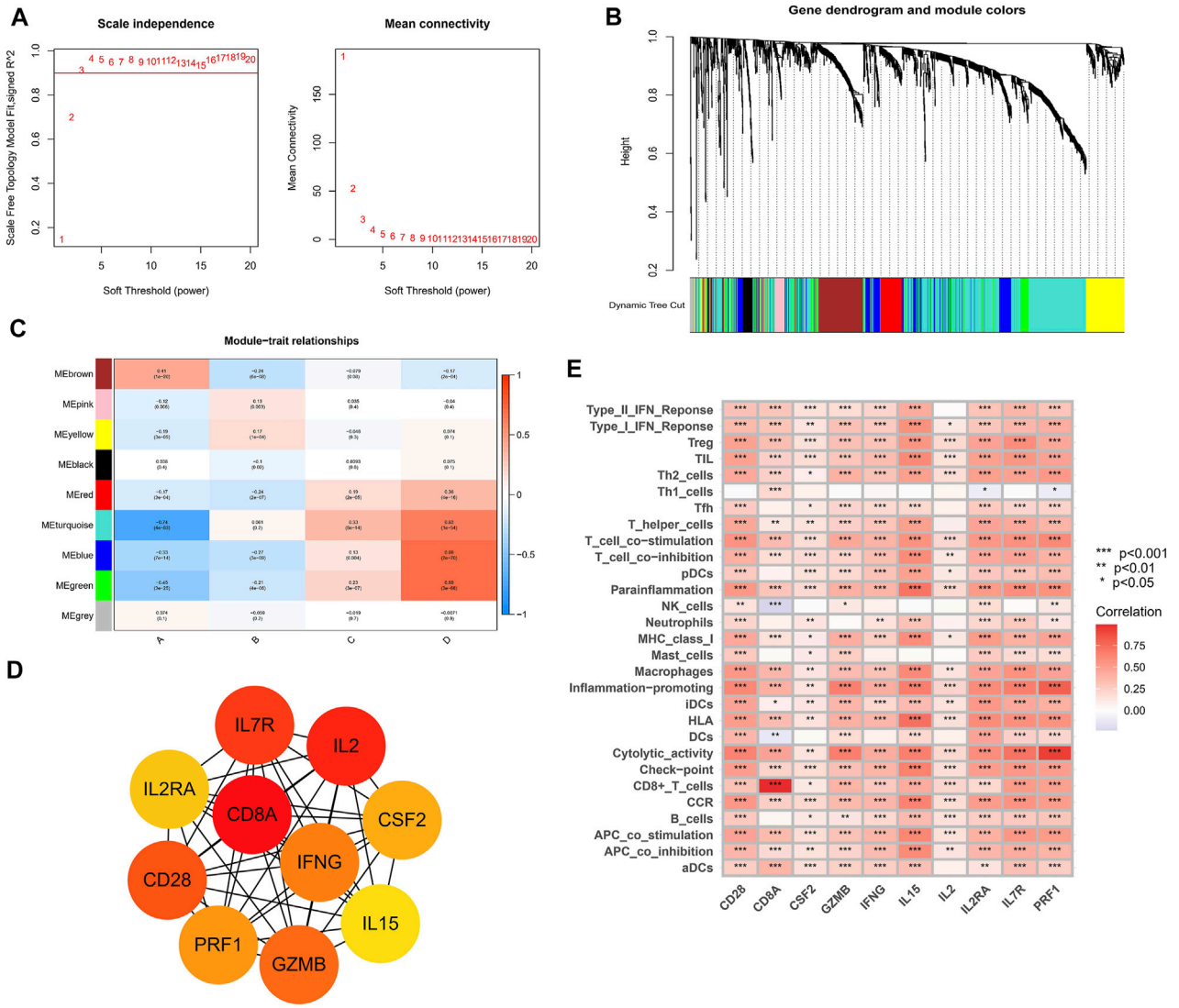


FIGURE 11 | Weighed co-expression analysis of immuno-related genes in the TCGA cohort. **(A)** Selection of soft thresholding power **(B)** Gene dendrogram and the correlation modules. **(C)** Heat map of module-trait relationships. **(D)** Hub genes and their internal correlation network. **(E)** Correlation heat map for hub genes and immune gene sets.

and green modules were positively correlated with the "high-immune infiltration" subgroup (clusters C and D) and negatively correlated with the "low-immune infiltration" subgroup (clusters A and B), $p < 0.05$ was used as the threshold. The turquoise module was positively correlated with clusters C and D and negatively correlated with cluster A. The brown module was positively correlated with cluster A and negatively correlated with clusters B and D. Pink and yellow modules were positively correlated with cluster B and negatively correlated with cluster A. In the blue module, which was most correlated with cluster D, the top ten hub genes selected by the MCC algorithm in the PPI network were CD28, CD8A, CSF2, GZMB, IFNG, IL15, IL2, IL2RA, IL7R, and PRF1(**Figure 11D**). Hub genes were positively correlated with most of the immune cells and immune functions, such as HLA, CCR, etc., (**Figure 11E**). A survival analysis showed that high expression of CD28, CD8A, IFNG, IL2RA, IL7R, IL15, and PRF1 had a poor prognosis whereas a better prognosis was found in IL2 and GZMB (**Figure 12**).

DISCUSSION AND CONCLUSION

The prognosis of LGG patients had few significant improvements in the past 30 years. Individualized therapeutic schedules were because of the natural intrinsic heterogeneity of LGG (Deng et al., 2020). Li et al. (2021) provided a metabolic signature-based subgrouping method for LGG and Zhou et al. (2021) divided LGG into three clusters based on a tertiary lymphoid structure to provide potential treating options. Since the existence of an afferent system between the brain and peripheral immune

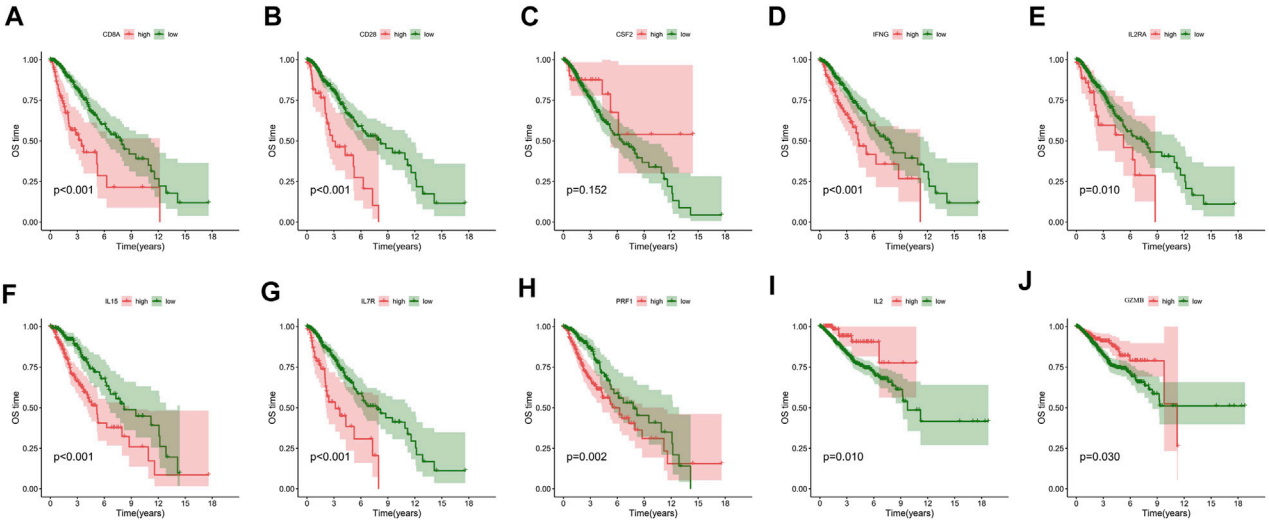


FIGURE 12 | Overall survival analysis of **(A)** CD8A, **(B)** CD28, **(C)** CSF2, **(D)** IFNG, **(E)** IL2RA, **(F)** IL15, **(G)** IL7R, **(H)** PRF1, **(I)** IL2, and **(J)** GZMB.

system had been demonstrated (Qi et al., 2020), immunotherapy would be a promising strategy for its ability to penetrate the blood–brain barrier (Xu et al., 2020). Wu et al. (2020) classified diffuse LGGs into three immunological subtypes and found that the high lymphocytic and macrophage M2 infiltrate subtype indicated a worse prognosis. The immune microenvironment of LGG remained complicated. We studied the correlations between immunogenomic changes and immunity infiltration features in LGG to identify proper immune clusters and hub genes for precision treatment.

Our study identified four immune clusters in TCGA dataset and they were verified in two CGGA datasets. Cluster A, which had the lowest immune infiltration, was regarded as an "immune cold" phenotype. Clusters A and B could be considered a "low-immune infiltration" subgroup and both of them were closely correlated with "Immunologically Quiet" (C5) functional subtypes. Clusters C and D were considered the "high-immune infiltration" subgroup when cluster D was inferred as an "immune-rich" phenotype with the highest immune infiltration degree. In general, the expression of immune checkpoints such as PDCD1, CD274, PDCD1LG2, CTLA-4, LAG3, and HAVCR2 increased along with the order (cluster D > C > B > A), indicating higher hazards of immune escape in high-immune infiltration clusters. LGGs in clusters A and B tended to have a lower WHO grade, higher IDH1 mutation, and better overall survival than those in the "high-immune infiltration" subgroup. The LGGs in "immune-rich" cluster D showed significant PTEN and EGFR mutation frequencies and notable sensitivity to anti-immune checkpoint PD-1 therapy and the chemotherapy of cisplatin and gemcitabine. On the contrary, LGGs in the "low-immune infiltration" subgroup (clusters A and B) were more sensitive to bleomycin and doxorubicin. The results would provide potential individualized treatment recommendations for LGGs. Cluster D was enriched in KEGG pathways such as cytokine–cytokine receptor interaction, antigen processing and presentation, cell adhesion molecules (CAMs) and ECM-receptor interaction, reminding us that different immunophenotypes may be caused by changes in the aforementioned pathways. Meanwhile, RTK-RAS and TP53 pathways were affected in cluster D. In the blue module of WGCNA networks, CD28, CD8A, CSF2, GZMB, IFNG, IL15, IL2, IL2RA, IL7R, and PRF1 were selected as the hub genes which were closely correlated with most of the immune cells. Seven of them were correlated with a poor prognosis, two of them were protective prognostic factors and one of them had no significant association with prognosis.

An immune clustering analysis of our study indicated that high-immune infiltration would lead to a worse prognosis with immune checkpoint activation. This distinct feature might result from the immunocyte recoding by cytokines and chemokines in the LGG microenvironment (Hinshaw and Shevde, 2019). Immunocytes were turned into tumor-promoting phenotypes and conversely promoted tumor growth and immune evasion. In addition, the relatively lower IDH1 mutation and higher PTEN and EGFR mutation frequency in high immune infiltration clusters also supported the aforementioned inference. Although LGG patients in cluster D would suffer poor prognosis expectations, they might benefit from immune checkpoint PD-1 inhibitors and chemotherapeutic drugs of cisplatin and gemcitabine. Cisplatin and gemcitabine had shown encouraging tolerance and efficacy in clinical trials (Gertler et al., 2000; Massimino et al., 2002; Massimino et al., 2005; Hall et al., 2019). The TME in LGG appears to be different from other solid tumors because of the presence of the blood–brain barrier or properties of macrophages. In the present research, the M2-type macrophage was significantly enriched in primary LGG, and the proportions of macrophages can still constitute up to 50% in the TME of LGG. Some researchers demonstrated that high levels of M2-type macrophages were defined as the adverse prognostic factors

in LGG. Conversely, high levels of M1-type macrophages and CD8+T cells were identified as protective factors. Apart from that, the revolution of drug delivery methods in nanoplatforms and liposomes had shown a promising future to precisely deliver the individualized chemotherapeutic drugs for LGGs (Shein et al., 2016; Renault-Mahieux et al., 2021; Wang et al., 2021). Most of the hub genes correlated with cluster D had the function of immunocyte activation. CD28 is involved in T-cell activation, cell proliferation induction, and T-cell survival. CD8 mediates efficient cell–cell interactions within the immune system. IFNG can activate effector immune cells and enhance antigen presentation. IL2RA is involved in the regulation of immune tolerance by controlling regulatory T cells. IL7R mediates the proliferation of lymphoid progenitors. IL15 stimulates the proliferation of T-lymphocytes. PRF1 plays a key role in defence against neoplastic cells. IL2 can stimulate B-cells, monocytes, lymphokine-activated killer cells, natural killer cells, and glioma cells. GZMB mediates target cell death and CSF2 promotes the production, differentiation, and function of granulocytes and macrophages. Although the majority of hub genes play a role in tumor promotion in the microenvironment of LGG, which was consistent with the poor prognosis expectation in cluster D, two hub genes termed IL2 and GZMB exerted a protective role in prognosis. It revealed the complex inherent interconnections of immunogenomic changes.

There are still some limitations in our study. First, we were unable to conduct an external validation in native cohorts. Second, we only used the ssGSEA and MCPcounter algorithms to corroborate our findings, and we will need to conduct assays to confirm our conclusion in the future. In conclusion, immunotyping of LGGs revealed the heterogeneity of the immune microenvironment and genomics changes. Our classifications would be beneficial for individualized prognostic prediction and anti-tumor therapy.

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.

AUTHOR CONTRIBUTIONS

YZ, SF, and GC conceived the study. YZ and ZS conducted the data analysis. SF and YZ analyzed and visualized the results. YZ wrote the manuscript. ZW supervised the study. GC reviewed the manuscript. All authors approved the final version of the manuscript.

ACKNOWLEDGMENTS

The authors acknowledge TCGA, CGGA, Gene Pattern, and GDSC databases for their public platforms and contributors for uploading their valuable datasets.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.894865/full#supplementary-material

REFERENCES

Appolloni, I., Alessandrini, F., Ceresa, D., Marubbi, D., Gambini, E., Reverberi, D., et al. (2019). Progression from Low- to High-Grade in a Glioblastoma Model Reveals the Pivotal Role of Immunoediting. *Cancer Lett.* 442, 213–221. doi:10.1016/j.canlet.2018.10.006

Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the Population Abundance of Tissue-Infiltrating Immune and Stromal Cell Populations Using Gene Expression. *Genome Biol.* 17 (1), 218. doi:10.1186/s13059-016-1070-5

Deng, X., Lin, D., Zhang, X., Shen, X., Yang, Z., Yang, L., et al. (2020). Profiles of Immune-related Genes and Immune Cell Infiltration in the Tumor Microenvironment of Diffuse Lower-grade Gliomas. *J. Cell Physiol.* 235 (10), 7321–7331. doi:10.1002/jcp.29633

Eckel-Passow, J. E., Lachance, D. H., Molinaro, A. M., Walsh, K. M., Decker, P. A., Sicotte, H., et al. (2015). Glioma Groups Based on 1p/19q,IDH, andTERTPromoter Mutations in Tumors. *N. Engl. J. Med.* 372 (26), 2499–2508. doi:10.1056/NEJMoa1407279

Emens, L. A., Ascierto, P. A., Darcy, P. K., Demaria, S., Eggermont, A. M. M., Redmond, W. L., et al. (2017). Cancer Immunotherapy: Opportunities and Challenges in the Rapidly Evolving Clinical Landscape. *Eur. J. Cancer* 81, 116–129. doi:10.1016/j.ejca.2017.01.035

Gertler, S. Z., MacDonald, D., Goodyear, M., Forsyth, P., Stewart, D. J., Belanger, K., et al. (2000). NCIC-CTG Phase II Study of Gemcitabine in Patients with Malignant Glioma (IND.94). *Ann. Oncol.* 11 (3), 315–318. doi:10.1023/a:1008336607135

Hall, P. E., Lewis, R., Syed, N., Shaffer, R., Evanson, J., Ellis, S., et al. (2019). A Phase I Study of Pegylated Arginine Deiminase (Pegargiminase), Cisplatin, and Pemetrexed in Argininosuccinate Synthetase 1-Deficient Recurrent High-Grade Glioma. *Clin. Cancer Res.* 25 (9), 2708–2716. doi:10.1158/1078-0432.ccr-18-3729

Han, S., Liu, Y., Cai, S. J., Qian, M., Ding, J., Larion, M., et al. (2020). IDH Mutation in Glioma: Molecular Mechanisms and Potential Therapeutic Targets. *Br. J. Cancer* 122 (11), 1580–1589. doi:10.1038/s41416-020-0814-x

Hinshaw, D. C., and Shevde, L. A. (2019). The Tumor Microenvironment Innately Modulates Cancer Progression. *Cancer Res.* 79 (18), 4557–4566. doi:10.1158/0008-5472.can-18-3962

Hoshida, Y., Brunet, J.-P., Tamayo, P., Golub, T. R., and Mesirov, J. P. (2007). Subclass Mapping: Identifying Common Subtypes in Independent Disease Data Sets. *PLoS One* 2 (11), e1195. doi:10.1371/journal.pone.0001195

Li, G., Wu, Z., Gu, J., Zhu, Y., Zhang, T., Wang, F., et al. (2021). Metabolic Signature-Based Subtypes May Pave Novel Ways for Low-Grade Glioma Prognosis and Therapy. *Front. Cel Dev. Biol.* 9, 755776. doi:10.3389/fcell.2021.755776

Lipson, E. J., Forde, P. M., Hammers, H.-J., Emens, L. A., Taube, J. M., and Topalian, S. L. (2015). Antagonists of PD-1 and PD-L1 in Cancer Treatment. *Semin. Oncol.* 42 (4), 587–600. doi:10.1053/j.seminoncol.2015.05.013

Louis, D. N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W. K., et al. (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a Summary. *Acta Neuropathol.* 131 (6), 803–820. doi:10.1007/s00401-016-1545-1

Louis, D. N., Perry, A., Wesseling, P., Brat, D. J., Cree, I. A., Figarella-Branger, D., et al. (2021). The 2021 WHO Classification of Tumors of the Central Nervous

System: a Summary. *Neuro. Oncol.* 23 (8), 1231–1251. doi:10.1093/neuonc/noab106

Massimino, M., Spreafico, F., Cefalo, G., Riccardi, R., Tesoro-Tess, J. D., Gandola, L., et al. (2002). High Response Rate to Cisplatin/etoposide Regimen in Childhood Low-Grade Glioma. *Jco* 20 (20), 4209–4216. doi:10.1200/jco.2002.08.087

Massimino, M., Gandola, L., Luksch, R., Spreafico, F., Riva, D., Solero, C., et al. (2005). Sequential Chemotherapy, High-Dose Thiotepa, Circulating Progenitor Cell rescue, and Radiotherapy for Childhood High-Grade Glioma. *Neuro Oncol.* 7 (1), 41–48. doi:10.1215/s1152851704000304

Nejo, T., Matsushita, H., Karasaki, T., Nomura, M., Saito, K., Tanaka, S., et al. (2019). Reduced Neoantigen Expression Revealed by Longitudinal Multiomics as a Possible Immune Evasion Mechanism in Glioma. *Cancer Immunol. Res.* 7 (7), 1148–1161. doi:10.1158/2326-6066.cir-18-0599

Ohba, S., Kuwahara, K., Yamada, S., Abe, M., and Hirose, Y. (2020). Correlation between IDH, ATRX, and TERT Promoter Mutations in Glioma. *Brain Tumor Pathol.* 37 (2), 33–40. doi:10.1007/s10014-020-00360-4

Qi, Y., Liu, B., Sun, Q., Xiong, X., and Chen, Q. (2020). Immune Checkpoint Targeted Therapy in Glioma: Status and Hopes. *Front. Immunol.* 11, 578877. doi:10.3389/fimmu.2020.578877

Renault-Mahieux, M., Vieillard, V., Seguin, J., Espeau, P., Le, D. T., Lai-Kuen, R., et al. (2021). Co-Encapsulation of Fisetin and Cisplatin into Liposomes for Glioma Therapy: From Formulation to Cell Evaluation. *Pharmaceutics* 13 (7), 970. doi:10.3390/pharmaceutics13070970

Roh, W., Chen, P.-L., Reuben, A., Spencer, C. N., Prieto, P. A., Miller, J. P., et al. (2017). Integrated Molecular Analysis of Tumor Biopsies on Sequential CTLA-4 and PD-1 Blockade Reveals Markers of Response and Resistance. *Sci. Transl. Med.* 9 (379), eaah3560. doi:10.1126/scitranslmed.aah3560

Shein, S. A., KuznetsovII, Abakumova, T. O., Chelushkin, P. S., Melnikov, P. A., Korchagina, A. A., et al. (2016). VEGF- and VEGFR2-Targeted Liposomes for Cisplatin Delivery to Glioma Cells. *Mol. Pharm.* 13 (11), 3712–3723. doi:10.1021/acs.molpharmaceut.6b00519

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., et al. (2011). The STRING Database in 2011: Functional Interaction Networks of Proteins, Globally Integrated and Scored. *Nucleic Acids Res.* 39 (Database issue), D561–D568. doi:10.1093/nar/gkq973

Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Ou Yang, T. H., et al. (2018). The Immune Landscape of Cancer. *Immunity* 48 (4), 812–830.e14. doi:10.1016/j.immuni.2018.03.023

Wainwright, D. A., Chang, A. L., Dey, M., Balyasnikova, I. V., Kim, C. K., Tobias, A., et al. (2014). Durable Therapeutic Efficacy Utilizing Combinatorial Blockade against IDO, CTLA-4, and PD-L1 in Mice with Brain Tumors. *Clin. Cancer Res.* 20 (20), 5290–5301. doi:10.1158/1078-0432.ccr-14-0514

Wang, J., Tang, W., Yang, M., Yin, Y., Li, H., Hu, F., et al. (2021). Inflammatory Tumor Microenvironment Responsive Neutrophil Exosomes-Based Drug Delivery System for Targeted Glioma Therapy. *Biomaterials* 273, 120784. doi:10.1016/j.biomaterials.2021.120784

Wu, F., Wang, Z. L., Wang, K. Y., Li, G. Z., Chai, R. C., Liu, Y. Q., et al. (2020). Classification of Diffuse Lower-grade Glioma Based on Immunological Profiling. *Mol. Oncol.* 14 (9), 2081–2095. doi:10.1002/1878-0261.12707

Xu, S., Tang, L., Li, X., Fan, F., and Liu, Z. (2020). Immunotherapy for Glioma: Current Management and Future Application. *Cancer Lett.* 476, 1–12. doi:10.1016/j.canlet.2020.02.002

Yin, W., Jiang, X., Tan, J., Xin, Z., Zhou, Q., Zhan, C., et al. (2020). Development and Validation of a Tumor Mutation Burden-Related Immune Prognostic Model for Lower-Grade Glioma. *Front. Oncol.* 10, 1409. doi:10.3389/fonc.2020.01409

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

Zeng, W.-J., Yang, Y.-L., Liu, Z.-Z., Wen, Z.-P., Chen, Y.-H., Hu, X.-L., et al. (2018). Integrative Analysis of DNA Methylation and Gene Expression Identify a Three-Gene Signature for Predicting Prognosis in Lower-Grade Gliomas. *Cell Physiol. Biochem.* 47 (1), 428–439. doi:10.1159/000489954

Zhou, X., Li, W., Yang, J., Qi, X., Chen, Y., Yang, H., et al. (2021). Tertiary Lymphoid Structure Stratifies Glioma into Three Distinct Tumor Subtypes. *Aging* 13 (24), 26063–26094. doi:10.18632/aging.203798

Zindl, C. L., and Chaplin, D. D. (2010). Immunology. Tumor Immune Evasion. *Science* 328 (5979), 697–698. doi:10.1126/science.1190310

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 Genetics

ORIGINAL RESEARCH
published: 16 May 2022
doi: 10.3389/fgene.2022.905027

Evaluation of Biomarkers and Immune Microenvironment of Osteoarthritis: Evidence From Omics Data and Machine Learning

Zhixin Liu[1], Heng Liu[2,3], Deqiang Li[1], Liang Ma[1], Tongxin Lu[1], Hao Sun[1], Yuankai Zhang[1] and Hui Yang[4]**

[1]Department of Orthopedics, Qilu Hospital of Shandong University, Jinan, China, [2]NHC Key Laboratory of Otorhinolaryngology, Qilu Hospital of Shandong University, Jinan, China, [3]Department of Otorhinolaryngology, Qilu Hospital of Shandong University, Jinan, China, [4]Department of Radiology, Qilu Hospital of Shandong University, Jinan, China

Objectives: This study aimed to identify novel biomarkers for osteoarthritis (OA) and explore potential pathological immune cell infiltration.

Methods: We identified differentially expressed genes (DEGs) between OA and normal synovial tissues using the *limma* package in R, and performed enrichment analyses to understand the functions and enriched pathways of DEGs. Weighted gene co-expression network analysis (WGCNA) and distinct machine-learning algorithms were then used to identify hub modules and candidate biomarkers. We assessed the diagnostic value of the candidate biomarkers using receiver operating characteristic (ROC) analysis. We then used the CIBERSORT algorithm to analyze immune cell infiltration patterns, and the Wilcoxon test to screen out hub immune cells that might affect OA occurrence. Finally, the expression levels of hub biomarkers were confirmed by quantitative reverse transcription-polymerase chain reaction (qRT-PCR).

Results: We identified 102 up-regulated genes and 110 down-regulated genes. The functional enrichment analysis results showed that DEGs are enriched mainly in immune response pathways. Combining the results of the algorithms and ROC analysis, we identified GUCA1A and NELL1 as potential diagnostic biomarkers for OA, and validated their diagnosibility using an external dataset. Construction of a TF-mRNA-miRNA network enabled prediction of potential candidate compounds targeting hub biomarkers. Immune cell infiltration analyses revealed the expression of hub biomarkers to be correlated with CD8 T cells, memory B cells, M0/M2 macrophages, resting mast cells and resting dendritic cells. qRT-PCR results showed both GUCA1A and NELL1 were significantly increased in OA samples ($p < 0.01$). All validations are consistent with the microarray hybridization, indicating that GUCA1A and NELL1 may be involved in the pathogenesis of OA.

Conclusion: The findings suggest that GUCA1A and NELL1, closely related to OA occurrence and progression, represent new OA candidate markers, and that immune cell infiltration plays a significant role in the progression of OA.

Keywords: osteoarthritis, immune microenvironment, biomarkers, machine learning, OMICS data

OPEN ACCESS

Edited by:
Tao Huang,
Shanghai Institute of Nutrition and Health (CAS), China

Reviewed by:
Kexin Yan,
China Medical University, China
Tao Yan,
Fudan University, China

****Correspondence:***
Yuankai Zhang
drzhangyk@163.com
Hui Yang
hermes0@163.com

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *26 March 2022*
Accepted: *11 April 2022*
Published: *16 May 2022*

Citation:
Liu Z, Liu H, Li D, Ma L, Lu T, Sun H, Zhang Y and Yang H (2022) Evaluation of Biomarkers and Immune Microenvironment of Osteoarthritis: Evidence From Omics Data and Machine Learning. Front. Genet. 13:905027. doi: 10.3389/fgene.2022.905027

INTRODUCTION

Characterized by cartilage degeneration, sclerosis of subchondral bone and osteophyte formation, osteoarthritis (OA) is the most common degenerative joint disease (Mathiessen and Conaghan, 2017). Patients with OA experience chronic pain, swelling, malformation and joint stiffness, which may lead to progressive disability and deterioration of patients' quality of life (Parkinson et al., 2017). It is estimated that approximately 9.6% of men and 18% of women worldwide aged 60 years or over suffer from OA, and by 2030 nearly 67 million people living in the United States will have been diagnosed with the disease (Hootman and Helmick, 2006; Li et al., 2017). Unfortunately, current OA therapies cannot prevent or reverse the progress of the disease, and are limited to inhibiting pain and alleviating inflammation (Seed et al., 2011). Advanced patients undergo joint replacement surgery.

Considerable attention has been given to identifying promising biomarkers for disease diagnosis and therapy through transcriptomic and microarray analyses (Demircioğlu et al., 2019; Carr et al., 2020). A noteworthy study has found that the m6A demethylase FTO, which plays a tumor-suppressing role, may be a prospective risk biomarker for thyroid cancer (Tian et al., 2020). Based on the Gene Expression Omnibus (GEO) database, GZMA, PRC1 and TTK were enriched in the innate immune cell-mediated immune response and immune-related biological processes, validating them as potential targets for rheumatoid arthritis (RA) therapy (Cheng et al., 2021). It has also been reported that IFI27 may play a vital role in the occurrence of systemic lupus erythematosus (SLE), and may be a possible target for SLE diagnosis (Zhao et al., 2021). Therefore, it is vital to explore the molecular mechanisms underlying the development and progression of OA, and to identify new and effective biomarkers for its diagnosis and treatment.

In this study, we first acquired differentially expressed genes (DEGs) in OA and normal synovial tissue by mining four GEO datasets (GSE55235, GSE55457, GSE12021 and GSE82107). Next, we conducted a series of enrichment analyses of functions and pathways for these DEGs. To evaluate the key module and to screen out hub biomarkers highly correlated with OA, we performed weighted gene co-expression network analysis (WGCNA) and applied three machine-learning algorithms: least absolute shrinkage and selection operator (LASSO), support vector machine-recursive feature elimination (SVM-RFE) and logistic regression. We validated the selected hub genes using GEO datasets (GSE89408), and verified their diagnostic value with receiver operating characteristic (ROC) curves. A TF-mRNA-miRNA network was then constructed, and potential candidate compounds targeting the biomarkers were predicted. We used the CIBERSORT algorithm and the Wilcoxon test to analyze the difference in immune infiltration between OA and normal tissues and the relationship between biomarkers and infiltrating immune cells, and to identify hub immune cells that might affect OA. Finally, the expression levels of hub biomarkers were confirmed by qRT-PCR. This study strengthens understanding of the mechanisms of development and pathogenesis in OA at the transcriptome level, and provides new insights into potential biomarkers for diagnosis and treatment of OA.

MATERIALS AND METHODS

Data Collection

Gene expression profiles of OA and normal synovial tissue were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) (Barrett et al., 2013). To be eligible for selection, the profiles must have been produced with *Homo sapiens* expression profile analysis using array, and be of OA or normal synovial tissue from joint synovial biopsies, the datasets must contain more than five samples and complete sample information, and each subject had to have one biopsy sample analyzed without duplication. Three GPL96 datasets (GSE55235, GSE55457 and GSE12021), and GSE82107 based on the GPL570 platform, were selected as test sets, including 40 OA samples and 36 normal samples. We downloaded the original GSE89408 count data, a dataset based on the GPL11154 platform, as a validation set (22 OA and 28 normal synovial tissue samples). Patients' clinical features are detailed in **Supplementary Material S1**.

Data Processing and Identification of DEGs

The datasets were combined, and the *sva* package (Leek et al., 2012) was used to normalize the original data (**Supplementary Figure S1**). The DEGs were screened in the batch calibrated test set using the *limma* package (Ritchie et al., 2015). We selected $|\log2 \text{ fold change FC}| > 1$ and adj. p. value < 0.05 as truncation criteria.

Functional Enrichment Analyses

The *GOplot* program package (Walter et al., 2015) was used to visualize the gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and disease ontology (DO) analysis. Terms and pathways with $p < 0.05$ were considered statistically significant. We used the *clusterprofiler* R package (Yu et al., 2012) to conduct GSEA on the DEGs using sequencing data, and the *GSVA* R program (Hänzelmann et al., 2013) to identify pathways most closely associated with DEGs, with p values <0.05 being considered statistically significant. The h. all.v7.4. symbols gene set was downloaded from MSigDB (Liberzon et al., 2015), and GSEA analysis was performed on the gene set and gene expression matrix to explore possible regulatory pathways involved.

WGCNA Network Analysis and Key Module Identification

A co-expression network targeting DEGs was constructed using the *WGCNA* package (Langfelder and Horvath, 2008). In WGCNA analysis, all DEGs with an adjusted p value <0.05 and $|\log2 \text{ fold change FC}| > 1$ in the OA and normal samples were taken as inputs for topology calculation, with soft threshold values ranging from 1 to 20. The β value is determined from the lowest value near scale-free network, and the optimal soft threshold was determined to be 8. Following the optimal soft threshold, the relation matrix was converted into an

adjacency matrix and then into a topological overlap matrix (TOM). We carried out average link hierarchy clustering, and classified relevant modules according to the TOM, with the number of genes in each module being no less than 50. Similar modules were then merged. The Pearson method was used to calculate correlation between the combined module and OA, and hub modules and potential hub genes relating to clinical traits were identified.

Machine Learning-Based Hub Biomarker Screening

Machine-learning classification algorithms are increasingly being used to predict feature genes associated with diseases. LASSO (Engebretsen and Bohlin, 2019) is a regression analysis method for both gene selection and gene classification. In order to avoid collinearity generated by high-dimensional data, redundant genes were eliminated using LASSO's 10-fold cross validation (*GLMNET* package) on the genes screened by WGCNA. SVM-RFE (Yoon and Kim, 2009; Lin et al., 2012) is a machine-learning method based on the support vector machine (SVM), which finds the optimal variable by subtracting the feature vector generated by SVM. The SVM-RFE method was then used on the genes processed by LASSO for further screening to produce the optimal number of genes, while minimizing classification errors and overfitting. We then used univariate logistic regression analysis to screen the genes, with $p < 0.001$ as the threshold. Finally, the DEGs, SVM-RFE-screened genes and logistic regression-screened genes were overlapped to identify hub biomarkers.

Validation of Hub Biomarkers

Expression analysis of the hub biomarkers was performed on the test set. The ROC curves were then plotted using the *pROC* R package (Robin et al., 2011), and the area under the curve (AUC) was calculated separately to evaluate the predictive utility of identified hub genes. Values of AUC >0.7 and $p < 0.05$ indicated that the genes were highly predictive for OA diagnosis. The validation set GSE89408 based on the GPL11154 platform was used to verify the analysis results.

Construction of Regulatory Network

The mirDIP database (Tokar et al., 2018) was used to predict the potential miRNA of targeted hub genes and identify the miR regulatory network. TF-hub gene interactions with p values <0.05 were selected from the TRRUST database (Han et al., 2018) to establish upstream regulatory networks. In addition, compounds with potential relationships to hub genes were searched in the Comparative Toxicogenomics database (Davis et al., 2019). Finally, the hub genes regulatory network was visualized based on the Networkanalyst database (Zhou et al., 2019).

Analysis of Immune Cell Infiltration, Correlations Between Hub Genes and Immune Cell Infiltration

The CIBERSORT algorithm (Chen et al., 2018) was used to calculate the proportions of different immune cell types, according to the expression levels of immune cell-related genes. The output results for 22 infiltrated immune cell types were integrated to generate an immune cell component matrix for analysis. Relationships between hub biomarkers' expression levels and immune cell infiltration were examined using Pearson's rank correlation analyses, conducted and visualized with the *ggpubr* R package.

Identification of Hub Immune Cells

The Wilcoxon test was used to investigate differences in immune cell content between different tissues. The random forest program package was used to construct a random forest tree of the 22 kinds of immune cells to identify the point with the minimum error, and the immune cells were ranked by importance. Genes with importance scores greater than two were selected for screening. The identified immune cells were overlapped to screen out hub immune cells that might affect the occurrence of OA.

qRT-PCR Validation of Hub Biomarkers

In order to confirm the results of bioinformatics analysis, we collected synovial tissues from 10 OA patients and 10 patients without OA for qRT-PCR verification. The study protocol was approved by the ethics committee of Qilu Hospital of Shandong University, and all patients signed informed consent. Total RNA was extracted from synovial tissue using TRIzol® Reagent (15596026, Thermo Fisher Scientific, Inc.). An qRT-PCR kit (K1005S, Promega Biotech Co.) was used to synthesize the first strand of cDNA from equal amounts of total RNA samples, and real-time fluorescence PCR was performed with SYBR Green Realtime PCR Master Mix (QPK-201, TOYOBO Co., Ltd., Kita-ku, Osaka, Japan) according to the manufacturer's protocol. We selected ß-actin as the inner control and employed the $2^{-\Delta\Delta Ct}$ method to quantify the relative mRNA level. The sequences of NELL1 were as follows: TCACAGGAAGCCACTGCGAGAA (sense) and CCATCGTCATGGAAACCGCTTC (antisense). The sequences of GUCA1A were as follows: GCAGAGGAGTTC ACCGATACAG (sense) and GTCAGTGTGTCCAGGAGC ATCT (antisense). The sequences of ß-actin were as follows: CACCATTGGCAATGAGCGGTTC (sense) and AGGTCTTTGCGGATGTCCACGT (antisense). One-way analysis of variance was used for the statistical analysis, and $p < 0.05$ indicated a significant difference.

RESULTS

Differentially Expressed Genes Between OA and Normal Synovial Tissues

We analyzed the DEGs of 40 OA and 36 normal synovium samples in the test set (GSE55235, GSE55457, GSE12021 and GSE82107 datasets), and identified a total of 212 DEGs in the OA samples compared with the normal group (**Supplementary Material S2**). **Figure 1A** shows a heat map of the top 20 differential genes by log fold change, and the volcano map in **Figure 1B** illustrates that 102 genes were significantly up-regulated and 110 significantly down-regulated.

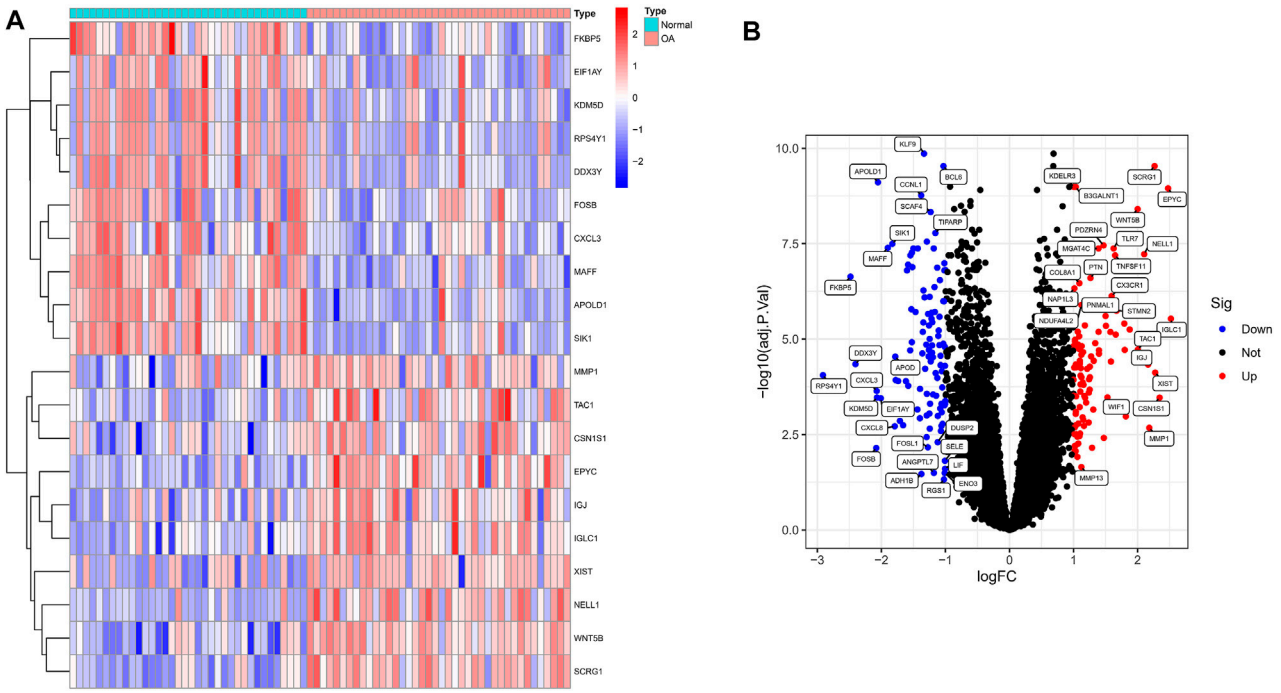


FIGURE 1 | Identification of significant differentially expressed genes (DEGs) in OA. **(A)** Heatmap of DEGs between OA and normal samples. **(B)** Volcano plot of DEGs between OA and normal samples. Red rectangles/plots represent up-regulated genes and blue rectangles/plots represent down-regulated genes.

Enrichment Analysis of DEGs

Next, in order to explore the potential biological mechanism of OA progression, we observed the enrichment pathway of DEGs from multiple perspectives. DO analysis revealed disease types that may share a common pathogenesis with OA, such as pre-eclampsia, periodontal disease and dental disease (**Figure 2A**). GO enrichment analysis of the DEGs showed that the immune response in OA samples was stronger than in the normal sample, including regulation of leukocyte migration and myeloid leukocyte migration. The top 10 biological processes were selected, with *Q* values <0.05, as shown in **Figure 2B**. KEGG pathway enrichment analysis showed related genes involved in, for example, the IL-17 signaling pathway, cytokine-cytokine receptor interaction and the TNF signaling pathway (**Figure 2C**). These results indicated that immune-related factors may affect the progression of OA, GSEA analysis was performed on the gene set and expression matrix. The results demonstrated that hypoxia, IL-2-STAT5 signaling and other pathways play an important role (**Figure 2D**). These strong evidence chains suggest that OA may be regulated by immune pathways.

Further Screening With WGCNA Analysis

To further correlate clinical information with key genes, we performed WGCNA analysis. The clustering of each sample was good, with no outlier samples. A soft threshold from 1 to 20 was used for topology calculation, and the optimal soft threshold was determined to be 8 (**Figure 3A**). Using the soft threshold, the relational matrix was transformed into an adjacency matrix, and then into a topological overlap matrix (TOM) to determine average link hierarchical clustering. Related modules were classified according to the TOM, with the number of genes in each module being no less than 50 (**Figure 3B**). Similar gene modules were merged, resulting in eight modules (**Figure 3C**). Correlation between genes and clinical traits in the module was calculated, revealing that the blue module containing 1,776 genes exhibited the highest positive correlation with OA occurrence (r = 0.73), and the grey module containing 128 genes had the highest negative correlation with OA occurrence (R = −0.84). A total of 1,904 potential core genes were identified.

Exploration of Hub Biomarkers

Next, we applied a series of machine-learning algorithms—LASSO, SVM-RFE and logistic regression—to screen the most significant genes associated with OA. A total of 1,904 potential hub genes screened by WGCNA in OA patients were selected between the two groups to fit the LASSO regression model. The next step was to use LASSO's 10-fold cross-validation to remove any further redundant genes, as a result of which 33 potential genes with non-zero coefficients identified in OA and normal cohorts were screened out (**Figure 3D**). We used the SVM-RFE algorithm for in-depth screening of these 33 genes. The results showed that the RMSE value was lowest when 19 genes were selected as variables (**Figure 4A**). Taking occurrence of OA as the dependent variable, univariate logistic analysis was then carried out. The forest map produced 25 genes with *p* values less than 0.001 (**Figure 4B**). Finally, we overlapped the genes of the two previous identification algorithms with 212 DEGs, and

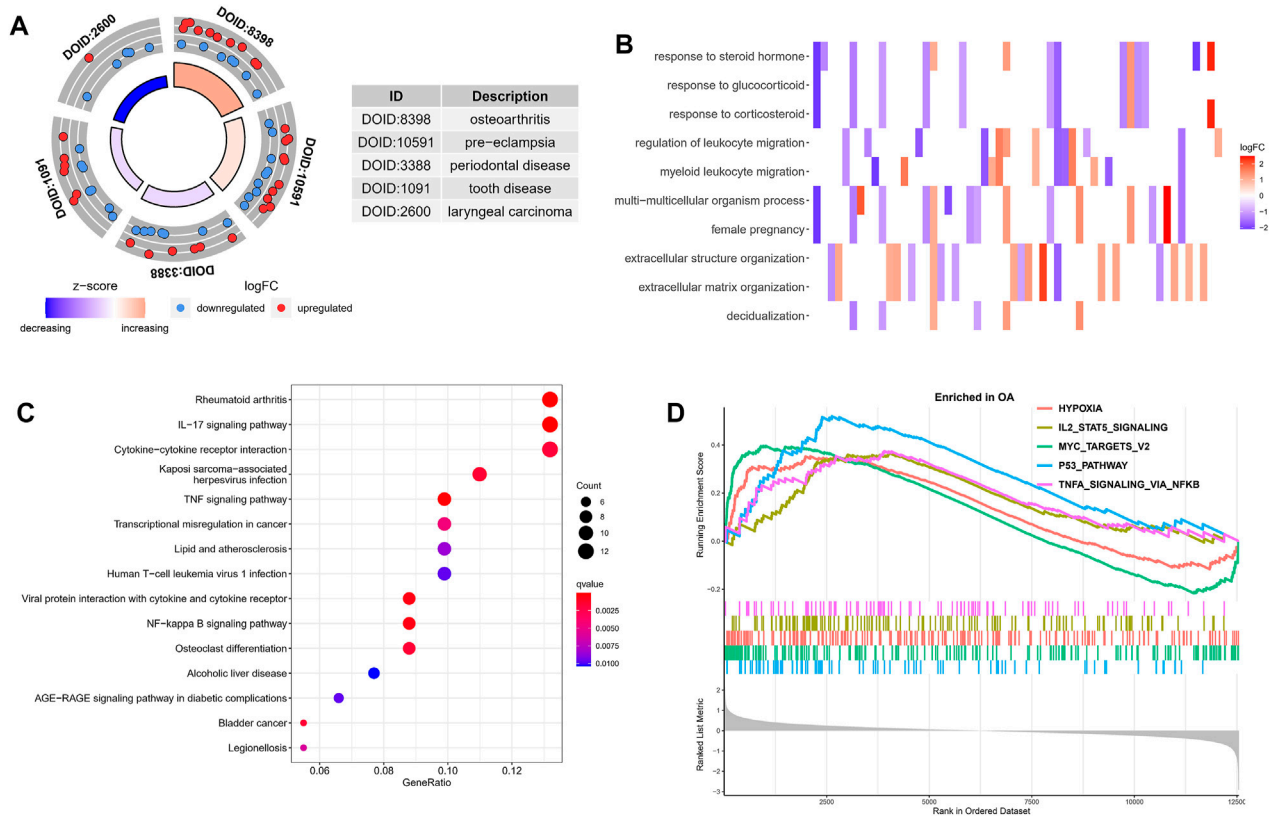


FIGURE 2 | Functional enrichment analysis of DEGs. **(A)** DO analysis results for disease types that may share a common pathogenesis with OA. **(B)** The top 10 enriched biological processes of DEGs identified using GO analysis. **(C)** KEGG enrichment analysis based on the DEGs. The gradual bubble color represents the adjusted *p* value, and the bubble size represents the gene count. **(D)** GSEA analysis of DEGs in the OA and normal groups.

identified GUCA1A and NELL1 as our hub biomarkers (**Figure 4C**).

Validation of Hub Biomarkers

ROC and differential expression analysis were performed on GUCA1A and NELL1, respectively. The results showed that these genes had good predictive performance in the test set: both GUCA1A (AUC = 0.822) and NELL1 (AUC = 0.871) were significantly over-expressed in the OA samples (**Figures 5A,B**). External validation using the GSE89408 dataset showed that the expressions of GUCA1A and NELL1 were similar to the test set, with both being up-regulated in OA tissues, and also had strong diagnostic performance (GUCA1A, AUC = 0.747; NELL1, AUC = 0.713) (**Figures 5C,D**). These results indicated that expressions of GUCA1A and NELL1 were highly correlated with OA progression, and that these genes may act as biomarkers to diagnose and verify effective treatment of OA.

TF-mRNA-miRNA Network Analysis and Prediction of Potential Candidate Compounds

Regulatory networks play a key role in understanding disease mechanisms. We used the TRRUST and mirDIP databases to predict interactions between hub biomarkers and transcription factors (TFs) as well as miRNA. A TF-mRNA-miRNA triple network was then constructed. We found 3 TFs and 26 miRNAs targeting NELL1, and identified two TF–GUCA1A pairs and three miRNA–GUCA1A pairs (**Figure 6A**). This network revealed hub nodes and their interactions associated with the molecular mechanisms of OA, and indicated that NELL1 and GUCA1A are related to multiple regulatory networks in OA progression. These two hub biomarkers may play a crucial role in the pathological process of OA. This enabled us to predict potential candidate compounds targeting GUCA1A and NELL1 that may alleviate OA patients' symptoms (**Figure 6B**).

Analysis of Differences in Immune Microenvironment

In view of the important role of immune-related pathways in the occurrence of OA in enrichment analysis (**Figure 2**), the CIBERSORT algorithm was used to analyze the content of immune cells in the different samples. The bar chart in **Figure 7A** illustrates the overall landscape of immune cell distribution, and the heat map in **Figure 7B** details the correlations of 22 immune cell types. The Wilcoxon test showed that the OA samples contained more memory B cells,

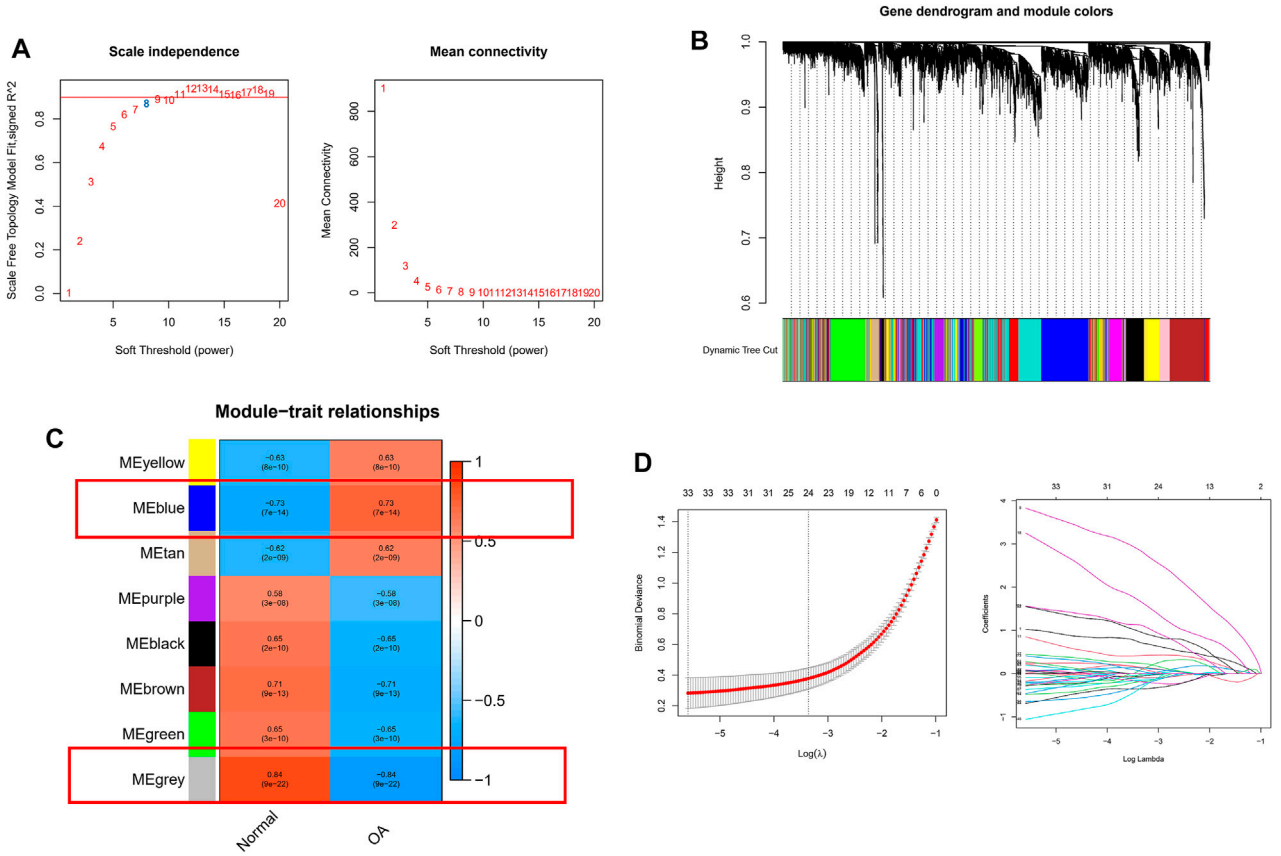


FIGURE 3 | WGCNA-based identification of co-expression modules from merged datasets. **(A)** Soft thresholds (β) and scale-free topology fitting indices (R^2). To maximize model fit, a β value of eight was chosen. **(B)** Dendrograms generated via average linkage hierarchical clustering of identified modules. **(C)** Associations between modules' clinical status in normal and OA patient samples, with individual rows corresponding to module eigengenes and columns corresponding to clinical characteristics. Correlations and p values are shown in the first and second lines of each cell, respectively. **(D)** Coefficient profile plot generated against the log(lambda) sequence, using the LASSO logistic regression algorithm to screen diagnostic markers. Different colors represent different genes.

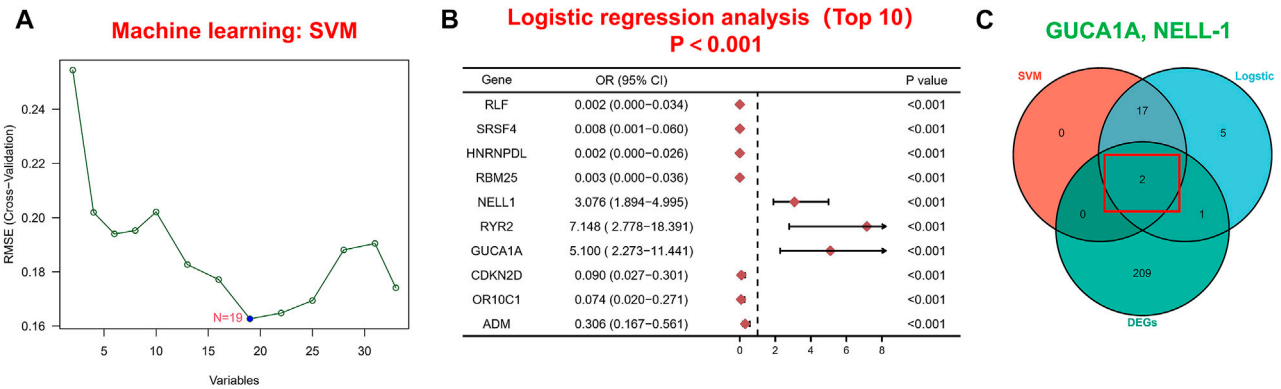


FIGURE 4 | Machine learning-based hub biomarker screening. **(A)** Diagnostic marker screening using the SVM-RFE algorithm. **(B)** The 10 most significant genes identified by univariate logistic analysis (p < 0.001). **(C)** Overlapping genes predicted by the DEGs, logistic regression and SVM-RFE algorithms.

plasma cells, M0 macrophages and resting mast cells. Compared with OA tissues, normal tissues had higher contents only of resting CD4 memory T cells and activated NK cells (**Figure 8A**). In addition, to identify hub immune cells that alter the immune microenvironment in OA synovial tissues, we performed random forest tree analysis on 22 immune cells (**Figures 8B,C**) and

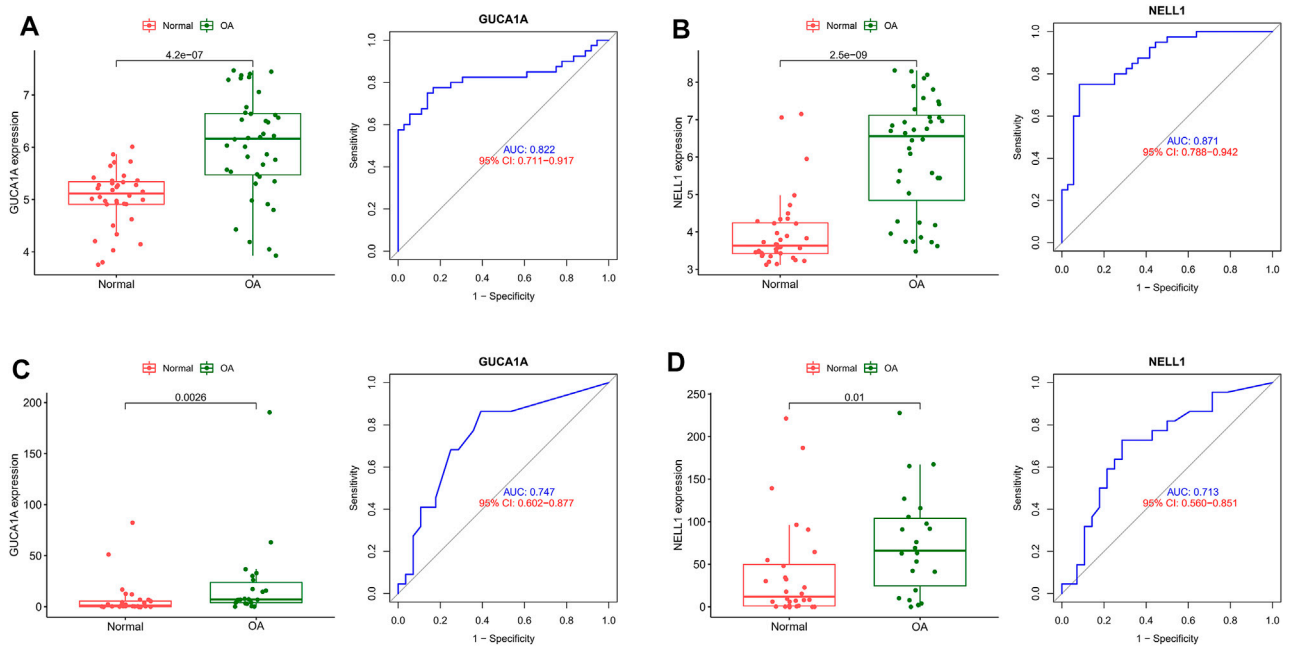


FIGURE 5 | Validation of hub biomarkers. **(A)** Expression and diagnostic value of GUCA1A in OA using the test set. **(B)** Expression and diagnostic value of NELL1 in OA using the test set. **(C)** Expression and diagnostic value of GUCA1A in OA using the validation set. **(D)** Expression and diagnostic value of NELL1 in OA using the validation set.

overlapped the Wilcoxon test with the immune cells identified in random forest trees. Finally, we identified six types of hub immune cells that may affect the occurrence of OA (**Figure 8D**): activated NK cells, activated mast cells, plasma cells, M0 macrophages, resting mast cells and memory B cells.

Correlation Analysis of Immune Cells and Hub Biomarkers

Correlation analysis between 22 kinds of immune cells and two hub biomarkers in OA tissue produced statistically significant results. GUCA1A is negatively correlated with memory B cells and resting mast cells, while M0 and M2 macrophages are positively correlated (**Figure 9A**). In addition, NELL1 is positively correlated with CD8 T cells and M0 macrophages, and negatively correlated with resting dendritic cells (**Figure 9B**).

qRT-PCR Validation of Hub Biomarkers

We performed radiological evaluations on the knee joints of different groups of patients to verify the typical imaging manifestations of OA. The results showed that compared with healthy donors, preoperative knee X-ray and gross images of OA patients showed bone spur, subchondral sclerosis and narrowing of joint space (**Figure 10A**). Then, qRT-PCR was used to detect the expression level of hub biomarkers in the OA synovium of the knee joint and normal control groups. Statistical analysis proved that both GUCA1A and NELL1 were significantly over-expressed in the synovium of OA samples ($p < 0.01$) (**Figure 10B**). All validations are consistent with the microarray hybridization, indicating that GUCA1A and NELL1 may be involved in the pathogenesis of OA.

DISCUSSION

OA is a chronic degenerative joint disease that causes irreversible bone erosion and cartilage destruction, and is one of the most common causes of disability (Wight et al., 2017; Smolen et al., 2018). However, because the pathophysiological mechanism of OA is unclear and effective biomarkers are lacking, diagnosis and treatment of OA is difficult. This study is the first to integrate WGCNA and machine-learning algorithms to identify new biomarkers associated with OA, and to explore the role of immune cell infiltration in OA using CIBERSORT and Wilcoxon tools.

In this study, we downloaded four gene expression profiles (GSE55235, GSE55457, GSE12021, and GSE82107) from the GEO database, and identified a total of 212 DEGs, including 102 up-regulated and 110 down-regulated genes in the OA sample compared with the normal sample. Then, we investigated the biological functions of these common DEGs and KEGG analysis revealed these genes to be enriched in the IL-17 signaling pathway and TNF signaling pathway, both of which are inflammatory. IL-17 and TNF are pro-inflammatory cytokines that are closely associated with cartilage destruction, cartilage matrix degradation and bone resorption (Kenna and Brown, 2013; Wang and He, 2018), both of which are promising therapeutic targets related to OA, which is consistent with our findings and highlights the correlation between these gene and the pathogenesis of OA. Our GO enrichment analysis of DEGs suggested that immune responses, such as regulation of leukocyte migration and myeloid leukocyte migration, were stronger in OA samples than in normal tissues. OA is a chronic inflammation

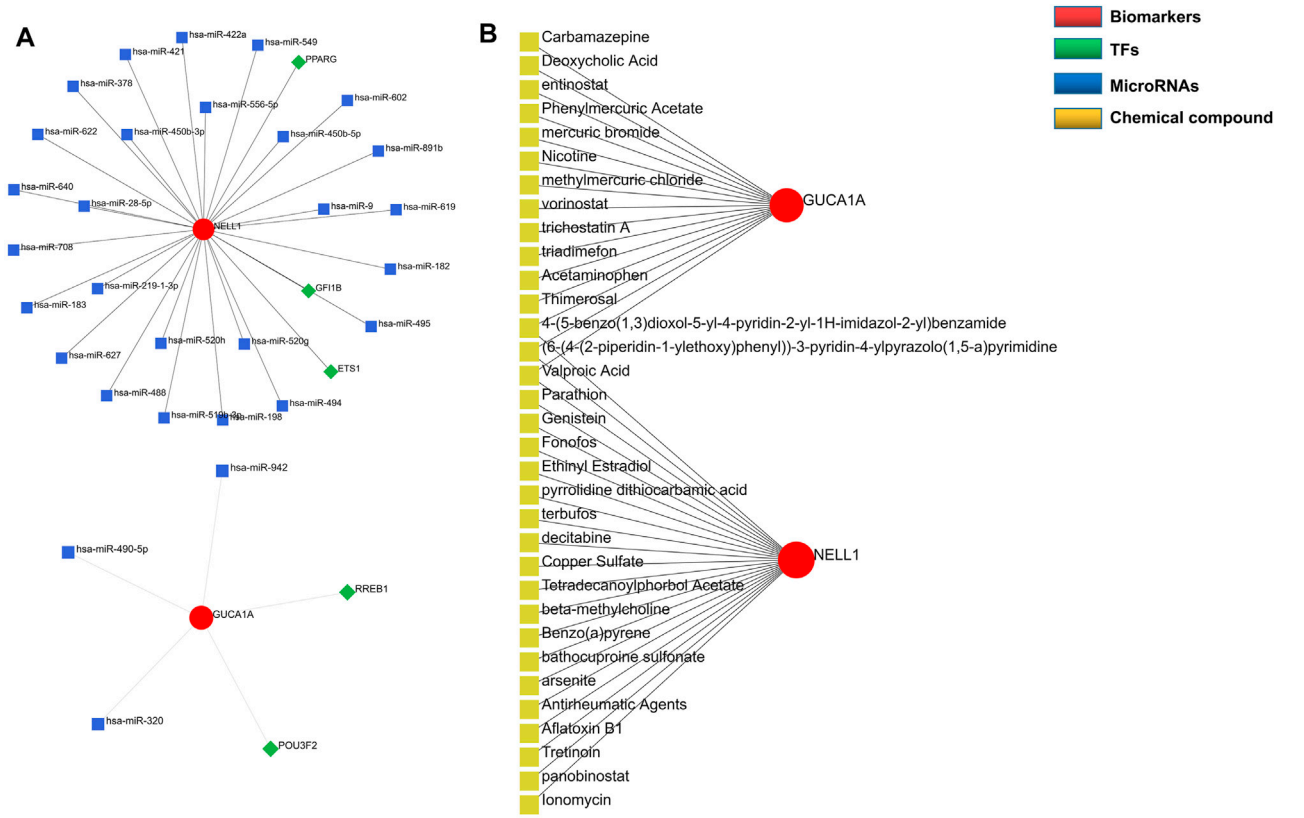


FIGURE 6 | Construction of regulatory network. **(A)** TF-mRNA-miRNA network of hub biomarkers. **(B)** Potential candidate compounds targeting GUCA1A and NELL1.

disease, which closely related to immune cell infiltration of bone and cartilage (Haseeb and Haqqi, 2013), and immune cell infiltration of OA synovial tissue correlates with OA disease progression and pain (Lopes et al., 2017). These results confirm that immune-related factors may affect the progression of OA, so continued efforts to identify OA-related immune cell infiltration may be of value for treatment of this disease.

To better understand the progression of OA, hub modules and candidate biomarkers of OA were identified using WGCNA. WGCNA analysis indicated that the blue and grey modules containing 1,904 genes were most strongly correlated with clinical characteristics of OA. Using three different machine-learning algorithms (LASSO, SVM-RFE and logistic regression), GUCA1A and NELL1 were identified as hub genes, which were statistically significantly over-expressed in the OA samples ($p < 0.05$). ROC curve analysis showed that GUCA1A and NELL1 had high sensitivity and specificity in OA diagnosis in both the test set and the validation set. We also constructed a TF-mRNA-miRNA network, enabling us to predict potential candidate compounds targeting GUCA1A and NELL1 to elucidate the pathogenesis of OA at the transcriptome level. NEL-like molecule-1 (NELL1), a new secretory protein originally identified in unilateral coronal craniosynostosis in humans (Ting et al., 1999), plays an important role in osteogenic differentiation, bone regeneration, chondrogenesis and inflammation (Aghaloo et al., 2006; Lee et al., 2010; Cao et al., 2021). Recent research has reported the ability of NELL1 to induce chondrogenesis and an anti-inflammatory response in OA through up-regulation of runt-related transcription factor 1 (RUNX1), making it a potential candidate for articular cartilage repair (Li et al., 2020). In our study, compared with healthy patients, we found that the expression level of NELL1 increases in OA synovial tissues. The ROC curve for NELL1 indicated that it has good predictive performance in OA (AUC >0.87), suggesting that NELL1 plays a significant role in the progression of OA. OA is considered to be an inflammatory disease of the joint cartilage and is caused by multiple factors. Inflammatory cytokines are mainly expressed in OA, but under pathological conditions, the body still has some anti-inflammatory gene expression for self-protection and repair. When the ultimate balance between anti-inflammatory and pro-inflammatory is broken, anti-inflammatory genes can't work as effectively (Wojdasiewicz et al., 2014). NELL1 may play this role, and the expression of NELL1 may be the body's self-protection regulation. Li et al. found that NELL1-haploinsufficient (NELL1+/6R) mice showed elevated inflammatory markers and accelerated progression of OA. After intra-articular injection of NELL1, the IL1β-induced inflammation and cartilage degradation were rescued obviously (Li et al., 2020). The above results remind us NELL1 may be a

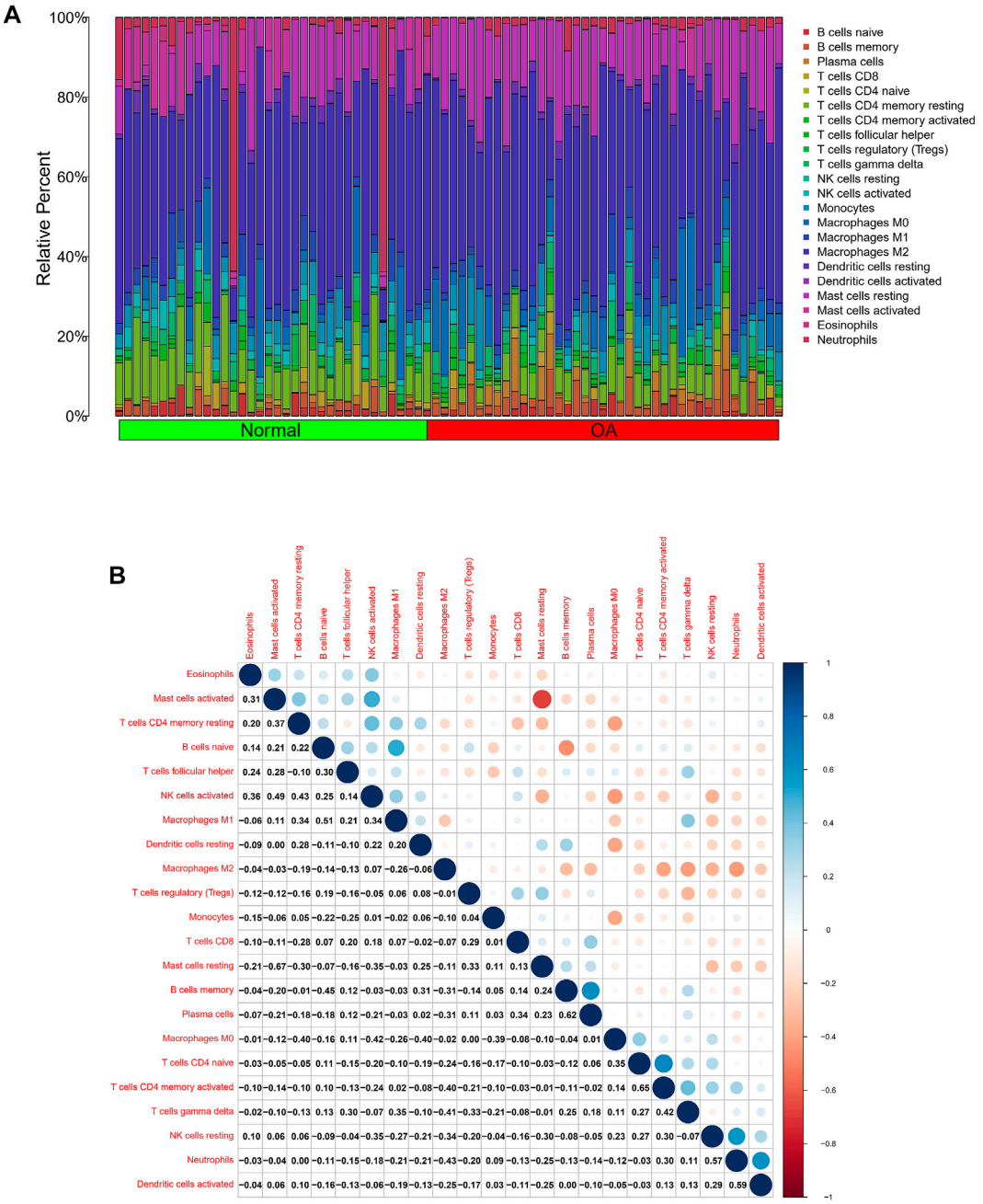


FIGURE 7 | Analysis of differences in immune microenvironment. **(A)** Relative proportions of synovial tissue infiltration by 22 different immune cell subtypes. **(B)** Correlations among 22 different immune cell populations, with blue and red indicating positive and negative correlations, respectively. White indicates absence of any correlation between the indicated immune cell populations.

promising target for precise treatment of OA for suppressing inflammation and arthritis-related cartilage damage. The guanylate cyclase activator 1A gene (GUCA1A), located in 6p21.1, encodes guanylyl cyclase-activating protein 1 (GCAP1), and has been identified as being involved in dominant cone dystrophy, cone-rod dystrophy and macular dystrophy (Payne et al., 1998; Wilkie et al., 2000). However, GUCA1A has not previously been reported in OA-related studies. We have identified that GUCA1A is also highly specifically expressed in the synovial membrane of OA and has a high diagnostic value (AUC >0.82) for OA. These results were validated using the GSE89408 dataset. Finally, the results of qRT-PCR showed both GUCA1A and NELL1 were significantly increased in OA samples ($p < 0.01$). All validations are consistent with the microarray hybridization, indicating that GUCA1A and NELL1 may be involved in the

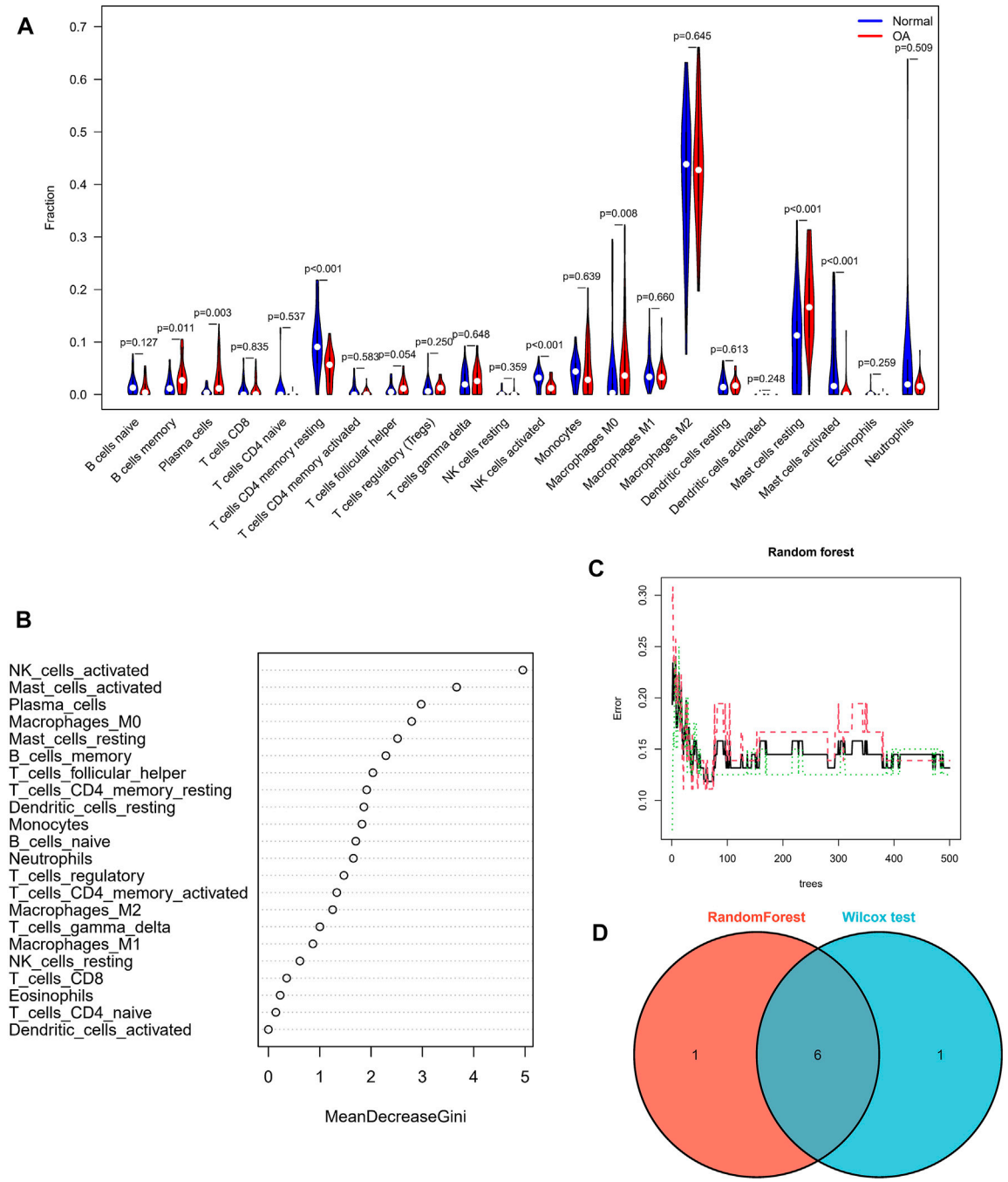


FIGURE 8 | Assessment of immune cell infiltration. **(A)** Comparisons of 22 immune cell types, with blue and red indicating normal and OA tissues, respectively. **(B,C)** Random forest tree analysis of 22 immune cells. **(D)** Overlapping immune cells predicted by the immune cell and Wilcoxon test.

pathogenesis of OA, and thus we consider NELL1 and GUCA1A to be very effective biomarkers for OA diagnosis.

To explore the potential pathological relevance of immune cell infiltration in this disease, we used the CIBERSORT algorithm to conduct a comprehensive evaluation of OA immune infiltration, which provided insights into how these immune cells affect OA pathology. We found that increased infiltration of memory B cells, plasma cells, resting mast cells and M0 macrophages, and decreased infiltration of resting memory CD4 T cells and activated NK cells may be related to the occurrence and development of OA. De Lange-Brokaar et al. have found that mast cell content is significantly higher in OA samples compared with RA, and is associated with structural damage in OA patients, suggesting the role of mast cells in this disease (De Lange-Brokaar et al., 2016). Previous studies have shown that

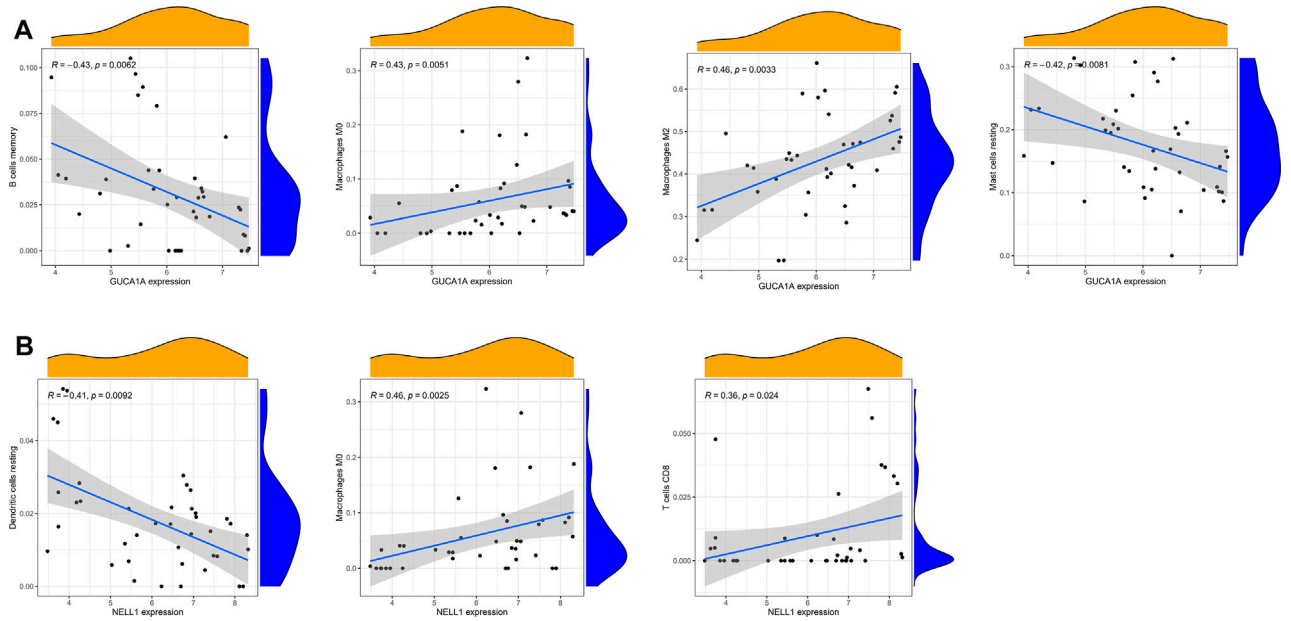


FIGURE 9 | Correlation analysis of immune cells and hub biomarkers. **(A)** Correlation between GUCA1A and infiltrating immune cells. **(B)** Correlation between NELL1 and infiltrating immune cells. $p < 0.05$ considered statistically significant.

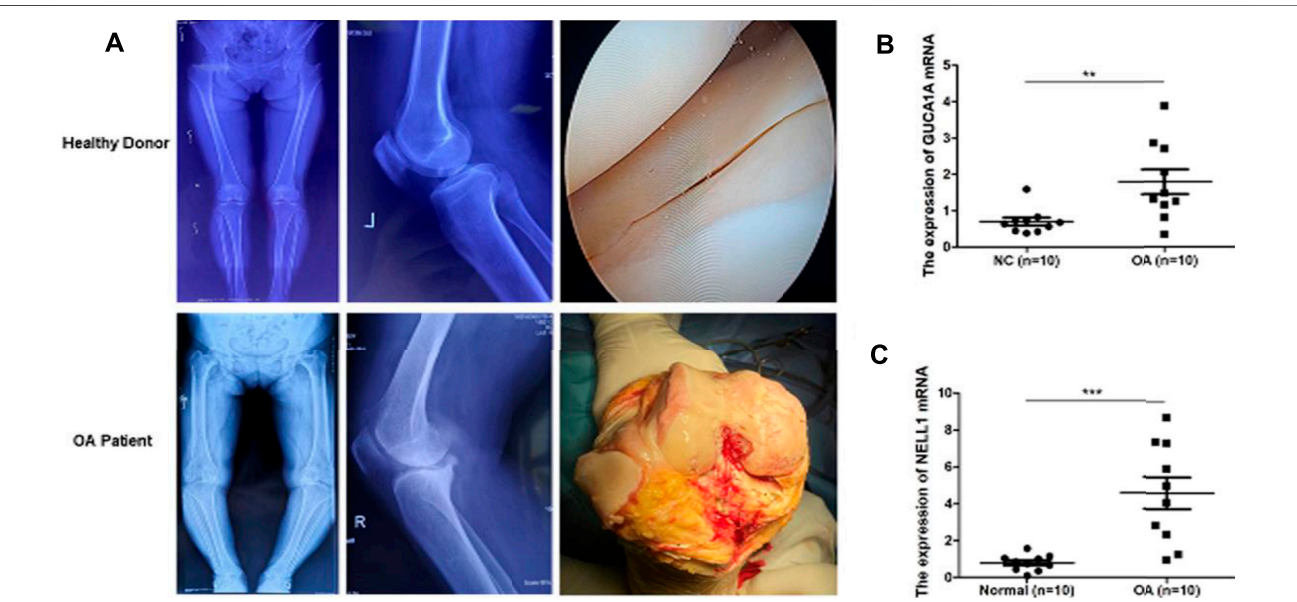


FIGURE 10 | qRT-PCR validation of hub biomarkers. **(A)** X ray images and macroscopic views (arthroscopic image and intraoperative image) of knee joint from healthy donors and OA patients. **(B,C)** Validation of GUCA1A and NELL1 by qRT-PCR between the OA group ($n = 10$) and the control group ($n = 10$). Data are mean ± SEM; $^{**}p < 0.01$, and $^{***}p < 0.001$.

macrophages may regulate joint inflammation and OA severity through various secretory mediators, and the modulation of macrophage functional phenotypes appears to be an effective treatment option to prevent OA or enhance cartilage repair and regeneration (Wu et al., 2020; Zhang et al., 2020). Studies indicate that accumulation of memory CD4 T cells is a common phenomenon during the local inflammatory response of RA and OA joints, and is involved in the progression of OA (Ezawa et al., 1997). Increasing evidence suggests that NK cells are key to promoting immune cells in OA, and that their interaction is promoted by the CXCL10/CXCR3 axis (Benigni et al., 2017). Our analysis results combined with the above literature evidence have shown that resting mast cells, M0

macrophages, resting memory CD4 T cells and activated NK cells play important roles in OA, which should be the focus of further research. However, no research has been conducted on the role of memory B cells and plasma cells in OA, and further experimental data are required. In our study, to screen out hub immune cells that might alter the immune microenvironment in OA synovial tissues, we performed random forest tree analysis on 22 immune cells, and overlapped the Wilcoxon test with the immune cells identified in random forest trees. We have identified six types of hub immune cells that may affect the occurrence of OA: activated NK cells, activated mast cells, resting mast cells, memory B cells, plasma cells and M0 macrophages. In addition, associations between GUCA1A, NELL1 and immune cells revealed these genes to be correlated with levels of CD8 T cells, memory B cells, resting mast cells, resting dendritic cells, and M0 and M2 macrophages. We hypothesize that GUCA1A and NELL1 may be involved in the occurrence and progression of OA by mediating the above immune cells, and further studies are needed to clarify the complex interaction between genes and immune cells. The above results suggest that various infiltrating immune cells play key roles in OA pathogenesis.

Some potential limitations of this study must be considered when interpreting the results. The CIBERSORT analysis was based on limited genetic data that may deviate from heterotypic interactions of cells, disease-induced disorders or phenotypic plasticity. The exact mechanisms of NELL1 and GUCA1A in regulating the initiation and progression of OA require further investigation, and further experimental studies are needed to validate the findings of this study.

CONCLUSION

In conclusion, we identify GUCA1A and NELL1 as diagnostic biomarkers of OA, and find that memory B cells, plasma cells, resting mast cells, M0 macrophages, resting CD4 memory T cells and activated NK cells may relate to the occurrence and progression of OA. These immune cells and immune-related genes may be potential immunotherapeutic targets for patients with OA.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethics committee of Qilu Hospital of Shandong University. 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

YZ and HY designed experiments and wrote the paper. ZL, HL, DL, and LM carried out bioinformatics analysis and prepared the manuscript. TL and HS helped to perform qRT-PCR validation. All authors read and approved the final manuscript.

FUNDING

This work was supported by Key R&D projects in Shandong Province (grant number: 2019GSF108205).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.905027/full#supplementary-material

REFERENCES

Aghaloo, T., Cowan, C. M., Chou, Y.-F., Zhang, X., Lee, H., Miao, S., et al. (2006). Nell-1-induced Bone Regeneration in Calvarial Defects. *Am. J. Pathol.* 169, 903–915. doi:10.2353/ajpath.2006.051210

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

Benigni, G., Dimitrova, P., Antonangeli, F., Sanseviero, E., Milanova, V., Blom, A., et al. (2017). CXCR3/CXCL10 axis Regulates Neutrophil-NK Cell Cross-Talk Determining the Severity of Experimental Osteoarthritis. *J.I.* 198, 2115–2124. doi:10.4049/jimmunol.1601359

Cao, R., Wang, Q., Wu, J., Liu, M., Han, Q., and Wang, X. (2021). Nell-1 Attenuates Lipopolysaccharide-Induced Inflammation in Human Dental Pulp Cells. *J. Mol. Histol.* 52, 671–680. doi:10.1007/s10735-021-09976-y

Carr, H. L., Turner, J. D., Major, T., Scheel-Toellner, D., and Filer, A. (2020). New Developments in Transcriptomic Analysis of Synovial Tissue. *Front. Med.* 7, 21. doi:10.3389/fmed.2020.00021

Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., and Alizadeh, A. A. (2018). Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol. Biol.* 1711, 243–259. doi:10.1007/978-1-4939-7493-1_12

Cheng, Q., Chen, X., Wu, H., and Du, Y. (2021). Three Hematologic/immune System-specific Expressed Genes Are Considered as the Potential Biomarkers for the Diagnosis of Early Rheumatoid Arthritis through Bioinformatics Analysis. *J. Transl Med.* 19, 18. doi:10.1186/s12967-020-02689-y

Davis, A. P., Grondin, C. J., Johnson, R. J., Sciaky, D., McMorran, R., Wiegers, J., et al. (2019). The Comparative Toxicogenomics Database: Update 2019. *Nucleic Acids Res.* 47, D948–D954. doi:10.1093/nar/gky868

De Lange-Brokaar, B. J. E., Kloppenburg, M., Andersen, S. N., Dorjée, A. L., Yusuf, E., Herb-van Toorn, L., et al. (2016). Characterization of Synovial Mast Cells in Knee Osteoarthritis: Association with Clinical Parameters. *Osteoarthritis and Cartilage* 24, 664–671. doi:10.1016/j.joca.2015.11.011

Demircioğlu, D., Cukuroglu, E., Kindermans, M., Nandi, T., Calabrese, C., Fonseca, N. A., et al. (2019). A Pan-Cancer Transcriptome Analysis Reveals Pervasive Regulation through Alternative Promoters. *Cell* 178, 1465–1477. doi:10.1016/j.cell.2019.08.018

Engebretsen, S., and Bohlin, J. (2019). Statistical Predictions with Glmnet. *Clin. Epigenet* 11, 123. doi:10.1186/s13148-019-0730-1

Ezawa, K., Yamamura, M., Matsui, H., Ota, Z., and Makino, H. (1997). Comparative Analysis of CD45RA- and CD45RO-Positive CD4+T Cells in Peripheral Blood, Synovial Fluid, and Synovial Tissue in Patients with Rheumatoid Arthritis and Osteoarthritis. *Acta Med. Okayama* 51, 25–31. doi:10.18926/amo/30810

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

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

Haseeb, A., and Haqqi, T. M. (2013). Immunopathogenesis of Osteoarthritis. *Clin. Immunol.* 146, 185–196. doi:10.1016/j.clim.2012.12.011

Hootman, J. M., and Helmick, C. G. (2006). Projections of US Prevalence of Arthritis and Associated Activity Limitations. *Arthritis Rheum.* 54, 226–229. doi:10.1002/art.21562

Kenna, T. J., and Brown, M. A. (2013). The Role of IL-17-secreting Mast Cells in Inflammatory Joint Disease. *Nat. Rev. Rheumatol.* 9, 375–379. doi:10.1038/nrrheum.2012.205

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

Lee, M., Siu, R. K., Ting, K., and Wu, B. M. (2010). Effect of Nell-1 Delivery on Chondrocyte Proliferation and Cartilaginous Extracellular Matrix Deposition. *Tissue Eng. A* 16, 1791–1800. doi:10.1089/ten.tea.2009.0384

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

Li, C., Zheng, Z., Ha, P., Jiang, W., Berthiaume, E. A., Lee, S., et al. (2020). Neural EGFL like 1 as a Potential Pro-chondrogenic, Anti-inflammatory Dual-Functional Disease-Modifying Osteoarthritis Drug. *Biomaterials* 226, 119541. doi:10.1016/j.biomaterials.2019.119541

Li, M. H., Xiao, R., Li, J. B., and Zhu, Q. (2017). Regenerative Approaches for Cartilage Repair in the Treatment of Osteoarthritis. *Osteoarthritis and Cartilage* 25, 1577–1587. doi:10.1016/j.joca.2017.07.004

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cel Syst.* 1, 417–425. doi:10.1016/j.cels.2015.12.004

Lin, X., Yang, F., Zhou, L., Yin, P., Kong, H., Xing, W., et al. (2012). A Support Vector Machine-Recursive Feature Elimination Feature Selection Method Based on Artificial Contrast Variables and Mutual Information. *J. Chromatogr. B* 910, 149–155. doi:10.1016/j.jchromb.2012.05.020

Lopes, E. B. P., Filiberti, A., Husain, S. A., and Humphrey, M. B. (2017). Immune Contributions to Osteoarthritis. *Curr. Osteoporos. Rep.* 15, 593–600. doi:10.1007/s11914-017-0411-y

Mathiessen, A., and Conaghan, P. G. (2017). Synovitis in Osteoarthritis: Current Understanding with Therapeutic Implications. *Arthritis Res. Ther.* 19, 18. doi:10.1186/s13075-017-1229-9

Parkinson, L., Waters, D. L., and Franck, L. (2017). Systematic Review of the Impact of Osteoarthritis on Health Outcomes for Comorbid Disease in Older People. *Osteoarthritis and Cartilage* 25, 1751–1770. doi:10.1016/j.joca.2017.07.008

Payne, A., Downes, S. M., Bessant, D. A., Taylor, R., Holder, G. E., Warren, M. J., et al. (1998). A Mutation in Guanylate Cyclase Activator 1A (GUCA1A) in an Autosomal Dominant Cone Dystrophy Pedigree Mapping to a New Locus on Chromosome 6p21.1. *Hum. Mol. Genet.* 7, 273–277. doi:10.1093/hmg/7.2.273

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

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., et al. (2011). pROC: an Open-Source Package for R and S+ to Analyze and Compare ROC Curves. *BMC Bioinformatics* 12, 7. doi:10.1186/1471-2105-12-77

Seed, S. M., Dunican, K. C., and Lynch, A. M. (2011). Treatment Options for Osteoarthritis: Considerations for Older Adults. *Hosp. Pract.* 39, 62–73. doi:10.3810/hp.2011.02.375

Smolen, J. S., Aletaha, D., Barton, A., Burmester, G. R., Emery, P., Firestein, G. S., et al. (2018). Rheumatoid Arthritis. *Nat. Rev. Dis. Primers* 4, 18001. doi:10.1038/nrdp.2018.1

Tian, R., Zhang, S., Sun, D., Bei, C., Li, D., Zheng, C., et al. (2020). M6A Demethylase FTO Plays a Tumor Suppressor Role in Thyroid Cancer. *DNA Cel Biol.* 39, 2184–2193. doi:10.1089/dna.2020.5956

Ting, K., Vastardis, H., Mulliken, J. B., Soo, C., Tieu, A., Do, H., et al. (1999). Human Nell-1 Expressed in Unilateral Coronal Synostosis. *J. Bone Miner Res.* 14, 80–89. doi:10.1359/jbmr.1999.14.1.80

Tokar, T., Pastrello, C., Rossos, A. E. M., Abovsky, M., Hauschild, A.-C., Tsay, M., et al. (2018). mirDIP 4.1-integrative Database of Human microRNA Target Predictions. *Nucleic Acids Res.* 46, D360–D370. doi:10.1093/nar/gkx1144

Walter, W., Sánchez-Cabo, F., and Ricote, M. (2015). GOplot: an R Package for Visually Combining Expression Data with Functional Analysis: Fig. 1. *Bioinformatics* 31, 2912–2914. doi:10.1093/bioinformatics/btv300

Wang, T., and He, C. (2018). Pro-inflammatory Cytokines: The Link between Obesity and Osteoarthritis. *Cytokine Growth Factor. Rev.* 44, 38–50. doi:10.1016/j.cytogfr.2018.10.002

Wight, L., Owen, D., Goldbloom, D., and Knupp, M. (2017). Pure Ankle Dislocation: a Systematic Review of the Literature and Estimation of Incidence. *Injury* 48, 2027–2034. doi:10.1016/j.injury.2017.08.011

Wilkie, S. E., Newbold, R. J., Deery, E., Walker, C. E., Stinton, I., Ramamurthy, V., et al. (2000). Functional Characterization of Missense Mutations at Codon 838 in Retinal Guanylate Cyclase Correlates with Disease Severity in Patients with Autosomal Dominant Cone-Rod Dystrophy. *Hum. Mol. Genet.* 9, 3065–3073. doi:10.1093/hmg/9.20.3065

Wojdasiewicz, P., Poniatowski, Ł. A., and Szukiewicz, D. (2014). The Role of Inflammatory and Anti-inflammatory Cytokines in the Pathogenesis of Osteoarthritis. *Mediators Inflamm.* 2014, 1–19. doi:10.1155/2014/561459

Wu, C.-L., Harasymowicz, N. S., Klimak, M. A., Collins, K. H., and Guilak, F. (2020). The Role of Macrophages in Osteoarthritis and Cartilage Repair. *Osteoarthritis and Cartilage* 28, 544–554. doi:10.1016/j.joca.2019.12.007

Yoon, S., and Kim, S. (2009). AdaBoost-based Multiple SVM-RFE for Classification of Mammograms in DDSM. *BMC Med. Inform. Decis. Mak* 9, S1. doi:10.1186/1472-6947-9-S1-S1

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

Zhang, H., Cai, D., and Bai, X. (2020). Macrophages Regulate the Progression of Osteoarthritis. *Osteoarthritis and Cartilage* 28, 555–561. doi:10.1016/j.joca.2020.01.007

Zhao, X., Zhang, L., Wang, J., Zhang, M., Song, Z., Ni, B., et al. (2021). Identification of Key Biomarkers and Immune Infiltration in Systemic Lupus Erythematosus by Integrated Bioinformatics Analysis. *J. Transl Med.* 19, 35. doi:10.1186/s12967-020-02698-x

Zhou, G., Soufan, O., Ewald, J., Hancock, R. E. W., Basu, N., and Xia, J. (2019). NetworkAnalyst 3.0: a Visual Analytics Platform for Comprehensive Gene Expression Profiling and Meta-Analysis. *Nucleic Acids Res.* 47, W234–W241. doi:10.1093/nar/gkz240

Conflict of Interests: 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 Genetics

ORIGINAL RESEARCH
published: 23 May 2022
doi: 10.3389/fgene.2022.902577

Characterization of the Immune Cell Infiltration Landscape Uncovers Prognostic and Immunogenic Characteristics in Lung Adenocarcinoma

Xin Wang[1†], Zhenyi Xu[2†], Zhilin Liu[3], Weihao Lin[1], Zheng Cao[4], Xiaoli Feng[4], Yibo Gao[1,5,6,7] and Jie He[1,6]**

[1]Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [2]Department of Epidemiology and Biostatistics, School of Public Health, Harbin Medical University, Harbin, China, [3]Department of Biostatistics, Peking University, Beijing, China, [4]Department of Pathology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [5]Laboratory of Translational Medicine, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [6]State Key Laboratory of Molecular Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [7]Central Laboratory, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen, China

OPEN ACCESS

Edited by:
Tao Huang,
Chinese Academy of Sciences (CAS), China

Reviewed by:
Min Peng,
Renmin Hospital of Wuhan University, China
Xin Ye,
Shandong University, China

***Correspondence:**
Yibo Gao
gaoyibo@cicams.ac.cn
Jie He
hejie@cicams.ac.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: 23 March 2022
Accepted: 14 April 2022
Published: 23 May 2022

Citation:
Wang X, Xu Z, Liu Z, Lin W, Cao Z, Feng X, Gao Y and He J (2022) Characterization of the Immune Cell Infiltration Landscape Uncovers Prognostic and Immunogenic Characteristics in Lung Adenocarcinoma. Front. Genet. 13:902577. doi: 10.3389/fgene.2022.902577

The immune cell infiltration in TME has been reported to be associated with prognosis and immunotherapy efficiency of lung cancers. However, to date, the immune infiltrative landscape of lung adenocarcinoma (LUAD) has not been elucidated yet. Therefore, this study aimed to identify a new transcriptomic-based TME classification and develop a risk scoring system to predict the clinical outcomes of patients with LUAD. We applied "CIBERSORT" algorithm to analyze the transcriptomic data of LUAD samples and classified LUAD into four discrete subtypes according to the distinct immune cell infiltration patterns. Furthermore, we established a novel predictive tool (TMEscore) to quantify the immune infiltration patterns for each LUAD patient by principal component analysis. The TMEscore displayed as a reliable and independent prognostic biomarker for LUAD, with worse survival in TMEscrore-high patients and better survival in TMEscrore-low patients in both TCGA and other five GEO cohorts. In addition, enriched pathways and genomic alterations were also analyzed and compared in different TMEscore subgroups, and we observed that high TMEscore was significantly correlated with more aggressive molecular changes, while the low TMEscore subgroup enriched in immune active-related pathways. The TMEscore-low subtype showed overexpression of PD-1, CTLA4, and associations of other markers of sensitivity to immunotherapy, including TMB,

Abbreviations: CI, confidence interval; DEGs, differential expression genes; ESTIMATE, Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data; FDR, false discovery rate; GEO, Gene Expression Omnibus; GO, Gene Ontology; GSEA, gene set enrichment analysis; HR, hazard ratio; ICIs, immune checkpoint inhibitors; IPS, immunophenoscore; KEGG, Kyoto Encyclopedia of Genes and Genomes; LUAD, lung adenocarcinoma; PCA, principal component analysis; RNA-seq, RNA sequencing; TCGA, The Cancer Genome Atlas; TIDE, Tumor Immune Dysfunction and Exclusion; TIICs, tumor-infiltrating immune cells; TMB, tumor mutation burden; TME, tumor microenvironment.

immunophenoscore (IPS) analysis, and tumor immune dysfunction and exclusion (TIDE) algorithm. Conclusively, TMEscore is a promising and reliable biomarker to distinguish the prognosis, the molecular and immune characteristics, and the benefit from ICIs treatments in LUAD.

Keywords: immune cell infiltration, prognosis, lung adenocarcinoma, tumor microenvironment, immunotherapy

INTRODUCTION

Although great advances have been achieved in both basic and clinical cancer research (Gu et al., 2020; Jiao and Yang, 2020), cancer still caused approximately 10 million of deaths in 2020 (Sung et al., 2021). With the high prevalence and poor prognosis, lung cancer is ranked as the first leading cause of cancer-related deaths worldwide, becoming a major global health problem (Miller et al., 2019; Siegel et al., 2019; Sung et al., 2021). Recently, the emergence of checkpoint blockade immunotherapy (Pardoll, 2012; Topalian et al., 2015) has significantly improved the strategies of LUAD. However, the minority of response and resistance to these treatments frequently impedes the clinical outcomes. Additionally, the effects of ICIs are not only driven by genetic and epigenetic alterations in tumor cells, but the tumor microenvironment (TME) has also been reported to be a crucial regulator in tumorigenesis (Dejima et al., 2021; Ye et al., 2022), development, metastasis (Quail and Joyce, 2013), and resistance to therapies (Ostman, 2012; Lu et al., 2020).

TME chiefly consists of multiple subpopulations of T and B lymphocytes, dendritic cells (DCs), macrophages, neutrophils, and myeloid-derived suppressor cells (MDSCs) (Belli et al., 2018). The balance between pro-tumorigenic and anti-tumor factors in the TME conducts tumor growth (Wellenstein and de Visser, 2018; Hinshaw and Shevde, 2019). Accumulating evidence has indicated the TME immune composition is generally correlated with prognosis and responsiveness to various cancer treatments. On one hand, tumor-infiltrating lymphocytes (TILs), such as $CD4^+$ and $CD8^+$ T cells, have been associated with longer survival and better response to immunotherapy (Kawai et al., 2008; Fridman et al., 2012). On the other hand, the tumor cells can promote a suppressive TME, which challenges anti-tumor immunity by inducing upregulation of inhibitory immune signaling, suppressive cytokine secretion, and recruitment of suppressive immune cells, such as tumor-associated macrophages (TAMs) presenting pro-tumor effects by secreting immunosuppressive cytokines, including interleukin-10 (IL-10) and transforming growth factor-β (TGF-β) (Mantovani et al., 2017), as well as immunomodulatory cells, such as myeloid-derived suppressor cells (MDSCs) (Ostrand-Rosenberg and Fenselau, 2018)and regulatory T cells (Tregs) (Shimizu et al., 2010), which are all associated with unfavorable prognosis. To be specific, focusing on cellular diversity shows that TME heterogeneity could impact clinical outcomes and provide a challenge for immunotherapy of LUAD (Wu F. et al., 2021; Nguyen et al., 2021). Therefore, investigating the effects of TME composition on the tumor cells will help us decode the regulation of the microenvironment by the tumor.

To date, the emerging predictors for immunotherapy in NSCLC are still imperfect, such as programmed death-ligand 1 (PD-L1) expression (Dempke et al., 2018) is thought to be induced by interferon-γ (IFN-γ)- mediated immune responses and tumor mutational burden (TMB) (Klein et al., 2021) is reported to determine the tumor immunogenicity. It is suggested that only reflecting the tumor cell intrinsic features but ignoring the extrinsic factor, especially TME, is attributed to inconsistencies. Thus, the characteristics of TME should be further comprehensively explored to determine effective biomarkers that precisely predict prognosis and considerably optimize personalized immunotherapy.

Progress has been recently achieved by immunotherapy, emphasizing the importance of TME in LUAD. It elucidates that TME is not the single-cell population but a complex interface among cancer cells, stroma, and infiltrating immune cells. Deeper analyses of the NSCLC TME are necessary to refine the potential application of these findings to clinical care. We applied "CIBERSORT" algorithm to analyze the transcriptomic data of 500 LUAD samples in TCGA and classified the LUAD into four discrete subtypes according to the distinct immune cell infiltration patterns. Furthermore, we established the TME scores to characterize and quantify the immune infiltration patterns for each LUAD patient based on the mRNA expression profiles. Conclusively, we investigated and validated the association between TME score and the clinical outcomes, as well as the efficacy of anti-PD- (L)1 treatment in LUAD, which can facilitate the identification of ideal candidates for personalized immunotherapeutic strategies.

METHODS

Datasets and Preprocessing

A total of 1,518 lung adenocarcinoma (LUAD) and 59 normal tissue samples were retrieved and downloaded from the corresponding datasets, including TCGA LUAD from TCGA data portal (https://xenabrowser.net/datapages/) and GSE31210, GSE37745, GSE50081, GSE68465, and GSE13213 from the NCBI Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/). The somatic mutation data (SNPs and small INDELs) were downloaded from TCGA database (MuTect2 Variant Aggregation and Masking). The raw data of the dataset from Affymetrix were processed using the RMA algorithm in the "Affy" package. The data from Agilent were downloaded with the processed version. For TCGA dataset, RNA-sequencing data (FPKM values) were transformed into transcripts per kilobase million (TPM) values, which are more similar to those resulting from microarrays and more comparable between samples

(Wagner et al., 2012). The following inclusion criteria were used: *1*) histologically confirmed LUAD, *2*) simultaneously available information on mRNA expression profile data and OS, *3*) the sample tissue was collected from the primary solid tumor ("01"), and there was no duplication sample in TCGA, *4*) genes were recorded in all datasets, and *5*) genes with more than 70% of the missing value or 0 value were deleted. The remaining missing values were imputed with KNN imputation approaches. Therefore, 10,320 mRNAs were included in the analysis.

Consensus Clustering for the Tumor Microenvironment

Distinguishing between tumor and normal tissue difference expression genes (DEGs) in TCGA was performed with "limma" (FDR <0.05 and |log2FC| > 1), which better identifies the characteristics of tumor. Furthermore, the tumor microenvironment was quantified by CIBERSORT (https://cibersort.stanford.edu/)(Newman et al., 2015), a deconvolution method for inference of tumor-infiltrating immune components from bulk tissue gene expression profiles. Tumors with qualitatively different immune cell infiltration patterns were grouped using consensus clustering (100 iterations, resample rate of 80%, and hierarchical cluster). This procedure was performed with the "ConsensusClusterPlus" R package.

Identification of DEGs Associated With the TME Phenotype

To functionally elucidate the biological characteristics of the TME subtypes in LUAD, we employed random forest (RF), an efficient and reliable machine learning method to identify DEGs between subtypes of TME. We run RF 100 times with different seeds to find the duplicated variables with at least 80% repetition rate to further ensure the stability of variable selection.

Generation of TMEscore

To further elucidate the comprehensive profile of TME characteristics, the construction of TME metagenes was performed as follows: first, we further screened candidate prognostic genes from DEGs. Next, a consensus clustering algorithm was employed to define the cluster of genes. Then, a principal component analysis (PCA) was performed, and principal component 1 was extracted to serve as the signature score. After obtaining the prognostic value of each gene signature score, we applied a method similar to GGI (Sotiriou et al., 2006) to define the TMEscore of each patient:

$$TMEscore = \sum PC1_i - \sum PC1_j,$$

where i is the signature score of clusters whose Cox coefficient is positive and j is the expression of genes whose Cox coefficient is negative.

Functional and Pathway Enrichment Analysis

To further analyze the biological significance of the genes related to TMEscore with KEGG and GO function analysis, the "clusterProfiler" R package was adopted to annotate gene patterns (Wu T. et al., 2021). The Benjamini–Hochberg procedure was used to control the false discovery rate (FDR). We set the cut-off of adj. p-values to 0.2 so that we could find more relevant pathways and functions based on the small number of DEGs. Gene set enrichment analysis (GSEA) illustrated the significantly different enriched pathways in the high- and low-TMEscore groups. Gene sets were downloaded from the MSigDB database of the Broad Institute (Subramanian et al., 2005) and employed the Hallmark gene sets and 1,000 permutations. An enrichment pathway between two subtypes was determined with an FDR of <0.25 and the normalized enrichment score (NES).

Predicting the Patients' Response to ICIs

The Cancer Immunome Atlas (https://tcia.at/) analyzed the immune landscapes and antigenomes of 20 solid tumors that were quantified by Immunophenoscore (IPS, a superior immune response molecular marker) (Charoentong et al., 2017). The IPS value, which ranged from 0 to 10, was positively correlated to tumor immunogenicity and could predict the patients' response to immune checkpoint inhibitors (ICI treatment). Tumor Immune Dysfunction and Exclusion (TIDE, http://tide.dfci.harvard.edu/), a computational method to predict immune checkpoint blockade response, was developed by Jiang et al. (2018). TIDE uses a set of expression markers to profile two primary mechanisms of tumor immune evasion: T-cell dysfunction and T-cell exclusion. Patients with higher TIDE prediction scores represent a greater potential of tumor immune escape; therefore, TIDE could evaluate patients who are more likely to benefit from ICI. In addition, the mRNA expression of immune checkpoints was analyzed in different prognosis groups.

Statistical Analysis

Continuous variables were summarized as mean ± SD, and categorized variables were described by frequency (n) and proportion (%). Differences among variables were tested by the Wilcoxon rank-sum test and Fisher's exact tests. The relationship between variables was tested by Spearman rank correlation analysis. The cut-off value of TMEscore was calculated based on the correlation between the patients' survival and the TMEscore in TCGA with the "survminer" package. Univariate and multivariate Cox regression analyses were used to assess prognostic analysis. Batch effects from non-biological technical biases were corrected using the "ComBat" algorithm of the "sva" package. The "Maftools" package was used to present the mutation landscape and identify the differential gene mutations between groups. The heatmap was produced by the R package "ComplexHeatmap." A two-sided $p < 0.05$ was regarded as statistically significant. All data processing was performed in R 4.0.2 software.

RESULTS

Landscape of Lung Adenocarcinoma TME

This study was conducted as per the flow chart shown in **Supplementary Figure S1**. The information of 1,518 LUAD patients is detailed in **Supplementary Table S1**. To classify the LUAD TME, the consensus clustering algorithm was used to cluster TME information obtained by CIBERSORT in TCGA-LUAD

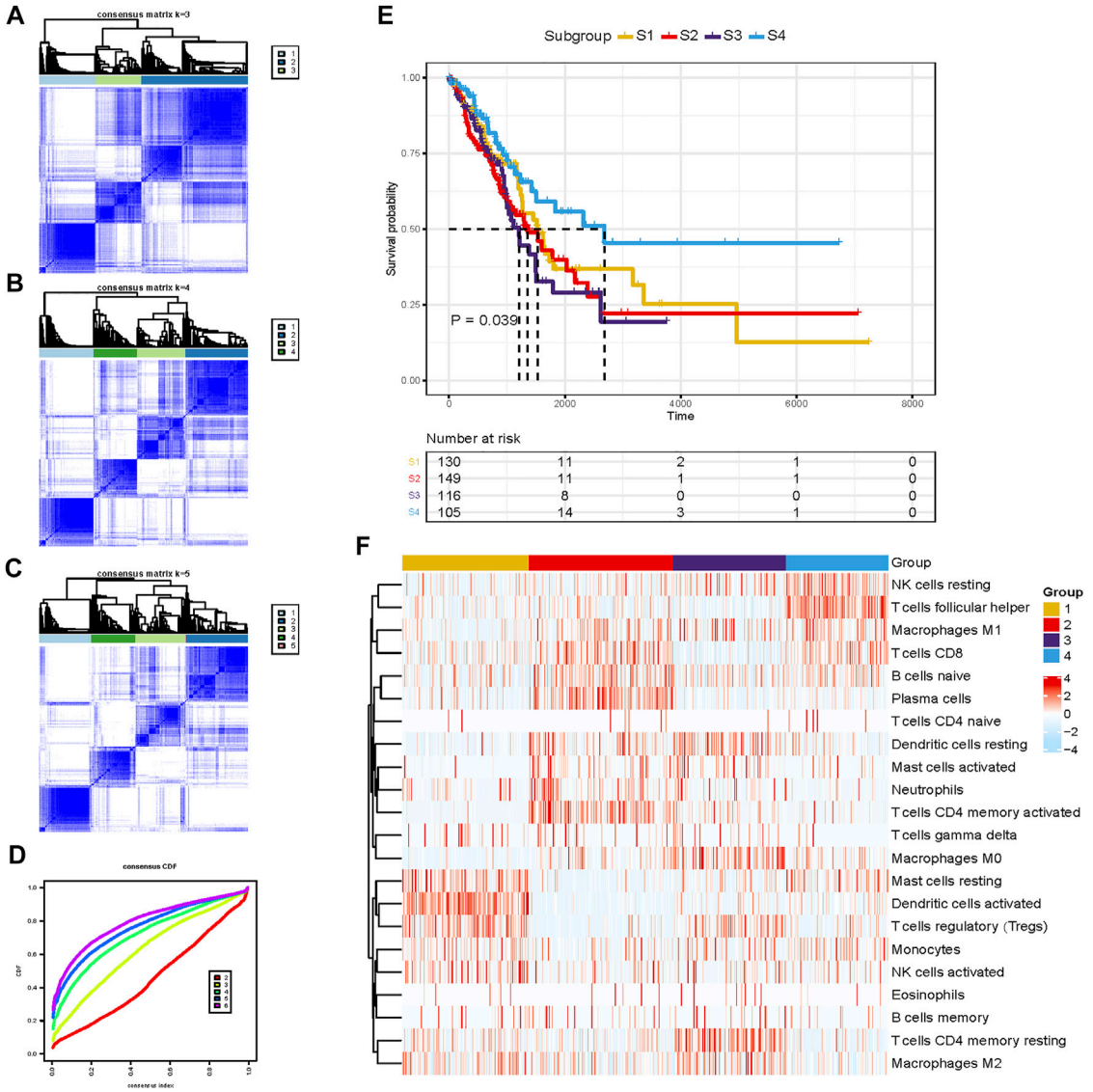


FIGURE 1 | Unsupervised clustering of the tumor microenvironment (TME) cells for 500 patients in the TCGA-LUAD cohort. **(A–C)** Consensus matrices of different clusters. **(D)** Consensus cumulative distribution function (CDF) curve. **(E)** Kaplan–Meier (K-M) curves for overall survival (OS) of four different subtypes (log-rank test, p =0.039). **(F)** Abundance pattern of 22 TME cell types in four TME subtypes.

dataset (**Supplementary Table S2**). The most appropriate clustering number was four (**Figures 1A–D**), which was selected by consensus matrices and consensus cumulative distribution function (CDF) curve. This analysis revealed that LUAD can be clustered into four distinct TME subtypes termed S1–4. The patients with subtype S4 had significantly longer overall survival (OS) than patients with subtypes S1 and S2, and subtype S3 demonstrated the worst survival (**Figure 1E**). These four TME subtypes varied significantly based on the expression levels of LM22 gene signatures (**Figure 1F**). The S4 subtype was characterized by increases in the infiltration of CD 8+ T cells, resting NK cells, follicular helper T cells, and M1 macrophages, displaying S4 was significantly associated with immune activation. Meanwhile, resting mast cells, activated dendritic cells, and regulatory T cells (Tregs) were enriched in the S1 subtype, and the S2 subtype showed significant increases in the infiltration of naïve B cells, plasma cells, and $CD4^+$ memory-activated T cells; on the contrary, M0 macrophages, M2 macrophages, and $CD4^+$ memory-resting T cells showed high infiltration in the S3 subtype, indicating an immunosuppressive milieu. Taken together, we demonstrated that the four TME subtypes were characterized by distinct immune cell infiltration and prognosis.

Identification of DEGs and Functional Annotation

To further identify the biological characteristics and differences among TME subtypes, RF algorithm was employed to extract the

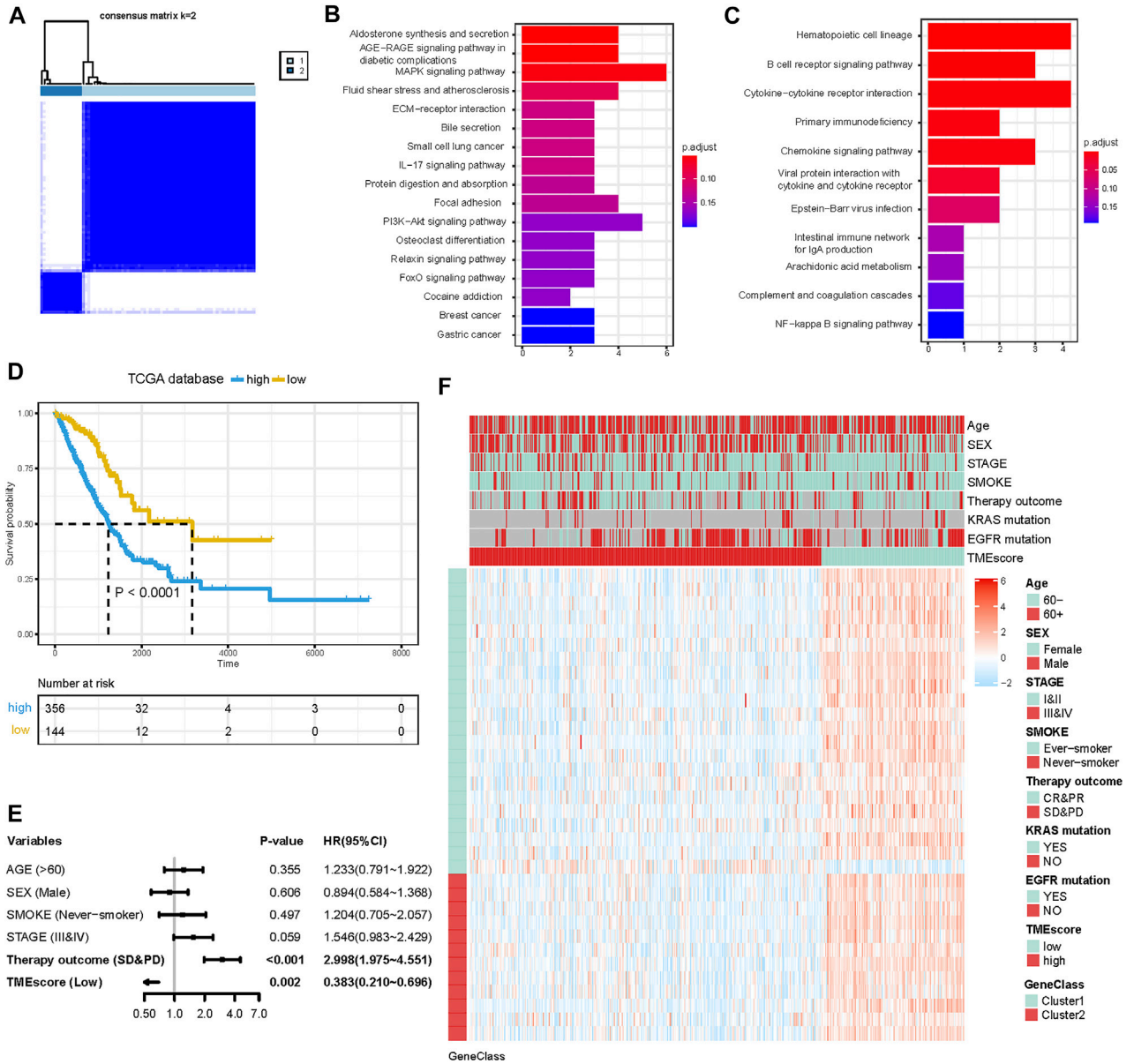


FIGURE 2 | Construction of the TMEscore for LUAD patients. **(A)** Consensus matrices of differentially expressed genes (DEGs) among TME subtypes. **(B** and **C)** KEGG pathway enrichment analysis results in G1 and G2. **(D)** K-M curve for OS of different TMEscore groups (log-rank test, $p < 0.001$). **(E)** Forest plots illustrating the results of multivariate Cox proportional hazards model of clinical feature in TCGA cohort. **(F)** Heatmap of DEG expression and clinical characteristics. TMEscore, age, sex, stage, smoke, therapy outcome, mutation of KARS, and mutation of EGFR are shown as patient annotations. Gene clusters are shown as gene annotations. Top legend, gray indicates a missing value.

phenotype signatures. By 100 times analysis, a total of 77 DEGs duplicated at least 80 times were identified (**Supplementary Table S3**). Through consensus clustering analysis based on the expression of the 77 most representative DEGs, we divided DEGs into two different clusters termed G1 (62 DEGs) and G2 (15 DEGs) (**Figure 2A**). These two gene clusters were closely related to distinct TME and played different biological roles. Then, GO and KEGG enrichment analyses were performed with the "clusterProfiler" R package. The G1 cluster was mainly enriched in the MAPK signaling pathway, PI3K-Akt signaling pathway, aldosterone syntheses, and focal adhesion pathways (**Figure 2B**). The G2 cluster was mainly enriched in hematopoietic cell lineage, B-cell receptor signaling pathway, cytokine–cytokine receptor interaction, and primary immunodeficiency (**Figure 2C**). Significantly enriched pathways and molecular functions are summarized in **Supplementary Tables S4 and S5**. Collectively, the coherence between the prognostic and biological features in the two gene

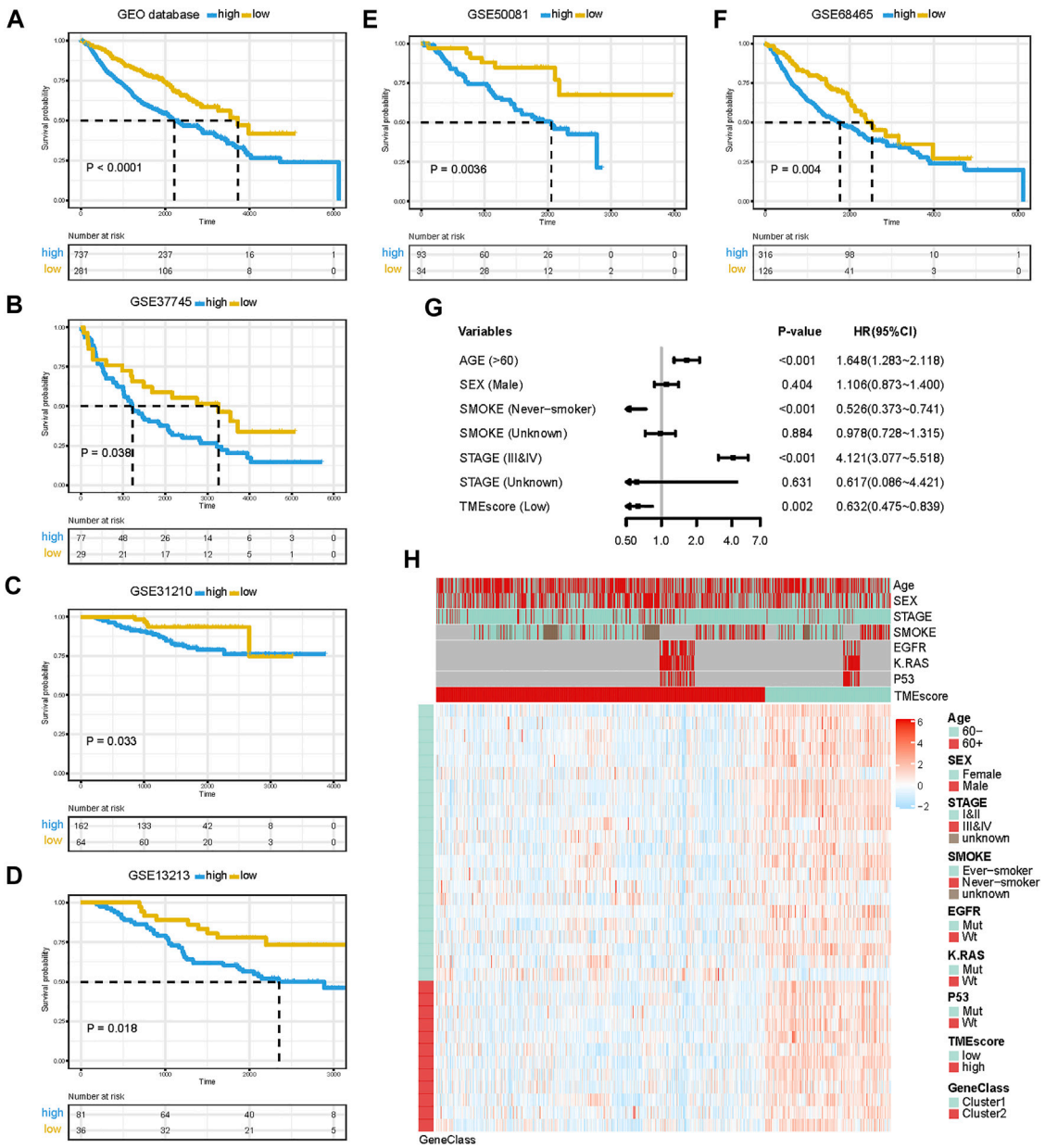


FIGURE 3 | Prognostic value of TMEscore for LUAD patients in five GEO cohorts. **(A)** K-M curve of all 1,018 patients in the GEO database between different TMEscore groups (log-rank test, $p < 0.001$). **(B–F)** K-M curves of five independent GEO datasets in different TMEscore subgroups. **(G)** Forest plots illustrating the results of the multivariate Cox proportional hazards model of clinical features in the GEO database. **(H)** Heatmap of DEG expression and clinical characteristics. TMEscore, age, sex, stage, smoke, mutation of KARS, mutation of EGFR, and mutation of P53 are shown as patient annotations. Gene clusters are shown as gene annotations. Top legend, gray indicates a missing value.

subgroups indicated that this classification was reliable and reasonable.

Construction and Validation of the TMEscore in Six Independent Cohorts

Although the four TME subtypes were identified, their clinical significance needed to be further evaluated and quantified. Therefore, we build TMEscore based on TME information in the TCGA-LUAD cohort to assess the prognostic value. Association with a prognosis of 34 genes (G1:22, G2:12) was confirmed by Cox regression analysis. First, principal component analysis (PCA) was used to compute two aggregate scores, TMBscore A from G1 and TMBscore B from G2. Then, we performed univariate Cox regression on each TMEscore to evaluate the prognostic value. Finally, TMEscore A and TMEscore B were integrated to obtain TMEscore for each sample. The

prognostic value of the TMEscore was further assessed by the log-rank test after classification as high-risk and low-risk groups based on the corresponding optimal cut-off value (−0.92) acquired by the "survminer" R package in the TCGA-LUAD cohort. We visualized gene expression and clinical features distribution in different risk groups with a heatmap in TCGA and GEO datasets, respectively (**Figures 2F** and **3H**). The Kaplan–Meier curve of TMEscore subgroups showed that the patients in the low TMEscore group (median survival time 3,169 days) had significantly better overall survival than the high-TMEscore group (median survival time 1,235 days; log-rank test, $p < 0.0001$; **Figure 2D**). Moreover, the prognostic value of the TMEscore was further assessed with five external datasets in the GEO database: GSE37745, GSE31210, GSE13213, GSE50081 and GSE68465. Similar results were found that the survival advantage in the low TMEscore group in above cohorts, with the corresponding p-value of 0.038, 0.033, 0.018, 0.0036, and 0.004 (**Figures 3B–F**). Meanwhile, we integrated a total of 1,018 samples in the GEO datasets to evaluate prognostic efficiency, indicating the low-TMEscore group patients had better overall survival compared to the high-TMEscore group (log-rank test, $p < 0.0001$; **Figure 3A**). These findings suggested that TMEscore possessed a reliable and robust capacity for predicting the prognosis for LUAD patients.

TMEscore Was an Independent Prognostic Factor for LUAD Patients

In addition to the TMEscore, other prognostic factors such as individual and clinicopathological features were included. After multivariable adjustments with age, sex, smoke situation, TNM stage, and therapy outcomes in the TCGA cohort, the TMEscore was confirmed as an independent prognostic indicator with a hazard ratio of 0.383 [95% CI: 0.210–0.696] in the TCGA-LUAD cohort (**Figure 2E**), 0.632 [95% CI: 0.475–0.839] in the GEO datasets (**Figure 3G**). Elder, ever-smoker, advanced stage, and non-response to therapy were also suggested to be independent risk factors in different datasets, respectively.

Recent studies have reported that specific gene alterations, such as TP53 (Sun et al., 2020), KRAS (Hamarsheh et al., 2020), EGFR (Chen et al., 2015), and STK11 (Mazzaschi et al., 2021) have an important role in the regulation of the tumor immune microenvironment (TIME) and served as biomarkers to tumor therapeutics (Lee et al., 2017; Krishnamurthy et al., 2021). We further explore the predictive value of this TMEscore in LUAD patients with EGFR/KRAS mutation (MUT) or wild type (WT). Remarkably, this risk model had predictive power for both EGFR wild type and EGFR mutation LUAD patients, except for patients with TP53/EGFR co-mutations (**Supplementary Figure S2**). Similarly, this risk model exhibited a robustly predictive value in both KRAS wild type and KRAS mutation LUAD patients, except for patients with KRAS/STK11 co-mutations LUAD patients (**Supplementary Figure S3**). Among the EGFR/KRAS wild-type/mutation population, the beneficial trends of low TMEscore in the prognosis of LUAD patients were observed in distinct subgroups, suggesting that TMEscore was an independent and reliable prognostic indicator.

Different Biological Processes Between the High-TMEscore Group and the Low-TMEscore Group

For a comprehensive analysis of the potential regulatory mechanisms resulting in different TMEscore groups, we performed GSEA analysis between high and low TMEscore subgroups. The results showed that 23 pathways were enriched in different subgroups with FDR<0.25 (**Supplementary Table S6**). In high TMEscore group, MYC targets V1 (NES = 2.28 and FDR = 0.001), MTORC1 signaling (NES = 2.08 and FDR = 0.011), MYC targets V2 (NES = 2.08 and FDR = 0.012), G2M checkpoint (NES = 2.00 and FDR = 0.021), glycolysis (NES = 1.79 and FDR = 0.014), and other pathways were enriched (**Figures 4A–D**). Meanwhile, the results revealed that complement (NES = −1.74 and FDR = 0.231), inflammatory response (NES = −1.69 and FDR = 0.212), IL6/JAK/STAT3 signaling (NES = −1.64 and FDR = 0.181), IL2/STAT5 signaling up (NES = −1.58 and FDR = 0.212) and interferon gamma response (NES = −1.42 and FDR = 0.223), and other pathways were correlated with the low TMEscore (**Figures 4E–H**). It is suggested that the gene sets of the TMEscore high samples were enriched in cancer and tumor metabolism-related pathways, while the gene sets of the TMEscore low samples were enriched in DNA repair and immune response-related pathways.

The Molecular Characteristics of Distinct TMEscore Subgroups

Genomic alterations and oncogenic signaling within the tumors have been reported to affect anti/pro-tumor immunity and TME activity (Hamarsheh et al., 2020; Kumagai et al., 2020; Zhou et al., 2020; Fountzilas et al., 2021), links between tumor mutations and TME subtypes needed to be investigated. To illustrate the somatic variants and acquire further biological insights into the immunological characteristics of LUAD between TMEscore subgroups, we utilized the Mutation Annotation Format (MAF) files and performed the variants annotation. We found higher mutation counts in the TME-high subgroup than in the TME-low subgroup. Missense variations were the most common mutation subtype, followed by nonsense and frameshift deletions. The oncoplot of tumor somatic mutation in the TCGA-LUAD cohort showed that TP53, TTN, and MUC16 gene mutations in the high-TMEscore group were approximately 20% higher than those in the low TMEscore group (**Figures 5A and B**). Among a total of 54 differential mutated genes between two groups ($p < 0.01$; **Figure 5C**), CMA1, HSPA12B, and FAM196A showed a higher mutation frequency in the low-TMEscore group. The other genes, such as TP53, TTN, and FBXL7, had a higher mutation frequency in the high-TMEscore group.

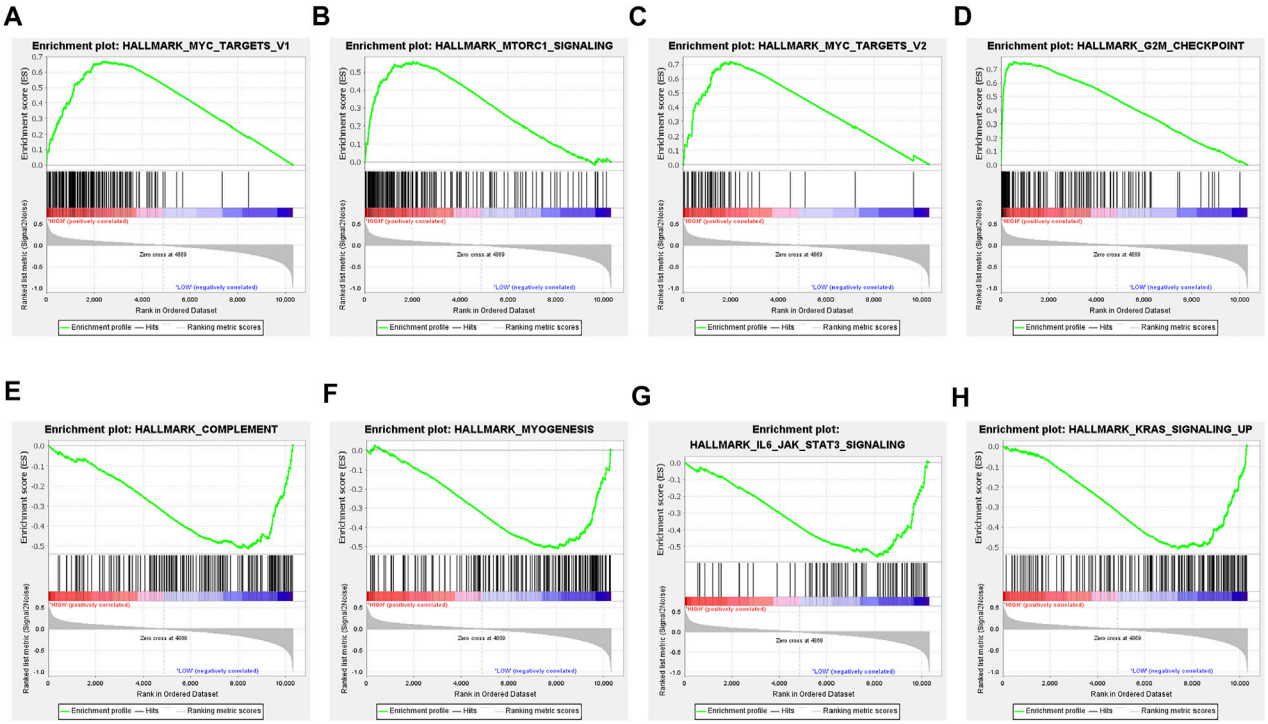


FIGURE 4 | Gene set enrichment analysis (GSEA) in TMEscore groups. **(A–D)** Enrichment plots showing MYC targets V1, MTORC1 signaling, MYC targets V2, and G2M checkpoint in the high-TMEscore group. **(E–H)** Enrichment plots showing complement, myogenesis, IL6/JAK/STAT3 signaling, and KRAS signaling up in the low-TMEscore group.

Collectively, this analysis indicated that transcriptomic-based TME classification coupled with genomics analysis can be exploited for further studies.

Combinations of TMEscore, Immune Checkpoints, and TMB Improve Risk Stratification and Survival Prediction

Previous studies have emphasized the importance of immune checkpoint genes in modulating immune infiltration (Keir et al., 2008; Andrews et al., 2019). Thus, we first compared the expression pattern of immune checkpoint genes between different patient groups delaminated by the TMEscore in TCGA-LUAD and GEO datasets. PDCD1, CD86, CD80, and CTLA4 showed significantly high expression in the low-TMEscore group than in the high-TMEscore group (**Figures 6A–D, F–I**), which was further confirmed in five independent validating cohorts (**Supplementary Figure S4**).

Considering the correlations between immune checkpoint genes and TMEscore, we next combined TMEscore with immune checkpoints expression to test whether they have an influence on OS in LUAD patients. Though survival analyses among four subgroups stratified by TMEscore and immune checkpoint gene expression, we displayed that patients with low PD-L1 and low TMEscore have prolonged OS compared to those with low PD-L1 and high TMEscore (p = 0.005), and among patients with high PD-L1 expression, a lower TMEscore signified a remarkably better survival ($p < 0.001$) (**Figure 7A**). We also found similar survival patterns among four patient subgroups stratified by TMEscore and PD1/CTLA-4 expressions in the TCGA cohort (**Figure 7A**). We then confirmed the results in the other five validation cohorts (**Figures 7B,C** and **Supplementary Figure S5**). In concordance with the TCGA dataset, patients with low TMEscore have significantly better survival relative to the high TMEscore group, even though with similar expression levels of immune checkpoint genes (**Figures 7B,C** and **Supplementary Figure S5**). In addition, TMB has been shown to have the potential to generate a larger number of neoantigens and make them more immunogenic (Schumacher and Schreiber, 2015), which is strongly associated with clinical outcomes and response of immune checkpoint blockade response (Yarchoan et al., 2017; Chan et al., 2019). We found that patients with low TMB and high TMEscore had the worst prognosis (**Figure 8A**). It is suggested that TMEscore, immune checkpoint genes, and TMB can complement each other as prognostic biomarkers.

The TMEscore Predicts Clinical Outcomes of Immunotherapy

Given the linkage between TMEscore and immune checkpoint genes as well as TMB, we further explore the predictive potential of TMEscore for immune checkpoint blockade response through analyzing the correlation of TMEscore and published

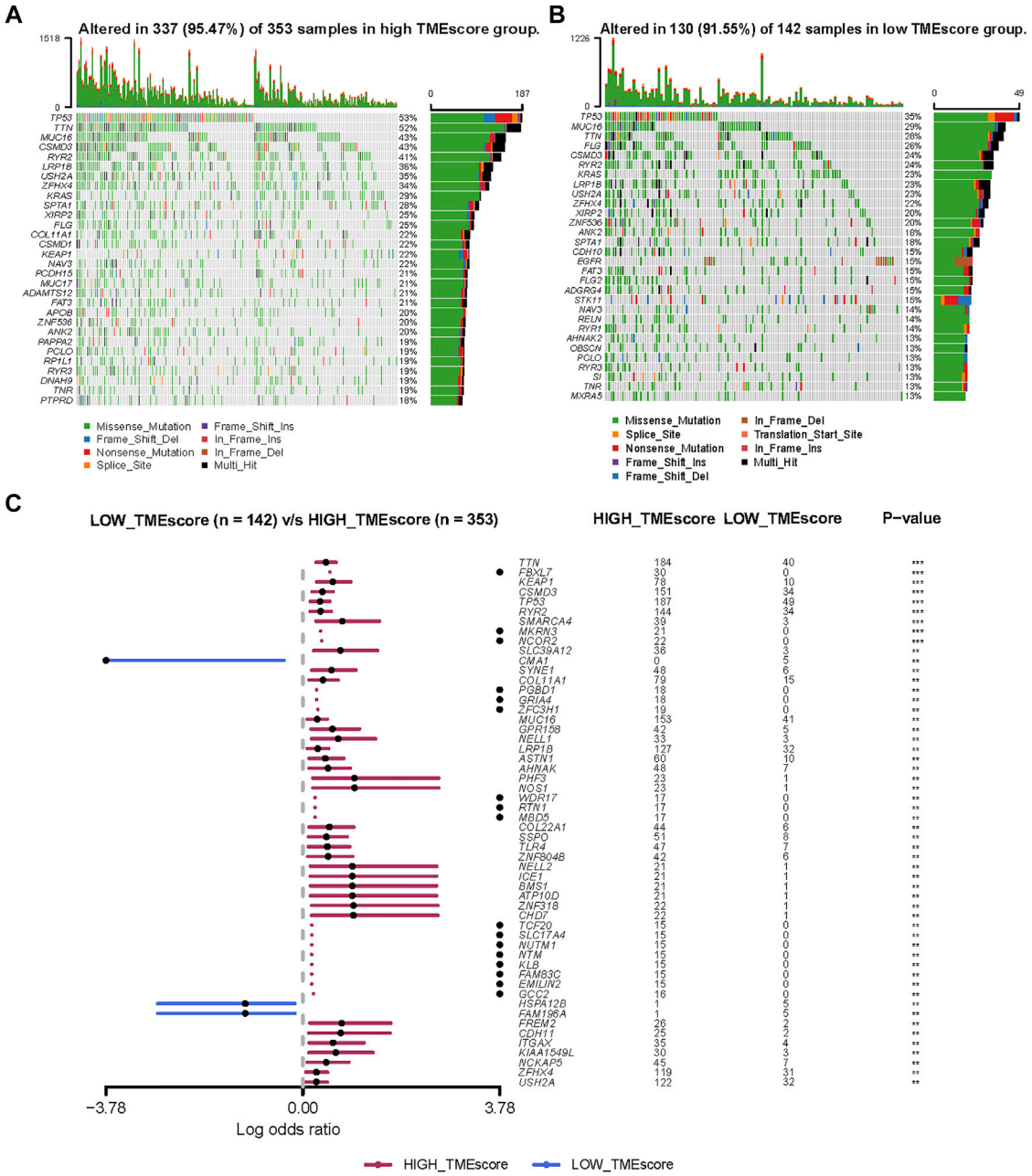


FIGURE 5 | Molecular variations between low-TMEscore and high-TMEscore groups. **(A** and **B)** Mutation profiles of high-TMEscore and low-TMEscore groups. **(C)** Comparing differentially mutated genes between two subgroups by Fisher's exact tests.

immunotherapy predictors, including TIDE and IPS. The relationship between TIDE and TMEscore was investigated in TCGA and GEO datasets. As expected, the high-TMEscore group was characterized by a significantly higher TIDE score (**Figures 6E and J**). The IPS values (IPS-PD-1/PD-L1/PD-L2_pos and IPS-CTLA-4_pos) increased in the low-TMEscore group compared to the high-TMEscore group in TCGA (**Figures 6K and L**). It is likely that the patients in the low-TMEscore group may have a better immune microenvironment and respond better to ICIs than those in the high-TMEscore group.

Furthermore, the practicability of the TMEscore was further evaluated for speculation of the therapeutic benefit for ICI treated patients. The patients who received anti-PD-L1 immunotherapy in the IMvigor210 cohort were assigned based on high and low TME scores. Given the contraindicatory prognostic and predictive value of TMEscore, TMB, and immune checkpoint gene expression (PD-L1, PD-1, and CTLA4), we next evaluated the synergistic effect of these biomarkers in the prognostic and predictive stratification of LUAD. Consistent with previous results, stratified survival analysis revealed that the TMB status did not interfere with TMEscore-based

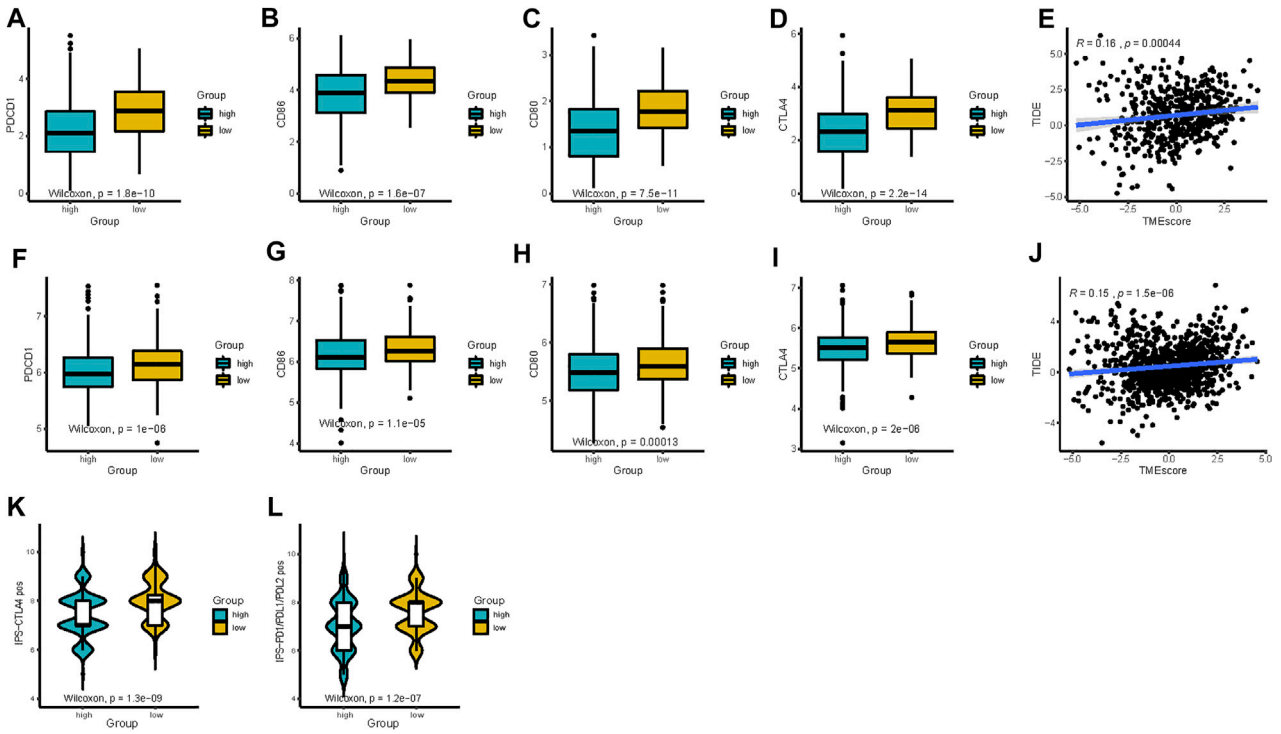


FIGURE 6 | TMEscore in the prediction of immunotherapeutic benefits. **(A–D)** Expression of immune-checkpoint-relevant genes (PDCD1, CD86, CD80, and CTLA4) between high- and low-TMEscore groups in TCGA and **(F–I)** all 1,018 patients in GEO datasets after batch correction. **(E** and **J)** Relationships between TIDE and TMEscore in TCGA and GEO datasets. **(K** and **L)** Relative probabilities of response to anti-CTLA-4 and anti-PD-1/PD-L1 treatment (IPS score) in the low-TMEscore and high-TMEscore groups in TCGA cohort.

predictions. Subtypes of the combination of TMEscore and TMB showed significant survival differences (log-rank test, p = 0.0076; **Figure 8B**). On the other hand, Kaplan–Meier analysis revealed patients in the IMvigor210 cohort with TMEscore low and PD-L1 high obtained most favorable OS than either single positive (TME low or PD-L1 high) or dual negative (TMEscore high PD-L1 low, p = 0.0018, **Figure 8D**). In addition, analysis of objective response also supported that TMEscore low and PD-L1 high subgroup represented an increased proportion of PR/CR/SD than either single positive (TMEscore low or PD-L1 high) or dual negative (p = 0.047, **Figure 8E**). Taken together, these findings indicate the TME classification system and scoring system may explain the effectiveness of immunotherapy in patients with low TMB and low PD-L1, and these distinct classification systems, TMEscore, PD-L1, and TMB might function as complementary factors for the prediction of immunotherapy.

DISCUSSION

Although immune checkpoint inhibitors (ICIs) have revolutionized treatment strategies of lung cancer, the overall response rate of ICI monotherapies is still limited and no more than 20% in NSCLC patients with EGFR/ALK wild-type (Doroshow et al., 2019). It has been reported that TME plays a crucial role in cancer development and anti-tumor process, especially the immunotherapy response in cancers (Lu et al., 2020; Ye et al., 2022). Therefore, characterizing the tumor–immune microenvironment can improve the personalized immunotherapeutic strategies.

Multi-omics data are often used for generating various predictive or prognostic models through machine learning or statistical modeling methods (Xu et al., 2021). However, to date, comprehensive analyses based on integrated genomic and transcriptomic profiles of the tumor and its TME remain rare and lack efficient and useful models. Therefore, we constructed a scoring system to classify and quantify the comprehensive tumor immune landscape based on an immune-cell phenotype algorithm and validation in external LUAD cohorts.

Transcriptomic analysis offers an opportunity to dissect the complexity of tumors, including TME, dynamically regulating cancer progression and influencing therapeutic outcomes (Cieslik and Chinnaiyan, 2018; Thorsson et al., 2018). In our study, we identified four distinct immune subtypes characterized by different biological processes and prognosis, using "CIBERSORT" algorithm to analyze the transcriptomic data of TCGA-LUAD samples. Furthermore, we established the TME scores to characterize and quantify the immune infiltration patterns for each LUAD patient based on the DEGs among the distinct subtypes. The TMEscore displayed as a reliable prognostic immune-related biomarker for LUAD, with worse survival in TMEscore-high patients and better survival in

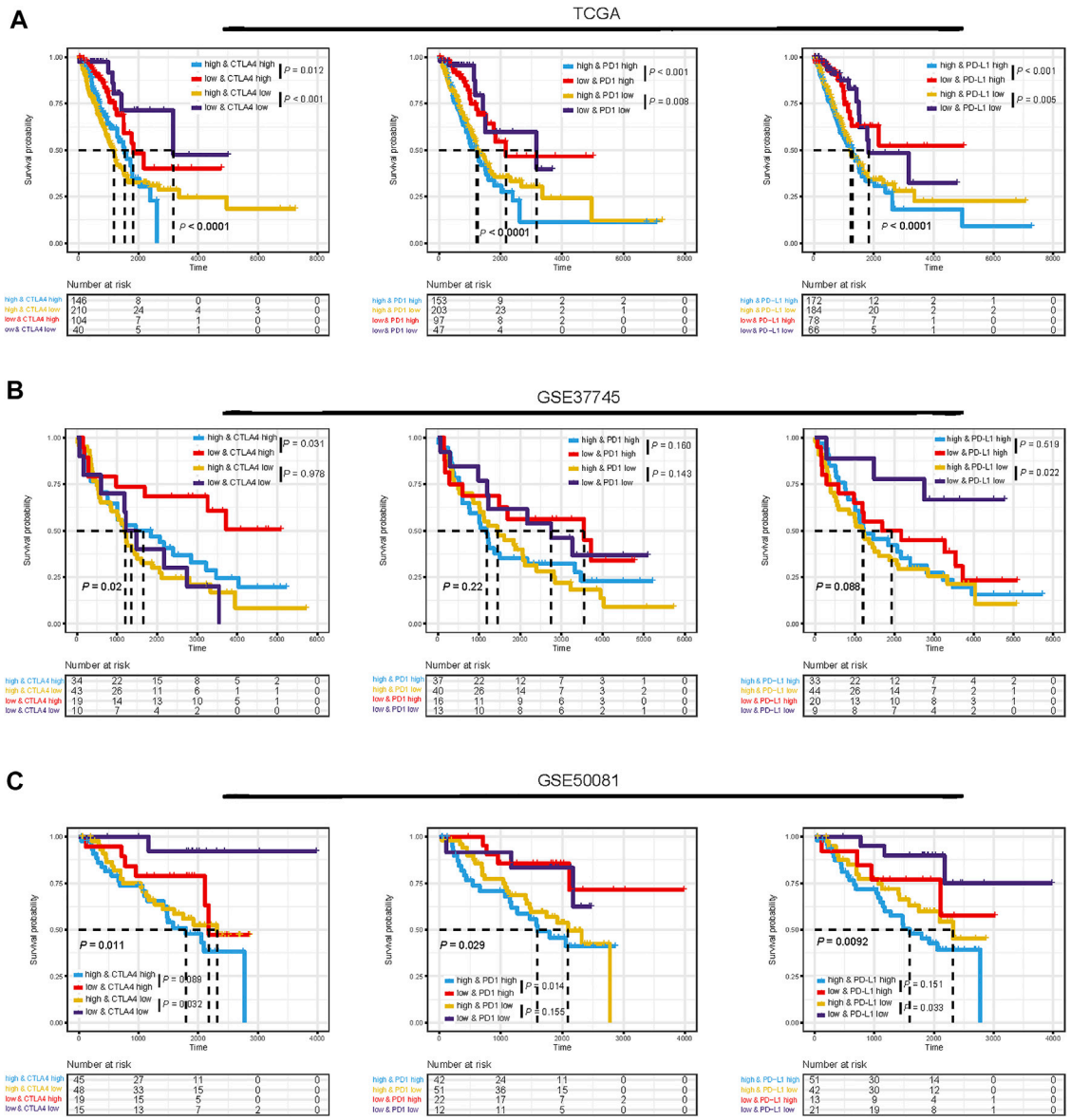


FIGURE 7 | Impact of immune checkpoint gene expressions and TMEscore on clinical outcome. Kaplan–Meier survival curves of overall survival among four patient groups stratified by TMEscore and immune checkpoint genes (PD1, PD-L1, and CTLA-4) in TCGA dataset **(A)**, GSE37745 dataset **(B)**, and GSE50081 dataset **(C)**.

TMEscore-low patients and in both TCGA and other five independent GEO cohorts. In addition, enriched pathways and genomic alterations were also analyzed and compared in different TMEscore subgroups, and we observed that a high TMEscore was significantly correlated with more aggressive molecular changes such as TP53 mutations. As expected from its increased immune gene expression, the TMEscore-low subtype showed overexpression of PD-L1, PD-1, CTLA4, and associations of other markers of sensitivity to immunotherapy, including IPS score and TIDE score. Our findings also revealed that the TMEscore is a robust and reliable prognostic tool and predictive indicator of the response to immunotherapy in the IMvigor210 cohort. With further in-depth investigation, our TMEscore might be utilized as an important supplementary predictor to LUAD immunotherapy.

The tumor microenvironment (TME) is a complex interface between cancer cells, stroma, and infiltrating immune cells (Fridman et al., 2012). A previous study demonstrated that the tumor microenvironment contexture plays a key role in tumor development and immunotherapeutic efficacy (Stankovic et al., 2018). TME heterogeneity, which impacts tumor progression and prognosis, has been identified in cancers, especially LUAD (Lavin et al., 2017; Zhang et al., 2019; Chen et al., 2020; Nguyen et al., 2021). In addition, the difference in TME patterns was found to be correlated to tumor heterogeneity and treatment diversity (Jia et al., 2018; Vitale et al., 2021). Considering the individual

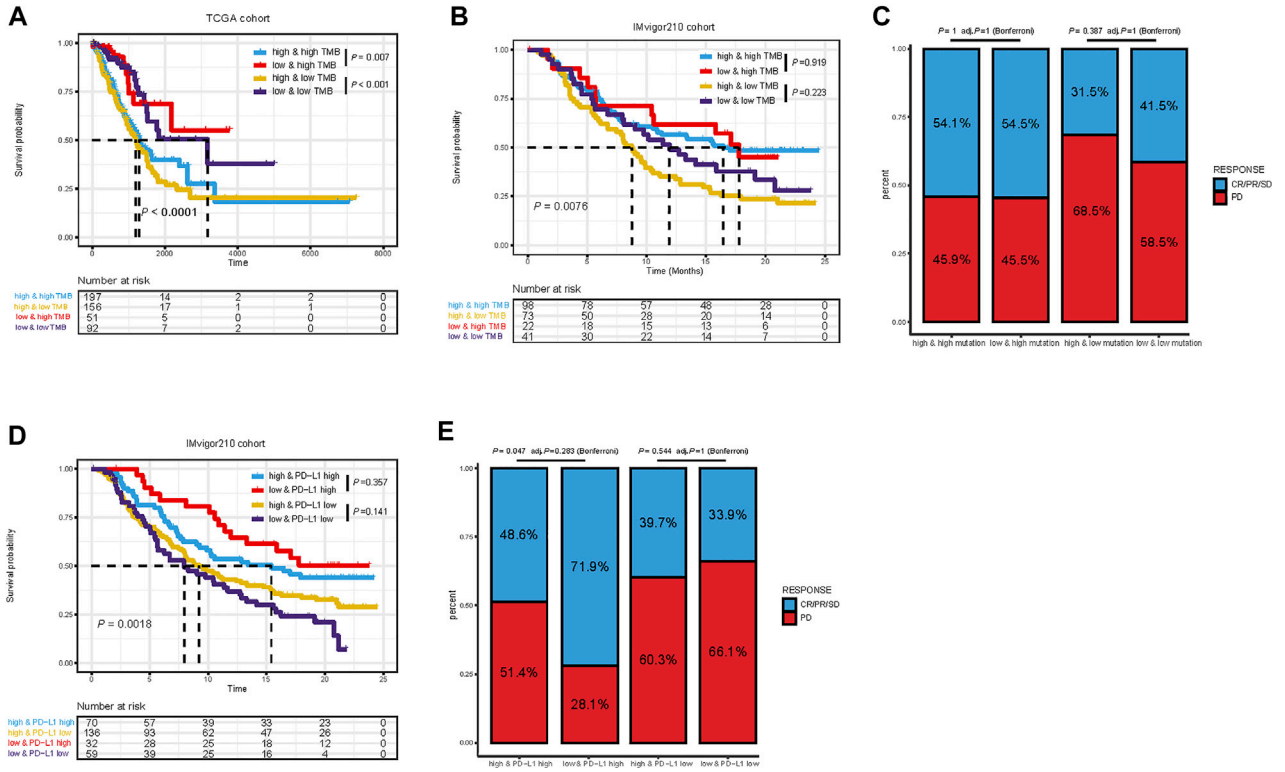


FIGURE 8 | Kaplan–Meier survival curves of overall survival among four patient groups stratified by TMEscore and TMB in TCGA dataset **(A)** and IMvigor210 cohort **(B)**. Proportional representation of the objective response rate among subgroups categorized by TMEscore and TMB in the IMvigor210 cohort **(C)**. Kaplan–Meier survival curves of overall survival among four patient groups stratified by TMEscore and PD-L1 in the IMvigor210 cohort. **(D)** Proportional representation of the objective response rate among subgroups categorized by TMEscore and PD-L1 in the IMvigor210 cohort **(E)**. CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease.

heterogeneity of the immune milieu, it was demanded to quantify the TME patterns of individual tumors. Here, using the "CIBERSORT" algorithm, we identified 22 human immune-cell phenotypes and generated an individualized TMEscore to assess TME patterns. Our study represents an essential step toward understanding the crosstalk between malignant cells and immune cells in LUAD.

Our findings of the TCGA molecular mutations displayed significant differences in distributions across the TMEscore subgroups. The largest difference in mutations between subgroups was in TP53 mutations, which were more common in TME-high samples than in TME-low samples (53 vs. 35%). TP53 mutation is not only the most common genetic event in NSCLC but also reported to be associated with poor prognosis in cancers, especially non-small cell lung cancer (Ozaki and Nakagawara, 2011). TP53 mutation could affect disease progression, tumor cell characteristics, and the therapeutic effect of different therapeutics (Wu and Hwang, 2019). In addition, the more enrichment of KEAP1 mutations in TMEscore high tumors than TMEscore low may be one potential explanation for the distinct performance of ICI efficacy in LUAD. KEAP1 mutations were reported to be enriched in patients with high TMB lacking T-cell infiltration and immunologically cold (Marinelli et al., 2020), which have been associated with decreased efficacy of ICIs in NSCLC in published studies (Papillon-Cavanagh et al., 2020; Di Federico et al., 2021). The differences in their molecular characteristics between TMEscore subgroups might contribute to the diverse immunogenic features and consequently varied responses to immunotherapy.

To acquire a deeper insight into the biological feature of the TMEscore subgroups, we further investigated enriched pathways and immune characteristics of different TMEscore subgroups. Patients with the low-TMEscore subtype, whose molecular traits, including an abundance of infiltration immune active cells, enhanced enrichment of immune-related pathways, such as interferon gamma response, complement, inflammatory response, were previously reported to predict the efficacy of pembrolizumab. In addition, we also observed elevated IL6/JAK/STAT3 signaling pathway in the low-TMEscore group, modulating the IFN-γ-induced expression of PD-L1 (Zhang et al., 2021). Collectively, the designated distinct TMEscore subtypes of LUAD were identified, and the crucial insights into the immunologic features of these subtypes were provided. Meanwhile, we proved that the TMEscore showed significant correlations with immune checkpoint genes (PD-L1, PD-1, and CTLA-4), TMB, and other biomarkers of immunotherapy, including IPS and TIDE, indicating that TMEscore possessed the potential to predict the response to immunotherapy. Previous studies have demonstrated that

some biomarkers, such as TIDE and IPS, could predict patient response to immunotherapy. TIDE, a creative computational method to identify the induction of T-cell dysfunction in tumors with high infiltration of cytotoxic T lymphocytes and the prevention of T-cell infiltration in tumors with low-CTL levels (Jiang et al., 2018), has been proven to predict the outcomes of cancers treated with ICIs (Jiang et al., 2018). In addition, IPS was developed to quantitatively predict patients' response to anti-PD-1/PD-L1 and anti-CTLA-4 therapies based on an 18-gene signature including genes that reflect an ongoing adaptive Th1 and cytotoxic CD8 T-cell response (Charoentong et al., 2017). Thus, the low-TMEscore patients presenting high IPS and low TIDE scores may have a better response to immunotherapy. However, both TIDE and IPS focused on the function and status of T cells, which could not fully reflect the complexity of the TME involved in the response to immunotherapy. Therefore, our scoring system exhibits promising clinical flexibility for the predictive value of anti-PD-(L)1 therapy.

Furthermore, patients in the high-risk subgroup presented with a higher level of immune checkpoint molecules and showed higher immunogenicity. However, PD-L1 expression and TMB are neither the only nor the satisfying tool to identify NSCLC patients that might benefit from therapy with immune checkpoints inhibitors (Klein et al., 2021). One critical obstacle impeding the extensive utility of TMB and PD-L1 expression is the determination of feasible cut-off values. Moreover, these two predictors only focused on the intrinsic features of tumors and may not cover other situations involved in antitumor immune responses such as TME. Notably, we demonstrated that it is reasonable to combine TMEscore with PD-L1 or TMB together, and thus it might help make clinical decisions in LUAD. Patients with TMEscore-low PD-L1 high or TMB high should be preferentially recommended for ICI treatment, while patients with TMEscore-low PD-L1 low, or TMEscore-high PD-L1 high can optionally consider anti-PD- (L)1 therapy; however, patients with TMEscore high/low PD-L1 or low TMB should carefully choose anti-PD-(L)1 therapy. Taking this step further, it is suggested that TMEscore can identify either potential sensitive patients with low PD-L1 expression/low TMB who may benefit or patients who do not respond to ICIs despite having a high PD-L1 expression/high TMB. In addition, we also explored the stability of our TMEscore model. We found that patients with a lower TMEscore were more likely to respond to ICB and had improved overall survival in the IMvigor210 cohort treated with checkpoint blockade. Collectively, combinations of TMEscore, TMB, and PD-L1 could be applied not only as refined prognostic stratification tools but also as more reliable predictive biomarkers for personalized immunotherapy treatment.

Our study provides a translational rationale for evaluating TME based on transcriptomic data and TMEscore as a biomarker for immunotherapy response in patients with LUAD. However, this study still has several limitations. First, while the composition of the TME has been recognized as a determinant of cancer progression and response to therapy, most analyses have focused on a limited proportion of cell types. Nonetheless, there are still numerous cellular and molecular mechanisms involved in immunotherapy, and our TMEscore may not cover the possible intra/extracellular situations involved in antitumor immune responses. Second, since this study was a retrospective analysis, the ability of the TMEscore in predicting survival and response to immunotherapy should be validated in a large-cohort, multi-center, and prospective study in the future. Third, all quantifications of gene expression are relative values, which makes it difficult to determine the absolute threshold and cut-off values for clinical application. Therefore, quantitative determinations of gene expression are also needed. Specifically, the underlying molecular mechanisms remain to be elucidated in LUAD *in vivo and in vitro*.

CONCLUSION

In conclusion, our translational rationale for TME classification may help in distinguishing immune and molecular characteristics and predicting clinical outcomes of LUAD patients. These findings will further improve the implementation and utility of precisely personalized immunotherapeutic strategies in LUAD.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

YG and JH designed the study. XW, ZX, ZL, WL, ZC, and XF collected the data. XW and ZX performed the data analysis and interpreted the data. XW and ZX drafted the manuscript. XW, ZX, ZL, WL, ZC, XF, YG, and JH revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This study was supported by the National Key R&D Program of China (2020AAA0109500), R&D Program of Beijing Municipal Education Commission (KJZD20191002302). National Natural Science Foundation of China (82122053, 82188102), the Beijing Municipal Science & Technology Commission (Z191100006619115), CAMS Initiative for Innovative Medicine (2021-I2M-1-012 2021-I2M-1-015), Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (2021-PT310-001), Key-Area Research and Development Program of Guangdong Province (2021B0101420005), and Aiyou Foundation (KY201701).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.902577/full#supplementary-material

REFERENCES

Andrews, L. P., Yano, H., and Vignali, D. A. A. (2019). Inhibitory Receptors and Ligands beyond PD-1, PD-L1 and CTLA-4: Breakthroughs or Backups. *Nat. Immunol.* 20, 1425–1434. doi:10.1038/s41590-019-0512-0

Belli, C., Trapani, D., Viale, G., D'Amico, P., Duso, B. A., Della Vigna, P., et al. (2018). Targeting the Microenvironment in Solid Tumors. *Cancer Treat. Rev.* 65, 22–32. doi:10.1016/j.ctrv.2018.02.004

Chan, T. A., Yarchoan, M., Jaffee, E., Swanton, C., Quezada, S. A., Stenzinger, A., et al. (2019). Development of Tumor Mutation burden as an Immunotherapy Biomarker: Utility for the Oncology Clinic. *Ann. Oncol.* 30, 44–56. doi:10.1093/annonc/mdy495

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

Chen, N., Fang, W., Zhan, J., Hong, S., Tang, Y., Kang, S., et al. (2015). Upregulation of PD-L1 by EGFR Activation Mediates the Immune Escape in EGFR-Driven NSCLC: Implication for Optional Immune Targeted Therapy for NSCLC Patients with EGFR Mutation. *J. Thorac. Oncol.* 10, 910–923. doi:10.1097/jto.0000000000000500

Chen, Y. P., Yin, J. H., Li, W. F., Li, H. J., Chen, D. P., Zhang, C. J., et al. (2020). Single-cell Transcriptomics Reveals Regulators Underlying Immune Cell Diversity and Immune Subtypes Associated with Prognosis in Nasopharyngeal Carcinoma. *Cell Res.* 30, 1024–1042. doi:10.1038/s41422-020-0374-x

Cieslik, M., and Chinnaiyan, A. M. (2018). Cancer Transcriptome Profiling at the Juncture of Clinical Translation. *Nat. Rev. Genet.* 19, 93–109. doi:10.1038/nrg.2017.96

Dejima, H., Hu, X., Chen, R., Zhang, J., Fujimoto, J., Parra, E. R., et al. (2021). Immune Evolution from Preneoplasia to Invasive Lung Adenocarcinomas and Underlying Molecular Features. *Nat. Commun.* 12, 2722. doi:10.1038/s41467-021-22890-x

Dempke, W. C. M., Fenchel, K., and Dale, S. P. (2018). Programmed Cell Death Ligand-1 (PD-L1) as a Biomarker for Non-small Cell Lung Cancer (NSCLC) Treatment-Are We Barking up the Wrong Tree? *Transl. Lung Cancer Res.* 7, S275–S279. doi:10.21037/tlcr.2018.04.18

Di Federico, A., De Giglio, A., Parisi, C., and Gelsomino, F. (2021). STK11/LKB1 and KEAP1 Mutations in Non-small Cell Lung Cancer: Prognostic rather than Predictive? *Eur. J. Cancer* 157, 108–113. doi:10.1016/j.ejca.2021.08.011

Doroshow, D. B., Sanmamed, M. F., Hastings, K., Politi, K., Rimm, D. L., Chen, L., et al. (2019). Immunotherapy in Non-small Cell Lung Cancer: Facts and Hopes. *Clin. Cancer Res.* 25, 4592–4602. doi:10.1158/1078-0432.ccr-18-1538

Fountzilas, E., Kurzrock, R., Hiep Vo, H., and Tsimberidou, A. M. (2021). Wedding of Molecular Alterations and Immune Checkpoint Blockade: Genomics as a Matchmaker. *J. Natl. Cancer Inst.* 113, 1634–1647. doi:10.1093/jnci/djab067

Fridman, W. H., Pages, F., Sautes-Fridman, C., and Galon, J. (2012). The Immune Contexture in Human Tumours: Impact on Clinical Outcome. *Nat. Rev. Cancer* 12, 298–306. doi:10.1038/nrc3245

Gu, C., Shi, X., Dai, C., Shen, F., Rocco, G., Chen, J., et al. (2020). RNA M(6)A Modification in Cancers: Molecular Mechanisms and Potential Clinical Applications. *Innovation (N Y)* 1, 100066. doi:10.1016/j.xinn.2020.100066

Hamarsheh, S., Gross, O., Brummer, T., and Zeiser, R. (2020). Immune Modulatory Effects of Oncogenic KRAS in Cancer. *Nat. Commun.* 11, 5439. doi:10.1038/s41467-020-19288-6

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

Jia, Q., Wu, W., Wang, Y., Alexander, P. B., Sun, C., Gong, Z., et al. (2018). Local Mutational Diversity Drives Intratumoral Immune Heterogeneity in Non-small Cell Lung Cancer. *Nat. Commun.* 9, 5361. doi:10.1038/s41467-018-07767-w

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

Jiao, D., and Yang, S. (2020). Overcoming Resistance to Drugs Targeting $KRAS^{G12C}$ Mutation. *Innovation (N Y)* 1, 100035. doi:10.1016/j.xinn.2020.100035

Kawai, O., Ishii, G., Kubota, K., Murata, Y., Naito, Y., Mizuno, T., et al. (2008). Predominant Infiltration of Macrophages and CD8(+) T Cells in Cancer Nests Is a Significant Predictor of Survival in Stage IV Nonsmall Cell Lung Cancer. *Cancer* 113, 1387–1395. doi:10.1002/cncr.23712

Keir, M. E., Butte, M. J., Freeman, G. J., and Sharpe, A. H. (2008). PD-1 and its Ligands in Tolerance and Immunity. *Annu. Rev. Immunol.* 26, 677–704. doi:10.1146/annurev.immunol.26.021607.090331

Klein, O., Kee, D., Markman, B., Carlino, M. S., Underhill, C., Palmer, J., et al. (2021). Evaluation of TMB as a Predictive Biomarker in Patients with Solid Cancers Treated with Anti-PD-1/CTLA-4 Combination Immunotherapy. *Cancer Cell* 39, 592–593. doi:10.1016/j.ccell.2021.04.005

Krishnamurthy, N., Goodman, A. M., Barkauskas, D. A., and Kurzrock, R. (2021). STK11 Alterations in the Pan-Cancer Setting: Prognostic and Therapeutic Implications. *Eur. J. Cancer* 148, 215–229. doi:10.1016/j.ejca.2021.01.050

Kumagai, S., Togashi, Y., Sakai, C., Kawazoe, A., Kawazu, M., Ueno, T., et al. (2020). An Oncogenic Alteration Creates a Microenvironment that Promotes Tumor Progression by Conferring a Metabolic Advantage to Regulatory T Cells. *Immunity* 53, 187–203. doi:10.1016/j.immuni.2020.06.016

Lavin, Y., Kobayashi, S., Leader, A., Amir, E. D., Elefant, N., Bigenwald, C., et al. (2017). Innate Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell Analyses. *Cell* 169, 750–765. doi:10.1016/j.cell.2017.04.014

Lee, C. K., Man, J., Lord, S., Links, M., Gebski, V., Mok, T., et al. (2017). Checkpoint Inhibitors in Metastatic EGFR-Mutated Non-small Cell Lung Cancer-A Meta-Analysis. *J. Thorac. Oncol.* 12, 403–407. doi:10.1016/j.jtho.2016.10.007

Lu, Z., Peng, Z., Liu, C., Wang, Z., Wang, Y., Jiao, X., et al. (2020). Current Status and Future Perspective of Immunotherapy in Gastrointestinal Cancers. *Innovation (N Y)* 1, 100041. doi:10.1016/j.xinn.2020.100041

Mantovani, A., Marchesi, F., Malesci, A., Laghi, L., and Allavena, P. (2017). Tumour-associated Macrophages as Treatment Targets in Oncology. *Nat. Rev. Clin. Oncol.* 14, 399–416. doi:10.1038/nrclinonc.2016.217

Marinelli, D., Mazzotta, M., Scalera, S., Terrenato, I., Sperati, F., D'Ambrosio, L., et al. (2020). KEAP1-driven Co-mutations in Lung Adenocarcinoma Unresponsive to Immunotherapy Despite High Tumor Mutational burden. *Ann. Oncol.* 31, 1746–1754. doi:10.1016/j.annonc.2020.08.2105

Mazzaschi, G., Leonetti, A., Minari, R., Gnetti, L., Quaini, F., Tiseo, M., et al. (2021). Modulating Tumor Microenvironment: A Review on STK11 Immune Properties and Predictive vs Prognostic Role for Non-small-cell Lung Cancer Immunotherapy. *Curr. Treat. Options. Oncol.* 22, 96. doi:10.1007/s11864-021-00891-8

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

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

Nguyen, P. H. D., Ma, S., Phua, C. Z. J., Kaya, N. A., Lai, H. L. H., Lim, C. J., et al. (2021). Intratumoural Immune Heterogeneity as a Hallmark of Tumour Evolution and Progression in Hepatocellular Carcinoma. *Nat. Commun.* 12, 227. doi:10.1038/s41467-020-20171-7

Ostman, A. (2012). The Tumor Microenvironment Controls Drug Sensitivity. *Nat. Med.* 18, 1332–1334. doi:10.1038/nm.2938

Ostrand-Rosenberg, S., and Fenselau, C. (2018). Myeloid-Derived Suppressor Cells: Immune-Suppressive Cells that Impair Antitumor Immunity and Are Sculpted by Their Environment. *J. Immunol.* 200, 422–431. doi:10.4049/jimmunol.1701019

Ozaki, T., and Nakagawara, A. (2011). Role of P53 in Cell Death and Human Cancers. *Cancers (Basel)* 3, 994–1013. doi:10.3390/cancers3010994

Papillon-Cavanagh, S., Doshi, P., Dobrin, R., Szustakowski, J., and Walsh, A. M. (2020). STK11 and KEAP1 Mutations as Prognostic Biomarkers in an Observational Real-World Lung Adenocarcinoma Cohort. *ESMO Open* 5, e000706. doi:10.1136/esmoopen-2020-000706

Pardoll, D. M. (2012). The Blockade of Immune Checkpoints in Cancer Immunotherapy. *Nat. Rev. Cancer* 12, 252–264. doi:10.1038/nrc3239

Quail, D. F., and Joyce, J. A. (2013). Microenvironmental Regulation of Tumor Progression and Metastasis. *Nat. Med.* 19, 1423–1437. doi:10.1038/nm.3394

Schumacher, T. N., and Schreiber, R. D. (2015). Neoantigens in Cancer Immunotherapy. *Science* 348, 69–74. doi:10.1126/science.aaa4971

Shimizu, K., Nakata, M., Hirami, Y., Yukawa, T., Maeda, A., and Tanemoto, K. (2010). Tumor-infiltrating Foxp3+ Regulatory T Cells are Correlated with Cyclooxygenase-2 Expression and Are Associated with Recurrence in

Resected Non-small Cell Lung Cancer. *J. Thorac. Oncol.* 5, 585–590. doi:10.1097/jto.0b013e3181d60fd7

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

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. *J. Natl. Cancer Inst.* 98, 262–272. doi:10.1093/jnci/djj052

Stankovic, B., Bjorhovde, H. A. K., Skarshaug, R., Aamodt, H., Frafjord, A., Muller, E., et al. (2018). Immune Cell Composition in Human Non-small Cell Lung Cancer. *Front. Immunol.* 9, 3101. doi:10.3389/fimmu.2018.03101

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

Sun, H., Liu, S. Y., Zhou, J. Y., Xu, J. T., Zhang, H. K., Yan, H. H., et al. (2020). Specific TP53 Subtype as Biomarker for Immune Checkpoint Inhibitors in Lung Adenocarcinoma. *EBioMedicine* 60, 102990. doi:10.1016/j.ebiom.2020.102990

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

Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Ou Yang, T. H., et al. (2018). The Immune Landscape of Cancer. *Immunity* 48, 812–830. doi:10.1016/j.immuni.2018.03.023

Topalian, S. L., Drake, C. G., and Pardoll, D. M. (2015). Immune Checkpoint Blockade: a Common Denominator Approach to Cancer Therapy. *Cancer Cell* 27, 450–461. doi:10.1016/j.ccell.2015.03.001

Vitale, I., Shema, E., Loi, S., and Galluzzi, L. (2021). Intratumoral Heterogeneity in Cancer Progression and Response to Immunotherapy. *Nat. Med.* 27, 212–224. doi:10.1038/s41591-021-01233-9

Wagner, G. P., Kin, K., and Lynch, V. J. (2012). Measurement of mRNA Abundance Using RNA-Seq Data: RPKM Measure Is Inconsistent Among Samples. *Theor. biosciences* 131, 281–285. doi:10.1007/s12064-012-0162-3

Wellenstein, M. D., and de Visser, K. E. (2018). Cancer-Cell-Intrinsic Mechanisms Shaping the Tumor Immune Landscape. *Immunity* 48, 399–416. doi:10.1016/j.immuni.2018.03.004

Wu, C. H., and Hwang, M. J. (2019). Risk Stratification for Lung Adenocarcinoma on EGFR and TP53 Mutation Status, Chemotherapy, and PD-L1 Immunotherapy. *Cancer Med.* 8, 5850–5861. doi:10.1002/cam4.2492

Wu, F., Fan, J., He, Y., Xiong, A., Yu, J., Li, Y., et al. (2021a). Single-cell Profiling of Tumor Heterogeneity and the Microenvironment in Advanced Non-small Cell Lung Cancer. *Nat. Commun.* 12, 2540. doi:10.1038/s41467-021-22801-0

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

Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., et al. (2021). Artificial Intelligence: A Powerful Paradigm for Scientific Research. *Innovation (N Y)* 2, 100179. doi:10.1016/j.xinn.2021.100179

Yarchoan, M., Hopkins, A., and Jaffee, E. M. (2017). Tumor Mutational Burden and Response Rate to PD-1 Inhibition. *N. Engl. J. Med.* 377, 2500–2501. doi:10.1056/nejmc1713444

Ye, Y., Zhang, Y., Yang, N., Gao, Q., Ding, X., Kuang, X., et al. (2022). Profiling of Immune Features to Predict Immunotherapy Efficacy. *Innovation (N Y)* 3, 100194. doi:10.1016/j.xinn.2021.100194

Zhang, Q., He, Y., Luo, N., Patel, S. J., Han, Y., Gao, R., et al. (2019). Landscape and Dynamics of Single Immune Cells in Hepatocellular Carcinoma. *Cell* 179, 829–845 e20. doi:10.1016/j.cell.2019.10.003

Zhang, L., Kuca, K., You, L., Zhao, Y., Musilek, K., Nepovimova, E., et al. (2021). Signal Transducer and Activator of Transcription 3 Signaling in Tumor Immune Evasion. *Pharmacol. Ther.* 230, 107969. doi:10.1016/j.pharmthera.2021.107969

Zhou, Y., Bian, S., Zhou, X., Cui, Y., Wang, W., Wen, L., et al. (2020). Single-Cell Multiomics Sequencing Reveals Prevalent Genomic Alterations in Tumor Stromal Cells of Human Colorectal Cancer. *Cancer Cell* 38, 818–828. doi:10.1016/j.ccell.2020.09.015

Author Disclaimer: The funders had no roles in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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 Genetics

ORIGINAL RESEARCH
published: 26 May 2022
doi: 10.3389/fgene.2022.897886

Identification of Immune-Related Genes for Risk Stratification in Multiple Myeloma Based on Whole Bone Marrow Gene Expression Profiling

Qiang-Sheng Wang[1], *Qi-Qin Shi*[2], *Ye Meng*[3], *Meng-Ping Chen*[4] *and Jian Hou*[4]*

[1]Department of Hematology, Ningbo Hangzhou Bay Hospital, Ningbo, China, [2]Department of Ophthalmology, Ningbo Hangzhou Bay Hospital, Ningbo, China, [3]Bone Marrow Transplantation Center, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China, [4]Department of Hematology, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China

OPEN ACCESS

Edited by:
Tao Huang,
Shanghai Institute of Nutrition and Health (CAS), China

Reviewed by:
Ram Vinay Pandey,
Karolinska University Hospital, Sweden
Chenqi Wang,
University of South Florida, United States
Ting Wang,
Complete Genomics, United States
Gregory Taylor,
Duke University, United States

***Correspondence:**
Jian Hou
houjian@medmail.com.cn

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: 16 March 2022
Accepted: 10 May 2022
Published: 26 May 2022

Citation:
Wang Q-S, Shi Q-Q, Meng Y, Chen M-P and Hou J (2022) Identification of Immune-Related Genes for Risk Stratification in Multiple Myeloma Based on Whole Bone Marrow Gene Expression Profiling. Front. Genet. 13:897886. doi: 10.3389/fgene.2022.897886

Background: Multiple myeloma (MM) is characterized by abnormal proliferation of bone marrow clonal plasma cells. Tumor immunotherapy, a new therapy that has emerged in recent years, offers hope to patients, and studying the expression characteristics of immune-related genes (IRGs) based on whole bone marrow gene expression profiling (GEP) in MM patients can help guide personalized immunotherapy.

Methods: In this study, we explored the potential prognostic value of IRGs in MM by combining GEP and clinical data from the GEO database. We identified hub IRGs and transcription factors (TFs) associated with disease progression by Weighted Gene Co-expression Network Analysis (WGCNA), and modeled immune-related prognostic signature by univariate and multivariate Cox and least absolute shrinkage and selection operator (LASSO) regression analysis. Subsequently, the prognostic ability of signature was verified by multiple statistical methods. Moreover, ssGSEA and GSEA algorithm reveled different immunological characteristics and biological function variation in different risk groups. We mapped the hub IRGs by protein-protein interaction network (PPI) and extracted the top 10 ranked genes. Finally, we conducted vitro assays on two alternative IRGs.

Results: Our study identified a total of 14 TFs and 88 IRGs associated with International Staging System (ISS). Ten IRGs were identified by Cox -LASSO regression analysis, and used to develop optimal prognostic signature for overall survival (OS) in MM patients. The 10-IRGs were BDNF, CETP, CD70, LMBR, LTBP1, NENF, NR1D1, NR1H2, PTK2B and SEMA4. In different groups, risk signatures showed excellent survival prediction ability, and MM patients also could be stratified at survival risk. In addition, IRF7 and SHC1 were hub IRGs in PPI network, and the vitro assays proved that they could promote tumor progression. Notably, ssGSEA and GSEA results confirmed that different risk groups could accurately indicate the status of tumor microenvironment (TME) and activation of biological pathways.

Conclusion: Our study suggested that immune-related signature could be used as prognostic markers in MM patients.

Keywords: multiple myeloma, immune-related genes, whole bone marrow sequencing, prognostic model, IRF7, SHC1

INTRODUCTION

Multiple myeloma (MM) is a B-cell malignancy characterized by abnormal proliferation of clonal plasma cells in the bone marrow. In recent years, its incidence has been on the rise and it has become the second most common hematologic malignancy (Mireles-Cano et al., 2020; Moser-Katz et al., 2021). MM Patients face multiple inevitable relapses after remission with multidrug combination therapy. The higher the number of relapses, the shorter the remission period and eventually the refractory relapse period, which seriously affects the prognosis (Gerecke et al., 2016). The occurrence of such condition is one of the greatest challenges in the treatment of MM, as it leads to incurable MM. Therefore, it is clinically important to explore the pathogenesis of MM in depth and to discover new therapeutic targets to provide more effective means for the treatment of MM.

Immunotherapy is a new therapeutic option and its efficacy in the treatment of MM needs to be further investigated and improved. The TME is closely related to the immunotherapeutic response (Hou et al., 2019). Studies have shown that dendritic cells (DCs) isolated from MM patients not only have impaired function but also express/produce low levels of key molecules that initiate the immune response, including IL-12, human leukocyte antigen DR (HLA-DR), CD40, CD86, and CD80 (Kawano et al., 2015). The immune checkpoint cytotoxic T lymphocyte-associated protein-4 (CTLA-4) on chromosome two interacts with CD80/CD86 on DCs and negatively regulates the CD28 signaling pathway. Although the killing of MM cells by $CD4^+$ T cells is mediated by resident myeloid macrophages (Haabeth et al., 2020), myeloid macrophages in MM is mainly derived from TNF-α and immunosuppressive cytokines IL-10 and IL-1β in the tumor microenvironment, which not only produce angiogenic factors that contribute to tumor growth and invasion, such as vascular endothelial growth factor (VEGF), IL-8, fibroblast growth factor-2, metalloproteinase and cyclooxygenase-2, and colony-stimulating factor-1, but also increase drug resistance in myeloma through direct cell-to-cell interactions (Kawano et al., 2017). PD-L1 is expressed in most MM plasma cells. Increased IFN-γ and toll-like receptor (TLR) ligands induce PD-L1 expression in isolated MM plasma cells (Tamura et al., 2020). Myeloid differentiation factor 88 (MyD88) and TNF receptor-associated factor 6 (TRAAF6) bridging proteins inhibit TLR pathway and suppress not only TLR ligand-induced PD-L1 expression but also IFN-γ-mediated PD-L1 expression (Liu et al., 2007). The above findings suggest that the immune microenvironment plays a key role in MM progression. In this study, we will reveal the abnormal expression of immune-related genes (IRGs) in tumor progression to provide effective diagnosis and treatment for the disease. Nowadays, there have been several studies on the prognosis prediction of MM, such as gene expression inflammatory signature (Botta et al., 2016), EMC-92-gene signature (Kuiper et al., 2012), and genome-wide association studies (GWAS) of MM (Went et al., 2019), etc. Although these study all predict survival status in MM patients, we found that most of research either used $CD138^+$ selected cells microarray or mixed samples from various time points. More importantly, prognostic signature based solely on IRGs have not yet been developed in MM patients. Therefore, an in-depth study of the treatment and prognosis of IRGs and individualized immunotherapy is essential to improve the prognosis of MM patients.

In this study, we investigated the potential prognostic value of IRGs in MM by integrating clinical data and pre-treatment gene expression profiling (GEP). Firstly, we identified 102 IRGs and transcription factors (TFs) driving MM progression, and performed gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses. Subsequently, immune-related prognostic signature was developed in training cohort and validate in training and testing cohorts. In addition, the protein-protein interaction (PPI) network were extracted 10 Top IRGs. The results of the bioinformatic analysis were supported by the identification of IRF7 and SHC1 genes as hub IRGs, and the vitro assays demonstrated that IRF7 and SHC1 have a function in promoting tumor progression. These results suggested that prognostic signature and hub IRGs may be promising and molecular markers, which in turn provide targets for the diagnosis and prognosis of MM.

MATERIALS AND METHODS

Data Collection and Pre-processing

Whole bone marrow GEP and corresponding clinical features were obtained from The Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/). Importantly, the whole bone marrow samples in the GSE136400 dataset contains five time points, such as before treatment, post Induction, post transplant, post consolidation, and post maintenance. The aim of the study was pre-treatment gene signature prediction, hence we retained only before treatment 354 patients for bioinformatics analysis. Samples were omitted genes with mean expression values less than 0.1 to ensure the significance of the analysis. Detailed clinical information for each sample is provided in **Supplementary File S1**. We annotated 1,594 TFs and IRGs based on the cis-chromosome and IMMPORT database

(http://cistrome.org/CistromeCancer/CancerTarget/; https://www.immport.org/home).

Weighted Gene Co-Expression Network Analysis (WGCNA)

The 'WGCNA' package (Langfelder and Horvath, 2008) screened the genes that were significantly associated with clinical features. According to our previous study (Shen et al., 2021), a soft threshold was determined, an adjacency matrix was clustered, and a hub module was determined. The strongest positive correlation was selected for further analysis by calculating the Pearson correlation coefficient between the modules and International Staging System (ISS). In this study, we classified the transcriptome data into genes modules based on the topological overlap matrix (TOM) and optimal soft threshold ($\beta = 7$).

Functional Enrichment Analysis

We used the 'cluster Profiler' package for gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of TFs and IRGs involved in disease progression. All MM patients were divided into high- and low-expression groups based on median expression for subsequent GSEA analysis. In addition, we used the c2. cp.kegg.v7.4. symbols and c5. go.v7.4. symbols gene sets from the Molecular Signature Database (MSigDB) for GSEA analysis. The number of permutations was set to 1,000. The criteria for screening statistically significant pathways were set as p-value less than 0.05 and FDR less than 0.05.

Construction and Validation of IRGs-Related Signature

We randomly divided the 354 pre-treatment patients in the GSE136400 dataset by a ratio of 6:4 (caret package in R software). Of these, 214 patients were used as the training set and the remaining 120 patients included in the testing set. In addition, we also added a validation cohort (n = 134), including post maintenance patients GEP. Univariate and multivariate Cox regression analyses were used to investigate the relationship between the expression of IRGs and clinical prognosis in training set. Specifically, we selected genes that were significantly associated with clinical prognosis ($p < 0.05$). Subsequently, the LASSO-Cox regression method was used to select the IRGs involved in the prediction model from the above IRGs. In the training and testing cohorts, the risk score of each individual was analyzed by regression coefficients and their expression in multivariate Cox analysis. MM patients in different sets were classified according to median risk score in training set and survival analysis was used to compare the clinical prognosis of high-risk and low-risk patients. The diversity of clinical information between the different risk groups and the prognostic significance were assessed. The accuracy of the prognostic model was verified using ROC curves with $p < 0.05$ as the significance criterion.

Comprehensive Analysis of Signature

Cox regression analysis was used to assess the independent prognostic value. We analyzed differences in risk subgroups and clinical characteristics. In addition, we used the 'rms' package to construct nomogram containing each IRG. The assessment of the accuracy of model was achieved. In addition, we performed a two-dimensional principal component analysis (PCA) to explore the differences in the discrete state distributions of different risk groups. We combined the top10 genes in the PPI network with prognosis-related IRGs from univariate Cox regression analysis in the entire cohort to obtain two hub IRGs by Venn plot.

Immune Infiltration Assessment

The 'GSVA' package in R software was used to perform a gene set enrichment analysis ssGSEA algorithm to unambiguously present the infiltrating score of 29 tumor-infiltrating immune cells and pathways (aDCs, APC co-inhibition, APC co-stimulation, B cells, CCR, $CD8^+$ T cells, Check-point, Cytolytic activity, DCs, HLA, iDCs, Inflammation-promoting, Macrophages, Mast cells, MHC class I, Neutrophils, NK cells, Parainflammation, pDCs, T cell co-inhibition, T cell co-stimulation, T helper cells, Tfh, Th1 cells, Th2 cells, TIL, Treg, Type I IFN Response, and Type II IFN Response). Also, the relationship between the risk subgroups and parameters related to immune cell infiltration in MM was explored.

CCK-8 Cell Proliferation Detection

RPMI8226 and MM1S cells (2000 cells/well) were inoculated in 24-well plates and cultured for 24 h using Cell Counting Kit-8 (CCK-8 Kit) from Beyotime (Shanghai, China) (sort code C0037). CCK-8 Cell Proliferation Assay Kit (C0037) was purchased from Beyotime (Shanghai, China) and cells were assayed according to its instructions viability. The human MM cell lines RPMI8226 and MM1S were donated by the Department of Hematology, Renji Hospital, Shanghai Jiaotong University School of Medicine, China.

EdU Detection

EdU-488 cell proliferation assay kit Beyotime (catalog number C0071S), RPMI8226 and MM1S cells (104 cells/well) were placed in 24-well plates using BeyoClick™ and cultured for 24 h. Cell proliferation capacity was detected using the EdU-488 Cell Proliferation Assay Kit (C0071S) purchased from Beyotime (Shanghai, China) according to the instructions.

Western Blot

RPMI8226 and MM1S cells (5×10^5 cells/well) were inoculated in 6-well plates and cultured for 24 h. After transfection and growth to 95%, cells were lysed and harvested, and protein concentrations were determined. Primary antibodies and their dilution working solutions were as follows: anti-IRF7 (1:1000), SCH1 (1:1000), HRP-conjugated secondary antibody (1:2000). The Ultra Enhanced ECL kit (G3308, GBCBIO) was used to amplify the exposure signal for western blot (WB) analysis. Grayscale analysis of WB bands was performed using ImageJ software.

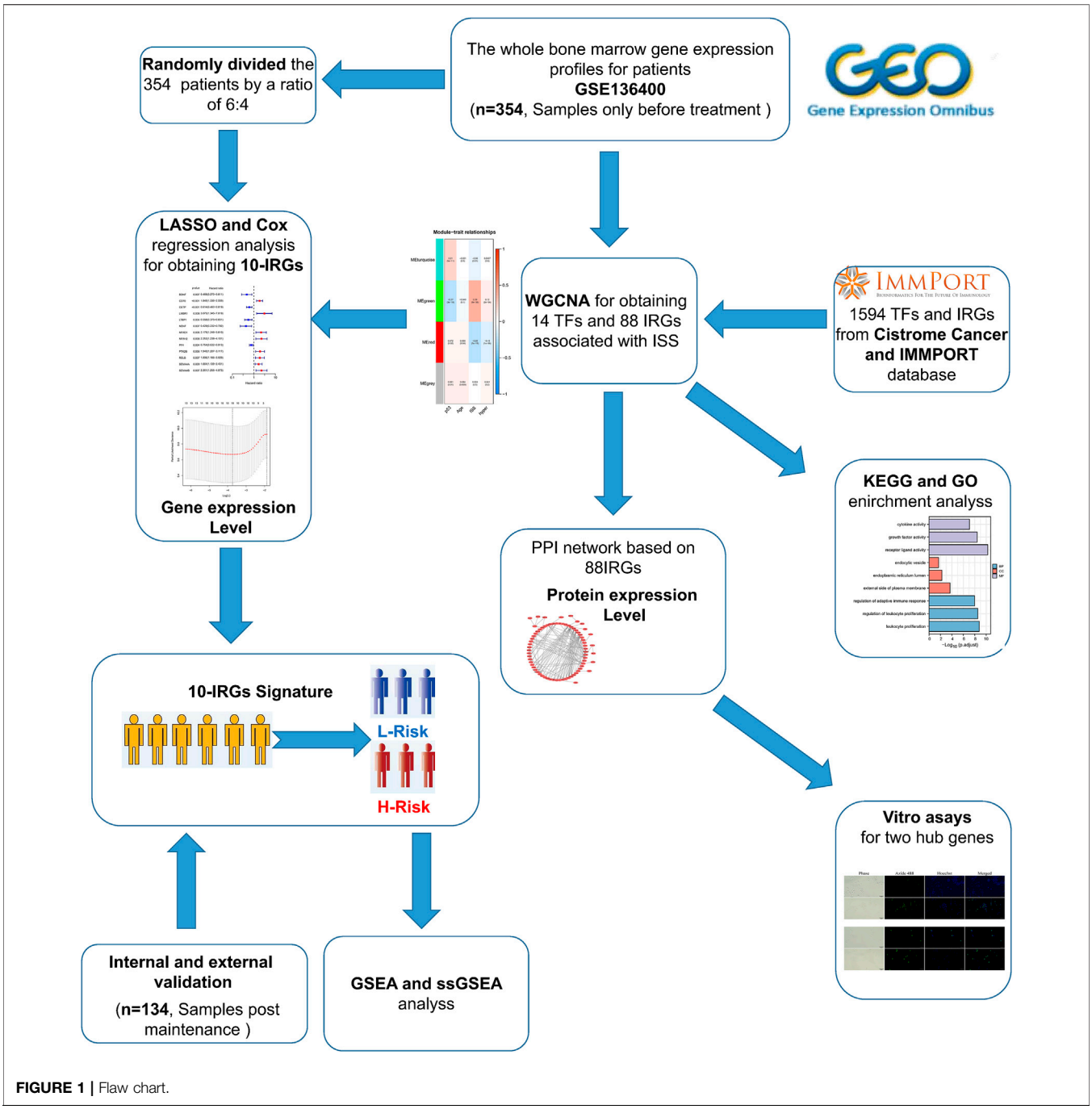


FIGURE 1 | Flaw chart.

qPCR

Total microarray was extracted from hepatocellular carcinoma cells and tissues using the Total microarray Extraction Kit (R4107; GBCBIO, Guangzhou, China). then, the microarray concentration was measured by nanodrop. Transcript First Strand cDNA Synthesis Kit (0489703000; Roche) was used for the reverse transcription reaction of microarray. Finally, qRT PCR of IRF7 and SCH1 was performed using the Light Cycle 480 SYBR Green I Master Kit (04707516001; Roche) on a Light Cycle 480®II instrument with internal microarray control for GAPDH. we used the 2-ΔCT method to infer the relative expression levels of microarray. All primers for microarray are listed below:

IRF7: F primer-CTTCGTGATGCTGCGGGATA, R primer-TTCTCGCCAGCACAGCTC, Product length 85bp. SHC1: F primer-AGGTCCAACCAGGCTAAGGG, R-primer: GGGGGC AGGAGATCCATAGT, Product length 120bp.

Statistical Analysis

All statistical analyses were performed using the R software (v.4.0.1). The Wilcoxon test was applied for pairwise

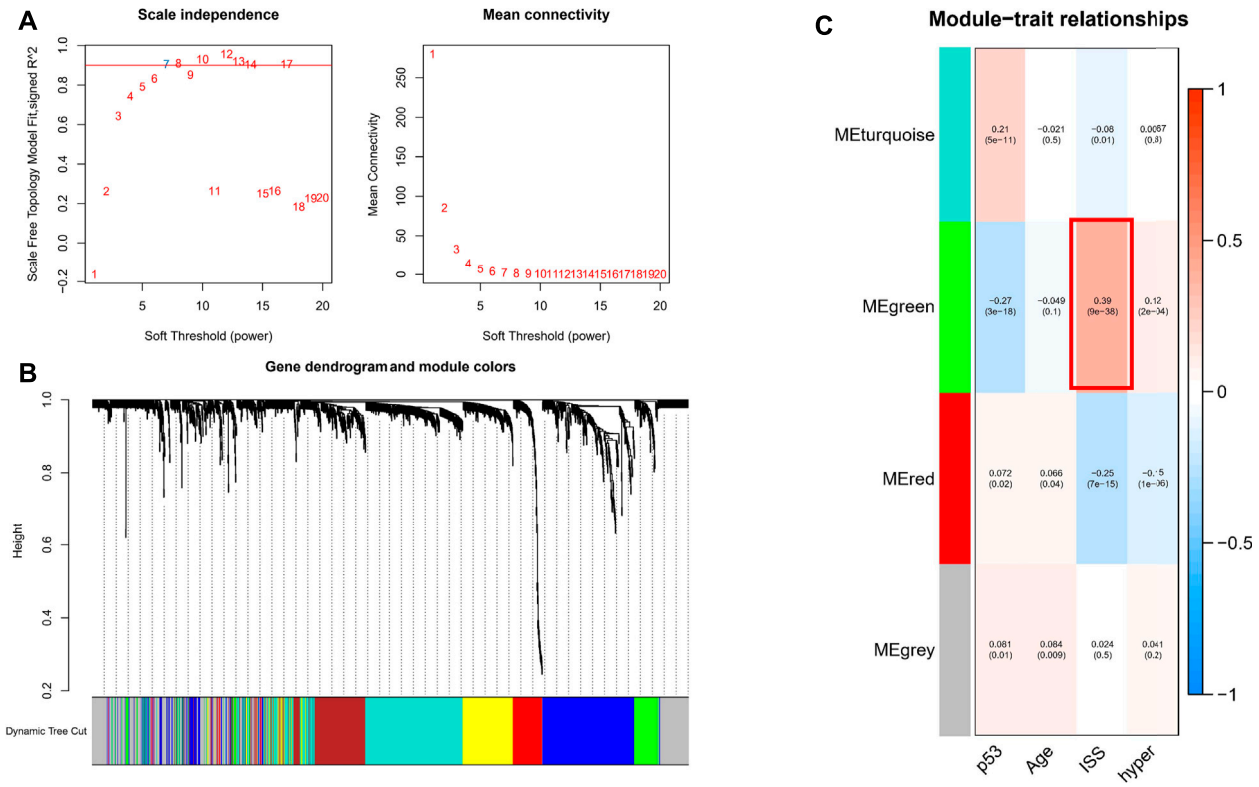


FIGURE 2 | WGCNA of all samples. **(A)** Soft threshold was identified by scale independence and mean connectivity. **(B)** Transcriptome data was classified into different modules. **(C)** Association between the modules and clinical traits.

comparisons. The Kaplan-Meier analysis with the log-rank test was adopted for overall survival comparisons. More detailed statistical methods for transcriptome data processing are covered in the above section. $p < 0.05$ was considered statistically significant.

RESULTS

Data Pre-processing

The flow chart of our study was shown in **Figure 1**. To investigate the immune-related features of MM and their prognostic associations, we downloaded the whole bone marrow transcriptome microarray dataset and clinical information of MM patients from the GEO database. Subsequently, we randomly divided the 354 pre-treatment patients in the GSE136400 dataset by a ratio of 6:4. Of these, 214 patients were used as the training set and the remaining 120 patients included in the internal validation set. Bioinformatics analysis was subsequently performed.

Identification of MM Progression-Related IRGs and TFs

To investigate genes that may be involved in MM disease progression, we focused on IRGs and TFs, as they may play a major role in tumor progression. We constructed WGCNA for all patient samples and excluded one sample (GSM4045581) based on the clustering. We classified the transcriptome data into genes modules based on the topological overlap matrix (TOM) and optimal soft threshold ($\beta = 7$) (**Figures 2A,B**). The TOM was obtained from: the adjacency matrix (matrix of weighted correlation values between genes) was converted to a topological overlap matrix to reduce noise and false correlation, and the new distance matrix was obtained. Subsequently, we calculated the correlation between modules and clinical features using Pearson method. The genes of the entire data were divided into four modules, with the green module being the hub module for International Staging System (ISS) ($r = 0.39$, $p < 0.05$) (**Figure 2C**). Moreover, the classification categories from WGCNA present different proportions of cell types. Most of all, green module had the strongest positive correlation with Treg ($r = 0.57$), and the strongest negative correlation with macrophages ($r = -0.65$), as shown in **Supplementary Figure S1**. We overlapped the genes in the green module with the known TFs and IRGs of the database, 101 TFs and IRGs associated with ISS were identified, which included 88 IRGs (**Figure 3A**) and 14 TFs (**Figure 3C**). This result suggested that these 88 IRGs and 14 TFs may drive disease progression in MM.

Functional Enrichment Analysis

To investigate the biological functions of these 101 TFs and IRGs that may be involved in disease progression, GO and KEGG analysis were performed on the above-mentioned IRGs and TFs.

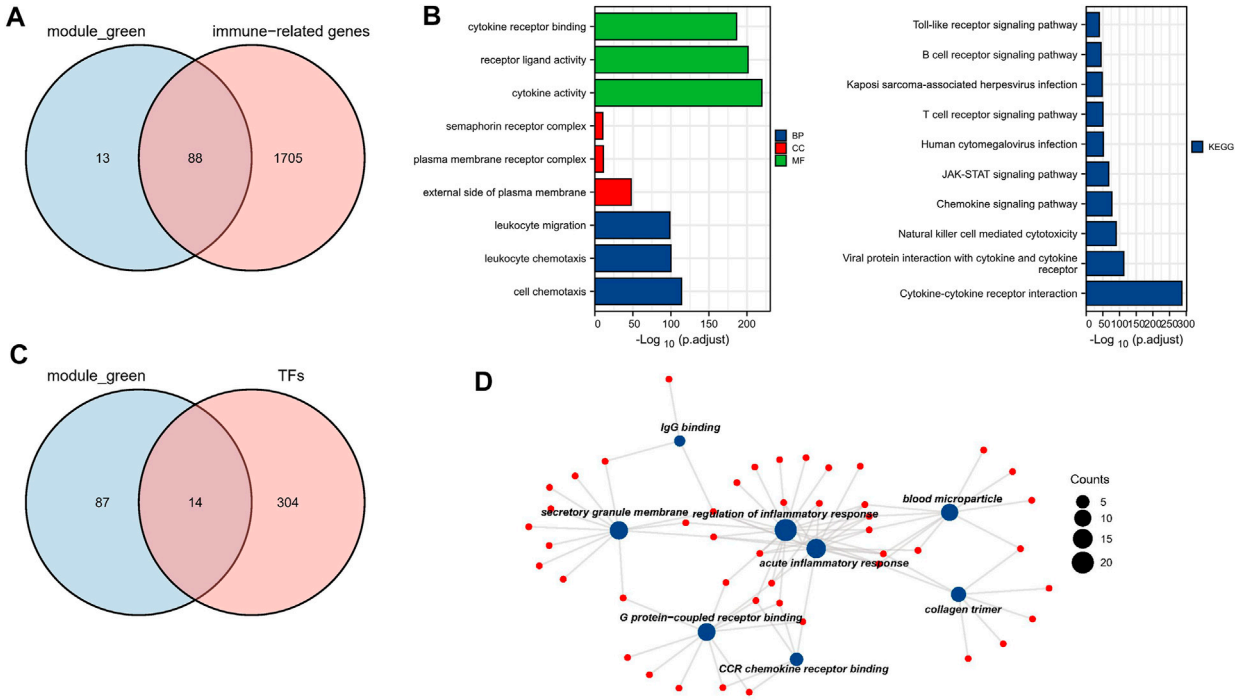


FIGURE 3 | Biological function analysis of TFs and IRGs. **(A)** The venn plot of green module and IRGs. **(B)** IRGs for GO and KEGG pathway enrichment analysis **(C)** The venn plot of green module and TFs. **(D)** The main biological process of TFs enrichment.

GO analysis showed that genes were enriched mainly in cytokine activity, receptor ligand activity, and leukocyte migration, etc. (**Figure 3B**). In KEGG analysis, genes were mainly enriched in immune-related pathways such as cytokine-cytokine receptor interaction and natural killer cell mediated cytotoxicity (**Figure 3C**), suggesting that these genes may influence tumor immunity and thus regulate MM progression. Meanwhile, TFs were mainly enriched in biological processes such as IgG binding, G protein-coupled receptor binding and CCR chemokine receptor binding (**Figure 3D**), and their functions were highly diversified, suggesting that these TFs may have pro-cancer potential.

Construction and Validation of Immune-Related Prognostic Signature

Firstly, univariate Cox regression analysis was performed on all IRGs to identify potential survival-related IRGs (**Figure 4A**) in the training cohort. LASSO regression analysis (10-flods) was used to select IRGs to avoid potential over-fitting (**Figure 4B**). The coefficient of 10-IRGs were identified by multivariate Cox regression analysis and used to develop optimal prognostic characteristics for OS in MM patients (**Figure 4C**). The risk score formula was obtained based on 10-IRGs = (-0.2138*BDNF) + (0.6791*CD70) + (-0.3790*CETP) + (0.8628*LMBR1) + (-0.1201*LTBP1) + (-1.0512*NENF) + (0.2852* NR1D1) + (0.5262*NR1H2) + (0.1247*PTK2B) + (0.0623*SEMA4B). Subsequently, patients were divided into high-risk and low-risk groups by median risk score. In addition, we constructed a nomogram based on the 10 IRGs (**Figure 4D**). In both the training and internal testing sets, the calibration curves showed that the one-year, three-year and five-year survival predictions were consistent with the actual observations, indicating that the prediction models were likely to be accurate (**Figures 4E,F**). Subsequently, we performed PCA analysis to explore the discrete distribution between the high-risk and low-risk groups, and the results indicated that risk profile was able to accurately differentiate patients (**Figures 4G,H**). To further validate the reliability of the prognostic model, we plotted the distribution of risk scores, survival status and corresponding gene expression levels of the selected individuals in the training (**Figures 5A,B**) and internal testing sets (**Figures 5E,F**). In the training set, the AUC values for survival prediction at 1, 3 and 5 years were 0.681, 0.676, and 0.724 (**Figure 5C**). Kaplan-Meier analysis showed a better prognosis for MM patients in the low-risk group compared to the high-risk group (**Figure 5D**). In the internal testing set, the AUC values for survival prediction at 1, 3 and 5 years were 0.550, 0.609, and 0.600 (**Figure 5G**). Kaplan-Meier analysis showed that the risk stratification system was still discriminating for OS in MM patients (**Figure 5H**), although it may be inappropriate for one-year survival prediction. Moreover, we validated the predictive power of our signature for long-term prognosis in post maintenance patients (n = 134). In another testing set, we also plotted the distribution of risk scores, and survival status (**Supplementary Figure S2A**). As the risk score increased, more patients died. Especially, for long-term survival prediction, ROC curve analysis showed that risk score had high predictive ability (AUC >0.7) (**Supplementary Figure S2B**). Moreover, Kaplan-Meier analysis also revealed that the risk

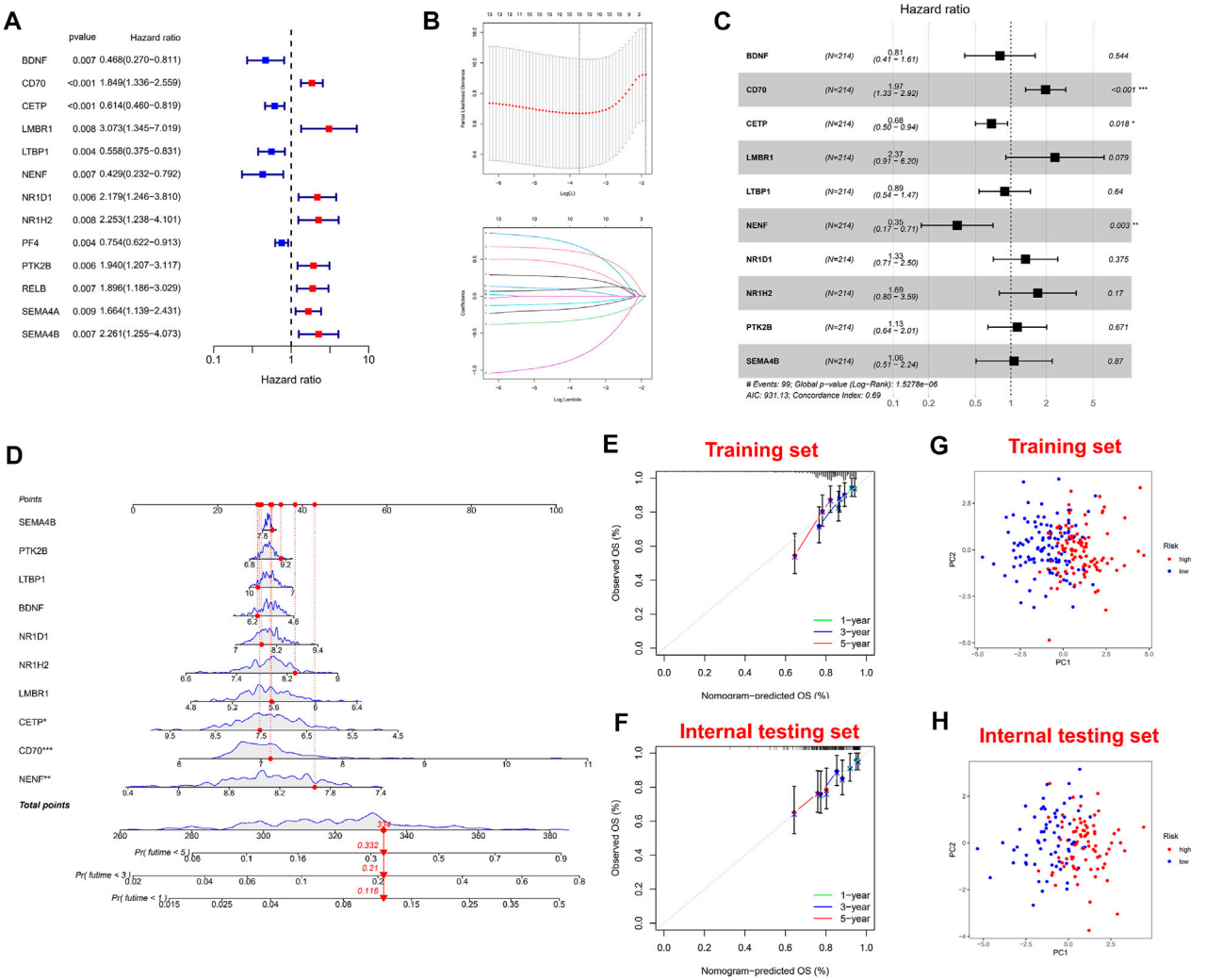


FIGURE 4 | Construction and validation of immune-related signature. **(A)** A forest of univariate cox regression analysis in training set. **(B)** LASSO regression analysis for most suitable λ. **(C)** A forest of multivariate cox regression analysis. **(D)** Nomogram based on 10-IRGs. **(E)** Calibration curve in the training set. **(F)** Calibration curve in the internal testing set. **(G)** PCA plot in the training set. **(H)** PCA plot in the internal testing set. The red dots represent high-risk patients and blue dots represent low-risk patients.

stratification system was still discriminating (**Supplementary Figure S1C**).

It is worth noting we also conducted a web calculator to identify survival possibility in MM patients (http://www.empowerstats.net/pmodel/?m=0_immunesignatureFORmm).

To assess the independent prognostic value in the prognostic model, we performed univariate and multivariate Cox regression analyses. Risk score was associated with OS in MM individuals in either the training or validation set (**Figures 6A,B**). Similarly, risk score was an independent prognostic factor for survival in MM patients (**Figures 6C,D**). Subsequently, in entire cohort, we explored the correlation between risk scores and clinicopathological parameters. The results showed that our risk scores were significantly correlated with age (**Supplementary Figure S3A**), ISS (**Supplementary Figure S3B**), p53 mutation status (**Supplementary Figure S3C**), albumin (**Supplementary Figure S3D**), β2-MG (**Supplementary Figure S3E**), and LDH (**Supplementary Figure S3F**). The feasibility of progression to advanced tumors gradually increased with increasing risk score, suggesting risk score could be as a indicator in MM progression.

Immune Infiltration and Biological Pathways in Patients With Different Risk

A growing number of studies suggest that the tumor microenvironment has an important and essential role in the response to immunotherapy. The tumor microenvironment can be reflected in the type and number of immune cells in the tumor. To further understand the relationship between risk characteristics, we used ssGSEA algorithm to explore the TME in MM. Notably, risk scores were significantly associated with 10

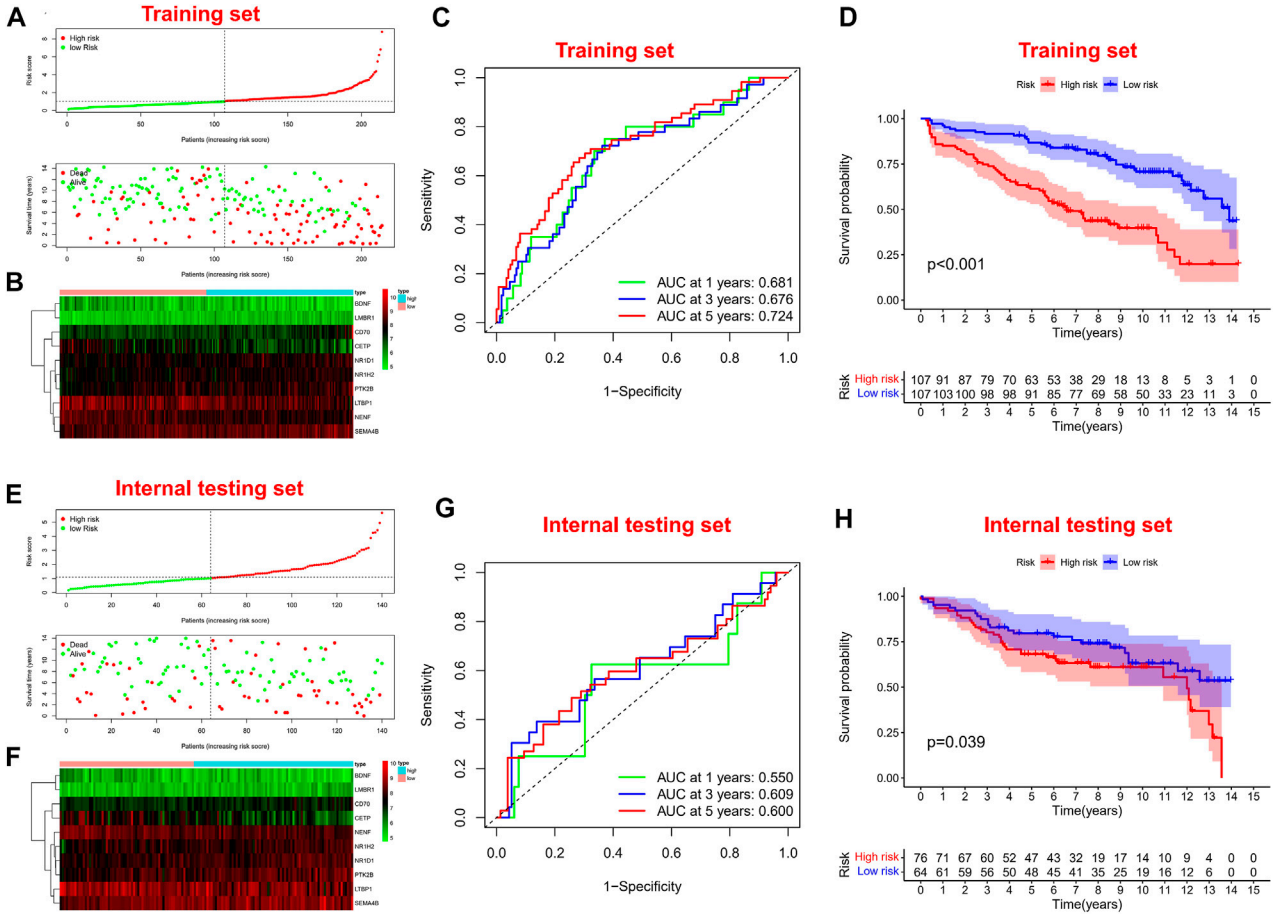


FIGURE 5 | Survival prognostic prediction to test the prognostic model. **(A, B)** Distribution of risk scores, survival status and corresponding gene expression levels of patients in the training set. **(C)** ROC analysis about one, three, and five-year survival prediction in the training set. **(D)** Kaplan–Meier analysis in the training set. **(E, F)** Distribution of risk scores, survival status and corresponding gene expression levels of patients in the internal testing set. **(G)** ROC analysis about one, three, and five-year survival prediction in the internal testing set. **(H)** Kaplan–Meier analysis in the internal testing set.

immune cell types, including aDCs, B cells, and Treg, among others (**Figure 7A**). 8 relevant immune pathways were significantly associated with the expression of risk scores, including APC co-stimulation, CCR signaling, and immune checkpoints (**Figure 7B**). The ssGSEA results further confirmed that the risk score could indicate the immune status of the TME. Next, GSEA was used to investigate potential biological pathways differences between high-risk and low-risk MM patients. Humoral immune responses and functional pathways, such as cell cycle and DNA synthesis were significantly enriched in the low-risk group (**Figures 7C,E**). Biological processes such as viral defense responses, endoplasmic reticulum protein transport and signaling pathways such as microarray degradation and protein transport were significantly enriched in the high-risk group (**Figures 7D,F**).

Identify Hub IRGs in PPI Network

To identify potential interaction networks in protein level between 88 IRGs, a circular PPI network (STRING database) was mapped using Cytoscape software (**Figure 8A**). Also, the top 10 IRGs in topology degree were screened (**Figure 8B**). Subsequently, by Cox regression analysis based on 88 IRGs, we found a total of 16 IRGs significantly associated with OS for the entire cohort (**Figure 8C**). Then, we overlapped the Top10 genes in the PPI network and prognostic genes (**Figure 8D**). Finally, two genes overlapped at the Venn plot, including IRF7 and SHC1. Hence, the above two IRGs were identified as hub IRGs associated with prognosis in protein-protein interaction level. In addition, TFs were identified as important molecules directly regulating the expression of other genes. Therefore, we explored the potential interactions between the 14 TFs in WGCNA and the hub IRGs (IRF7 and SHC1). Excitingly, the results suggested that interactions between TFs and hub IRGs indeed exist (**Figure 8E**).

IRF7 and SHC1 Promote Tumor Cell Proliferation

The results of our analysis suggest that IRF7 and SHC1 genes are core IRGs associated with MM prognosis and may be key factors

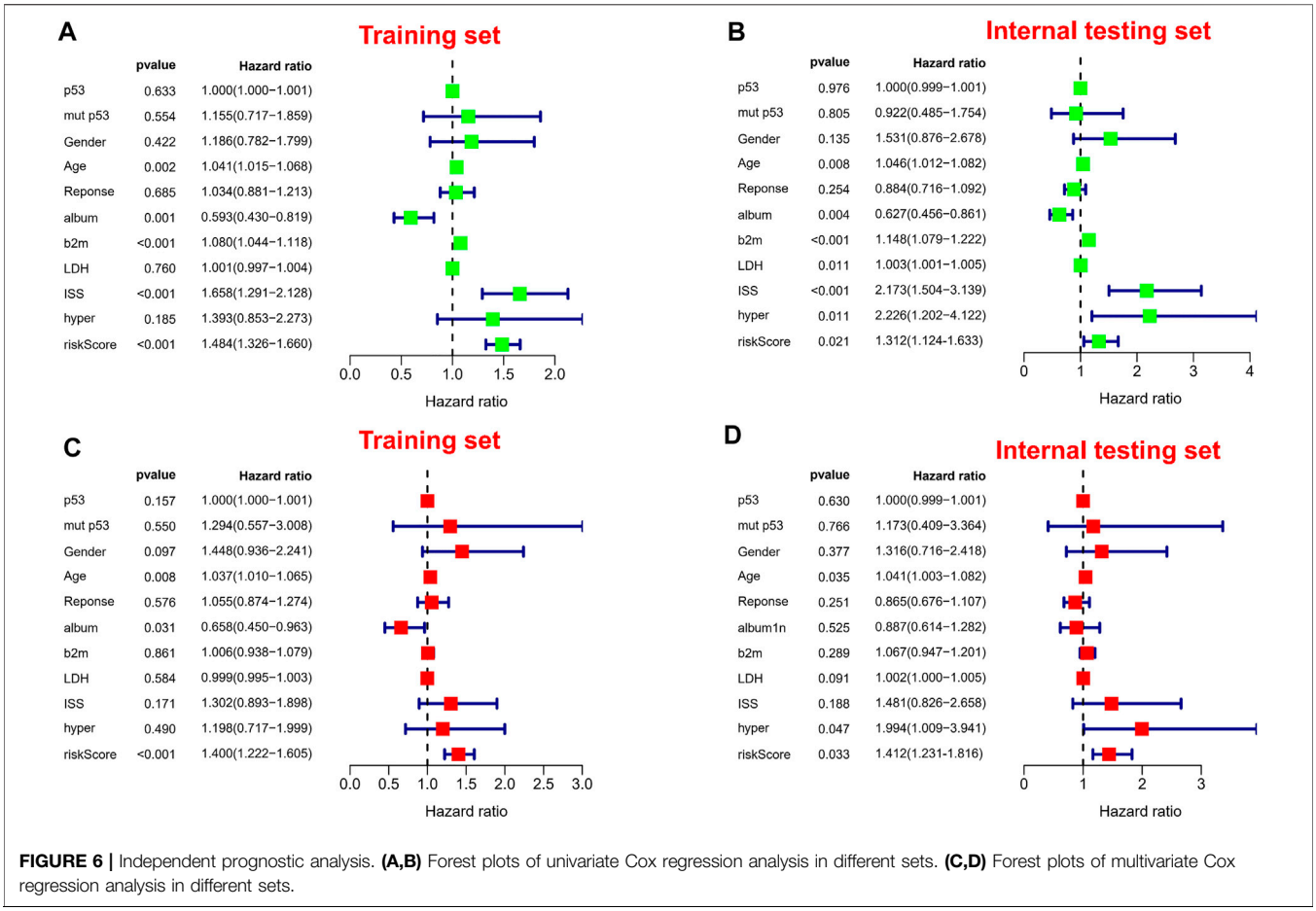


FIGURE 6 | Independent prognostic analysis. **(A,B)** Forest plots of univariate Cox regression analysis in different sets. **(C,D)** Forest plots of multivariate Cox regression analysis in different sets.

in MM disease progression. To validate this analysis, we overexpressed IRF7 and SHC1 in MM cell lines *in vitro*. QPCR experiments and Western Blot assays verified the successful overexpression of IRF7 and SHC1 (**Figures 9A–D**). CCK8 immunofluorescence staining and EDU experiments showed that compared to controls, overexpression of both IRF7 and SHC1 promoted tumor cell proliferation (**Figures 9E–H**). These results support the conclusions of our bioinformatic analysis.

DISCUSSION

As a clonal plasma cell abnormal proliferative tumor in the bone marrow, MM is accompanied by the secretion of large amounts of M proteins and is highly heterogeneous, leading to symptoms such as hypercalcemia, renal damage, anemia, bone destruction, and pathological clinical signs (Yanai et al., 2012; Corre et al., 2021). With the advent of novel drugs such as immunomodulators and proteasome inhibitors, the prognosis of MM patients has improved significantly, but patients are still repeatedly admitted to hospital for relapse and progression, so multiple myeloma remains an incurable type of disease. Therefore, it is clinically important to explore new molecular biological markers to track the treatment effect of MM, predict the disease progression, and provide more effective treatment options for MM. In our study, we screened out 10-IRGs involved in signature (BDNF, CETP, CD70, LMBR, LTBP1, NENF, NR1D1, NR1H2, PTK2B and SEMA4) and two hub IRGs in PPI network (IRF7 and SHC1). *In vitro* experiments showed that IRF7 and SHC1 could promote the proliferation of MM cell lines. It is suggested that IRF7 and SHC1 may play an important role in promoting the progression of MM. We believed that the above 12 novel markers could provide more possibilities for future MM therapies.

IRF7 is a major regulator of viral immune responses, which is type I interferon-dependent and tumorigenic (Lan et al., 2019). IRF7 not only affects tumor growth and malignant transformation of various tumor populations, but also regulates the development of myeloid-derived suppressor cells in cancer (Robak et al., 2018). Previous reports have shown that IRF7 is highly necessary for monocytes to differentiate them from macrophages (Lu and Pitha, 2001). In IRF7-deficient mice, it has potential effects on the accumulation of immature myeloid cells and on the dynamics of IRF7 expression in myeloid cell differentiation. Factors from tumors can prevent IRF7 expression in myeloid progenitor cells, which may lead to the accumulation of G-MDSC (Yang et al., 2017). Targeting IRF7 may help to reverse the abnormal differentiation of myeloid cells and thus play a role in tumor immunotherapy. This suggests that

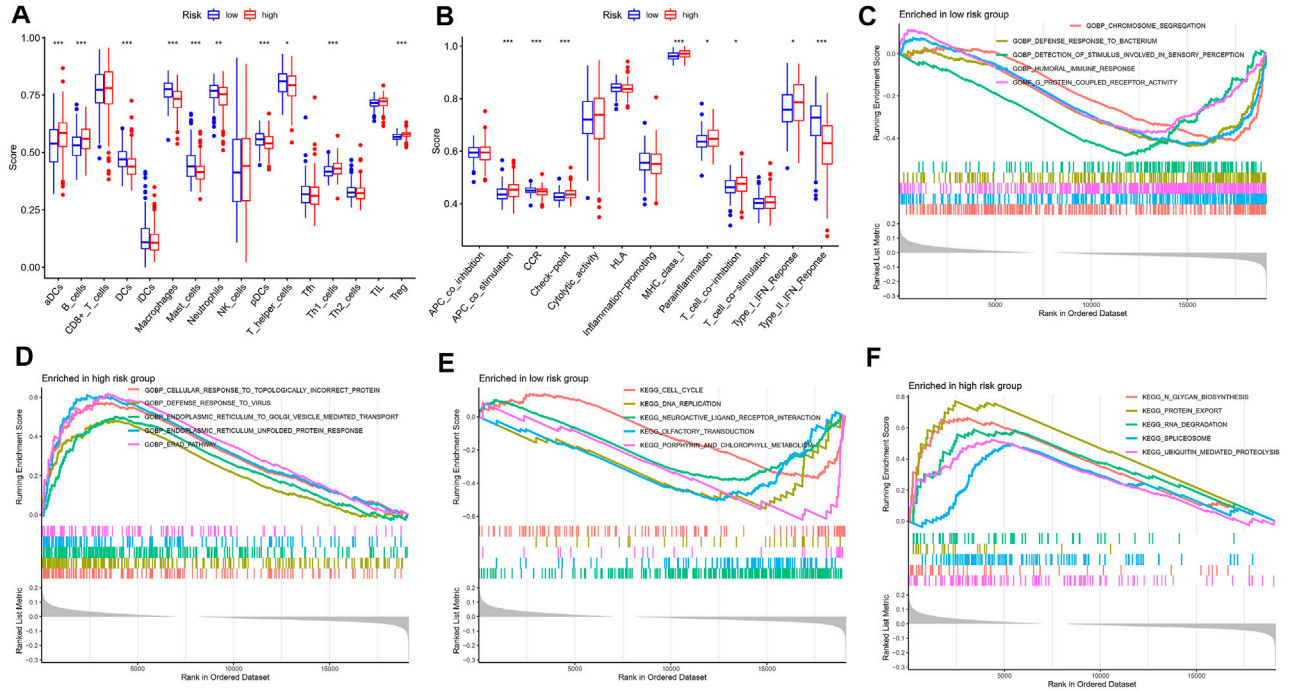


FIGURE 7 | Immune infiltration analysis and GSEA. **(A)** A box plot showed difference of 21 immune cells in different risk subgroups. **(B)** A box plot showed difference of eight immune related pathways in different risk subgroups. **(C,D)** GSEA analysis in the high-risk group. **(E,F)** GSEA analysis in the low-risk group.

IRF7 may be the key to MM immunotherapy. Overexpression of SHC1 promotes activation of MM cell lines and progression of MM. the SHC1 gene encodes an adaptor protein that is an important regulator of several tyrosine kinase signaling pathways. In other oncology studies, it has been suggested to promote immunosuppression and is a key regulator of breast cancer (Ahn et al., 2017). Furthermore, overexpression of SHC1 is associated with low survival rates in stage IIA colon cancer (Grossman et al., 2007). Previous studies have suggested that SHC1 associated with imbalance in integrin expression may be a prognostic predictor of clear cell renal cell carcinoma (ccRCC) (Lu et al., 2016). Interestingly, in our present study, SHC1 was an important hub IRGs in the PPI network, suggesting that SHC1 may play a general broad-spectrum function in tumor progression.

In recent years, it has been found that the bone marrow microenvironment plays a key role in the development of MM. The bone marrow microenvironment is composed of immune cells, fibroblasts, bone marrow-derived inflammatory cells and lymphocytes. Under normal conditions, natural killer cells (NK cells) and cytotoxic1 lymphocytes are present in the bone marrow environment and can exert a powerful anti-tumor response. However, the immunosuppressive microenvironment arises due to the presence of tumor cells, which can be of great benefit in expanding the immunosuppressive cell population (Haabeth et al., 2020). A better understanding of the tumor microenvironment can help to determine the prognostic value and therapeutic outcome of MM patients. Immunotherapy is an important and effective treatment for a large number of tumors, and IRGs are closely associated with tumor progression (Murray and Anagnostou, 2021). Currently, MM remains a difficult area of treatment due to recurrence and repeated hospital admissions. Therefore, the discovery of a more powerful tool is an urgent need, and immunotherapy has become a new focus of public attention. Although there has been an increasing number of studies on the relevance of immunotherapy to MM in recent years, more in-depth basic exploration and clinical trials are still needed to apply IRGs to clinical diagnosis and treatment. In our study, we developed a IRGs signature, and the important role of our signature in prognosis was confirmed by various statistical methods. In both the training and testing sets, the calibration curves showed that the one-year, three-year and five-year survival predictions were consistent with the actual observations. PCA analysis to explore the discrete distribution between the high-risk and low-risk groups, In the training set, the AUC values for survival prediction at 1, 3 and 5 years were 0.681, 0.676, and 0.724. Kaplan-Meier analysis showed a better prognosis for MM patients in the low-risk group compared to the high-risk group. Moreover, we validated the predictive power of our signature for long-term prognosis in post maintenance patients (n = 134). As the risk score increased, more patients died. Especially, for long-term survival prediction, ROC curve analysis showed that risk score had high predictive ability. Compared with $CD138^{+}$ selected cells microarray, whole bone marrow microarray is cheaper and easier to promote to clinical practice (Kuiper et al., 2012; Botta et al., 2016; Went et al., 2019). In our study, ssGSEA algorithm was used to show the changes of immune cells and immune function using the whole bone marrow samples. These are all analyses that CD138 + selected cells microarray can't do.

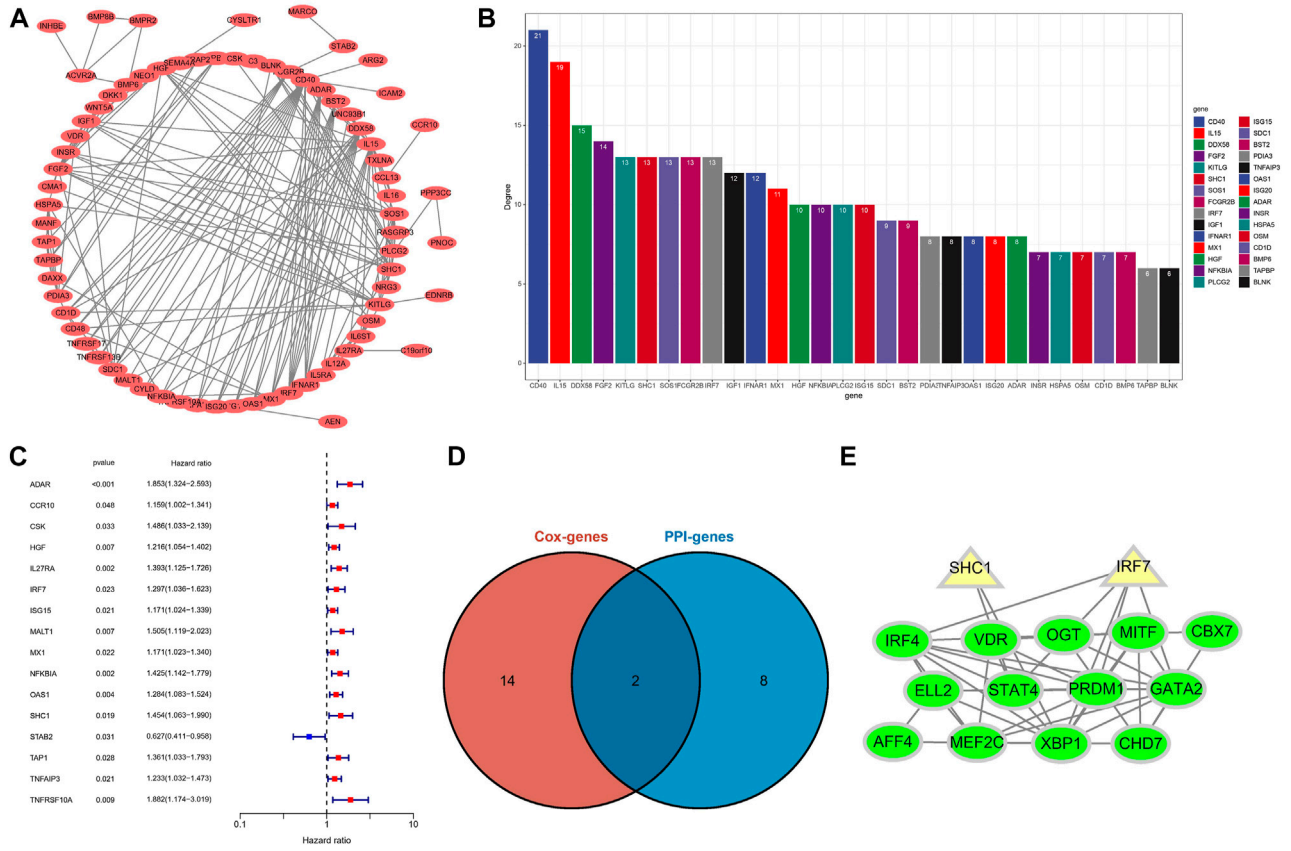


FIGURE 8 | Identification of hub IRGs related to prognosis in PPI network. **(A)** Based on the STRING database, 88 IRGs ring PPI networks are constructed. **(B)** PPI network analyzes topological degree. **(C)** A forest plot of univariate Cox regression analysis in the entire set. **(D)** Venn plot of the Top10 IRGs of PPI network and prognostic-related IRGs. **(E)** Analysis of the interaction between TFs and core IRGs.

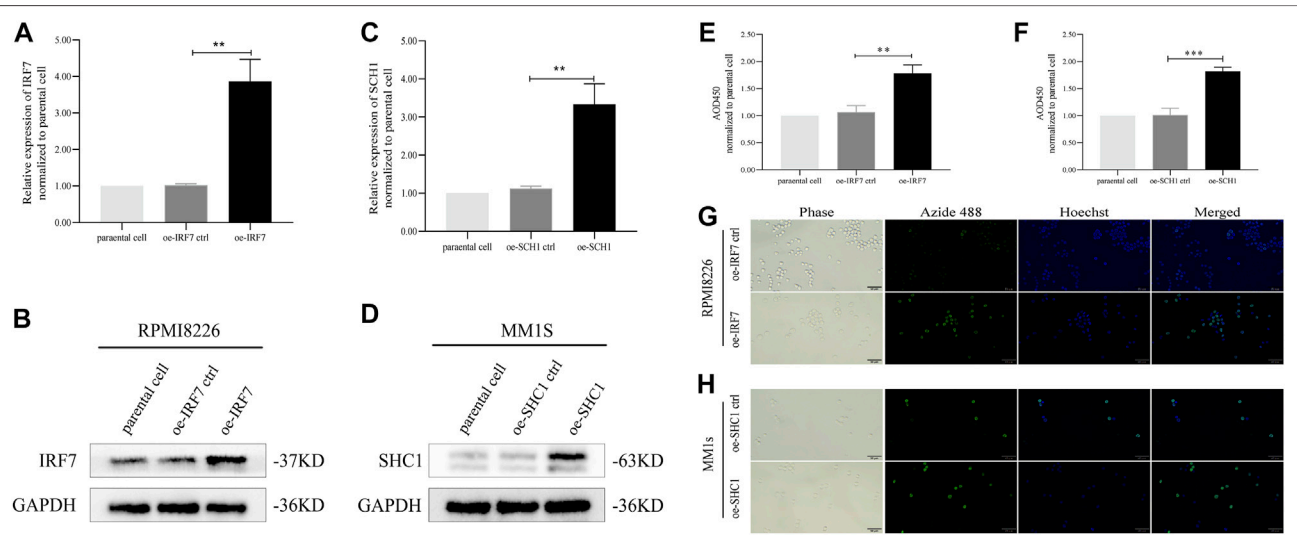


FIGURE 9 | IRF7, SHC1 promote tumor cell proliferation. **(A–D)** qPCR experiment and Western Blot experiment to detect the expression of IRF7 and SHC1. **(E–H)** CCK8 immunofluorescence staining and EDU test to detect the proliferation of tumor cells in the experiment and the control group MM. Protein expression was determined by western blotting and representative results from one of the three independent experiments are presented. Bar graphs were average of experimental replicates from three independent experiments. Error bars represent mean ± s.d.; $^{**}p < 0.01$; $^{***}p < 0.001$; by unpaired two-sided Student's t-test.

However, there are numerous limitations to our study that should be considered. Firstly, our research was only based on the GSE16400 dataset, and only pre-treatment whole bone marrow GEP can be used for survival prediction. More independent data sets are needed to verify the risk model we identified. When extending our findings to different treatment or GEP, caution is advised. Moreover, two hub genes in PPI network were validated *in vitro*, and the other 10-IRGs were not further explored. Hence, we will need to conduct more experiments in the future to confirm our conclusion. In conclusion, our study identified a risk model associated with MM prognosis through a series of bioinformatics analyses, and this risk score may have important implications for MM progression.

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

Q-SW, Q-QS, YM and M-PC were responsible for drawing the pictures. Furthermore, we have one corresponding authors in this manuscript. Q-SW, Q-QS and JH have contributed to information interpretation, editing and critical revision of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Ningbo Science and Technology Bureau (No. 202002N3174).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.897886/full#supplementary-material

REFERENCES

Ahn, R., Sabourin, V., Bolt, A. M., Hébert, S., Totten, S., De Jay, N., et al. (2017). The Shc1 Adaptor Simultaneously Balances Stat1 and Stat3 Activity to Promote Breast Cancer Immune Suppression. *Nat. Commun.* 8, 14638. doi:10.1038/ncomms14638

Botta, C., Di Martino, M. T., Ciliberto, D., Cucè, M., Correale, P., Rossi, M., et al. (2016). A Gene Expression Inflammatory Signature Specifically Predicts Multiple Myeloma Evolution and Patients Survival. *Blood Cancer J.* 6 (12), e511. doi:10.1038/bcj.2016.118

Corre, J., Munshi, N. C., and Avet-Loiseau, H. (2021). Risk Factors in Multiple Myeloma: Is it Time for a Revision? *Blood* 137 (1), 16–19. doi:10.1182/blood.2019004309

Gerecke, C., Fuhrmann, S., Strifler, S., Schmidt-Hieber, M., Einsele, H., and Knop, S. (2016). The Diagnosis and Treatment of Multiple Myeloma. *Dtsch. Arztebl Int.* 113 (27-28), 470–476. doi:10.3238/arztebl.2016.0470

Grossman, S. R., Lyle, S., Resnick, M. B., Sabo, E., Lis, R. T., Rosinha, E., et al. (2007). p66 Shc Tumor Levels Show a Strong Prognostic Correlation with Disease Outcome in Stage IIA Colon Cancer. *Clin. Cancer Res.* 13 (19), 5798–5804. doi:10.1158/1078-0432.CCR-07-0073

Haabeth, O. A. W., Hennig, K., Fauskanger, M., Løset, G. Å., Bogen, B., and Tveita, A. (2020). CD4+ T-Cell Killing of Multiple Myeloma Cells is Mediated by Resident Bone Marrow Macrophages. *Blood Adv.* 4 (12), 2595–2605. doi:10.1182/bloodadvances.2020001434

Hou, J., Wei, R., Qian, J., Wang, R., Fan, Z., Gu, C., et al. (2019). The Impact of the Bone Marrow Microenvironment on Multiple Myeloma (Review). *Oncol. Rep.* 42 (4), 1272–1282. doi:10.3892/or.2019.7261

Kawano, Y., Moschetta, M., Manier, S., Glavey, S., Görgün, G. T., Roccaro, A. M., et al. (2015). Targeting the Bone Marrow Microenvironment in Multiple Myeloma. *Immunol. Rev.* 263 (1), 160–172. doi:10.1111/imr.12233

Kawano, Y., Roccaro, A., Azzi, J., and Ghobrial, I. (2017). Multiple Myeloma and the Immune Microenvironment. *Curr. Cancer Drug Targets* 17 (9), 1. doi:10.2174/1568009617666170214102301

Kuiper, R., Broyl, A., de Knegt, Y., van Vliet, M. H., van Beers, E. H., van der Holt, B., et al. (2012). A Gene Expression Signature for High-Risk Multiple Myeloma. *Leukemia* 26 (11), 2406–2413. doi:10.1038/leu.2012.127

Lan, Q., Peyvandi, S., Duffey, N., Huang, Y.-T., Barras, D., Held, W., et al. (2019). Type I interferon/IRF7 Axis Instigates Chemotherapy-Induced Immunological Dormancy in Breast Cancer. *Oncogene* 38 (15), 2814–2829. doi:10.1038/s41388-018-0624-2

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

Liu, J., Hamrouni, A., Wolowiec, D., Coiteux, V., Kuliczkowski, K., Hetuin, D., et al. (2007). Plasma Cells from Multiple Myeloma Patients Express B7-H1 (PD-L1) and Increase Expression after Stimulation with IFN-{gamma} and TLR Ligands via a MyD88-, TRAF6-, and MEK-Dependent Pathway. *Blood* 110 (1), 296–304. doi:10.1182/blood-2006-10-051482

Lu, R., and Pitha, P. M. (2001). Monocyte Differentiation to Macrophage Requires Interferon Regulatory Factor 7. *J. Biol. Chem.* 276 (48), 45491–45496. doi:10.1074/jbc.C100421200

Lu, X., Wan, F., Zhang, H., Shi, G., and Ye, D. (2016). ITGA2B and ITGA8 are Predictive of Prognosis in Clear Cell Renal Cell Carcinoma Patients. *Tumor Biol.* 37 (1), 253–262. doi:10.1007/s13277-015-3792-5

Mireles-Cano, J. N., Escoto-Venegas, E., García-González, O. G., Miranda-González, A., González-Ramírez, J., Hernández-Sepúlveda, E., et al. (2020). A Management Algorithm for Vertebral Destruction Syndrome by Multiple Myeloma and Metastatic Spinal Cord Compression. *Acta Ortop. Mex.* 34 (5), 293–297. doi:10.35366/97991

Moser-Katz, T., Joseph, N. S., Dhodapkar, M. V., Lee, K. P., and Boise, L. H. (2021). Game of Bones: How Myeloma Manipulates its Microenvironment. *Front. Oncol.* 10, 625199. doi:10.3389/fonc.2020.625199

Murray, J. C., and Anagnostou, V. (2021). Translating Noninvasive Molecular Responses into Clinical Reality for Cancer Immunotherapy. *Nat. Rev. Clin. Oncol.* 18 (2), 65–66. doi:10.1038/s41571-020-00450-4

Robak, P., Drozdz, I., Szemraj, J., and Robak, T. (2018). Drug Resistance in Multiple Myeloma. *Cancer Treat. Rev.* 70, 199–208. doi:10.1016/j.ctrv.2018.09.001

Shen, X., Yang, Z., Feng, S., and Li, Y. (2021). Identification of Uterine Leiomyosarcoma-Associated Hub Genes and Immune Cell Infiltration Pattern Using Weighted Co-expression Network Analysis and CIBERSORT Algorithm. *World J. Surg. Onc* 19 (1), 223. doi:10.1186/s12957-021-02333-z

Tamura, H., Ishibashi, M., Sunakawa-Kii, M., and Inokuchi, K. (2020). PD-L1-PD-1 Pathway in the Pathophysiology of Multiple Myeloma. *Cancers* 12 (4), 924. doi:10.3390/cancers12040924

Went, M., Kinnersley, B., Sud, A., Johnson, D. C., Weinhold, N., Försti, A., et al. (2019). Transcriptome-Wide Association Study of Multiple Myeloma Identifies Candidate Susceptibility Genes. *Hum. Genomics* 13 (1), 37. doi:10.1186/s40246-019-0231-5

Yanai, H., Negishi, H., and Taniguchi, T. (2012). The IRF Family of Transcription Factors. *Oncoimmunology* 1 (8), 1376–1386. doi:10.4161/onci.22475

Yang, Q., Li, X., Chen, H., Cao, Y., Xiao, Q., He, Y., et al. (2017). IRF7 Regulates the Development of Granulocytic Myeloid-Derived Suppressor Cells through S100A9 Transrepression in Cancer. *Oncogene* 36 (21), 2969–2980. doi:10.1038/onc.2016.448

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 Genetics

ORIGINAL RESEARCH
published: 30 May 2022
doi: 10.3389/fgene.2022.904513

How the Replication and Transcription Complex Functions in Jumping Transcription of SARS-CoV-2

Jianguang Liang[1†], Jinsong Shi[2†], Shunmei Chen[3], Guangyou Duan[4], Fan Yang[2], Zhi Cheng[5], Xin Li[5], Jishou Ruan[6], Dong Mi[7] and Shan Gao[5]**

[1]School of Pharmacy, Changzhou University, Changzhou, China, [2]National Clinical Research Center of Kidney Disease, Jinling Hospital, Nanjing University School of Medicine, Nanjing, China, [3]Yunnan Key Laboratory of Stem Cell and Regenerative Medicine, Biomedical Engineering Research Center, Kunming Medical University, Kunming, China, [4]School of Life Sciences, Qilu Normal University, Jinan, China, [5]College of Life Sciences, Nankai University, Tianjin, China, [6]School of Mathematical Sciences, Nankai University, Tianjin, China, [7]Department of Clinical Laboratory, Affiliated Maternity Hospital, Nankai University, Tianjin, China

OPEN ACCESS

Edited by:
Tao Huang,
Shanghai Institute of Nutrition and Health (CAS), China

Reviewed by:
Xue Yao,
Harvard University, United States
Yanna Shen,
Tianjin Medical University, China

***Correspondence:**
Dong Mi
mrmidong@163.com
Shan Gao
gao_shan@mail.nankai.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: 25 March 2022
Accepted: 13 May 2022
Published: 30 May 2022

Citation:
Liang J, Shi J, Chen S, Duan G, Yang F, Cheng Z, Li X, Ruan J, Mi D and Gao S (2022) How the Replication and Transcription Complex Functions in Jumping Transcription of SARS-CoV-2. Front. Genet. 13:904513. doi: 10.3389/fgene.2022.904513

Background: Coronavirus disease 2019 (COVID-19) is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Although unprecedented efforts are underway to develop therapeutic strategies against this disease, scientists have acquired only a little knowledge regarding the structures and functions of the CoV replication and transcription complex (RTC). Ascertaining all the RTC components and the arrangement of them is an indispensably step for the eventual determination of its global structure, leading to completely understanding all of its functions at the molecular level.

Results: The main results include: 1) hairpins containing the canonical and non-canonical NSP15 cleavage motifs are canonical and non-canonical transcription regulatory sequence (TRS) hairpins; 2) TRS hairpins can be used to identify recombination regions in CoV genomes; 3) RNA methylation participates in the determination of the local RNA structures in CoVs by affecting the formation of base pairing; and 4) The eventual determination of the CoV RTC global structure needs to consider METTL3 in the experimental design.

Conclusions: In the present study, we proposed the theoretical arrangement of NSP12-15 and METTL3 in the global RTC structure and constructed a model to answer how the RTC functions in the jumping transcription of CoVs. As the most important finding, TRS hairpins were reported for the first time to interpret NSP15 cleavage, RNA methylation of CoVs and their association at the molecular level. Our findings enrich fundamental knowledge in the field of gene expression and its regulation, providing a crucial basis for future studies.

Keywords: coronavirus, RNA methylation, nanopore, TRS hairpin, METTL3

INTRODUCTION

Coronavirus disease 2019 (COVID-19) is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Li et al., 2020) (Duan et al., 2020) with a genome of ~30 kb (Jiayuan et al., 2020). By reanalyzing public data (Kim et al., 2020a), we determined that a SARS-CoV-2 genome has 12 genes, which are *spike (S)*, *envelope (E)*, *membrane (M)*, *nucleocapsid (N)*, and *ORF1a*, *1b*, *3a*, *6*, *7a*, *7b*, *8* and *10* (Li et al., 2021a). The *ORF1a* and *1b* genes encode 16 non-structural proteins (NSPs), named NSP1 through NSP16 (Silva et al., 2020), while the other 10 genes encode four structural proteins (S, E, M and N) and six accessory proteins (ORF3a, 6, 7a, 7b, 8 and 10). Among the above 26 proteins, NSP4-16 are highly conserved in all known CoVs and have been experimentally demonstrated or predicted to be critical enzymes in CoV RNA synthesis and modification (Denison et al., 2011), particularly including: NSP12, RNA-dependent RNA polymerase (RdRP) (Yan et al., 2020); NSP13, RNA helicase-ATPase (Hel); NSP14, RNA exoribonuclease (ExoN) and N7 methyltransferase (MTase); NSP15 endoribonuclease (EndoU) (Kim et al., 2020b); and NSP16, RNA 2′-O-MTase.

NSP1-16 assemble into a replication and transcription complex (RTC) (Yan et al., 2020). The basic function of the RTC is RNA synthesis: it synthesizes genomic RNAs (gRNAs) for replication or transcription of the *ORF1a*, *1b* genes, while it synthesizes subgenomic RNAs (sgRNAs) for jumping transcription of the other 10 genes (Kim et al., 2020a). In 1998, the "leader-to-body fusion" model (Sawicki et al., 1998) was proposed to explain the jumping transcription, however, the molecular basis of this model was unknown until our previous study in 2020 (Li et al., 2021a). In our previous study (Li et al., 2021a), we provided a molecular basis for the "leader-to-body fusion" model by identifying the cleavage sites of NSP15 and proposed a negative feedback model to explain the regulation of CoV replication and transcription. In addition, we revealed that the jumping transcription and recombination of CoVs share the same molecular mechanism (Li et al., 2021a), which causes rapid mutation and inevitably outbreaks of CoVs. These findings are vital for the further investigation of CoV transcription and recombination. However, there will be a long way to completely understand how the RTC functions in the jumping transcription at the molecular level.

For a complete understanding of CoV replication and transcription, particularly the jumping transcription, much research (Yan et al., 2020) (Kim et al., 2020b) (Hillen et al., 2020) has been conducted to determine the global structure of the SARS-CoV-2 RTC, since the outbreak of SARS-CoV-2 in 2019. Although some single protein structures (*e.g.*, NSP15 (Kim et al., 2020b)) and local structures of the RTC (i.e. NSP7&8&12&13 (Yan et al., 2020) and NSP7&8&12 (Hillen et al., 2020)) have been determined, the global structure and all components of RTC are still unknown. As the global structure of the CoV RTC cannot be determined by simple use any one of current methods (i.e., X-ray, NMR and Cryo-EM), ascertaining all the RTC components and the arrangement of them is an indispensably step for the eventual determination of its global structure, leading to completely understanding all of its functions at the molecular level. In the present study, we aimed to determine the theoretical arrangement of NSP12-16 in the global structure of the CoV RTC by comprehensive analysis of data from different sources, and to preliminarily elucidate how the RTC functions in the jumping transcription of CoVs at the molecular level.

RESULTS

Jumping Transcription, TRS and NSP15 Cleavage Site

First, we provide a brief introduction to the jumping transcription of CoVs, the "leader-to-body fusion" model proposed in an early study (Sawicki et al., 1998) and its molecular basis proposed in our recent study (Li et al., 2021a). In the "leader-to-body fusion" model, the realization of jumping transcription requires transcription regulatory sequences (TRSs), which include leader TRSs (TRS-Ls) and body transcription regulatory sequences (TRS-Bs). Each CoV genome contains a TRS-L in the 5′ untranslated region (UTR) and several TRS-Bs located in the upstreams of genes except *ORF1a and 1b*. CoV replication and transcription require gRNAs(+) as templates for the synthesis of antisense genomic RNAs [gRNAs(-)] and antisense subgenomic RNAs [sgRNAs(-)] by RdRP. When RdRP pauses, as it crosses a TRS-B and switches the template to the TRS-L, sgRNAs(-) are formed through jumping transcription (also referred to as discontinuous transcription, polymerase jumping or template switching). Otherwise, RdRP reads gRNAs(+) continuously, without interruption, resulting in gRNAs(-). Thereafter, gRNAs(-) and sgRNAs(-) are used as templates to synthesize gRNAs(+) and sgRNAs(+), respectively; gRNAs(+) and sgRNAs(+) are used as templates for the translation of NSP1-16 and the other 10 proteins (S, E, M, N, and ORF3a, 6, 7a, 7b, 8 and 10), respectively. In our previous study (Li et al., 2021a), we provided a molecular basis for the "leader-to-body fusion" model by identifying the reverse complimentary sequences of TRS-Bs [denoted as TRS-Bs(-)] as the NSP15 cleavage sites, which actually functions in the regulation of CoV regulation. NSP15 cleaves gRNAs(-) and sgRNAs(-) at TRS-Bs(-). Then, the free 3′ ends (~6 nt) of TRS-Bs(-) hybridize TRS-Ls to realize "leader-to-body fusion". These findings associated the investigation of TRSs to that of NSP15 cleavage sites.

In our previous study (Bei et al., 2022), we made a generalization that a TRS motif is a (6~8-nt long for CoVs) consensus sequence beginning with at least an adenosine residue (**A**), and enriched with A and followed by C, based on the analysis of 1,265 CoV genome sequences (**Materials and Methods**). We defined that the antisense sequence of a TRS motif as the motif of the corresponding NSP15 cleavage site (the NSP15 cleavage motif). For example, the canonical TRS motif of SARS-CoV-2 and the corresponding NSP15 cleavage motif are **A**CGAAC and GTTCG**T**, respectively. We defined the TRS motif in the TRS-L as the canonical TRS motif. Thus, the canonical TRS motif is unique to a CoV genome, while the TRS motifs in TRS-Bs can be canonical TRS motifs or non-canonical TRS motifs with little nucleotide (nt) differences. By these definitions, we determined canonical TRS motifs of all viruses in the order *Nidovirales*

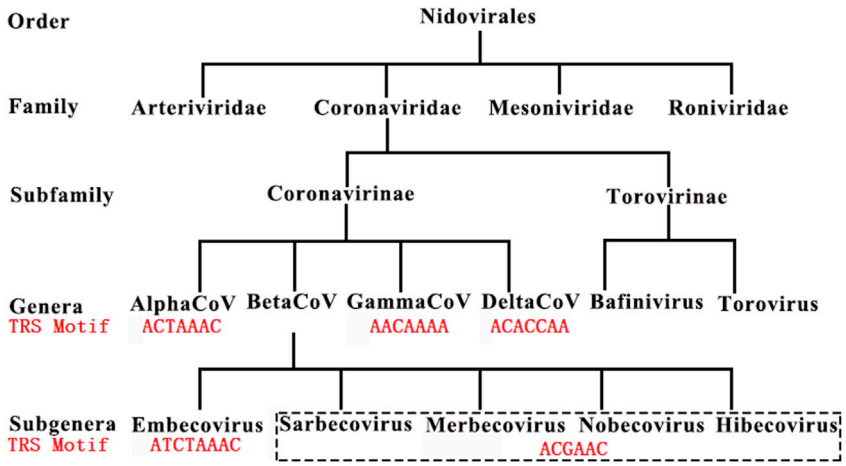


FIGURE 1 | Canonical TRS motifs in Coronaviridae. *Embecovirus*, *Sarbecovirus*, *Merbecovirus*, *Nobecovirus* and *Hibecovirus* are also defined as *Betacoronavirus* subgroups A, B, C, D and E. SARS-CoV and SARS-CoV-2 belong to *Betacoronavirus* subgroup B. These canonical TRS motifs (in red color) of viruses in Coronaviridae have been reported in our previous study (Bei et al., 2022).

(**Figure 1**) and corrected some canonical TRS motifs reported in the previous studies. For instance, the canonical TRS motifs of mouse hepatitis virus (MHV), transmissible gastroenteritis virus (TGEV), canada goose coronavirus (Goose-CoV) and beluga whale coronavirus (BWCoV) were corrected from CTAAAC (Grossoehme et al., 2009), CTAAAC (Sola et al., 2005), CTTAACAAA (Papineau et al., 2019) and AAACA (Mihindukulasuriya et al., 2008) to **A**TCTAAAC, **A**CTAAAC, **A**ACAAAA and **A**ACAAAA, respectively. Canonical TRS motifs are highly conserved in *Alphacoronavirus*, *Gammacoronavirus*, *Deltacoronavirus* and *Betacoronavirus* genera except the subgroup A (**Figure 1**). *Betacoronavirus* subgroup A has the canonical TRS motif **A**TCTAAAC, which is different from **A**CGAAC in *Betacoronavirus* subgroup B, C, D and E. Different from *Betacoronavirus* subgroup B, *Betacoronavirus* subgroup A, C, D and E, *Alphacoronavirus*, *Gammacoronavirus* and *Deltacoronavirus* have non-canonical TRS motifs in the TRS-Bs of four structural genes (*S*, *E*, *M* and *N*), which were caused by mutations during evolution. These TRS motif mutations resulted in the attenuation of CoVs in *Betacoronavirus* subgroup A, D and E by down-regulating the transcription of CoV genes except *ORF1a* and *1b* (Li et al., 2021b). This confirmed that TRSs (Actually revealed as the NSP15 cleavage sites (Li et al., 2021a)) function in the regulation of CoV transcription (Yount et al., 2006). Furthermore, a previous study reported that the recognition of a TRS (Actually revealed as the NSP15 cleavage site (Li et al., 2021a)) is independent on its motif, but dependent on its context (Yount et al., 2006).

NSP15 Cleavage, RNA Methylation and TRS Hairpin

A previous study (Kim et al., 2020a) reported that RNA methylation sites containing the "AAGAA-like" motif (including AAGAA and other A/G-rich sequences) are present throughout the SARS-CoV-2 genome, particularly enriched in genomic positions 28,500-29,500. This study used Nanopore RNA-seq (Xu et al., 2019), a direct RNA sequencing method, which can be used to measure RNA methylation at 1-nt resolution although it has a high error rate. By analyzing the Nanopore RNA-seq data, the previous study (Kim et al., 2020a) concluded that the methylated RNAs have shorter 3′ polyA tails than the unmethylated ones in SARS-CoV-2. Although the type of RNA methylation was unknown, the previous study (Kim et al., 2020a) proposed that the "AAGAA-like" motif associates with the lengths of 3′ polyA tails of gRNAs and sgRNAs. However, the previous study did not answer the following questions: 1) it was not explained that what functions the internal methylation sites have, as they are far from 3′ ends, thus unlikely to contribute to the lengths of 3′ polyA tails; and 2) the extremely high ratio between sense and antisense reads (Li et al., 2021a) may result from quick degradation of the antisense nascent RNAs due to their shorter 3′ polyA tails, however, the "AAGAA-like" motif occurs in both sense and antisense strands at a similar frequency. Notably, the previous study (Kim et al., 2020a) shouldn't have neglected the analysis of the "AAGAA-like" motif on the antisense strand, since only very a few antisense reads from the Nanopore RNA-seq data were obtained for analysis. Therefore, we proposed that RNA methylation sites containing the "AAGAA-like" motif may have other biological functions and conducted further analysis.

Different from the previous study (Kim et al., 2020a), our study focused on the analysis of the "AAGAA-like" motif on the antisense strand of the SARS-CoV-2 genome, particularly the association between the "AAGAA-like" motif and the TRS or corresponding NSP15 cleavage motifs. As a result, we discovered that the "AAGAA-like" motif co-occurred with the NSP15 cleavage motif GTTCG**T** of four genes (*S, ORF6, 7a* and *8*). In our previous study (Liu et al., 2018), complemented palindrome sequences in genomes of viruses in *Betacoronavirus* subgroup B have been investigated and most of them are semipalindromic or heteropalindromic. These complemented palindrome sequences

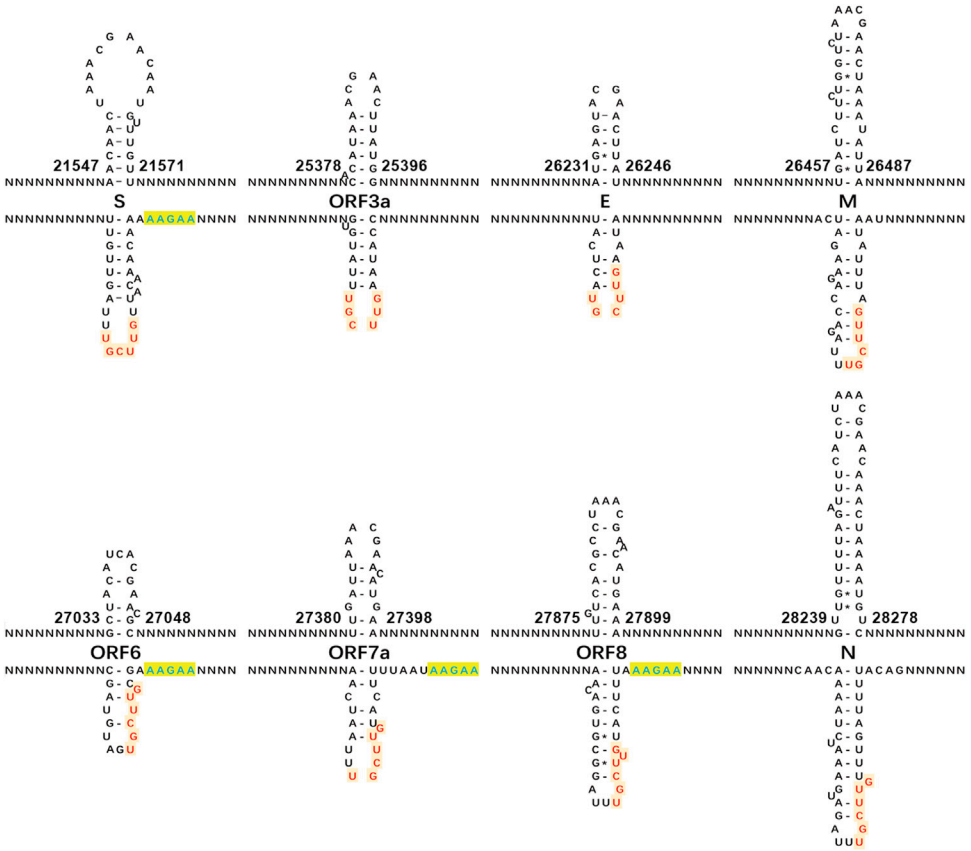


FIGURE 2 | Canonical TRS hairpins in SARS-CoV-2. The canonical transcription regulatory sequence (TRS) motif ACGAAC is present in the upstreams of eight genes (*S*, *E*, *M*, *N*, and *ORF3a*, *6*, *7a* and *8*). Read on the antisense strands of the SARS-CoV-2 genome (GenBank: MN908947.3), "AAGAA" (in blue color) or "AAACH" (**Supplementary Table S1**) represents an RNA methylation site, while "GUUCGU" (in red color) represents a NSP15 cleavage site. The positions are the start and end positions of hairpins in the SARS-CoV-2 genome. NSP15 cleaves a single-strand RNA after U (indicated by arrows). In the present study, we defined: (1) the hairpins containing the canonical and non-canonical NSP15 cleavage sites are canonical and non-canonical TRS hairpins, respectively; and (2) the hairpins opposite to the TRS hairpins as the opposite TRS hairpins.

containing A-rich and T-rich regions form hairpins. The "AAGAA-like" and GTTCG**T** motifs are located in the A-rich and T-rich regions. Thus, the association between the "AAGAA-like" and GTTCG**T** motifs was discovered by analysis of TRS hairpins of the four genes (**Figure 2**). For analysis of TRS hairpins, we defined: 1) hairpins containing the canonical and non-canonical NSP15 cleavage sites are canonical and non-canonical TRS hairpins, respectively; and 2) hairpins opposite to TRS hairpins are opposite TRS hairpins (**Figure 2**). However, the formation of opposite TRS hairpins is uncertain, as all complemented palindrome sequences forming the TRS and opposite TRS hairpins are asymmetric (semipalindromic or heteropalindromic). Among the 10 genes, eight (*S*, *E*, *M*, *N*, *ORF1a*, *1b*, *3a*, *6*, *7a*, and *8*) have canonical TRS hairpins and two (*ORF7b* and *10*) may have non-canonical TRS hairpins (**Supplementary Table S1**). Non-canonical TRS hairpins have been reported in seven common recombination regions in one of our previous studies (Li et al., 2021b) and identified in five recombination events (**Figure 3**) in another one of our previous studies (Li et al., 2021a). Therefore, TRS hairpins can be used to identify recombination regions in CoV genomes. NSP15 cleaves the canonical TRS hairpins of the seven genes at canonical breakpoints, whereas it cleaves the canonical TRS hairpin of *ORF3a* at an unexpected breakpoint "**GTTCGT**TTAT|N" (the NSP15 cleavage motif is underlined; the vertical line indicates the breakpoint and N represents any nt), rather than the end of the canonical NSP15 cleavage motif "**GTTCGT**|TTATN". According to our definitions, "**GTTCGT**|TTATN" and "**GTTCGT**TTAT|N″ are canonical and non-canonical NSP15 breakpoints, respectively. The discovery of non-canonical TRS hairpins and non-canonical NSP15 breakpoints indicated that the recognition of NSP15 cleavage sites is structure-based rather than sequence-based.

How RTC Functions in Jumping Transcription

Since several A-rich and T-rich regions are alternatively present around each NSP15 cleavage site, many hypothetical TRS hairpins (**Figure 4A–C**) containing the NSP15 cleavage site

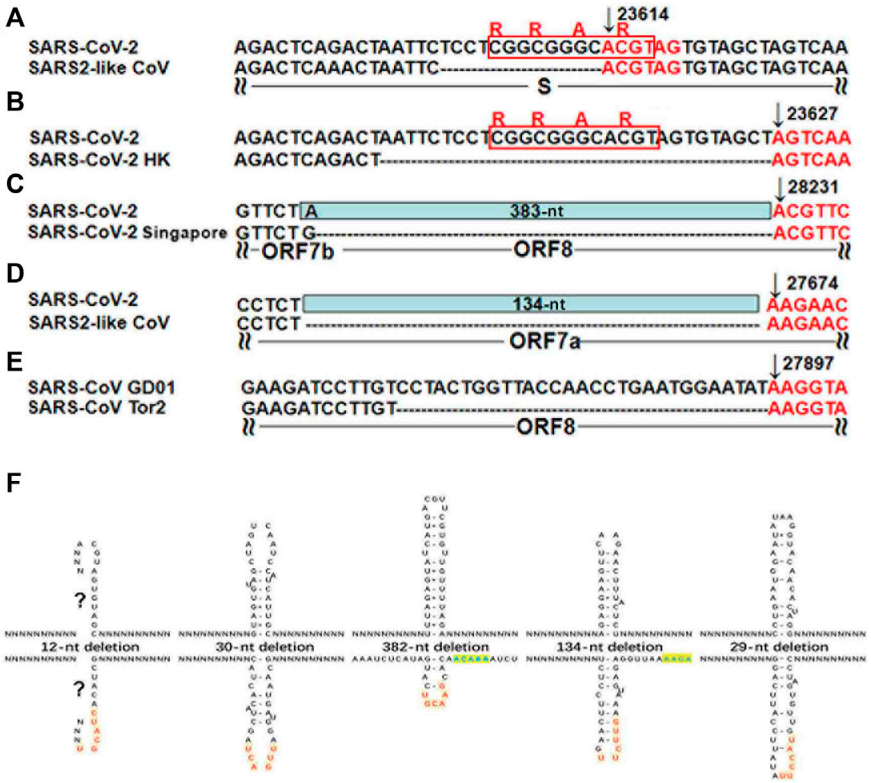


FIGURE 3 | TRS hairpins in five recombination regions. **(A-E)** have already been published in our previous study (Li et al., 2021a). N represents any nt. All the positions were annotated on the SARS-CoV (GenBank: AY278489) or SARS-CoV-2 (GenBank: MN908947) genomes. **(A)**. The genome (GenBank: MN996532) of the SARS2-like CoV strain RaTG13 from bats is used to show the 12-nt deletion; **(B)**. The genome (GISAID: EPI_ISL_417443) of the SARS-CoV-2 strain Hongkong is used to show the 30-nt deletion; **(C)**. The genomes (GISAID: EPI_ISL_414378, EPI_ISL_414379 and EPI_ISL_414380) of three SARS-CoV-2 strains from Singapore are used to show the 382-nt deletion; **(D)**. The genome (GenBank: MT457390) of the mink SARS2-like CoV strain is used to show the 134-nt deletion; **(E)**. The genome (GenBank: AY274119) of the SARS-CoV strain Tor2 is used to show the 29-nt deletion . **(F)**. These recombinant events occurred at the non-canonical NSP15 breakpoints that also end with at least an uridine residue ("U"), due to the cleavage of the non-canonical TRS hairpins.

can form. Thus, to investigate if a unique TRS hairpin can be formed, we further analyzed the association between the "AAGAA-like" and GTTCGT motifs in all possible TRS hairpins of the eight genes (**Supplementary Table S1**) using 1,265 CoV genome sequences (Materials and Methods), leading to discovery of the association between RNA methylation and NSP15 cleavage. Here, we illustrate how the association was discovered, using the *M* gene of SARS-CoV-2 as an example (**Figure 4**). The minimum free energies (MFEs) of three possible TRS hairpins in the *M* gene were estimated as -2.50, -4.00 and -4.90 kcal/mol (Materials and Methods). Although the third hairpin (**Figure 4C**) is the most stable one, the difference of MFEs between the second (**Figure 4B**) and third hairpins is marginal. The first (**Figure 4A**) and third hairpins require the "AAGAA-like" and AAACH (Detailed later) motifs involved in the base pairing, respectively. However, RNA methylation (*e.g.*, m6A) of these motifs is not in favour of base pairing in the first and third hairpins. Thus, only the second hairpin was able to form. We proposed that RNA methylation participates in the determination of the local RNA structures in CoVs by affecting the formation of base pairing. RNA methylation of sequences containing the "AAGAA-like" or AAACH motifs significantly reduces the possibility of formation of many hairpins, ensuring the formation of a unique TRS hairpin (**Figure 4B**) in all likelihood. In the unique TRS hairpin, the NSP15 cleavage site exposes in a small loop, which facilitates the contacts of NSP15, while the loop of the opposite TRS hairpin may not contain uridine residues for NSP15 cleavage. The structure of this small loop can be used to explain the results of mutation experiments in a previous study (Yount et al., 2006) that the recognition of a TRS (Actually revealed as the NSP15 cleavage site (Li et al., 2021a)) is independent on its motif, but dependent on its context. The TRS hairpin can be used to explain the discovery that the recognition of NSP15 cleavage sites is structure-based (TRS hairpin) rather than sequence-based (NSP15 cleavage motif). The above results indicated that TRS hairpins in nascent gRNAs(-) are indispensable for the functions of the RTC in jumping transcription (**Figure 4D**).

The following topic is which enzyme is responsible for the internal methylation of CoV RNAs, which is supposed to be done before the NSP15 cleavage for jumping transcription. A recent study reported that NSP14 (no structure data available) and NSP10&16 (PDB: 7BQ7), as N7 and 2′-O-MTase respectively (Introduction), are crucial for RNA cap

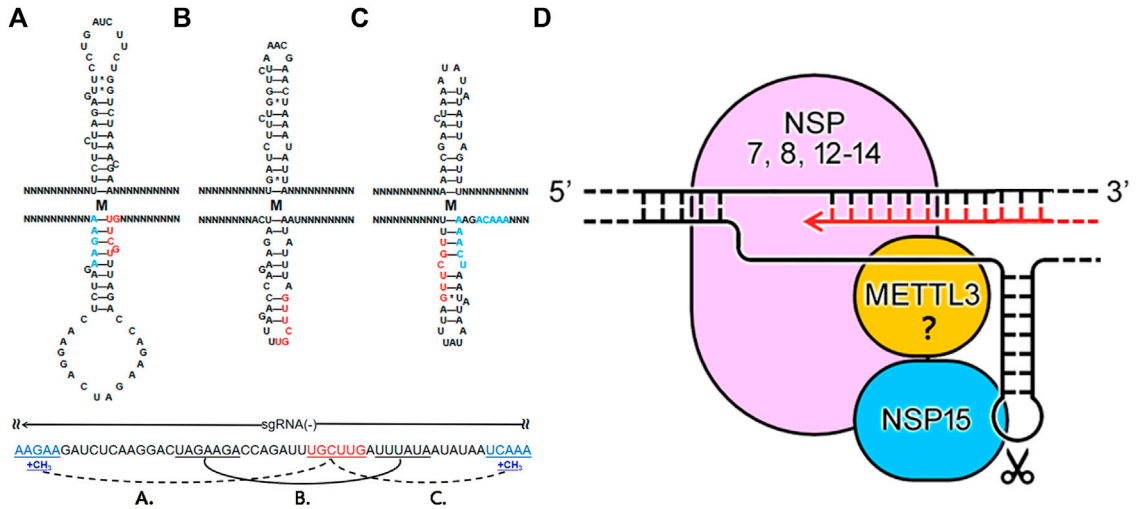


FIGURE 4 | How RTC functions in jumping transcription. N represents any nt. Using the *M* gene of SARS-CoV-2 as an example, the first **(A)** and third **(C)** hairpins require the "AAGAA-like" or AAACH motifs involved in the base pairing. RNA methylation of sequences containing the "AAGAA-like" or AAACH (in blue color) motifs is not in favour of base pairing, ensuring the formation of a unique TRS hairpin **(B)** containing a NSP15 cleavage site in the loop **(D)** 5′-3′ represents the strand of the SARS-CoV-2 genome. NSP12-14 form the main structure of the RTC; NSP7 and NSP8, acting as the cofactors of NSP12, may be also included in the main structure of the RTC (Yan et al., 2020); NSP15 and METTL3 are coupled with the main structure. The RTC processes the double-strand RNAs (dsRNAs) and single-strand RNAs (ssRNAs) in two situations. Nascent RNAs are synthesized in one route using unwound ssRNAs(+) or ssRNAs(-) as templates. In the other route, ssRNAs(-) can be uncleaved or cleaved for jumping transcription or degraded, which is regulated by a negative feedback mechanism (Li et al., 2021a). NSP15 cleaves a ssRNA in a small loop in the second route.

formation (Krafcikova et al., 2020). This suggested that NSP14 and NSP10&16 are unlikely to function in the internal methylation of CoV RNAs. Although the previous study excluded METTL3-mediated RNA (m6A) methylation for lack of the canonical motif RR**A**CH (R and H represent A/G and A/C/T, respectively) (Kim et al., 2020a), we still found many internal methylation sites containing the AAACH motif in the SARS-CoV-2 genome by reanalyzing the Nanopore RNA-seq data. Notable instances include "ag**T**tt" (AAACT on the antisense strand) at the positions 29408 and 29444 (corresponding to the underlined capital letter), and "tgTtt" at the position 29170. Particularly, "tgTtt", "cgTtt", "agTtt" and "tgTtt" located at the positions 25402, 26258, 26494 (**Figure 4C**) and 28235 co-occurred with the NSP15 cleavage motif of four genes (*ORF3a*, *E*, *M* and *N*). In addition, "tgTtt", "tgTtt", "ttctT" (the "AAGAA-like" motif on the antisense strand) and "tgTtt" were located at the positions 21566, 21570, 21577 and 21579 (**Supplementary Table S1**), which are closely linked and flanking the GTTCGT motif of the *S* gene, which merits investigation in the future. The above findings indicated that METTL3 functions in RNA (m6A) methylation of sequences containing the AAACH motif for *ORF3a*, *E*, *M* and *N*, and possibly the "AAGAA-like" motif for *S, ORF6, 7a* and *8*. Finally, we proposed the theoretical arrangement of NSP12-15 and METTL3 in the global RTC structure (**Figure 4D**) by the integration of information from many aspects, particularly including: 1) identification of NSP15 cleavage sites in our previous study (Li et al., 2021a); 2) discovery of the AAACH motif co-occurred with the NSP15 cleavage motif of four genes; 3) discovery of the association between RNA methylation and NSP15 cleavage; and 4) discovery of the TRS hairpins of eight genes (*S, E, M, N,* and *ORF3a, 6, 7a* and *8*).

By comprehensive analysis of the above results, we constructed a model to answer how the RTC functions in the jumping transcription of CoVs. In this model, the RTC processes double-strand RNAs (dsRNAs) and single-strand RNAs (ssRNAs) in two situations (**Figure 4D**), respectively. In the first situation, NSP13 unwinds dsRNAs (Yan et al., 2020) to produce ssRNAs(+) or ssRNAs(-), which are processed in two routes. In one route, NSP12 synthesizes RNAs with error correction by NSP14 to produce dsRNAs using unwound ssRNAs(+) or ssRNAs(-) as templates (Knoops et al., 2008). The other route processes ssRNAs(+) or ssRNAs(-), which can be methylated at internal sites and cleaved by NSP15 for jumping transcription. Then, the ssRNAs(+) and ssRNAs(-) are further processed in different ways: most ssRNAs(+) are uncleaved and packaged by the N protein (this is still not clear), while ssRNAs(-) can be uncleaved or cleaved for jumping transcription or degraded, which is regulated by a negative feedback mechanism (Li et al., 2021a). In the second situation, the RTC processes ssRNAs: uncleaved ssRNAs(+) and ssRNAs(-) are used as templates for replication; cleaved ssRNAs(-) are used as templates for transcription. The model can be used to explain the extremely high ratio between sense and antisense reads analyzed in our previous study (Li et al., 2021a) and the experimental result that knockdown of NSP15 by mutation increases the accumulation of viral dsRNA in another previous study (Deng et al., 2017). According to our model, knockdown of NSP15 increases the uncleaved gRNAs(-), which continue to be templates to produce more dsRNAs.

CONCLUSION AND DISCUSSION

In the present study, we proposed the theoretical arrangement of NSP12-15 and METTL3 in the global RTC structure and constructed a model to answer how the RTC functions in the jumping transcription of CoVs. More importantly, our results reveal the complex associations between RNA methylation, NSP15 cleavage, CoV replication and transcription at the molecular level. Our findings enrich fundamental knowledge in the field of gene expression and its regulation, providing a crucial basis for future studies. NSP12-14 form the main structure of the RTC; NSP7 and NSP8, acting as the cofactors of NSP12, may be also included in the main structure of the RTC (Yan et al., 2020); NSP15 and METTL3 are coupled with the main structure. The results of previous experiments suggest that NSP8 is able to interact with NSP15 (Lianqi et al., 2018). Future research needs to be conducted to determine the structures of NSP12&14, NSP12&15, NSP12&METTL3 and NSP15&METTL3 complexes by Cryo-EM. These local RTC structures can be used to assemble a global RTC structure by protein-protein docking calculation. Our model does not rule out the involvement of other proteins (*e.g.*, ORF8) in the global RTC structure or other proteins in the internal methylation of the "AAGAA-like" motif. Future drug design targeting SARS-CoV-2 needs to consider protein-protein and protein-RNA interactions in the RTC, particularly the structure of NSP15 and the TRS hairpin complex.

MATERIALS AND METHODS

The *Betacoronavirus* genus includes five subgenera (*Embecovirus*, *Sarbecovirus*, *Merbecovirus*, *Nobecovirus* and *Hibecovirus*), which were defined as subgroups A, B, C, D and E (Bei et al., 2022). In our previous study (Li et al., 2021b), 1,265 genome sequences of viruses in the *Embecovirus*, *Sarbecovirus*, *Merbecovirus*, *Nobecovirus* subgenera were downloaded from the NCBI Virus database (https://www.ncbi.nlm.nih.gov/labs/virus). Two genome sequences (RefSeq: NC_025217 and GenBank: KY352407) of viruses in the *Hibecovirus* subgenus were also downloaded. Among 1,265 genomes, 292 belongs to *Betacoronavirus* subgroup B (including SARS-CoV and SARS-CoV-2). 1,178, 480 and 194 genome sequences of viruses in the *Alphacoronavirus*, *Gammacoronavirus* and *Deltacoronavirus* genera were downloaded to validate the TRS motifs (**Figure 1**). Nanopore RNA-seq data was downloaded from the website (https://osf.io/8f6n9/files/) for reanalysis. Data cleaning and quality control were performed using Fastq_clean (Zhang et al., 2014). Statistics and plotting were conducted using the software R v2. 15.3 with the Bioconductor packages (Gao et al., 2014). Protein structure data (PDB: 6 × 1B, 7BQ7 and 7CXN) were used to analyze NSP15, NSP10&16 and NSP7&8&12&13, respectively. The structures of NSP12-16 were predicted using trRosetta (Yang et al., 2020). The minimum free energies (MFEs) of hairpins were estimated by RNAeval v2.4.17 with parameters by manual adjustment.

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

SG conceived the project. SG and DM supervised this study. JL, SC, FY and ZC downloaded, managed and processed the data. GD and JS performed programming. XL predicted and analyzed the protein structures. SG drafted the main manuscript text. SG and JR revised the manuscript.

FUNDING

This work was supported by the Yunnan Applied Basic Research—Yunnan Provincial Science and Technology Department—Kunming Medical University joint projects (202101AY070001-073), Shandong Province Natural Science Foundation (ZR2020QC071) to GD and National Natural Science Foundation of China (31900444) to ZC.

ACKNOWLEDGMENTS

We are grateful for the help from the following faculty members of College of Life Sciences at Nankai University: Qiang Zhao, Jia Chang, Jianyi Yang and Bingjun He and that from Jinlong Bei and Tung On Yau. We would like to thank Editage (www.editage.cn) for polishing part of this manuscript in English language. This manuscript was online as a preprint on 18 Feb 2021 at https://biorxiv.org/cgi/content/short/2021.02.17.431652v1.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.904513/full#supplementary-material

REFERENCES

Bei, J., Xu, G., Chang, J., Wang, X., Tung, O. Y., Ruan, Ji., et al. (2022). SARS-CoV-2 with Transcription Regulatory Sequence Motif Mutation Poses a Greater Threat. *J. South Med. Univ. (In Chin.* 42 (3), 399–404. doi:10.12122/j.issn.1673-4254.2022.03.12

Deng, X., Hackbart, M., Mettelman, R. C., O'Brien, A., Mielech, A. M., Yi, G., et al. (2017). Coronavirus Nonstructural Protein 15 Mediates Evasion of dsRNA Sensors and Limits Apoptosis in Macrophages. *Proc. Natl. Acad. Sci. U. S. A.* 114 (21), E4251–E4260. doi:10.1073/pnas.1618310114

Denison, M. R., Graham, R. L., Donaldson, E. F., Eckerle, L. D., and Baric, R. S. (2011). Coronaviruses. *RNA Biol.* 8 (2), 270–279. doi:10.4161/rna.8.2.15013

Duan, G., Shi, J., Xuan, Y., Chen, J., Liu, C., Ruan, J., et al. (2020). 5' UTR Barcode of the 2019 Novel Coronavirus Leads to Insights into its Virulence. *Chin. J. Virology (In Chinese)* 36 (3), 365–369. doi:10.13242/j.cnki.bingduxuebao.003681

Gao, S., Ou, J., and Xiao, K. (2014). *R Language and Bioconductor in Bioinformatics Applications*. Chinese Edition. Tianjin: Tianjin Science and Technology Translation Publishing Ltd.

Grossoehme, N. E., Li, L., Keane, S. C., Liu, P., Dann, C. E., Leibowitz, J. L., et al. (2009). Coronavirus N Protein N-Terminal Domain (NTD) Specifically Binds the Transcriptional Regulatory Sequence (TRS) and Melts TRS-cTRS RNA Duplexes. *Journal of molecular biology* 394, 544–557. doi:10.1016/j.jmb.2009.09.040

Hillen, H. S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D., and Cramer, P. (2020). Structure of Replicating SARS-CoV-2 Polymerase. *Nature* 584 (7819), 154–156. doi:10.1038/s41586-020-2368-8

Jiayuan, C., Jinsong, S., Yau Tung, O., Chang, L., Xin, L., Qiang, Z., et al. (2020). Bioinformatics Analysis of the 2019 Novel Coronavirus Genome. *Chinese Journal of Bioinformatics (In Chinese)* 18 (2), 96–102. doi:10.12113/202001007

Kim, D., Lee, J.-Y., Yang, J.-S., Kim, J. W., Kim, V. N., and Chang, H. (2020). The Architecture of SARS-CoV-2 Transcriptome. *Cell.* 181 (4), 914–921. doi:10.1016/j.cell.2020.04.011

Kim, Y., Jedrzejczak, R., Maltseva, N. I., Wilamowski, M., Endres, M., Godzik, A., et al. (2020). Crystal Structure of Nsp15 Endoribonuclease NendoU from SARS-CoV -2. *Protein Science* 29 (7), 1596–1605. doi:10.1002/pro.3873

Knoops, K., Kikkert, M., Worm, S. H. E. v. d., Zevenhoven-Dobbe, J. C., van der Meer, Y., Koster, A. J., et al. (2008). SARS-coronavirus Replication Is Supported by a Reticulovesicular Network of Modified Endoplasmic Reticulum. *PLoS Biol* 6 (9), e226. doi:10.1371/journal.pbio.0060226

Krafcikova, P., Silhan, J., Nencka, R., and Boura, E. (2020). Structural Analysis of the SARS-CoV-2 Methyltransferase Complex Involved in RNA Cap Creation Bound to Sinefungin. *Nat Commun* 11 (1), 3717. doi:10.1038/s41467-020-17495-9

Li, X., Duan, G., Zhang, W., Shi, J., Chen, J., Chen, S., et al. (2020). A Furin Cleavage Site Was Discovered in the S Protein of the 2019 Novel Coronavirus. *Chinese Journal of Bioinformatics (In Chinese)* 18 (2), 103–108. doi:10.12113/202002001

Li, X., Chang, J., Chen, S., Wang, L., Tung, O. Y., Zhao, Q., et al. (2021). Genomic Feature Analysis of Betacoronavirus Provides Insights into SARS and COVID-19 Pandemics. *Frontiers Microbiology* 10, 1–11. doi:10.3389/fmicb.2021.614494

Li, X., Cheng, Z., Wang, F., Chang, Jia., Zhao, Q., Zhou, H., et al. (2021). A Negative Feedback Model to Explain Regulation of SARS-CoV-2 Replication and Transcription. *Frontiers in Genetics* 10, 1–11. doi:10.3389/fgene.2021.641445

Lianqi, Z., Lei, L., Liming, Y., Zhenhua, M., Zhihui, J., Zhiyong, L., et al. (2018). Structural and Biochemical Characterization of Endoribonuclease Nsp15 Encoded by Middle East Respiratory Syndrome Coronavirus. *Journal of Virology* 92(22):e00893-18. doi:10.1128/JVI.00893-18

Liu, C., Chen, Z., Hu, Y., Ji, H., Yu, D., Shen, W., et al. (2018). Complemented Palindromic Small RNAs First Discovered from SARS Coronavirus. *Genes. (Basel)* 9 (9), 1–11. doi:10.3390/genes9090442

Mihindukulasuriya, K. A., Wu, G., St. Leger, J., Nordhausen, R. W., and Wang, D. (2008). Identification of a Novel Coronavirus from a Beluga Whale by Using a Panviral Microarray. *J Virol* 82, 5084–5088. doi:10.1128/jvi.02722-07

Papineau, A., Berhane, Y., Wylie, T. N., Wylie, K. M., Sharpe, S., and Lung, O. (2019). Genome Organization of Canada Goose Coronavirus, A Novel Species Identified in a Mass Die-Off of Canada Geese. *Sci Rep* 9, 5954. doi:10.1038/s41598-019-42355-y

Sawicki, S. G., and Sawicki, D. L. (1998). "A New Model for Coronavirus Transcription," in *Coronaviruses and Arteriviruses*. Editors L. Enjuanes, S. G. Siddell, and W. Spaan (Boston, MA: Springer US), 215–219. doi:10.1007/978-1-4615-5331-1_26

Silva, S. J. R., Alves da Silva, C. T., Mendes, R. P. G., and Pena, L. (2020). Role of Nonstructural Proteins in the Pathogenesis of SARS-CoV-2. *J Med Virol* 92, 1427–1429. doi:10.1002/jmv.25858

Sola, I., Moreno, J. L., Zúñiga, S., Alonso, S., and Enjuanes, L. (2005). Role of Nucleotides Immediately Flanking the Transcription-Regulating Sequence Core in Coronavirus Subgenomic mRNA Synthesis. *J Virol* 79, 2506–2516. doi:10.1128/JVI.79.4.2506-2516.2005

Xu, X., Ji, H., Jin, X., Cheng, Z., Yao, X., Liu, Y., et al. (2019). Using pan RNA-Seq Analysis to Reveal the Ubiquitous Existence of 5' and 3' End Small RNAs. *Front Genet* 10, 105–111. doi:10.3389/fgene.2019.00105

Yan, L., Zhang, Y., Ge, J., Zheng, L., Gao, Y., Wang, T., et al. (2020). Architecture of a SARS-CoV-2 Mini Replication and Transcription Complex. *Nat Commun* 11 (2020), 5874–5876. doi:10.1038/s41467-020-19770-1

Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., and Baker, D. (2020). Improved Protein Structure Prediction Using Predicted Interresidue Orientations. *Proc. Natl. Acad. Sci. U.S.A.* 117 (3), 1496–1503. doi:10.1073/pnas.1914677117

Yount, B., Roberts, R. S., Lindesmith, L., and Baric, R. S. (2006). Rewiring the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) Transcription Circuit: Engineering a Recombination-Resistant Genome. *Proc. Natl. Acad. Sci. U.S.A.* 103 (33), 12546–12551. doi:10.1073/pnas.0605438103

Zhang, M., Zhan, F., Sun, H., Gong, X., Fei, Z., and Gao, S. (2014). "Fastq_clean: An Optimized Pipeline to Clean the Illumina Sequencing Data with Quality Control," in Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Belfast, UK, 2-5 Nov. 2014, 44–48.doi:10.1109/BIBM.2014.6999309

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 Genetics

ORIGINAL RESEARCH
published: 06 June 2022
doi: 10.3389/fgene.2022.921925

Rare Variants in Novel Candidate Genes Associated With Nonsyndromic Patent Ductus Arteriosus Identified With Whole-Exome Sequencing

*Ying Gao[1†], Dan Wu[1†], Bo Chen[2], Yinghui Chen[3], Qi Zhang[3] and Pengjun Zhao[3]**

[1]Department of Pediatric, Shidong Hospital, Shanghai, China, [2]Department of Cardiothoracic Surgery, School of Medicine, Heart Center, Shanghai Children's Medical Center, School of Medicine, Shanghai Jiao Tong University, Shanghai, China, [3]Department of Pediatric Cardiology, Xin Hua Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

Background: Patent ductus arteriosus (PDA) is one of the most common congenital heart defects causing pulmonary hypertension, infective endocarditis, and even death. The important role of genetics in determining spontaneous ductal closure has been well-established. However, as many of the identified variants are rare, thorough identification of the associated genetic factors is necessary to further explore the genetic etiology of PDA.

Methods: We performed whole-exome sequencing (WES) on 39 isolated nonsyndromic PDA patients and 100 healthy controls. Rare variants and novel genes were identified through bioinformatic filtering strategies. The expression patterns of candidate genes were explored in human embryo heart samples.

Results: Eighteen rare damaging variants of six novel PDA-associated genes (*SOX8*, *NES*, *CDH2*, *ANK3*, *EIF4G1*, and *HIPK1*) were newly identified, which were highly expressed in human embryo hearts.

Conclusions: WES is an efficient diagnostic tool for exploring the genetic pathogenesis of PDA. These findings contribute new insights into the molecular basis of PDA and may inform further studies on genetic risk factors for congenital heart defects.

Keywords: congenital heart defects, patent ductus arteriosus, whole-exome sequencing, rare variants, single-nucleotide polymorphism

OPEN ACCESS

Edited by:
Tao Huang,
Shanghai Institute of Nutrition and Health (CAS), China

Reviewed by:
Liping Liu,
Hunan Province People's Hospital, China
Jie Huang,
Children's Hospital of Soochow University, China

****Correspondence:***
Pengjun Zhao
Pjunzhao@sina.com

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

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *16 April 2022*
Accepted: *09 May 2022*
Published: *06 June 2022*

Citation:
Gao Y, Wu D, Chen B, Chen Y, Zhang Q and Zhao P (2022) Rare Variants in Novel Candidate Genes Associated With Nonsyndromic Patent Ductus Arteriosus Identified With Whole-Exome Sequencing. Front. Genet. 13:921925. doi: 10.3389/fgene.2022.921925

INTRODUCTION

The ductus arteriosus (DA) is a normal fetal structure that connects the pulmonary artery and descending aorta to maintain blood circulation during the fetal period (Benitz, W. E. et al., 2016). From the perspective of cardiac development, the DA functionally shuts down 15 h after birth in healthy, full-term infants (Crockett, S. L. et al., 2019). This process involves abrupt contraction of the muscular wall of the DA, which is associated with a proper balance among neurohumoral factors. An increase in the levels of contractile elements, such as peroxidase O2 and endothelin-1, and decrease in the levels of relaxants, such as prostaglandin E2 and nitric oxide, are the main events causing closure of the DA (Crockett, S. L. et al., 2019). Neural crest-derived cells migrate into the subendothelial space under the action of these hormones and transform into vascular smooth muscle cells (VSMCs). With contraction of the medial membrane and circular muscle in the DA, the lumen is shortened and

finally closed (Li, N. et al., 2016). However, the maintenance of DA patency after birth has a pathological effect (Benitz, W. E. et al., 2016).

Failure of the DA to close after birth is termed patent DA (PDA), which is one of the most common heart defects, affecting approximately 1 in 2000 full-term infants and 8 in 1000 premature infants (Hoffman, J. I. et al., 2002). Persistent ductal shunting may lead to pulmonary overcirculation and induce systemic hypoperfusion, thereby increasing the risk of pulmonary hypertension, infective endocarditis, heart failure, and even death (Mitra, S. et al., 2018). However, its etiology and pathogenesis remain unclear.

PDA has both inherited and acquired causes. The preliminary understanding of the genetic mechanism of PDA was based on the studies in patients with syndrome. Previous studies have confirmed the association of several chromosomal syndromes, including Turner (45, XO), Kartagener, and Klinefelter (47, XXY), with PDA (Groth, K. A. et al., 2013; Gravholt, C. H. et al., 2019; Yang, D. et al., 2019). In addition to chromosomal rearrangements, a single gene mutation can also cause syndromic PDA, including Noonan (*PTPN11* mutation), Holt-Oram (*TBX5* mutation), and char (*TFAP2B* mutation) syndrome (Satoda, M. et al., 2000; Pannone, L. et al., 2017; Vanlerberghe, C. et al., 2019). However, the genetic mechanism of nonsyndromic PDA (isolated findings without other abnormalities) remains unclear. Rare damaging mutations in *MYH11* and *TFAP2B* were detected in several isolated nonsyndromic PDA patients (Harakalova, M. et al., 2013). Erdogan et al. (Erdogan, F. et al., 2008) performed an array comparative genome hybridization analysis of 105 patients with congenital heart defects and identified a 1.92 Mb deletion of chromosome 1q21.1 (CJA5) in a PDA patient. Genetic determinants of nonsyndromic PDA is still unknown.

Therefore, in this study, we recruited 39 unrelated nonsyndromic PDA patients and 100 healthy children for WES. Using a series of bioinformatics filtering steps, we identified 18 rare damaging variants in six candidate PDA-associated genes (*SOX8*, *NES*, *CDH2*, *ANK3*, *EIF4G1*, and *HIPK1*). Notably, these candidate genes were also highly expressed in human embryonic hearts. This identification of new pathogenic genes could help to elucidate the detailed underlying mechanism of PDA and promote further experimental analyses.

MATERIAL AND METHODS

Patients and Consent

Thirty-nine isolated nonsyndromic PDA patients of Han Chinese ethnicity and 100 healthy children (aged between 2 months and 13 years) were recruited from Xinhua Hospital affiliated with Shanghai Jiao Tong University (Shanghai, China). The structural heart phenotypes of all participants were assessed using echocardiography or cardiac catheterization. A diagnosis of PDA was made in the patient group by cardiac catheterization or surgery. Patients with a history of complex congenital heart disease were excluded from the study. The study protocol and ethics were approved by the Medical Ethics Committee of Xinhua Hospital. Informed consent was obtained from the parents of all participants. The study was conducted in accordance with the Declaration of Helsinki and the International Ethical Guidelines for Health-Related Research Involving Humans.

DNA Extraction and Whole-Exome Sequencing

The genomic DNA of all participants was extracted from blood samples using QIAamp DNA Blood Mini Kit (QIAGEN, Germany). DNA samples were stored at –80°C until further use. Genomic DNA was eluted, purified, amplified by ligation-mediated polymerase chain reaction, and then subjected to DNA sequencing on an Illumina platform. The target depth of the DNA sequencing was x100. Qualified DNA samples from the PDA and control groups were subjected to WES to detect rare variations. Read quality was checked using Fastp software(Chen, S. et al., 2018) and raw sequence data were aligned to human genome (human_glk_v37) using BWA (v0.7.12-r1039). Duplicated and low-quality reads (Per base sequence quality <20) were removed by using Picard software (https://broadinstitute.github.io/picard). Alignment quality was assessed using qualimap software (Okonechnikov, K. et al., 2016).

Single-Nucleotide Polymorphism Identification and Quality Filtering

Single nucleotide polymorphisms (SNPs) account for much of the phenotypic diversity among individuals.

SNPs and insertions/deletions were detected using the HaplotypeCaller module of GATK4 software (Mckenna, A. et al., 2010), based on sequence alignment of the clinical samples to the reference genome. Before detection, we recalibrated the base qualities using the BaseRecalibrator module of GATK4 software (Mckenna, A. et al., 2010) to improve variant detection accuracy based on the quality with a depth (QD) criterion >2. The resulting BAM files were then sorted, indexed, and processed using base quality score recalibration (Okonechnikov, K. et al., 2016). The GATK HaplotypeCaller module was then used for variant calling. We used ANNOVAR53 (Wang, K. et al., 2010) to annotate the variants for functional and population frequency information with the 1000 Genomes (Clarke, L. et al., 2012), Refseq (O'leary, N. A. et al., 2016), ExAC (Karczewski, K. J. et al., 2017), ESP6500 (Liang, Y. et al., 2019), gnomAD, SIFT (Flanagan, S. E. et al., 2010), clinvar (Landrum, M. J. et al., 2020), PolyPhen (Flanagan, S. E. et al., 2010), MutationTaster (Steinhaus, R. et al., 2021), COSMIC (Forbes, S. A. et al., 2011), gwasCatalog, and OMIM databases (Amberger, J. S. et al., 2017). All potentially damaging variants of the candidate genes were classified into five groups: pathogenic, likely pathogenic, variant of uncertain significance, likely benign, and benign (Richards, S. et al., 2015). Finally, the rare damaging variants were filtered according to the American College of Medical Genetics criteria guidelines (**Figure 1**).

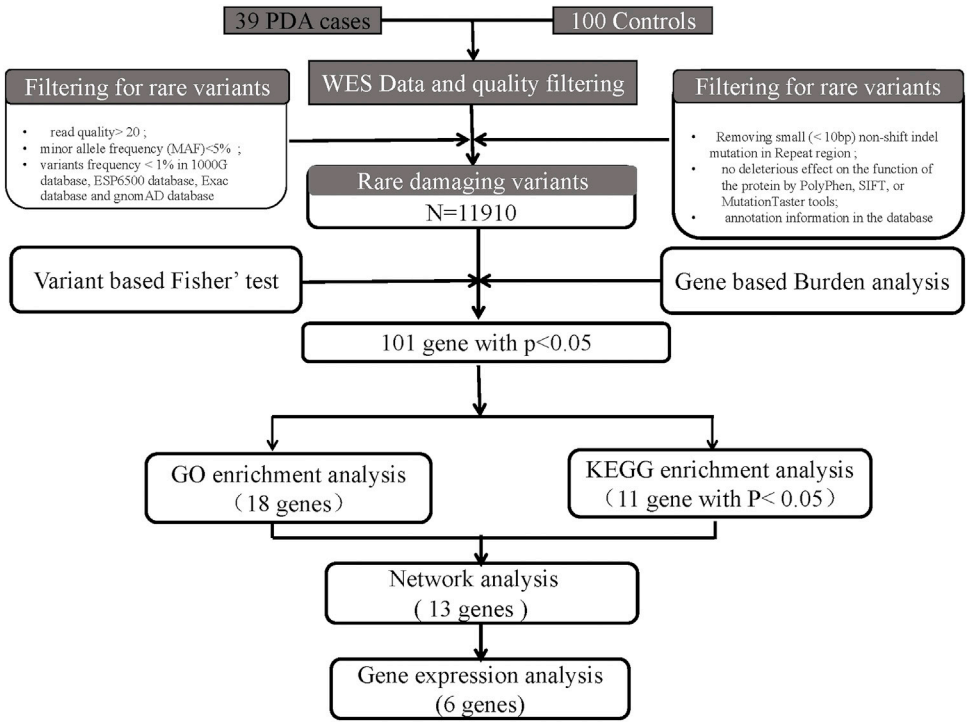


FIGURE 1 | Bioinformatics filtering strategy workflow for the candidate genes. Through a series of filtering methods, we finally identified 6 candidate genes. The potentially damaging variants in candidate genes were subjected to validation *via* human embryonic heart expression analysis.

Variant Filtering Based on Fisher's Exact Test and Burden Analysis

The difference in allele frequency for each SNP between cases and controls was compared using the Fisher's exact test with R statistical software packages; a p-value < 0.05 was considered statistically significant. Subsequently, we aggregated the SNP data based on gene expression levels and conducted a gene-based burden analysis to increase statistical power. Candidate pathogenic genes were filtered based on the results of burden analysis according to the following criteria: 1) p-value or false-discovery rate (FDR) < 0.05, 2) hit for at least one variant in three cases, and 3) not found in any sample of the control group. We then prioritized genes based on the p value of Fisher's exact test and burden analysis.

Functional Enrichment and Network Analysis

To further filter the candidate genes associated with PDA, we performed functional enrichment analysis to identify the functions of candidate genes identified through the aforementioned filtering steps. Pathway analysis of the candidate gene profiling results was performed using Gene Ontology (GO; version 30.10.2017) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway (http://www.genome.jp/kegg/pathway.html) mapping within the web-based tool Database for Annotation, Visualization, and Integrated Discovery (Gene Ontology, C. 2015; Kanehisa, M. et al., 2017). GO terms represent a network of biological processes that overlap in space and are clustered according to their relationships (Gene Ontology, C. 2015). The threshold was set to an adjusted p-value < 0.05. In addition, we prioritized these genes based on functional enrichment analysis. Furthermore, to detect the relationship between the candidate genes and known disease-causing genes, we constructed the protein–protein interaction (PPI) network (Brohee, S. et al., 2008) using Cytoscape software based on the STRING database.

Tissue Collection and Expression Detection

In addition to the genes prioritized using the steps described above, we further prioritized genes according to their expression levels in the human embryonic heart. Previous studies have divided eight embryonic weeks (56 days) into 23 internationally accepted Carnegie stages (O'rahilly, R. 1987). To further investigate the potential function of our candidate genes, human embryonic hearts in different Carnegie stages (S10–S16) were collected after medical termination of pregnancy from patients at Xinhua Hospital. RNA was extracted and purified using the Experion automated gel electrophoresis system and RNeasy MinElute Cleanup Kit. The expression patterns of candidate genes were subsequently detected using the Affymetrix HTA 2.0 microarray.

RESULTS

Population

Among the 39 patients, 28% had common cardiac defects, including atrial septal defect (n = 7), ventricular septal defect

TABLE 1 | Characteristics of 39 PDA patients.

Patients characteristics	Numbers
Age (year)	2.92 ± 2.44
Male-to-female ration (%)	62%
BMI (kg/m2)	16.58 ± 4.34
PDA size (mm)	2.87 ± 1.68
Birth weight (kg)	2.96 ± 0.73
Gestational age (week)	39.04 ± 1.46
Associated cardiac defect n (%)	
VSD n (%)	2 (5%)
ASD n (%)	7 (18%)
Others n (%)	2 (5%)

All values are expressed as mean ± SD or n (%)

($n = 2$), and others ($n = 2$) (**Table 1**). All subjects were born at full term, and no other major cardiac structural abnormalities or developmental syndromes were identified. WES, with an average depth of coverage of approximately x105 per base, identified 411,344 single-nucleotide variants and 23,101 insertions/deletions across the genome. Through a series of filtering strategies (see **Figure 1**), rare damaging variants were screened with a threshold of 0.5% minor allele frequency. As illustrated in **Figure 2**, we found more rare damaging variants in the PDA group than in the control group, including splice-site, nonsense, and missense mutations. Consistently, the C > T and G > A substitutions accounted for the majority of single-base mutations compared with other types (**Figure 2**). Based on these mutations, we adopted a bioinformatics filtering strategy to identify candidate genes associated with PDA.

Variants Identified Based on Fisher's Exact Test

Based on the results of Fisher's exact test, we identified 44 variants that were more frequently detected in the PDA group than in the control group (FDR <0.05, $p < 0.05$), as presented in **Table 2** ($p < 0.01$). We then prioritized these variants based on the p-value; the top 10 variants with statistical significance are shown in **Figure 3**. Notably, we found that the SNPs rs103826685 and rs32552095 located in *SLC9B1* and *HLA-DRB1*, respectively, had the most significantly different frequencies between the patient and control groups ($p < 0.0001$).

Candidate Genes Identified Based on Burden Analysis

To further increase statistical power, we aggregated the SNP data at the gene level and performed burden analysis. Under a significance threshold of 0.05, we observed 57 genes with potential pathogenicity as candidate PDA-associated genes (**Table 3** ($p < 0.01$)). We then prioritized these genes based

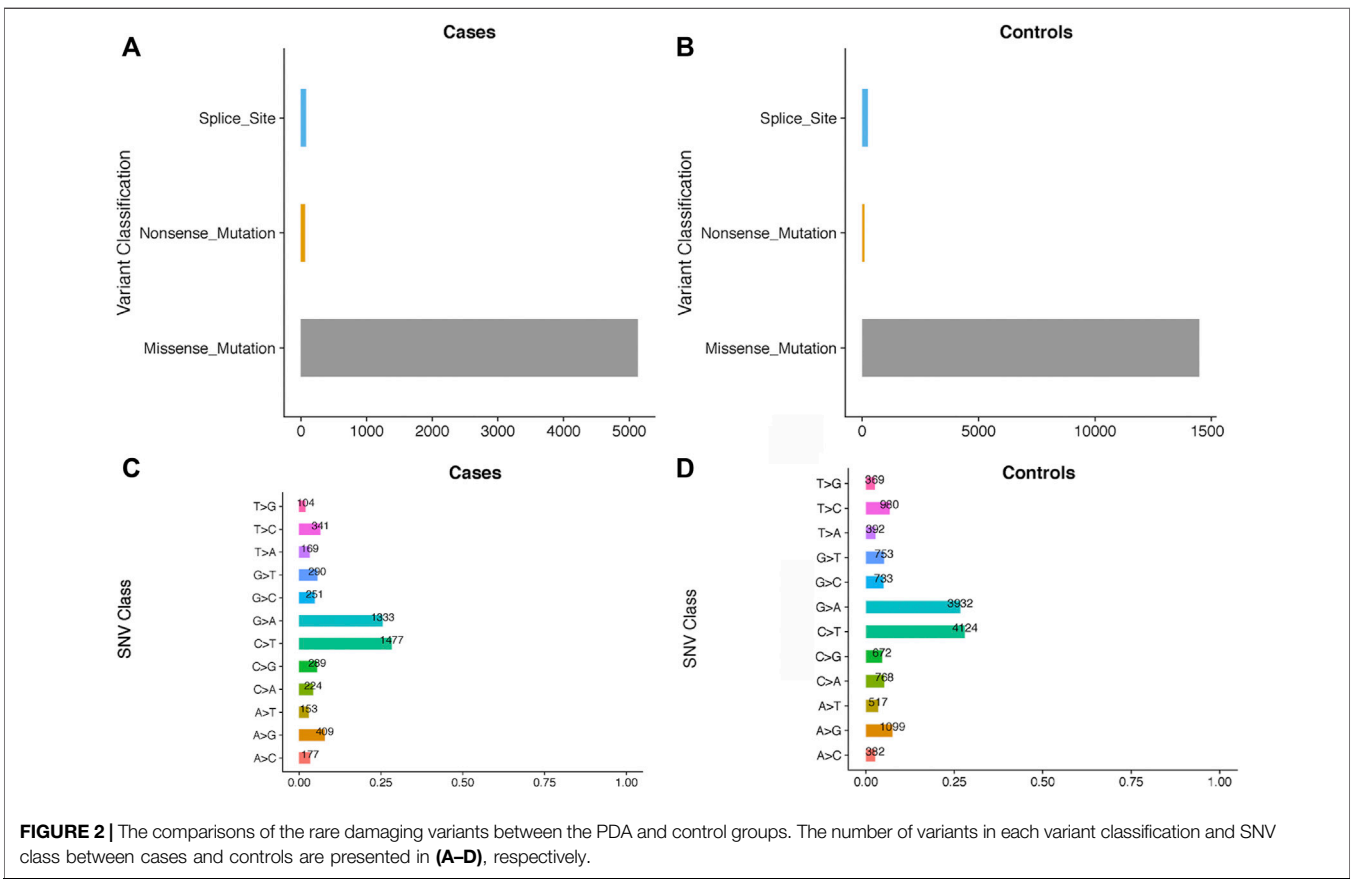


FIGURE 2 | The comparisons of the rare damaging variants between the PDA and control groups. The number of variants in each variant classification and SNV class between cases and controls are presented in **(A–D)**, respectively.

TABLE 2 | SNP filtering Based on Fisher Exact Test.

Chromosome	Gene	Mutation position	Mutation type	*p*-value
1	LRRC8C	90179703	T > G	0.006
1	NES	156640657	A > C	0.000
11	LRRC4C	40136434	C > T	0.006
14	SLC7A8	23612372	T > G	0.001
16	SOX8	1034733	A > C	0.006
16	NPIPA1	15045634	T > C	0.000
16	NPIPB5	22545658	A > C	0.000
16	NPIPB5	22546505	G > T	0.000
16	NPIPB5	22546506	A > C	0.000
19	MAP3K10	40719910	C > G	0.006
3	ZNF717	75786264	G > T	0.001
3	EIF4G1	184033621	G > C	0.000
3	MUC4	195506722	G > A	0.001
3	MUC4	195506723	T > G	0.001
3	MUC4	195514174	G > T	0.000
4	USP17L20	9217567	C > A	0.006
4	USP17L17	9246041	C > A	0.006
4	SLC9B1	103826685	T > A	0.000
4	LRBA	151770608	A > C	0.006
6	VARS	31746821	G > A	0.006
6	HLA-DRB5	32487344	T > C	0.000
6	HLA-DRB1	32552095	C > T	0.000
7	TCAF2	143400090	G > A	0.001

on the *p*-value from the burden analysis; the top 10 genes with statistical significance are displayed as a heatmap in **Figure 4**. The top three genes with high confidence were *NPIPB5*, *SLC9B1*, and *HLA-DRB1*. Notably, *SLC9B1* and *HLA-DRB1* were also in the top significant genes based on Fisher's exact test.

Functional Analysis

Functional enrichment analysis of the 101 candidate differentially expressed genes identified through Fisher's exact test and burden analysis revealed that the main enriched GO terms in the upregulated gene set were thiol-dependent ubiquitinyl hydrolase activity (TermID: GO:0036459), peptide antigen binding (TermID: GO:0042605), and ubiquitin-dependent protein catabolic process (TermID: GO:0006511). Particular focus was placed on terms representing prostaglandin, apoptosis, and heart development (**Figure 5**). Moreover, KEGG analysis of the direct gene targets in PDA patients revealed enrichment in pathways related to cell adhesion molecules (TermID: path: hsa04514, $p < 0.001$), viral myocarditis (TermID: path: hsa05416, $p = 0.0035$), and asthma (TermID: path: hsa05310, $p = 0.01$; **Figure 6**). Based on functional enrichment analysis, 29 pathway genes related to cardiovascular development were screened.

Network Analysis

To further explore their roles, the 29 candidate genes were mapped to construct a PPI network along with 240 known pathogenic genes involved in cardiovascular development (**Supplementary Table S1**). The 240 known genes from the literature were divided into two groups related to cardiovascular development and PDA, respectively. In the network, the candidate genes *NES* and *CDH2* showed the most direct and strongest relationship with known pathogenic genes in both groups. Moreover, CDH2 and NES had the highest molecular weights and were located at the center of the PPI network (**Figures 7**, **8**). Therefore, based on the degree of correlation, we screened out 11 candidate genes for final verification.

Detection of Candidate Gene Expression in the Human Embryonic Heart

To further investigate the potential function of our candidate genes, we detected the expression levels of the 11 screened out genes in human embryonic hearts at different Carnegie stages. After prioritizing the candidate genes based on expression levels, the final six pathogenic genes (*SOX8*, *NES*, *CDH2*, *ANK3*, *EIF4G1*, and *HIPK1*) were identified (**Figure 9**). Among them, *CDH2* was the most highly expressed in the embryonic heart (**Figure 10**).

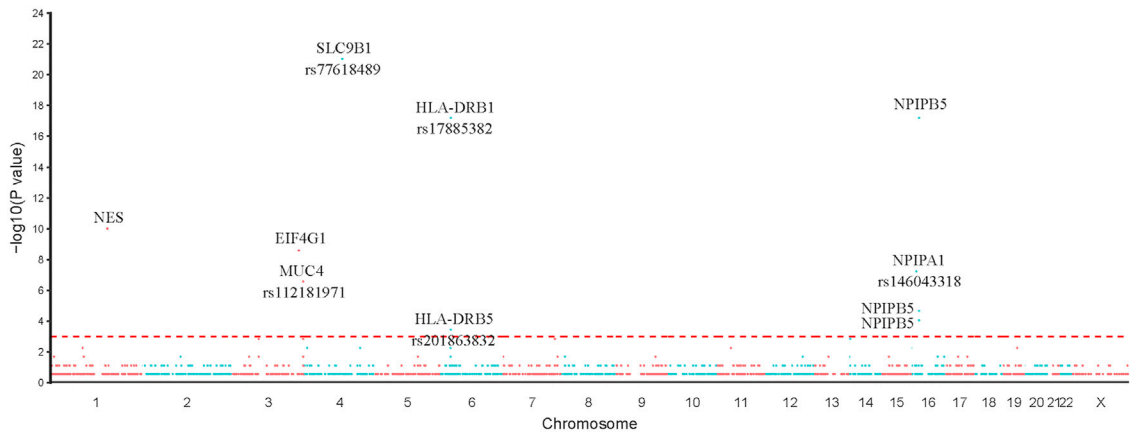


FIGURE 3 | Single SNP allele frequency and genotype frequency *p*-values were obtained using the fisher exact test. X-axis represents the position of each snp (represented in circles) on human chromosome, Y-axis is the–log *p*-value of Fisher Exact test. Top 10 variants in our study were represented in the figure.

TABLE 3 | Gene filtering based on Burden analysis.

Gene	Case mutation	Case normal	Control mutation	Control normal	*p*-value
ASIC3	5	34	0	100	0.001
CFAP45	4	35	0	100	0.006
CYP21A2	4	35	0	100	0.006
EVI5	4	35	0	100	0.006
HIPK1	4	35	0	100	0.006
HLA-DRB1	25	14	0	100	0.000
HLA-DRB5	6	33	0	100	0.000
KRT39	5	34	0	100	0.001
LRRC4C	4	35	0	100	0.006
MAP3K10	4	35	0	100	0.006
NPIPA1	13	26	0	100	0.000
NPIPB5	31	8	0	100	0.000
POTEE	5	34	0	100	0.001
SLC9B1	29	10	0	100	0.000
SLX4	4	35	0	100	0.006
SOX8	4	35	0	100	0.006
TBC1D3F	4	35	0	100	0.006
TCAF2	5	34	0	100	0.001
USP17L11	5	34	0	100	0.001
USP17L17	6	33	0	100	0.000
USP17L18	5	34	0	100	0.001
USP17L2	6	33	0	100	0.000
USP17L20	5	34	0	100	0.001
VARS	4	35	0	100	0.006
ZNF717	6	33	0	100	0.000

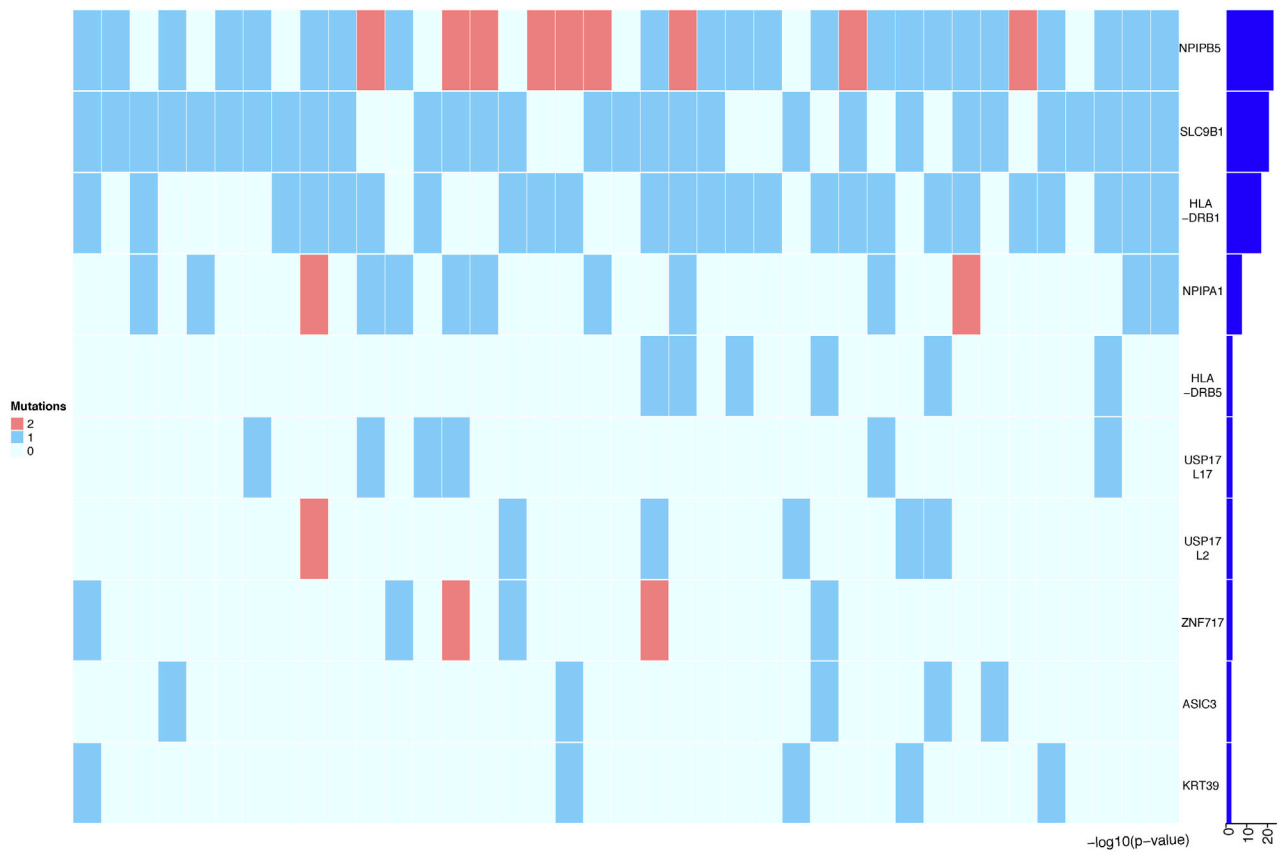


FIGURE 4 | Heatmap representing the top 10 genes identified in Burden analysis. Heatmap that shows the mutational burden (*p*-value< 0.05) of the top 10 gene based on gene-based burden analysis in PDA patients. The heatmap was generated by using R package, the mutation values were normalized per gene over all PDA samples. Each box in the heatmap represent a single variant in a case, with the dark red indicating high gene mutation ration in gene-based Burden analysis.

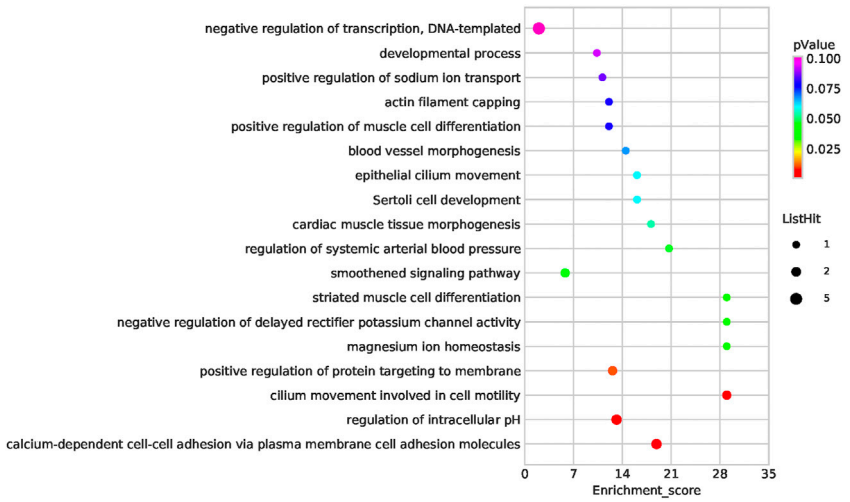


FIGURE 5 | Bubble plot of the GO analysis. Bubble plot summarizing enrichment for the most significant biological process GO terms associated to differentially expressed genes. The bubble size indicates the frequency of the GO term, while the color indicates the *p*-value.

DISCUSSION

The underlying molecular genetic mechanisms of PDA remain largely unknown as one of the most common congenital heart defects. In this study, we explored the clinical characteristics of 39 PDA patients and 100 healthy controls by performing WES to identify rare variants and candidate PDA-associated genes. Through a series of bioinformatic filtering strategies, we prioritized the candidate genes *via* Fisher's exact test, mutation burden analysis, gene network construction, and expression levels in embryonic hearts. Finally, we identified 18 rare damaging variants in six novel candidate genes (*SOX8*, *NES*, *CDH2*, *ANK3*, *EIF4G1*, and *HIPK1*) associated with PDA. Among these, *CDH2* was highly expressed in the human embryonic heart and appears to be the most important candidate gene identified in our study.

CDH2 encodes N-cadherin, a member of a protein family regulating cadherin-mediated cell–cell adhesion in multiple tissues. The structure comprises a single transmembrane domain, cytoplasmic domain, and five conserved extracellular cadherin domains (ECI–V) (Alimperti, S. et al., 2015). We found two variants (rs25565020 and rs25532304) in *CDH2* in four patients with PDA. In addition, CDH2 had the highest molecular weight and was located at the center of the PPI network, both among known CHD- and PDA-related genes. Further investigation showed that *CDH2* is highly expressed in human embryonic hearts. Previous studies in mice have also noted the importance of CDH2 in the proper development of the heart, brain, and skeletal structures (Radice, G. L. et al., 1997). Moreover, genetic analyses in zebrafish revealed that mutation in the EC-I or EC-IV domains of *cdh2* play important role in embryonic development (Masai, I. et al., 2003). Mayosi, B. M. et al. (2017) used WES to detect novel rare variants in patients with arrhythmogenic cardiomyopathy and found that *CDH2* mutation changes the conserved amino acids of CDH2 protein.

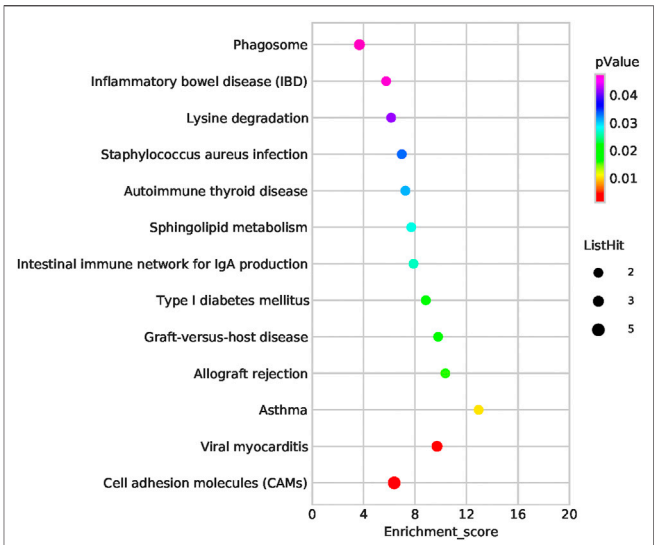


FIGURE 6 | Bubble plot of the KEGG pathway analysis. The representative enriched pathways shown by KEGG analysis. The bubble size indicates the frequency of the KEGG term, while the color indicates the *p*-value.

Since the relationship between *CDH2* and PDA is unclear, additional studies are needed to determine how genetic perturbations of *CDH2* contribute to PDA.

In our study, 16 patients (42%) had the same variant (rs156646936) in *NES*. In the network analysis, we observed a strong correlation between *NES* and known pathogenic genes. NES belongs to the human tissue kallikrein family of secreted serine proteases (Luo, L. et al., 1998), which play an important role in carcinogenesis, including in breast, prostate, and testicular cancers, and leukemia (Luo, L. Y. et al., 2001). Further experimental evidence suggests that the function of *NES* as a tumor suppressor

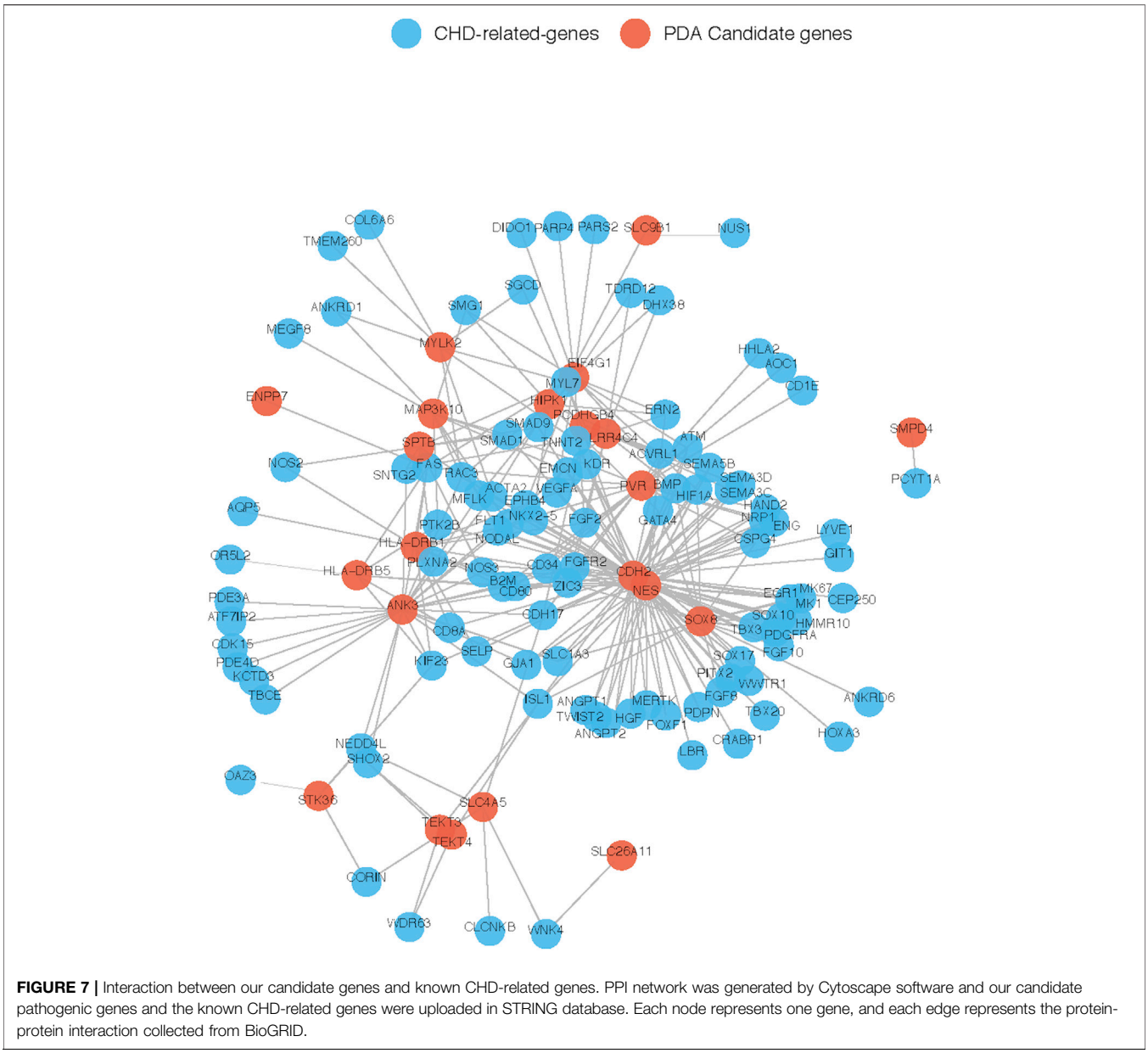


FIGURE 7 | Interaction between our candidate genes and known CHD-related genes. PPI network was generated by Cytoscape software and our candidate pathogenic genes and the known CHD-related genes were uploaded in STRING database. Each node represents one gene, and each edge represents the protein-protein interaction collected from BioGRID.

may be achieved by hypermethylation of the CpG islands (Li, B. et al., 2001). However, this is the first report of *NES* mutations in PDA. ANK3 is a member of the ankyrin family, which is expressed in several different isoforms in many tissues. ANK3 plays key roles in cell motility, activation, proliferation, contact, and the maintenance of specialized membrane domains. In our study, eight patients (10%) had variants in *ANK3*. *ANK3* variants have previously been associated with schizophrenia, autism, epilepsy, and intellectual disability (Leussis, M. P. et al., 2013; Wirgenes, K. V. et al., 2014). Studies from knockout mouse models have revealed that loss of ANK3 function leads to defects in cardiac calcium handling and arrhythmias (Mohler, P. J. et al., 2004). Although the roles of *NES* and *ANK3* in the pathogenesis of PDA are supported by bioinformatic analyses, our study was limited by the lack of experimental evidence to validate the deleteriousness of the variants.

EIF4G1 encodes a protein, that is, a component of the multi-subunit protein complex EIF4F. EIF4G plays a crucial role in translation initiation and serves as a scaffolding protein that binds several initiation factors (the cap-binding protein eIF4E, the RNA helicase eIF4A, and eIF3) (Haimov, O. et al., 2018). In our study, 15 patients had three types of variants in *EIF4G1*, and the same variant (rs184033621) was detected in 14 patients. EIF4G1 modulates the proliferation, apoptosis, and angiogenesis of most tumor types by limiting steps during the initiation phase of protein synthesis and interacting with ubiquitin-specific protease 10 (USP10) (Cao, Y. et al., 2016). Moreover, EIF4G1 phosphorylation specifically activates the PKC-Ras-ERK

FIGURE 8 | Interaction between our candidate genes and known PDA-related genes. PPI network was generated by Cytoscape software and Our candidate pathogenic genes and the known CHD-related genes were uploaded in STRING database. Each node represents one gene, and each edge represents the protein-protein interaction collected from BioGRID.

signaling pathway, which is involved in the control of cell growth and proliferation (Dobrikov, M. et al., 2011). Diseases associated with EIF4G1 include Parkinson's disease, nonsmall cell lung carcinoma, and prostate cancer (Cao, Y. et al., 2016). Although the relationship between EIF4G1 and cardiovascular development remains unknown, our results suggest that EIF4G1 might be potentially pathogenic in terms of PDA.

HIPK1 belongs to the Ser/Thr family of protein kinases as part of the HIPK subfamily. HIPK1 is related to pathways involved in the regulation of TP53 activity and cardiac conduction. The homeodomain-interacting protein kinases HIPK1 and HIPK2 play key roles in embryonic development by regulating transforming growth factor β-dependent angiogenesis (Aikawa, Y. et al., 2006; Shang, Y. et al., 2013). HIPK1 loss-of-function conditional knockout mice exhibit defects in primitive/definitive hematopoiesis, vasculogenesis, angiogenesis, and neural tube closure (Shang, Y. et al., 2013). In addition, HIPK1 can interact with homeobox proteins and other transcription factors to regulate various biological processes, including signal transduction, apoptosis, embryonic development, and retinal vascular dysfunction (Aikawa, Y. et al., 2006). In our study, only two *HIPK1* variants (rs114516009 and rs114506069) were detected in four individuals with PDA; these are novel variants that have not been reported previously. Further investigation showed that *HIPK1* is highly expressed in human embryonic hearts. However, additional experiments are needed to determine the genetic mechanism by which *HIPK1* contributes to PDA.

SOX8 is a member of the SRY-related HMG-box (SOX) family of transcription factors, which are involved in the regulation of embryonic development and in determining cell fate (Haseeb, A. et al., 2019). In our study, the same rare variant (rs1034733) was detected in three patients with PDA. SOX8 expression is essential in the developing heart, which correlates with heart septation and differentiation of the connective tissue of the valve leaflets (Montero, J. A. et al., 2002). Moreover, a previous study revealed that SOX8 overexpression might be associated with hypoxia-induced cell injury by activating the PI3K/AKT/mTOR and MAPK pathways (Gong, L. C. et al., 2017). Interestingly, DA closure after birth is closely related to the blood oxygenation level, and hypoxia can lead to an increase in endogenous PGE2 release, which directly leads to opening of the DA (Benitz, W. E. et al., 2016). Therefore, *SOX8* may be a novel candidate gene involved in the pathogenesis of PDA.

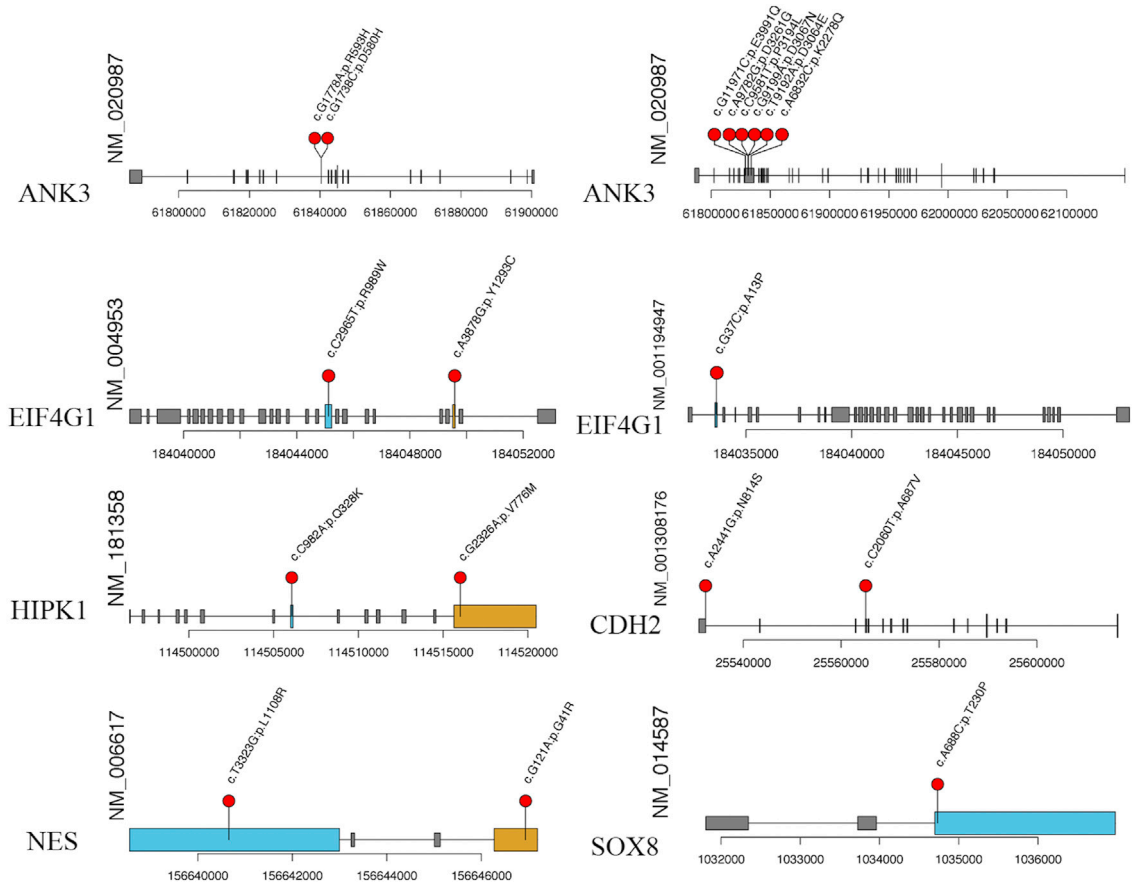


FIGURE 9 | The specific amino acid sites of variants of our candidate gene. The red balls represent the location of rare variant on the encoded proteins or protein domains.

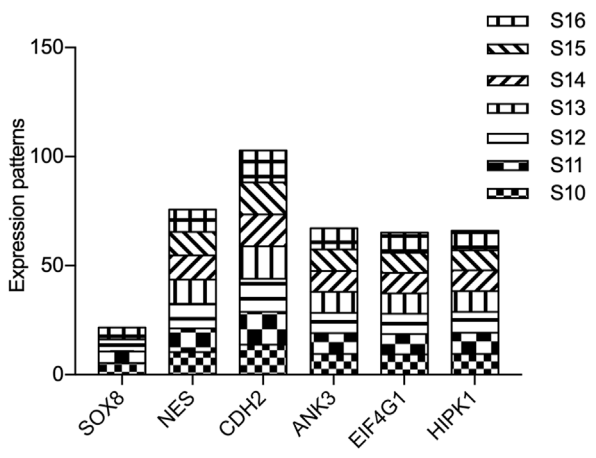


FIGURE 10 | Expression of candidate genes in human embryonic heart. The expression patterns of candidate genes in human embryonic heart at different stages of S10 to S16 were analyzed by microarray. X-axis represents the different stages of human embryonic heart, while the Y-axis indicates the level of gene expression.

In conclusion, through a series of bioinformatics filtering steps, we identified 18 rare damaging variants in six novel candidate genes (*SOX8*, *NES*, *CDH2*, *ANK3*, *EIF4G1*, and *HIPK1*) associated with PDA. The discovery of these genes opens up a new field for genetic research on PDA and provides new ideas for understanding the pathogenesis of PDA. Nevertheless, our study has some limitations. The lack of parental samples and the small sample size limited our ability to identify the detailed genetic background of PDA. Thus, more fundamental research is needed to determine candidate genes that contribute to PDA. We hope to confirm these findings with larger sample sizes.

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/, SRP288538.

ETHICS STATEMENT

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

AUTHOR CONTRIBUTIONS

PZ contributed to design of the study and performed the statistical analysis. YC, BC, and QZ collected the blood samples from all subjects. YG and DW wrote the first draft of the manuscript and contributed to this study equally. PZ revised the manuscript. All authors have read and agreed to the published version of the manuscript.

FUNDING

This study received financial supports from National Natural Science Foundation of China (82070386), the Project of Shanghai Municipal Health Commission (Grant No.201940393).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.921925/full#supplementary-material

REFERENCES

Aikawa, Y., Nguyen, L. A., Isono, K., Takakura, N., Tagata, Y., Schmitz, M. L., et al. (2006). Roles of HIPK1 and HIPK2 in AML1- and P300-dependent Transcription, Hematopoiesis and Blood Vessel Formation. *EMBO J.* 25 (17), 3955–3965. doi:10.1038/sj.emboj.7601273

Alimperti, S., and Andreadis, S. T. (2015). CDH2 and CDH11 Act as Regulators of Stem Cell Fate Decisions. *Stem Cell. Res.* 14 (3), 270–282. doi:10.1016/j.scr.2015.02.002

Amberger, J. S., and Hamosh, A. (2017). Searching Online Mendelian Inheritance in Man (OMIM): A Knowledgebase of Human Genes and Genetic Phenotypes. *Curr. Protoc. Bioinforma.* 58, 1–12. doi:10.1002/cpbi.27

Benitz, W. E., Watterberg, K. L., Cummings, S. J. J., Eichenwald, E. C., Goldsmith, J., Poindexter, B. B., et al. (2016). Patent Ductus Arteriosus in Preterm Infants. *Pediatrics* 137 (1). 1. doi:10.1542/peds.2015-3730

Brohée, S., Faust, K., Lima-Mendez, G., Vanderstocken, G., and van Helden, J. (2008). Network Analysis Tools: from Biological Networks to Clusters and Pathways. *Nat. Protoc.* 3 (10), 1616–1629. doi:10.1038/nprot.2008.100

Cao, Y., Wei, M., Li, B., Liu, Y., Lu, Y., Tang, Z., et al. (2016). Functional Role of Eukaryotic Translation Initiation Factor 4 Gamma 1 (EIF4G1) in NSCLC. *Oncotarget* 7 (17), 24242–24251. doi:10.18632/oncotarget.8168

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: an Ultra-fast All-In-One FASTQ Preprocessor. *Bioinformatics* 34 (17), i884–i890. doi:10.1093/bioinformatics/bty560

Clarke, L., Zheng-Bradley, X., Zheng-Bradley, X., Smith, R., Kulesha, E., Xiao, C., et al. (2012). The 1000 Genomes Project: Data Management and Community Access. *Nat. Methods* 9 (5), 459–462. doi:10.1038/nmeth.1974

Crockett, S. L., Berger, C. D., Shelton, E. L., and Reese, J. (2019). Molecular and Mechanical Factors Contributing to Ductus Arteriosus Patency and Closure. *Congenit. Heart Dis.* 14 (1), 15–20. doi:10.1111/chd.12714

Dobrikov, M., Dobrikova, E., Shveygert, M., and Gromeier, M. (2011). Phosphorylation of Eukaryotic Translation Initiation Factor 4G1 (eIF4G1) by Protein Kinase Cα Regulates eIF4G1 Binding to Mnk1. *Mol. Cell. Biol.* 31 (14), 2947–2959. doi:10.1128/MCB.05589-11

Erdogan, F., Larsen, L. A., Zhang, L., Tumer, Z., Tommerup, N., Chen, W., et al. (2008). High Frequency of Submicroscopic Genomic Aberrations Detected by Tiling Path Array Comparative Genome Hybridisation in Patients with Isolated Congenital Heart Disease. *J. Med. Genet.* 45 (11), 704–709. doi:10.1136/jmg.2008.058776

Flanagan, S. E., Patch, A.-M., and Ellard, S. (2010). Using SIFT and PolyPhen to Predict Loss-Of-Function and Gain-Of-Function Mutations. *Genet. Test. Mol. Biomarkers* 14 (4), 533–537. doi:10.1089/gtmb.2010.0036

Forbes, S. A., Bindal, N., Bamford, S., Cole, C., Kok, C. Y., Beare, D., et al. (2011). COSMIC: Mining Complete Cancer Genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* 39 (Database issue), D945–D950. doi:10.1093/nar/gkq929

Gene Ontology, C. (2015). Gene Ontology Consortium: Going Forward. *Nucleic Acids Res.* 43, D1049–D1056. doi:10.1093/nar/gku1179

Gong, L.-C., Xu, H.-M., Guo, G.-L., Zhang, T., Shi, J.-W., and Chang, C. (2017). Long Non-coding RNA H19 Protects H9c2 Cells against Hypoxia-Induced Injury by Targeting MicroRNA-139. *Cell. Physiol. Biochem.* 44 (3), 857–869. doi:10.1159/000485354

Gravholt, C. H., Viuff, M. H., Brun, S., Stochholm, K., and Andersen, N. H. (2019). Turner Syndrome: Mechanisms and Management. *Nat. Rev. Endocrinol.* 15 (10), 601–614. doi:10.1038/s41574-019-0224-4

Groth, K. A., Skakkebæk, A., Høst, C., Gravholt, C. H., and Bojesen, A. (2013). Klinefelter Syndrome-A Clinical Update. *J. Clin. Endocrinol. Metabolism* 98 (1), 20–30. doi:10.1210/jc.2012-2382

Haimov, O., Sehrawat, U., Tamarkin-Ben Harush, A., Bahat, A., Uzonyi, A., Will, A., et al. (2018). Dynamic Interaction of Eukaryotic Initiation Factor 4G1 (eIF4G1) with eIF4E and eIF1 Underlies Scanning-dependent and -Independent Translation. *Mol. Cell. Biol.* 38 (18). 1. doi:10.1128/MCB.00139-18

Harakalova, M., van der Smagt, J., de Kovel, C. G. F., Van't Slot, R., Poot, M., Nijman, I. J., et al. (2013). Incomplete Segregation of MYH11 Variants with Thoracic Aortic Aneurysms and Dissections and Patent Ductus Arteriosus. *Eur. J. Hum. Genet.* 21 (5), 487–493. doi:10.1038/ejhg.2012.206

Haseeb, A., and Lefebvre, V. (2019). The SOXE Transcription Factors-SOX8, SOX9 and SOX10-Share a Bi-partite Transactivation Mechanism. *Nucleic Acids Res.* 47 (13), 6917–6931. doi:10.1093/nar/gkz523

Hoffman, J. I. E., and Kaplan, S. (2002). The Incidence of Congenital Heart Disease. *J. Am. Coll. Cardiol.* 39 (12), 1890–1900. doi:10.1016/s0735-1097(02)01886-7

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs. *Nucleic Acids Res.* 45 (D1), D353–D361. doi:10.1093/nar/gkw1092

Karczewski, K. J., Weisburd, B., Thomas, B., Solomonson, M., Ruderfer, D. M., Kavanagh, D., et al. (2017). The ExAC Browser: Displaying Reference Data Information from over 60 000 Exomes. *Nucleic Acids Res.* 45 (D1), D840–D845. doi:10.1093/nar/gkw971

Landrum, M. J., Chitipiralla, S., Brown, G. R., Chen, C., Gu, B., Hart, J., et al. (2020). ClinVar: Improvements to Accessing Data. *Nucleic Acids Res.* 48 (D1), D835–D844. doi:10.1093/nar/gkz972

Leussis, M. P., Berry-Scott, E. M., Saito, M., Jhuang, H., de Haan, G., Alkan, O., et al. (2013). The ANK3 Bipolar Disorder Gene Regulates Psychiatric-Related Behaviors that Are Modulated by Lithium and Stress. *Biol. Psychiatry* 73 (7), 683–690. doi:10.1016/j.biopsych.2012.10.016

Li, B., Goyal, J., Dhar, S., Dimri, G., Evron, E., Sukumar, S., et al. (2001). CpG Methylation as a Basis for Breast Tumor-specific Loss of NES1/kallikrein 10 Expression. *Cancer Res.* 61 (21), 8014–8021.

Li, N., Subrahmanyan, L., Smith, E., Yu, X., Zaidi, S., Choi, M., et al. (2016). Mutations in the Histone Modifier PRDM6 Are Associated with Isolated Nonsyndromic Patent Ductus Arteriosus. *Am. J. Hum. Genet.* 98 (6), 1082–1091. doi:10.1016/j.ajhg.2016.03.022

Liang, Y., Jiang, L., Zhong, X., Hochwald, S. N., Wang, Y., Huang, L., et al. (2019). Discovery of Aberrant Alteration of Genome in Colorectal Cancer by Exome Sequencing. *Am. J. Med. Sci.* 358 (5), 340–349. doi:10.1016/j.amjms.2019.07.012

Luo, L.-Y., Meyts, E. R.-D., Jung, K., and Diamandis, E. P. (2001). Expression of the Normal Epithelial Cell-specific 1 (NES1; KLK10) Candidate Tumour Suppressor Gene in Normal and Malignant Testicular Tissue. *Br. J. Cancer* 85 (2), 220–224. doi:10.1054/bjoc.2001.1870

Luo, L., Herbrick, J.-A., Scherer, S. W., Beatty, B., Squire, J., and Diamandis, E. P. (1998). Structural Characterization and Mapping of the Normal Epithelial Cell-specific 1 Gene. *Biochem. Biophysical Res. Commun.* 247 (3), 580–586. doi:10.1006/bbrc.1998.8793

Masai, I., Lele, Z., Yamaguchi, M., Komori, A., Nakata, A., Nishiwaki, Y., et al. (2003). N-cadherin Mediates Retinal Lamination, Maintenance of Forebrain Compartments and Patterning of Retinal Neurites. *Development* 130 (11), 2479–2494. doi:10.1242/dev.00465

Mayosi, B. M., Fish, M., Shaboodien, G., Mastantuono, E., Kraus, S., Wieland, T., et al. (2017). Identification of Cadherin 2 (CDH2) Mutations in Arrhythmogenic Right Ventricular Cardiomyopathy. *Circ. Cardiovasc Genet.* 10 (2). 1. doi:10.1161/CIRCGENETICS.116.001605

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data. *Genome Res.* 20 (9), 1297–1303. doi:10.1101/gr.107524.110

Mitra, S., Florez, I. D., Tamayo, M. E., Mbuagbaw, L., Vanniyasingam, T., Veroniki, A. A., et al. (2018). Association of Placebo, Indomethacin, Ibuprofen, and Acetaminophen with Closure of Hemodynamically Significant Patent Ductus Arteriosus in Preterm Infants. *JAMA* 319 (12), 1221–1238. doi:10.1001/jama.2018.1896

Mohler, P. J., Splawski, I., Napolitano, C., Bottelli, G., Sharpe, L., Timothy, K., et al. (2004). A Cardiac Arrhythmia Syndrome Caused by Loss of Ankyrin-B Function. *Proc. Natl. Acad. Sci. U.S.A.* 101 (24), 9137–9142. doi:10.1073/pnas.0402546101

Montero, J. A., Giron, B., Arrechedera, H., Cheng, Y. C., Scotting, P., Chimal-Monroy, J., et al. (2002). Expression of Sox8, Sox9 and Sox10 in the Developing Valves and Autonomic Nerves of the Embryonic Heart. *Mech. Dev.* 118 (1-2), 199–202. doi:10.1016/s0925-4773(02)00249-6

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., et al. (2016). Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation. *Nucleic Acids Res.* 44 (D1), D733–D745. doi:10.1093/nar/gkv1189

O'Rahilly, R. (1987). Human Embryo. *Nature* 329 (6138), 385. doi:10.1038/329385e0

Okonechnikov, K., Conesa, A., and García-Alcalde, F. (2016). Qualimap 2: Advanced Multi-Sample Quality Control for High-Throughput Sequencing Data. *Bioinformatics* 32 (2), btv566–294. doi:10.1093/bioinformatics/btv566

Pannone, L., Bocchinfuso, G., Flex, E., Rossi, C., Baldassarre, G., Lissewski, C., et al. (2017). Structural, Functional, and Clinical Characterization of a NovelPTPN11Mutation Cluster Underlying Noonan Syndrome. *Hum. Mutat.* 38 (4), 451–459. doi:10.1002/humu.23175

Radice, G. L., Rayburn, H., Matsunami, H., Knudsen, K. A., Takeichi, M., and Hynes, R. O. (1997). Developmental Defects in Mouse Embryos Lacking N-Cadherin. *Dev. Biol.* 181 (1), 64–78. doi:10.1006/dbio.1996.8443

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (2015). Standards and Guidelines for the Interpretation of Sequence Variants: a Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* 17 (5), 405–424. doi:10.1038/gim.2015.30

Satoda, M., Zhao, F., Diaz, G. A., Burn, J., Goodship, J., Davidson, H. R., et al. (2000). Mutations in TFAP2B Cause Char Syndrome, a Familial Form of Patent Ductus Arteriosus. *Nat. Genet.* 25 (1), 42–46. doi:10.1038/75578

Shang, Y., Doan, C. N., Arnold, T. D., Lee, S., Tang, A. A., Reichardt, L. F., et al. (2013). Transcriptional Corepressors HIPK1 and HIPK2 Control Angiogenesis *via* TGF-β-TAK1-dependent Mechanism. *PLoS Biol.* 11 (4), e1001527. doi:10.1371/journal.pbio.1001527

Steinhaus, R., Proft, S., Schuelke, M., Cooper, D. N., Schwarz, J. M., and Seelow, D. (2021). MutationTaster2021. *Nucleic Acids Res.* 49 (W1), W446–W451. doi:10.1093/nar/gkab266

Vanlerberghe, C., Jourdain, A.-S., Ghoumid, J., Frenois, F., Mezel, A., Vaksmann, G., et al. (2019). Holt-oram Syndrome: Clinical and Molecular Description of 78 Patients with TBX5 Variants. *Eur. J. Hum. Genet.* 27 (3), 360–368. doi:10.1038/s41431-018-0303-3

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data. *Nucleic Acids Res.* 38 (16), e164. doi:10.1093/nar/gkq603

Wirgenes, K. V., Tesli, M., Inderhaug, E., Athanasiu, L., Agartz, I., Melle, I., et al. (2014). ANK3 Gene Expression in Bipolar Disorder and Schizophrenia. *Br. J. Psychiatry* 205 (3), 244–245. doi:10.1192/bjp.bp.114.145433

Yang, D., Liu, B. C., Luo, J., Huang, T. X., and Liu, C. T. (2019). Kartagener Syndrome. *QJM* 112 (4), 297–298. doi:10.1093/qjmed/hcy242

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 Genetics

ORIGINAL RESEARCH
published: 09 June 2022
doi: 10.3389/fgene.2022.908510

Identification of Important Modules and Biomarkers That Are Related to Immune Infiltration Cells in Severe Burns Based on Weighted Gene Co-Expression Network Analysis

Zexin Zhang[†,1], Yan He[†,1], Rongjie Lin[†,3], Junhong Lan[1], Yueying Fan[1], Peng Wang[,1,2] and Chiyu Jia[*,1]*

[1]Department of Burns and Plastic and Wound Repair Surgery, Xiang'an Hospital of Xiamen University, School of Medicine, Xiamen University, Xiamen, China, [2]Department of Burns and Plastic and Cosmetic Surgery, The Ninth Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China, [3]Department of Orthopedics, The 900th Hospital of Joint Logistic Support Force, Fuzhou, China

OPEN ACCESS

Edited by:
Tao Huang,
Shanghai Institute of Nutrition and Health (CAS), China

Reviewed by:
Michael Poidinger,
Royal Children's Hospital, Australia
Jian Wang,
Kunming Medical University, China
Jianyong Zhang,
Zunyi Medical University, China

***Correspondence:**
Chiyu Jia
jiachiyu@qq.com
Peng Wang
nxmuwp@163.com

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

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: 30 March 2022
Accepted: 22 April 2022
Published: 09 June 2022

Citation:
Zhang Z, He Y, Lin R, Lan J, Fan Y, Wang P and Jia C (2022) Identification of Important Modules and Biomarkers That Are Related to Immune Infiltration Cells in Severe Burns Based on Weighted Gene Co-Expression Network Analysis. Front. Genet. 13:908510. doi: 10.3389/fgene.2022.908510

Background: Immunosuppression is an important trigger for infection and a significant cause of death in patients with severe burns. Nevertheless, the prognostic value of immune-related genes remains unclear. This study aimed to identify the biomarkers related to immunosuppression in severe burns.

Methods: The gene expression profile and clinical data of 185 burn and 75 healthy samples were obtained from the GEO database. Immune infiltration analysis and gene set variation analysis were utilized to identify the disorder of circulating immune cells. A weighted gene co-expression network analysis (WGCNA) was carried out to select immune-related gene modules. Enrichment analysis and protein–protein interaction (PPI) network were performed to select hub genes. Next, LASSO and logistic regression were utilized to construct the hazard regression model with a survival state. Finally, we investigated the correlation between high- and low-risk patients in total burn surface area (TBSA), age, and inhalation injury.

Results: Gene set variation analysis (GSVA) and immune infiltration analysis showed that neutrophils increased and T cells decreased in severe burns. In WGCNA, four modular differently expressed in burns and controls were related to immune cells. Based on PPI and enrichment analysis, 210 immune-related genes were identified, mainly involved in T-cell inhibition and neutrophil activation. In LASSO and logistic regression, we screened out key genes, including *LCK, SKAP1* and *GZMB,* and *LY9*. In the ROC analysis, the area under the curve (AUC) of key genes was 0.945, indicating that the key genes had excellent diagnostic value. Finally, we discovered that the key

Abbreviations: DAMP, damage-associated molecular patterns; GSVA, gene set variation analysis; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; KICs, key immune cells; LASSO, Least Absolute Shrinkage and Selectionator Operator; PPI, protein–protein interaction; PAMP, pathogen-associated molecular patterns; TBSA, total burn surface area; and WGCNA, weighted gene co-expression network analysis.

genes were related to T cells, and the regression model performed well when accompanied by TBSA and age.

Conclusion: We identified LCK, SKAP1, GZMB, and LY9 as good prognostic biomarkers that may play a role in post-burn immunosuppression against T-cell dysfunction and as potential immunotherapeutic targets for transformed T-cell dysfunction.

Keywords: immunosuppression, burns, WGCNA, LASSO, GSVA, CIBERSORT, prognostic biomarker

INTRODUCTION

There are 180,000 people who die as a result of burns, with 47 percent of those fatalities linked to infection-related complications. Infections in severe burns are often caused by an overactive inflammatory response and immunosuppression. In severe burns, adaptive immune functions represented by T cells are suppressed and inflammatory responses represented by neutrophils and dendritic cells are activated (Miller et al., 2007; Sood et al., 2016; Hampson et al., 2017). Impaired skin and intestinal mucosal barriers are exposed to pathogens, and the disorder of peripheral blood cells results in a low response to pathogens (Neely et al., 2004; Wrba et al., 2017). All of that mentioned above lead to uncontrollable infection and death (Fitzwater et al., 2003).

Immunosuppression in severe burns is considered to be significantly correlated with prognosis. In the early stages of burns, PAMP-mediated innate immune translation was enhanced, such as macrophages, dendritic cells, and neutrophils being recruited to the injured site to clear necrotic tissue. Subsequently, the adaptive immune response is impaired, such as Th cell subtype imbalance, where Th1 cells are inhibited and Th2 cells are activated, resulting in immunosuppression. Inflammatory factors, cytokines, and immune cells have a significant prognostic value in severe burns (Hur et al., 2015; Osuka et al., 2019). Previous studies have explored the prognostic value of platelets, inflammatory factors, immune-related cytokines, and scoring scales. However, the accuracy and clinical practicability need to be improved. These prognostic factors cannot explain the disturbance of homeostasis after severe burns, especially immunosuppression (Hur et al., 2015; Lip et al., 2019; Geng et al., 2020). Alterations in gene expression profiles underlie disease development and can reflect changes in homeostasis from a pathophysiological perspective, explaining the mechanisms that affect prognosis and providing therapeutic targets for subsequent studies (Gaetani et al., 2019; Zou and Wang 2019). In addition, gene detection is convenient, economical, and has a strong stability in the application of prognosis. Genes have been used as biomarkers to model the prognosis of a variety of diseases, showing strong prognostic power. However, their application to burns is rare (Sandquist and Wong 2014; Foth et al., 2016; Gavrielatou et al., 2020). Therefore, it is meaningful to explore immune-related genes in severe burns for revealing their value as prognostic biomarkers and immune therapeutic targets.

This study is a large population-based prognostic study involving 185 burn patients. We investigated the relationship between gene expression profiles and the prognosis of severe burns using machine learning algorithms (Newman et al., 2015). The disorder of immune cells in severe burns was investigated by CIBERSORT and gene set variation analysis (GSVA) (Hänzelmann et al., 2013). The genes related to the disorder of immune cells were identified in WGCNA. We used the Least Absolute Shrinkage and Selectionator operator (LASSO) and logistic regression to create a prognostic model for immune-related genes. Additionally, we identified the correlation between cellular subtypes and genes, which were associated with immune abnormalities following severe burns. The research aimed to provide a certain basis and reference value for revealing the prognostic value of genes associated with immunosuppression.

METHODS

Acquisition of RNA Data

We downloaded three microarray expression profiles and clinical data of severe burns (GSE19743, GSE77791, and GSE37069) from the GEO database (http://www.ncbi.nlm.nih.gov/geo/). Patients >18 years or <55 years and sampling time between 280 and 705 h were selected to remove the influence of age and sampling time. (**Table 1**). Meanwhile, we downloaded clinical information (survival, burn area, sampling time, and age) from three datasets. GSE37069 was utilized to screen immune- and prognostic-related genes between burn and control (**Supplementary Data Sheet S1**), GSE77791 and GSE19743 were utilized to be the training cohort and validating cohort, between survival and non-survival, respectively (**Supplementary Data Sheets S2, S3**). There was no need for patients' consent and ethical approval as all data were taken from public databases. The experimental procedure was as shown in **Figure 1**.

TABLE 1 | Clinical data of burn patients and health controls in GSE37069 and GSE19743.

Group (burn)	Sex			Age	Time of sampling
	N	Male	Female		
GSE19743	28	24	4	37.61 ± 7.98	439.28 ± 117.86
GSE37069	81	57	24	37.41 ± 10.45	411.92 ± 124.76
P		0.109		0.271	0.311
Group (health)	**Sex**			**Age**	**Time of sampling**
	N	**Male**	**Female**		
GSE19743	25	14	11	30.21 ± 8.16	—
GSE37069	37	17	20	32.59 ± 11.03	—
P		0.709		0.364	—

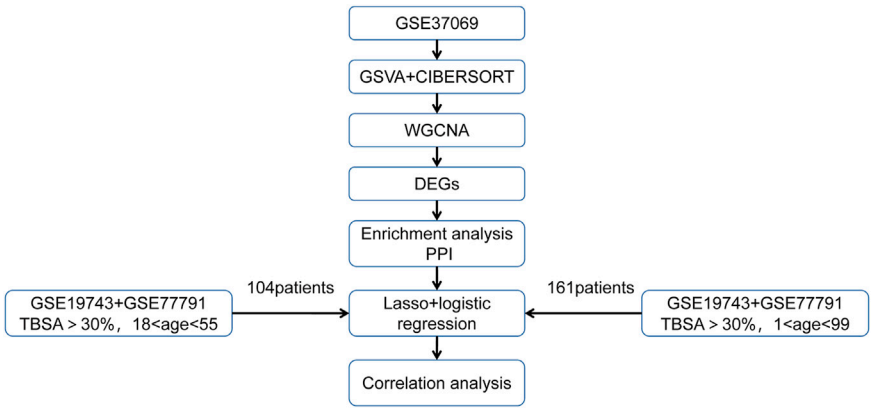


FIGURE 1 | A graphical summary of the research design.

TABLE 2 | Clinical data of burn patients in GSE77791 and GSE19743.

Group	Sex			Age	TBSA		Time of sampling (GSE19743)
	N	Male	Female		Severe (30–49)	Major (49–100)	
Death	23	19	4	40.73 ± 7.67	3	20	422.24 ± 122.32
Survival	81	70	11	40.67 ± 10.74	21	60	393.33 ± 113.19
P		0.147		0.974	0.001		0.544

Data Processing

The expression profiles were processed by the package of R software, "*Affy*." The background correction of the expression value and normalization of the expression profile data were performed, including conversion of original data format, the supplement of missing value, background correction, and data standardization by using the Quantile method.

GSVA

GO term related to immune cells was selected from the GSEA website (http://www.gsea-msigdb.org/gsea/index.jsp). The GSVA analysis was performed between burns and healthy controls in GSE37069 by the "*GSVA*" package in R software. Unqualified samples were removed prior to the variance analysis. The result of GSVA was analyzed by the R package "limma" to calculate the differences in enrichment results between severe burns and healthy controls.

Immune Infiltration Analysis

"*CIBERSORT*" is a machine learning algorithm that can analyze the proportion of immune cells from RNA-seq (Newman et al., 2015). We downloaded the expression profile of GSE37069, GSE19743, and GSE77791 to select 81, 28, and 76 burn patients and 37, 25, and 13 healthy controls, respectively, and performed an immune infiltration analysis by using the "*CIBERSORT*" package in R software. Cell subtypes with $p < 0.05$ in three gene sets were considered as key immune cells (KICs) for further analysis. Next, the "*ggplot2*" package in the R software was used to visualize the different proportion of KICs between burn patients (Group2) and healthy controls (Group1).

WGCNA

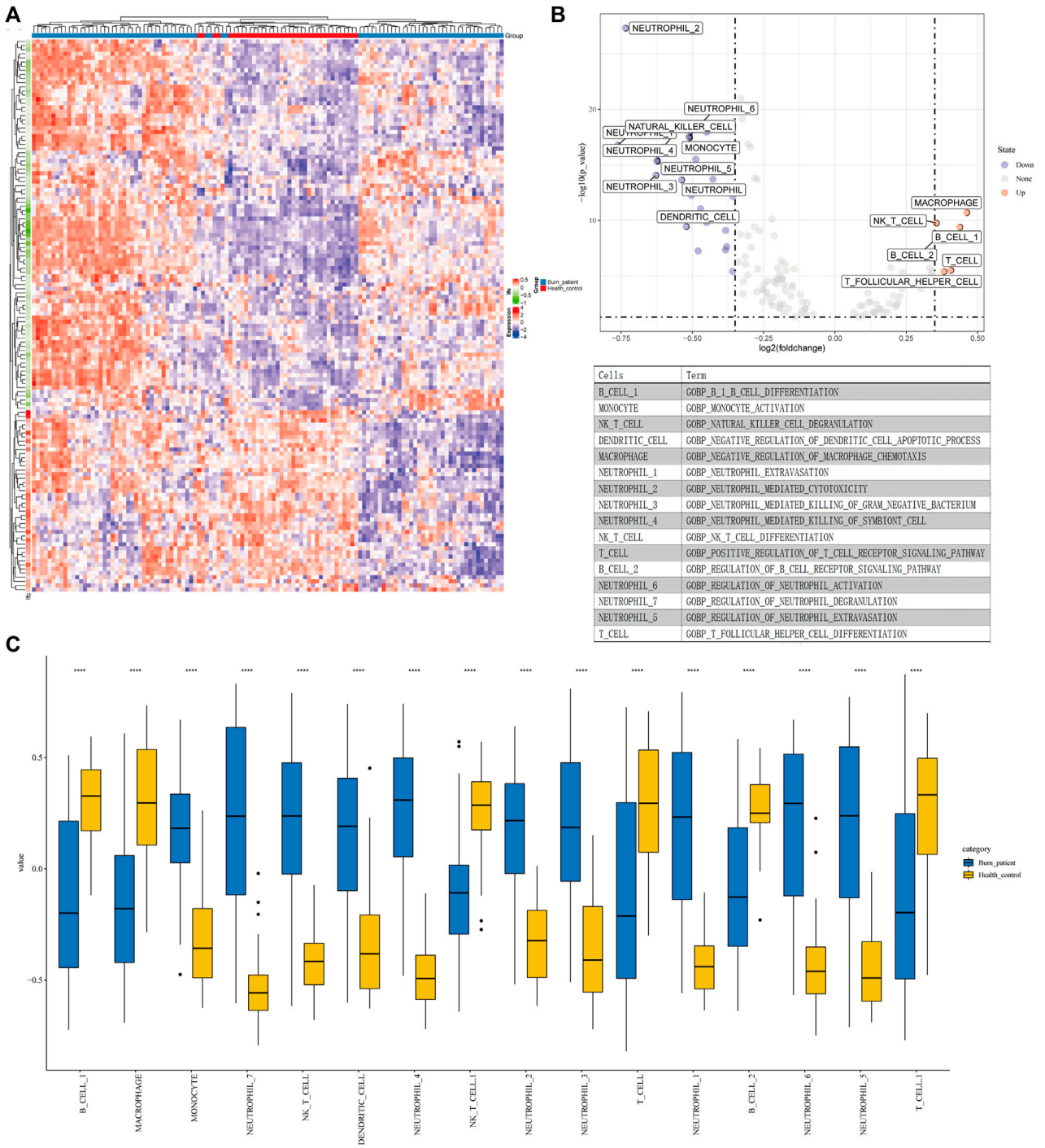
WGCNA could construct a scale-free distribution network by using soft power to classify genes with the same expression trend and analyze the correlation between genes and traits (Langfelder and Horvath 2008). We performed WGCNA on the gene set GSE37069 by using the "*WGCNA*" package in the R software and identified gene modules associated with key immune cells. The intersection of GSVA and immune infiltration analysis are defined as clinical traits to identify immune-related gene modules.

Differential Expression Analysis

We utilized the "*LIMMA*" package in R software (version 4.0.5) to analyze DEGs in GSE19743, GSE77791, and GSE37069 datasets (FDR<0.05, |logFC|>1) and took the intersection of genes in GSE19743, GSE77791, GSE37069, and immune-related modules.

Enrichment Analysis and PPI

We performed an enrichment analysis of differently expressed genes in the immune-related gene modules.(modules with differential genes more than 20 were selected). We used the DAVID6.8 online tool (https://david.ncifcrf.gov) to perform the enrichment analysis of Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) and the "*ggplot2*" package in R software was used to draw a bubble chart. We constructed the interaction network between the enrichment results (FDR<0.05). We selected immune-related genes in the enrichment analysis and constructed the PPI network which was visualized by Cytoscape and hub genes selected by MCODE.



Cells	Term
B_CELL_1	GOBP_B_1_B_CELL_DIFFERENTIATION
MONOCYTE	GOBP_MONOCYTE_ACTIVATION
NK_T_CELL	GOBP_NATURAL_KILLER_CELL_DEGRANULATION
DENDRITIC_CELL	GOBP_NEGATIVE_REGULATION_OF_DENDRITIC_CELL_APOPTOTIC_PROCESS
MACROPHAGE	GOBP_NEGATIVE_REGULATION_OF_MACROPHAGE_CHEMOTAXIS
NEUTROPHIL_1	GOBP_NEUTROPHIL_EXTRAVASATION
NEUTROPHIL_2	GOBP_NEUTROPHIL_MEDIATED_CYTOTOXICITY
NEUTROPHIL_3	GOBP_NEUTROPHIL_MEDIATED_KILLING_OF_GRAM_NEGATIVE_BACTERIUM
NEUTROPHIL_4	GOBP_NEUTROPHIL_MEDIATED_KILLING_OF_SYMBIONT_CELL
NK_T_CELL	GOBP_NK_T_CELL_DIFFERENTIATION
T_CELL	GOBP_POSITIVE_REGULATION_OF_T_CELL_RECEPTOR_SIGNALING_PATHWAY
B_CELL_2	GOBP_REGULATION_OF_B_CELL_RECEPTOR_SIGNALING_PATHWAY
NEUTROPHIL_6	GOBP_REGULATION_OF_NEUTROPHIL_ACTIVATION
NEUTROPHIL_7	GOBP_REGULATION_OF_NEUTROPHIL_DEGRANULATION
NEUTROPHIL_5	GOBP_REGULATION_OF_NEUTROPHIL_EXTRAVASATION
T_CELL	GOBP_T_FOLLICULAR_HELPER_CELL_DIFFERENTIATION

FIGURE 2 | Results of GSVA and CIBERSORT **(A)** Heat map of 203 GO terms between severe burns and controls. **(B,C)** Different terms of GSVA with |logFC| >0.35, $p < 0.05$.

LASSO and Logistic Regression

In GSE19743 and GSE77791 (age 18–55, TBSA >30%, sampling time 280–705 h), we performed a LASSO regression analysis by using the "*lasso*" package of R software to screen hub genes. The hub genes were analyzed by ROC curves and the variables with AUC>0.6 were selected for a logistic regression to establish the regression model. The regression model used GSE77791 as the training cohort and GSE19743 as the validation cohort. The nomogram plot was utilized to calculate the risk score, and the model was evaluated by the ROC curve and calibration curve.

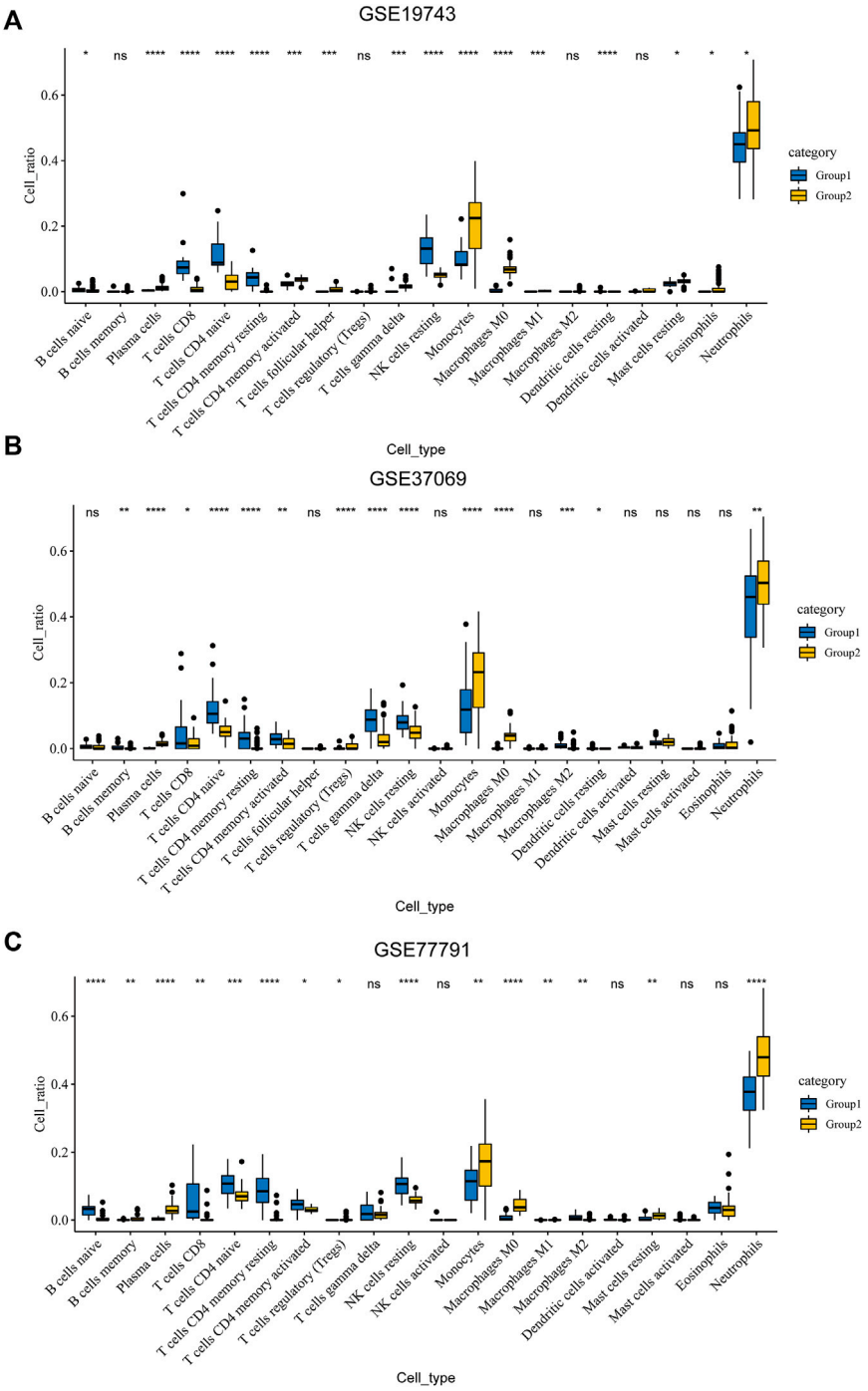


FIGURE 3 | Results of CIBERSORT. **(A)** Different ratios of the 22 immune cells between severe burns and controls in GSE19743. **(B)** Different ratios of the 22 immune cells between severe burns and controls in GSE37069. **(C)** Different ratios of the 22 immune cells between severe burns and controls in GSE77791.

Multidimensional Validation

To explore the effects of burn area and age on the model, we reincorporated 161 patients of GSE19743 and GSE77791 to validate the prognostic ability of the key genes. Burn area, age, and key genes were included in for modeling. Regression models were constructed by random grouping (70% training cohort and 30% validation cohort). A nomogram was drawn to calculate the risk score, and the ROC curve was used to evaluate the accuracy.

Correlation Analysis

A correlation analysis was performed between key immune cell subtypes and genes, LCK, SKAP1, GZMB, and LY9. The patients

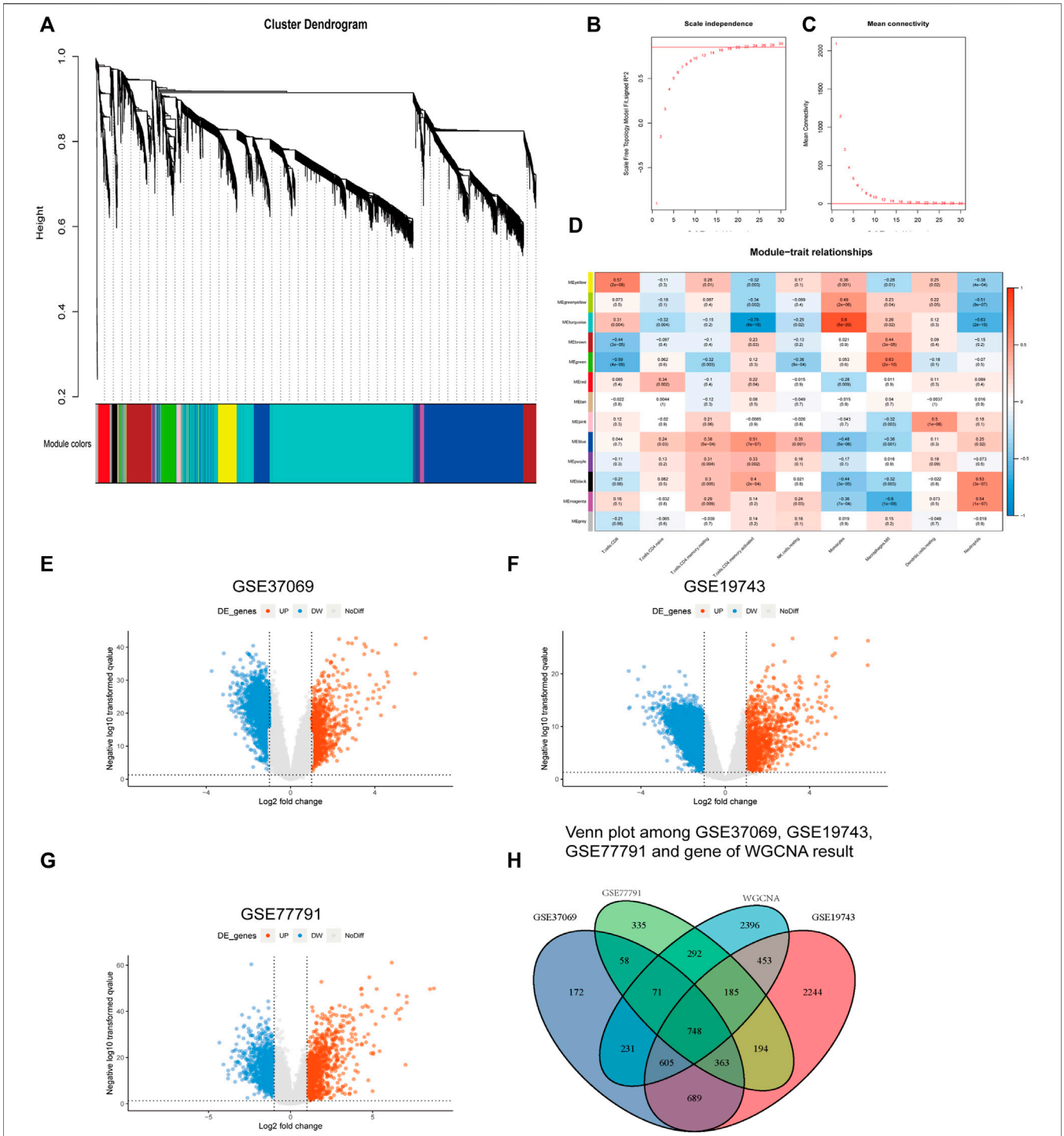


FIGURE 4 | Results of WGCNA and different expression analysis. **(A)** Genes in GSE37069 were divided into 12 modules in different colors. Genes in the same color module have similar expression patterns. **(B,C)** Soft power of WGCNA. **(D)** Correlation genes and immune cells. The ordinate is the gene module; the abscissa is the cell type. Blue are negative correlations and red are positive correlations. **(E–G)** The ordinate is the -$\log_{10}$*p* value; the abscissa is the logFC. **(H)** Intersection of genes in GSE19743, GSE37069, GSE77791, and immune cell-related genes in WGCNA.

were divided into two groups according to the risk score (high risk: score>0.5, low risk: score<0.5). The prognosis, age, gender, TBSA, and inhalation injury were contrasted between the high- and low-risk groups.

Validation of Key Gene Expression

To further verify the prognostic value of the key genes, we verified the expression profiles of key genes in death and survivors of burn patients in independent cohorts, GSE19743 and GSE77791. We converted the fluorescence values of each sample to log2 and averaged the different probes of the same gene.

Statistics Method

Variables are represented by mean±σ. Comparisons of data were made by using the Chi-square test for categorical data or Student's t-test for normalized quantitative data as appropriate. The multiple logistic regression model was utilized to examine the relationship between mortality and variables. The criterion variable was death as the outcome. The explanatory variables included age, gender, %TBSA, and expression of genes. ROC and calibration curves were utilized to process prognostic ability logistic models.

RESULT

Acquisition of RNA Data

We selected 28, 76, and 81 burn patients with a control group of 25, 13, and 37 health controls in the GSE19743, GSE77791, and GSE37069 datasets. GSE19743 and GSE77791 (training and validating cohorts) had 104 severe burns, with 81 survivors and 23 deaths (**Table 2**).

GSVA

We selected 203 immune-related GO terms which had significant differences in immune-related pathways between normal and severe burns (**Figure 2A**). Adj. $p < 0.05$, $|logFC| > 0.35$ are considered to be different GO terms in the differential analysis (**Figure 2B**). In severe burns, the enrichment score of neutrophil, dendritic, monocyte, and NKT cell-related pathways were increased while T cells, B cells, and macrophages were decreased (**Figure 2C**).

Immune Infiltration Analysis

An analysis of immune infiltration showed that plasma cells, T cells CD8, T cells CD4 naive, T cells CD4 memory resting, T cells CD4 memory activated, NK cells resting, monocytes, macrophages M0, dendritic cells resting and neutrophils were KICs (**Figures 3A–C**). We took the intersection of GSVA and CIBERSORT results. All seven kinds of immune cells, which were utilized for WGCNA, had a significant difference between burns and healthy controls.

WGCNA

Genes in GSE37069 were divided into twelve modules (**Figure 4A**). The soft power was 22 ($R > 0.85$) (**Figures 4B,C**). Yellow, turquoise, green, and blue modules were related to the proportion of immune cells and named immune-related modules (correlated to T cells, $p < 0.05$, correlation coefficient > 0.5) (**Figure 4D**).

Differential Expression Analysis

We obtained 2,937, 5,481, and 3,233 differential expression genes from GSE37069, GSE19743, and GSE77791 ($|logFC|>2$, *adj.* $p < 0.05$) and there are 748 differential expression genes in immune-related modules (**Figures 4E–G**).

Enrichment Analysis and PPI

The blue module was mainly enriched in T-cell activation, lymphocyte differentiation, and T-cell differentiation (**Figure 5A**). The green module was mainly enriched in neutrophil degranulation and neutrophil activation involved in immune response (**Figure 5B**). Turquoise was mainly enriched in neutrophil degranulation and neutrophil activation involved in immune response and regulation of inflammatory response (**Figure 5C**). Yellow was mainly enriched in the antigen receptor-mediated signaling pathway, immune response-activating cell surface receptor signaling pathway, and immune response-activating signal transduction pathway (**Figure 5D**). A total of 210 immune-related genes were found in the aforementioned immune-related enrichment results, which were mainly related to the function and cell structure of immune-related genes such as T cells, immune response, MHC II class protein complex, CD4 receptor, and Ca^{2+} signal pathway (**Figures 5E–G**). Three core modules were selected in the MCODE module, which were marked by blue, red, and orange, with a total of 53 hub genes (**Figures 6A–D**).

LASSO and Logistic Regression

26 variables were screened out in LASSO and logistic regression, and 7 genes that had AUC >0.6 were selected. Four immune-related genes LCK, SKAP1, GZMB, and LY9 were obtained by logistic regression modeling (**Figures 7A–D**). A nomogram plot was drawn to calculate the risk score of each patient, and the ROC curves were used to evaluate the prognostic ability of the risk score. AUC was 0.930 in the training cohort and AUC was 0.919 in the validation cohort (**Figures 7E–G**). The aclibration curve shows that the regression model has a good prediction ability (**Figure 7H**). 161patientswere randomly divided into two groups, training cohort (70%) and validation cohort (30%). Results showed that $AUC^{training} = 0.946$, $AUC^{validation} = 0.902$ (**Figures 7I,J**). The risk of non-survival was calculated by the nomogram containing risk score, age, and TBSA. $AUC^{risk\ score+TBSA+age} = 0.945 > AUC^{risk\ score} = 0.933$ ($p < 0.05$) (**Figures 7K,L**). Incorporating TBSA and age improves model predictive ability. Dead patients were older in age, had larger TBSA, and were not different in inhalation injury (**Figures 8A,B**) (**Table 3**). LCK, SKAP1, GZMB, and LY9 are associated with T cell CD4 naive, T cell CD4 memory activated, and T cell CD8 (**Figures 8C,D**).

Validation of Key Gene Expression

In GSE19743 cohort, key genes, LY9, SKAP1, GZMB, and LCK, were highly expressed in survival patients. The same result was presented in GSE77791 (**Figures 9A–H**).

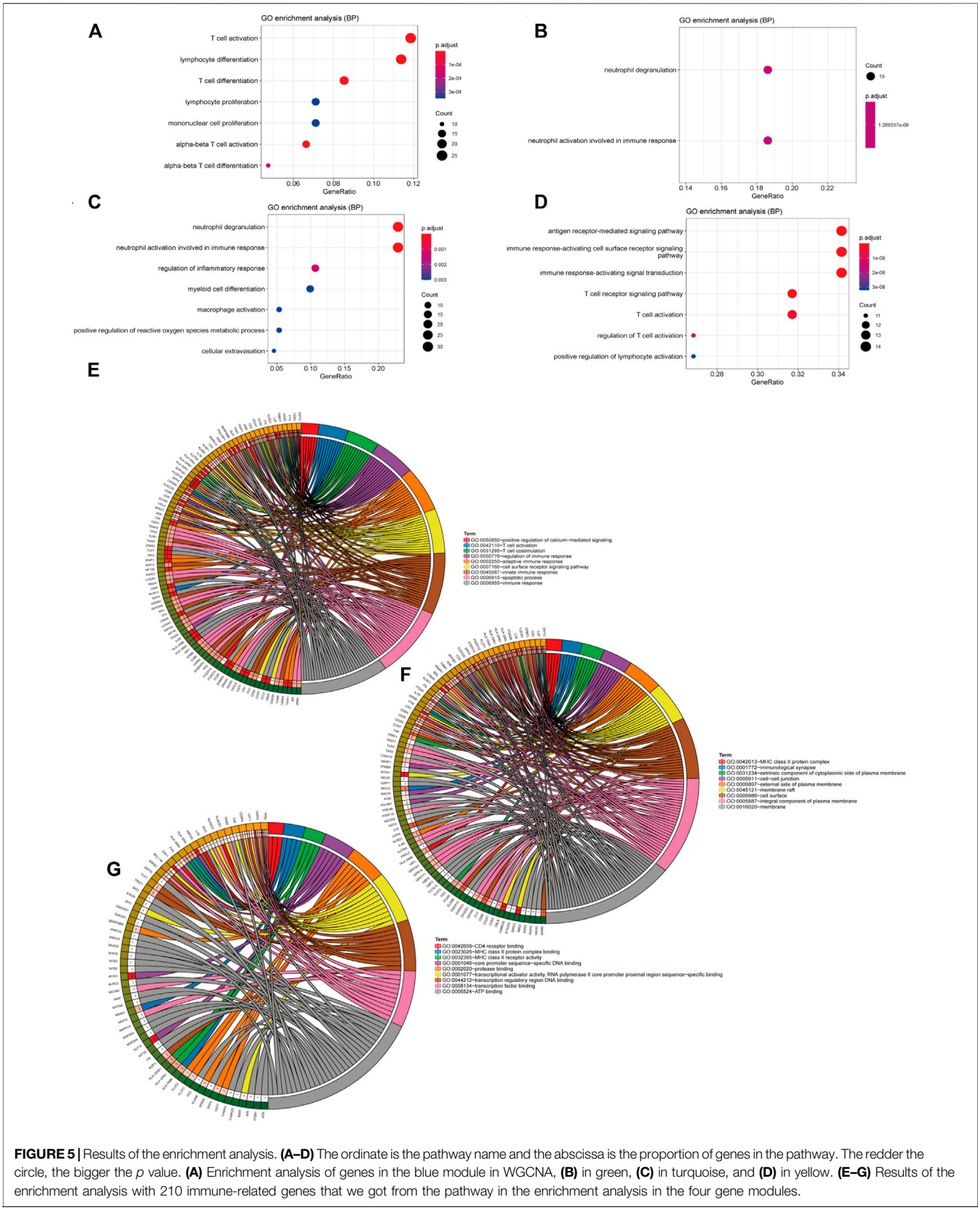


FIGURE 5 | Results of the enrichment analysis. **(A–D)** The ordinate is the pathway name and the abscissa is the proportion of genes in the pathway. The redder the circle, the bigger the *p* value. **(A)** Enrichment analysis of genes in the blue module in WGCNA, **(B)** in green, **(C)** in turquoise, and **(D)** in yellow. **(E–G)** Results of the enrichment analysis with 210 immune-related genes that we got from the pathway in the enrichment analysis in the four gene modules.

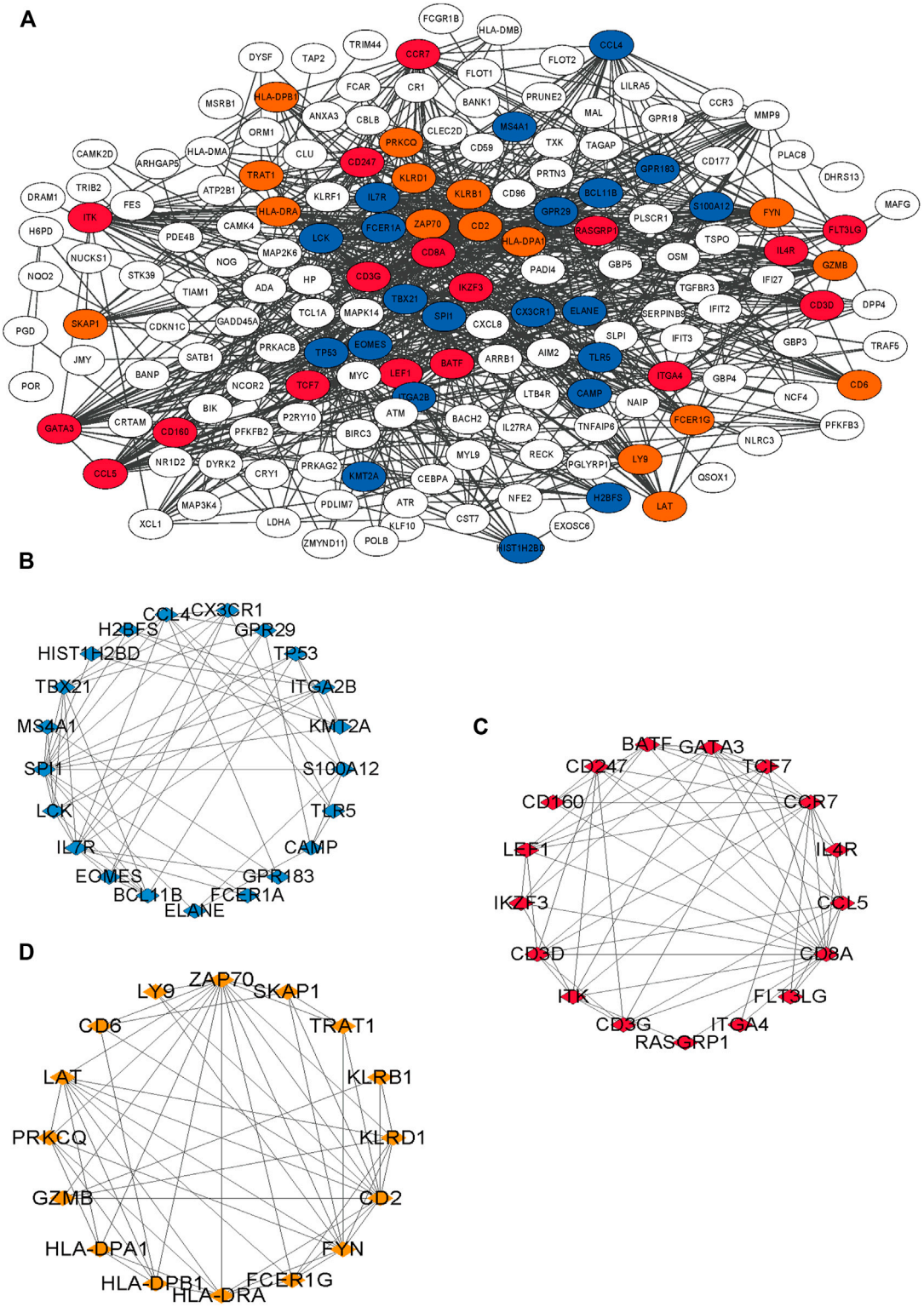


FIGURE 6 | Results of PPI network of 210 immune-related genes. **(A)** The interaction of 210 genes, different colors means different interaction groups. **(B–D)** Blue is module 1, orange is 2 and red is 3.

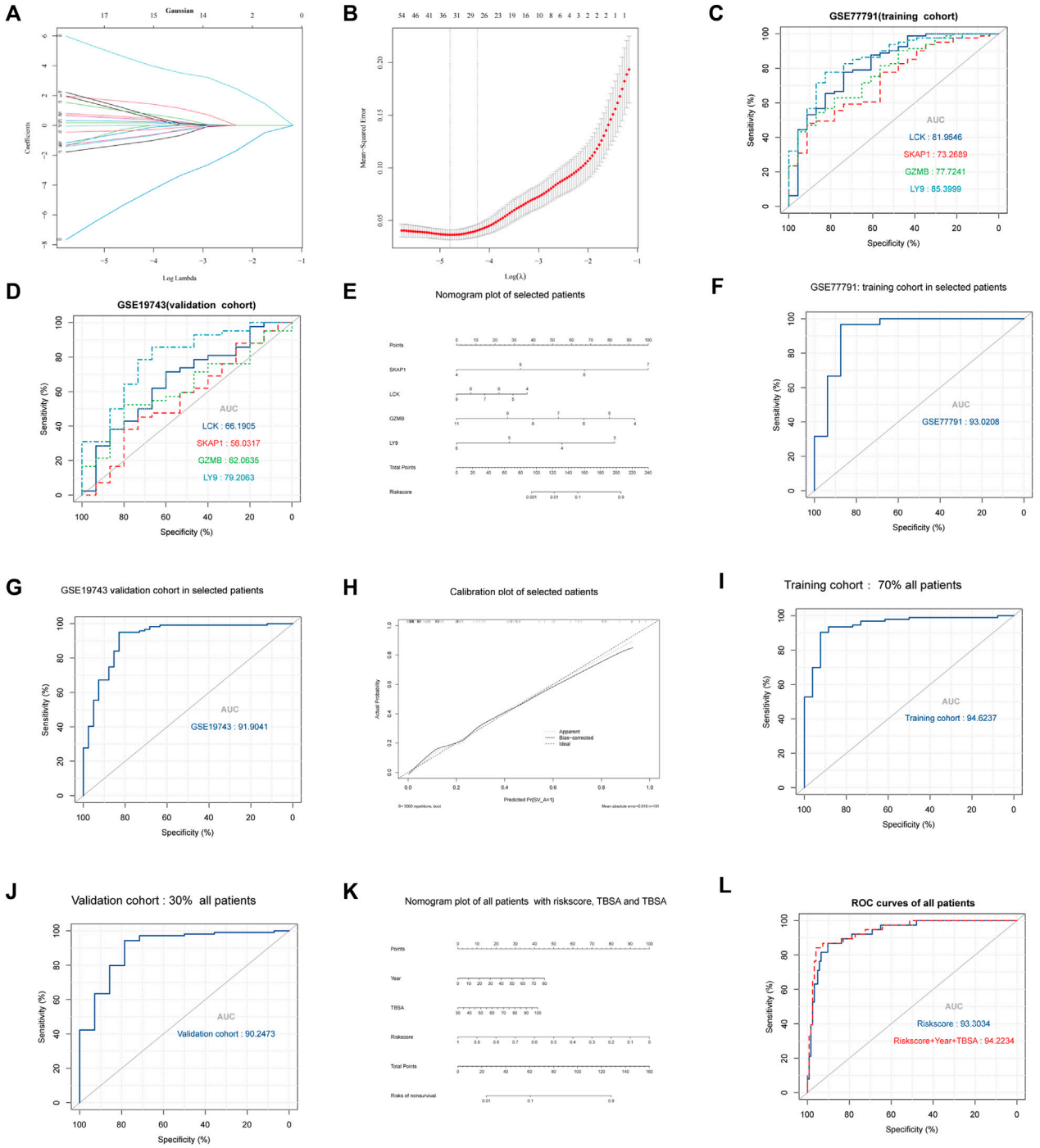


FIGURE 7 | Results of regression, multidimensional verification, and correlation between key genes and immune cells. **(A,B)** The ordinate is the correlation coefficient between gene expression and prognosis. λ was utilized to screened genes and the number aforementioned is the number of genes. **(C)** AUC of the four key genes in GSE77791. The larger the area under the curve, the stronger the prediction ability. **(D)** AUC of the four key genes in GSE19743. **(E)** Nomogram plot of logistic regression which can be utilized to calculate the risk score between severe burns and controls. **(F)** AUC of logistic regression with selected patients (18 < age<55, TBSA>30, 280 h < sample times <706) in the training cohort (GSE77791). **(G)** AUC of logistic regression with selected patients (18 < age<55, TBSA>30%, 280 h < sample times <706) in the validation cohort (GSE19743). **(H)** Calibration curve of logistic regression. The closer bias-corrected curve and ideal curve are, the better predictive regression model is. **(I,J)** Selected patients (*N* = 104, 18 < age<55, TBSA>30, 280 h < sample times <706) were divided into two cohorts in randomly to train and validate regression model, the training cohort had 70% patients, and the validation cohort had 30%. **(K,L)** All patients (*N* = 161, 1 < age<99, TBSA>30%, 280 h < sample times <706) in GSE77791 and GSE19743 were divided into two cohorts randomly to train and validate the regression model, the training cohort had 70% patients, and the validation cohort had 30%.

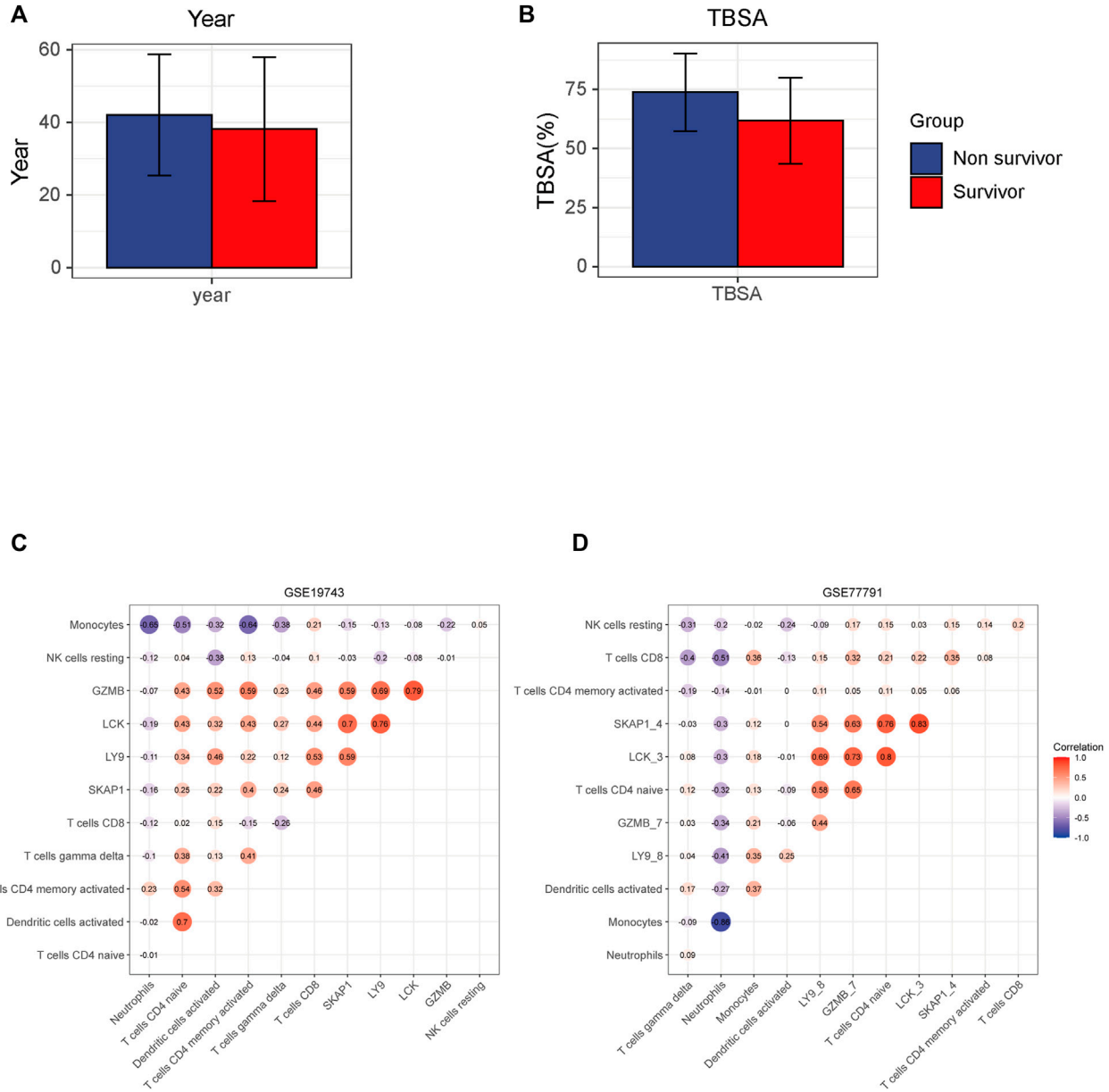


FIGURE 8 | Results of the correlation between key genes and immune cells. **(A,B)** Different TBSA and years between survival and non-survival patients. Non-survival is older and has a larger TBSA. **(C,D)** Horizontal and ordinate are the names of the genes and cells. The figure in the circle means correlation and red means positive correlation and blue means negative.

TABLE 3 | Differences between high- and low-risk burns that were divided by the regression model in inhalation injury.

Group	In inhalation injury		Value of chi-square	P
	Yes	No		
High risk	6	9	0.216	0.624
Low risk	14	28		

DISCUSSION

Physiological characteristics of the inflammatory response and immunosuppression in severe burns are the disorder of the number and proportion of immune cells. After severe burns, monocytes, macrophages, and neutrophils are activated. DAMP and PAMP recognize TLR to activate NF-κB. NF-κB is involved in the activation of inflammatory factors such as IL-1, IL-6, IL-8,

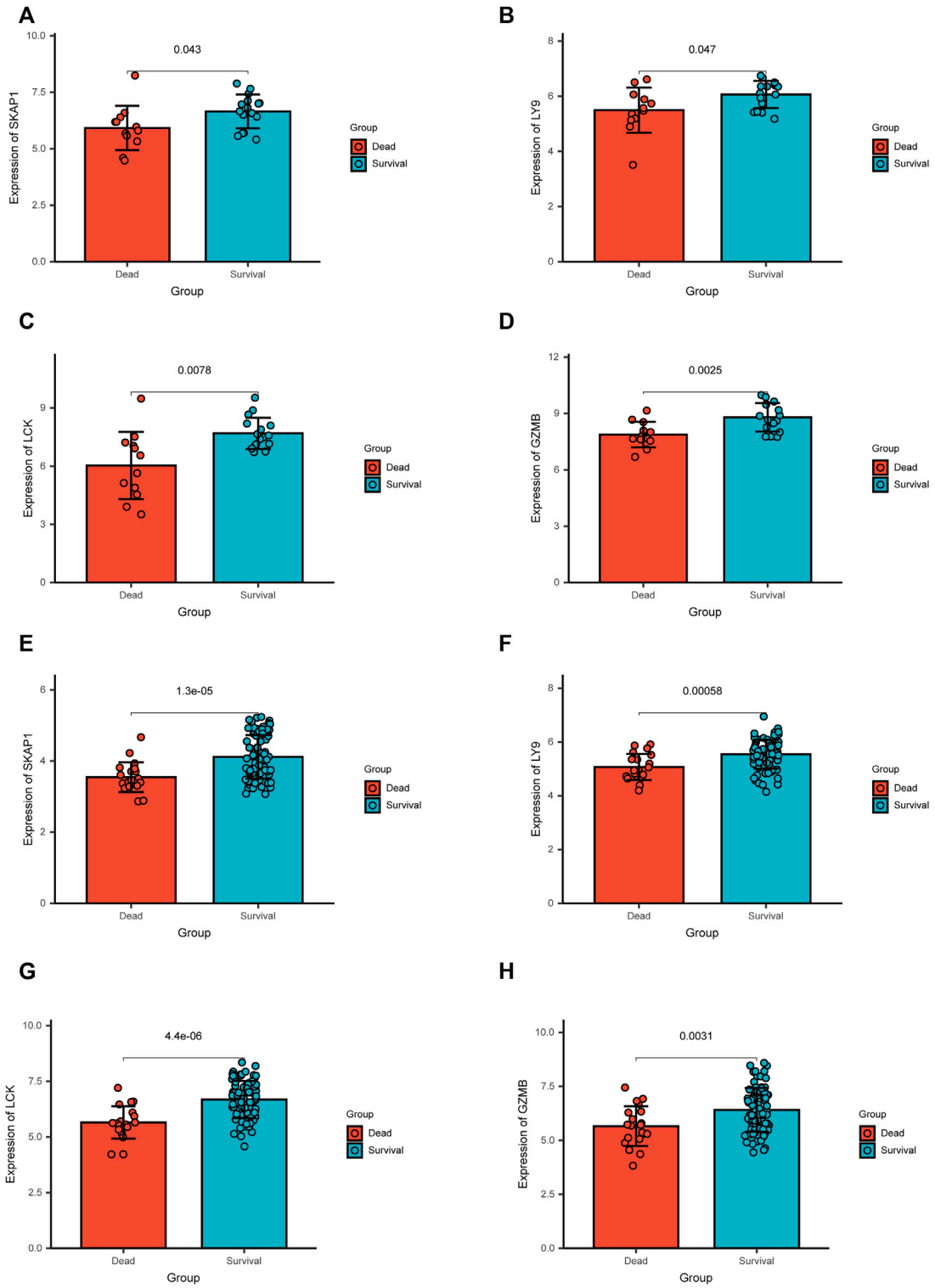


FIGURE 9 | Results of the differential expression analysis of the microarray data in two independent cohorts.**(A–D)** In GSE19743, key genes, SKAP1, LY9, LCK, and GZMB, were up-regulated in survival patients. **(E–H)** In GSE77791, key genes, SKAP1, LY9, LCK, and GZMB, were up-regulated in survival patients.

IL-18, and TNF, resulting in a strong inflammatory response (D'Arpa and Leung 2017). Subsequently, immune function is inhibited. The antigen-presenting function of the macrophages and the killing function of the neutrophils are weakened, followed by the decreasing proliferation of T cells, particularly in the differentiation, proliferation, and function of Th cells (Miller

et al., 2007). The main manifestation is the inhibition of Th1 cell differentiation, and a relative increase of Th2 cell differentiation leads to pro-inflammatory inhibition and anti-inflammatory activation. The changes in cytokines and cell proportion are not only the results of severe burns but also the important causes of immunosuppression and inflammatory response syndrome, which are related to the prognosis of patients and play an important role in the development of immunotherapy targets (Hur et al., 2015). In the GSVA results (**Figure 2C**), we found that the pathway of the T cell, and macrophage (negative regulated) were activated in the controls contrast to severe burns, and the pathways of the neutrophil and monocyte were inhibited, which may be an important reason for the activation of neutrophil, monocyte, and macrophages and the inhibition of T-cell function in severe burns. All the aforementioned data are consistent with the previous research. In addition, we also found that the pathways of B cells, NK cells, and T follicular-assistedpara-cellular were inhibited while dendritic cells were activated. These cells are the key to the immune response, but have not been revealed in the immunosuppression of severe burns.

The intersection of Cibersort and GSVA results showed that the disorder of immune cell subtypes in severe burns included T cell CD8, T cell CD4 naive, T cell CD4 memory resting, T cell CD4 memory activated, NK cell resting, monocytes, macrophages M0, dendritic cells resting, and neutrophils. Subtype disorders of immune cells are an important basis for immune dysfunction. Although the quantitative changes and pathway activation/inhibition of these subtypes have not been studied in the burn, they play an important role in the proliferation, differentiation, and function of T cells, NK cells, mononuclear macrophages, dendritic cells, and neutrophils. The decrease of CD4 naive T cells directly leads to the decrease of Th cells. In our study, we found that CD4 naive T cells decreased in severe burn patients, which could differentiate into Th1, Th2, Th17, and Treg cells (Zhou et al., 2009). In addition to the decrease in cell number, the proportion of different cell subtypes is also imbalanced. For example, Th1/Th2 decreased and Treg/Th17 increased in severe burns, which are important causes of adaptive immune dysfunction (O'Sullivan et al., 1995; MacConmara et al., 2011; Rendon and Choudhry 2012; Valvis et al., 2015). Our study found that CD8T cells and NK cells decreased in severe burns, which are principal cells against pathogens. In the early stages of burns, an inflammatory response will lead to an increase in the number of CD8 cells and NK cells, but a significant decrease will soon follow (Sherwood and Toliver-Kinsky 2004; Patil et al., 2016). Although the reason for the depletion of NK cells has not been found, excessive stress can lead to the depletion of CD8T cells, which may be the reason for the significant reduction of CD8T cells in the mid-burn stage (Sherwood and Toliver-Kinsky 2004; Kurachi 2019). Our study also found that monocytes, macrophages, dendritic cells, and neutrophils were significantly increased in patients with severe burns. The aforementioned cells were the key cells of the inflammatory response, connecting innate immunity, and adaptive immunity. Neutrophils, dendritic cells, and mononuclear macrophages are activated after a burn, causing a strong inflammatory response, which is an important cause of subsequent multiple organ failure and sepsis (Sherwood and Toliver-Kinsky 2004). In addition, over-activated neutrophils will also suppress the function of T cells and affect the adaptive immune response (Aarts et al., 2019). We believe that the changes of these cell subtypes play an important role in the immunosuppression of severe burns, and have a significant prognostic and therapeutic value. Therefore, we used WGCNA to analyze gene changes associated with these immune cells.

Considering that T cells are the main effector cells of adaptive immune response and play an important role in immunosuppression after burns, we selected modules correlated to T cells. Four modules containing differently expressed genes were found to be associated with cell subtypes in the WGCNA analysis (**Figures 4A–D**), with 748 genes related to the immune pathway (**Figures 4E–H**). The related pathways are mainly related to the immune-related gene functions and cell structures such as T cells, immune response, MHC II class protein complex, CD4 receptor, and Ca^{2+} signal pathway (**Figures 5A–D**). The main physiological manifestation of immunosuppression in patients with severe burns is a decrease in the adaptive immune response. T cells are key cells for adaptive immune response. Th1/Th2 ratio is an important factor in immune function. Th-1 produces IL-2 and IFN-γ and activates the immune response. Th-2 produces IL-4 and IL-10 and inhibits the immune response (Abbas et al., 1996). The Ca^{2+} signal pathway is associated with IL-2 production and plays an important role in immune function (Sayeed 1996). In addition, Th17 secretes IL-22 to active epithelial cells, participating in chemotaxis, tissue repair, and antimicrobial peptide expression to prevent bacterial invasion and epithelial cell proliferation and differentiation (Rendon and Choudhry 2012). This effect of Th-17 cells is important because severe burns can induce mucosal atrophy and apoptosis, as well as damage to the homeostasis of intestinal epithelial cells (Magnotti and Deitch 2005). Intestinal mucosal barrier is impaired as early as 5 minutes after severe burns, which increases the risk of bacterial translocation and sepsis. Th-17 cells have been proved to be able to prevent local and systemic proliferation of common infection sources after burning, such as *Bacteroides* fragileus, *Klebsiella pneumoniae*, and *Candida albicans* (Rendon and Choudhry 2012; Rani et al., 2014). T cells are not only an important manifestation of immunosuppression, but also an important therapeutic target for improving immune function. IL-15 treatment can improve the reduction of CD4 + T (Th) cells. Blocking CD47/CD172 signaling pathway can improve the reduction of CD4 + T cells and CD8 + T cells, thereby improving immune function (Patil et al., 2016; Beckmann et al., 2020). Flt3 ligand treatment can alleviate T-cell dysfunction and significantly improve the prognosis of septic mice. Therefore, we believe that T-cell-related genes play an important role in the development of T-cell function inhibition, have an important prognostic value, and are likely to be targeted for immunotherapy.

Among 210 immune-related genes in PPI network, 53 genes are hub genes (**Figures 6A–D**). Twenty-six genes were finally selected by LASSO regression, of which seven genes had significant indigenous prognosis (AUC > 0.7, $p < 0.05$), namely LCK, SKAP1, CX3CR1, FYN, GZMB, LY9, and FYN. The genes were incorporated into the logistic regression to construct the regression that included 4 key genes, LCK, SKAP1, GZMB, and LY9 (**Figures 7A–D**). In the GSE77791 (training cohort), AUC = 0.930 (**Figure 7F**), and calibration curve indicated that the model had an excellent prediction ability (**Figure 7H**). In the GSE19743 (validation cohort), AUC = 0.919 (**Figure 7G**). We included age and TBSA in the regression model for all patients (161), $AUC^{\text{risk score+TBSA+age}} = 0.945 > AUC^{\text{risk score}} = 0.933$ (**Figure 7L**). The prediction ability of the model was improved. In addition, there were significant differences in the burn area and age between survival and non-survival patients (**Figures 8A,B**). So, the prognostic function of our regression includes the effects of burn area and age.

LCK, SKAP1, GZMB, and LY9 are related to T cells, such as T cells CD4 naive, T cells CD4 memory activated, and T cells CD8. The correlation between gene expression and cell proportion indicates that genes may be potential biomarkers that characterize the number and function of cells.

The protein encoded by the LCK gene is a key molecule for differentiation and maturation of developing T cells. LCK exists in all normal T cells. In the cells, LCK is located in the plasma membrane and vesicles around the centrosome, which is related to the cytoplasmic tail of CD4 co-receptors on helper T cells and CD8 co-receptors on cytotoxic T cells, to help T-cell receptor (TCR) complexes signal and participate in the TCR-mediated T-cell activation (Shebzukhov et al., 2017). Human somatic cell experiments showed that the inhibition of LCK expression led to the inhibition of the TCR pathway, thereby hindering the differentiation and development of T cells. Targeted destruction of LCK can lead to T-cell development stagnation in the thymus (van Oers et al., 1996). Although there are few studies on the LCK gene in severe burns, the expression of the LCK gene is significantly related to T-cell subtypes, which is an important molecule to characterize the number and activity of T cells. In addition, therapies targeting LCK have been shown to promote/inhibit T-cell growth and development in a variety of diseases such as type 1 diabetes, colon cancer, asthma, and organ transplant rejection, thereby altering disease outcomes (Sabat et al., 2006; Gholamin et al., 2015). Therefore, LCK, depending on its correlation with T cells, is expected to provide a predictive value for T-cell function and become an important target for the treatment of T-cell dysfunction in severe burns.

SKAP1 gene encodes T-cell adaptor protein which is a key regulator of TCR-mediated LFA-1 signaling (inside-out/outside-in signaling), T-cell receptor-induced activation of LFA-1 to promote T-cell adhesion and interaction with antigen-presenting cells (APCs) (Witte et al., 2017). SKAP1 deficiency affects TCR activation (Lim et al., 2016). The expression of SKAP1 was correlated with T-cell function and disease development. In SKAP1 deficient mice, it was found that IL17 cytokines decreased and T-cell infiltration decreased, which alleviated collagen-induced osteoarthritis (Smith et al., 2016). Th17 is an important mucosal immune cell, which has an important relationship with intestinal flora translocation after burns. SKAP1 deficiency may lead to Th17 deficiency and promote the development of the disease. Although SKAP1 has not been studied in immunosuppression after burns, we believe that SKAP1 can characterize T-cell function and is a promising immunotherapy target for improving T-cell function.

GZMB encodes Granzyme B which is mainly secreted by natural killer (NK) cells and cytotoxic T lymphocytes (CTLs). Granzyme B induces target cell apoptosis and can impact processes such as tissue remodeling, barrier function, autoantigen generation, and angiogenesis. It plays an important role in wound healing, chronic inflammation, and scar formation (Śledzińska et al., 2020). Therefore, the expression of GZMB reflects the differentiation of T-cell subtypes to some extent. LY9 encodes a homocellular surface receptor that exists on all thymocytes and is highly expressed on innate lymphocytes such as iNKT cells. LY9 plays an important role in maintaining T-cell subtype differentiation. The level of IL-4 in LY9-deficient mice was significantly increased, and IL-4 was mainly secreted by Th2 cells, which inhibited the inflammation and immune responses (Cuenca et al., 2018). In our experiment, we found that LY9 was significantly down-regulated, which may be one of the molecular mechanisms of the Th cell subtype disorder. Although GZMB and LY9 have not been further studied in the immunosuppression of severe burns, the proteins encoded by GZMB and LY9 play an important role in T-cell immune function, T-cell subtype differentiation, and wound healing. Obviously, GZMB and LY9 can be used as prognostic factors which can characterize physiological changes after severe burns.

Key genes have great potential in post-burn immunosuppression, which will be a meaningful research direction. In addition, in the differential expression analysis, it was found that in the two independent cohorts, the expression of key genes in survival patients was significantly increased (**Figures 9A–H**), which may indicate that the down-regulation of key genes is an important factor leading to immunosuppression and death, which needs further research.

Our experiment is the first to use WGCNA, GSVA, and LASSO regression to construct a gene prognosis model with genes in three severe burns cohorts (185 patients). In contrast to prognostic models for platelets, coagulation disorders, IFN-γ, IL-2, IL-4, Burn Severity Index (ABSI) score, Ryan score, Belgium Outcome Burn Injury (BOBI) score, and modified Baux score, our prognostic model was based on gene expression profile, which had a higher accuracy and was more convenient for clinical operation (Hur et al., 2015; Lip et al., 2019; Geng et al., 2020). Others use bioinformatics methods to study the pathophysiology of severe burns, but most are limited to animal models or have unstable and inaccurate prognostic indicators (Li et al., 2016; Fang et al., 2020; Liu et al., 2021). We first introduced WGCNA,

CIBERSORT, GSVA, and LASSO into the analysis to find prognostic factors from the pathophysiological mechanism of immunosuppression after severe burns, so our prognostic model is more stable and reliable. In addition, we found that key genes were associated with immunosuppression after severe burns and were related to the ratio of specific immune cell types, which provided an important direction for the future development of immunotherapy targets. Of course, our experiment is also insufficient. We need to collect more information about patients, such as whether shock resuscitation or sepsis occurs, to further stabilize our model. Nevertheless, compared with other clinical prognosis models, our model showed a good prognosis ability in collaboration with age and burn area, and the gene expression and prognosis model were verified multi-dimensionally (three large cohorts, sequencing datasets, and multiple groupings).

SUMMARY

Our study found that immunosuppressive-related genes after severe burns had important prognostic value. The prognostic ability of LCK, SKAP1, GZMB, and LY9 in the gene expression profiles of 185 severe burns was superior to the current prognostic models and scale score.

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

CJ: Conceived and designed the study. ZZ, YH, and PW: Consulted the literature and prepared materials. ZZ, PW, RL: Drawn up the manuscript and analysed the data. JL analyzed the data, prepared figures and/or tables, and approved the final draft. YF performed the experiments, prepared figures and/or tables, and approved the final draft.

FUNDING

This project was supported by the Natural Science Foundation of Fujian Province of China (No. 2019J01011); Open project of Provincial Key Laboratory of Union Hospital Affiliated to Fujian Medical University in 2020 (Nos. XHZDSYS202004 and XHZDSYS202005); and Starting Package of Xiang'an Hospital of Xiamen University (No. PM201809170010).

ACKNOWLEDGMENTS

The authors would like to thank the staff of the Department of Burns and Plastic and Wound Repair Surgery, Xiang'an Hospital of Xiamen University, for their help in this study. They would also like to thank Yi Liao of the Eye Institute of Xiamen University for her help in statistics.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.908510/full#supplementary-material

REFERENCES

Aarts, C. E. M., Hiemstra, I. H., Tool, A. T. J., van den Berg, T. K., Mul, E., van Bruggen, R., et al. (2019). Neutrophils as Suppressors of T Cell Proliferation: Does Age Matter? *Front. Immunol.* 10, 2144. doi:10.3389/fimmu.2019.02144

Abbas, A. K., Murphy, K. M., and Sher, A. (1996). Functional Diversity of Helper T Lymphocytes. *Nature* 383, 787–793. doi:10.1038/383787a0

Beckmann, N., Huber, F., Hanschen, M., St. Pierre Schneider, B., Nomellini, V., and Caldwell, C. C. (2020). Scald Injury-Induced T Cell Dysfunction Can Be Mitigated by Gr1+ Cell Depletion and Blockage of CD47/CD172a Signaling. *Front. Immunol.* 11, 876. doi:10.3389/fimmu.2020.00876

Cuenca, M., Puñet-Ortiz, J., Ruart, M., Terhorst, C., and Engel, P. (2018). Ly9 (SLAMF3) Receptor Differentially Regulates iNKT Cell Development and Activation in Mice. *Eur. J. Immunol.* 48, 99–105. doi:10.1002/eji.201746925

D'Arpa, P., and Leung, K. P. (2017). Toll-Like Receptor Signaling in Burn Wound Healing and Scarring. *Adv. Wound Care* 6, 330–343. doi:10.1089/wound.2017.0733

Fang, X., Duan, S.-F., Gong, Y.-Z., Wang, F., and Chen, X.-L. (2020). Identification of Key Genes Associated with Changes in the Host Response to Severe Burn Shock: A Bioinformatics Analysis with Data from the Gene Expression Omnibus (GEO) Database. *Jir* 13, 1029–1041. doi:10.2147/jir.s282722

Fitzwater, J., Purdue, G. F., Hunt, J. L., and O?Keefe, G. E. (2003). The Risk Factors and Time Course of Sepsis and Organ Dysfunction after Burn Trauma. *J. Trauma Inj. Infect. Crit. Care* 54, 959–966. doi:10.1097/01.ta.0000029382.26295.ab

Foth, M., Wouters, J., de Chaumont, C., Dynoodt, P., and Gallagher, W. M. (2016). Prognostic and Predictive Biomarkers in Melanoma: an Update. *Expert Rev. Mol. Diagnostics* 16, 223–237. doi:10.1586/14737159.2016.1126511

Gaetani, L., Blennow, K., Calabresi, P., Di Filippo, M., Parnetti, L., and Zetterberg, H. (2019). Neurofilament Light Chain as a Biomarker in Neurological Disorders. *J. Neurol. Neurosurg. Psychiatry* 90, 870–881. doi:10.1136/jnnp-2018-320106

Gavrielatou, N., Doumas, S., Economopoulou, P., Foukas, P. G., and Psyrri, A. (2020). Biomarkers for Immunotherapy Response in Head and Neck Cancer. *Cancer Treat. Rev.* 84, 101977. doi:10.1016/j.ctrv.2020.101977

Geng, K., Liu, Y., Yang, Y., Ding, X., Tian, X., Liu, H., et al. (2020). Incidence and Prognostic Value of Acute Coagulopathy after Extensive Severe Burns. *J. Burn Care Res.* 41, 544–549. doi:10.1093/jbcr/irz178

Gholamin, M., Bazi, A., and Abbaszadegan, M. R. (2015). Idiopathic Lymphocytopenia. *Curr. Opin. Hematol.* 22, 46–52. doi:10.1097/MOH.0000000000000102

Hampson, P., Dinsdale, R. J., Wearn, C. M., Bamford, A. L., Bishop, J. R. B., Hazeldine, J., et al. (2017). Neutrophil Dysfunction, Immature Granulocytes, and Cell-free DNA Are Early Biomarkers of Sepsis in Burn-Injured Patients. *Ann. Surg.* 265, 1241–1249. doi:10.1097/sla.0000000000001807

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

Hur, J., Yang, H. T., Chun, W., Kim, J.-H., Shin, S.-H., Kang, H. J., et al. (2015). Inflammatory Cytokines and Their Prognostic Ability in Cases of Major Burn Injury. *Ann. Lab. Med.* 35, 105–110. doi:10.3343/alm.2015.35.1.105

Kurachi, M. (2019). CD8+ T Cell Exhaustion. *Semin. Immunopathol.* 41, 327–337. doi:10.1007/s00281-019-00744-5

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

Li, Z., Wang, Q., Yu, H., Zou, K., Xi, Y., Mi, W., et al. (2016). Screening of Key Genes in Severe Burn Injury at Different Stages via Analyzing Gene Expression Data. *J. Burn Care & Res.* 37, e254–e262. doi:10.1097/bcr.0000000000000179

Lim, D., Lu, Y., and Rudd, C. E. (2016). Non-cleavable Talin Rescues Defect in the T-Cell Conjugation of T-Cells Deficient in the Immune Adaptor SKAP1. *Immunol. Lett.* 172, 40–46. doi:10.1016/j.imlet.2016.02.004

Lip, H. T. C., Idris, M. A. M., Imran, F.-H., Azmah, T. N., Huei, T. J., and Thomas, M. (2019). Predictors of Mortality and Validation of Burn Mortality Prognostic Scores in a Malaysian Burns Intensive Care Unit. *BMC Emerg. Med.* 19, 66. doi:10.1186/s12873-019-0284-8

Liu, X., Rong, Y., Huang, D., Liang, Z., Yi, X., Wu, F., et al. (2021). Altered Genes and Biological Functions in Response to Severe Burns. *Biomed. Res. Int.* 2021, 8836243. doi:10.1155/2021/8836243

MacConmara, M. P., Tajima, G., O'Leary, F., Delisle, A. J., McKenna, A. M., Stallwood, C. G., et al. (2011). Regulatory T Cells Suppress Antigen-Driven CD4 T Cell Reactivity Following Injury. *J. Leukoc. Biol.* 89, 137–147. doi:10.1189/jlb.0210082

Magnotti, L. J., and Deitch, E. A. (2005). Burns, Bacterial Translocation, Gut Barrier Function, and Failure. *J. Burn Care & Rehabilitation* 26, 383–391. doi:10.1097/01.bcr.0000176878.79267.e8

Miller, A. C., Rashid, R. M., and Elamin, E. M. (2007). The "T" in Trauma: the Helper T-Cell Response and the Role of Immunomodulation in Trauma and Burn Patients. *J. Trauma* 63, 1407–1417. doi:10.1097/ta.0b013e31815b839e

Neely, A. N., Fowler, L. A., Kagan, R. J., and Warden, G. D. (2004). Procalcitonin in Pediatric Burn Patients: an Early Indicator of Sepsis? *J. Burn Care Rehabil.* 25, 76–80. doi:10.1097/01.bcr.0000105095.94766.89

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

O'Sullivan, S. T., Lederer, J. A., Horgan, A. F., Chin, D. H., Mannick, J. A., and Rodrick, M. L. (1995). Major Injury Leads to Predominance of the T Helper-2 Lymphocyte Phenotype and Diminished Interleukin-12 Production Associated with Decreased Resistance to Infection. *Ann. Surg.* 222, 482–2. doi:10.1097/00000658-199522240-00006

Osuka, A., Ishihara, T., Shimizu, K., Shintani, A., Ogura, H., and Ueyama, M. (2019). Natural Kinetics of Blood Cells Following Major Burn: Impact of Early Decreases in White Blood Cells and Platelets as Prognostic Markers of Mortality. *Burns* 45, 1901–1907. doi:10.1016/j.burns.2019.07.014

Patil, N. K., Luan, L., Bohannon, J. K., Guo, Y., Hernandez, A., Fensterheim, B., et al. (2016). IL-15 Superagonist Expands mCD8+ T, NK and NKT Cells after Burn Injury but Fails to Improve Outcome during Burn Wound Infection. *PLoS One* 11, e0148452. doi:10.1371/journal.pone.0148452

Rani, M., Zhang, Q., and Schwacha, M. G. (2014). Burn Wound γδ T-Cells Support a Th2 and Th17 Immune Response. *J. Burn Care & Res.* 35, 46–53. doi:10.1097/01.bcr.0000440705.91099.cc

Rendon, J. L., and Choudhry, M. A. (2012). Th17 Cells: Critical Mediators of Host Responses to Burn Injury and Sepsis. *J. Leukoc. Biol.* 92, 529–538. doi:10.1189/jlb.0212083

Sabat, M., Vanrens, J. C., Brugel, T. A., Maier, J., Laufersweiler, M. J., Golebiowski, A., et al. (2006). The Development of Novel 1,2-Dihydro-Pyrimido[4,5-C] pyridazine Based Inhibitors of Lymphocyte Specific Kinase (Lck). *Bioorg. Med. Chem. Lett.* 16, 4257–4261. doi:10.1016/j.bmcl.2006.05.072

Sandquist, M., and Wong, H. R. (2014). Biomarkers of Sepsis and Their Potential Value in Diagnosis, Prognosis and Treatment. *Expert Rev. Clin. Immunol.* 10, 1349–1356. doi:10.1586/1744666x.2014.949675

Sayeed, M. M. (1996). Alterations in Cell Signaling and Related Effector Functions in T Lymphocytes in Burn/trauma/septic Injuries. *Shock* 5, 157–166. doi:10.1097/00024382-199603000-00001

Shebzukhov, Y. V., Stanislawiak, S., Bezhaeva, T. R., Nedospasov, S. A., and Kuprash, D. V. (2017). Low Level of Lck Kinase in Th2 Cells Limits Expression of CD4 Co-receptor and S73 Phosphorylation of Transcription Factor C-Jun. *Sci. Rep.* 7, 2339. doi:10.1038/s41598-017-02553-y

Sherwood, E. R., and Toliver-Kinsky, T. (2004). Mechanisms of the Inflammatory Response. *Best Pract. Res. Clin. Anaesthesiol.* 18, 385–405. doi:10.1016/j.bpa.2003.12.002

Śledzińska, A., Vila de Mucha, M., Bergerhoff, K., Hotblack, A., Demane, D. F., Ghorani, E., et al. (2020). Regulatory T Cells Restrain Interleukin-2- and Blimp-1-dependent Acquisition of Cytotoxic Function by CD4+ T Cells. *Immunity* 52, 151–e6. e156. doi:10.1016/j.immuni.2019.12.007

Smith, X., Taylor, A., and Rudd, C. E. (2016). T-cell Immune Adaptor SKAP1 Regulates the Induction of Collagen-Induced Arthritis in Mice. *Immunol. Lett.* 176, 122–127. doi:10.1016/j.imlet.2016.04.007

Sood, R. F., Gibran, N. S., Arnoldo, B. D., Gamelli, R. L., Herndon, D. N., and Tompkins, R. G. (2016). Early Leukocyte Gene Expression Associated with Age, Burn Size, and Inhalation Injury in Severely Burned Adults. *J. Trauma Acute Care Surg.* 80, 250–257. doi:10.1097/ta.0000000000000905

Valvis, S. M., Waithman, J., Wood, F. M., Fear, M. W., and Fear, V. S. (2015). The Immune Response to Skin Trauma Is Dependent on the Etiology of Injury in a Mouse Model of Burn and Excision. *J. Investigative Dermatology* 135, 2119–2128. doi:10.1038/jid.2015.123

van Oers, N. S. C., Lowin-Kropf, B., Finlay, D., Connolly, K., and Weiss, A. (1996). Aβ T Cell Development Is Abolished in Mice Lacking Both Lck and Fyn Protein Tyrosine Kinases. *Immunity* 5, 429–436. doi:10.1016/s1074-7613(00)80499-9

Witte, A., Meineke, B., Sticht, J., Philipsen, L., Kuropka, B., Müller, A. J., et al. (2017). D120 and K152 within the PH Domain of T Cell Adapter SKAP55 Regulate Plasma Membrane Targeting of SKAP55 and LFA-1 Affinity Modulation in Human T Lymphocytes. *Mol. Cell Biol.* 37, e00509–16. doi:10.1128/MCB.00509-16

Wrba, L., Palmer, A., Braun, C. K., and Huber-Lang, M. (2017). Evaluation of Gut-Blood Barrier Dysfunction in Various Models of Trauma, Hemorrhagic Shock, and Burn Injury. *J. Trauma Acute Care Surg.* 83, 944–953. doi:10.1097/ta.0000000000001654

Zhou, L., Chong, M. M. W., and Littman, D. R. (2009). Plasticity of CD4+ T Cell Lineage Differentiation. *Immunity* 30, 646–655. doi:10.1016/j.immuni.2009.05.001

Zou, J., and Wang, E. (2019). Cancer Biomarker Discovery for Precision Medicine: New Progress. *Curr. Med. Chem.* 26, 7655–7671. doi:10.2174/0929867325666180718164712

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 Genetics

ORIGINAL RESEARCH
published: 14 June 2022
doi: 10.3389/fgene.2022.895587

Identification of Immune-Related Hub Genes in Thymoma: Defects in CD247 and Characteristics of Paraneoplastic Syndrome

Lin-Fang Deng[1,2]*

[1]College of Sciences, Shanghai University, Shanghai, China, [2]College of Medicine, Shanghai University, Shanghai, China

Background: Thymomas (Ts) and thymic carcinomas (TCs) are rare primary tumors of the mediastinum. Paraneoplastic syndrome (PNS) is an important feature of thymoma, which presents great challenges to clinicians.

Methods: The present study uses the weighted gene co-expression network analysis (WGCNA) to identify possible immunologic mechanisms of thymoma. RNA sequencing data from thymoma samples were downloaded from the TCGA. Core genes were taken from the module that is closely related to the WHO's stage of classification. Enhanced analysis using the online database "Metascape" and an overall survival (OS) analysis were carried out via the Kaplan–Meier method. The hub genes were obtained from the protein–protein interaction (PPI) network. In addition, we jointly analyzed multiple sets of PNS data related to thymomas from other sources to verify the correlation between thymomas and PNS. The impact of hub genes on the prognosis of PNS was evaluated via the ROC curve, with simultaneous analysis of immune infiltration by CIBERSORT.

Findings: The 14 immune hub genes closely related to thymomas were found to be jointly involved in the T-cell receptor signaling pathway. Compared to the normal thymus and type B1/B2 thymoma, there is a lower number of T-cells in type A/B3 thymoma and thymic carcinoma. The expression of genes related to the T-cell receptor signaling pathway appeared defective. The low expression of CD247 and the decrease in the number of mature T-cells are common features among thymomas, specific pulmonary fibrosis, rheumatoid arthritis, and systemic lupus erythematosus.

Keywords: Thymomas, immune, WGCNA, paraneoplastic syndrome (PNS), CD247

OPEN ACCESS

Edited by:
Tao Huang,
Shanghai Institute of Nutrition and Health (CAS), China

Reviewed by:
Xin Du,
Capital Medical University, China
Jiankang Yang,
Dali University, China

***Correspondence:**
Lin-Fang Deng
lfdeng@shu.edu.cn

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *14 March 2022*
Accepted: *06 May 2022*
Published: *14 June 2022*

Citation:
Deng L-F (2022) Identification of Immune-Related Hub Genes in Thymoma: Defects in CD247 and Characteristics of Paraneoplastic Syndrome. Front. Genet. 13:895587. doi: 10.3389/fgene.2022.895587

INTRODUCTION

Thymomas (Ts) and thymic carcinomas (TCs) are rare primary tumors of the mediastinum, originating from the thymic epithelium (Scorsetti et al., 2016). The World Health Organization (WHO) divides thymoma into type A, AB, B1, B2, B3, and C according to the morphology of thymoma epithelial cells and the ratio of lymphocytes to epithelial cells in tissues (Marx et al., 2015; Travis et al., 2015). Type C is a thymic carcinoma, and others are thymomas. Thymomas are relatively common primary anterior mediastinal mass, while thymic carcinoma is rare. These two can be synchronous. Thymoma sometimes can develop into carcinomas, but it usually takes 10 to 14 years (Ettinger et al., 2013). In addition to the WHO classification system, the Masaoka staging

system and TNM staging system are also commonly used in clinical thymoma (Ried et al., 2018). Masaoka classifies thymomas into stages I, II (IIa and IIb), III, and IV (IVa and IVb) according to their comprehensive capsule infiltration and tumor metastasis (Masaoka, 2010). However, the TNM staging system includes lymph nodes in the evaluation criteria, which is of great significance for thymoma. In general, there appears to be a clear correlation and correspondence between several staging methods.

Due to the insidious onset and slow progression of thymoma, the first diagnosis is usually found by chance in physical examination, or physical discomfort caused by paraneoplastic syndrome (PNS). PNS is an important feature of thymoma, bringing great challenges to clinicians (Blum et al., 2020). The common PNS of thymoma are myasthenia gravis, total red cell aplasia, polymyositis, systemic lupus erythematosus, rheumatoid arthritis, Cushing syndrome, and syndrome of inappropriate antidiuretic hormone secretion (Rajan and Zhao, 2019; Rubinstein et al., 2019; Hashimoto et al., 2020; Sideris and Huang, 2020). A sizeable percentage (25–40%) of thymoma patients with myasthenia gravis, and more than 15% of patients with fever and immunodeficiency such as syndromes (Zhao et al., 2020; Liao et al., 2021). Thus, deciphering the relationship between this particular clinical symptom and tumor is crucial for the treatment of thymoma. The treatment and management of thymomas include chemoradiotherapy, the use of corticosteroids, immunotherapy, tyrosine kinase inhibitors, and surgical resection. However, the complications, caused by surgery and radiotherapy, such as radiation pericarditis, radiation pneumonia, and pulmonary fibrosis can hasten the patients' death (Ettinger et al., 2013; Jeffrey Yang et al., 2020; Tian et al., 2020; Yu et al., 2020).

In recent years, with the development of systems biology, an increasing number of researchers have explored and predicted the prognostic targets of tumors from the perspectives of gene expression and molecular interaction. However, the etiology of thymoma is still unclear, and the association between PNS and thymoma is also ambiguous. Therefore, this study explores the biochemical mechanism of thymoma from the perspective of systems biology. Weighted correlation network analysis (WGCNA) is a high-quality method for finding clusters (modules) of highly correlated genes (Langfelder and Horvath, 2008; Langfelder and Horvath, 2012), which is a method of data reduction and unsupervised classification. This method can form a module of co-expressed genes to simplify the complex data matrix. Moreover, we can further explore the correlation between gene network and phenotype of concern and explore the hub genes in the network. This method has been widely used since its development. For example, Tian et al. used WGCNA to identify two gene co-expression modules involved in the process of lung squamous cell carcinoma metastasis and suggested that CFTR, SCTR, and FIGF genes could be used as a potential prognostic biomarker (Tian et al., 2017). Magdalena Niemira et al. (2019) applied WGCNA for exploring molecular networks associated with clinical traits such as tumor size, SUV max, BMI, smoking status, recurrence-free survival, and disease-free survival (Niemira et al., 2019). In addition, WGCNA has been applied to breast cancer, liver cancer, colon adenocarcinoma, and other diseases (Yin et al., 2018; Zhou et al., 2018; Wang et al., 2019).

This study analyzes the immunologic mechanism of thymoma by WGCNA. RNA sequencing data from thymoma samples were downloaded from the TCGA. The thymomas datasets GSE177522, GSE57892, and GSE29695 from GEO were used as the validation set to verify the reliability of the results. In addition, we jointly analyze multiple sets of PNS data related to thymomas (GSE33566, GSE93272, and GSE138458) to explore the internal relationship between thymomas and PNS.

MATERIALS AND METHODS

Data Sources and Pre-Processing

RNA sequencing data and clinical information for thymomas were downloaded using TCGAbiolinks, a third-party tool officially recommended by GDC in R language (Colaprico et al., 2016; Silva et al., 2016; Mounir et al., 2019). The TCGAbiolinks tool downloads data through GDC official API. There were 119 cases of thymomas (including 36 cases of type AB thymoma, 15 of type A thymoma, 14 of type B1 thymoma, 30 of type B2 thymoma, 13 of type B3 thymoma, and 11 of thymic carcinomas) and two cases of normal thymus tissue. The validation sets GSE177522, GSE57892, and GSE29695 were downloaded from the GEO official website (https://www.ncbi.nlm.nih.gov/geo/). Some common PNS data were also downloaded from the GEO official website, including rheumatoid arthritis (GSE93272), systemic lupus erythematosus (GSE138458), and pulmonary fibrosis (GSE33566).

The workflow of the study is shown in **Figure 1**.

Screening of Immune Gene and Noise Gene

The purpose of this study is to find out the relationship between thymoma and immunity. Therefore, a total of 2,381 genes in all immune-related pathways were collected from the KEGG website (https://www.genome.jp/kegg/). First, we extract the expression profiles of these immune genes from thymoma data. For details of the gene list, cf. Table 1.XLSX of the **Supplementary Materials** (SM). Second, noise genes were screened by the correlation method. Specifically, the Spearman correlation coefficient matrix of immune genes was calculated. It was determined that the two genes are not correlated when the correlation coefficient r is in the range of (−0.2, 0.2). When a gene is not correlated with 70 % of the remaining genes, it is determined to be a noise gene (cf. SI-1 of the Data Sheet 1.PDF). The data matrix of the optimized immune-related genes is 2023 * 121.

Weighted Gene Co-Expression Network Analysis

This study uses WGCNA (version: 1.70–3), downloaded by BiocManager (version: 1.30.10) in the R suite, to construct the immune co-expression network (Langfelder and Horvath, 2008; Langfelder and Horvath, 2012). After screening and testing the immune expression matrix, the Pearson correlation coefficient matrix is calculated and a suitable soft threshold is selected.

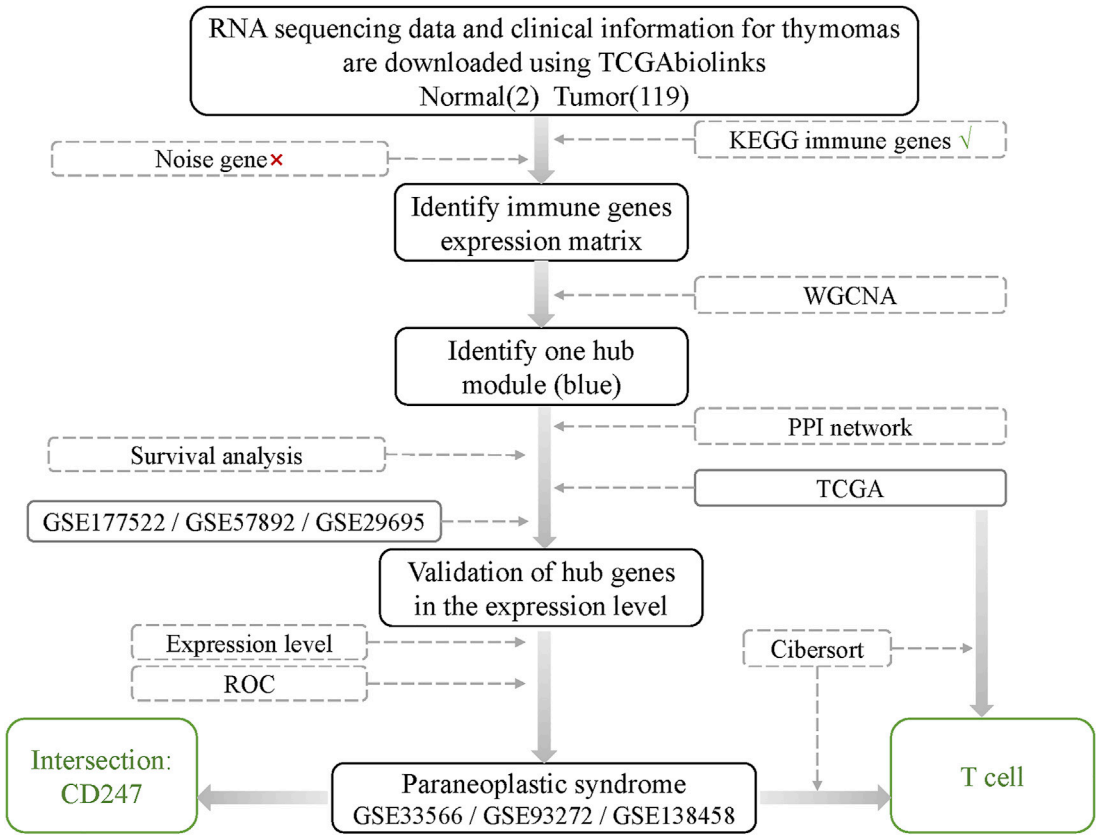


FIGURE 1 | Workflow of searching hub genes in thymoma.

$$c_{ij} = \frac{cov(i,j)}{\sigma_i \sigma_j},$$

where $cov(i, j)$ refers to the covariance of genes i and j, σ represents the standard deviation of gene, and c_{ij} is the Pearson correlation coefficient between genes i and j. Then, the adjacency matrix, a matrix of weighted correlation between genes, was constructed using the power function. This method strengthens the strong correlation and weakens the weak correlation or negative correlation, which makes the correlation value more in line with the scale-free network characteristics and more biological significance.

$$a_{ij} = |c_{ij}|^{\beta},$$

where a_{ij} is adjacency between those two genes. Then, the topological overlap matrix (TOM) was constructed using the adjacency function to reduce noise and false correlation.

$$TOM_{i,j} = \frac{I_{ij} + a_{ij}}{min(k_i + k_j) + 1 - a_{ij}},$$

where I_{ij} is the product's sum of the adjacency coefficients of the nodes connected by genes i and j, and k refers to the sum of the adjacency coefficients of the given gene with all other genes in the weighted network. The TOM is a method to quantitatively describe the similarity in nodes by comparing the weighted correlation between two nodes and other nodes.

Identification of Clinically Significant Modules and Immune Hub Genes in Thymoma

The co-expression module is a collection of genes with high topological overlap similarity. First, principal component analysis (PCA) was used to find the first principal component of each module (module eigengene, ME) to represent the expression pattern of the module. Then, the correlation between these modules and clinical data was calculated to determine the concerned clinical information and gene modules. Module membership (MM) and gene significance (GS) were used to describe the reliability of a gene in the module. The intramodular connectivity may be interpreted as a measure of MM. Genes with intramodular connectivity greater than 0.8 were selected as hub genes (highly connected genes).

Survival Analysis of Hub Genes

The ability of hub genes to predict survival is based on Kaplan–Meier analysis by using the "survival" package (version: 3.2–7) in the R suite. First, the expression profile of hub genes was extracted from TCGA data. Second, the median

expression value of each hub gene was determined. Third, the tumor samples were divided into high-expression groups and low-expression groups with the median of each gene as the boundary. The median sample is divided into high-expression groups. Finally, differences in survival between high- or low-expression groups were assessed for significance by the log-rank test. If $p < 0.05$, we consider the gene to be a validated hub gene.

Functional Enrichment of Hub Genes

In order to analyze the biological functions and signaling pathways of differentially expressed genes in thymoma-related modules, we perform enrichment analysis using the online database "Metascape" (https://metascape.org/gp/index.html#/main/step1) (Zhou et al., 2019). At the same time, protein–protein interaction (PPI) network can be given by Metascape, which uses the molecular complex detection (MCODE) algorithm.

CIBERSORT

CIBERSORT (Newman et al., 2015) is an R package/webpage tool for deconvolution of the expression matrix of human immune cell subtypes based on the principle of linear support vector regression. The proportions of the 22 tumor-infiltrating immune cells (TIICs) from each sample were determined by using the "CIBERSORT" (R package). CIBERSORT was used to analyze the relative expression levels of 547 genes in individual tissue samples according to their gene expression profiles, to predict the proportion of 22 types of TIICs in each tissue. CIBERSORT derived a *p*-value for the deconvolution of each sample, which provided a measure of confidence in the results, and $p < 0.05$ was considered accurate. Significant results ($p < 0.05$) were selected for subsequent analysis.

The CIBERSORT results of TCGA thymoma data were downloaded from the GDC website (https://gdc.cancer.gov/about-data/publications/panimmune) (Thorsson et al., 2018). The CIBERSORT result of GEO data was computed using the "cibersort" package of R (the number of permutations: perm =1,000). For details of the list, see **Supplementary Material**: *results_cibersort.xlsx*.

RESULTS

Immune Gene Expression Profile Data in Thymoma

A total of 2,381 genes related to immune pathways were obtained from the KEGG website. The data matrix of the immune gene expression profile of thymoma was obtained by taking intersection with TCGA thymoma expression profile data, and the matrix dimension was 2247 * 121. Then the noise genes were screened by the correlation method, and the optimized immune gene expression data matrix of thymoma was obtained, with the matrix dimension of 2023 * 121. The optimized matrix avoided the interference of noise in the analysis, and there was a strong correlation between the 2,023 immune genes that remained. The correlation heat map of the immune expression data before and after noise reduction is shown in **Figure 2**.

Construction of Weighted Gene Co-Expression Network

WGCNA was used to construct a network based on the expression matrix of 2023 immune genes and clinical data from 121 thymoma samples.

A dendrogram of samples was clustered by the average linkage method and Pearson's correlation method to check the quality of the data from the 121 samples, and no outliers were identified for removal (**Figure 3A**). To construct a scale-free network, we set the soft threshold power β to 8, and the independence degree to 0.9 (**Figure 3B**). The result showed that there were a total of five co-expression modules; the gray module contains genes that could not be incorporated into any other module (**Figures 3C–E**).

According to the topological overlap matrix (TOM), the connection relationship between genes in each module could be obtained. Different colors indicate that the weight in different modules was greater than 0.35, and the immune gene interaction network of thymoma was visualized by Cytoscape. Among them, LEF1, RHOH, APBB1IP, CD1B, CAMK4, and TCF7 interact with more than 30 immune genes, which are the core genes of the immune network of thymoma, shown as **Figure 4**.

Association of Modules With Clinical Traits and Determination of Core Genes

In addition, we calculated the correlation between module genes and clinical traits of thymoma. For each module, the gene co-expression was summarized by the eigengene, and the correlations of each eigengene with clinical traits were calculated, such as history myasthenia gravis, radiation therapy, gender, OS, OS time, age, tumor, tissue or organ, Masaoka stage, and WHO stage. We found that the blue module had the highest correlation with the WHO stage ($cor = 0.72$, $p = 8e-21$) (**Figures 5A,B**).

Therefore, the blue module was analyzed for core genes. The first 87 intersection genes of the blue module with the highest correlation (intramodular connectivity >0.7) were selected as the core genes for subsequent study and enrichment analysis. For details of the core gene list, cf. Table 4.XLSX of the **Supplementary Material**. The results of functional enrichment analysis were obtained by the online database Metascape; terms with a *p*-value < 0.01, a minimum count of 3, and an enrichment factor >1.5 (the enrichment factor is the ratio between the observed counts and the counts expected by chance) were collected and grouped into clusters based on their membership similarities. We selected the top 20 clusters with their representative enriched terms, mainly involving T-cell activation (GO:0042110), T-cell receptor (TCR) signaling pathway (WP69), human immunodeficiency virus 1 infection (hsa05170), hematopoietic cell lineage (hsa04640), primary immunodeficiency (hsa05340), hemostasis (R-HSA-109582), adaptive immune response (GO:0002250), human T-cell leukemia virus 1 infection (hsa05166), and leukocyte activation involved in immune response (GO:0002366) (**Figure 5C**). Metascape combines DisGeNET, a comprehensive platform integrating information on human disease–associated genes

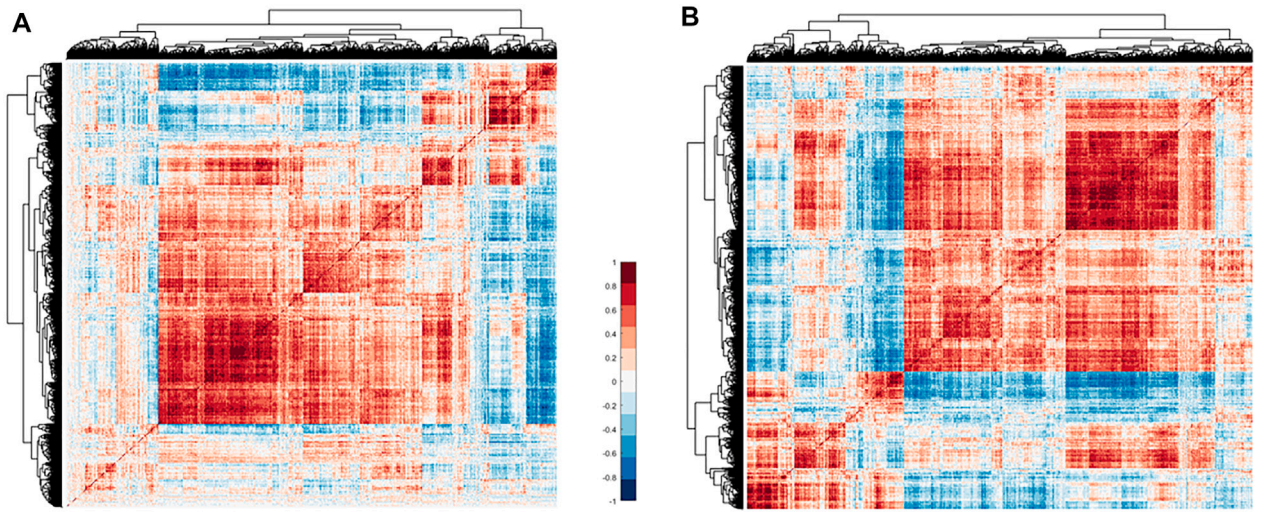


FIGURE 2 | Heat map of immune gene correlation matrix clustering. **(A)** Correlation matrix clustering hot map of 2,247 immune genes in TCGA thymoma data. **(B)** The 2,023 immune gene correlation matrix clustering heat map. Of these, 224 noise genes are eliminated by the correlation method. Red means positive correlation, blue means negative correlation, and white means no correlation.

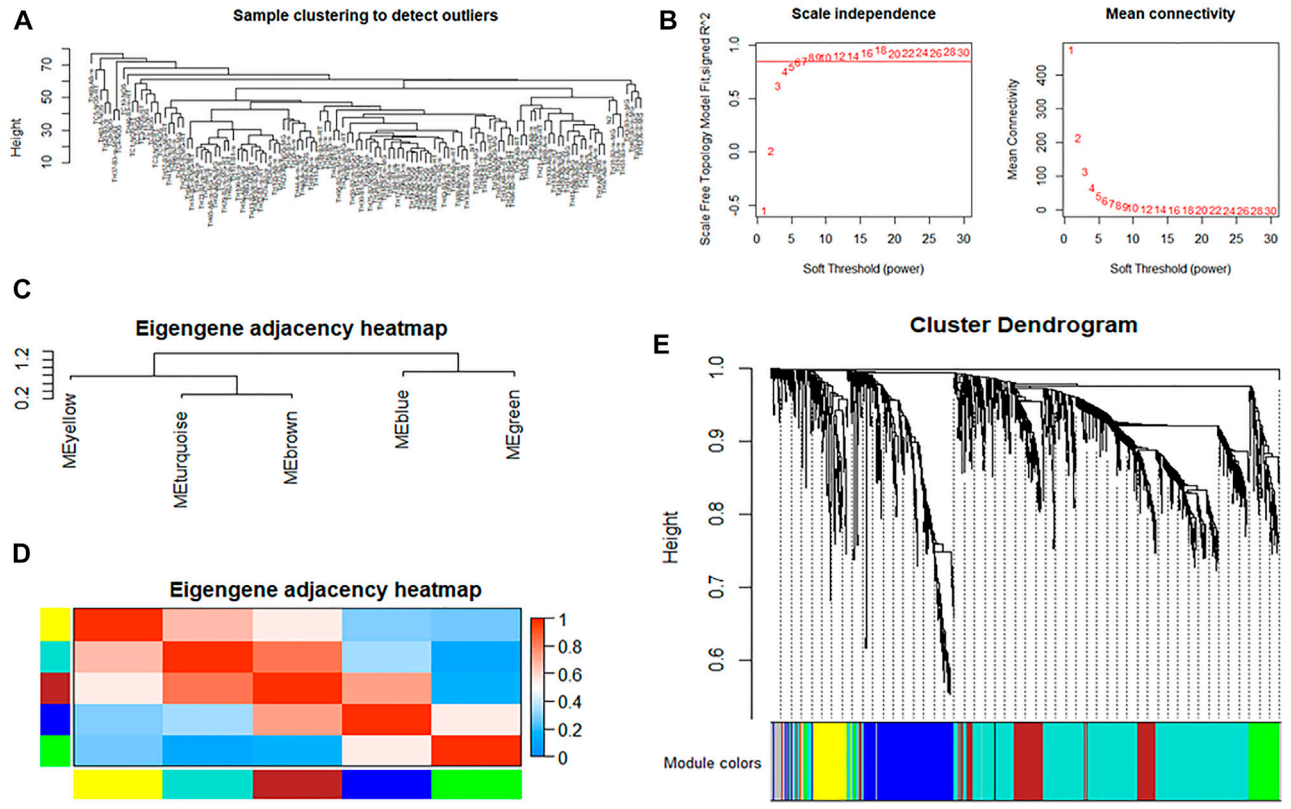


FIGURE 3 | WGCNA of immune genes in thymoma. **(A)** Cluster analysis of samples to detect outliers. **(B)** Determination of soft-thresholding power in weighted gene co-expression network analysis. The left shows the scale-free fit index (*y*-axis) as a function of the soft-thresholding power (*x*-axis). The right shows the average connectivity (degree, *y*-axis) as a function of the soft-thresholding power (*x*-axis). **(C,D)** Module eigengene dendrogram and heatmap of eigengene adjacency. **(E)** Clustering dendrogram of genes, with dissimilarity based on the topological overlap, together with assigned module colors.

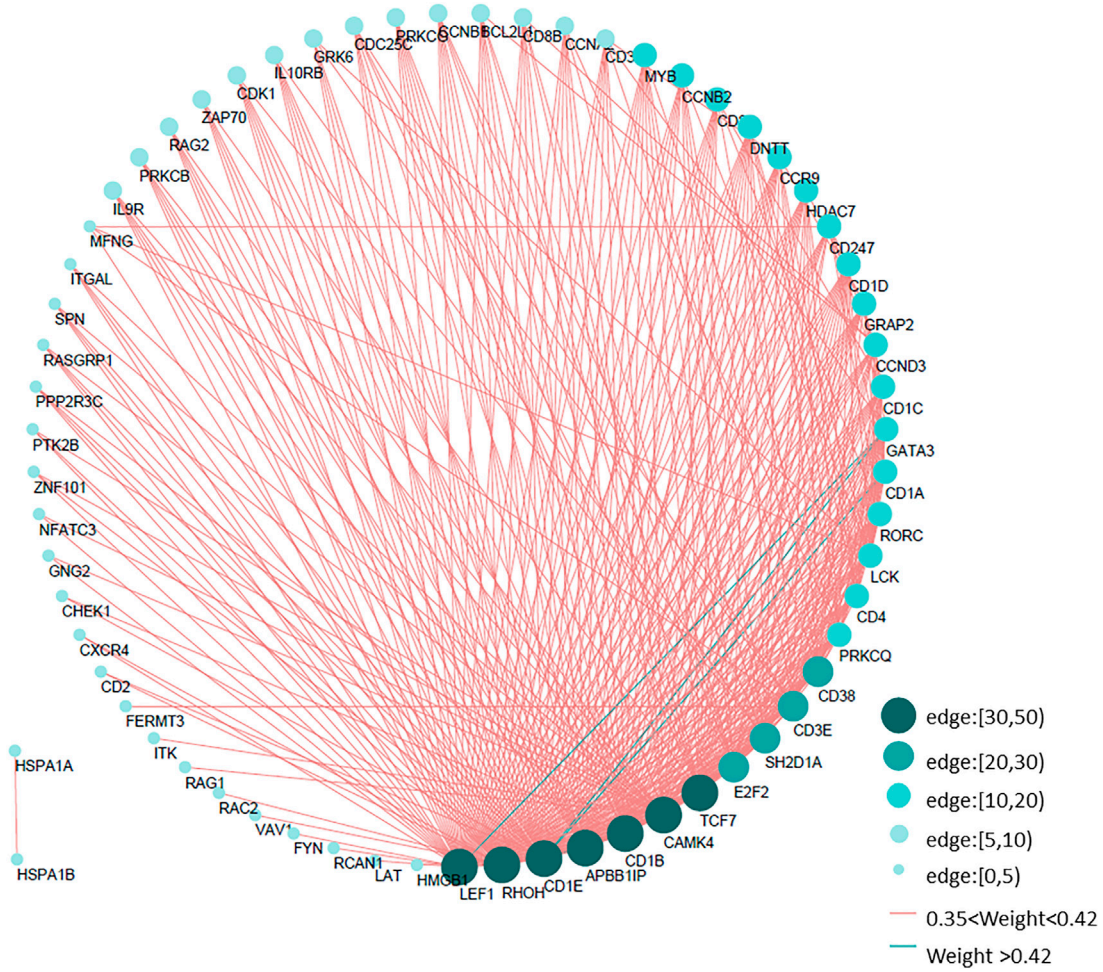


FIGURE 4 | Immune gene interaction network of thymoma. According to the gene interaction table obtained by WGCNA, the edges with a weight greater than 0.35 in five modules are selected to draw the immune network of thymoma. The size of the circle and the depth of the color indicate the different number of edges, cf. Table 3.XLSX of the **Supplementary Material**.

and variants, and gives the disease information related to these core genes (**Figure 5D**), most of which are autoimmune diseases or T-cell-related diseases, such as precursor T-Cell lymphoblastic leukemia-lymphoma, T-cell lymphoma, combined immunodeficiency, and lupus erythematosus. Metascape also gives a summary of enrichment analysis in PaGenBase. PaGenBase is a novel repository for the collection of tissue- and time-specific pattern genes. The results show that most of these core genes were specifically expressed in the thymus (**Figure 5E**), which proved the correctness of these core genes to some extent.

At the same time, the protein–protein interaction (PPI) network of core genes was given by Metascape (**Figure 5F**), which applied a mature complex identification algorithm called MCODE to automatically extract protein complexes embedded in such a large network. Where the T-cell receptor signaling pathway was the core cluster in this PPI network ($log10(p) = -37.4$), the hub genes of this cluster include CD247, CD8A, CD8B, PRKCQ, CD3E, CD3G, GRAP2, VAV1, CD4, LCK, ZAP70, LCP2, ITK, FYN, and LAT.

Validation of Hub Genes in the Expression Level

In the 87 core genes screened by WGCNA, which have a strong correlation with WHO's classification of thymomas, a core cluster, the T-cell receptor signaling pathway, was obtained by PPI enrichment analysis, which contained 15 interrelated hub genes. We observed the expression levels of these 15 genes in TCGA data using a thermal map (**Figure 6A**), and found that these hub genes that were lowly expressed in type A and type B3 thymomas and thymic carcinomas are highly expressed in type B1 and type B2 thymomas. There is no consistent expression pattern in AB thymomas, but thymoma patients with MG have relatively low-expression in AB (cf. SI-2 of the Data Sheet 1.PDF). To verify these results, three groups of thymic tumor data were

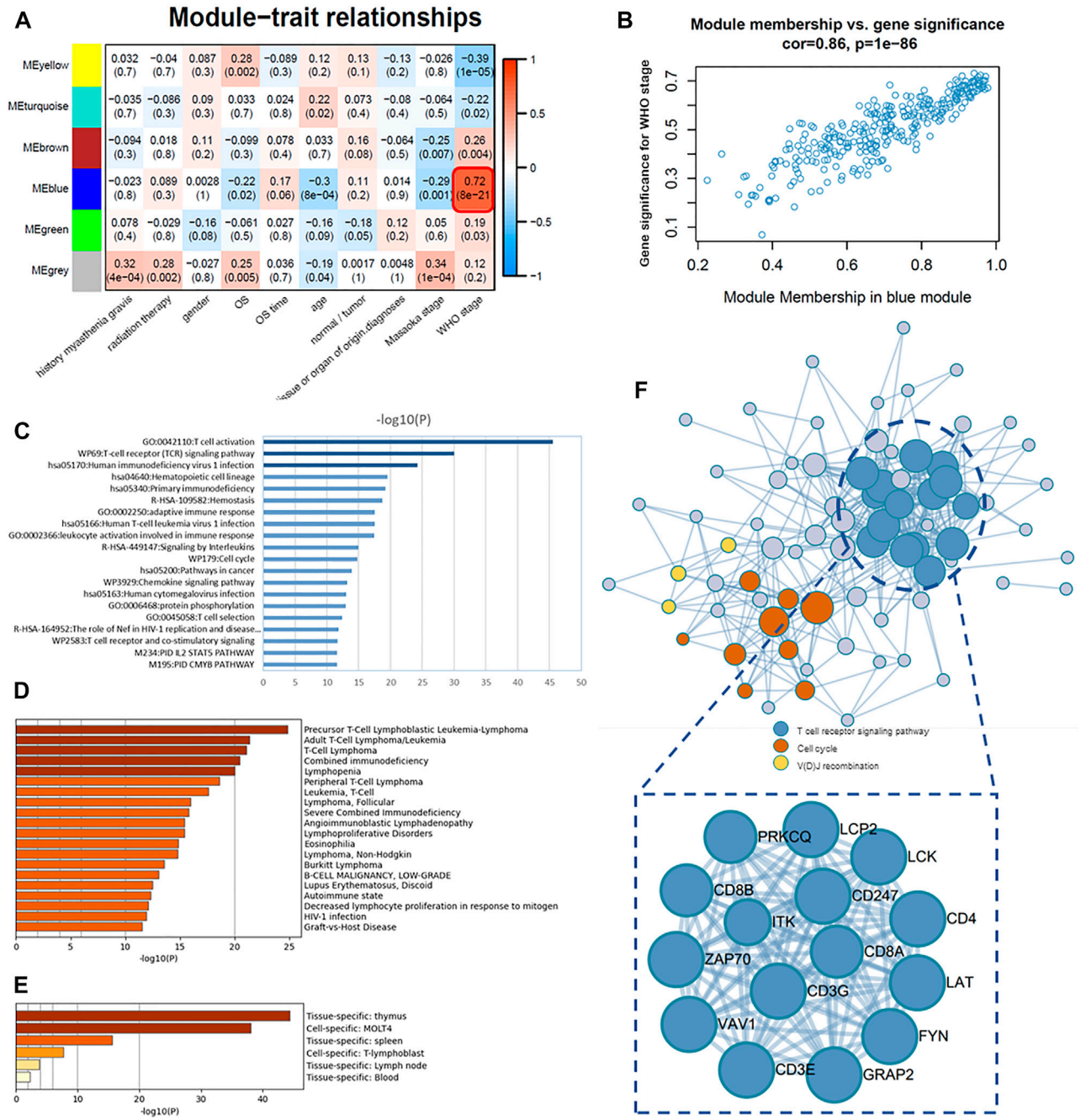


FIGURE 5 | Association of modules with clinical traits and determination of core genes. **(A)** Module–trait associations: each row corresponds to a module eigengene and each column to a trait. Each cell contains the corresponding correlation and *p*-value. **(B)** Scatter plots of GS score and MM for genes in the blue module. **(C)** Pathway and process enrichment analysis: bar graph of enriched terms across 87 core genes (intramodular connectivity >0.7) in the blue module, colored by *p*-values. **(D)** Summary of enrichment analysis in DisGeNET, colored by *p*-values. **(E)** Summary of enrichment analysis in PaGenBase, colored by *p*-values. **(F)** Protein–protein interaction (PPI) network of 87 core genes.

used as validation data, and the data processing method was consistent with TCGA data, which were standardized. In GSE57892, type A and B3 thymoma and thymic carcinoma has almost low expression in these 15 hub genes, while type B2 thymoma has a large expression level (**Figure 6C**). The data GSE177522 also verify the result that the 15 hub genes were almost low expressed in thymic carcinoma (**Figure 6D**). The data GSE29695 clearly show that B3 thymoma cannot be expressed in these 15 hub genes, and the relative expression levels of B1 and B2 were higher (**Figure 6E**).

The WHO stage is mainly based on the morphology of thymoma epithelial cells and the ratio of lymphocytes to

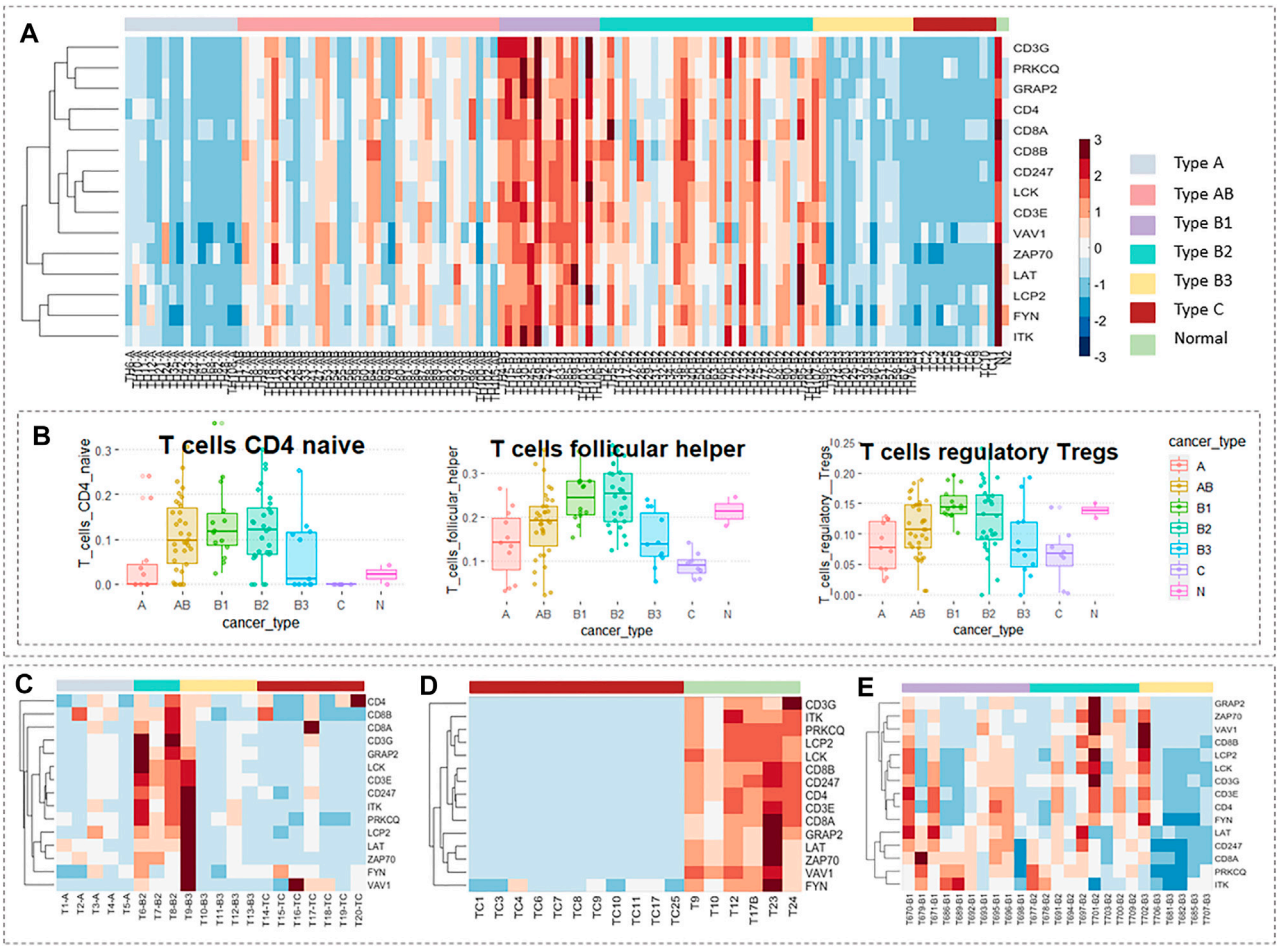


FIGURE 6 | Validation of hub genes in the expression level. Expression levels of 15 hub genes in TCGA data. **(B)** The proportion of T-cells of each thymic tumor subtype in TCGA data obtained by the CIBERSORT algorithm. **(C)** Expression levels of 15 hub genes in GSE57892 data. There are 25 samples in total, including two cases of type AB thymoma, five of type A thymoma, three of type B2 thymoma, five of type B3 thymoma, seven of thymic carcinomas, and three of cell line. **(D)** Expression levels of 15 hub genes in GSE177522 data. There are 19 samples in total, including 11 cases of thymic carcinoids, two cases of thymoma, and six cases of the normal thymus. **(E)** Expression levels of 15 hub genes in GSE29695 data. There are 41 samples in total, including nine cases of type AB thymoma, one of type A thymoma, 10 of type B1 thymoma, nine of type B2 thymoma, six of type B3 thymoma, one of type A/B thymoma, one of type B1/B2 thymoma, and four of the cell line. In particular, cell lines and subtypes of less than three samples were not included in the study.

epithelial cells in tissues. We know that these 15 genes which have a strong correlation with WHO classification belong to the T-cell receptor signaling pathway. Interestingly, the variation trend of the proportions of three T-cells in different thymic tumor subtypes obtained from the CIBERSORT algorithm in **Figure 6B** was very consistent with the variation trend of the expression levels of these 15 hub genes (cf. Table 2.XLSX of the **Supplementary Materials**). A similar trend of expression spectrum and immune infiltration can be used as an inspiration for the phenomenological hypothesis in this group of hub genes.

Certainly, for sake of comprehending whether the 15 hub genes are related to the prognosis of patients, we performed a Kaplan–Meier analysis of these genes according to the clinical data of 119 cases of thymomas in TCGA. It was found that 14 genes were associated with prognosis except FYN ($p < 0.05$). The Kaplan–Meier survival curves of 15 hub genes are given in SI-6 of the Data Sheet 1.PDF.

Taken together, these validation analyses confirm that the more severe the subtype of thymomas is, the less the number of T-cells is, and the expression of related genes in the T-cell receptor signaling pathway is defective.

DISCUSSION

In this study, we obtained 14 hub immune genes of thymomas through statistical analysis. The common correlation pathway of these 14 genes was the T-cell receptor signaling pathway, and their expression was closely related to the WHO stage.

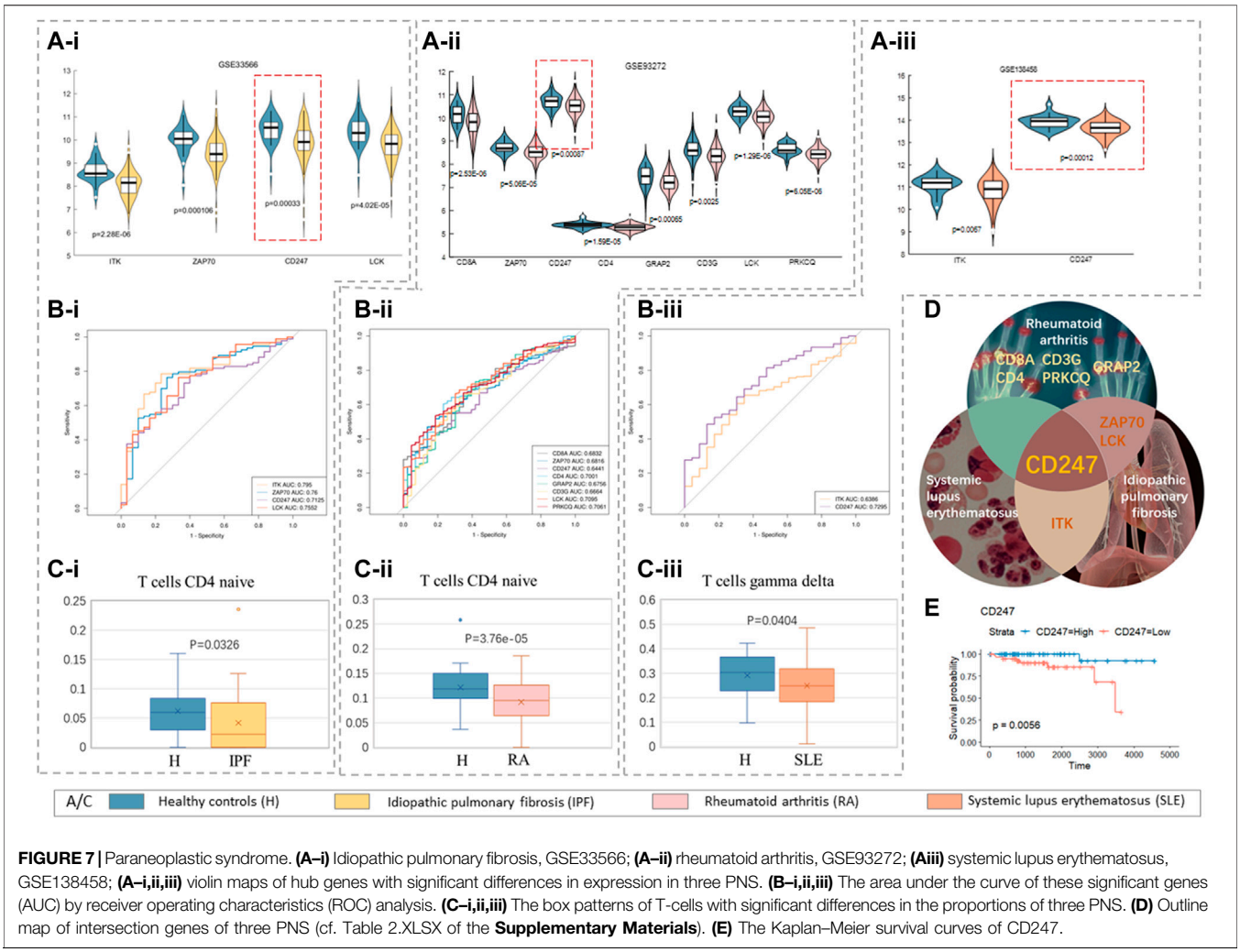


FIGURE 7 | Paraneoplastic syndrome. **(A–i)** Idiopathic pulmonary fibrosis, GSE33566; **(A–ii)** rheumatoid arthritis, GSE93272; **(Aiii)** systemic lupus erythematosus, GSE138458; **(A–i,ii,iii)** violin maps of hub genes with significant differences in expression in three PNS. **(B–i,ii,iii)** The area under the curve of these significant genes (AUC) by receiver operating characteristics (ROC) analysis. **(C–i,ii,iii)** The box patterns of T-cells with significant differences in the proportions of three PNS. **(D)** Outline map of intersection genes of three PNS (cf. Table 2.XLSX of the **Supplementary Materials**). **(E)** The Kaplan–Meier survival curves of CD247.

It is well known that there are generally no specific symptoms for patients with thymomas, but they may have nonspecific symptoms such as chest pain, chest tightness, palpitation, fatigue, and cough. Therefore, they are often ignored by patients who miss the best opportunity for early diagnosis. The first diagnosis is usually found by chance in physical examination or because of the physical discomfort caused by PNS. Thymoma is closely related to autoimmune disorders, and most of its concomitant PNS are autoimmune diseases. Therefore, in this study, we selected two kinds of autoimmune PNS, namely, systemic lupus erythematosus (GSE138458) and rheumatoid arthritis (GSE93272), as the objects of discussion. At the same time, specific pulmonary fibrosis (GSE33566), a severe complication that often occurs after thymic tumor surgery, was also selected. There were 330 samples in dataset GSE138458, including 307 cases of systemic lupus erythematosus (SLE) and 23 cases of healthy control. For dataset GSE93272, there were 232 cases of rheumatoid arthritis (RA) and 43 cases of healthy control. The dataset GSE33566 includes 93 cases of idiopathic pulmonary fibrosis (IPF) and 30 cases of healthy control.

Considering the condition that the *p*-value of survival analysis was less than 0.01 and the *p*-value of expression difference was less than 0.01, we found that ITK, ZAP70, CD247, and LCK were lower expressed in patients with specific pulmonary fibrosis than in healthy controls (**Figure 7A–i**). The expression of CD8 A, ZAP70, CD247, CD4, GRAP2, CD3 G, LCK, and PRKCQ in patients with rheumatoid arthritis was lower than that in healthy controls (**Figure 7A–ii**). Compared with the healthy control group, the expression levels of ITK and CD247 in patients with systemic lupus erythematosus decreases (**Figure 7A–iii**). All hub genes expression levels were significantly different between the highest and lowest quartiles, and the area under the curve (AUCs) of these genes was higher than 0.6 (**Figures 7B–i,ii,iii**), confirming the authenticity of the differences in these genes in their respective PNS. We calculated the proportion of immune cells in patients with three types of by CIBERSORT concomitantly, and found that the proportion of T-cells that were

CD4 naive in patients with specific pulmonary fibrosis and rheumatoid arthritis was significantly different, while the number of T-cells gamma delta in patients with systemic lupus erythematosus was significantly different.

In summary, the decreased proportion of T-cells and the lack of hub gene expression are the common links between thymomas and specific pulmonary fibrosis, rheumatoid arthritis, and systemic lupus erythematosus. The lack of expression of different core genes may be the reason why different patients tend to have different PNS.

Obviously, CD247 was at a low expression level in all three PNS (**Figure 7D**), and the Kaplan–Meier survival curves of CD247 in TCGA data ($p = 0.0056$) were given in (**Figure 7E**). The protein encoded by CD247 (also called CD3ζ) is T-cell receptor zeta. The zeta chain plays an important role in coupling antigen recognition to several intracellular signal-transduction pathways. Low expression of the antigen results in impaired immune response (Call et al., 2006). Moreover, this gene plays an important role in intrathymic T-cell differentiation, and its lack of expression may lead to the reduction of mature T-cells. Petros Christopoulos et al. (2015) proposed a novel thymoma-associated immunodeficiency in 2015. Its characteristics are an accumulation of CD247-deficient, hyporesponsive naive γδ and αβ T-cells and an increased susceptibility to infections (Christopoulos et al., 2015). In 2018, his team further suggested that deficient CD247 expression was a typical histopathological characteristic of thymomas with cortical features (Christopoulos et al., 2018). Recent evidence has demonstrated that CD247 is a potential T-cell–derived disease severity and prognostic biomarker in patients with idiopathic pulmonary fibrosis (Li et al., 2021). Some studies have shown that CD3ζ plays a vital role in multiple autoimmune diseases, such as the gene expression assays showing that CD3ζ mRNA levels are downregulated in PBMCs of patients with RA when compared to healthy controls (Li et al., 2016). Moreover, available evidence suggests that SLE is associated with a deficiency in a cluster of differentiation 247 (Takeuchi and Suzuki, 2013). CD247 is shared by various autoimmune disorders and supports a common T-cell–mediated mechanism. A classic T-cell phenotype in SLE is the downregulation and replacement of the CD3ζ chain that alters T-cell receptor signaling (Martins et al., 2015).

In summary, the common characteristics of thymomas and these three PNS are the low expression of CD247 and the inhibition of T-cell differentiation.

CONCLUSION

In this study, 14 hub immune genes closely related to thymomas, jointly involved in the T-cell receptor signaling pathway, were found by analysis of the expression data of immune genes in thymomas. Compared with normal thymus and type B1/ B2 thymoma, the number of T-cells in type A/B3 thymoma and thymic carcinoma is less, and the expression of genes related to T-cell receptor signaling pathway is low. Then, we also discussed the expression of these 14 core genes in three PNS, and found that CD247 has not only minimal expression in multiple subtypes of thymomas, but also low expression in specific pulmonary fibrosis, rheumatoid arthritis, and systemic lupus erythematosus. The low expression of CD247 and the decrease in the number of mature T-cells are the common features of thymomas and these three PNS.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found here: TCGA-THYM https://portal.gdc.cancer.gov/ GSE177522, GSE57892, GSE29695, GSE93272, GSE138458, and GSE33566 https://www.ncbi.nlm.nih.gov/geo/.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor to this work and has approved it for publication.

ACKNOWLEDGMENTS

I thank Prof. Yong-Cong Chen for his critical reading and suggestions on various parts of the work. I thank Dr. Zhang He-Yao for his support and encouragement. I am grateful to Dr Shi Chun-Xiao, Dr Cheng Run-Tan, and Dr Yao Meng-Chao for valuable discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.895587/full#supplementary-material

REFERENCES

Blum, T. G., Misch, D., Kollmeier, J., Thiel, S., and Bauer, T. T. (2020). Autoimmune Disorders and Paraneoplastic Syndromes in Thymoma. *J. Thorac. Dis.* 12 (12), 7571–7590. doi:10.21037/jtd-2019-thym-10

Call, M. E., Schnell, J. R., Xu, C., Lutz, R. A., Chou, J. J., and Wucherpfennig, K. W. (2006). The Structure of the ζζ Transmembrane Dimer Reveals Features Essential for its Assembly with the T Cell Receptor. *Cell* 127 (2), 355–368. doi:10.1016/j.cell.2006.08.044

Christopoulos, P., Chung, I., Bozorgmehr, F., Muley, T., Meister, M., Kobinger, S., et al. (2018). Deficient CD247 Expression Is a Typical Histopathological Characteristic of Thymomas with Cortical Features. *Histopathology* 73 (6), 1040–1043. doi:10.1111/his.13724

Christopoulos, P., Dopfer, E. P., Malkovsky, M., Esser, P. R., Schaefer, H.-E., Marx, A., et al. (2015). A Novel Thymoma-Associated Immunodeficiency with Increased Naive T Cells and Reduced CD247 Expression. *J. Immunol.* 194 (7), 3045–3053. doi:10.4049/jimmunol.1402805

Colaprico, A., Silva, T. C., Olsen, C., Garofano, L., Cava, C., Garolini, D., et al. (2016). TCGAbiolinks: an R/Bioconductor Package for Integrative Analysis of TCGA Data. *Nucleic Acids Res.* 44 (8), e71. doi:10.1093/nar/gkv1507

Ettinger, D. S., Riely, G. J., Akerley, W., Borghaei, H., Chang, A. C., Cheney, R. T., et al. (2013). Thymomas and Thymic Carcinomas: Clinical Practice Guidelines

in Oncology. *J. Natl. Compr. Canc Netw.* 11 (5), 562–576. doi:10.6004/jnccn.2013.0072

Hashimoto, S., Hayasaka, K., Suzuki, K., Endoh, M., Yanagawa, N., and Shiono, S. (2020). Thymic Small Cell Carcinoma Associated with Lambert-Eaton Myasthenic Syndrome. *Ann. Thorac. Surg.* 109 (5), e347–e348. doi:10.1016/j.athoracsur.2019.08.080

Jeffrey Yang, C-F., HurdYang, J., Shah, S. A., Liou, D., Wang, H., Backhus, L. M., et al. (2020). A National Analysis of Open versus Minimally Invasive Thymectomy for Stage I to III Thymoma. *J. Thorac. Cardiovasc Surg.* 160 (2), 555–e15. doi:10.1016/j.jtcvs.2019.11.114

Langfelder, P., and Horvath, S. (2012). Fast R Functions for Robust Correlations and Hierarchical Clustering. *J. Stat. Softw.* 46 (11), i11. doi:10.18637/jss.v046.i11

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

Li, P., Wang, X., Zhao, M.-Q., Li, L.-J., Zhang, C., Li, B.-Z., et al. (2016). TCR-CD3ζ Gene Polymorphisms and Expression Profile in Rheumatoid Arthritis. *Autoimmunity* 49 (7), 466–471. doi:10.1080/08916934.2016.1174855

Li, Y., Chen, S., Li, X., Wang, X., Li, H., Ning, S., et al. (2021). CD247, a Potential T Cell-Derived Disease Severity and Prognostic Biomarker in Patients with Idiopathic Pulmonary Fibrosis. *Front. Immunol.* 12, 762594. doi:10.3389/fimmu.2021.762594

Liao, C.-H., Lyu, S.-Y., Chen, H.-C., Chang, D.-M., and Lu, C.-C. (2021). Thymoma-Related Paraneoplastic Syndrome Mimicking Reactive Arthritis. *Medicina* 57 (9), 932. doi:10.3390/medicina57090932

Martins, M., Williams, A. H., Comeau, M., Marion, M., Ziegler, J. T., Freedman, B. I., et al. (2015). Genetic Association of CD247 (CD3ζ) with SLE in a Large-Scale Multiethnic Study. *Genes Immun.* 16 (2), 142–150. doi:10.1038/gene.2014.73

Marx, A., Chan, J. K. C., Coindre, J.-M., Detterbeck, F., Girard, N., Harris, N. L., et al. (2015). The 2015 World Health Organization Classification of Tumors of the Thymus: Continuity and Changes. *J. Thorac. Oncol.* 10 (10), 1383–1395. doi:10.1097/JTO.0000000000000654

Masaoka, A. (2010). Staging System of Thymoma. *J. Thorac. Oncol.* 5 (10 Suppl. 4), S304–S312. doi:10.1097/JTO.0b013e3181f20c05

Mounir, M., Lucchetta, M., Silva, T. C., Olsen, C., Bontempi, G., Chen, X., et al. (2019). New Functionalities in the TCGAbiolinks Package for the Study and Integration of Cancer Data from GDC and GTEx. *PLoS Comput. Biol.* 15 (3), e1006701. doi:10.1371/journal.pcbi.1006701

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

Niemira, M., Collin, F., Szalkowska, A., Bielska, A., Chwialkowska, K., Reszec, J., et al. (2019). Molecular Signature of Subtypes of Non-small-cell Lung Cancer by Large-Scale Transcriptional Profiling: Identification of Key Modules and Genes by Weighted Gene Co-expression Network Analysis (WGCNA). *Cancers* 12 (1), 37. doi:10.3390/cancers12010037

Rajan, A., and Zhao, C. (2019). Deciphering the Biology of Thymic Epithelial Tumors. *Mediastinum* 3, 36. doi:10.21037/med.2019.08.03

Ried, M., Eicher, M.-M., Neu, R., Kraus, D., Inderhees, S., Marx, A., et al. (2018). Vergleich der Masaoka-Koga-Klassifikation mit der neuen TNM-Klassifikation der IASLC/ITMIG für Thymome und Thymuskarzinome. *Zentralbl Chir.* 143 (S 01), S44–S50. doi:10.1055/a-0606-5603

Rubinstein, M. M., Goss, C., Avecilla, S., Dubé, G. P., Riely, G. J., and Mones, J. V. (2019). Management of Thymoma-associated Pure Red Cell Aplasia: A Novel Use of Blood Substitute HBOC-201 in a Jehovah's Witness. *Clin. Case Rep.* 8 (2), 289–292. doi:10.1002/ccr3.2626

Scorsetti, M., Leo, F., Trama, A., D'Angelillo, R., Serpico, D., Macerelli, M., et al. (2016). Thymoma and Thymic Carcinomas. *Crit. Rev. Oncology/Hematology* 99, 332–350. doi:10.1016/j.critrevonc.2016.01.012

Sideris, A., and Huang, J. (2020). Commentary: Toward a Better Understanding of the Natural History of Surgically Treated Thymoma Associated with Paraneoplastic Syndromes. *J. Thorac. Cardiovasc. Surg.* 160 (1), 316–317. doi:10.1016/j.jtcvs.2019.12.059

Silva, T. C., Colaprico, A., Olsen, C., D'Angelo, F., Bontempi, G., Ceccarelli, M., et al. (2016). TCGA Workflow: Analyze Cancer Genomics and Epigenomics Data Using Bioconductor Packages. *F1000Res* 5, 1542. doi:10.12688/f1000research.8923.2

Takeuchi, T., and Suzuki, K. (2013). CD247 Variants and Single-Nucleotide Polymorphisms Observed in Systemic Lupus Erythematosus Patients. *Rheumatology* 52 (9), 1551–1555. doi:10.1093/rheumatology/ket119

Thorsson, V., Gibbs, D. L., Brown, S., Wolf, D., Bortone, D. S., Ou Yang, T. H., et al. (2018). The Immune Landscape of Cancer. *Immunity* 48 (4), 812–e14. doi:10.1016/j.immuni.2018.03.023

Tian, F., Zhao, J., Fan, X., and Kang, Z. (2017). Weighted Gene Co-expression Network Analysis in Identification of Metastasis-Related Genes of Lung Squamous Cell Carcinoma Based on the Cancer Genome Atlas Database. *J. Thorac. Dis.* 9 (1), 42–53. doi:10.21037/jtd.2017.01.04

Tian, W., Li, X., Tong, H., Weng, W., Yang, F., Jiang, G., et al. (2020). Surgical Effect and Prognostic Factors of Myasthenia Gravis with Thymomas. *Thorac. Cancer* 11 (5), 1288–1296. doi:10.1111/1759-7714.13396

Travis, W. D., Brambilla, E., Burke, A. P., Marx, A., and Nicholson, A. G. (2015). Introduction to the 2015 World Health Organization Classification of Tumors of the Lung, Pleura, Thymus, and Heart. *J. Thorac. Oncol.* 10 (9), 1240–1242. doi:10.1097/JTO.0000000000000663

Wang, C. N., Li, C. Y., Cai, J-H., Sheu, P., Tsai, J., Wu, M., et al. (2019). Identification of Prognostic Candidate Genes in Breast Cancer by Integrated Bioinformatic Analysis. *Jcm* 8 (8), 1160. doi:10.3390/jcm8081160

Yin, L., Cai, Z., Zhu, B., and Xu, C. (2018). Identification of Key Pathways and Genes in the Dynamic Progression of HCC Based on WGCNA. *Genes* 9 (2), 92. doi:10.3390/genes9020092

Yu, L., Zhang, B. X., Du, X., Yu, Z., Yang, X. G., and Jiang, Y. X. (2020). Evaluating the Effectiveness of Chemotherapy for Thymic Epithelial Tumors Using the CD-DST Method. *Thorac. Cancer* 11 (5), 1160–1169. doi:10.1111/1759-7714.13362

Zhao, J., Bhatnagar, V., Ding, L., Atay, S. M., David, E. A., McFadden, P. M., et al. (2020). A Systematic Review of Paraneoplastic Syndromes Associated with Thymoma: Treatment Modalities, Recurrence, and Outcomes in Resected Cases. *J. Thorac. Cardiovasc. Surg.* 160 (1), 306–314. doi:10.1016/j.jtcvs.2019.11.052

Zhou, X.-g., Huang, X.-l., Liang, S.-y., Tang, S.-m., Wu, S.-k., Huang, T.-t., et al. (2018). Identifying miRNA and Gene Modules of Colon Cancer Associated with Pathological Stage by Weighted Gene Co-expression Network Analysis. *Ott* 11, 2815–2830. doi:10.2147/OTT.S163891

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets. *Nat. Commun.* 10 (1), 1523. doi:10.1038/s41467-019-09234-6

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

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 17 June 2022
doi: 10.3389/fgene.2022.919389

Identification of Hypoxia-Related Subtypes, Establishment of Prognostic Models, and Characteristics of Tumor Microenvironment Infiltration in Colon Cancer

Changjing Wang[1†], Yujie Tang[1†], Hongqing Ma[2], Sisi Wei[3], Xuhua Hu[2], Lianmei Zhao[3] and Guiying Wang[1,2]**

[1]Department of Gastrointestinal Surgery, The Third Hospital of Hebei Medical University, Shijiazhuang, China, [2]The Second Department of Surgery, The Fourth Hospital of Hebei Medical University, Shijiazhuang, China, [3]Research Center, The Fourth Hospital of Hebei Medical University, Shijiazhuang, China

OPEN ACCESS

Edited by:
Tao Huang,
Shanghai Institute of Nutrition and Health, (CAS), China

Reviewed by:
Yisha Zhao,
Zhejiang University, China
Jiaxi Lu,
Chongqing University, China

***Correspondence:**
Guiying Wang
wangguiying@hebmu.edu.cn
Lianmei Zhao
lianmeizhmail@163.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: 13 April 2022
Accepted: 25 April 2022
Published: 17 June 2022

Citation:
Wang C, Tang Y, Ma H, Wei S, Hu X, Zhao L and Wang G (2022) Identification of Hypoxia-Related Subtypes, Establishment of Prognostic Models, and Characteristics of Tumor Microenvironment Infiltration in Colon Cancer. Front. Genet. 13:919389. doi: 10.3389/fgene.2022.919389

Background: Immunotherapy is a treatment that can significantly improve the prognosis of patients with colon cancer, but the response to immunotherapy is different in patients with colon cancer because of the heterogeneity of colon carcinoma and the complex nature of the tumor microenvironment (TME). In the precision therapy mode, finding predictive biomarkers that can accurately identify immunotherapy-sensitive types of colon cancer is essential. Hypoxia plays an important role in tumor proliferation, apoptosis, angiogenesis, invasion and metastasis, energy metabolism, and chemotherapy and immunotherapy resistance. Thus, understanding the mechanism of hypoxia-related genes (HRGs) in colon cancer progression and constructing hypoxia-related signatures will help enrich our treatment strategies and improve patient prognosis.

Methods: We obtained the gene expression data and corresponding clinical information of 1,025 colon carcinoma patients from The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO) databases, respectively. We identified two distinct hypoxia subtypes (subtype A and subtype B) according to unsupervised clustering analysis and assessed the clinical parameters, prognosis, and TME cell-infiltrating characteristics of patients in the two subtypes. We identified 1,132 differentially expressed genes (DEGs) between the two hypoxia subtypes, and all patients were randomly divided into the training group (n = 513) and testing groups (n = 512). Following univariate Cox regression with DEGs, we construct the prognostic model (HRG-score) including six genes (S1PR3, ETV5, CD36, FOXC1, CXCL10, and MMP12) through the LASSO–multivariate cox method in the training group. We comprehensively evaluated the sensitivity and applicability of the HRG-score model from the training group and the testing group, respectively. We explored the correlation between HRG-score and clinical parameters, tumor microenvironment, cancer stem cells (CSCs), and MMR status. In order to evaluate the value of the risk model in clinical application, we further analyzed the sensitivity of chemotherapeutics and

immunotherapy between the low-risk group and high-risk group and constructed a nomogram for improving the clinical application of the HRG-score.

Result: Subtype A was significantly enriched in metabolism-related pathways, and subtype B was significantly enriched in immune activation and several tumor-associated pathways. The level of immune cell infiltration and immune checkpoint-related genes, stromal score, estimate score, and immune dysfunction and exclusion (TIDE) prediction score was significantly different in subtype A and subtype B. The level of immune checkpoint-related genes and TIDE score was significantly lower in subtype A than that in subtype B, indicating that subtype A might benefit from immune checkpoint inhibitors. Finally, an HRG-score signature for predicting prognosis was constructed through the training group, and the predictive capability was validated through the testing group. The survival analysis and correlation analysis of clinical parameters revealed that the prognosis of patients in the high-risk group was significantly worse than that in the low-risk group. There were also significant differences in immune status, mismatch repair status (MMR), and cancer stem cell index (CSC), between the two risk groups. The correlation analysis of risk scores with IC_{50} and IPS showed that patients in the low-risk group had a higher benefit from chemotherapy and immunotherapy than those in the high-risk group, and the external validation IMvigor210 demonstrated that patients with low risk were more sensitive to immunotherapy.

Conclusion: We identified two novel molecular subgroups based on HRGs and constructed an HRG-score model consisting of six genes, which can help us to better understand the mechanisms of hypoxia-related genes in the progression of colon cancer and identify patients susceptible to chemotherapy or immunotherapy, so as to achieve precision therapy for colon cancer.

Keywords: colon cancer, hypoxia-related genes, molecular subtype, tumor microenvironment, immunotherapy, immune checkpoint blockade, HRG-score

INTRODUCTION

Colon cancer is the fifth most common malignancy, with more than 1 million new cases every year (Sung et al., 2021). Metastasis and recurrence have always been the main problems leading to refractory colon cancer (Bekaii-Saab et al., 2019; Mayer et al., 2015; Sartore-Bianchi et al., 2016), and about 30–50% of patients with primary colon cancer will relapse and die from metastatic cancer (Arnold et al., 2015; Siegel et al., 2021). Surgical treatment is the main treatment for colon cancer, and the 5-year survival rate is about 50% (Ferlay et al., 2010). The 5-year survival rate for patients with distal metastasis is even worse at about 14%. With the advances in treatments such as surgery, radiation therapy, chemotherapy, and immunotherapy, the survival rate in colon cancer patients has improved significantly (Jahanafrooz et al., 2020). Up to now, the tumor stage has been the most important factor in judging the severity of a patient's disease, specifying treatment strategy, and predicting the prognosis (Compton et al., 2000).

Moreover, given the high heterogeneity in molecular genetics and histopathology, the treatment strategies based on the tumor-node-metastasis staging system may not be effective across all individuals. With the advance in genomic technology, many epigenetic changes have been identified as potential prognostic biomarkers in colon cancer patients, such as aberrant DNA methylation processes, noncoding RNA and microRNA disorders, and histone modification changes (Kandimalla et al., 2021; Vymetalkova et al., 2019). However, genetic changes still play a key role in the progression of colon cancer. Therefore, the construction of prognostic markers based on changes in genes is vital to enable individualized treatment decisions, which may then guide the choice of treatment strategy and the accurate prediction of patient prognosis.

Tumor cells are metabolically active, so hypoxia often occurs in the center. Hypoxia affects the tumor immune microenvironment (TIME) directly and indirectly, with much evidence favoring an immunosuppressive effect (Chouaib et al., 2018; You et al., 2021). For tumor cells, hypoxia enhances angiogenesis and remodeling by inducing hypoxia-inducible factor (HIF) expression, which is a marker of tumor proliferation, metastasis, and recurrence (King et al., 2021). Potential mechanisms include altered gene expression, oncogene activation, inactivation of anti-oncogenes, decreased genome stability, and clonal selection (Emami Nejad et al., 2021).

Under normal oxygen tension, the HIF protein is unstable and easily degraded by proteasome (Semenza et al., 2010; Semenza et al., 2021). In hypoxic cells, HIF proteins are not easily degraded, thereby creating an immune-unfavorable microenvironment by regulating the transcription of downstream genes, ultimately leading to immune resistance (Chouaib et al., 2018; Noman et al., 2019). Hypoxia can regulate the status of the tumor immune microenvironment by promoting the recruitment of innate immune cells and interfering with the differentiation and function of adaptive immune cells (Palazon et al., 2014). For colon cancer, hypoxia also promotes epithelial–mesenchymal transformation (EMT) and ultimately leads to further migration and invasion of tumor cells (Choietal et al., 2017).

In the study, we systematically evaluated the patterns of hypoxia-related genes and tumor immune microenvironment characteristics of COAD patients by clustering the expression of hypoxia genes. We identified two subtypes with distinct clinical and immune characteristics in COAD and constructed an HRG-score signature based on the expression profile of HRGs.HRG-score serves as a reliable predictor of overall survival, clinical characteristics, and immune cell infiltration, which has the potential to be applied as a valuable biomarker for COAD immunotherapy.

MATERIALS AND METHODS

Dataset Collection and Processing

The gene expression data (fragments per kilobase million, FPKM) and the corresponding clinicopathological information of colon carcinoma were downloaded from TCGA-COAD project (https://portal.gdc.cancer.gov/) databases and the GSE39582 cohort (https://www.ncbi.nlm.nih.gov/geo/).

In order to obtain reliable results, samples with no information on survival outcomes were excluded, and a total of 1,025 COAD patients were eventually included in the follow-up analysis. Details of these 1,025 COAD patients are presented in **Supplementary Table S1**. Before merging the expression matrices of TCGA-COAD project and GSE39582 cohort, the FPKM values of TCGA-COAD were transformed into transcripts per kilobase million (TPM), which were considered to be more comparable with the microarray data. In addition, all raw data were normalized and standardized to eliminate batch effects by using the R software package. Meanwhile, we downloaded the IMvigor210 cohort from the website, which was a cohort study for evaluating the clinical response of atezolizumab in metastatic urothelial cancer (mUC) (Mariathasan et al., 2018). In the IMvigor210 cohort, we excluded the patients with no clinical response information and a total of 298 patients for subsequent validation (**Supplementary Table S2**). In total, 200 hypoxia-related genes (HRGs) were retrieved from the MSigDB database (http://www.broad.mit.edu/gsea/msigdb/), and the full details of these genes are shown in **Supplementary Table S3**.

Consensus Clustering Analysis Based on Hypoxia-Related Genes

Unsupervised clustering analysis was employed to classify patients into distinct molecular subtypes according to the expression of 200 HRGs. In order to increase the intra-class correlation and decrease the correlation, the consensus clustering algorithm was performed and repeated 1,000 times to ensure the stability of the clusters, which we plotted using the R package "ConsensusClusterPlus."

Relationship Between Molecular Subtypes With the Clinical Parameters and Prognosis of Colon Carcinoma

We compared the relationships between molecular subtypes, clinical parameters, and prognosis to examine the clinical value of the two subtypes identified by consensus clustering. Furthermore, we also analyzed the expression of the HRGs among the two subtypes. The clinical parameters included age, sex, T stage, N stage, M stage, and TNM stage. Kaplan–Meier curves were used to assess the differences in overall survival among different molecular subtypes.

Molecular and Immune Features Between Subtypes

GSVA enrichment analysis was employed to assess and compare the difference in biological pathways between the distinct molecular subtypes. and the hallmark gene set (c2.cp.kegg.v7.2) was retrieved from the MSigDB database. Meanwhile, we estimated the relative abundance of 23 immune cells in colon carcinoma using a single-sample gene set enrichment analysis (ssGSEA) algorithm, which was performed using the GSVA R package.

Considering the role of the tumor microenvironment (TME) in tumor progression, we also evaluated the Stromal, Immune, and ESTIMATE scores of each sample by the ESTIMATE algorithm to determine the degree of immune cell infiltration of each subtype. We not only estimated and compared the expression level of six common immune checkpoint-related genes, such as CD274 (PD-L1), PDCD1LG2 (PD-L2), PDCD1 (PD-1), CTLA4, LAG3, and TIGI, but also calculated the patient TIDE score to evaluate the immunotherapy response.

Identification of Differentially Expressed Genes

DEGs between the two hypoxia-related subtypes were identified using the "limma" R package, and the significance criterion for defining DEGs was |log fold change (FC)| > 0.585 and adjusted p-value < 0.05. Furthermore, we performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis on DEGs to identify the related gene functions and enriched pathways through the "clusterProfiler" R package with a cut-off p value < 0.05 and an adjusted p value < 0.05.

Construction of the Prognostic Hypoxia-Related Gene Score

First, univariate Cox regression analysis was performed on DEGs to identify those linked to the prognostic value with a p-value <

0.05. Second, a total of 1,025 patients were randomly categorized into the training group (n = 513) and testing group (n = 512) at a ratio of 1:1; then, the patients in the training group were used to construct the hypoxia-related prognostic HRG-score, and the testing group was used for validation. Finally, based on hypoxia-related prognostic DEGs, The LASSO–Cox regression analysis was then utilized to develop the prognostic HRG-score in the training group, which was performed using the "glmnet" R packet. The HRG-score formula is as follows: HRG-score = Σ (Expi * Coefi), where Coefi and Expi denote the risk coefficient and expression of each gene, respectively. Based on the HRG-score formula, each patient can get a specific risk score. A total of 513 patients in the training group were assigned, based on a median value, to the high-risk group (n = 256) and low-risk group (n = 257). Similarly, a total of 512 patients in the testing group were assigned to HRG-score-related subgroups based on the formula constructed by the training group. The receiver operating characteristic (ROC) curve, which is used to judge the accuracy of the prognostic risk model, was generated by the "timeROC" R package, and principal component analysis (PCA) was performed using the "ggplot2" R package.

RNAseq data (level3) and the corresponding clinical information for 450 colon cancer tumors were obtained from The Cancer Genome Atlas (TCGA) dataset (https://portal.gdc.com). First univariate and multivariate cox regression analyses and forest plots were used to display each variable (*p*-value, HR, and 95% CI) via the "forest plot" package. Based on the results of multivariate Cox proportional risk analysis, column line plots were created using the "rms" package to predict the total recurrence rate in 1, 2, and 3 years. The line graphs provide graphical results for these factors, allowing the prognostic risk of individual patients to be calculated by the points associated with each risk factor.

Correlation Analysis of the HPR-Score With Clinical Parameters

A Chi-square test was applied to explore the correlation between the HRG-score and the clinical parameters (age, gender, T stage, N stage, M stage, and TNM stage). To assess whether the HRG-score is an independent prognostic factor associated with prognosis, we performed univariate analysis and multivariate analysis on the training group and testing group. Kaplan-Meier analysis was used to compare survival outcomes of patients between high- and low-risk and assessed the correlation between the survival outcome and HRG-score. We further analyzed the relationship between HRG-score and molecular subtypes through a boxplot.

Evaluation of Immune Status and Mismatch Repair Status Between the High- and Low-Risk Groups

The CIBERSORT algorithm was used to calculate the relative abundance of 22 infiltrating immune cells per sample in the low- and high-risk groups (**Supplementary Table S11**). We explored the correlation between the 22 infiltrating immune cell fractions and the 7 genes in the PRG scores. In addition, we compared the expression levels of immune checkpoints between the low- and high-scoring groups and analyzed the relationship between the HRG score and the cancer stem cell (CSC) index.

Sensitivity Analysis of Chemotherapy and Immunotherapy

In a project to evaluate the difference in the treatment effect of five chemotherapeutic agents in patients in the high-risk and low-risk groups, the semi-inhibitory concentration (IC_{50}) values of chemotherapeutic agents were analyzed by the R package "pRRophetic." We acquired the IPS of colon cancer patients in TCGA-COAD project from TCIA database and compared the IPS of the distinct risk group to evaluate the response to immune checkpoint-blocking therapy. We further explored the relationship between immunotherapy sensitivity and HRG-score by the IMvigo210 cohort.

Statistical Analysis

All statistical analyses were performed using R software (v4.0.2). *p*-values <0.05 were considered statistically significant if not explicitly stated.

RESULT

Identification of Hypoxia Gene-Related Subtypes in Colon Carcinoma

A total of 1,025 patient samples with complete survival information from TCGA-COAD project and GEO-GES39582 were included in our study. To further investigate the expression characteristics of HRGs in colon carcinoma, we used a consensus clustering algorithm to cluster the patients based on the expression of the 200 HRGs. Our result found that when K = 2, the intra-group correlations were the highest, and the inter-group correlations were the lowest, indicating sorting the entire patients into two subtypes may be the most optimal selection (**Figure 1A**). PCA analysis revealed the significant differences between the two subtypes (**Figure 1B**), suggesting there existed significant heterogeneity in the expression of hypoxia genes in patients with colon carcinoma. The Kaplan–Meier curves showed an obvious difference in the prognosis between the two hypoxia subtypes, and the prognosis in patients with subtype A was significantly better than that in patients with subtype B (log-rank test, $p = 0.011$; **Figure 1C**). Furthermore, we compared the correlations of the two subtypes with clinical parameters and the expression of hypoxia genes. As the heatmap showed (**Figure 1D**), there were no significant differences in clinical parameters between the two subtypes; however, compared with subtype A, most of the hypoxia-related genes were highly expressed in subtype B.

Function Enrichment of the Molecular Subtypes

GSVA enrichment analysis showed that metabolism-related and DNA synthesis-related pathways including butanoate

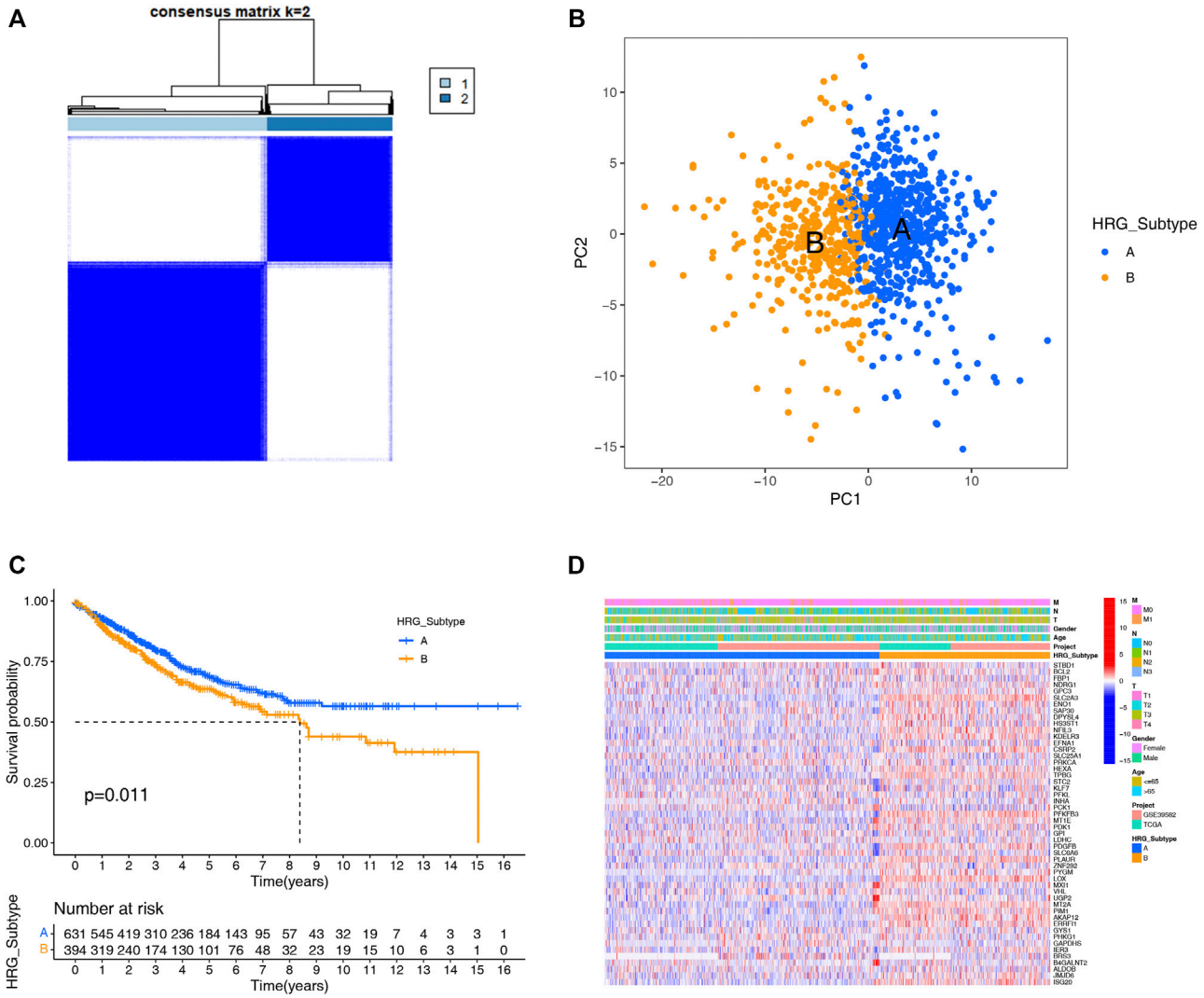


FIGURE 1 | HRG subtypes and clinical parameters and biological characteristics of two distinct subtypes of samples divided by consistent clustering. **(A)** Consensus matrix heatmap defining two subtypes (k = 2) and their correlation area. **(B)** PCA showing a remarkable difference in transcriptomes between the distinct HRG-subtypes, and each dot represents a single sample. **(C)** KM survival curve analysis showed that the overall survival time of the distinct HRG-subtypes was different (log-rank tests, $p < 0.001$). **(D)** Differences in clinical parameters and HRG expression levels between the two distinct HRG-subtypes.

metabolism, propanoate metabolism, pyruvate metabolism, fatty acid metabolism, nonhomologous end joining, base excision repair, DNA replication-related pathway were upregulated in subtype A, while T- and B-cell receptor signaling pathway, natural killer cell-mediated cytotoxicity, antigen processing and presentation, checkpoint signaling pathway, and NOD-like, RIG-I-like, and Toll-like receptor signaling pathways were upregulated in subtype B(**Figure 2A**, **Supplementary Table S4**).

Characteristics of the Tumor Microenvironment in Distinct Subtypes

The tumor microenvironment (TME) has been proved to play an important role in tumor progression and immune response. We evaluated the 23 immune cells' infiltration levels of each patient by applying the ssGSEA (**Supplementary Table S5**) and found significant differences in the infiltration of most immune cells between the two subtypes (**Figure 2B**). The infiltrate levels of 20 immune cell types, including activated B cells, activated $CD4^+$ T cells, activated $CD8^+$ T cells, natural killer T cells, and regulatory T cells, were significantly higher in the subtype B than those in the subtype A. The ESTIMATE algorithm was used to evaluate the TME score (stromal score, immune score, and ESTIMATE score) of each patient through the "ESTIMATE" R package (**Supplementary Table S6**), and we found that the stromal score, immune score, and ESTIMATE score were significantly higher in subtype B than subtype A (**Figure 2C**). Recently, the immune checkpoint blockade has achieved promising results in the immunotherapy of tumors. Therefore, we subsequently analyzed the expression levels of several

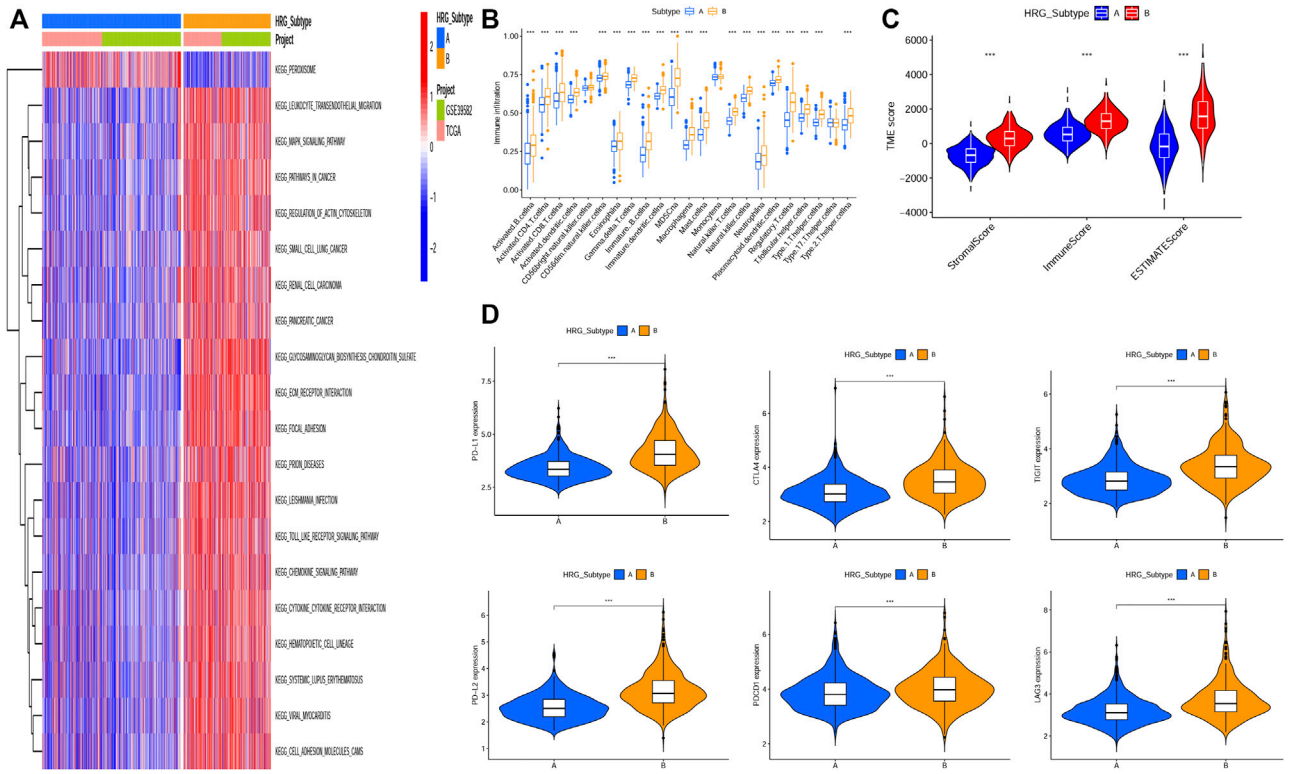


FIGURE 2 | Correlations of tumor immune cell microenvironments and two HRG-subtypes. **(A)** GSVA of biological pathways between two distinct subtypes, in which red and blue represent activated pathways and blue represents inhibited pathways, respectively. **(B)** Relative abundance of 23 infiltrating immune cell types in the two HRG-subtypes. **(C)**Correlations between the two CRC subtypes and TME score. **(D–I)** Expression levels of PD-L1, PD-L2, PDCD1, LAG3, TIGIT, and CTLA4 in two distinct HRG-subtypes. (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$).

important immune checkpoint-related genes, such as CD274 (PD-L1), PDCD1LG2, PDCD1, CTLA4, LAG3, and TIGIT (**Figures 2D–I**). We found that the expression levels of six immune checkpoint-related genes in subtype B were higher than those in subtype A, indicating that patients in subtype B were more likely to form an immunosuppressive microenvironment and escape from immune surveillance.

Construction and Validation of the Prognostic Hypoxia Related Gene-Score

We identified 1,132 DEGs between the two HRG-related subtypes, of which 139 genes were upregulated in subtype A and 993 genes were upregulated in subtype B (**Figure 3A**, **Supplementary Table S7**). Then, we conducted GO and KEGG enrichment analysis on the 1,132 DEGs to explore the potential function and pathway through the "clusterProfiler" R package. In the GO analysis, the top 5 most significantly enriched terms were collagen-containing extracellular matrix, extracellular matrix organization, extracellular structure organization, positive regulation of cell adhesion, and negative regulation of immune system process (**Figure 3B**, **Supplementary Table S8**).In the KEGG analysis, the top 5 most significantly enriched terms were PI3K-Akt signaling pathway, cytokine–cytokine receptor interaction, cell adhesion molecules, phagosome, and focal adhesion (**Figure 3C**, **Supplementary Table S9**). Univariate Cox regression analysis was employed on the 1,132 DEGs and 437 genes associated with the prognostic value with a p-value <0.05 and were identified as candidate genes for subsequent analysis (**Supplementary Table S10**). Then, all patients were classified into training group (n = 513) and testing group (n = 512) at a ratio of 1:1 randomly, the training group for developing the prognostic signature and the testing group for validation. LASSO regression analysis on the 437 candidate genes was performed to exclude overlapping genes and reduce the fitting effect of the signature (**Figures 3D,E**). Finally, six genes were included to construct the risk model after multivariate Cox proportional risk regression analysis, four of which were associated with high risk and two with low risk (**Figure 3F**). According to the results of the multivariate Cox proportional risk regression analysis, the HRG-score was constructed as follows: Risk score = (0.2665 * expression of S1PR3) + (0.2478* expression of ETV5) + (0.2115* expression of CD36) + (0.2808* expression of FOXC1) + (−0.1735* expression of CXCL10) + (−0.0976* expression of MMP12). According to the median risk score, patients in the training group were classified into high-risk group (n = 256) and low-risk group (n = 257) (**Figure 3G**). When compared to the low-risk group, we found that more patients died and a shorter survival time in the high-risk group (**Figure 3H**). The expression levels of six genes involved in the

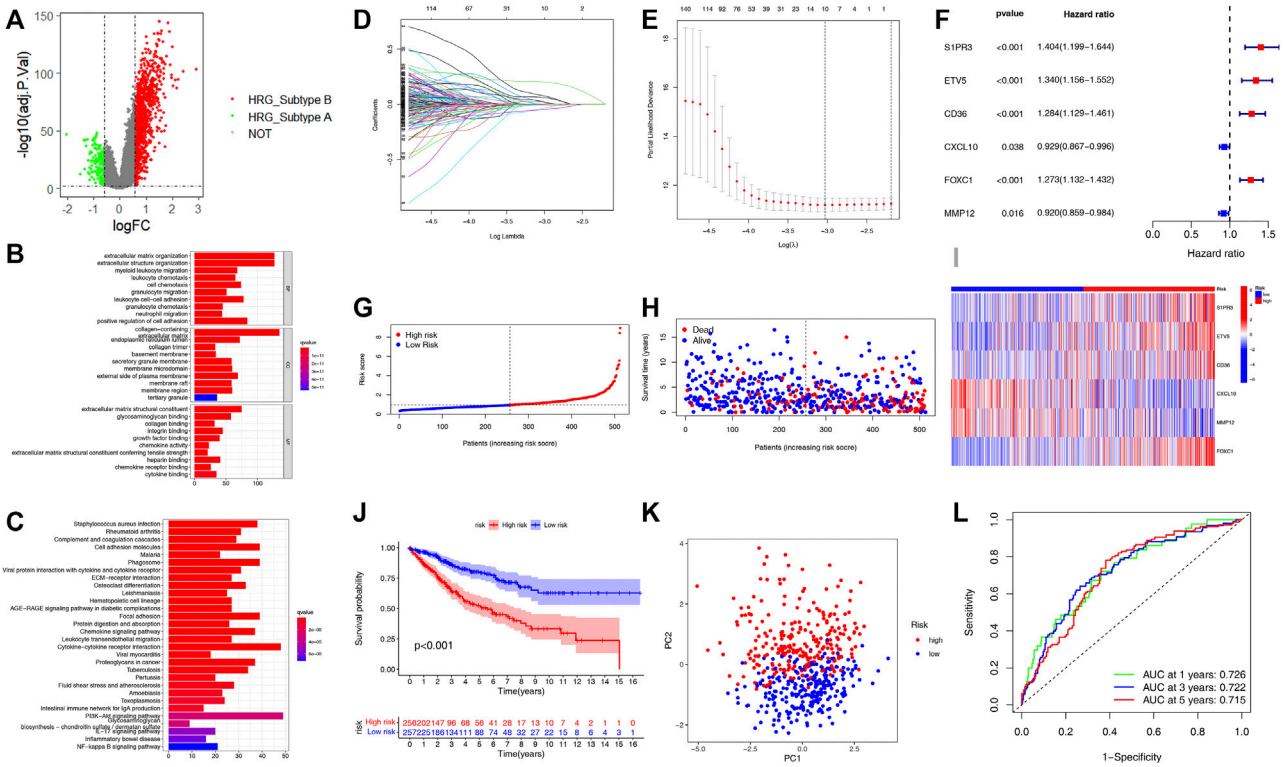


FIGURE 3 | Construction of the HRG-score in the training set based on the differentially expressed genes of two distinct HRG-subtypes. **(A)** Volcano plot of differentially expressed genes between the two distinct HRG-subtype. Gray dots represent not significant genes, green dots represent upregulated genes in HRG-subtype A, and red dots represent upregulated genes in HRG-subtype B **(B–C)** GO and KEGG enrichment analyses of DEGs among two distinct HRG-subtypes. **(D–E)** LASSO regression analysis and partial likelihood deviance on the prognostic genes. **(F)** Forest plot of multivariate cox regression analysis for prognostic genes. **(G–H)** Ranked dot and scatter plots showing the HRG-score distribution and patient survival status. **(I)** Heatmap of the expression of six genes involved in the HRG-score in low- and high-risk groups. **(J)** Survival analysis of the patients in low- and high-risk groups. **(K)** PCA based on the prognostic signature. (L) ROC curves to predict the sensitivity and specificity of 1-, 3-, and 5-year survival according to the HRG-score.

construction of our HRG-score signature are shown in **Figure 3I**. Kaplan–Meier survival analysis revealed that there existed a significant difference in survival time between the low- and high-risk group, and the patients in the low-risk group had a longer survival time ($p < 0.001$) (**Figure 3J**). The principal component analysis (PCA) showed that patients with different risks were well separated into two clusters (**Figure 3K**). The AUC values for the 1-, 3-, and 5-year survival were 0.726, 0.722, and 0.715, respectively (**Figure 3L**).

Validation of the Hypoxia Related Gene-Score Signature

In order to verify the practicality and credibility of the model, we performed the same analysis for internal validation using a testing group (n = 512). Based on the median risk score in the training group, all patients in the testing group were classified into the low-risk group (n = 244) and high-risk group (n = 268) (**Figure 4A**). Compared to the low-risk group, the proportion of patient deaths tended to be high in the high-risk group (**Figure 4B**). Heatmap was also plotted to analyze the expression of the six genes involved in the HRG-score signature between the high- and low-risk groups (**Figure 4C**). Kaplan–Meier analysis showed that the survival probability of the high-risk group was significantly lower than that of the low-risk group ($p < 0.04$) (**Figure 4D**). The principal component analysis (PCA) showed that the patients with different risk scores can be stratified into two clusters distinctly (**Figure 4E**). The AUC values for the 1, 3, and 5 years of ROC were 0.748, 0.727, and 0.726 respectively, indicating our model's good predictive efficacy (**Figure 4F**). Nomograms of S1PR3, ETV5, CD36, FOXC1, CXCL10, and MMP12 expression and independent clinical risk factors (age and pathological stage) were constructed (**Supplementary Figure S4**). A higher total number of points in the nomogram represents a worse prognosis. In addition, the C-index value was 0.779 ($p < 0.001$). The deviation-corrected line in the calibration plot was close to the ideal curve (i.e., 45° line), indicating good agreement between the predicted and observed results.

Correlation Analysis of Hypoxia Related Gene-Score and Clinical Parameters

We plotted a heatmap of clinical parameters for the patients in the training group and found statistically significant differences in T, N, M, and TNM stages between high- and low-risk groups

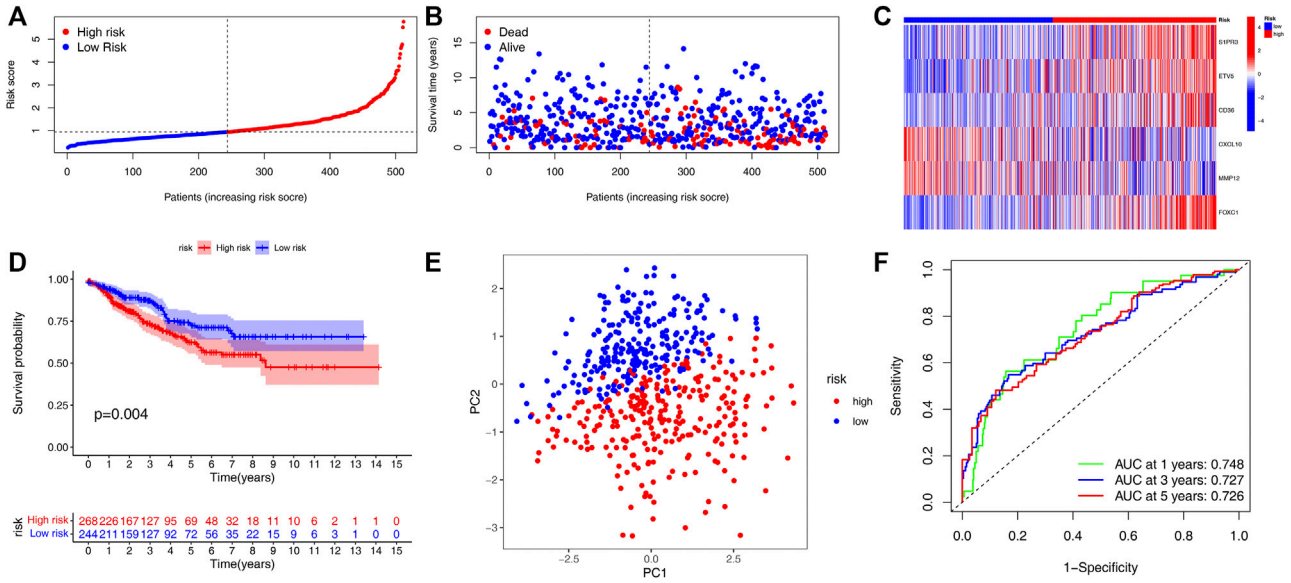


FIGURE 4 | Validation of the HRG-score signature in the testing set. **(A,B)** Ranked dot plot indicates the PRG-score distribution, and the scatter plot presents the patients' survival status. **(C)** Heatmap of the expression of six genes involved in the HRG-score in low- and high-risk groups. **(D)** KM analysis of the OS between the low- and high-risk groups. **(E)** PCA demonstrated that the patients in the different risk groups were distributed in two directions. **(F)** ROC curves to predict the sensitivity and specificity of 1-, 3-, and 5-year survival according to the PRG-score.

(**Figure 5A**). We further analyzed the relationship between the T stage, N stage, M stage, and TNM stage and risk score separately. As shown in **Figure 5B**, we found significant differences in risk scores for T, N, M, and TNM stages, and patients' clinical stage deteriorated as risk scores increased, suggesting that high-risk scores predicted poor outcomes for patients. In addition, we also analyzed the correlation between the risk score of the testing group and the clinical parameters (**Supplementary Figure S1A**) and obtained the same result that the risk score can be used to evaluate the prognosis of patients (**Supplementary Figure S1B–E**). Univariate and multivariate Cox regression analyses were employed to assess whether HRG-score could be used as an independent prognostic factor. The univariate Cox regression analysis indicated that the HRG-score was an independent factor predicting poor survival in the training group (HR = 1.701, 95% CI: 1.485–1.948) (**Figure 5C**). After adjusting for other confounding factors, the multivariate analysis yielded similar results that the HRG-score can be a prognostic factor for patients in the training group (HR = 1.419, 95% CI: 1,226–1.641) (**Figure 5D**). Univariate and multivariate Cox regression analyses were also employed in the testing group, and we also got the same result (HR = 1.505, 95% CI: 1.344–1.686 and HR = 1.297, 95% CI: 1.147–1.467, **Supplementary Figure S2A,B**).

Evaluation of Tumor Microenvironment and Checkpoints Between the High- and Low-Risk Groups

CIBERSORT algorithm was performed to assess the association between the HRG-score and the abundance of immune cells. The scatter diagrams showed that the HRG-score was positively correlated with macrophage M2, neutrophils, and macrophages M0 and negatively correlated with macrophages M1, plasma cells, T cell CD4 memory activated, T-cell follicular helper, and T cell CD8 (**Figure 6A**). We observed that the stromal score and ESTIMATE score were significantly higher in the low-risk group than the high-risk group (**Figure 6B**). **Figure 6C** shows that 22 immune checkpoints were differentially expressed in the two groups, and the expression of most immune checkpoint-related genes was higher in the low-risk group than that in the high-risk group. We also assessed the correlation between the six genes of the HRG-score signature and the abundance of immune cells. We observed that most immune cells were significantly correlated with the six genes (**Figure 6D**).

Correlation Analysis of PRG-Score With the MMR Status and CSC Index

Inactivating mutations in mismatch repair genes such as MLH1, MSH2, MSH6, and PMS2 can cause mismatch repair (MMR) dysfunction and then lead to microsatellite high instability (MSI-H). Patients with high microsatellite instability (MSI-H) are more sensitive to immunotherapy and can benefit from immunotherapy drugs. Correlation analyses revealed that a high HRG-score was significantly correlated with proficient mismatch repair status (pMMR), while a low HRG-score was associated with deficient mismatch repair (dMMR) status (**Figure 7A**), suggesting that patients with low-risk scores benefit from immunotherapy better than those with high-risk scores. Stem cells (CSCs) are a small subset of undifferentiated cells in tumor tissues, which have strong self-renewal potential

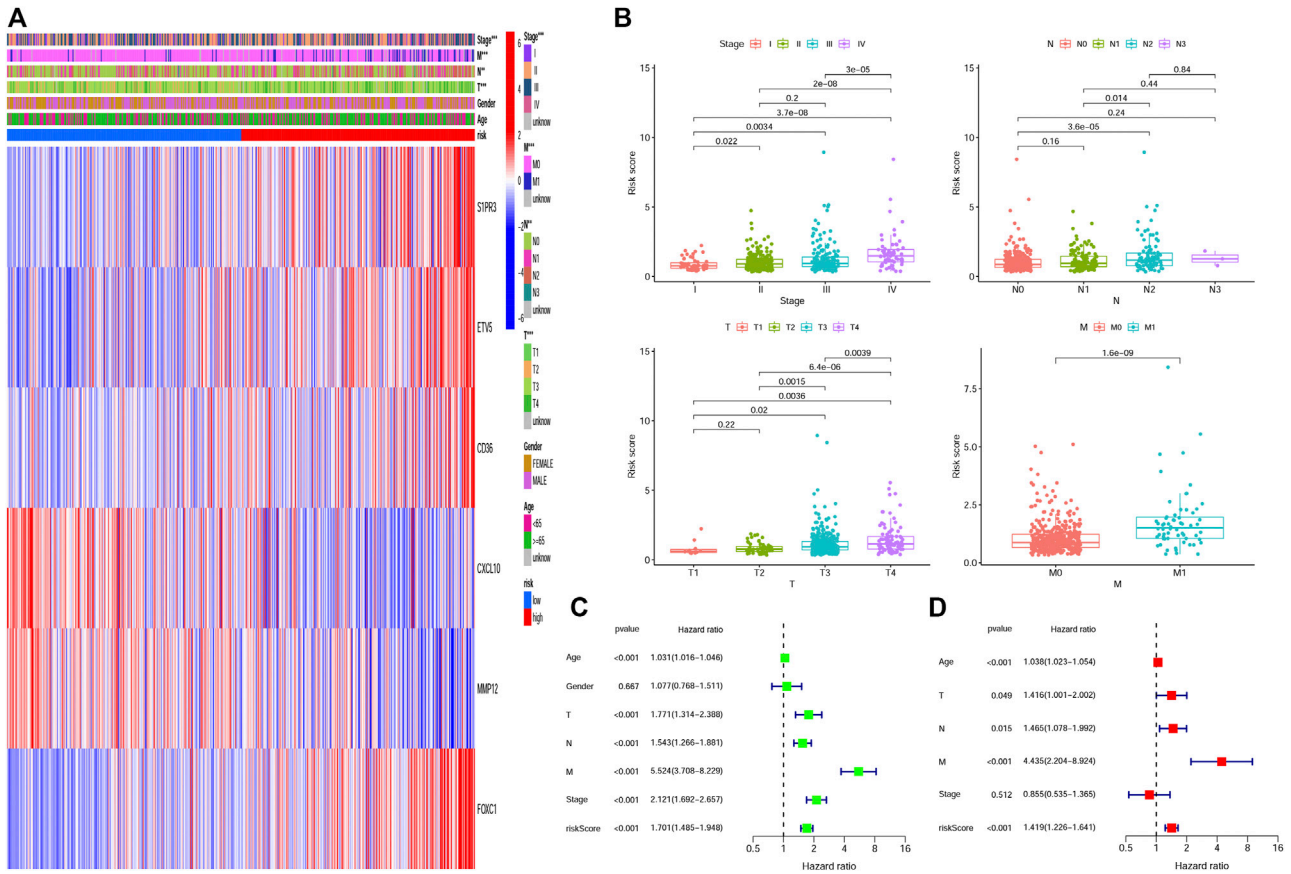


FIGURE 5 | Correlation and independent prognosis analysis of HRG-score and clinical parameters in the training set. **(A,B)** Univariate and multivariate analyses of the prognostic value of the HRG-score. **(C)** Relationships between clinical parameters and the low- and high-risk groups. **(D)** Clinical application value of HRG-score in predicting T stage, N stage, M stage, and TNM stage, respectively (*p < 0.05; **p < 0.01; ***p < 0.001).

and tumorigenic potential, and can form tumors in a low number *in vivo*. The correlation analysis between the PRG-score and CSC index showed that PRG-score was negatively correlated with the CSC index (R = −0.31, p < 0.001), indicating that tumor cells with lower HRG-score had a lower degree of cell differentiation and distinct stem cell properties (**Figure 7B**).

Analysis of the Sensitivity of Chemotherapeutics and Immunotherapy Based on Hypoxia Related Gene-Score

We next selected four chemotherapy drugs currently used for the treatment of colon carcinoma to assess the sensitivity of patients in the low- and high-risk groups to these drugs. As shown in **Figure 7C**, we found that the patients in the low-risk group showed more sensitivity to chemotherapy drugs indicating that the low-risk group may benefit more from chemotherapy drugs. Meanwhile, the applicability of different HRG-score samples to combined therapy of anit-CTLA4 and anti-PD1 was compared by IPS. The analysis showed a significant difference (p = 0.00023 < 0.05) that the IPS of the low-risk group treated with the combination of anti-CTLA4 and anti-PD1 was relatively higher than that of the high-risk group, indicating that the patients with low HRG-score had a better therapeutic effect on Immunotherapy (**Figure 7D**). To further evaluate the robustness of our HRG-score signature, we calculated the risk score of patients in the IMvigor210 cohort based on the formula of HRG-score and analyzed the correlation of risk score with the effect of immunotherapy. As shown in **Figure 7E**, there existed significant differences in risk scores between the complete remission/partial remission (CR/PR) group and stable disease/progressive disease (SD/PD) group, and the risk score of patients in the CR/PR group was significantly lower than that of patients in the SD/PD group (p = 0.0031 < 0.05). To further improve the clinical application of our model, we constructed a nomogram containing HRG-score and clinical parameters to predict overall survival at 1, 3, and 5 years (**Supplementary Figure S3A**), and the calibration plots suggested that the nomogram had a good performance in predicting the survival of colon cancer patients (**Supplementary Figure S3B**).

DISCUSSION

CRC is an extremely common malignant tumor. In recent years, there is a tendency to develop to the right half of the colon, which

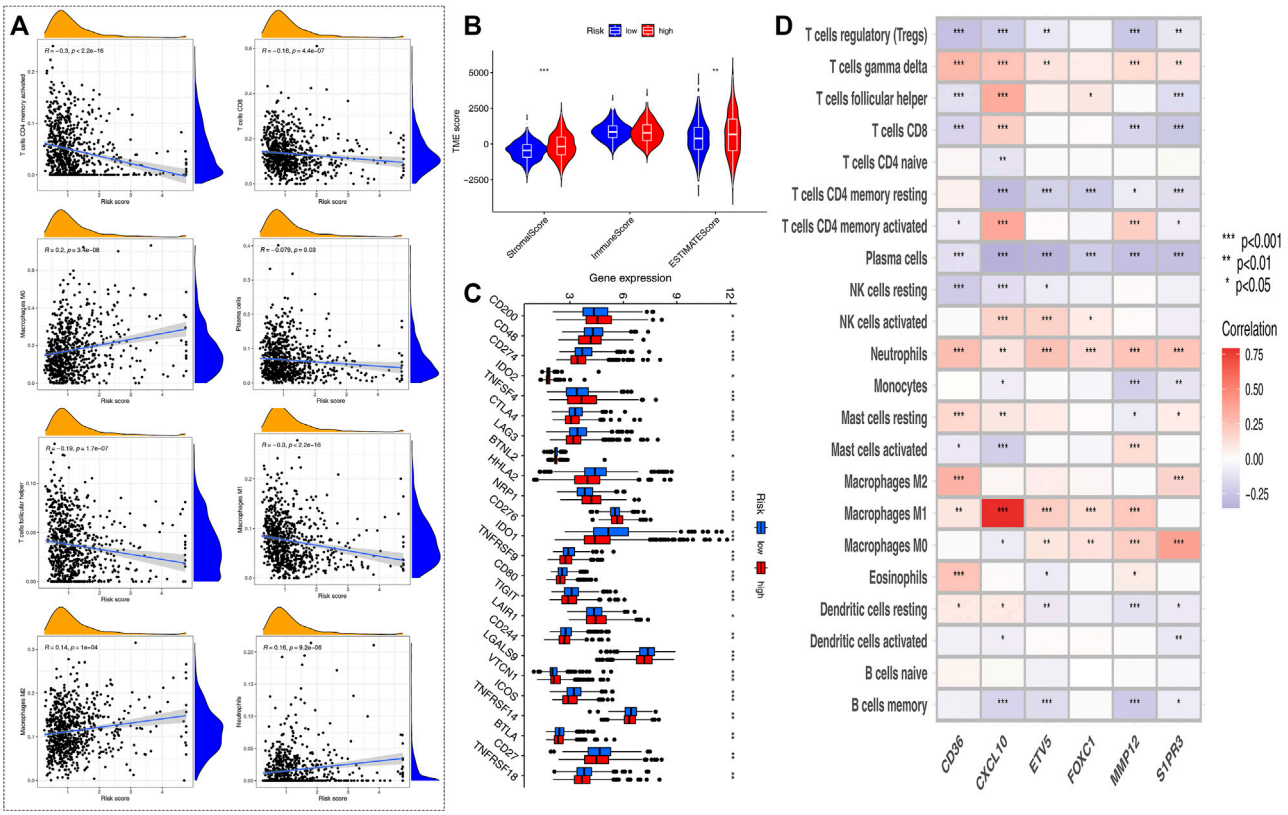


FIGURE 6 | Evaluation of the TME and checkpoints between the two risk groups. **(A)** Correlations between HRG-score and immune cell types. **(B)** Correlations between HRG-score and TME score. **(C)** Expression of immune checkpoint-related genes in the low- and high-risk groups. **(D)** Correlations between the relative abundance of immune cells and six genes involved in the HRG-score. (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$)

is closely related to heredity, living habits, and colorectal adenoma (Zhang et al., 2020). According to the latest data, the global incidence rate of CRC is the second only to breast cancer and lung cancer, and the mortality rate is the second only to lung cancer. At present, the main treatment of CRC is surgical treatment, supplemented by neoadjuvant radiotherapy and chemotherapy, postoperative radiotherapy and chemotherapy, and immunotherapy. The main prognostic key issues affecting CRC are currently the need for timely surgical intervention and effective radiotherapy treatment. Unfortunately, more than 50% of CRC patients experience tumor recurrence, metastasis, invasion, and resistance to chemotherapy drugs at the time of diagnosis or during their follow-up treatment (Song et al., 2021), thus losing the standard of care of surgical treatment with radiotherapy and subsequently having a poor prognosis as well as poor quality of survival. Chemotherapy is a relative option for patients with CRC who cannot tolerate surgical intervention; however, there are still no specific chemotherapeutic agents for CRC. A growing body of evidence suggests that multiple genes and cellular pathways are involved in the development of CRC. To date, the lack of knowledge about the exact molecular mechanisms underlying CRC progression has limited the ability to treat advanced diseases. Therefore, it is necessary to identify the key genes and pathways of CRC in order to understand its molecular mechanism, explore potential biomarkers, and develop more effective diagnostic and therapeutic strategies.

Hypoxia-inducible factor (HIF) played an important role in cancer biology, including angiogenesis, cell survival, glucose metabolism, and invasion (Zhang et al., 2021). HIF can facilitate metabolic metastasis and enhance the non-mitochondrial mechanism of ATP production, thus providing energy for tumor cells (Gatenby et al., 2004). In addition, HIF stabilization can lead to inhibition of apoptotic pathways through silencing of mitochondrial activity. Hypoxia can mitigate the infiltration rate of immune cells and their function in the TME (You et al., 2021). Glycolysis can lead to acid TME with a pH as low as 5.8 to 6.5, and the acidic environments can inhibit immune cell differentiation and function. With the advance of high-throughput sequencing, identification of molecular characterization gradually becomes a significant method for biomedical research, which can be used for identifying biomarkers for prognosis predicting, recurrence monitoring, and clinical risk stratification (Wang et al., 2009; Xiao et al., 2018).

The growth and progression of malignant tumors are associated with immunosuppression, and tumor cells evade immune surveillance through different mechanisms, including the activation of immune checkpoints pathways that suppress

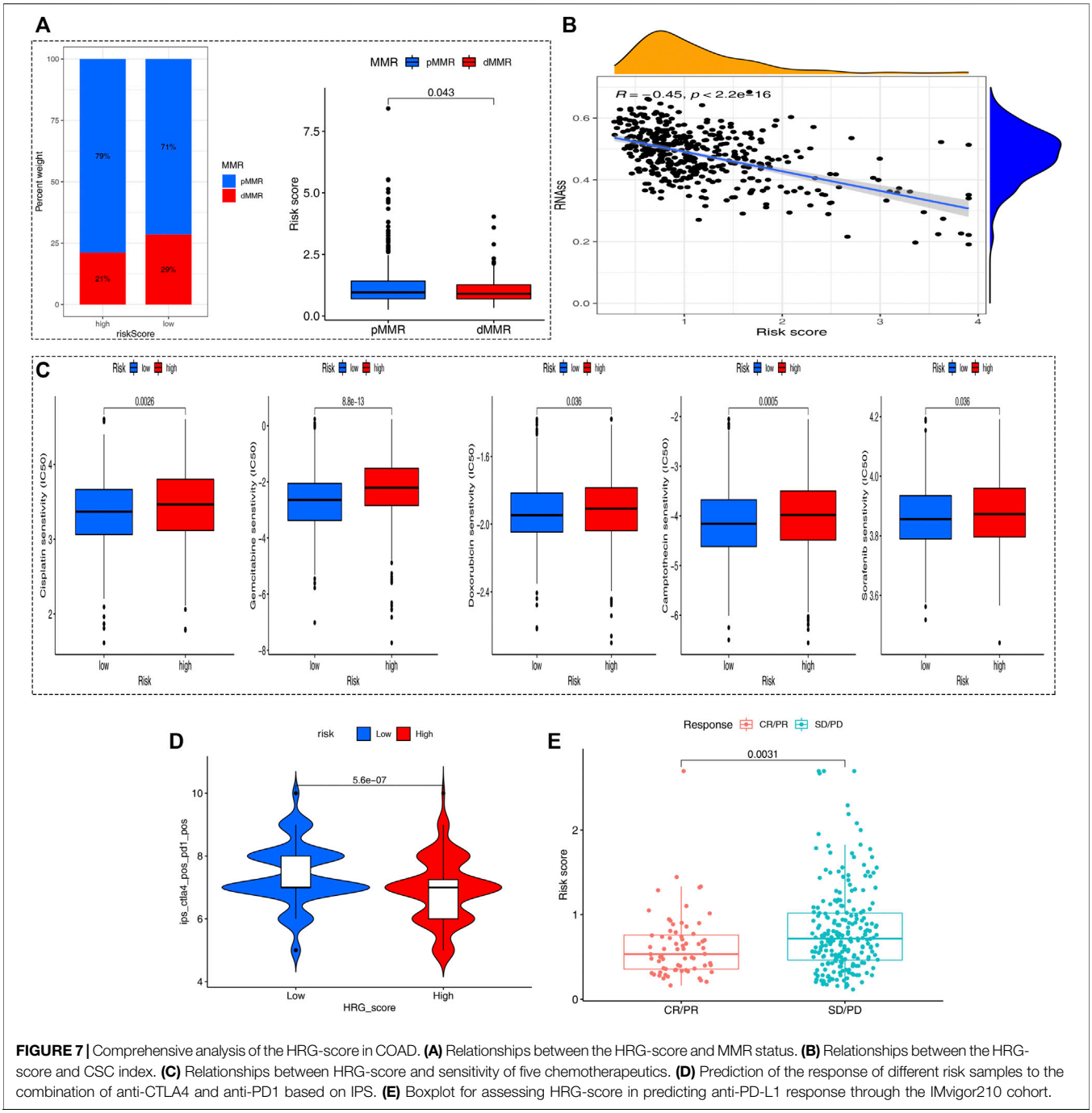


FIGURE 7 | Comprehensive analysis of the HRG-score in COAD. **(A)** Relationships between the HRG-score and MMR status. **(B)** Relationships between the HRG-score and CSC index. **(C)** Relationships between HRG-score and sensitivity of five chemotherapeutics. **(D)** Prediction of the response of different risk samples to the combination of anti-CTLA4 and anti-PD1 based on IPS. **(E)** Boxplot for assessing HRG-score in predicting anti-PD-L1 response through the IMvigor210 cohort.

anti-tumor immune responses. The successful development of immune checkpoint genes (ICGs) was a milestone event in tumor immunotherapy and was named one of the top 10 scientific discoveries by Nature in 2013 (Wolchok et al., 2014). ICGs inhibit and kill tumor cells by enhancing the body's anti-tumor immune function and have shown significant clinical efficacy in the treatment of a variety of malignancies, becoming an important tool in tumor therapeutics (Wang et al., 2018). Based on the expression of 120 hypoxia-related genes, 1,025 colon cancer samples from TCGA-COAD project and GEO-GSE39582 were separated into two heterogeneous subtypes, with significant differences in OS between the two subtypes. Hypoxia is an important factor in the poor prognosis of tumor by regulating cancer hallmark, thus creating physical barriers conducive to tumor survival (Abou Khouzam et al., 2022). We found most hypoxia-related genes are highly expressed in subtype B, and the patients in subtype B had a worse survival outcome than those in subtype A. We then compared the several expression levels of six known immune checkpoint genes (PD-L1, PD-L2, PD-1, LAG3, TIGIT, IDO1, and CTLA-4)between the two subtypes, and the

expression level of the six genes was significantly higher in subtypes B than subtype A. The previous studies reported that the high expression level of immune checkpoint genes was more likely to form an immunosuppressive microenvironment and promote tumor immune escape (Dunn et al., 2022); meanwhile, the upregulation of immune checkpoint genes (ICGs) was positively correlated with high immune cell infiltration (Hu et al., 2021). The TME score and immune cell infiltration have been reported to be tightly associated with the immunotherapy of cancers and the prognosis (Luo et al., 2020). Thus, we also analyzed the relationship between subtype and immune cell infiltration. Compared with subtype A, the expression level of most immune cells including activated B cells, activated $CD4^+$T cells and activated CD^+8 T cells was significantly higher in subtype B. In addition, we also observed that the stromal score and ESTIMATE score were higher in B than A. These results suggest that patients in subtype A may benefit from immune checkpoint inhibitor therapies. TIDE comparison between the two groups showed that patients with subtype B were more likely to form immune escape than patients with subtype A, which further confirmed our previous results.

Our findings suggest that hypoxic genes differ in the course of changes in the colon. Therefore, we constructed a robust and effective prognostic HRG-score and validated its predictive ability. We explored the expression level of six genes of our HRG-score and found a significant difference between the risk groups. There were significant differences in clinical parameters, prognosis, TME, ICGs, MMR status, CSC index, and drug sensitivity between low- and high-risk HRG-score patients. It will help to better understand the molecular mechanism of colorectal cancer and provide new ideas for targeted therapy (Bai et al., 2020; Huo et al., 2021; Yan et al., 2021).

Immunotherapy is a promising method in cancer treatment and has achieved remarkable efficacy in the treatment of colorectal cancer (Ganesh et al., 2019). Due to the high heterogeneity of molecular genetics and histopathology of colon cancer, immunotherapy still has limitations and obstacles (Makaremi et al., 2021). TME plays a crucial role in the tumorigenesis and progression of COAD, and the immunosuppressive function is one of the causes of poor response to treatment. Immune cells of TME are involved in tumor suppression and progression. Immune-infiltrating cells in TME are mainly composed of dendritic cells, macrophages, NK cells, T cells, and B cells (Koi et al., 2017). Surveillance and elimination of abnormal antigens is an essential feature of the normal function of the immune system. Macrophages and NK cells play a crucial role in stimulating the adaptive immune system that targets tumor cells (Markman et al., 2015), and a higher level of NK cells and $CD8^+$ T-cell infiltration often predicts a better prognosis (Sconocchia et al., 2014). In our study, we discovered that the relative abundance of B cells, $CD8^+$T cells, NK cells, and macrophages cells was significantly higher in the low-risk group.

In summary, this study conducted a comprehensive bioinformatic analysis of two new molecular subgroups of hypoxic genes and colorectal cancer patients and constructed an HRG-score model consisting of six genes. However, due to the limitations of bioinformatics analysis, further clinical sample testing and cellular and animal experiments are needed to explore the function of hypoxia genes in colorectal cancer and the related molecular mechanisms in depth.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

GW conceived and designed the study. CW, YT, HM, SW, XH, LZ and GW performed the analyses. All authors wrote the manuscript and read and approved the manuscript.

ACKNOWLEDGMENTS

We would like to acknowledge TCGA for free use.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.919389/full#supplementary-material

Supplementary Figure S1 | Nomograms and calibration plots predicting the probability of 1-year, 2-year, and 3-years DFS in COAD patients.

Supplementary Table S1 | Clinical and prognostic information of 1,025 patients.

Supplementary Table S2 | 298 patients inside IMvigor120 for follow-up analysis.

Supplementary Table S3 | 200 hypoxia genes.

Supplementary Table S4 | GSVA analysis results.

Supplementary Table S5 | ssGSEA calculates the content of 23 immune cells.

Supplementary Table S6 | TME SCORE score.

Supplementary Table S7 | 1,132 differential genes.

Supplementary Table S8 | GO analysis.

Supplementary Table S9 | KEGG analysis.

Supplementary Table S10 | 437 one-way prognostic genes.

Supplementary Table S11 | CIBERSORT algorithm calculates the relative infiltration content of 22 immune cells.

REFERENCES

Abou Khouzam, R., Zaarour, R. F., Brodaczewska, K., Azakir, B., Venkatesh, G. H., Thiery, J., et al. (2022). The Effect of Hypoxia and Hypoxia-Associated Pathways in the Regulation of Antitumor Response: Friends or Foes? *Front. Immunol.* 13, 828875. doi:10.3389/fimmu.2022.828875

Arnold, M., Pandeya, N., Byrnes, G., Renehan, A. G., Stevens, G. A., Ezzati, M., et al. (2015). Global Burden of Cancer Attributable to High Body-Mass Index in 2012: a Population-Based Study. *Lancet. Oncol.* 16 (1), 36–46. doi:10.1016/S1470-2045(14)71123-4

Bai, Y., Wei, C., Zhong, Y., Zhang, Y., Long, J., Huang, S., et al. (2020). Development and Validation of a Prognostic Nomogram for Gastric Cancer Based on DNA Methylation-Driven Differentially Expressed Genes. *Int. J. Biol. Sci.* 16 (7), 1153–1165. doi:10.7150/ijbs.41587

Bekaii-Saab, T. S., Ou, F.-S., Ahn, D. H., Boland, P. M., Ciombor, K. K., Heying, E. N., et al. (2019). Regorafenib Dose-Optimisation in Patients with Refractory Metastatic Colorectal Cancer (ReDOS): a Randomised, Multicentre, Open-Label, Phase 2 Study. *Lancet. Oncol.* 20 (8), 1070–1082. doi:10.1016/S1470-2045(19)30272-4

Chouaib, S., Umansky, V., and Kieda, C. (2018). The Role of Hypoxia in Shaping the Recruitment of Proangiogenic and Immunosuppressive Cells in the Tumor Microenvironment. *Contemp Oncol (Pozn)*, 2018(1A), 7–13. doi:10.5114/wo.2018.73874

Choi, B.-J., Park, S.-A., Lee, S.-Y., Cha, Y. N., and Surh, Y.-J. (2017). Hypoxia Induces Epithelial-Mesenchymal Transition in Colorectal Cancer Cells through Ubiquitin-specific Protease 47-mediated Stabilization of Snail: A Potential Role of Sox9. *Sci. Rep.* 7 (1), 15918. doi:10.1038/s41598-017-15139-5

Compton, C., Fenoglio-Preiser, C. M., Pettigrew, N., and Fielding, L. P. (2000). American Joint Committee on Cancer Prognostic Factors Consensus Conference. *Cancer*, 88(7), 1739–1757. doi:10.1002/(sici)1097-0142(20000401)88:7<1739::aid-cncr30>3.0.co;2-t

Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J., and Schreiber, R. D. (2002). Cancer Immunoediting: from Immunosurveillance to Tumor Escape. *Nat. Immunol.* 3 (11), 991–998. doi:10.1038/ni1102-991

Emami Nejad, A., Najafgholian, S., Rostami, A., Sistani, A., Shojaeifar, S., Esparvarinha, M., et al. (2021). The Role of Hypoxia in the Tumor Microenvironment and Development of Cancer Stem Cell: a Novel Approach to Developing Treatment. *Cancer Cell Int.* 21 (1), 62. doi:10.1186/s12935-020-01719-5

Ferlay, J., Shin, H.-R., Bray, F., Forman, D., Mathers, C., and Parkin, D. M. (2010). Estimates of Worldwide Burden of Cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer* 127 (12), 2893–2917. doi:10.1002/ijc.25516

Ganesh, K., Stadler, Z. K., Cercek, A., Mendelsohn, R. B., Shia, J., Segal, N. H., et al. (2019). Immunotherapy in Colorectal Cancer: Rationale, Challenges and Potential. *Nat Rev Gastroenterol HepatolGastroenterology hepatology* 16 (6), 361–375. doi:10.1038/s41575-019-0126-x

Gatenby, R. A., and Gillies, R. J. (2004). Why Do Cancers Have High Aerobic Glycolysis? *Nat. Rev. Cancer* 4 (11), 891–899. doi:10.1038/nrc1478

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. *Briefings Bioinforma.* 22 (3), bbaa176. doi:10.1093/bib/bbaa176

Huo, J., Wu, L., and Zang, Y. (2021). Eight-gene Prognostic Signature Associated with Hypoxia and Ferroptosis for Gastric Cancer with General Applicability. *Epigenomics* 13 (11), 875–890. doi:10.2217/epi-2020-0411

Jahanafrooz, Z., Mosafer, J., Akbari, M., Hashemzaei, M., Mokhtarzadeh, A., and Baradaran, B. (2020). Colon Cancer Therapy by Focusing on Colon Cancer Stem Cells and Their Tumor Microenvironment. *J. Cell Physiol.* 235 (5), 4153–4166. doi:10.1002/jcp.29337

Kandimalla, R., Xu, J., Link, A., Matsuyama, T., Yamamura, K., Parker, M. I., et al. (2021). EpiPanGI Dx: A Cell-free DNA Methylation Fingerprint for the Early Detection of Gastrointestinal Cancers. *Clin. cancer Res. official J. Am. Assoc. Cancer Res.* 27 (22), 6135–6144. doi:10.1158/1078-0432.CCR-21-1982

King, R., Hayes, C., Donohoe, C. L., Dunne, M. R., Davern, M., and Donlon, N. E. (2021). Hypoxia and its Impact on the Tumour Microenvironment of Gastroesophageal Cancers. *Wjgo* 13 (5), 312–331. doi:10.4251/wjgo.v13.i5.312

Koi, M., and Carethers, J. M. (2017). The Colorectal Cancer Immune Microenvironment and Approach to Immunotherapies. *Future Oncol.* 13 (18), 1633–1647. doi:10.2217/fon-2017-0145

Luo, J., Xie, Y., Zheng, Y., Wang, C., Qi, F., Hu, J., et al. (2020). Comprehensive Insights on Pivotal Prognostic Signature Involved in Clear Cell Renal Cell Carcinoma Microenvironment Using the ESTIMATE Algorithm. *Cancer Med.* 9 (12), 4310–4323. doi:10.1002/cam4.2983

Makaremi, S., Asadzadeh, Z., Hemmat, N., Baghbanzadeh, A., Sgambato, A., Ghorbaninezhad, F., et al. (2021). Immune Checkpoint Inhibitors in Colorectal Cancer: Challenges and Future Prospects. *Biomedicines* 9 (9), 1075. doi:10.3390/biomedicines9091075

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

Markman, J. L., and Shiao, S. L. (2015). Impact of the Immune System and Immunotherapy in Colorectal Cancer. *J. Gastrointest. Oncol.* 6 (2), 208–223. doi:10.3978/j.issn.2078-6891.2014.077

Mayer, R. J., Van Cutsem, E., Falcone, A., Yoshino, T., Garcia-Carbonero, R., Mizunuma, N., et al. (2015). RECOURSE Study GroupRandomized Trial of TAS-102 for Refractory Metastatic Colorectal Cancer. *N. Engl. J. Med.* 372 (20), 1909–1919. doi:10.1056/NEJMoa1414325

Noman, M. Z., Hasmim, M., Lequeux, A., Xiao, M., Duhem, C., Chouaib, S., et al. (2019). Improving Cancer Immunotherapy by Targeting the Hypoxic Tumor Microenvironment: New Opportunities and Challenges. *Cells* 8 (9), 1083. doi:10.3390/cells8091083

Palazon, A., Goldrath, A. W., Nizet, V., and Johnson, R. S. (2014). HIF Transcription Factors, Inflammation, and Immunity. *Immunity* 41 (4), 518–528. doi:10.1016/j.immuni.2014.09.008

Sartore-Bianchi, A., Trusolino, L., Martino, C., Bencardino, K., Lonardi, S., Bergamo, F., et al. (2016). Dual-targeted Therapy with Trastuzumab and Lapatinib in Treatment-Refractory, KRAS Codon 12/13 Wild-type, HER2-Positive Metastatic Colorectal Cancer (HERACLES): a Proof-Of-Concept, Multicentre, Open-Label, Phase 2 Trial. *Lancet. Oncol.* 17 (6), 738–746. doi:10.1016/S1470-2045(16)00150-9

Sconocchia, G., Eppenberger, S., Spagnoli, G. C., Tornillo, L., Droeser, R., Caratelli, S., et al. (2014). NK Cells and T Cells Cooperate during the Clinical Course of Colorectal Cancer. *Oncoimmunology* 3 (8), e952197. doi:10.4161/21624011.2014.952197

Semenza, G. L. (2010). Defining the Role of Hypoxia-Inducible Factor 1 in Cancer Biology and Therapeutics. *Oncogene* 29 (5), 625–634. doi:10.1038/onc.2009.441

Semenza, G. L. (2021). Intratumoral Hypoxia and Mechanisms of Immune Evasion Mediated by Hypoxia-Inducible Factors. *Physiology* 36 (2), 73–83. doi:10.1152/physiol.00034.2020

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2022). Cancer Statistics, 2022. *CA A Cancer J. Clin.* 72 (1), 7–33. doi:10.3322/caac.21708

Song, W., Ren, J., Xiang, R., Kong, C., and Fu, T. (2021). Identification of Pyroptosis-Related Subtypes, the Development of a Prognosis Model, and Characterization of Tumor Microenvironment Infiltration in Colorectal Cancer. *Oncoimmunology* 10 (1), 1987636. doi:10.1080/2162402X.2021.1987636

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

Vymetalkova, V., Vodicka, P., Vodenkova, S., Alonso, S., and Schneider-Stock, R. (2019). DNA Methylation and Chromatin Modifiers in Colorectal Cancer. *Mol. aspects Med.* 69, 73–92. doi:10.1016/j.mam.2019.04.002

Wang, D. Y., Salem, J.-E., Cohen, J. V., Chandra, S., Menzer, C., Ye, F., et al. (2018). Fatal Toxic Effects Associated with Immune Checkpoint Inhibitors. *JAMA Oncol.* 4 (12), 1721–1728. doi:10.1001/jamaoncol.2018.3923

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-seq: a Revolutionary Tool for Transcriptomics. *Nat. Rev. Genet.* 10 (1), 57–63. doi:10.1038/nrg2484

Wolchok, J. D., and Chan, T. A. (2014). Antitumour Immunity Gets a Boost. *Nature* 515 (7528), 496–498. doi:10.1038/515496a

Xiao, M., Liu, L., Zhang, S., Yang, X., and Wang, Y. a. (2018). Cancer Stem Cell Biomarkers for Head and Neck Squamous Cell Carcinoma: A Bioinformatic Analysis. *Oncol. Rep.* 40 (6), 3843–3851. doi:10.3892/or.2018.6771

Yan, Z., He, M., He, L., Wei, L., and Zhang, Y. (2021). Identification and Validation of a Novel Six-Gene Expression Signature for Predicting Hepatocellular Carcinoma Prognosis. *Front. Immunol.* 12, 723271. doi:10.3389/fimmu.2021.723271

You, L., Wu, W., Wang, X., Fang, L., Adam, V., Nepovimova, E., et al. (2021). The Role of Hypoxia-inducible Factor 1 in Tumor Immune Evasion. *Med. Res. Rev.* 41 (3), 1622–1643. doi:10.1002/med.21771

Zhang, S., Han, Q., Zhu, K., and Wang, Q. (2020). The Association of miR-27a Rs895819 Polymorphism with Colorectal Cancer Risk in Chinese Population. *J. Clin. Lab. Anal.* 34 (11), e23497. doi:10.1002/jcla.23497

Zhang, Y., Yapryntseva, M. A., Vdovin, A., Maximchik, P., Zhivotovsky, B., and Gogvadze, V. (2021). Modeling Hypoxia Facilitates Cancer Cell Survival through Downregulation of P53 Expression. *Chemico-biological Interact.* 345, 109553. doi:10.1016/j.cbi.2021.109553

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 Genetics

ORIGINAL RESEARCH
published: 28 June 2022
doi: 10.3389/fgene.2022.939751

Exploration of Shared Gene Signatures and Molecular Mechanisms Between Periodontitis and Nonalcoholic Fatty Liver Disease

Wanqiu Xu[1†], *Zhengwei Zhang*[2†], *Lihong Yao*[1], *Bing Xue*[1], *Hualei Xi*[1], *Xiumei Wang*[1*] and *Shibo Sun*[2*]

[1]Department of Dentistry, The Second Affiliated Hospital of Harbin Medical University, Harbin, China, [2]Ward 7, Department of General Surgery, The Second Affiliated Hospital of Harbin Medical University, Harbin, China

OPEN ACCESS

Edited by:
Tao Huang,
(CAS), China

Reviewed by:
Lifei Ma,
Chinese Academy of Medical Sciences and Peking Union Medical College, China
BL Gan,
Guangxi Medical University, China

***Correspondence:**
Xiumei Wang
hrbwangmei@163.com
Shibo Sun
shibosun8@hrbmu.edu.cn

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

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: 09 May 2022
Accepted: 16 May 2022
Published: 28 June 2022

Citation:
Xu W, Zhang Z, Yao L, Xue B, Xi H, Wang X and Sun S (2022) Exploration of Shared Gene Signatures and Molecular Mechanisms Between Periodontitis and Nonalcoholic Fatty Liver Disease. Front. Genet. 13:939751. doi: 10.3389/fgene.2022.939751

Background: Periodontitis is associated with periodontal tissue damage and teeth loss. Nonalcoholic fatty liver disease (NAFLD) has an intimate relationship with periodontitis. Nevertheless, interacted mechanisms between them have not been clear. This study was intended for the exploration of shared gene signatures and latent therapeutic targets in periodontitis and NAFLD.

Methods: Microarray datasets of periodontitis and NAFLD were obtained from the Gene Expression Omnibus (GEO) database. The weighted gene co-expression network analysis (WGCNA) was utilized for the acquisition of modules bound up with NAFLD and periodontitis. We used ClueGO to carry out biological analysis on shared genes to search their latent effects in NAFLD and periodontitis. Another cohort composed of differential gene analysis verified the results. The common microRNAs (miRNAs) in NAFLD and periodontitis were acquired in the light of the Human microRNA Disease Database (HMDD). According to miRTarbase, miRDB, and Targetscan databases, latent target genes of miRNAs were forecasted. Finally, the miRNAs–mRNAs network was designed.

Results: Significant modules with periodontitis and NAFLD were obtained *via* WGCNA. GO enrichment analysis with GlueGo indicated that damaged migration of dendritic cells (DCs) might be a common pathophysiologic feature of NAFLD and periodontitis. In addition, we revealed common genes in NAFLD and periodontitis, including IGK, IGLJ3, IGHM, MME, SELL, ENPP2, VCAN, LCP1, IGHD, FCGR2C, ALOX5AP, IGJ, MMP9, FABP4, IL32, HBB, FMO1, ALPK2, PLA2G7, MNDA, HLA-DRA, and SLC16A7. The results of differential analysis in another cohort were highly accordant with the findings of WGCNA. We established a comorbidity model to explain the underlying mechanism of NAFLD secondary to periodontitis. Finally, the analysis of miRNA pointed out that hsa-mir-125b-5p, hsa-mir-17-5p, and hsa-mir-21-5p might provide potential therapeutic targets.

Conclusion: Our study initially established a comorbidity model to explain the underlying mechanism of NAFLD secondary to periodontitis, found that damaged migration of DCs

might be a common pathophysiological feature of NAFLD and periodontitis, and provided potential therapeutic targets.

Keywords: periodontitis, nonalcoholic fatty liver disease, WGCNA, dendritic cell migration, miRNAs–mRNAs

INTRODUCTION

Nonalcoholic fatty liver disease (NAFLD), accompanied by varying levels of hepatic fat accumulation, can gradually progress to nonalcoholic steatohepatitis, cirrhosis, and hepatocellular carcinoma, which has fatal consequences (Wesolowski et al., 2017). It was reported that the prevalence of NAFLD accounted roughly 25%, with the prospect of further increase according to expanding populations with metabolic syndrome (Younossi et al., 2016; Estes et al., 2018). According to the pathophysiology of NAFLD, some kinds of medical treatments with respective effects are being assessed in clinical trials. It is regrettable that these drug candidates have been found bringing unpalatable side effects or are limited by efficacy (Alkhouri et al., 2020; Younossi et al., 2021; Barritt et al., 2022). Currently, there has been a lively investigation over the participation of periodontitis in the occurrence and development of NAFLD. Some scholars even believed that there was a comorbidity effect between the two diseases (Rosato et al., 2019). Suffering from oral microbial imbalance brought about by anaerobic Gram-negative bacteria chiefly, periodontitis is associated with periodontal tissue damage and teeth loss (Kuraji et al., 2021). It was reported that there were 1.1 billion people with severe periodontitis worldwide in 2019 (Chen et al., 2021). Actually, mechanical debridement is hard to absolutely clear periodontitis infection and prolonged antibiotic exposure is effective but unsafe (Rotundo et al., 2010; Rams et al., 2020).

From the beginning, periodontitis has contributed to the development of NAFLD owing to systemic inflammation and oxidative stress on the basis of vitro study (Tomofuji et al., 2007). Then, Porphyromonas gingivalis, the main pathogenic bacteria of periodontitis, resulted in the development of NAFLD, above which academic discussion had continued ever since (Furusho et al., 2013; Nagasaki et al., 2021; Yamazaki et al., 2021). Epidemiological investigation reported that NAFLD incidence was increasing with the combination of periodontitis, which could increase the risk of progression to liver fibrosis as well (Akinkugbe et al., 2017a; Akinkugbe et al., 2017b; Iwasaki et al., 2018; Suominen et al., 2019; Kuroe et al., 2021). The potential associations between periodontitis and NAFLD has been discussed from *in vitro*, *in vivo*, and epidemiologic perspectives, but the genetic and biological mechanisms of connection between periodontitis and NAFLD is unknown. Although most studies suggest that periodontitis can affect NAFLD outcomes, the effect of genetic and biological mechanisms might be bidirectional and extremely valuable.

In order to have insights into the mechanisms of diseases, gene microarray technology is developed, which can generate thousands of gene expression data in various diseases. Despite periodontitis and NAFLD being two relatively independent pathological process, periodontitis feels more like a trigger, once it is lit, it will quicken NAFLD aggravation. To explain the trigger, the weighted gene co-expression network analysis (WGCNA) was applied to seek the clusters of shared genes in periodontitis and NAFLD. This method has been utilized to explain genetic mechanism related to various disease phenotypes effectively (Zhu et al., 2020; Yao et al., 2021). Through the deep analysis of the Gene Expression Omnibus (GEO) database, we found that genes related to "dendritic cell migration" were presented in modules hugely relevant to periodontitis and NAFLD, which meant that biological pathway "dendritic cell migration" might play a significant role in periodontitis and NAFLD. In addition, the unique gene signatures in periodontitis and NAFLD were also identified and microRNAs (miRNAs) might play a regulatory role. So far as we know, this is the first study to utilize the bioinformation technique to explain the gene signatures between periodontitis and NAFLD, which is expected to provide new diagnostic and therapeutic windows for these two diseases.

METHODS

Download and Preprocessing of the Gene Expression Omnibus Dataset

We used the key words "Nonalcoholic Fatty Liver Disease" or "periodontitis" to search NAFLD and periodontitis gene expression profiles in which the data at original or processed state could be for the return to analysis in the GEO database (Barrett et al., 2013). Finally, the GEO dataset numbered GSE16134 was accepted, which contained a total of 241 periodontitis samples and 69 healthy samples. The GSE48452 and GSE63067 microarray datasets were used for NAFLD, which contained raw transcriptomics data from the human liver tissue. In GSE48452 dataset, 73 samples of human liver grouped into C (control = 14), H (healthy obesity = 27), S (steatosis = 14), and N (NASH = 18) from original references. In GSE63067dataset, two human steatosis and nine human nonalcoholic steatohepatitis (NASH) together with their respective control patterns were analyzed from original references. The original data were processed with background correction, normalization, and relative expression calculation. Log2 transformation was applied to gene expression profiling and the probes were matched with their gene symbols on the basis of annotated files from relevant platforms. Ultimately, we acquired the genetic matrix with row and column defined as specimen names and gene symbols, respectively, for the following analysis.

Weighted Gene Co-Expression Network Analysis

A popular algorithm, WGCNA, is applied to seek gene co-expression modules with the great importance of biology and discover the relevance between diseases and gene networks

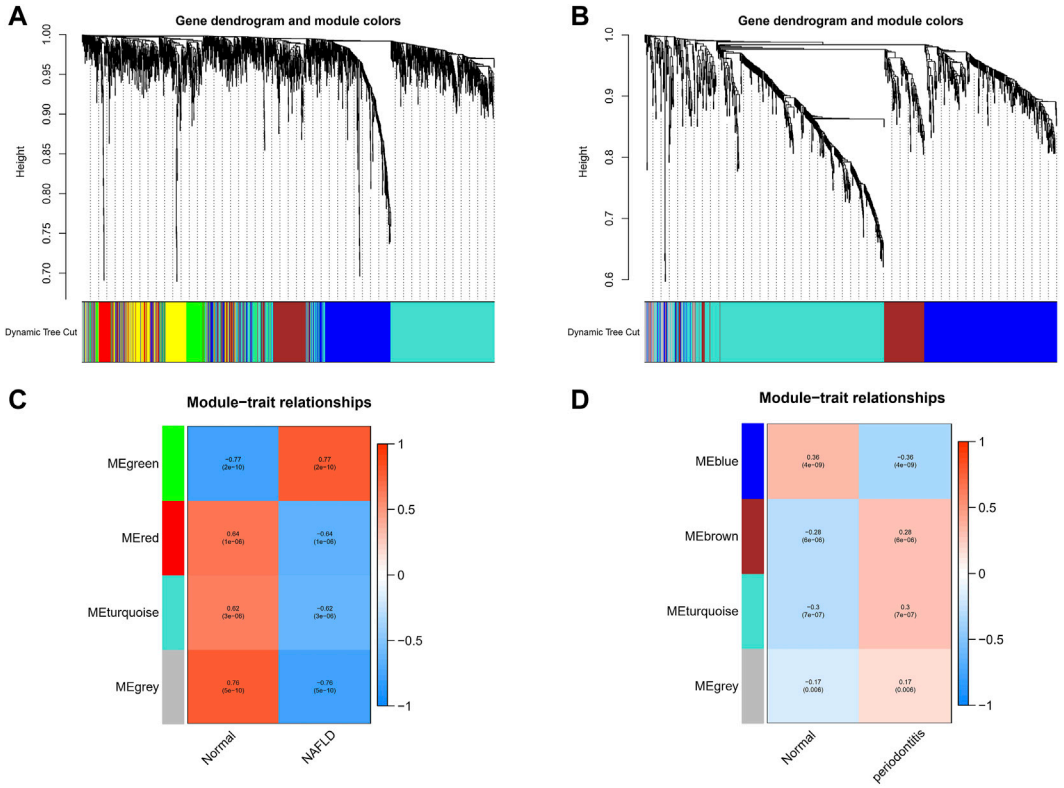


FIGURE 1 | Weighted gene co-expression network analysis (WGCNA). **(A)** The cluster dendrogram of co-expression genes in NAFLD. **(B)** The cluster dendrogram of co-expression genes in periodontitis. **(C)** Module–trait relationships in NAFLD. Each cell contains the corresponding correlation and *p*-value. **(D)** Module–trait relationships in periodontitis. Each cell contains the corresponding correlation and *p*-value. NAFLD, nonalcoholic fatty liver disease.

(Langfelder and Horvath, 2008). Consequently, WGCNA was utilized for the acquisition of modules bound up with NAFLD and periodontitis. All the differential genes (DEGs) from healthy and disease samples satisfying p value < 0.05 were collected for WGCNA analysis (**Supplementary Data S1**). Clustering of samples was doing well and the threshold of cutting line was 30. The soft thresholds ranging from 1 to 20 were used for topology calculation and optimum soft threshold was identified as 6. According to the soft threshold, the matrix of correlations was converted to the adjacency matrix and then into a topological overlap matrix (TOM). With the average-linkage hierarchical clustering method which followed, the genes were clustered. The modules were divided according to TOM, each of which contained at least 50 genes. The cutting height of gene module was 0.7 and similar modules were combined. After that, gene significance (GS) and module membership (MM) in every module were calculated for plotting the scatter plots. At last we applied Pearson correlation analysis to estimate the relevance of disease emergence with the merged modules.

Identification of Shared and Unique Gene Signatures

The modules with high correlation with NAFLD and periodontitis were chosen and the shared genes in modules positively related to NAFLD and periodontitis were crossed and overlapped through venn (Bardou et al., 2014). The nonredundant GO terms can be classified and visually arranged into networks grouped by functions through ClueGO, which is a Cytoscape plug-in unit (Bindea et al., 2009). Hence, we used ClueGO to carry out biological analysis on the shared genes to search their latent effects in NAFLD and periodontitis, in which the biological process (BP) of GO analysis was highlighted. The unique gene signatures in NAFLD and periodontitis were distinguished through the protein–protein interaction (PPI) network and cluster analysis, the latter of which was calculated by the "MCODE" algorithm with default parameters in Cytoscape software (version: 3.7.2).

RESULTS

The Co-Expression Modules in Periodontitis and Nonalcoholic Fatty Liver Disease

With the application of WGCNA, four modules in total were recognized in GSE48452 and GSE63067, each of which had different color betokening separate module. For the assessment of relevance between disease and each module, a heatmap was plotted on the basis of Spearman correlation coefficient, in which module "green" had the highest relevance to NAFLD (**Figures**

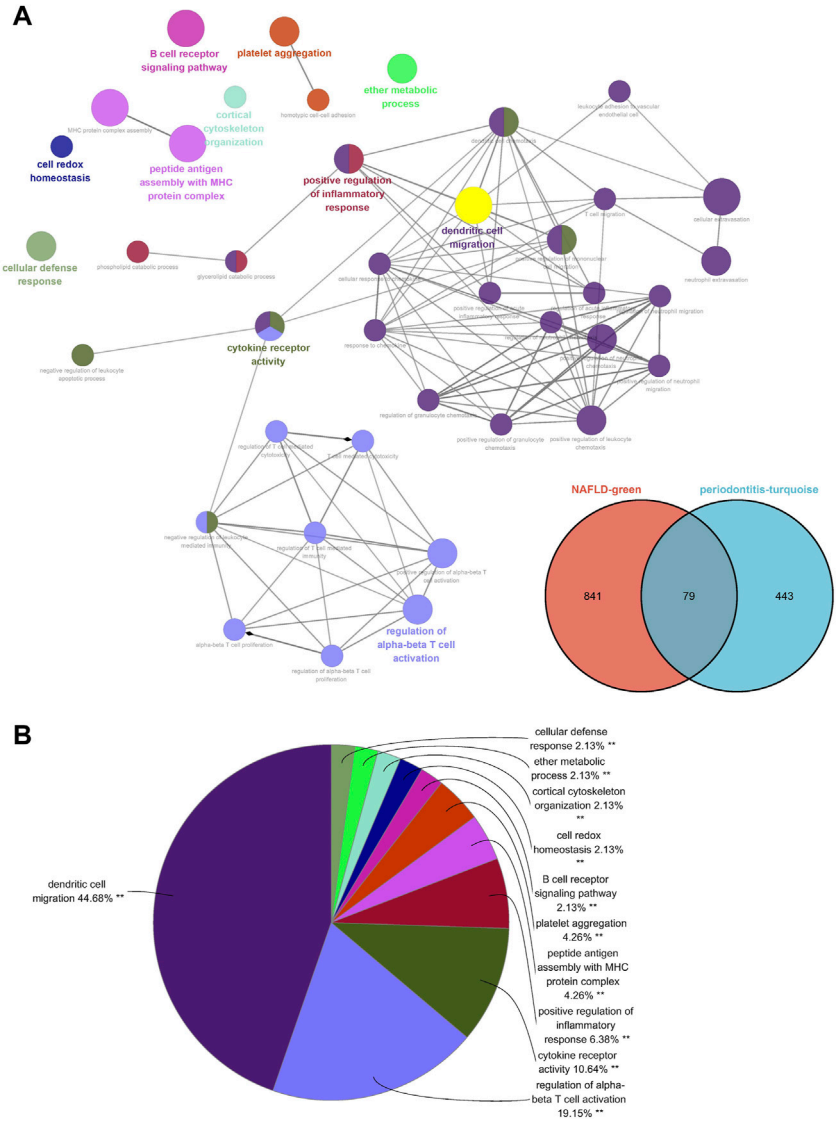


FIGURE 2 | ClueGO enrichment analysis. **(A)** The interaction network of GO terms generated by the Cytoscape plug-in ClueGO. **(B)** Proportion of each GO terms group in the total. GO, gene ontology. $^{**}p < 0.05$.

1A,C). The module, with the core (r = 0.77), was positively related to NAFLD, including 920 genes. Four modules in total were recognized equally in GSE16134, in which the module "cyan" was the strongest and positively related to periodontitis (r = 0.3), including 522 genes (**Figures 1B,D**).

The Common Gene Signatures in Periodontitis and Nonalcoholic Fatty Liver Disease

Seventy-nine genes were crossed and overlapped in the relevant core modules of NAFLD and periodontitis, which was recognized as gene set 1 (GS1). Periodontitis could be the important risk factor for the development of NAFLD according to current study. GlueGo was used to discuss the latent functions of GS1 through the GO enrichment analysis. The top three markedly enriched GO terms about BP were "dendritic cell migration," "regulation of alpha-beta T cell activation," and "cytokine receptor activity" (**Figure 2A**). Dendritic cell migration represented 44.68% of all the GO terms (**Figure 2B**), meaning that this pathway might be vital to both NAFLD and periodontitis.

The Unique Gene Signatures in Periodontitis and Nonalcoholic Fatty Liver Disease

A PPI network was subsequently established at protein levels for green module of NAFLD. MCODE analysis was applied to acquire clusters. There were 34 nodes and 274 edges in cluster 1 (score = 16.606) (**Figure 3A**). Cluster 2 embodied 13 nodes and

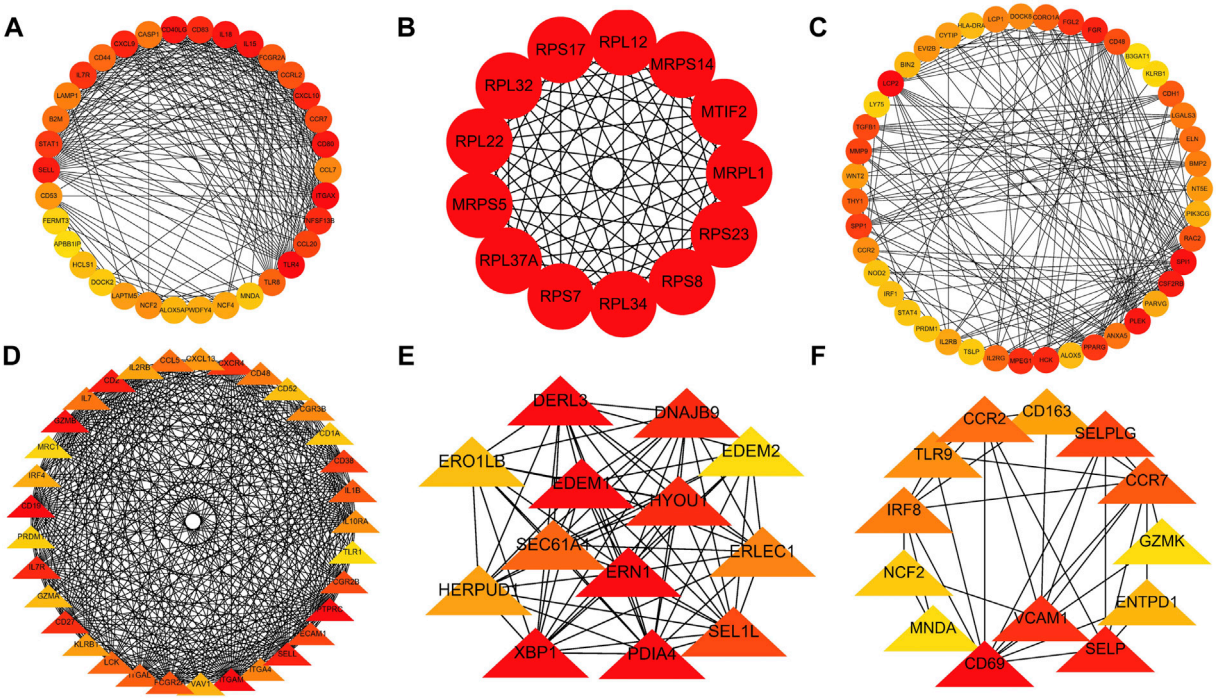


FIGURE 3 | The PPI network. **(A–C)** The clusters 1-3 extracted from green module in NAFLD. **(D–F)** The clusters 1-3 extracted from the cyan module in periodontitis. PPI, protein–protein network; NAFLD, nonalcoholic fatty liver disease.

78 edges (score = 13.000) (**Figure 3B**). Cluster 3 embodied 43 nodes and 209 edges (score = 9.952) (**Figure 3C**). Cluster 3 was primarily related to dendritic cell migration, which was represented with the functional enrichment analysis (**Figure 4A**). Consequently, it was inferred that cluster 3 pertained to common genes section from NAFLD and periodontitis. The other two clusters were recognized as unique gene signatures in NAFLD. The PPI network was established at protein levels for cyan module of periodontitis equally. MCODE analysis was also applied to acquire the clusters. There were 33 nodes and 387 edges in cluster 1 (score = 24.188) (**Figure 3D**). Cluster 2 embodied 13 nodes and 71 edges (score = 11.833) (**Figure 3E**). Cluster 3 embodied 13 nodes and 34 edges (score = 5.667) (**Figure 3F**). Coincidentally, cluster 3 was primarily related to dendritic cell migration, which was represented with the functional enrichment analysis (**Figure 4B**). The other two clusters were recognized as unique gene signatures in periodontitis.

The Differential Genes Analysis in Periodontitis and Nonalcoholic Fatty Liver Disease

There were 91 upregulated genes and 33 downregulated genes being represented in GSE48452 and GSE63067. Concurrently, there were 664 upregulated genes and 402 downregulated genes represented in GSE16134. Of all the upegulated genes, 21 overlapped genes were discovered (IGK, IGLJ3, IGHM, MME, SELL, ENPP2, VCAN, LCP1, IGHD, FCGR2C, ALOX5AP, IGJ, MMP9, FABP4, IL32, HBB, FMO1, ALPK2, PLA2G7, MNDA, and HLA-DRA). On the other hand, one downregulated gene was overlapped (SLC16A7), which were recognized as gene set 2 (GS2) (**Figure 5A**). Representing remarkable enrichment of dendritic cell migration, dendritic cell chemotaxis, and neutral lipid catabolic process, the genes of GS2 were explored through the functional enrichment analysis, which was highly accordant with the findings of WGCNA (**Figure 5B**).

Identification and Analysis of Common miRNAs in Periodontitis and Nonalcoholic Fatty Liver Disease

In the light of the Human microRNA Disease Database (HMDD) (Huang et al., 2019), 43 miRNAs were found to be related to NAFLD and 33 miRNAs were related to periodontitis (**Supplementary Data S2**). There were five overlapped miRNAs (hsa-mir-125b-5p, hsa-mir-155-5p, hsa-mir-17-5p, hsa-mir-200b-5p, and hsa-mir-21-5p) between NAFLD and periodontitis. There followed the enrichment analysis of five miRNAs, which revealed a variety of biological functions that these miRNAs are involved in. Similarly, "dendritic cell migration" got involved in these biological processes according to the heatmap, signifying that miRNAs associated with pathogenesis of NAFLD and periodontitis could also regulate dendritic cell migration (**Figure 6A**). Hence, our findings were proved again. According to miRTarbase (Chou et al., 2018), miRDB (Chen and Wang, 2020), and Targetscan (Morovat et al., 2022) databases, latent target genes of five miRNAs were

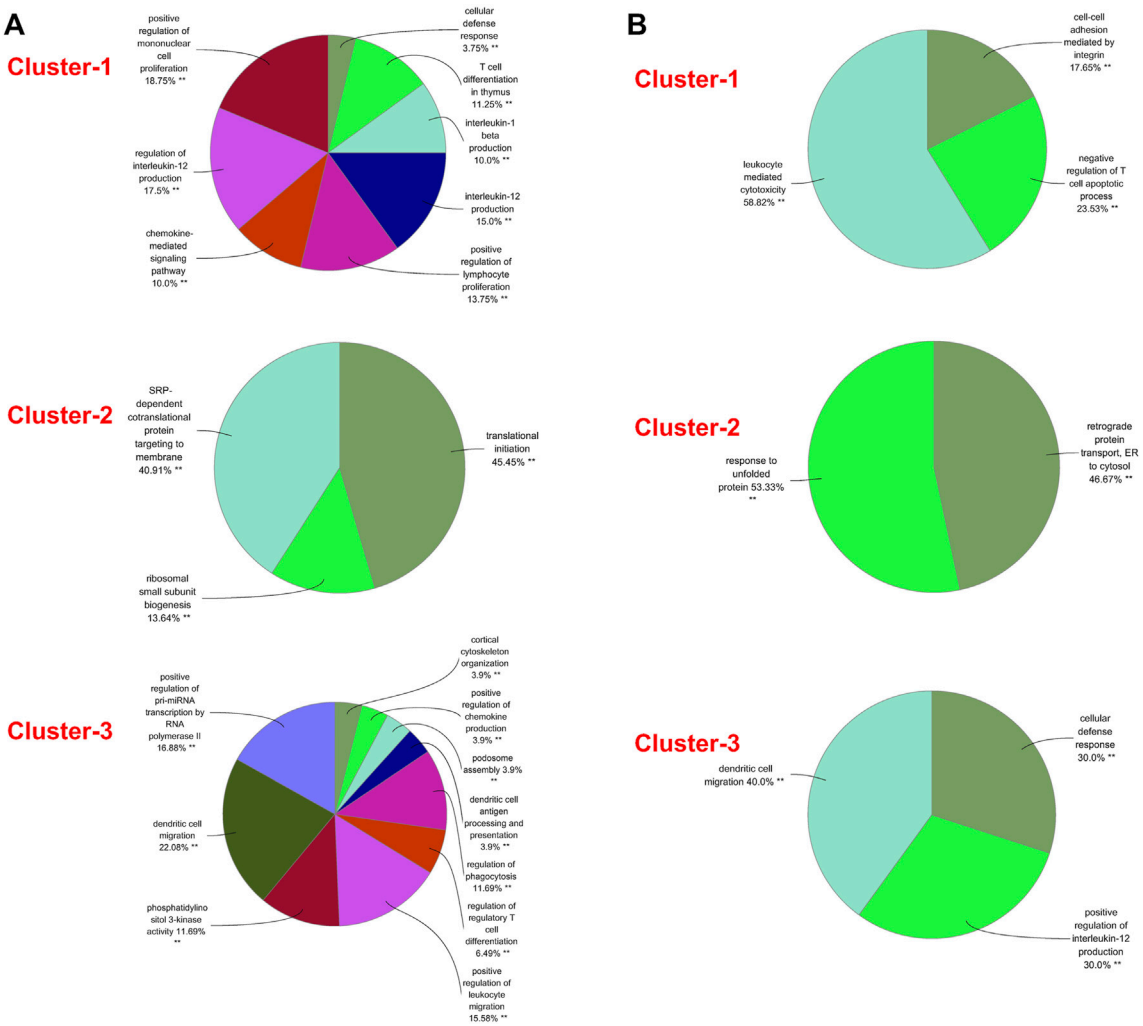


FIGURE 4 | GO biological process analyses of clusters. **(A)** The GO biological process analyses of three gene clusters in NAFLD. **(B)** The GO biological process analyses of three genes clusters in periodontitis. GO, gene ontology; NAFLD, nonalcoholic fatty liver disease. $^{**}p < 0.05$.

forecasted (**Figure 6B**). Unfortunately, hsa-mir-155-5p was not retrieved in the database, and hsa-mir-200b-5p had no overlapped target genes. Finally, the miRNAs–mRNAs network was designed (**Figure 6C**).

DISCUSSION

As noted earlier, NAFLD has a high prevalence in periodontitis, indicating that some susceptibility factors in periodontitis may trigger the initiation and progression of NAFLD. Although it is not yet clear that how hazardous factors are delivered to liver from periodontium, the following two routes have been highly accepted. Blood transmission of bacteria, endotoxin, and inflammatory mediators from the periodontal tissues is the first aspect correlating periodontitis and NAFLD. Delivery of oral bacteria *via* the digestive tract is the second aspect, which brings out the imbalance of the intestinal bacteria (Kuraji et al., 2021). Regardless of dangerous medium such as periodontal bacteria, lipopolysaccharide and proinflammatory mediators, or intestinal dysbacteriosis, the precise role of them in effect of periodontitis on NAFLD needs further studies. So far, no studies have discussed the susceptibility of NAFLD in periodontitis at the genetic level.

Drawing support from WGCNA, we first discussed the common mechanisms of periodontitis and NAFLD. The differentially expressed genes in common were found in the intersection of GS1 and GS2, such as VCAN, LCP1, and ENPP2. Functional enrichment analysis concerned included dendritic cell migration, regulation of alpha-beta T cell activation, cytokine receptor activity, dendritic cell chemotaxis, and neutral lipid catabolic process. Finally, the miRNAs–mRNAs network was designed. More importantly, genes related to "dendritic cell migration" were presented in modules hugely relevant to periodontitis and NAFLD and experienced repeated verification. In addition, miRNAs might play a regulatory role in periodontitis and NAFLD. Playing a major

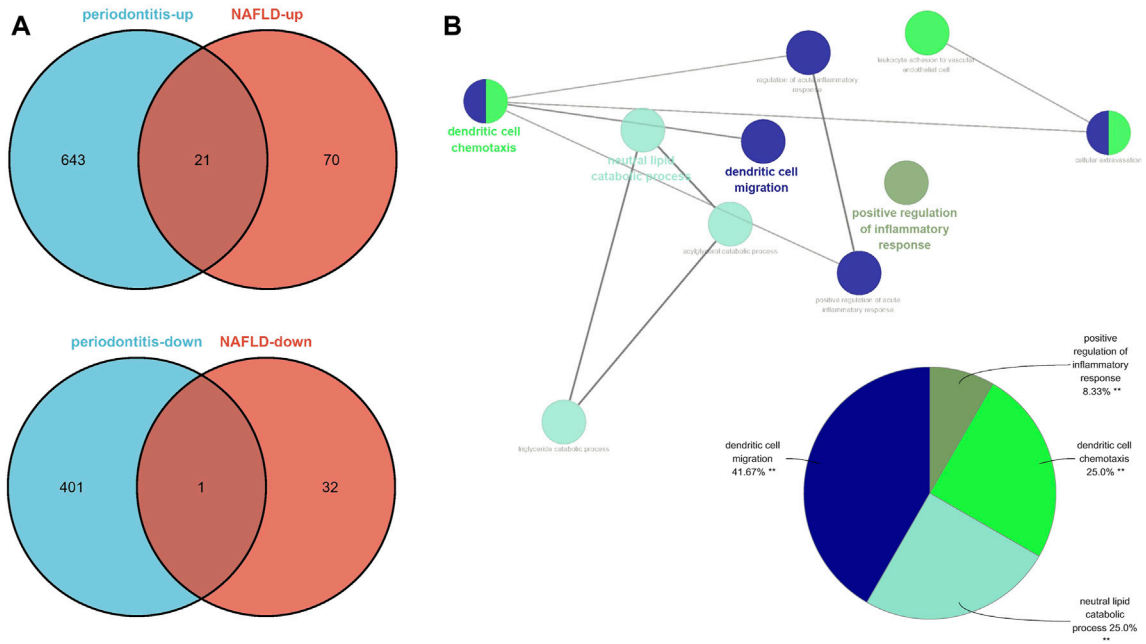


FIGURE 5 | Identification of the common DEGs and ClueGO enrichment analysis. **(A)** The Venn diagram of the upregulated and downregulated genes in periodontitis and NAFLD. **(B)** The interaction network of GO terms generated by the Cytoscape plug-in ClueGO and proportion of GO terms in the total. GO, gene ontology; NAFLD, nonalcoholic fatty liver disease. $^{**}p < 0.05$.

role in innate immunity, dendritic cells (DCs) could capture and present antigens, which are also the bond to adaptive immunity (Steinman, 2001). The research shows that transmission of bacteria from periodontal tissues to distant sites *via* systemic circulation might appear at highly migrated DCs (Carrion et al., 2012). Porphyromonas gingivalis, as major pathogens in periodontitis, can attack DCs, reduce the level of proapoptosis protein expression, and prolong the survival of DCs (Meghil et al., 2019). This type of bacteria not only damages immune homeostasis of DCs, but also disrupts DCs homing to secondary lymphoid organs, the latter of which makes the inflammation migrate to vascular circulation (Miles et al., 2014). Unfortunately, it could avoid intracellular killing in DCs by targeting to dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin (El-Awady et al., 2015). However, oral microbial diversity destines that Porphyromonas gingivalis do not fight alone. In a previous study, a union of three oral microorganisms, *Streptococcus gordonii, Fusobacterium nucleatum, and Porphyromonas gingivalis*, drove bacterial growth, attack and stability in DCs, and regressed DCs maturation *via* coordinated effects, which generated microbial transmission and inflammatory spread (El-Awady et al., 2019). After-effects of bacteria themselves are taken out, lipopolysaccharide or proinflammatory cytokines, coming from periodontitis and bringing about low-grade systemic inflammatory state, is closely related to DCs (Kanaya et al., 2004; Jardine et al., 2019; Psarras et al., 2021). With its receptors distributed extensively in the human body, inactive gingipains, as critical virulence factors of *Porphyromonas gingivalis*, leads to proinflammatory response in DCs (Ciaston et al., 2022). All in all, DCs not only play a central role in initiating and exacerbating periodontitis but also could be considered as potential contributing factors to the development of systemic diseases related to periodontitis, one of which is NAFLD.

It is generally known that intestinal microbial imbalance is intimately connected to NAFLD. First, anomalous abundance changes of bacterial phyla affect the severity of NAFLD (Boursier et al., 2016). Second, metabolite of intestinal bacteria results in fatty degeneration of liver cells, insulin resistance, and hepatic fibrosis (Ji et al., 2019). Third, endotoxemia attributed to the increase in intestinal permeability is related to pathogenesis of NAFLD (Wang et al., 2022). However, the mechanism of the pathology in which the intestinal flora imbalance induced by oral bacteria contributes to NAFLD has been unclear. Studies have pointed out that *Porphyromonas gingivalis* plays a major role in the process *via* interfering with the metabolic and immune profiles (Wang et al., 2022). It is not clear if DCs also affect the transmission of pathogenic bacteria and their toxic metabolites to the liver through the portal vein. The existing fact remains that the physiological action of DCs can be affected by the intestinal microbes (Yang et al., 2021). On the other hand, numerous researches have proved that migratory DCs could dominate induction of enteric T regulatory cells to manage commensal bacteria or to set up oral tolerance targeted at dietary antigens (Esterházy et al., 2016; Esterházy et al., 2019; Russler-Germain et al., 2021). Although the action mechanism of DCs in NAFLD is not completely clear, existing studies have confirmed the important role of DCs. DCs play a proinflammatory role in the animal models with nonalcoholic steatohepatitis (NASH). With CD11c+ DCs or CD103+ DCs

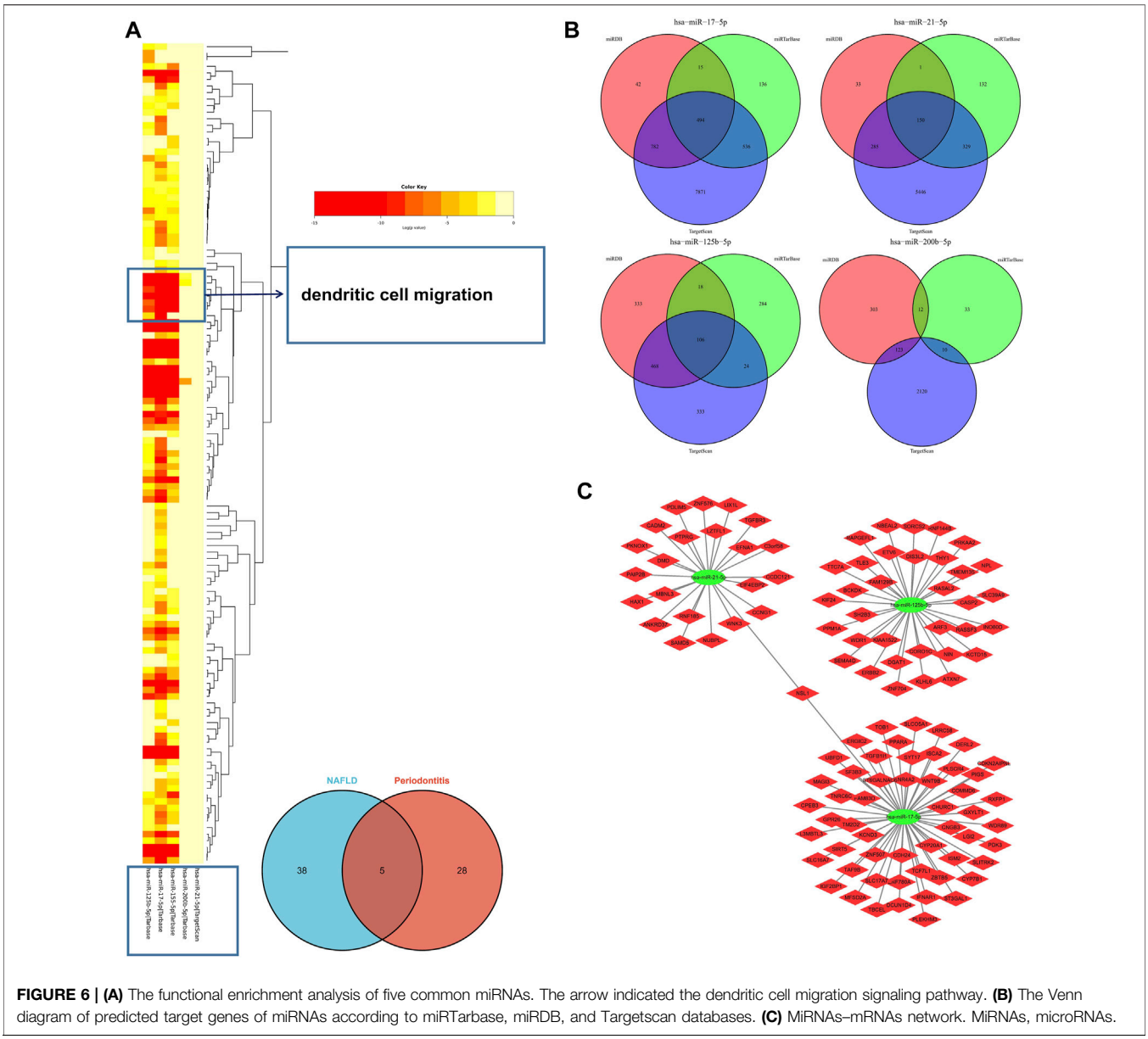


FIGURE 6 | (A) The functional enrichment analysis of five common miRNAs. The arrow indicated the dendritic cell migration signaling pathway. **(B)** The Venn diagram of predicted target genes of miRNAs according to miRTarbase, miRDB, and Targetscan databases. **(C)** MiRNAs–mRNAs network. MiRNAs, microRNAs.

consumption, decreased expression of proinflammatory cytokines and chemokines could prevent liver fibrosis (Nati et al., 2016; Schuster et al., 2018). Recent research has also shown that depletion of type 1 conventional DCs attenuates liver pathology in the NASH mouse models (Deczkowska et al., 2021). As noted previously, NAFLD has a high morbidity in periodontitis, indicating that predisposing factors in periodontitis could touch off NAFLD. In our modeling, both the discovery cohort and validation cohort reached the conclusion that dendritic cell migration played an important part in gene function enrichment analysis. Previous studies also support our view. Consequently, damaged migration of DCs might be a common pathophysiologic feature of NAFLD and periodontitis, which means that dendritic cell migration plays a key role and provides critical therapeutic target in the comorbidity model. MiRNA, as endogenous noncoding regulatory RNA, plays huge roles in the regulation of post-transcriptional gene. We have constructed the miRNAs–mRNAs network with the benefit of HMDD, miRTarbase, miRDB, and Targetscan databases. Interestingly, the target genes of common miRNAs, having no intersection with GS1 and GS2, still enriched in “dendritic cell migration”, which might be related to the indirect interaction of genes. Among these miRNAs, epigenetic silencing of miR-125b-5p resulted in liver fibrosis in NAFLD (Cai et al., 2020). Differential expression of miR-125b-5p influenced the functions of DCs (Hu et al., 2017). Mast cells had a close associate of periodontitis, and overexpressed miR-125b-5p in its own exosomes (Ekström et al., 2012; Tetè et al., 2021). We speculated that periodontitis might be affected by miR-125b-5p.

Similarly, miR-17-5p and miR-21-5p were reported to play a part in periodontitis and migration of DCs and were predicted to get involved in NAFLD (Du et al., 2016; Kim et al., 2017; Reis et al., 2018; Cui et al., 2019; Zhang et al., 2020; Lin et al., 2022). Although these miRNAs have not been verified in the microenvironment of comorbidity with periodontitis and NAFLD, they also provide important therapeutic targets.

Considering the reality of the situation, experimental validation is currently not possible because clinical specimens of NAFLD are extremely difficult to obtain. Therefore, this is a limitation of our study and we will gradually collect samples for vitro assays. All in all, our study has established a comorbidity model to explain the underlying mechanism of NAFLD secondary to periodontitis, found that damaged migration of DCs might be a common pathophysiologic feature of NAFLD and periodontitis, and provided potential therapeutic targets.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

WX and ZZ designed and conducted the whole research. WX and ZZ wrote the majority of the manuscript. LY, BX, and HX contributed to data collation. XW and SS revised and finalized the manuscript. All authors contributed to the article and approved the submitted version. WX and ZZ have contributed equally to this work.

FUNDING

This work was supported by the Key Program of Natural Science Foundation of Heilongjiang Province of China [ZD 2019H002 to X.M.W.].

ACKNOWLEDGMENTS

The authors thank the authors of the GSE16134, GSE48452, and GSE63067 datasets for their contribution.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.939751/full#supplementary-material

Supplementary Data S1 | All differential genes (DEGs) from health and disease samples.

Supplementary Data S2 | 43 miRNAs related to NAFLD and 33 miRNAs related to periodontitis according to HMDD. HMDD, Human microRNA Disease Database.

REFERENCES

Akinkugbe, A. A., Avery, C. L., Barritt, A. S., Cole, S. R., Lerch, M., Mayerle, J., et al. (2017). Do Genetic Markers of Inflammation Modify the Relationship between Periodontitis and Nonalcoholic Fatty Liver Disease? Findings from the SHIP Study. *J. Dent. Res.* 96 (12), 1392–1399. doi:10.1177/0022034517720924

Akinkugbe, A. A., Slade, G. D., Barritt, A. S., Cole, S. R., Offenbacher, S., Petersmann, A., et al. (2017). Periodontitis and Non-alcoholic Fatty Liver Disease, a Population-Based Cohort Investigation in the Study of Health in Pomerania. *J. Clin. Periodontol.* 44 (11), 1077–1087. doi:10.1111/jcpe.12800. Epub.2017.Sep.22

Alkhouri, N., Lawitz, E., Noureddin, M., DeFronzo, R., and Shulman, G. I. (2020). GS-0976 (Firsocostat): an Investigational Liver-Directed Acetyl-CoA Carboxylase (ACC) Inhibitor for the Treatment of Non-alcoholic Steatohepatitis (NASH). *Expert Opin. Investigational Drugs* 29 (2), 135–141. doi:10.1080/13543784.2020.1668374

Bardou, P., Mariette, J., Escudié, F., Djemiel, C., and Klopp, C. (2014). Jvenn: an Interactive Venn Diagram Viewer. *BMC Bioinforma.* 15 (1), 293. doi:10.1186/1471-2105-15-293

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 (Database issue), D991–D995. doi:10.1093/nar/gks1193

Barritt, A. S., 4th, Marshman, E., and Noureddin, M. (2022). Review Article: Role of Glucagon-like Peptide-1 Receptor Agonists in Non-alcoholic Steatohepatitis, Obesity and Diabetes-What Hepatologists Need to Know. *Aliment. Pharmacol. Ther.* 55 (8), 944–959. doi:10.1111/apt.16794

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 (8), 1091–1093. doi:10.1093/bioinformatics/btp101

Boursier, J., Mueller, O., Barret, M., Machado, M., Fizanne, L., Araujo-Perez, F., et al. (2016). The Severity of Nonalcoholic Fatty Liver Disease Is Associated with Gut Dysbiosis and Shift in the Metabolic Function of the Gut Microbiota. *Hepatology* 63 (3), 764–775. doi:10.1002/hep.28356

Cai, Q., Chen, F., Xu, F., Wang, K., Zhang, K., Li, G., et al. (2020). Epigenetic Silencing of microRNA-125b-5p Promotes Liver Fibrosis in Nonalcoholic Fatty Liver Disease via Integrin α8-mediated Activation of RhoA Signaling Pathway. *Metabolism* 104, 154140. doi:10.1016/j.metabol.2020.154140

Carrion, J., Scisci, E., Miles, B., Sabino, G. J., Zeituni, A. E., Gu, Y., et al. (2012). Microbial Carriage State of Peripheral Blood Dendritic Cells (DCs) in Chronic Periodontitis Influences DC Differentiation, Atherogenic Potential. *J. I.* 189 (6), 3178–3187. doi:10.4049/jimmunol.1201053

Chen, M. X., Zhong, Y. J., Dong, Q. Q., Wong, H. M., and Wen, Y. F. (2021). Global, Regional, and National Burden of Severe Periodontitis, 1990-2019: An Analysis of the Global Burden of Disease Study 2019. *J. Clin. Periodontol.* 48 (9), 1165–1188. doi:10.1111/jcpe.13506

Chen, Y., and Wang, X.(2020) miRDB: an Online Database for Prediction of Functional microRNA Targets. *Nucleic Acids Res.*, 48(D1):D127–D131. doi:10.1093/nar/gkz757

Chou, C.-H., Shrestha, S., Yang, C.-D., Chang, N.-W., Lin, Y.-L., Liao, K.-W., et al. (2018). miRTarBase Update 2018: a Resource for Experimentally Validated microRNA-Target interactionsmiRDB: an Online Database for Prediction of Functional microRNA Targets. *Nucleic Acids ResNucleic Acids Res.* 4648 (D1D1), D296D127–D302D131. doi:10.1093/nar/gkx1067.29.Chen10.1093/nar/gkz757

Ciaston, I., Budziaszek, J., Satala, D., Potempa, B., Fuchs, A., Rapala-Kozik, M., et al. (2022). Proteolytic Activity-independent Activation of the Immune Response by Gingipains from Porphyromonas Gingivalis. *mBio* 2022, e0378721. doi:10.1128/mbio.03787-21

Cui, Z.-J., Xie, X.-L., Qi, W., Yang, Y.-C., Bai, Y., Han, J., et al. (2019). Cell-free miR-17-5p as a Diagnostic Biomarker for Gastric Cancer Inhibits Dendritic Cell Maturation. *Ott* 12, 2661–2675. doi:10.2147/OTT.S197682

Deczkowska, A., David, E., Ramadori, P., Pfister, D., Safran, M., Li, B., et al. (2021). XCR1+ Type 1 Conventional Dendritic Cells Drive Liver Pathology in Non-alcoholic Steatohepatitis. *Nat. Med.* 27 (6), 1043–1054. doi:10.1038/s41591-021-01344-3

Du, A., Zhao, S., Wan, L., Liu, T., Peng, Z., Zhou, Z., et al. (2016). Micro RNA Expression Profile of Human Periodontal Ligament Cells under the Influence of Porphyromonas Gingivalis LPS. *J. Cell. Mol. Med.* 20 (7), 1329–1338. doi:10.1111/jcmm.12819

Ekström, K., Valadi, H., Sjöstrand, M., Malmhäll, C., Bossios, A., Eldh, M., et al. (2012). Characterization of mRNA and microRNA in Human Mast Cell-Derived Exosomes and Their Transfer to Other Mast Cells and Blood CD34 Progenitor Cells. *J. Extracell. Vesicles* 1, 18389. doi:10.3402/jev.v1i0.18389

El-Awady, A., de Sousa Rabelo, M., Meghil, M. M., Rajendran, M., Elashiry, M., Stadler, A. F., et al. (2019). Polymicrobial Synergy within Oral Biofilm Promotes Invasion of Dendritic Cells and Survival of Consortia Members. *NPJ Biofilms Microbiomes* 5 (1), 11. doi:10.1038/s41522-019-0084-7

El-Awady, A. R., Miles, B., Scisci, E., Kurago, Z. B., Palani, C. D., Arce, R. M., et al. (2015). Porphyromonas Gingivalis Evasion of Autophagy and Intracellular Killing by Human Myeloid Dendritic Cells Involves DC-SIGN-TLR2 Crosstalk. *PLoS Pathog.* 11 (2), e1004647. doi:10.1371/journal.ppat.1004647

Esterházy, D., Canesso, M. C. C., Mesin, L., Muller, P. A., de Castro, T. B. R., Lockhart, A., et al. (2019). Compartmentalized Gut Lymph Node Drainage Dictates Adaptive Immune Responses. *Nature* 569 (7754), 126–130. doi:10.1038/s41586-019-1125-3

Esterházy, D., Loschko, J., London, M., Jove, V., Oliveira, T. Y., and Mucida, D. (2016). Classical Dendritic Cells are Required for Dietary Antigen-Mediated Induction of Peripheral Treg Cells and Tolerance. *Nat. Immunol.* 17 (5), 545–555. doi:10.1038/ni.3408

Estes, C., Anstee, Q. M., Arias-Loste, M. T., Bantel, H., Bellentani, S., Caballeria, J., et al. (2018). Modeling NAFLD Disease Burden in China, France, Germany, Italy, Japan, Spain, United Kingdom, and United States for the Period 2016-2030. *J. Hepatology* 69 (4), 896–904. doi:10.1016/j.jhep.2018.05.036

Furusho, H., Miyauchi, M., Hyogo, H., Inubushi, T., Ao, M., Ouhara, K., et al. (2013). Dental Infection of Porphyromonas Gingivalis Exacerbates High Fat Diet-Induced Steatohepatitis in Mice. *J. Gastroenterol.* 48 (11), 1259–1270. doi:10.1007/s00535-012-0738-1

Hu, X., Chen, Q., Sowrirajan, B., Bosche, M., Imamichi, T., and Sherman, B. (2017). Genome-Wide Analyses of MicroRNA Profiling in Interleukin-27 Treated Monocyte-Derived Human Dendritic Cells Using Deep Sequencing: A Pilot Study. *Int. j. Mol. Sci.* 18 (5), 925. doi:10.3390/ijms18050925

Huang, Z., Shi, J., Gao, Y., Cui, C., Zhang, S., Li, J., et al. (2019). HMDD v3.0: a Database for Experimentally Supported Human microRNA-Disease Associations. *Nucleic Acids Res.* 47 (D1), D1013–D1017. doi:10.1093/nar/gky1010

Iwasaki, T., Hirose, A., Azuma, T., Ohashi, T., Watanabe, K., Obora, A., et al. (2018). Correlation between Ultrasound-Diagnosed Non-alcoholic Fatty Liver and Periodontal Condition in a Cross-Sectional Study in Japan. *Sci. Rep.* 8 (1), 7496. doi:10.1038/s41598-018-25857-z

Jardine, L., Wiscombe, S., Reynolds, G., McDonald, D., Fuller, A., Green, K., et al. (2019). Lipopolysaccharide Inhalation Recruits Monocytes and Dendritic Cell Subsets to the Alveolar Airspace. *Nat. Commun.* 10 (1), 1999. doi:10.1038/s41467-019-09913-4

Ji, Y., Yin, Y., Li, Z., and Zhang, W. (2019). Gut Microbiota-Derived Components and Metabolites in the Progression of Non-alcoholic Fatty Liver Disease (NAFLD). *Nutrients* 11 (8), 1712. doi:10.3390/nu11081712

Kanaya, S., Nemoto, E., Ogawa, T., and Shimauchi, H. (2004). Porphyromonas Gingivalis Lipopolysaccharides Induce Maturation of Dendritic Cells with $CD14^{+}CD16^{+}$ Phenotype. *Eur. J. Immunol.* 34 (5), 1451–1460. doi:10.1002/eji.200324549

Kim, J.-H., Lee, B.-R., Choi, E.-S., Lee, K.-M., Choi, S.-K., Cho, J. H., et al. (2017). Reverse Expression of Aging-Associated Molecules through Transfection of miRNAs to Aged Mice. *Mol. Ther. - Nucleic Acids* 6, 106–115. doi:10.1016/j.omtn.2016.11.005

Kuraji, R., Sekino, S., Kapila, Y., and Numabe, Y. (2021). Periodontal Disease-Related Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis: An Emerging Concept of Oral-liver axis. *Periodontol.* 87 (1), 204–240. doi:10.1111/prd.12387

Kuroe, K., Furuta, M., Takeuchi, K., Takeshita, T., Suma, S., Shinagawa, T., et al. (2021). Association between Periodontitis and Fibrotic Progression of Non-alcoholic Fatty Liver Among Japanese Adults. *J. Clin. Periodontol.* 48 (3), 368–377. doi:10.1111/jcpe.13415

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

Lin, H., Mercer, K. E., Ou, X., Mansfield, K., Buchmann, R., Børsheim, E., et al. (2022). Circulating microRNAs Are Associated with Metabolic Markers in Adolescents with Hepatosteatosis. *Front. Endocrinol.* 13, 856973. doi:10.3389/fendo.2022.856973

Meghil, M. M., Tawfik, O. K., Elashiry, M., Rajendran, M., Arce, R. M., Fulton, D. J., et al. (2019). Disruption of Immune Homeostasis in Human Dendritic Cells via Regulation of Autophagy and Apoptosis by Porphyromonas Gingivalis. *Front. Immunol.* 10, 2286. doi:10.3389/fimmu.2019.02286

Miles, B., Zakhary, I., El-Awady, A., Scisci, E., Carrion, J., O'Neill, J. C., et al. (2014). Secondary Lymphoid Organ Homing Phenotype of Human Myeloid Dendritic Cells Disrupted by an Intracellular Oral Pathogen. *Infect. Immun.* 82 (1), 101–111. Epub 2013 Oct 14. doi:10.1128/IAI.01157-13

Morovat, P., Morovat, S., Ashrafi, A. M., and Teimourian, S. (2022). Identification of Potentially Functional Circular RNAs Hsa_circ_0070934 and Hsa_circ_0004315 as Prognostic Factors of Hepatocellular Carcinoma by Integrated Bioinformatics Analysis. *Sci. Rep.* 12 (1), 4933. doi:10.1038/s41598-022-08867-w

Nagasaki, A., Sakamoto, S., Arai, T., Kato, M., Ishida, E., Furusho, H., et al. (2021). Elimination of Porphyromonas Gingivalis Inhibits Liver Fibrosis and Inflammation in NASH. *J. Clin. Periodontol.* 48 (10), 1367–1378. doi:10.1111/jcpe.13523

Nati, M., Haddad, D., Birkenfeld, A. L., Koch, C. A., Chavakis, T., and Chatzigeorgiou, A. (2016). The Role of Immune Cells in Metabolism-Related Liver Inflammation and Development of Non-alcoholic Steatohepatitis (NASH). *Rev. Endocr. Metab. Disord.* 17 (1), 29–39. doi:10.1007/s11154-016-9339-2

Psarras, A., Antanaviciute, A., Alase, A., Carr, I., Wittmann, M., Emery, P., et al. (2021). TNF-α Regulates Human Plasmacytoid Dendritic Cells by Suppressing IFN-α Production and Enhancing T Cell Activation. *J. Immunol.* 206 (4), 785–796. doi:10.4049/jimmunol.1901358

Rams, T. E., Sautter, J. D., and van Winkelhoff, A. J. (2020). Comparative *In Vitro* Resistance of Human Periodontal Bacterial Pathogens to Tinidazole and Four Other Antibiotics. *Antibiotics* 9 (2), 68. doi:10.3390/antibiotics9020068

Reis, M., Mavin, E., Nicholson, L., Green, K., Dickinson, A. M., and Wang, X.-n. (2018). Mesenchymal Stromal Cell-Derived Extracellular Vesicles Attenuate Dendritic Cell Maturation and Function. *Front. Immunol.* 9, 2538. doi:10.3389/fimmu.2018.02538

Rosato, V., Masarone, M., Dallio, M., Federico, A., Aglitti, A., and Persico, M. (2019). NAFLD and Extra-hepatic Comorbidities: Current Evidence on a Multi-Organ Metabolic Syndrome. *Ijerph* 16 (18), 3415. doi:10.3390/ijerph16183415

Rotundo, R., Nieri, M., Cairo, F., Franceschi, D., Mervelt, J., Bonaccini, D., et al. (2010). Lack of Adjunctive Benefit of Er:YAG Laser in Non-surgical Periodontal Treatment: a Randomized Split-Mouth Clinical Trial. *J. Clin. Periodontol.* 37 (6), 526–533. doi:10.1111/j.1600-051X.2010.01560.x

Russler-Germain, E. V., Yi, J., Young, S., Nutsch, K., Wong, H. S., Ai, T. L., et al. (2021). Gut Helicobacter Presentation by Multiple Dendritic Cell Subsets Enables Context-specific Regulatory T Cell Generation. *Elife* 10, e54792. doi:10.7554/eLife.54792

Schuster, S., Cabrera, D., Arrese, M., and Feldstein, A. E. (2018). Triggering and Resolution of Inflammation in NASH. *Nat. Rev. Gastroenterol. Hepatol.* 15 (6), 349–364. doi:10.1038/s41575-018-0009-6

Steinman, R. M. (2001). Dendritic Cells and the Control of Immunity: Enhancing the Efficiency of Antigen Presentation. *Mt. Sinai J. Med.* 68 (3), 160

Suominen, J., Ruokonen, A. L., Knuuttila, H., Puukka, M., and Jula, P. (2019). Periodontitis Is Associated with Incident Chronic Liver Disease-A Population-Based Cohort Study. *Liver Int.* 39 (3), 583–591. doi:10.1111/liv.13985

Tetè, G., D'orto, B., Ferrante, L., Polizzi, E., and Cattoni, F. (2021). Role of Mast Cells in Oral Inflammation. *J. Biol. Regul. Homeost. Agents* 35 (4 Suppl. 1), 65–70. doi:10.23812/21-4supp1-6

Tomofuji, T., Ekuni, D., Yamanaka, R., Kusano, H., Azuma, T., Sanbe, T., et al. (2007). Chronic Administration of Lipopolysaccharide and Proteases Induces Periodontal Inflammation and Hepatic Steatosis in Rats. *J. Periodontology* 78 (10), 1999–2006. doi:10.1902/jop.2007.070056

Wang, T., Ishikawa, T., Sasaki, M., and Chiba, T. (2022). Oral and Gut Microbial Dysbiosis and Non-alcoholic Fatty Liver Disease: The Central Role of Porphyromonas Gingivalis. *Front. Med.* 9, 822190. doi:10.3389/fmed.2022.822190

Wesolowski, S. R., Kasmi, K. C. E., Jonscher, K. R., and Friedman, J. E. (2017). Developmental Origins of NAFLD: a Womb with a Clue. *Nat. Rev. Gastroenterol. Hepatol.* 14 (2), 81–96. doi:10.1038/nrgastro.2016.160

Yamazaki, K., Kato, T., Tsuboi, Y., Miyauchi, E., Suda, W., Sato, K., et al. (2021). Oral Pathobiont-Induced Changes in Gut Microbiota Aggravate the Pathology of Nonalcoholic Fatty Liver Disease in Mice. *Front. Immunol.* 12, 766170. doi:10.3389/fimmu.2021.766170

Yang, Z.-J., Wang, B.-Y., Wang, T.-T., Wang, F.-F., Guo, Y.-X., Hua, R.-X., et al. (2021). Functions of Dendritic Cells and its Association with Intestinal Diseases. *Cells* 10 (3), 583. doi:10.3390/cells10030583

Yao, M., Zhang, C., Gao, C., Wang, Q., Dai, M., Yue, R., et al. (2021). Exploration of the Shared Gene Signatures and Molecular Mechanisms Between Systemic Lupus Erythematosus and Pulmonary Arterial Hypertension: Evidence from Transcriptome Data. *Front. Immunol.* 12, 658341. doi:10.3389/fimmu.2021.658341

Younossi, Z. M., Koenig, A. B., Abdelatif, D., Fazel, Y., Henry, L., and Wymer, M. (2016). Global Epidemiology of Nonalcoholic Fatty Liver Disease-Meta-Analytic Assessment of Prevalence, Incidence, and Outcomes. *Hepatology* 64 (1), 73–84. doi:10.1002/hep.28431

Younossi, Z. M., Stepanova, M., Nader, F., Loomba, R., Anstee, Q. M., Ratziu, V., et al. (2021). Obeticholic Acid Impact on Quality of Life in Patients with Nonalcoholic Steatohepatitis: REGENERATE 18-Month Interim Analysis. *Clin. Gastroenterology Hepatology* 15 (21), 00751–00755. S1542-3565. doi:10.1016/j.cgh.2021.07.020

Zhang, Z., Shuai, Y., Zhou, F., Yin, J., Hu, J., Guo, S., et al. (2020). PDLSCs Regulate Angiogenesis of Periodontal Ligaments via VEGF Transferred by Exosomes in Periodontitis. *Int. J. Med. Sci.* 17 (5), 558–567. doi:10.7150/ijms.40918

Zhu, Y., Ding, X., She, Z., Bai, X., Nie, Z., Wang, F., et al. (2020). Exploring Shared Pathogenesis of Alzheimer's Disease and Type 2 Diabetes Mellitus via Co-expression Networks Analysis. *Car* 17 (6), 566–575. doi:10.2174/1567205017666200810164932

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 Genetics

ORIGINAL RESEARCH
published: 14 July 2022
doi: 10.3389/fgene.2022.942153

Integrated Bioinformatics Analysis Identifies Robust Biomarkers and Its Correlation With Immune Microenvironment in Nonalcoholic Fatty Liver Disease

Feng Zhang†, Zhengwei Zhang†, Yapeng Li, Yi Sun, Xinliang Zhou, Xiaoning Chen and Shibo Sun**

Department of General Surgery, Second Affiliated Hospital of Harbin Medical University, Harbin, China

OPEN ACCESS

Edited by:
Tao Huang,
Shanghai Institute of Nutrition and Health (CAS), China

Reviewed by:
Feng Jiang,
Fudan University, China
Yi Li,
Nanjing Medical University, China

***Correspondence:**
Xiaoning Chen
wfqcxn@hrbmu.edu.cn
Shibo Sun
shibosun8@hrbmu.edu.cn

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

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *12 May 2022*
Accepted: *13 June 2022*
Published: *14 July 2022*

Citation:
Zhang F, Zhang Z, Li Y, Sun Y, Zhou X, Chen X and Sun S (2022) Integrated Bioinformatics Analysis Identifies Robust Biomarkers and Its Correlation With Immune Microenvironment in Nonalcoholic Fatty Liver Disease. Front. Genet. 13:942153. doi: 10.3389/fgene.2022.942153

Objective: Nonalcoholic fatty liver disease (NAFLD) is a serious threat to human health worldwide. In this study, the aim is to analyze diagnosis biomarkers in NAFLD and its relationship with the immune microenvironment based on bioinformatics analysis.

Methods: We downloaded microarray datasets (GSE48452 and GSE63067) from the Gene Expression Omnibus (GEO) database for screening differentially expressed genes (DEGs). The hub genes were screened by a series of machine learning analyses, such as support vector machine (SVM), least absolute shrinkage and selection operator (LASSO), and weighted gene co-expression network analysis (WGCNA). It is worth mentioning that we used the gene enrichment analysis to explore the driver pathways of NAFLD occurrence. Subsequently, the aforementioned genes were validated by external datasets (GSE66676). Moreover, the CIBERSORT algorithm was used to estimate the proportion of different types of immune cells. Finally, the Spearman analysis was used to verify the relationship between hub genes and immune cells.

Results: Hub genes (CAMK1D, CENPV, and TRHDE) were identified. In addition, we found that the pathogenesis of NAFLD is mainly related to nutrient metabolism and the immune system. In correlation analysis, CENPV expression had a strong negative correlation with resting memory CD4 T cells, and TRHDE expression had a strong positive correlation with naive B cells.

Conclusion: CAMK1D, CENPV, and TRHDE play regulatory roles in NAFLD. In particular, CENPV and TRHDE may regulate the immune microenvironment by mediating resting memory CD4 T cells and naive B cells, respectively, and thus influence disease progression.

Keywords: non-alcoholic fatty liver disease, integrated bioinformatics analysis, immune microenvironment, differentially expressed genes, robust biomarkers

INTRODUCTION

Nonalcoholic Fatty Liver Disease (NAFLD) is defined by the presence of hepatic steatosis in the absence of significant alcohol consumption or causes other than the metabolic disorders constituting the metabolic syndrome, which is a leading cause of chronic liver disease and affects about 10% of the world population (Sven et al., 2020). Hepatic pathologies of NAFLD range from simple hepatic steatosis to nonalcoholic steatohepatitis (NASH), even developing into liver fibrosis, liver cirrhosis, and hepatic carcinoma (Kabbany et al., 2017). With the increasing incidence of obesity, diabetes, hyperlipidemia, and cardiovascular disease, NAFLD has become increasingly prevalent, which represents the hepatic manifestation of metabolic syndrome. The global prevalence of NAFLD will be increasing soon. Despite the enormous burden on healthcare costs, there is no effective cure approved for NAFLD. Lifestyle interventions are recommended as first-line management in guidelines, but it is difficult to achieve favorable and persistent outcomes in the real world regrettably (Polyzos et al., 2019).

With the drastic development of generation sequencing technologies, systems biology techniques including genomics, metabolomics, transcriptomics, and proteomics provide new insight into solving this task. An increasing number of studies have indicated that NAFLD is linked to metabolic disorders (Huang and Kong, 2021; Luukkonen et al., 2021; Osborne et al., 2021). Immunity is also involved in the development and progression of NAFLD (Barrow et al., 2021; Huby and Gautier, 2021; Song et al., 2021). Nonetheless, there is still a lack of insensitive and targeted biomarkers that may be widely used in the clinical setting, which causes significant challenges for clinical diagnosis and treatment, especially for early diagnosis and follow-up strategy.

Therefore, an exploration into the molecular mechanism in NAFLD is necessary. To address these issues, we applied the GEO database to mine DEGs between NAFLD patients and normal patients, and determined the correlation between robust biomarkers, immune microenvironment and nutrient metabolism. Using various bioinformatics analysis methods, we described the differential genes and verified these genes in the external gene dataset, and finally screened CAMK1D, CENPV, and TRHDE. In addition, we found compounds or environmental poisons that might have a potential relationship with hub genes in the comparative toxicology database, which provided an important theoretical basis for the primary prevention and treatment of NAFLD.

MATERIALS AND METHODS

Datasets and Data Preprocessing

Raw transcriptomic data from two microarray datasets (GSE48452 and GSE63067) based on the GPL11532 and GPL570 platforms, both taken from the liver tissue, were downloaded from the GEO database. Normalization was performed on the raw data using the sva package. PCA showed that the aforementioned analysis method was better at eliminating batch effects (**Supplementary Figure.S1**). Twenty-one healthy liver control tissues, as well as 27 NAFLD liver tissue samples, were ultimately included in the screening set. In addition, we downloaded the GPL6244 platform–based GSE66676, derived from liver wedge biopsies, as an external validation dataset. Particular clinical characteristics of the patients in the dataset are presented in **Supplementary Datasheet S1**.

Screening and Validation of Hub Markers

As previous studies have done (Hu et al., 2022; Jiang et al., 2022), differentially expressed genes (DEGs) were screened in GSE48452 and GSE63067 in a batch-calibrated screening set. DEGs between NAFLD samples and normal samples were screened using the limma program package, with P. adj. value <0.05 selected as the cutoff criterion. Considering the situation of datasets, we did not set logFC as the threshold. Subsequently, the core genes were further screened in the aforementioned DEGs using a 10-fold cross-validation of LASSO (glmnet package). Alternatively, support vector machine–recursive feature elimination (SVM-REF) is a support vector machine–based machine learning method that builds on DEGs by removing support vector machine–generated feature vectors (e1071 and msvmRFE program packages) to find the optimal core genes. Simultaneously, we screened DEGs using one-way logistic regression with NAFLD as the dependent variable, using $p < 0.001$ as the threshold. In the WGCNA analysis, all DEGs satisfying p. value <0.05 in normal and NAFLD samples were used as input, and each sample clustered well, with a shear line of 30 as the threshold, and one outlier sample was excluded. Subsequently, a soft threshold from 1 to 20 was used for topology calculation to determine the optimal soft threshold of 6. The curve is smoothest when β was 6. Based on the soft threshold, the relationship matrix was converted to an adjacency matrix and then to a topological overlap matrix (TOM) for mean linkage hierarchical clustering, and the related modules were classified according to TOM with the number of genes in each module not less than 50. The gene module shear height in this study was 0.7, and similar module merging was performed. In addition, GS and GS of each module were calculated. In addition, GS and MM within each module were calculated for scatter plots. Finally, the Pearson method was used to calculate the correlation between the merged modules and the occurrence of NAFLD.

Enrichment Analysis

GO enrichment analysis is a common bioinformatics method used to search for comprehensive information on large-scale genetic data, including BP, CC, and MF. In addition, KEGG pathway enrichment analysis is widely used to understand biological mechanisms and functions. At the same time, DO enrichment analysis can be used to explore the diseases in which the genes of interest are predominantly involved. Finally, GO, KEGG pathway, and DO analyses were visualized using the GO plot program package. Finally, primary signaling pathways associated with core genes were further explored using the cluster profile package and the GSVA package. The h.all.v7.4.symbols.gmt gene set was downloaded from MSigDB,

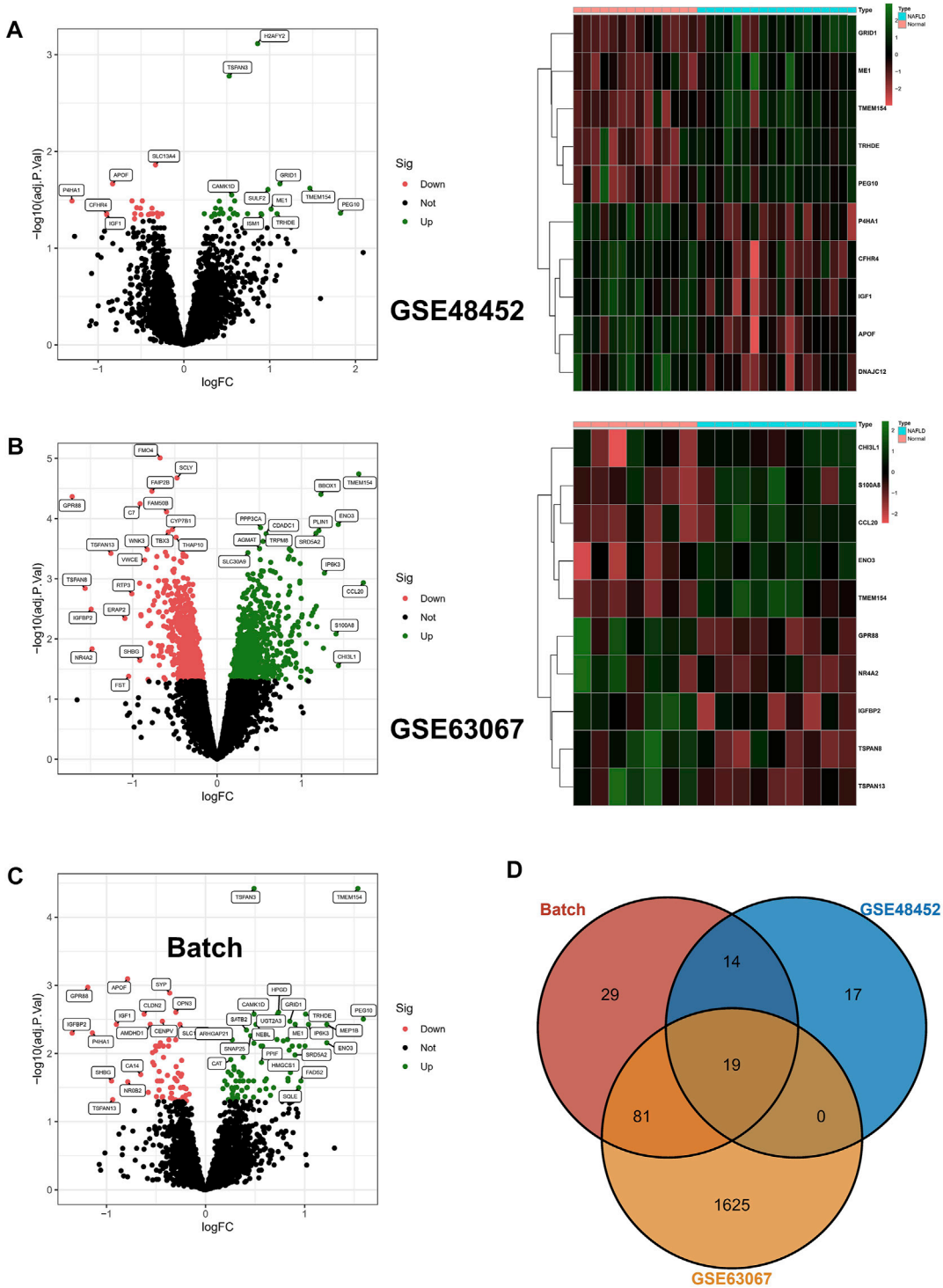


FIGURE 1 | Differentially expressed genes. **(A,B)** Differentially expressed genes (DEGs) were identified in GSE48452 and GSE63067, respectively, with upregulated genes indicated in red and downregulated genes in green in the volcano plot; in addition, the heat map shows the top 10 differentially expressed genes. **(C)** After batch correction, we again identified DEGs in the screening set using the limma package, and the volcano plot in red indicates upregulated genes and green indicates downregulated genes. **(D)** The DEGs from the three aforementioned screens were crossed, resulting in the identification of 19 core DEGs.

and the gene set was subjected to GSVA analysis with the gene expression matrix to explore the regulatory pathways that may be involved.

Construction of Hub Gene Regulatory Network

First, potential miRNAs targeting hub genes were predicted using mirDIP and Starbase databases, with the threshold set to minimum score = very high, to identify the regulatory network of miRs downstream of core genes. In addition, the TRRUST database contains 800 human transcription factors (TFs), and TF-core gene reciprocal pairs with p-values < 0.05 were selected to build upstream regulatory networks. In addition, we queried the Comparative Toxicogenomic database for compounds or environmental reads that might retain potential relationships with core genes. Finally, the core gene regulatory network was visualized based on the Networkanalyst database (**Supplementary Figures S2, S3**).

Immune Analysis Algorithm

As previous studies have done (Lu et al., 2021; Shen et al., 2021), the CIBERSORT algorithm calculates the proportion of different immune cell types based on the expression levels of immune cell–related genes. The output of the 22 infiltrated immune cells was integrated to generate a matrix of immune cell fractions for analysis (the CIBERSORT program package). The correlation of core genes with the content of the 22 immune cell types was calculated using the Spearman method.

Statistical Analysis

All statistical analyses were performed using R software (v.4.0.1). Detailed statistical methods for transcriptome data processing are covered in the aforementioned section. $p < 0.05$ was considered statistically significant.

RESULT

DEGs in Different Datasets

In the beginning, we identified 50 differentially expressed genes (DEGs) in GSE48452, and the volcano map shows 28 upregulated genes as well as 22 downregulated genes; in addition, the heat map shows the top 10 differentially expressed genes (**Figure 1A**). In addition, 1725 DEGs were identified in GSE63067, and the volcano map and heat map demonstrates 885 upregulated genes as well as 840 downregulated genes (**Figure 1B**). In addition, after batch correction, we again identified DEGs in the screening set using the limma package, with the volcano map demonstrating 77 upregulated genes as well as 66 downregulated genes (**Figure 1C**). Ultimately, we crossed DEGs from the three aforementioned screens and ultimately identified 19 core DEGs. In detail, TMEM154, TSPAN3, CAMK1D, TRHDE, PEG10, ME1, SATB2, SNAP25, ANKRD18A, ISM1, and SGCB were upregulated in NAFLD, while APOF, SYP, OPN3, CENPV, IGF1, AMDHD1, P4HA1, and MRPL21 were downregulated within NAFLD samples (**Figure 1D**).

Enrichment Analysis in DEGs

To explore the potential biological mechanisms of the 19 DEGs and the development of NAFLD, KEGG analysis illustrated the possible biological mechanisms of NAFLD development such as glioma, hypertrophic cardiomyopathy, and other disease processes (**Supplementary Figure S2A**). In addition, DO analysis revealed 19 differential genes that may have shared pathogenesis in diseases such as cell type benign neoplasm (**Supplementary Figure S2B**). Meanwhile, the BP section of the GO enrichment analysis suggested the important role of the dicarboxylic acid metabolic process, pyruvate metabolic process, etc. (**Supplementary Figure S2C**). Finally, we downloaded the corresponding gene sets from MSigDB and performed the GSVA analysis of the gene sets and gene expression matrices to explore the potential pathways involved in the pathogenesis of NAFLD, and the results showed that allograft rejection, cholesterol homeostasis, complement, and inflammatory response pathways have significant roles (**Supplementary Figure S2D**). Interestingly, taken together, the established chain of evidence suggests a possible involvement of the immune system with the nutritional metabolic system in NAFLD.

Integrated LASSO Analysis, Machine Learning Algorithm, and Logistic Analysis for Screening Hub Biomarkers

Among the aforementioned 19 DEGs, we further screened the core genes using a 10-fold cross-validation of LASSO and finally screened 12 potential genes (**Figures 2A,B**). At the same time, we performed an in-depth screening of the differential genes using a machine learning approach with SVM, and the results showed the lowest RMSE values when all 19 genes were included (**Figure 2C**). Finally, we performed a one-way logistic analysis of the expression of the 19 DEGs, with NAFLD occurrence as the dependent variable, and the final results of Moritu showed that 15 genes entered the subsequent analysis (**Figure 2D**).

WCGNA Analysis Was Used for Further Screening

To further link clinical information to key genes, the expression of only 1,989 genes that met p. value <0.05 in the analysis of differences between normal and NAFLD samples was used as the input matrix in WGCNA analysis. The samples clustered well, and one outlier sample was excluded using a shear line of 30 as the threshold (**Figure 3A**). Subsequently, a soft threshold of 1–20 was used for topological calculations, and the optimal soft threshold was determined to be 6 (**Figure 3B**). Based on the soft threshold, the relationship matrix was converted to an adjacency matrix and then to a topological overlap matrix (TOM) for mean linkage hierarchical clustering, and the related modules were classified according to TOM with no less than 50 genes in each module, and the similar gene modules were finally merged (**Figure 3C**), resulting in the identification of three modules. In addition, to calculate the correlation between genes within modules and clinical traits, we

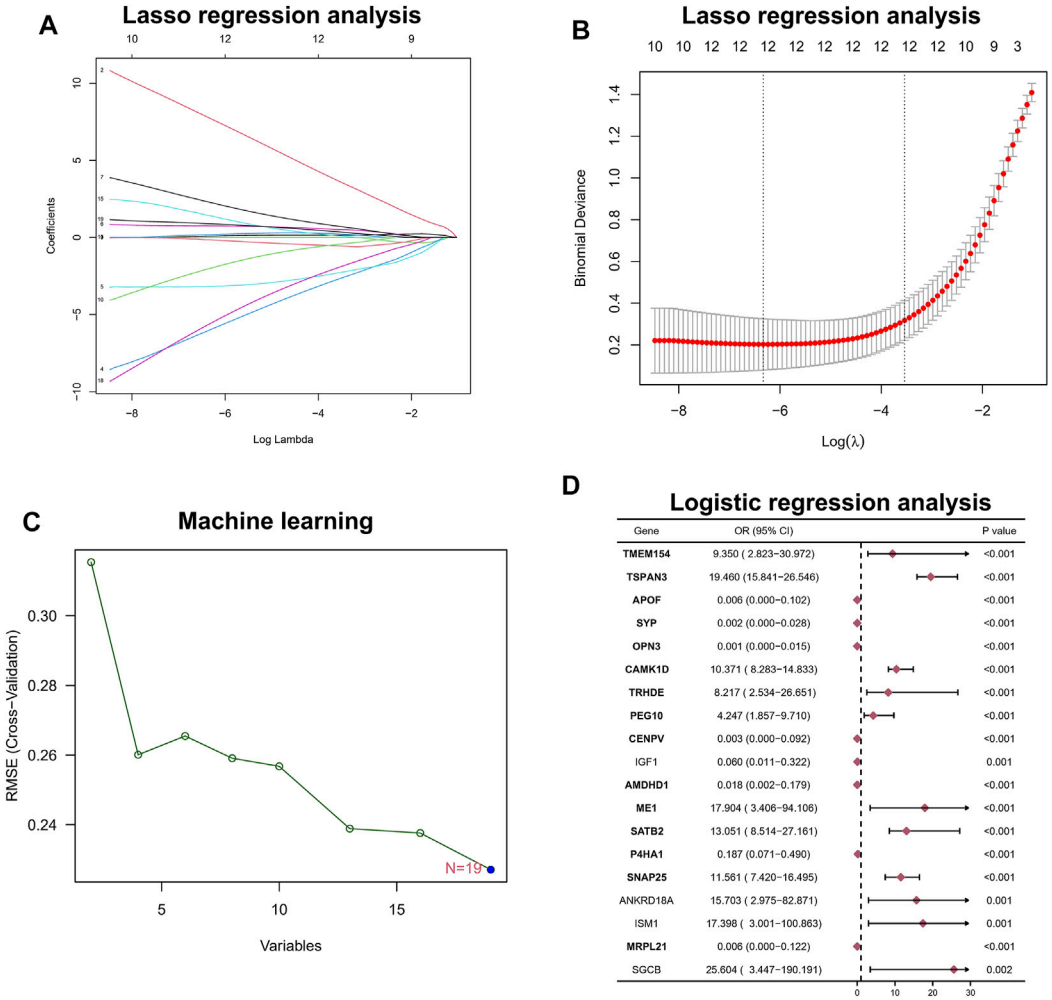


FIGURE 2 | Hub biomarker screening. **(A,B)** LASSO regression analysis. **(C)** Machine learning approach with SVM. **(D)** Logistic regression analysis.

found that the green module included the highest correlation with the occurrence of NAFLD ($p = 0.83$), so this was used as the core module (**Figure 3D**). In addition, GS and MM were calculated for 1,196 genes within the green module, and correlation scatter plots were drawn (**Figure 3E**). We found a direct correlation between GS and MM of genes within the core module, which verified our speculation from another perspective.

Exploring Predictive Value of Biomarkers

To identify core biomarkers, we cross-tabulated relevant genes from WGCNA, LASSO, Logistic, and machine learning, and finally identified seven biomarkers as our candidate genes (**Figure 4A**). We performed ROC analysis on each of these seven genes in the screening set, and the results showed that all genes had excellent predictive performance in the screening set: CAMK1D (AUC = 0.859, **Figure 4B**), CENPV (AUC = 0.864, **Figure 4C**), OPN3 (AUC = 0.891, **Figure 4D**), SATB2 (AUC = 0.840 (**Figure 4E**), SNAP25 (AUC = 0.868, **Figure 4F**), TRHED (AUC = 0.848, **Figure 4G**), and TSPAN3 (AUC = 0.926, **Figure 4H**).

The Validation of Hub Biomarkers

To validate the accuracy of seven aforementioned genes, we performed validation in an external validation set. In the dataset, also with liver tissue sequencing, only CAMK1D, CENPV, and TRHDE obtained differential expression between samples (**Figure 5A**), and in addition, as shown in **Figure 5B**, ROC analysis also demonstrated better predictive performance for three aforementioned biomarkers (CAMK1D, AUC = 0.632; CENPV, ADU = 0.651; TRHED, ACU = 0.676). In addition, we queried the Comparative Toxicogenomics database for compounds or environmental toxicants that may have potential relationships with core genes. Finally, the core gene regulatory network was visualized based on the Networkanalyst database (**Supplementary Figures S3, S4**).

The Analysis of Differences in Immune Microenvironment

Considering the important role of the immune pathway in NAFLD in the GSEA gene enrichment analysis, we used the CIBERSORT algorithm to analyze the immune cell content in various tissues.

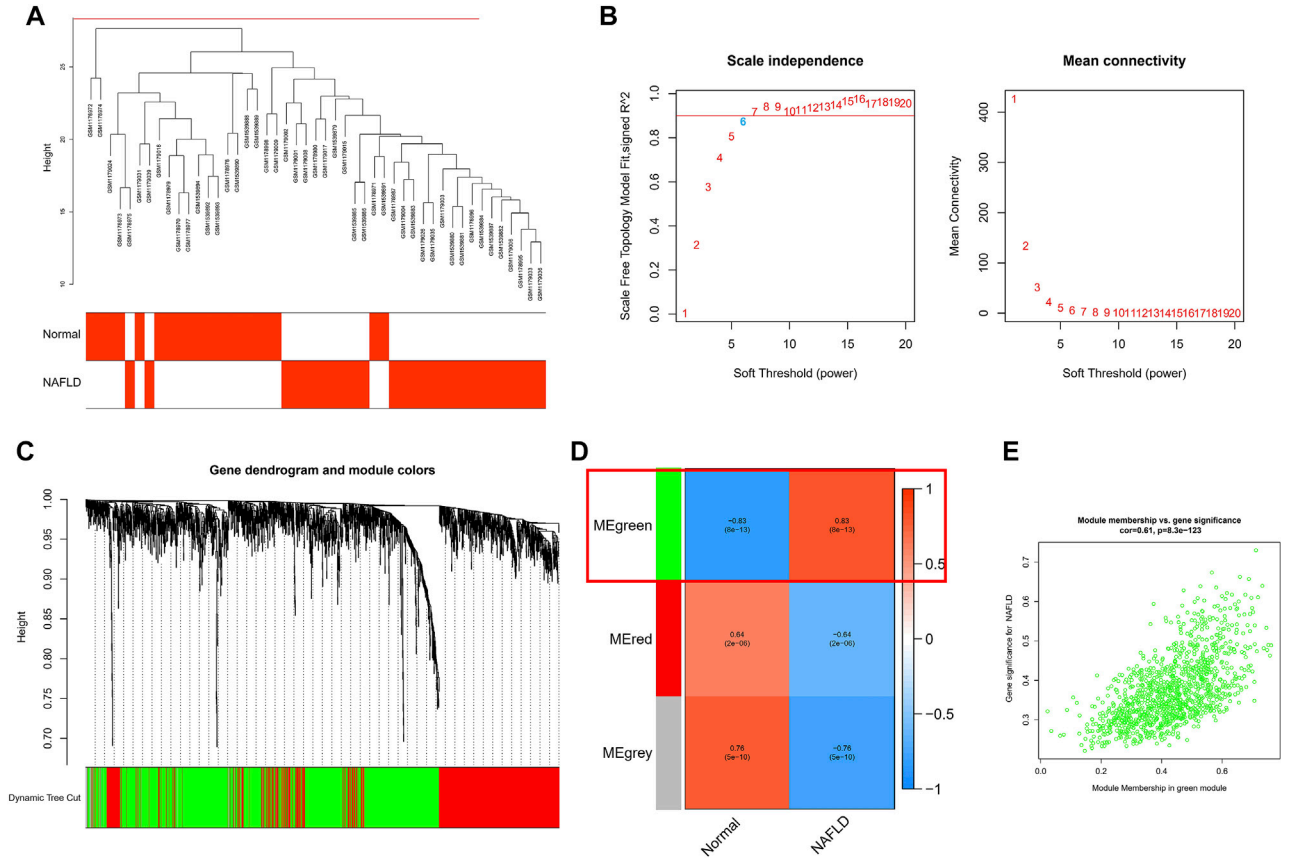


FIGURE 3 | WCGNA analysis was used for further screening. **(A)** Each sample clustered well, with a shear line of 30 as the threshold and one case of outlier removed. **(B)** Topological calculations with soft thresholds from 1 to 20 to determine the optimal soft threshold of 6. **(C)** Based on soft thresholds, the relationship matrix is converted to an adjacency matrix and then to a topological overlap matrix (TOM) for average link hierarchy clustering, which classifies the relevant modules according to TOM. **(D)** Calculating the correlation between genes within modules and clinical traits, we found that the green module had the highest correlation with the occurrence of NAFLD (p = 0.83). **(E)** The scatter plot of the correlation between GS and MM for 1,196 genes within the green module.

The results indicated higher levels of CD8 T cells, activated NK cells, and follicular-helper T cells in normal samples; in NAFLD tissues, only Macrophages M1 had a higher enrichment fraction compared to normal liver tissue (**Figure 6A**). In addition, the results of PCA analysis also showed a natural heterogeneity in the distribution of immune cells between the two tissues (**Figure 6B**). At the same time, the bar chart illustrates the general landscape of immune cell distribution between the different tissues (**Figure 6C**). Finally, as shown in **Figure 6D**, we performed a correlation analysis of all immune cells in the CIBERSORT algorithm, showing Macrophages M0 had the strongest negative correlation with T cells CD4 memory resting (r = −0.47) and T cells CD8 had the strongest positive correlation with Dendritic cells resting (r = 0.64).

Correlation Hub Biomarkers With Immune Infiltrating Cells

To explore the association of our identified core genes CAMK1D, CENPV, and TRHDE with immune cell content, we performed separate correlation analyses. Unfortunately, there was no statistically significant correlation between the CAMKD1 expression and the content of 22 types of aforementioned immune cells (**Figure 7A**). In addition, the CENPV expression had a significant negative correlation with resting memory CD4 T cells, r = −0.581 (**Figures 7B,D**). Simultaneously, TRHDE expression had a strong positive correlation with naive B cells, r = 0.538 (**Figures 7C,D**). Based on our results, we propose a speculation that the core genes CENPV and CRHDE may be involved in disease progression and regulate the immune microenvironment by mediating resting memory CD4 T cells and naive B cells, respectively.

DISCUSSION

NAFLD is a disease spectrum of a series of liver diseases, including simple fatty infiltration (steatosis) and fat and inflammation [nonalcoholic steatohepatitis (NASH)), and cirrhosis] without excessive alcohol consumption (<20 g a day for women and <30 g a day for men is adopted). NAFLD is associated with metabolic syndrome, including insulin resistance,

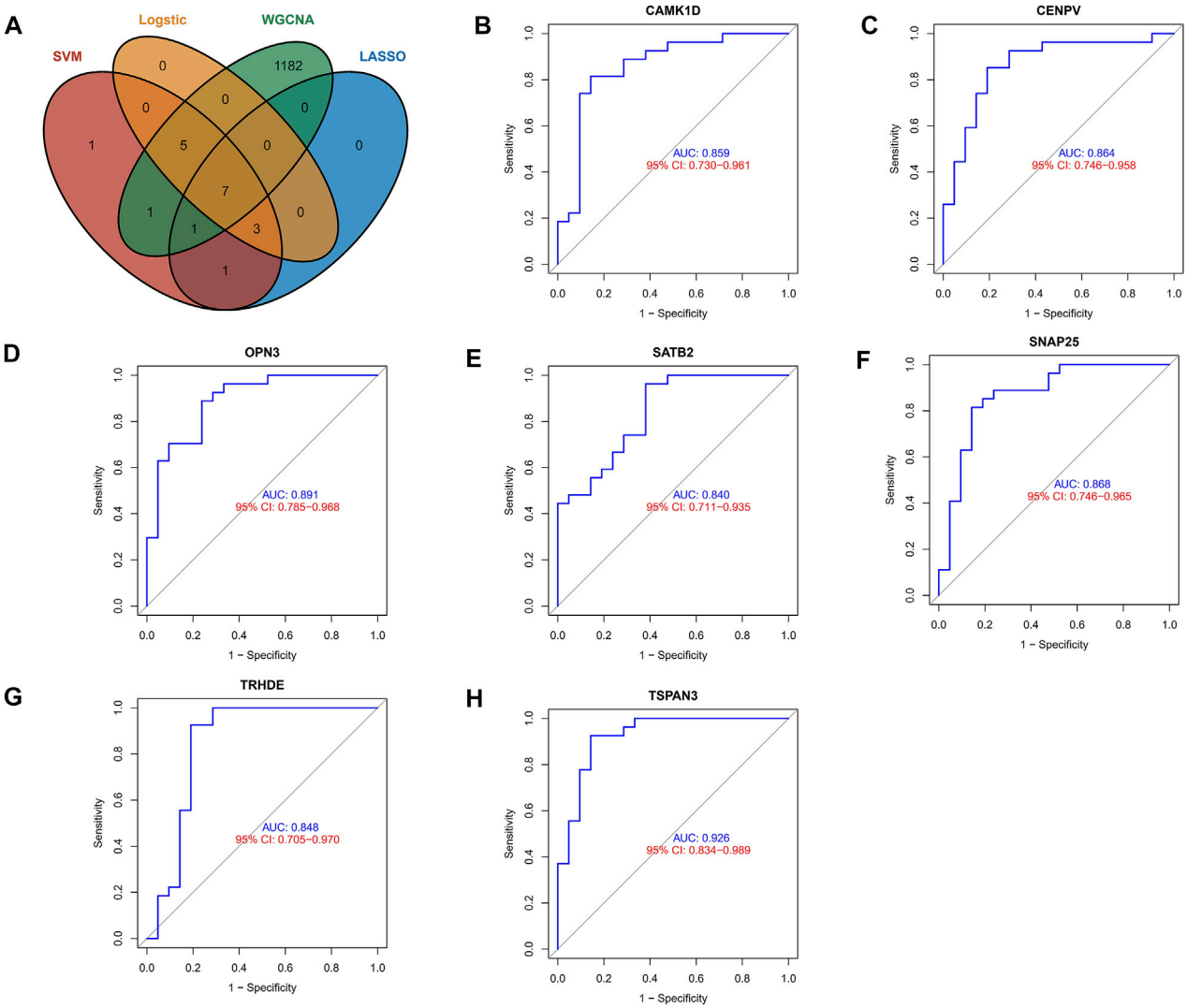


FIGURE 4 | Exploring the predictive value of biomarkers. **(A)** Cross-tabulation of relevant genes from WGCNA, LASSO, Logistic, and machine learning to identify candidate genes. **(B–H)** ROC analysis of the aforementioned genes.

hyperlipidemia, type 2 diabetes, and obesity. It is considered to represent the hepatic manifestation of this syndrome (de Alwis and Day, 2008; Anstee et al., 2011). An epidemiological model predicts that the prevalence of NAFLD/NASH will continue to increase and the mortality of associated diseases will double by 2030 (Estes et al., 2018). NAFLD is gradually becoming the fastest-growing cause of HCC; many risk factors for NAFLD are also independently associated with HCC, and screening for NAFLD-related HCC is difficult, so the exploration of the pathogenesis of NAFLD, related biomarkers, primary treatment, and prevention is urgent (Ioannou, 2021). In our study, we applied the GEO database to mine differential genes in NAFLD patients versus normal patients to identify strong biomarkers for correlation with the immune microenvironment. We characterized the differential genes in NAFLD patients versus normal patients by using various raw letter analysis methods and validated these differential genes in an external genetic dataset, finally screening for CAMK1D, CENPV, and TRHDE, and illustrating the relevance of these biomarkers to the immune microenvironment. In addition, we queried the Comparative Toxicogenomics database for compounds or environmental toxicants that may have potential relationships with core genes, providing a historic theoretical basis for the primary prevention and treatment of NAFLD.

Previous studies have shown that among multiple genetic risk factors, an SNP in the gene-encoding patatin-like phospholipase domain-containing 3 (PNPLA3) strongly predicts an increased risk of developing NAFLD. The G allele of the PNPLA3 rs738409 (148M) variant is associated with an increased risk of NAFLD development, and progression of NAFLD to NASH, liver fibrosis, and even cirrhosis (Romeo et al., 2008; Valenti et al., 2010). Epidemiological studies have shown that fatty acids (FAs) and

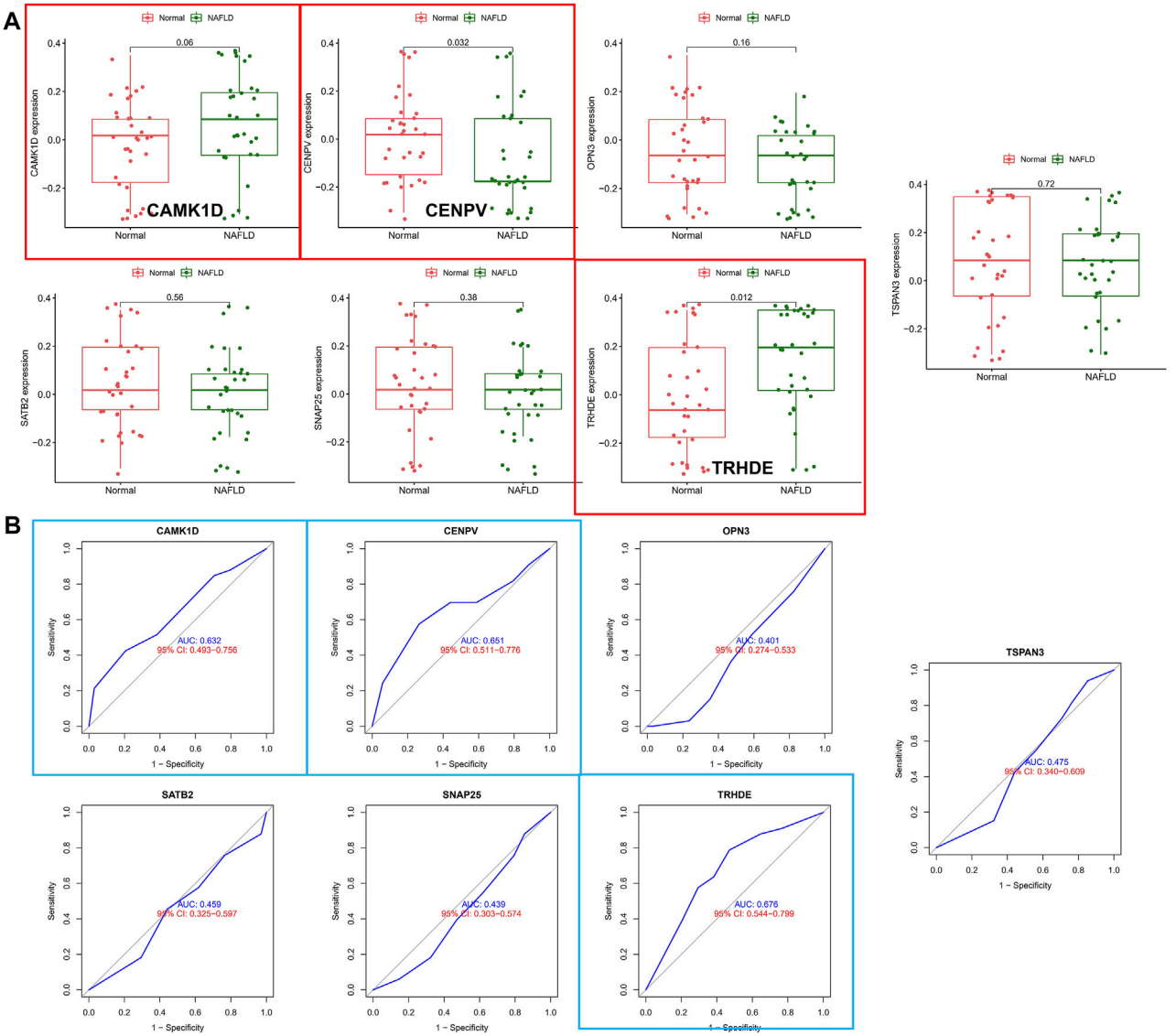


FIGURE 5 | Validation of hub biomarkers. **(A)** Candidate genes were validated in the external validation set, and those marked in red indicate differential expression between samples. **(B)** ROC analysis of candidate genes in the external dataset, with blue annotations indicating good predictive performance for the gene.

palmitoleic acid levels in NAFLD patients predict increased risk factors for CVD-related mortality and that the principal driver of CVD in NAFLD patients is a mutagenic lipid profile caused by increased hepatic lipogenesis. However, the specific pathogenesis of dilated cardiomyopathy and hypertrophic cardiomyopathy associated with the NAFLD process is unknown and needs to be further explored (Lai et al., 2019; Soehnlein and Libby, 2021). On the other hand, plasma amino acid concentrations have been associated with the pathogenesis of NAFLD and the progression of NASH, but the exact mechanisms are unclear. Concentrations of AA are altered in metabolic diseases such as T2DM, NAFLD, and obesity, and an established chain of evidence suggests that AA concentrations are associated with insulin resistance. BCAA has been of interest and may play a role in promoting peripheral and hepatic insulin resistance and in accelerating the T2DM process. In obese patients with NAFLD, fasting BCAA levels are elevated and associated with peripheral insulin resistance, possibly in relation to the liver being the site of protein and amino acid metabolism. In contrast, however, serine and glycine are found to be reduced in metabolic diseases such as NAFLD, suggesting that glycine metabolism is associated with the pathogenesis of NAFLD (Hyötyläinen et al., 2016; Gaggini et al., 2018). In our study, through the KEGG analysis of NAFLD, the DO analysis of the screened differential genes, GO analysis, and GSVA analysis, we found that the pathogenesis of NAFLD may involve the immune system and the nutritional metabolic system.

To further validate the correlation between DEGs screened from the database and NAFLD, LASSO regression analysis, machine learning, and logistic regression analysis were

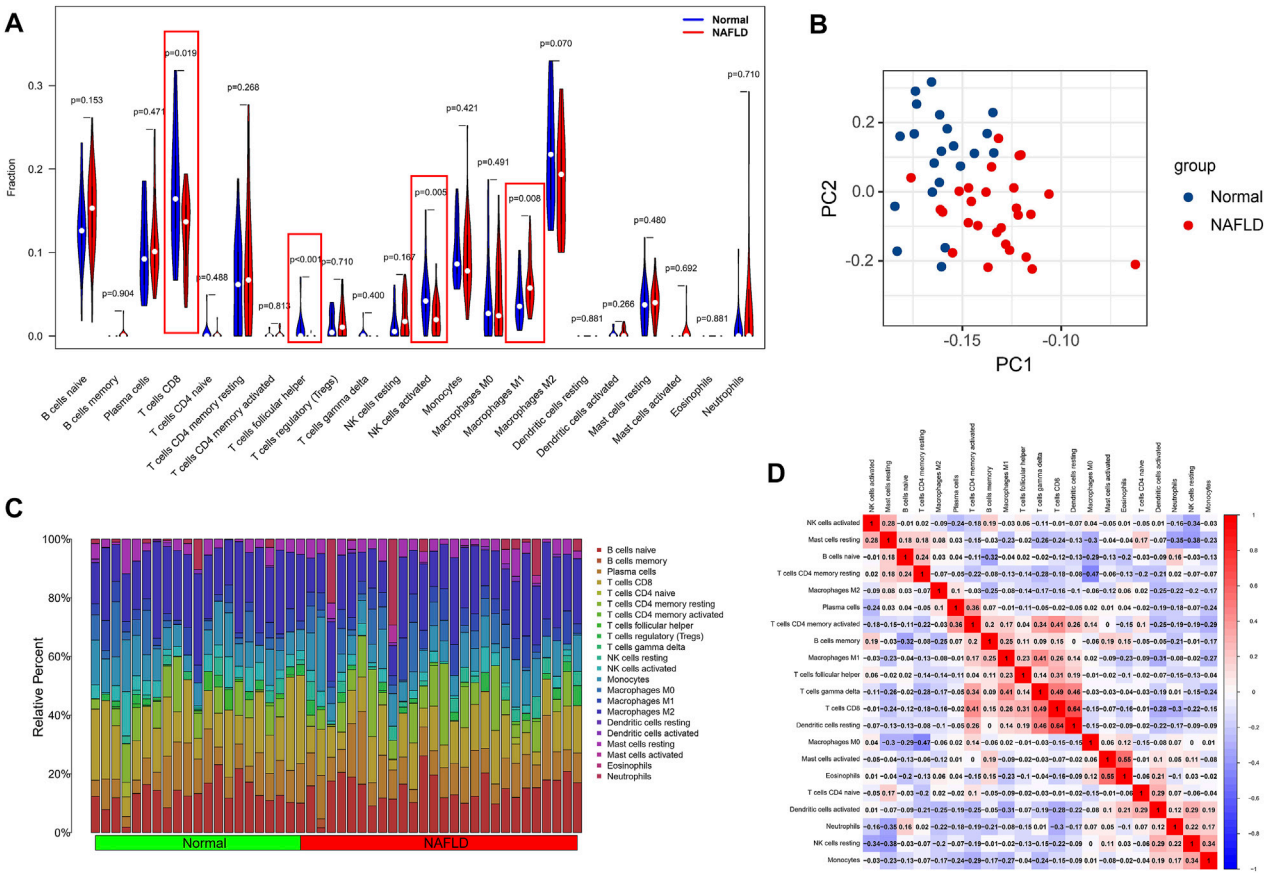


FIGURE 6 | Analysis of differences in the immune microenvironment. **(A)** The CIBERSORT algorithm was used to analyze the content of immune cells in different tissues, and the red markers are the immune cells with different content in the two different samples. **(B)** PCA analysis of the distribution of different immune cells in NAFLD and normal tissue samples. **(C)** A holistic view of the distribution of immune cells between different tissues. **(D)** Correlation analysis was performed on all immune cells in the CIBERSORT algorithm.

performed on 19 DEGs. Machine learning with SVM was chosen over other tools because of its ability to function with extraordinary accuracy and effective model. In addition, SVM machine learning has a nonlinear processing characteristic that allows it to produce a more accurate output than other algorithms, even when the data contain a large variability. However, in this study, when using the SVM's machine learning approach for in-depth screening of DEGs, RMSE only showed a minimum when all genes were included. The weighted gene coexpression network analysis (WGCNA) is a simple method that allows the construction of gene expression networks by aggregating highly related genes into modules, a method that allows visualization of the most representative AMI genes. These core elements of the biological network are more likely to represent essential genes with more critical functions (Langfelder et al., 2011). In order to identify the core biomarkers, we crossed the relevant genes screened by the previous methods and identified a total of seven candidate genes, and after ROC analysis, we found that all seven genes had strong predictive power. ROC analysis also demonstrated good predictive power for these three core biomarkers. We have predicted potential candidate compounds for these three hub genes, which are important for both primary prevention and subsequent targeted therapy in patients.

The CAMK1 family of calmodulin-dependent kinases is widely expressed in hepatocytes, endothelial cells, immune cells, and the essential nervous system (CNS) (Parkinson et al., 2007; Wayman et al., 2008). CAMK1D may play a role in hepatic gluconeogenesis (Rausch et al., 2018). Lina Xu et al., using integrated Hi-C, Nanopore, and RNA sequencing techniques to analyze liver tissues from normal and NAFLD mice, found thousands of regions in the genome with 3D chromatin organization and genomic rearrangements in the genome and revealed genetic dysregulation accompanying these variants. These genes were identified in NAFLD and were affected by genetic rearrangements and spatial organization disruption. Among them, CAMK1 expression was downregulated by alternating CNV and SV, chromatin loop, domains, and interaction matrix (Xu et al., 2021). In the type 2 diabetes CDC123/CAMK1D GWAS (genome-wide association studies) locus, rs11257655 affects transcriptional activity by altering the binding of the protein complexes of FOXA1 and FOXA2, a

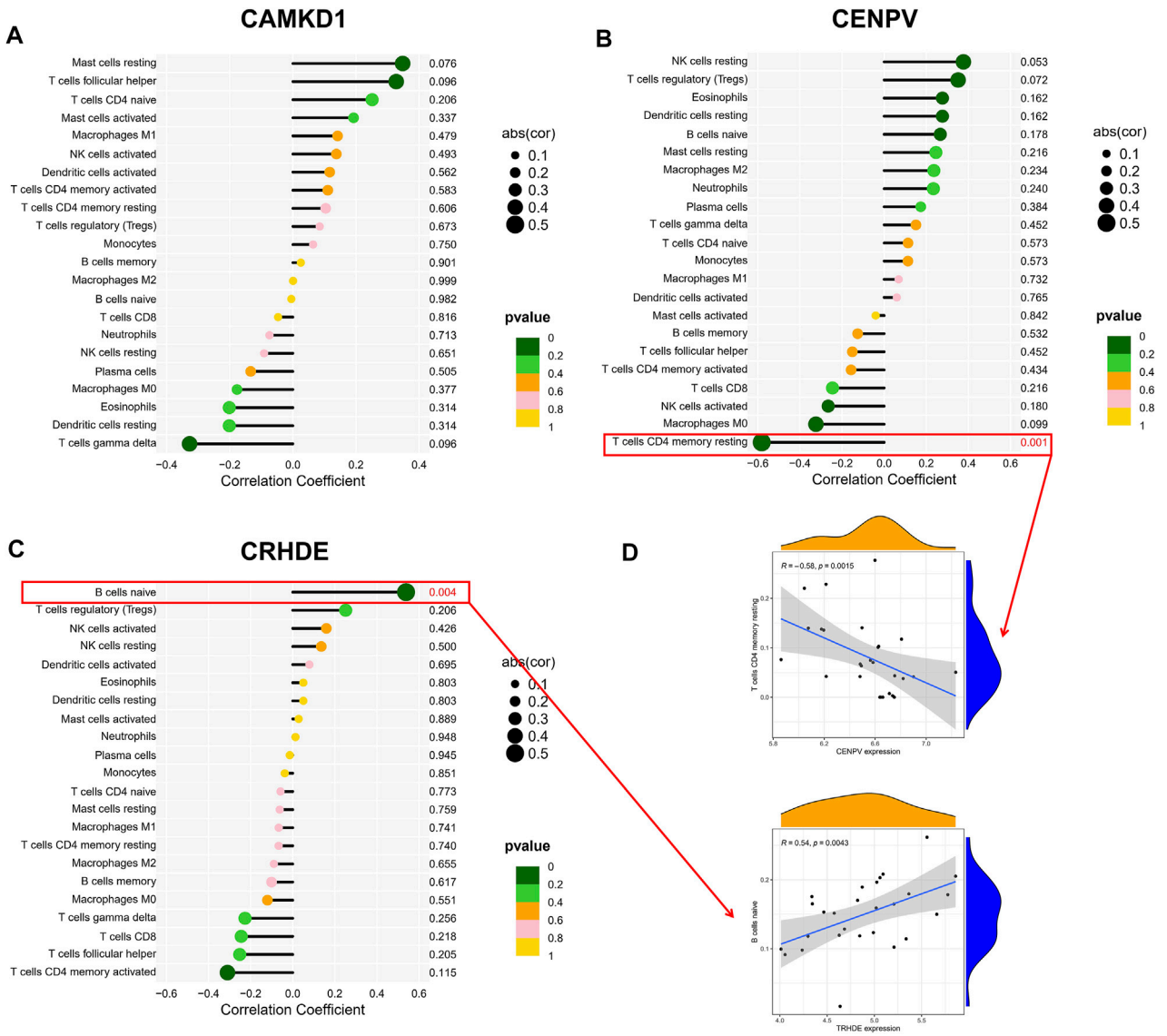


FIGURE 7 | Correlation hub biomarkers with immune-infiltrating cells. **(A–C)** Analysis of the correlation between the hub gene and immune cell content, with those marked in red indicating a statistically significant correlation between the gene and immune cells. **(D)** Once again, statistically significant correlations were made between hub genes and immune cell content.

potential molecular mechanism in type 2 diabetes (Fogarty et al., 2014). Christophe Fromont et al. first validated CAMK1D as a target for diabetes therapy in an *in vivo* experiment (Fromont et al., 2020). A single nucleotide polymorphism (SNP) genotyping of 11,530 cases pointed out that SNP rs10906115A of CDC123/CAMK1D was significantly associated with susceptibility to type 2 diabetes in the Japanese population (Imamura et al., 2011). However, the specific mechanism of regulation of NAFLD by CAMK1D is unclear. CENPV is a component of mitotic chromosomes associated with cytoplasmic microtubules. Elena Chiticariu et al. found that CENPV localizes to primary cilia in interphase, regulates cilia levels of acetylated microtubulin (α-tubulin), and is overexpressed in basal cell carcinomas and adnexal skin tumors (Chiticariu et al., 2020). CENPV levels are critical for cell viability, and either decreased or increased protein levels lead to cell death. cENPV provides an interesting link between the chromosomal passenger complex (CPC), primary contraction of mitotic chromosomes, and peristomal heterochromatin. The depletion of CENPV leads to a strong CPC phenotype (difficulties in chromosome bi-orientation and a failure to complete cytokinesis), followed rapidly by apoptotic cell death (Tadeu et al., 2008). The function of the CENPV gene is more organelle-specific and its role in the regulation of NAFLD has not yet been reported, and determining its detailed role and mechanism remains an exciting challenge for subsequent research. TRHDE was reported to be a DNA methylation marker for precancerous lesions in oral cancer (Shridhar et al., 2016). The overexpression of the noncoding long RNA TRHDE-

AS1 inhibits lung cancer progression via the miRNA-103/KLF4 axis (Zhuan et al., 2019). In a study of the genomic signature of gliomas, TRHDE was found to be positively correlated with the disease pathogenesis process (Liang et al., 2017). In a study of the genetic basis of thyrotropin receptor antibodies and hyperthyroidism in mice immunized with CXB recombinant inbred strains, it was uncovered that the TRHDE expression is controlled by thyroid hormones and is linked to genes related to thyroid function, which represents an extremely intriguing result (Aliesky et al., 2006). There are no studies on the role of TRHDE in NAFLD, but it has been reported in oral cancer, lung cancer, and glioma development, and its relationship with thyroid hormones could be explored in depth.

Inflammation is a hallmark of NAFLD that continues to progress to NASH and is characterized by a severe dysregulation of different innate and adaptive immune cell compartments, with immune cells regrouping in the liver and being activated (Parthasarathy et al., 2020). We have presently obtained two alternative views on the inflammatory response in NAFLD. While dysregulated immune cells can further exacerbate liver damage, the inflammatory response that occurs early in the process of liver injury may be substantial for tissue healing and repair (Wynn and Vannella, 2016; Eming et al., 2017). Our analysis of the immune cell content of NAFLD and normal tissues showed that normal tissues had higher levels of T cells CD8, activated NK cells, and follicular-helper T cells, while only Macrophages M1 had a relatively high enrichment fraction in NAFLD samples. A study in triple-transgenic model pigs suggests that CD8 T cells play a crucial role in adipose inflammation, recruiting and activating macrophages after activation in adipose tissue, which differs from our results for several reasons; we speculate, first, that CD8 T cells may not be consistently highly expressed throughout the development of the disease. Second, CD8 T cells may act more early in the development of NAFLD, and the exact cause and mechanism may need to be further explored (Zhang et al., 2021). NK cells perform a fundamentally meaningful role in liver fibrosis and are generally thought to reduce fibrotic events by eliminating activated hepatic stellate cells or altering the phenotype of hepatic macrophages. However, previous studies have found NK cell dysfunction in some patients with hepatocellular carcinoma and an association with a poor prognosis (Cai et al., 2008). In an analysis of the differences in NK cell surface markers and cell function correlations between NAFLD and ordinary volunteers, it was discovered that peripheral blood NK cells from NAFLD patients had reduced abundance and function (Sakamoto et al., 2021). The regulatory role of follicular-helper T cells (Tfh) is more in viral and autoimmune hepatitis. Xiaowen Wang et al. showed that in studies of blood from HBV-infected mice and patients with chronic HBV infection, the Tfh cell response to HBsAg was required for HBV clearance and that this response was blocked. The inhibition of Treg cell activity with anti-CTLA4 neutralizing antibodies restored the ability of Tfh cells to acquire HBV infection and could be implemented in the treatment of chronic HBV-infected patients. The dysregulation of the immune response to Tfh also induces lethal autoimmune hepatitis (Wang et al., 2018). The role of pro-inflammatory Macrophage M1 in NASH is primarily to exert immunomodulatory activity, with Macrophage M1 accumulating in areas of inflammation to secrete pro-inflammatory factors that exacerbate the progression of inflammation (Sun et al., 2021). The increase or decrease in the number of these immune cells can only suggest the occurrence of immune dysregulation in NAFLD tissues and the recruitment of some immune cells, which can be useful for subsequent studies and can be used clinically to slow down the progression of NAFLD or reverse the disease process to some extent by regulating the level of immune cells. The most significant negative correlation was demonstrated between Macrophages M0 and resting memory CD4 T cells, indicating that Macrophages M0 may be related to the activation of resting memory CD4 T cells, CD8 T cells, and dendritic cells. The most direct positive correlation between CD8 T cells and dendritic cells resting suggests that CD8 T cells inhibit dendritic cell recruitment through immune cell interactions during the NAFLD recruitment. However, this is only speculation and further experiments are needed to verify the exact relationship and mechanism of action.

Is there a relationship between the infiltration of immune cells and the screened hub biomarkers? To address this question, we analyzed the correlation between CAMK1D, CENPV, and TRHDE and immune cells, and finally found that CENPV expression had a direct negative correlation with resting memory CD4 T cells, and TRHDE expression had a strong positive correlation with naive B cells; we venture to guess that CENPV and TRHDE may regulate resting memory CD4 T cells and naive B cells through certain pathways, and have an impact on disease progression remission or recovery in NAFLD in terms of immune and inflammatory responses. Despite our findings, our conclusions need to be adopted with caution due to the limitations of our study. Our study is limited to the processing of previous data, and the timeliness and accuracy of our findings need to be verified, which may provide a reference for clinical diagnosis, but more detailed basic experiments and clinical trials are needed to support our findings before they can be applied to clinical treatment. Moreover, NAFLD samples are difficult to obtain, and it is difficult for us to conduct more assays.

In summary, we screened the GEO database for differential genes in two datasets, GSE48452 and GSE63067, and performed LASSO regression analysis, SVM machine learning analysis, logistic regression analysis, and WGCNA analysis on the differentially expressed genes. Seven candidate genes (CAMK1D, CENPV, OPN3, SATB2, SNAP25, TRHED, and TSPAN3) were finally screened, and three hub genes (CAMK1D, TRHDE, and CENPV) were identified after the external dataset validation. In GO analysis, we found that the disease process in NAFLD is strongly associated with nutritional metabolism and the immune system, and we identified more excessive levels of macrophage M1 in NAFLD than in normal tissue through immune cell content analysis. The ultimate analysis of hub genes and immune cell correlations suggests

that CENPV and TRHDE may influence the disease process in NAFLD by regulating resting memory CD4 T cells and naive B cells through certain pathways. This may additionally provide a theoretical basis for subsequent clinical treatment.

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

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

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.942153/full#supplementary-material

REFERENCES

Aliesky, H. A., Pichurin, P. N., Chen, C.-R., Williams, R. W., Rapoport, B., and McLachlan, S. M. (2006). Probing the Genetic Basis for Thyrotropin Receptor Antibodies and Hyperthyroidism in Immunized CXB Recombinant Inbred Mice. *Endocrinology* 147, 2789–2800. doi:10.1210/en.2006-0160

Anstee, Q. M., McPherson, S., and Day, C. P. (2011). How Big a Problem Is Non-alcoholic Fatty Liver Disease? *BMJ* 343, d3897. doi:10.1136/bmj.d3897

Barrow, F., Khan, S., Fredrickson, G., Wang, H., Dietsche, K., Parthiban, P., et al. (2021). "Microbiota-Driven Activation of Intrahepatic B Cells Aggravates Nonalcoholic Steatohepatitis through Innate and Adaptive Signaling," in *Hepatology* (Baltimore, Md).

Cai, L., Zhang, Z., Zhou, L., Wang, H., Fu, J., Zhang, S., et al. (2008). Functional Impairment in Circulating and Intrahepatic NK Cells and Relative Mechanism in Hepatocellular Carcinoma Patients. *Clin. Immunol.* 129, 428–437. doi:10.1016/j.clim.2008.08.012

Chiticariu, E., Regamey, A., Huber, M., and Hohl, D. (2020). CENPV Is a CYLD-Interacting Molecule Regulating Ciliary Acetylated α-Tubulin. *J. Invest. Dermatol.* 140, 66–74.e4. doi:10.1016/j.jid.2019.04.028

de Alwis, N. M. W., and Day, C. P. (2008). Non-alcoholic Fatty Liver Disease: the Mist Gradually Clears. *J. Hepatol.* 48, S104–S112. doi:10.1016/j.jhep.2008.01.009

Eming, S. A., Wynn, T. A., and Martin, P. (2017). Inflammation and Metabolism in Tissue Repair and Regeneration. *Science* 356, 1026–1030. doi:10.1126/science.aam7928

Estes, C., Anstee, Q. M., Arias-Loste, M. T., Bantel, H., Bellentani, S., Caballeria, J., et al. (2018). Modeling NAFLD Disease Burden in China, France, Germany, Italy, Japan, Spain, United Kingdom, and United States for the Period 2016-2030. *J. Hepatol.* 69, 896–904. doi:10.1016/j.jhep.2018.05.036

Fogarty, M. P., Cannon, M. E., Vadlamudi, S., Gaulton, K. J., and Mohlke, K. L. (2014). Identification of a Regulatory Variant that Binds FOXA1 and FOXA2 at the CDC123/CAMK1D Type 2 Diabetes GWAS Locus. *PLoS Genet.* 10, e1004633. doi:10.1371/journal.pgen.1004633

Fromont, C., Atzori, A., Kaur, D., Hashmi, L., Greco, G., Cabanillas, A., et al. (2020). Discovery of Highly Selective Inhibitors of Calmodulin-Dependent Kinases that Restore Insulin Sensitivity in the Diet-Induced Obesity *In Vivo* Mouse Model. *J. Med. Chem.* 63, 6784–6801. doi:10.1021/acs.jmedchem.9b01803,

Gaggini, M., Carli, F., Rosso, C., Buzzigoli, E., Marietti, M., Della Latta, V., et al. (2018). Altered Amino Acid Concentrations in NAFLD: Impact of Obesity and Insulin Resistance. *Hepatology* 67, 145–158. doi:10.1002/hep.29465

Hu, Y., Zeng, N., Ge, Y., Wang, D., Qin, X., Zhang, W., et al. (2022). Identification of the Shared Gene Signatures and Biological Mechanism in Type 2 Diabetes and Pancreatic Cancer. *Front. Endocrinol.* 13, 847760. doi:10.3389/fendo.2022.847760

Huang, W., and Kong, D. (2021). The Intestinal Microbiota as a Therapeutic Target in the Treatment of NAFLD and ALD. *Biomed. Pharmacother.* 135, 111235. doi:10.1016/j.biopha.2021.111235

Huby, T., and Gautier, E. (2021). Immune Cell-Mediated Features of Non-alcoholic Steatohepatitis. *Nat. Rev. Immunol.* 22 (7), 429–443. doi:10.1038/s41577-021-00639-3

Hyötyläinen, T., Jerby, L., Petäjä, E., Mattila, I., Jäntti, S., Auvinen, P., et al. (2016). Genome-Scale Study Reveals Reduced Metabolic Adaptability in Patients with Non-alcoholic Fatty Liver Disease. *Nat. Commun.* 7, 8994. doi:10.1038/ncomms9994

Imamura, M., Iwata, M., Maegawa, H., Watada, H., Hirose, H., Tanaka, Y., et al. (2011). Genetic Variants at CDC123/CAMK1D and SPRY2 Are Associated with Susceptibility to Type 2 Diabetes in the Japanese Population. *Diabetologia* 54, 3071–3077. doi:10.1007/s00125-011-2293-3

Ioannou, G. (2021). Epidemiology and Risk-Stratification of NAFLD-Associated HCC. *J. Hepatol.* 75, 1476–1484. doi:10.1016/j.jhep.2021.08.012

Jiang, F., Hu, Y., Liu, X., Wang, M., and Wu, C. (2022). Methylation Pattern Mediated by m6A Regulator and Tumor Microenvironment Invasion in Lung Adenocarcinoma. *Oxid. Med. Cell Longev.* 2022, 2930310. doi:10.1155/2022/2930310

Kabbany, M. N., Selvakumar, P. K. C., Watt, K., Lopez, R., Akras, Z., Zein, N., et al. (2017). Prevalence of Nonalcoholic Steatohepatitis-Associated Cirrhosis in the United States: An Analysis of National Health and Nutrition Examination Survey Data. *Am. J. Gastroenterol.* 112, 581–587. doi:10.1038/ajg.2017.5

Lai, H. T. M., de Oliveira Otto, M. C., Lee, Y., Wu, J. H. Y., Song, X., King, I. B., et al. (2019). Serial Plasma Phospholipid Fatty Acids in the De Novo Lipogenesis Pathway and Total Mortality, Cause-specific Mortality, and Cardiovascular Diseases in the Cardiovascular Health Study. *J. Am. Heart Assoc.* 8, e012881. doi:10.1161/JAHA.119.012881

Langfelder, P., Luo, R., Oldham, M. C., and Horvath, S. (2011). Is My Network Module Preserved and Reproducible? *PLoS Comput. Biol.* 7, e1001057. doi:10.1371/journal.pcbi.1001057

Liang, A., Zhou, B., and Sun, W. (2017). Integrated Genomic Characterization of Cancer Genes in Glioma. *Cancer Cell Int.* 17, 90. doi:10.1186/s12935-017-0458-y

Lu, L., Hu, Y., Wang, C., Jiang, F., and Wu, C. (2021). Methylation and Expression of the Exercise-Related TLR1 Gene Is Associated with Low Grade Glioma Prognosis and Outcome. *Front. Mol. Biosci.* 8, 747933. doi:10.3389/fmolb.2021.747933

Luukkonen, P., Qadri, S., Ahlholm, N., Porthan, K., Männistö, V., Sammalkorpi, H., et al. (2021). Distinct Contributions of Metabolic Dysfunction and Genetic Risk Factors in the Pathogenesis of Non-alcoholic Fatty Liver Disease. *J. Hepatol.* 76, 526–535. doi:10.1016/j.jhep.2021.10.013

Osborne, N., Leahy, C., Kwang Lee, Y., Rote, P., Song, B., and Hardwick, J. (2021). CYP4V2 Fatty Acid Omega Hydroxylase, a Druggable Target for the Treatment of Metabolic Associated Fatty Liver Disease (MAFLD). *Biochem. Pharmacol.* 195, 114841. doi:10.1016/j.bcp.2021.114841

Parkinson, H., Kapushesky, M., Shojatalab, M., Abeygunawardena, N., Coulson, R., Farne, A., et al. (2007). ArrayExpress--a Public Database of Microarray Experiments and Gene Expression Profiles. *Nucleic acids Res.* 35, D747–D750. doi:10.1093/nar/gkl995

Parthasarathy, G., Revelo, X., and Malhi, H. (2020). Pathogenesis of Nonalcoholic Steatohepatitis: An Overview. *Hepatol. Commun.* 4, 478–492. doi:10.1002/hep4.1479

Polyzos, S. A., Kountouras, J., and Mantzoros, C. S. (2019). Obesity and Nonalcoholic Fatty Liver Disease: From Pathophysiology to Therapeutics. *Metabolism* 92, 82–97. doi:10.1016/j.metabol.2018.11.014

Rausch, J. C., Lavine, J. E., Chalasani, N., Guo, X., Kwon, S., Schwimmer, J. B., et al. (2018). Genetic Variants Associated with Obesity and Insulin Resistance in

Hispanic Boys with Nonalcoholic Fatty Liver Disease. *J. Pediatr. Gastroenterol. Nutr.* 66, 789–796. doi:10.1097/mpg.0000000000001926

Romeo, S., Kozlitina, J., Xing, C., Pertsemlidis, A., Cox, D., Pennacchio, L. A., et al. (2008). Genetic Variation in PNPLA3 Confers Susceptibility to Nonalcoholic Fatty Liver Disease. *Nat. Genet.* 40, 1461–1465. doi:10.1038/ng.257

Sakamoto, Y., Yoshio, S., Doi, H., Mori, T., Matsuda, M., Kawai, H., et al. (2021). Increased Frequency of Dysfunctional Siglec-7–CD57+PD-1+ Natural Killer Cells in Patients with Non-alcoholic Fatty Liver Disease. *Front. Immunol.* 12, 603133. doi:10.3389/fimmu.2021.603133

Shen, X., Yang, Z., Feng, S., and Li, Y. (2021). Identification of Uterine Leiomyosarcoma-Associated Hub Genes and Immune Cell Infiltration Pattern Using Weighted Co-expression Network Analysis and CIBERSORT Algorithm. *World J. Surg. Onc.* 19 (1), 223. doi:10.1186/s12957-021-02333-z

Shridhar, K., Walia, G. K., Aggarwal, A., Gulati, S., Geetha, A. V., Prabhakaran, D., et al. (2016). DNA Methylation Markers for Oral Pre-cancer Progression: A Critical Review. *Oral Oncol.* 53, 1–9. doi:10.1016/j.oraloncology.2015.11.012

Soehnlein, O., and Libby, P. (2021). Targeting Inflammation in Atherosclerosis - from Experimental Insights to the Clinic. *Nat. Rev. Drug Discov.* 20, 589–610. doi:10.1038/s41573-021-00198-1

Song, Y., Zhang, J., Wang, H., Guo, D., Yuan, C., Liu, B., et al. (2021). A Novel Immune-Related Genes Signature after Bariatric Surgery Is Histologically Associated with Non-alcoholic Fatty Liver Disease. *Adipocyte* 10, 424–434. doi:10.1080/21623945.2021.1970341

Sun, G., Zhao, X., Li, M., Zhang, C., Jin, H., Li, C., et al. (2021). CD4 Derived Double Negative T Cells Prevent the Development and Progression of Nonalcoholic Steatohepatitis. *Nat. Commun.* 12, 650. doi:10.1038/s41467-021-20941-x

Sven, M. F., Pierre, B., Manal, F. A., Quentin, M. A., Elisabetta, B., Vlad, R., et al. (2020). A Randomised, Double-Blind, Placebo-Controlled, Multi-Centre, Dose-Range, Proof-Of-Concept, 24-week Treatment Study of Lanifibranor in Adult Subjects with Non-alcoholic Steatohepatitis: Design of the NATIVE Study. *Contemp. Clin. trials* 98, 106170. doi:10.1016/j.cct.2020.106170

Tadeu, A. M. B., Ribeiro, S., Johnston, J., Goldberg, I., Gerloff, D., and Earnshaw, W. C. (2008). CENP-V Is Required for Centromere Organization, Chromosome Alignment and Cytokinesis. *Embo J.* 27, 2510–2522. doi:10.1038/emboj.2008.175

Valenti, L., Al-Serri, A., Daly, A. K., Galmozzi, E., Rametta, R., Dongiovanni, P., et al. (2010). Homozygosity for the Patatin-like Phospholipase-3/adiponutrin I148M Polymorphism Influences Liver Fibrosis in Patients with Nonalcoholic Fatty Liver Disease. *Hepatology* 51, 1209–1217. doi:10.1002/hep.23622

Wang, X., Dong, Q., Li, Q., Li, Y., Zhao, D., Sun, J., et al. (2018). Dysregulated Response of Follicular Helper T Cells to Hepatitis B Surface Antigen Promotes HBV Persistence in Mice and Associates with Outcomes of Patients. *Gastroenterology* 154, 2222–2236. doi:10.1053/j.gastro.2018.03.021

Wayman, G. A., Lee, Y.-S., Tokumitsu, H., Silva, A., Soderling, T. R., and Soderling, T. (2008). Calmodulin-kinases: Modulators of Neuronal Development and Plasticity. *Neuron* 59, 914–931. doi:10.1016/j.neuron.2008.08.021

Wynn, T. A., and Vannella, K. M. (2016). Macrophages in Tissue Repair, Regeneration, and Fibrosis. *Immunity* 44, 450–462. doi:10.1016/j.immuni.2016.02.015

Xu, L., Yin, L., Qi, Y., Tan, X., Gao, M., and Peng, J. (2021). 3D Disorganization and Rearrangement of Genome Provide Insights into Pathogenesis of NAFLD by Integrated Hi-C, Nanopore, and RNA Sequencing. *Acta Pharm. Sin. B* 11, 3150–3164. doi:10.1016/j.apsb.2021.03.022

Zhang, K., Tao, C., Xu, J., Ruan, J., Xia, J., Zhu, W., et al. (2021). CD8 T Cells Involved in Metabolic Inflammation in Visceral Adipose Tissue and Liver of Transgenic Pigs. *Front. Immunol.* 12, 690069. doi:10.3389/fimmu.2021.690069

Zhuan, B., Lu, Y., Chen, Q., Zhao, X., Li, P., Yuan, Q., et al. (2019). Overexpression of the Long Noncoding RNA TRHDE-AS1 Inhibits the Progression of Lung Cancer via the miRNA-103/KLF4 axis. *J Cell. Biochem.* 120, 17616–17624. doi:10.1002/jcb.29029

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 Genetics

TYPE Original Research
PUBLISHED 08 September 2022
DOI 10.3389/fgene.2022.877833

OPEN ACCESS

EDITED BY
Tao Huang,
Chinese Academy of Sciences (CAS), China

REVIEWED BY
Wanze Chen,
Chinese Academy of Sciences (CAS), China
Bin Li,
Fifth Affiliated Hospital of Guangzhou Medical University, China

*CORRESPONDENCE
Kai Ye,
e5179518qihe9324@163.com
Pan Chi,
panchidoctor@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

RECEIVED 17 February 2022
ACCEPTED 26 July 2022
PUBLISHED 08 September 2022

CITATION
Ye K, Zhong W, Wang P, Chen Y and Chi P (2022), PD-1 blockage combined with vaccine therapy can facilitate immune infiltration in tumor microenvironment of Lynch syndrome colon cancer.
Front. Genet. 13:877833.
doi: 10.3389/fgene.2022.877833

PD-1 blockage combined with vaccine therapy can facilitate immune infiltration in tumor microenvironment of Lynch syndrome colon cancer

Kai Ye[1†]*, Wenjin Zhong[1†], Pengcheng Wang[1], Yanxin Chen[1] and Pan Chi[2]*

[1]Department of Gastrointestinal Surgery, Second Affiliated Hospital of Fujian Medical University, Quanzhou, Fujian Province, China, [2]Department of Colorectal Surgery, Union Hospital Affiliated of Fujian Medical University, Fuzhou, Fujian Province, China

Background: Lynch syndrome is a genetic disease resulting from mismatch repair gene mutation. Vaccine therapy can enhance the immunogenicity of Lynch syndrome and improve the therapeutic efficacy of immunotherapy. However, there is no approved Lynch syndrome vaccine coming onto the market.

Methods: Herein, we used gene knockdown method to construct Lynch syndrome cell model, paving way for us to develop Lynch syndrome tumor lysate vaccine. Then the isograft technique was employed for constructing the tumor-bearing mouse model of Lynch syndrome. And this isograft model was treated with PD-1 monoclonal antibody and tumor vaccine, respectively. Flow cytometry was used for detecting the proportion of immune cells and immunosuppressive cells, and ELISA was used for detecting the contents of chemokines and cytokines in the blood circulation system and tumor tissues of mice. Finally, IHC was used to detect the effects of tumor vaccines as well as PD-1 antibody on tumor tissue proliferation and angiogenesis.

Results: The results demonstrated that tumor vaccine could prolong the overall survival of mice, and improve the disease-free survival rate of mice. The vaccine could increase the proportion of inflammatory cells and decrease the proportion of anti-inflammatory cells in the blood circulation system of mice. In addition, tumor vaccine could also improve inflammatory infiltration in the tumor microenvironment and reduce the proportion of immunosuppressive cells. The results of IHC showed that tumor vaccine could inhibit angiogenesis and tumor cell proliferation in mouse tumor tissues.

Conclusion: In colon cancer associated with Lynch syndrome, tumor vaccine can hinder the growth of tumor cells, and assist immunotherapy whose therapeutic effect on this kind of cancer is thus enhanced.

KEYWORDS

Lynch syndrome, immune checkpoint-inhibitors, colon cancer, tumor vaccine, tumor microenvironment

Introduction

Immune checkpoint-inhibitors (ICIs) have been the most common clinical treatment for solid tumors, which are extensively applied in treating many cancers including melanoma, breast cancer and lung cancer due to the good therapeutic effect (Vaddepally et al., 2020). At present, many ICIs have got market approval, such as ipilimumab, nivolumab and pembrolizumab, which have shown great efficacy in clinical practice. For example, nivolumab combined with chemotherapy can prolong the median survival time of patients with esophageal cancer from 11.1 months (chemotherapy) to 13.1 months (nivolumab plus chemotherapy) (Janjigian et al., 2021). In a clinical study of patients with advanced triple-negative breast cancer, the researchers found that for patients with high tumor mutational burden (TMB) whose TMB over 10, pembrolizumab increased the objective response rate to 14.3% compared with chemotherapy (8.3%) (Winer et al., 2020). There are many ways by which ICIs work. Currently, ICIs drugs mainly target common immune checkpoints such as CTL4, PD-1 and PD-L1. By binding to the above targets, ICIs are able to mask immunosuppressive signals, thereby inhibiting the immune escape of tumor cells, alleviating the state of T cell depletion in the tumor microenvironment (TME) and improving the activity of the immune system (Jia et al., 2020). Much as ICIs have shown promising therapeutic effects in clinical practice, many patients do not respond well to the treatment with ICIs as a result of T cell depletion and poor immunogenicity of cancer cells (Blank and Mackensen, 2007). The way to improve the therapeutic effect of ICIs in such kind of patients is a pressing matter for scientists and clinicians.

Lynch syndrome is an autosomal dominant hereditary disease resulted from mutations in mismatch repair genes, and patients with Lynch syndrome have a higher risk of being diagnosed with colorectal and endometrial cancer. Lynch syndrome is the most common genetic syndrome giving rise to colorectal cancer, accounting for 3% of newly diagnosed colorectal cancer incidences (Sinicrope, 2018). Because patients with Lynch syndrome have inherited mutations in mismatch repair genes, they often exhibit microsatellite instability (MSI) and TMB-H after cancer onset, and interestingly, these two phenotypes are considered to have a fair response to immunotherapy (Rizzo et al., 2021). Current clinical practice has found favorable efficacy of immunotherapy in patients with Lynch syndrome. The results of ICIs treatment in patients with pan-cancer Lynch syndrome showed that the objective response rate is 94% (16/17), of which 94% (15/16) have persistent response without disease progression and relapse (Bari et al., 2020). Since cancer patients with Lynch syndrome have a good drug response to ICIs, the way to improve the therapeutic effect of ICIs on those patients is an urgent clinical problem to overcome, which is also the key to enhance the survival rate of patients with Lynch syndrome.

The tumor vaccine is a vaccine designed by using tumor tissue antigens, which contains tumor-specific antigens, aiming at eliciting an immune response against tumor antigens (Buonaguro and Tagliamonte, 2020). Recently, researchers have found that the therapeutic effect of ICIs can be enhanced once the patient's immune system is mobilized after the inoculation of tumor vaccine. For example, TAS0314 long-chain peptide vaccine has been found to have a synergistic anti-tumor immune effect with PD-1/PD-L1, which can enhance the therapeutic effect of PD-1/PD-L1 blockage by promoting the infiltration of cancer-specific cytotoxic T lymphocytes (CTLs) in tumor tissues (Tanaka et al., 2020). Since tumor cells from patients with Lynch syndrome are immunogenic and have a large number of mutation sequences, we speculated that tumor vaccines based on Lynch syndrome cells may help stimulate the patient's immune system and promote anti-tumor immunity to improve the sensitivity of ICIs. In this study, we first used gene knockout method to construct mouse MC38^{Mlh1KD} cells, which underwent homotransplantation to construct a mouse model of Lynch syndrome. Then mice were treated with tumor vaccine, and PD-1 blockage separately to explore the synergy of tumor vaccine on PD-1 blockage treatment. This study investigated the synergy of tumor vaccine on PD-1 blockage therapy for colon cancer in Lynch syndrome in the hope of providing more theoretical basis and reference for treating Lynch syndrome clinically.

Materials and methods

Cell culture and vaccine preparation

The murine colon carcinoma cell line MC38 (BNCC337716), purchased from BeNa Culture Collection (BNCC), was cultured in DMEM-H medium containing 10% FBS along with 1% P/S. The cells were cultured in a constant temperature incubator at 37°C with 5% CO_2. The lentiviral vector encoding Mlh1 shRNA was synthesized by GenePhama (China) and transfected into MC38 cells to construct dMMR CRC model cells with stable knockdown of Mlh1 (MC38$^{M1h1\ KD}$).

The tumor vaccine was prepared based on well-immunoreactive dMMR-type mouse tumor cells (MC38$^{Mlh1\ KD}$). MC38$^{Mlh1\ KD}$ cells were collected, and repeatedly frozen and thawed between −80°C and 37°C (5 minutes each, 4 cycles), followed by one heat shock

treatment (42°C, 5 min). Lysates from collected tumor cells were stored at −80°C.

In Vivo animal experiment

To detect the specific effect of tumor vaccine treatment on the survival time and tumor tissues of mice, 30 C57/BL6 mice aged 6–8 weeks acquired from Nanjing Institute of Model Animals were equally divided into two groups. The control group were injected with 2×10^6 (Sinicrope, 2018) $MC38^{Mlh1\ wildtype}$ cells in the right lower limb of mice, while the experimental group were injected with 2×10^6 (Sinicrope, 2018) $MC38^{Mlh1KD}$ cells. These mice were all treated with vaccine (injected with tumor cell lysate, 10 mg/kg bw, s.c., biweekly). The survival time experiment was enrolled 20 C57/BL6 mice with 10 mice in each group. Treatment continued until 70 days or until the mice died. The tumor tissues detection experiment was enrolled 10 C57/BL6 mice with 5 mice in each group. Mice were euthanized after 35 days of treatment, and then measured the tumor tissues volume and weight.

To detect the effect of tumor vaccine and PD-1 treatment on the survival time of mice, 40 C57/BL6 mice aged 6–8 weeks acquired from Nanjing Institute of Model Animals were injected with 2×10^6 (Sinicrope, 2018) $MC38^{Mlh1KD}$ cells in the right lower limb of mice to construct an isograft dMMR colon cancer model. The mice were divided into control group (treated with IgG) (Rockland Immunochemicals Inc., Boyertown, PA, USA), vaccine treatment group (injected with tumor cell lysate, 10 mg/kg bw, s.c., biweekly), PD-1 treatment group (injected with murine PD-1 monoclonal antibody 2.5 mg/kg bw, i.v., biweekly) or double treatment group (simultaneously injected with 10 mg/kg bw tumor cell lysate and 2.5 mg/kg PD-1 monoclonal antibody) according to different treatment methods, with 10 mice in each group. Treatment continued until 70 days or until the mice died.

To detect the changes of TME in mouse tumor tissues, 20 C57/BL6 mice aged 6–8 weeks obtained from Nanjing Institute of Model Animals were divided into groups according to the above treatments, 5 mice in each group. Mice were euthanized after 35 days of treatment, and whole blood of mice was obtained by eyeball blood sampling method. The tumor tissues were used for subsequent detection.

Tumor length and width were measured to calculate tumor volume by using the mathematical formula: volume (mm^3) = ($width^2$ (Janjigian et al., 2021) × length)/2.

Flow cytometry

Blood samples were obtained from previously collected whole blood. Blood samples were collected 150 μl/time from each mouse once every 2 weeks by eyeball blood sampling method. For flow cytometry detection of tumor tissues, tumor tissues were first collected from mice, mashed and passed through a 100-μm cell strainer and prepared into single cells suspensions. A panel of conjugated antibodies (Abcam, UK) was subsequently used for staining: FITC Anti-CD3 antibody, APC Anti-CD8 alpha antibody, PE Anti-CD4 antibody, PE Anti-NKR-P1C (NK1.1), FITC Anti-CD11b antibody, APC Anti-Ly6g antibody (gr1), FITC Anti-CD19 antibody, FITC Anti-CD83 antibody, PE/Cy7® Anti-PD1 antibody, PerCP/Cy5.5® Anti-PD-L1 antibody, PE Anti-LAG-3 antibody, PE Anti-CTLA4 antibody. Negative controls were stained by lymphocytes of the appropriate isotype. Flow cytometry was performed using BD FACSVerse™ and analysis of flow cytometric data was performed using BD FACSuite software.

ELISA test

Under the instructions of manufacturer, ELISA kits (MultiSciencesBiotech, China): Mouse TNF-a ELISA Kit, Mouse CCL4/MIP-1β ELISA Kit, Mouse IL-10 ELISA Kit, Mouse IL-13 ELISA Kit, Mouse CCL11/Eotaxin ELISA Kit, Mouse CCL5/RANTES ELISA Kit were employed for measuring expression of immune factors in serum.

Immunohistochemistry test and hematoxylin-eosin staining

Fresh tumor tissues were taken and fixed by immersion in formalin solution, and the formalin-fixed tissues were subsequently sectioned using paraffin embedding. Hematoxylin-eosin (H&E) staining was used for histopathological observation. For IHC staining, antigen recovery was first performed in EDTA buffer at pH = 9.0, followed by blocking of endogenous peroxidase activity as well as non-specific binding using ltraSen-sitive S-P kits (Maixin Biotechnology, China). Sections were incubated with rabbit anti-human anti-PD-L1 (Abcam, United Kingdom), rabbit anti-human anti-Ki-67 (Abcam, United Kingdom), and rabbit anti-human anti-CD31 monoclonal antibody at 4°C. Secondary antibody-horseradish peroxidase co-incubation was subsequently used (Maixin Biotechnology, China), and 3,3-diaminobenzidine-tetrahydrochloric acid was used as chromogen for coloration (Maixin Biotechnology, China).

The scoring criteria for staining intensity were as follows: staining (0), weak staining (1), moderate staining (2) and strong staining (3). The scoring criteria for the positive proportion of stained tumor cells were as follows: 0 (0%), 1 (1%–10%), 2 (11%–50%), 3 (51%–80%), and 4 (>80%). The staining results were evaluated semiquantitatively by multiplying the staining intensity by the percentage of positive tumor cells. Then the sections were scored by two experienced pathologists, and the average of the scores was selected as the final result. For evaluation of microtubule density, the number of microvessels per unit area was calculated using 3 random fields after staining with CD31, and the mean value was seen as the final result.

Quantitative real-time polymerase chain reaction

Trizol reagent (Takara, Japan) was employed for the extraction as well as purification of total DNA. qRT-PCR was performed in triplicate in the ABI 7500 fast real-time PCR System (Applied Biosystems, United States). The relative expression level of Mlh1 was calculated through normalization to GAPDH internal controls. The following primers were used for PCR detection:

Mlh1: forward: 5′-CTCCAAGATGAGGCTGTAGGAA-3′; reverse: 5′-CCTATGAGATGGAAGGCAAGA-3′; GAPDH forward: 5′-CTGGGCACTGAGCACC-3′; reverse: 5′-AAGTGG TCGTTGAGGGCAATG-3′.

Western blot

The extraction of total protein and the measurement of protein concentration were performed respectively by using radioimmunoprecipitation assay lysis buffer (Beyotime, China) and the BCA protein assay kit (Thermo Fisher Scientific, United States), respectively. Protein samples were separated by 10% SDS-PAGE and then transferred onto Polyvinylidene fluoride membrane (Millipore, United States). Afterwards, the membrane was blocked in 5% skim milk at 37°C for 1 h and then incubated with the primary antibodies at 4°C overnight. After incubating with Horseradish Peroxidase-conjugated secondary antibody for 1 h at 37°C, the membrane was washed in TBST and prepared for signal detection. The bands were visualized using an ECL chemiluminescent detection system (Thermo Fisher Scientific, United States). The primary antibodies and second antibody were purchasing from Abcam (United Kingdom), as anti-Mlh1 antibody (ab92312), anti-GAPDH antibody (ab9485), goat anti-rabbit IgG H&L antibody (ab6721).

Data analysis

In this paper, Graphpad Prism 8 was used to plot the data and data analysis, and one-way ANOVA was used for significance test in advance, and Student's *t*-test was used for intergroup analysis. $p < 0.05$ was considered statistically significant.

Results

Tumor vaccines combined with ICIs can remarkably prolong the survival time of Lynch syndrome model mice

We first constructed dMMR CRC model cells with stable knockdown of Mlh1 ($MC38^{Mlh1\ KD}$), and measured the transfection efficiency through qRT-PCR and western blot. The results exhibited the mRNA and protein expression levels of Mlh1 were significantly decreased in sh-Mlh1 group compared to sh-NC group (Supplementary Figures S1A, B). And then, we injected with 2×10^6 $MC38^{Mlh1\ KD}$ cells in the right lower limb of mice to construct an isograft dMMR colon cancer model. In order to verify whether the vaccine effect was specific to the Lynch syndrome model ($MC38^{Mlh\ KD}$), we treated with vaccine to control group and experimental group. The results exhibited tumor vaccine significantly improve the disease-free survival rate of $MC38^{Mlh\ KD}$ group mice (Supplementary Figures S2A–C), the tumor volume and weight in $MC38^{Mlh\ KD}$ group were significantly decreased compared to $MC38^{Mlh}$ wildtype group (Supplementary Figures S2D–F). These results clarified the vaccine effect was specific to the Lynch syndrome model. Subsequently, we explored whether vaccines made by tumor supernatant could increase the anti-tumor activity of PD-1 antibody in Lynch syndrome. Mice were treated with IgG, vaccine, Anti-PD-1 antibodies or vaccine plus Anti-PD-1 antibody on the third day after tumor inoculation in mice vaccinated with $MC38^{Mlh1\ KD}$ cells. (Figure 1A). The survival time of mice after treatment demonstrated that with a contrast to anti-PD-1 antibody treatment alone, anti-PD-1 antibody treatment combined with vaccine treatment of mice could improve the overall survival time of tumor mice (Figure 1B). The result also revealed that the combined therapy achieved 70-days survival in more mice than PD-1 antibody or vaccine therapy alone did (Figure 1C).

Tumor vaccines combined with ICIs can slow the growth rate of tumors in homografted mice

The above results illustrated that tumor vaccines combined with ICIs could elevate the overall survival time of mice, and we redesigned to treat mice with IgG, vaccine, Anti-PD-1 antibodies or vaccine plus Anti-PD-1 antibody, respectively. Mice were euthanized on day 35, tumor tissues were excised and tumor sizes were measured (Figure 2A). The results demonstrated that the size and weight of tumor tissues in mice were considerably reduced after treatment with vaccine or PD-1 antibody, and the size and weight of tumor tissues could be minimized when treating combined therapy (Figures 2B,C).

Tumor vaccines can alter the circulatory immune environment in mice

Here, we then analyzed the changes in the immune system of mice after treatment. Mouse blood was collected while euthanizing mice, half of which was used for the detection of chemokine content and the rest was used for analysing the content of specific immune cells in the blood using flow cytometry. The results showed that NK cell content was markedly increased, $CD19^+$ B cell and $CD8^+$ cell

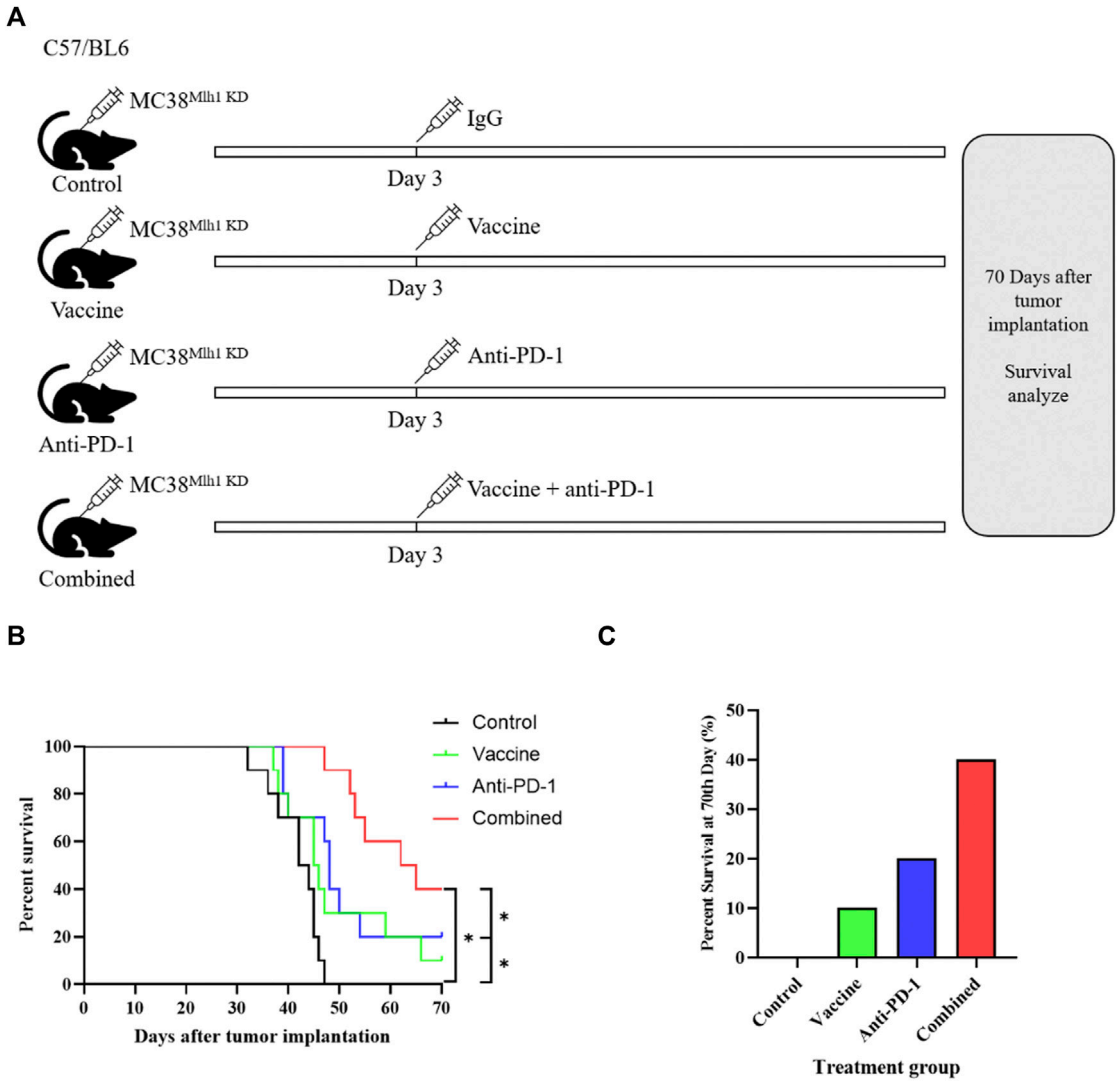


FIGURE 1
Combination therapy with vaccine and PD-1 blockage improves clinical outcomes in the Lynch syndrome colon cancer mouse model. **(A)** Schema of tumor implantation and treatment (C57/BL6 mice were injected with $MC38^{Mlh1\ KD}$ cells in the right lower limb and divided into four groups with ten mice in each group). **(B)** Kaplan-Meier survival curves of mice that were implanted with $MC38^{Mlh1\ KD}$ colon cells and were treated with different combinations of IgG, Vaccine and PD-1 antibody. **(C)** The percentages of mice that remained survival at 70th day following tumor implantation and therapy. Statistical significance was measured using Student's t-test and one-way ANOVA. *, $p < 0.05$.

contents were evidently decreased after treatment with the vaccine. After PD-1 antibody treatment, the content of $CD4^+$ T cells decreased, and the contents of $CD8^+$ T, NK, $CD11b^+/Gr1^+$ $CD19^+$ as well as $CD83^+$ B cells were significantly increased (Figure 3A). The results of combined treatment showed that blood $CD4^+$ T cells, NK cells, and $CD19^+$ B cells were significantly increased and $CD11b^+/Gr1$ cell content was remarkably decreased in mice compared with PD-1 antibody therapy alone (Figure 3A). The results showed that the combination therapy could promote the inflammatory response and improve the effect of PD-1 antibody therapy.

We also detected the contents of chemokines in the serum of mice, and the results suggested that the contents of TNF-α, MIP1β, and IL-13 in the serum were notably increased, and the content of Eotaxin was significantly decreased after treatment with the vaccine. After treatment with PD-1 monoclonal antibody, the levels of TNF-α, MIP1β and RANTES in serum were significantly increased. The content of MIP1β was remarkably increased while the contents of IL-10 and IL-13 were remarkably decreased in the combined treatment group compared with PD-1 monoclonal antibody treatment group (Figure 3B). To sum up, we found that the combination

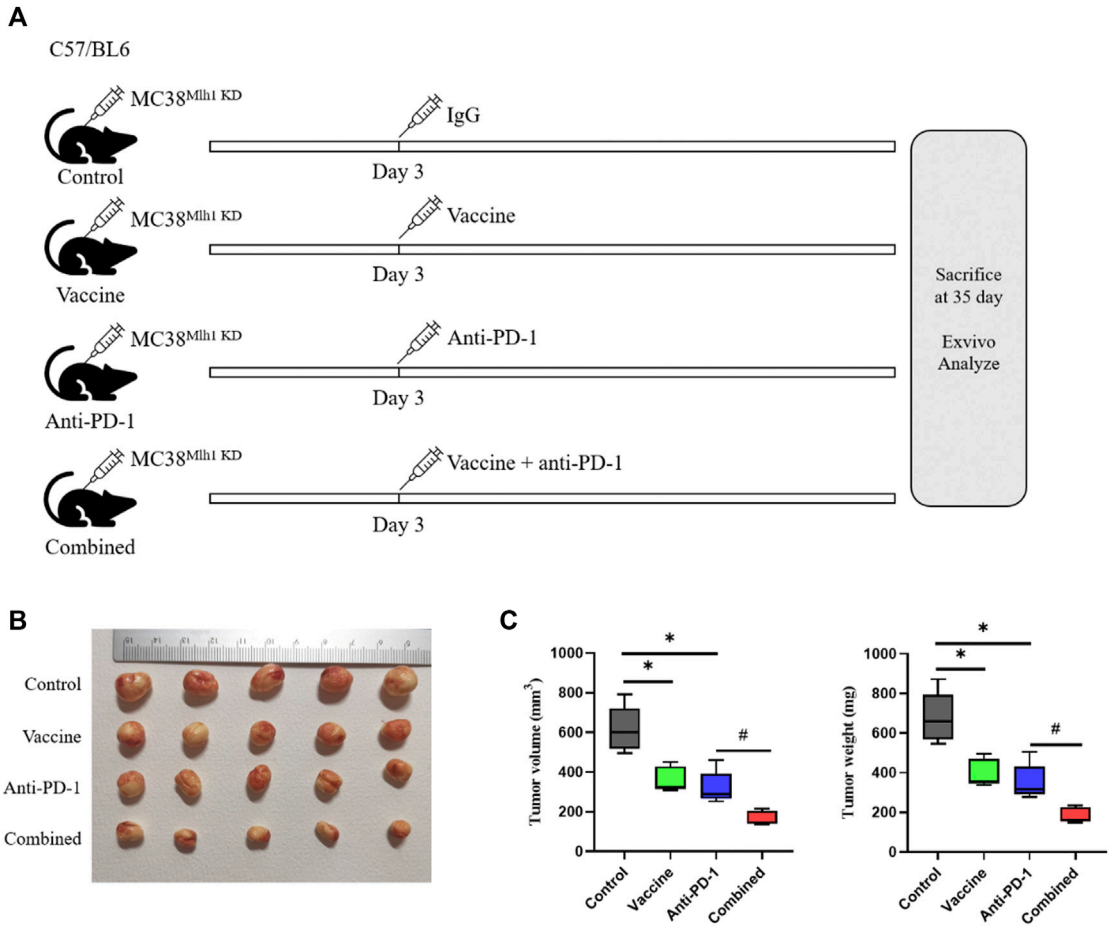


FIGURE 2
Combination therapy processed better anti-tumor effects than PD-1 blockage and vaccine alone. **(A)** Schema of tumor implantation and treatment (C57/BL6 mice were injected with MC38$^{Mlh1\ KD}$ cells in the right lower limb and divided into four groups with five mice in each group). **(B)** Tumors of each group were photographed after being stripped from mice. **(C)** Tumor size and weight were tested after removal from mice. Statistical significance was measured using Student's t-test and one-way ANOVA. *, $p < 0.05$ vs. Control; $^{\#}$, $p < 0.05$ vs. Anti PD-1.

therapy could promote the inflammatory response and reduce the immunosuppression of the circulatory system.

Cancer vaccines can alter the tumor microenvironment

After exploring the effects of PD-1 antibody, vaccines, and combination therapy on the immune system, we probed into the possible effects of various treatment combinations on the TME. In the study, the tumor tissues of mice were excised, and a part of fresh tissues were taken for IHC and H&E staining. The results of H&E staining demonstrated that the immune cell infiltration in the treatment group was more significant than that in the control group (Figure 4A). The result of PD-L1 IHC staining showed that PD-L1 expression was increased in the vaccine treatment and PD-1 antibody treatment groups, and the combination treatment group exhibited the strongest increasing (Figure 4A). Subsequently, in this study, the cell infiltration in tumor tissues was tested by flow cytometry. The result showed that the increase of various immune cells was not notable after vaccine treatment, while the $CD4^+$ T, $CD8^+$ T and $CD11b^+/Gr1^+$ cell infiltration was remarkably increased in PD-1 antibody treatment group. Compared with PD-1 antibody treatment group, the $CD8^+$ T cell content was significantly increased while $CD11b^+/Gr1^+$ cell infiltration was remarkably decreased in the combined treatment group (Figure 4B). After examining the infiltration of immune-related cells, we also detected the infiltration of cells carrying immune checkpoint-related proteins. The results showed that there was an increase in cells carrying $CTL4^+$ and $LAG3^+$ proteins after treatment with vaccine, and there was a significant increase in cells carrying PD-L1 and LAG3 proteins after PD-1 antibody treatment. The results showed that both vaccine treatment and PD-1 antibody treatment caused stress synthesis of

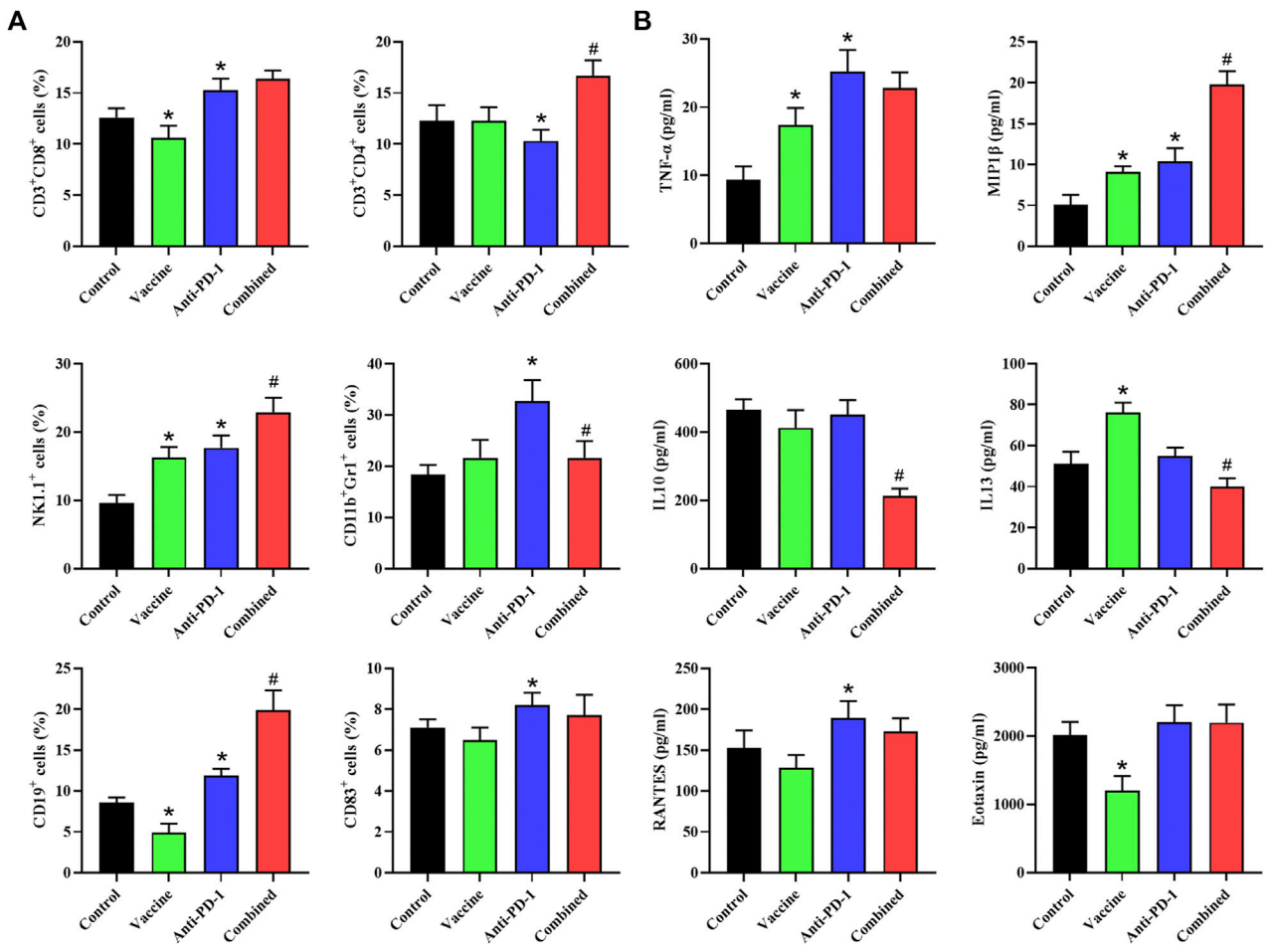


FIGURE 3
Flow cytometry is used to determine the levels of specific immune cells and chemokines in plasma. **(A)** Phenotyping of peripheral blood leukocytes from mice with different treatment combinations. **(B)** Plasma cytokine levels from mice with different treatment combinations. Statistical significance was measured using Student's t-test and one-way ANOVA. *, $p < 0.05$ vs. Control; #, $p < 0.05$ vs. Anti PD-1.

immunosuppressive proteins in cancer cells (Figure 4C). What's more, the results of combination therapy showed that compared with PD-1 antibody therapy alone, the expression of PD-L1 and LAG3 in tumor tissues were markedly decreased, and the expression of CTLA-4 was significantly increased. These data showed that the combination therapy exhibited a better inhibitory effect on immune escape than PD-1 monoclonal antibody therapy alone (Figure 4C).

Tumor vaccine combined with ICIs inhibited the proliferation of tumor tissue and angiogenesis

To further explore the effect of vaccine, PD-1 antibody and combination therapy on the treatment of Lynch syndrome, we applied IHC to detect the proliferation and angiogenesis of tumor tissues in mice. The result of Ki67 expression demonstrated that vaccine, PD-1 antibody, and combination therapy treatment could reduce the level of tumor cell proliferation in Lynch syndrome, and combination therapy acted the best effect (Figures 5A,B). The results of CD31 staining illustrated that vaccine, PD-1 antibody and combination therapy treatment reduced angiogenesis in Lynch syndrome mouse model, and combination therapy exhibited the best effect (Figures 5A,B). The above results suggested that vaccine and PD-1 antibody treatment could inhibit the proliferation level of tumor cells and reduce angiogenesis in tumor tissues, combination therapy treatment exerted the best inhibitory effect.

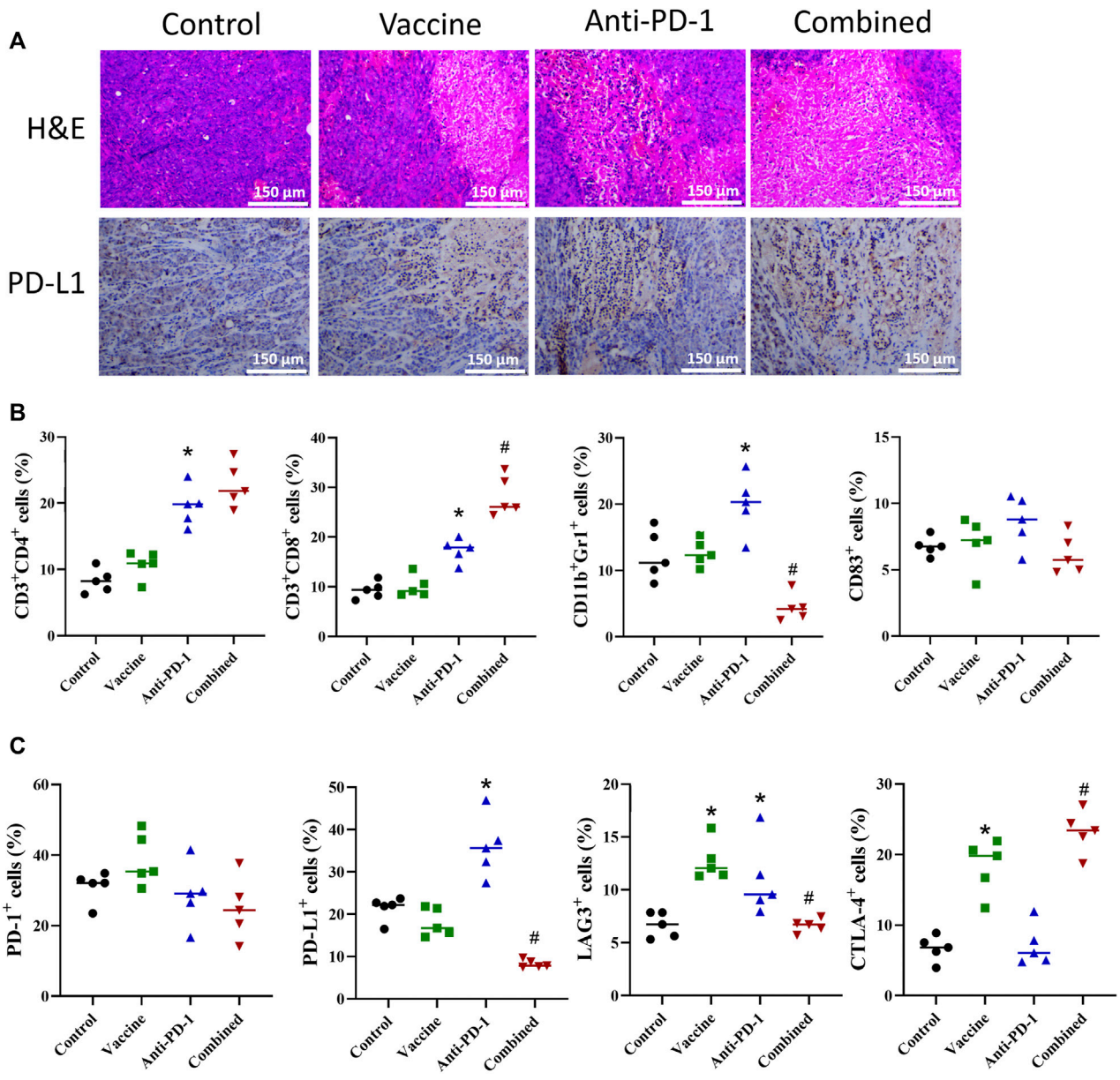


FIGURE 4
Effect of different combination of treatment on TME. **(A)** Representative H&E staining and IHC from mice with different treatment combinations. **(B–C)** Flow cytometric phenotyping of cells in colon cancer TME. Statistical significance was measured using Student's t-test and one-way ANOVA. *, $p < 0.05$ vs. Control; #, $p < 0.05$ vs. Anti PD-1.

Discussion

Cancer therapy for patients with Lynch syndrome has been a tricky point, and the advent of immunotherapy, especially ICIs, has enormously improved the survival of patients with Lynch syndrome (Lee and Le, 2016; Winer et al., 2019). Because of the strong immunogenicity of tumor tissues in patients with Lynch syndrome, patients have a great response to ICIs (Cerretelli et al., 2020). At present, in addition to the use of ICIs for treating cancer patients with Lynch syndrome, cancer vaccine is also an effective treatment method. At present, many studies have attempted to use tumor vaccines to prevent the incidence of tumors in patients with Lynch syndrome. For example, researchers have discovered that injection of immunogenic frameshiftpeptides (FSP) into patients with Lynch syndrome can prevent the incidence of Lynch syndrome-related cancers (von Knebel Doeberitz and Kloor, 2013; Majumder et al., 2018). In addition, tumor antigen peptide vaccines prepared by using

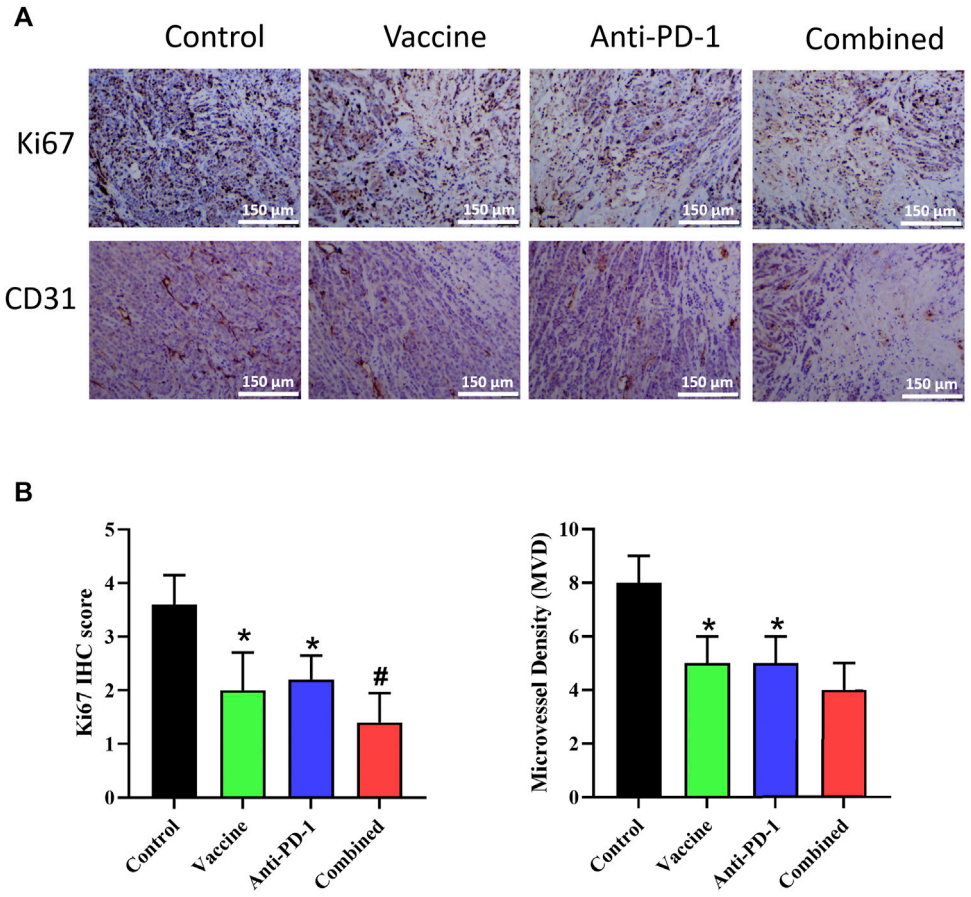


FIGURE 5
Effect of vaccine, PD-1 blockage and combination therapy on cellular proliferation, and vessel density in tumors. **(A,B)** Ki67^{+} cells and microvessel density are quantified in mice treated with different therapy. Statistical significance was measured using Student's t-test and one-way ANOVA. *, $p < 0.05$ vs. Control; #, $p < 0.05$ vs. Anti PD-1.

tumor-specific antigens can directly kill tumor tissue or improve the therapeutic effect of other immunotherapies by stimulating the patient's immune system and recruiting killer immune cells (Soares et al., 2015). At present, a large number of clinical trials related to cancer vaccines have been carried out, and encouraging results have been obtained in melanoma (Carreno et al., 2015; Li et al., 2017). In this study, we firstly clarified that tumor vaccine effect was specific to the Lynch syndrome model (MC38$^{\text{Mlh KD}}$). Besides, we prepared MC38$^{\text{Mlh1KD}}$ Lynch syndrome model cells into cell lysate vaccine and treated homograft mice with different methods, and found that tumor vaccine combined with PD-1 blocking therapy could increase the survival rate of mice and prolong the overall survival time of tumor-bearing mice. Subsequent experiments also revealed that the combined treatment could reduce the growth rate of tumor tissues in mice. The current study found that combining tumor vaccines with immunotherapy can hinder tumor growth by improving the therapeutic effect of traditional immunotherapy. For example, the team of Duraiswamy et al. (2013). combined tumor vaccines with PD-1/CTLA-4 blockage to successfully improve the activity of effector T cells in mouse tumor cells, weaken the inhibitory effect of T-reg and strengthen the therapeutic effect of PD-1/CTLA-4 blockage. Herein, we found that tumor vaccines prepared from Lynch syndrome cell lysates could assist the therapeutic effect of ICIs and markedly improve the survival time of tumor-bearing mice.

The TME is a collection of a series of cells in tumor tissues such as tumor cells, mechanistic cells, and immune cells, which can affect the growth of tumor tissues in a variety of ways, such as lower oxygen content can inhibit the immune response and promote the division of tumor tissues, in addition, lower pH can also curb the activity of immune cells, attenuate the effect of chemotherapeutic drugs, and facilitate immune escape of tumors (Vanichapol et al., 2018; Lei et al., 2020). Since the TME has a significant and auxiliary part in the growth of tumor tissue, its targeting role

in cancer therapy cannot be disregarded. There are many current treatments to target the TME, such as the study team using the combination therapy of Apatinib to promote the immune response of the immune system by improving blood supply to tumor tissues and reducing the hypoxic condition of tumor tissues (Zhao et al., 2019). In addition, there are pH-sensitive nanomedicines designed using the tumor pH microenvironment to inhibit tumor development by inducing apoptosis in tumor tissue (Wijesinghe et al., 2013). Recent studies have found that tumor vaccines are an effective means to improve the TME and avoid immune escape from tumor tissues, for example, Xu et al. (2020).s team regulated the TME through Listeria-based hepatocellular carcinoma vaccine and induced macrophage differentiation to M1 phenotype to promote immune system killing of tumor tissues. In this study, we examined the immune infiltration in the circulatory system and TME of homograft mice treated with tumor vaccine using flow cytometry and ELISA, and found that killer immune cells such as CD8, CD4, and NK were significantly increased after treatment, while immunosuppressive cells MDSC and immunosuppressive cells with PD-1 and LAG3 antigens were increased after PD-1treatment and significantly decreased after vaccine combination therapy. This suggests that tumor vaccines can reverse the drug resistance response of tumor tissues to PD-1 treatment and inhibit the resistance of tumor tissues to PD-1 blocking therapy. In addition, we detected the contents of immune factors and cytokines in the circulatory system of mice after vaccine treatment, and the results demonstrated that the contents of immunosuppressive cytokines were notably reduced in the combined treatment group, and the expression levels of inflammatory cytokines such as TNF-α and MIP-1β were remarkably elevated in the combined treatment group. In summary, this study found that cancer vaccines can increase the anti-tumor immune cell content in tumor tissues by improving the TME and regulating immune infiltration, which can not only inhibit the development of cancer itself, but also synergize ICIs immunotherapy.

Angiogenesis of the TME and proliferation of tumor tissue are the main causes of rapid tumor development (Zhao et al., 2019). Therefore, this study also examined the effects of immunotherapy as well as vaccine therapy on tumor cell proliferation and tumor tissue angiogenesis, and the results showed that vaccine therapy could enhance the inhibitory effect of immunotherapy on tumor cell proliferation and improve the inhibitory effect of immunotherapy on angiogenesis. In this study, we found that the combination therapy could greatly inhibit the proliferation of tumor cells and the angiogenesis of tumor tissues, thus inhibiting the growth of tumor tissues.

In summary, this paper found that tumor vaccines had a therapeutic effect on colon cancer induced by Lynch syndrome, which can also assist ICIs therapy to improve the therapeutic effect of ICIs. At the same time, tumor vaccines could regulate the TME, increase immune infiltration in tumor tissues, reduce the proportion of immunosuppressive cells, inhibit tumor growth by inhibiting tumor cell proliferation and angiogenesis in tumor tissues, and exert a therapeutic role in colon cancer caused by Lynch syndrome. Tumor vaccine was confirmed in our study to have a great therapeutic effect on colon cancer caused by Lynch syndrome. The therapeutic effect of ICIs treatment was improved by tumor vaccines and investigated by animal experiments. This study provides a reliable theoretical reference for the treatment of related cancers caused by Lynch syndrome. Although this study is enough to prove the role of cancer vaccines in treatment, there are still shortcomings. In this study, we applied the Lynch syndrome model with $MC38^{Mlh1\ KD}$ cells injected into the limb of mice, which lacked the content of the native colon cancer environment. We would try to construct modeling grafted colon cancer with a proper microenvironment in the follow-up study. Due to conditions and other reasons, this paper does not well simulate the actual incidence of Lynch syndrome, and does not further explore the molecular mechanism by which the vaccine works. The team intends to further construct an *in vivo* immune reconstitution model and further study the effect of tumor vaccines and the molecular mechanism using techniques such as gene editing on a model closer to Lynch syndrome.

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 Ethics Review of the Second Affiliated Hospital of Fujian Medical University (No. 2021287).

Author contributions

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

Funding

This study was supported by the funds from Medical Innovation Project of Fujian Province in 2021, Mechanism of

BTB structural protein 3 regulating the proliferation and migration of colorectal cancer cells (No. 2021CXA028); Project supported by Fujian Provincial Natural Science Foundation in 2020, The mechanism of hsa_circ_0060967 as a ceRNA affecting the invasion and metastasis of colorectal cancer cells by regulating miR-338-3p (Grant No. 2020J01209).

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/fgene.2022.877833/full#supplementary-material

References

Bari, S., Kim, R. D., Wang, X., Matejcic, M., and Muzaffar, J. (2020). Outcomes of Lynch syndrome (LS) patients treated with immune checkpoint inhibitors (ICI). *J. Clin. Oncol.* 38, 1548. doi:10.1200/JCO.2020.38.15_suppl.1548

Blank, C., and Mackensen, A. (2007). Contribution of the PD-L1/PD-1 pathway to T-cell exhaustion: An update on implications for chronic infections and tumor evasion. *Cancer Immunol. Immunother.* 56, 739–745. doi:10.1007/s00262-006-0272-1

Buonaguro, L., and Tagliamonte, M. (2020). Selecting target antigens for cancer vaccine development. *Vaccines* 8. doi:10.3390/vaccines8040615

Carreno, B. M., Magrini, V., Becker-Hapak, M., Kaabinejadian, S., Hundal, J., Petti, A. A., et al. (2015). Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* 348, 803–808. doi:10.1126/science.aaa3828

Cerretelli, G., Ager, A., Arends, M. J., and Frayling, I. M. (2020). Molecular pathology of Lynch syndrome. *J. Pathol.* 250, 518–531. doi:10.1002/path.5422

Duraiswamy, J., Kaluza, K. M., Freeman, G. J., and Coukos, G. (2013). Dual blockade of PD-1 and CTLA-4 combined with tumor vaccine effectively restores T-cell rejection function in tumors. *Cancer Res.* 73, 3591–3603. doi:10.1158/0008-5472.CAN-12-4100

Janjigian, Y. Y., Shitara, K., Moehler, M., Garrido, M., Salman, P., Shen, L., et al. (2021). First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): A randomised, open-label, phase 3 trial. *Lancet* 398, 27–40. doi:10.1016/S0140-6736(21)00797-2

Jia, Y., Liu, L., and Shan, B. (2020). Future of immune checkpoint inhibitors: Focus on tumor immune microenvironment. *Ann. Transl. Med.* 8, 1095. doi:10.21037/atm-20-3735

Lee, V., and Le, D. T. (2016). Efficacy of PD-1 blockade in tumors with MMR deficiency. *Immunotherapy* 8, 1–3. doi:10.2217/imt.15.97

Lei, X., Li, J-K., Du, W-X., Li, R-G., Yang, J., Li, 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, L., Goedegebuure, S. P., and Gillanders, W. E. (2017). Preclinical and clinical development of neoantigen vaccines. *Ann. Oncol.* 28, xii11–xii17. doi:10.1093/annonc/mdx681

Majumder, S., Shah, R., Elias, J., Manoharan, M., Shah, P., Kumari, A., et al. (2018). A cancer vaccine approach for personalized treatment of Lynch Syndrome. *Sci. Rep.* 8, 12122. doi:10.1038/s41598-018-30466-x

Rizzo, A., Ricci, A. D., and Brandi, G. (2021). PD-L1, TMB, MSI, and other predictors of response to immune checkpoint inhibitors in biliary tract cancer. *Cancers (Basel)* 13. doi:10.3390/cancers13030558

Sinicrope, F. A. (2018). Lynch syndrome-associated colorectal cancer. *N. Engl. J. Med.* 379, 764–773. doi:10.1056/NEJMcp1714533

Soares, K. C., Rucki, A. A., Wu, A. A., Olino, K., Xiao, Q., Chai, Y., et al. (2015). PD-1/PD-L1 blockade together with vaccine therapy facilitates effector T-cell infiltration into pancreatic tumors. *J. Immunother.* 38, 1–11. doi:10.1097/CJI.0000000000000062

Tanaka, Y., Wada, H., Goto, R., Osada, T., Yamamura, K., Fukaya, S., et al. (2020). TAS0314, a novel multi-epitope long peptide vaccine, showed synergistic antitumor immunity with PD-1/PD-L1 blockade in HLA-A*2402 mice. *Sci. Rep.* 10, 17284. doi:10.1038/s41598-020-74187-6

Vaddepally, R. K., Kharel, P., Pandey, R., Garje, R., and Chandra, A. B. (2020). Review of indications of FDA-approved immune checkpoint inhibitors per NCCN guidelines with the level of evidence. *Cancers (Basel)* 12. doi:10.3390/cancers12030738

Vanichapol, T., Chutipongtanate, S., Anurathapan, U., and Hongeng, S. (2018). Immune escape mechanisms and future prospects for immunotherapy in neuroblastoma. *Biomed. Res. Int.* 2018, 1812535. doi:10.1155/2018/1812535

von Knebel Doeberitz, M., and Kloor, M. (2013). Towards a vaccine to prevent cancer in Lynch syndrome patients. *Fam. Cancer* 12, 307–312. doi:10.1007/s10689-013-9662-7

Wijesinghe, D., Arachchige, M. C., Lu, A., Reshetnyak, Y. K., and Andreev, O. A. (2013). pH dependent transfer of nano-pores into membrane of cancer cells to induce apoptosis. *Sci. Rep.* 3, 3560. doi:10.1038/srep03560

Winer, A., Ghatalia, P., Bubes, N., Anari, F., Varshavsky, A., Kasireddy, V., et al. (2019). Dual checkpoint inhibition with ipilimumab plus nivolumab after progression on sequential PD-1/PDL-1 inhibitors pembrolizumab and atezolizumab in a patient with Lynch syndrome, metastatic colon, and localized urothelial cancer. *Oncologist* 24, 1416–1419. doi:10.1634/theoncologist.2018-0686

Winer, E. P., Lipatov, O., Im, S-A., Goncalves, A., Muñoz-Couselo, E., Lee, K. S., et al. (2020). Association of tumor mutational burden (TMB) and clinical outcomes with pembrolizumab (pembro) versus chemotherapy (chemo) in patients with metastatic triple-negative breast cancer (mTNBC) from KEYNOTE-119. *J. Clin. Oncol.* 38, 1013. doi:10.1200/JCO.2020.38.15_suppl.1013

Xu, G., Feng, D., Yao, Y., Li, P., Sun, H., Yang, H., et al. (2020). Listeria-based hepatocellular carcinoma vaccine facilitates anti-PD-1 therapy by regulating macrophage polarization. *Oncogene* 39, 1429–1444. doi:10.1038/s41388-019-1072-3

Zhao, S., Ren, S., Jiang, T., Zhu, B., Li, X., Zhao, C., et al. (2019). Low-dose Apatinib optimizes tumor microenvironment and potentiates antitumor effect of PD-1/PD-L1 blockade in lung cancer. *Cancer Immunol. Res.* 7, 630–643. doi:10.1158/2326-6066.CIR-17-0640

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 14 September 2022
DOI 10.3389/fgene.2022.973444

OPEN ACCESS

EDITED BY
Tao Huang,
Shanghai Institute of Nutrition and Health (CAS), China

REVIEWED BY
Qing Chang,
Shanghai University of Medicine and Health Sciences, China
Qi Sun,
Shenzhen University, China

*CORRESPONDENCE
Xiaxing Deng,
kejiadxx@hotmail.com
Baiyong Shen,
shenby@shsmu.edu.cn

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

SPECIALTY SECTION
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

RECEIVED 20 June 2022
ACCEPTED 01 August 2022
PUBLISHED 14 September 2022

CITATION
Ma D, Yang Y, Cai Q, Ye F, Deng X and Shen B (2022), Identification of a lncRNA based signature for pancreatic cancer survival to predict immune landscape and potential therapeutic drugs. *Front. Genet.* 13:973444. doi: 10.3389/fgene.2022.973444

Identification of a lncRNA based signature for pancreatic cancer survival to predict immune landscape and potential therapeutic drugs

Di Ma[1†], Yuchen Yang[1†], Qiang Cai[1], Feng Ye[1], Xiaxing Deng[1,2*] and Baiyong Shen[1,2*]

[1]Department of General Surgery, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [2]Shanghai Key Laboratory of Translational Reseach for Pancreatic Neoplasms, Research Institute of Pancreatic Diseases, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Pancreatic cancer is one major digestive malignancy with a poor prognosis. Given the clinical importance of lncRNAs, developing a novel molecular panel with lncRNAs for pancreatic cancer has great potential. As a result, an 8-lncRNA-based robust prognostic signature was constructed using a random survival forest model after examing the expression profile and prognostic significance of lncRNAs in the PAAD cohort from TCGA. The efficacy and effectiveness of the lncRNA-based signature were thoroughly assessed. Patients with high- and low-risk defined by the signature underwent significantly distinct OS expectancy. Most crucially the training group's AUCs of ROC approached 0.90 and the testing group similarly had the AUCs above 0.86. The lncRNA-based signature was shown to behave as a prognostic indicator of pancreatic cancer, either alone or simultaneously with other factors, after combined analysis with other clinical-pathological factors in Cox regression and nomogram. Additionally, using GSEA and CIBERSORT scoring methods, the immune landscape and variations in biological processes between high- and low-risk subgroups were investigated. Last but not least, drug databases were searched for prospective therapeutic molecules targeting high-risk patients. The most promising compound were Afatinib, LY-303511, and RO-90-7501 as a result. In conclusion, we developed a novel lncRNA based prognostic signature with high efficacy to stratify high-risk pancreatic cancer patients and screened prospective responsive drugs for targeting strategy.

KEYWORDS

drug response, immune landscape, prognostic model, long non-coding RNA, pancreatic cancer

Introduction

Pancreatic cancer is known as a highly lethal malignancy with a poor prognosis. Accounting for approximately 496,000 new patients per year worldwide, pancreatic cancer ranks 14th in new cases among 36 major types of cancer (Sung et al., 2021). However, the disease reaches as high as the 4th in cancer-related death, with the death number around 466,000 annually. On average, the 5-year survival rate of pancreatic cancer is below 10% at the time of diagnosis (Mizrahi et al., 2020; Zhu et al., 2021). Nowadays, surgical resection remains the optimal treatment for pancreatic cancer by increasing the 5-year survival rate to around 20% (Christenson et al., 2020). For unresectable tumors, nonetheless, the efficacy of other approaches such as chemotherapy, radiotherapy, and systemic therapies, despite receiving incremental progress during the last decade, requires further assessment. To note, due to lacking severe symptoms in the early stage, the diagnosis of pancreatic cancer is of great difficulty and often delayed. Therefore, novel molecular markers of satisfying sensitivity and accuracy become an urgent demand for diagnosis and prognosis evaluation purposes. Besides, new strategies for screening patients with higher long-term risk are also expected for better clinical decision-making.

Not surprisingly, considerable amounts of studies have revealed the possibility and value of molecular signatures in the diagnosis and prognosis of pancreatic cancer over the last few years. For instance, Wu et al. have developed a nine-gene (MET, KLK10, COL17A1, CEP55, ANKRD22, ITGB6, ARNTL2, MCOLN3, and SLC25A45) panel to predict the overall survival of pancreatic cancer (Wu et al., 2019). Other groups have also built several signatures based on genes associated with different biological aspects of pancreatic cancer including autophagy, methylation, and metabolic changes (Yu et al., 2021; Xiao et al., 2022; Zhang et al., 2022). Nowadays it is widely acknowledged that the dysregulation of non-coding RNA is closely correlated to different types of tumors including pancreatic cancer. Hence, signature classifiers generated from non-coding RNAs have also been carried out using micro RNA (miRNA), long non-coding RNA (lncRNA), circular RNA (circRNA) and so forth. Nevertheless, the value of non-coding RNAs in assessing pancreatic cancer has not been thoroughly explored, as most of them emphasized prognosis prediction but failed to provide detailed hints on clinical decision making.

Therefore, this study aims to identify the clinical significance of lncRNAs for pancreatic cancer evaluation and construct a comprehensive lncRNA-based signature with high prognostic efficacy to monitor outcomes of pancreatic cancer patients. Besides, the molecular signature is used to explore the immune landscape and potential therapeutic targets and small molecules between risk subgroups. In detail, expression and clinical data of pancreatic cancer patients were acquired from public databases including The Cancer Genome Atlas (TCGA), Cbioportal, and Cancer Cell Line Encyclopedia (CCLE). An 8-lncRNA classifier was then constructed by applying Cox and random survival forest (RSF) regression in differentially expressed lncRNAs (DElncRNAs). The capacity of the signature as a prognostic indicator was evaluated in different aspects. To emphasize, the immune feature landscape, possible therapeutic targets, and molecules were subsequently checked in patients with a high-risk score according to the signature in detail, holding the potential to expand the current therapeutic strategies for the pancreatic cancer population.

Materials and methods

RNA-sequencing cohorts

An RNA-seq dataset of 177 pancreatic cancer patients involving RNA expression value and matched clinical information was obtained from the TCGA data portal (http://portal.gdc.cacner.gov/repository) and the Cbioportal website (http://cbioportal.org). Fragments per million reads (FPKM) normalized expression value was used for further analysis. The cohort was then randomly split with a 2:1 ratio into a training group and a testing group.

Cancer cell line data

Expression profiles of human cancer cell lines (CCLs) were achieved from the Broad Institute CCLE project (http://portals.broadinstitute.org/ccle). To search for potential therapeutic agents, sensitivity data of compounds in CCLs were achieved from the Cancer Therapeutics Response Portal, Board institute (CTRP, http://portals.broadinstitute.org/ctrp) and PRISM repurposing dataset (http://depmap.org/portal/prism). The algorism of drug sensitivity was described in previous studies. Briefly, the database provided the area under the curve (AUC) values as the readout of drug sensitivity. The lower AUC values indicate higher drug sensitivity. Compounds with more than 20% missing data were excluded from the dataset, and the K-nearest neighbor algorithm (K-NN) was applied to estimate the AUC values. To further investigate the mechanism of actions (MoA) of the drugs screened out, the Connectivity Map tools database (CMap, http://clue.io) with more than 2000 small molecule perturbagen types was applied for specific analysis.

Construction of lncRNA-based prognostic signature

The human lncRNA annotation profile was obtained from the GENCODE website (GRCh38.p13, release 39, http://

gencodegenes.org/human). After the acquisition of lncRNA expression data as described above, the lncRNAs were separated from gene-coding RNA and other non-coding RNAs. Differentially expressed lncRNAs (DElncRNAs) were identified with the criteria of absolute log2 fold-change (log2FC) > 1 and adjusted p-value < 0.05 between tumor and control tissues. Afterward, univariate Cox regression was conducted to identify prognostic lncRNAs which shared a correlation with the overall survival (OS) time of the patients in the cohort ($p < 0.05$). Thus, the candidate prognostic lncRNAs were determined by overlapping the DElncRNAs and prognostic lncRNAs.

To develop the lncRNA-based signature, the univariate Cox proportional hazards regression was first applied to preliminarily narrow the candidates using the training group. Subsequently, an RSF regression based on minimal depth was used to finally identify the signature. The RSF regression model underwent iteration 1,000 times to construct a lncRNA-based OS classifier with the largest C-index value. Eventually, multivariate Cox regression was employed to select candidates as independent indicators to form the 8-lncRNA-based signature retained for the next analysis. According to the classifier, each sample in the cohort was endowed with a risk score following the equation:

$$RiskScore = \sum_{k=1}^{n} Coef_k \times Exp_k$$

in which $Coef_k$ was the coefficients of each sample, Exp_k was the expression of member lncRNAs of the signature. The cohort was divided into high-risk and low-risk groups by the mean value of the risk score. Afterward, the efficacy and effectiveness of the lncRNA classifiers in both training, and validation cohorts were evaluated by the Kaplan-Meier long-rank test, Time-dependent ROC curve analysis, multivariate Cox regression and nomogram scoring.

Immune function analysis

Algorithms including CIBERSORT and ssGSEA were applied to compare the pattern of immune infiltration between high-risk and low-risk groups. Moreover, the Tumor Immune Dysfunction and Exclusion (TIDE) scoring method were employed to assess the response to immunotherapy, extent of immune dysfunction, immune exclusion, and microsatellite instability (MSI) for patients in high-risk and low-risk groups.

Statistical analysis

All statistical analyses were conducted with the R software platform (v4.0.2, R Foundation for Statistical Computing, Vienna, Austria). Some major R packages included "edgeR," "limma," "survival," "ROCR," "ggplot2," "pRRophetic," and "randomForestSRC". To compare variables in multiple groups, Student's t-test and ANOVA analysis were used for parametric factors, whereas the Wilcoxon rank-sum test and Kruskal–Wallis test were applied for nonparametric factors. To measure the correlation of different variables, Spearman's rank-order correlation and Pearson's r correlation were set. Furthermore, Kaplan-Meier and the log-rank test were used for survival analysis. The area under the curve (AUC) was measured to judge the efficacy of the receiver operating characteristic curve. For all statistical calculations, a two-tailed $p < 0.05$ was considered significant.

Results

Construction of lncRNA based prognostic signature

The whole RNA transcriptome profile containing tumor tissue ($n = 177$) and adjacent control ($n = 4$) was obtained from the TCGA portal as described above. Of the 14,078 lncRNAs extracted from the RNA-seq dataset, 540 lncRNAs were identified as DElncRNA under the condition of absolute log2 FC > 1 and adjust $p < 0.05$ (Figure 1; Figure 2A; Supplementary Figure S1). On the other hand, a univariate Cox regression analysis was used to select 2676 prognostic significant lncRNAs ($p < 0.05$) that possibly correlated to the OS time of the patients. A Venn diagram was created by overlapping DElncRNAs and univariate COX positive lncRNAs to select the candidate lncRNAs (Figure 2B). Next, a random survival forest (RSF) model was built based on minimal depth to screen out lncRNAs were most relevant to the prognosis. The RSF went through 1,000 times iterations under the criterion of largest C-index value and eventually led to an 8-lncRNA signature as the prognostic signature model for pancreatic cancer (Figure 2C). Thus, the multivariate Cox regression was applied to determine the risk score for each patient by values of coefficient and expression as described above (Supplementary Figure S2A). The detailed information on the elemental lncRNAs was listed as follows (Table 1).

Assessment of the prognostic potentiality of lncRNA based signature

As the 8-lncRNAs-based classifier for pancreatic cancer was constructed, its efficacy for prognosis induction was evaluated in all aspects. First, patients were divided into high-risk and low-risk groups according to the median value of the risk score in all the training, validation and whole groups. Hence, the distribution of risk scores, the

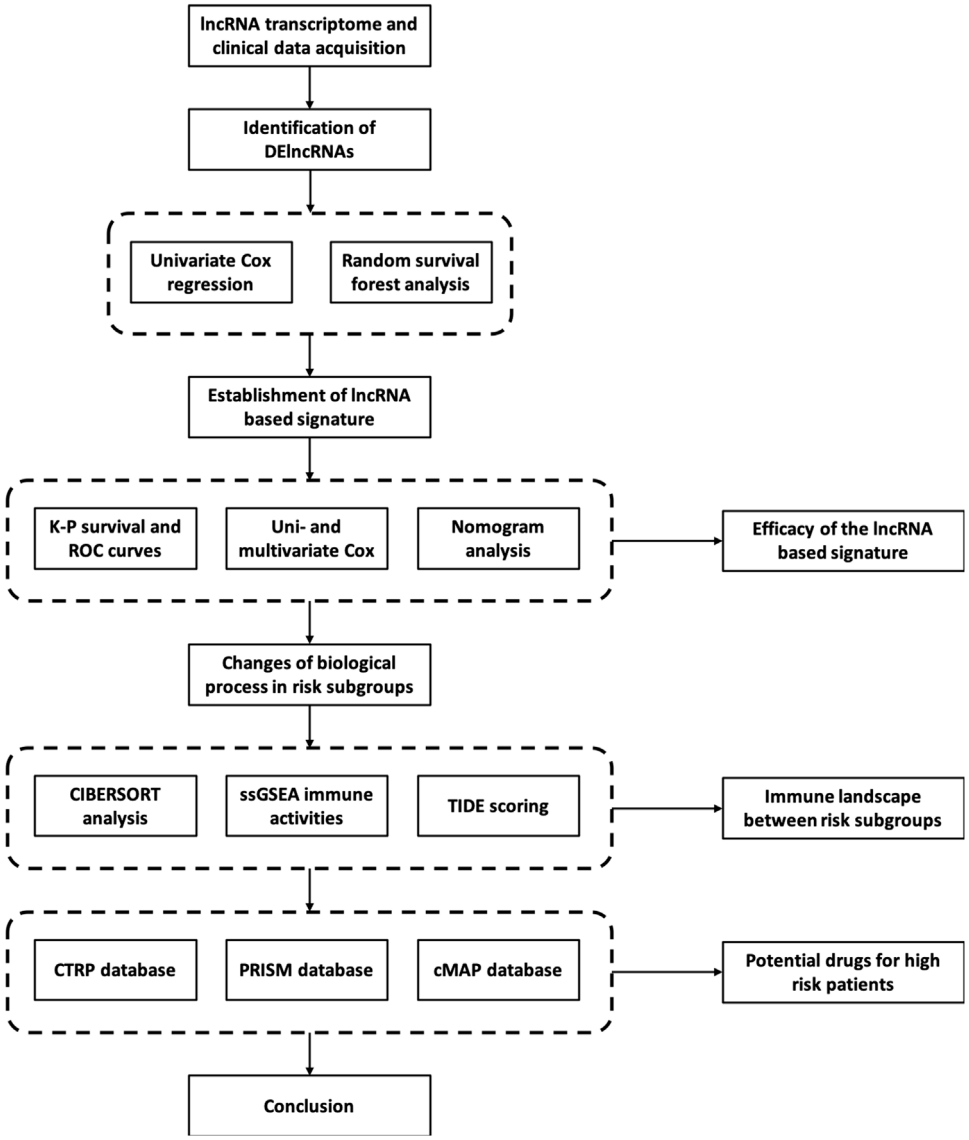


FIGURE 1
The work flow of the study.

correlation between vital status and risk score and the expression pattern of elemental lncRNAs were shown in detail (Figures 2D–F). To investigate the relationship between 8 prognostic lncRNAs and the risk score. Spearman's correlation analysis was conducted among the expression of element lncRNAs and the risk score (Supplementary Figure S2B). Interestingly, the expression of most members was found strongly correlated with the level of risk score, represented by MIR600HG with a Spearman's coefficient of −0.54 and CASC8 with 0.58. In addition, certain components of the classifier share a closer relationship in expression. For example, the correlation coefficient between Lnc-PQLC1-10 and CASC8 was more than 0.5, suggesting a potential biological relevance might exist between them in the development of pancreatic cancer.

Afterward, the expression of lncRNAs of the prognostic classifier was compared in groups with different risk levels to further evaluate the differential expression pattern accompanied by the risk score (Supplementary Figure S2C). As a result, 7 out of 8 components (Lnc-ROBO2-3 excluded) of the signature expressed differently between groups with different risk levels. Among them, 5 lncRNAs (Lnc-PQLC1-10, CASC8, Lnc-KAT7-3, MIR924HG and Lnc-PDK2-5) were found up-regulated in patients with a higher risk score, while MIR600HG and Lnc-

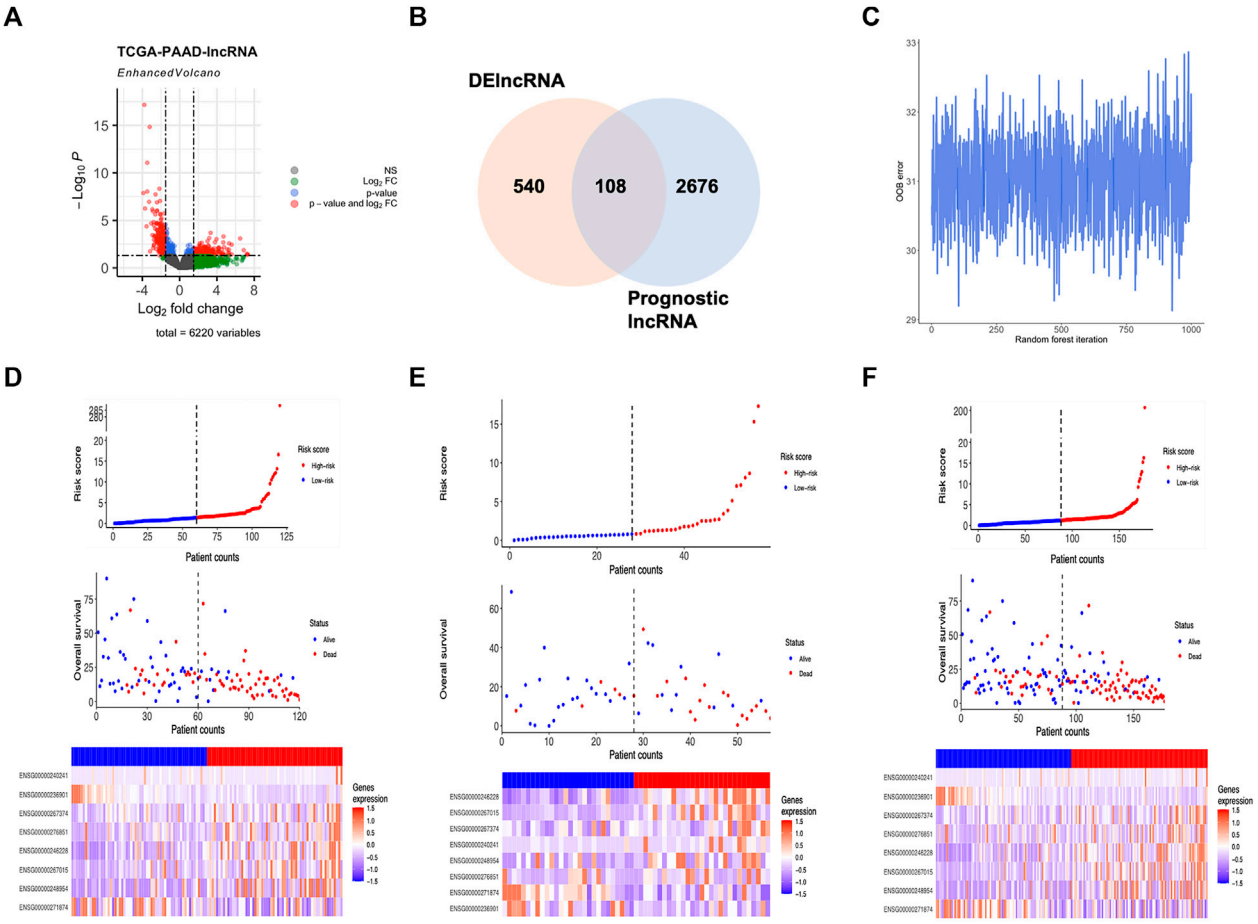


FIGURE 2
Development of lncRNA based molecular signature. **(A)** Volcano plot showed DElncRNAs identified from the TCGA-PAAD dataset. **(B)** Venn diagram of candidate lncRNAs obtained by overlapping DElncRNAs and Cox positive lncRNAs in the training cohort. **(C)** OOB error in 1,000 iteration of the random survival forest regression. **(D)** Distribution of the lncRNA based signature and expression of component lncRNAs in training group. **(E)** Distribution of the lncRNA based signature and expression of component lncRNAs in testing group. **(F)** Distribution of the lncRNA based signature and expression of component lncRNAs in whole group.

TABLE 1 Information of the lncRNAs in pancreatic cancer prognostic signature. 8 lncRNA-based molecular classifier.

Gene id	Gene name	Chromosome	Start point	End point
ENSG00000236901	MIR600HG	9q33.3	125871773	125877756
ENSG00000267015	LncPQLC1-10	18q23	79337837	79344139
ENSG00000246228	CASC8	8q24.21	127277047	127482140
ENSG00000248954	Lnc-KAT7-3	17q21.33	49887597	49936831
ENSG00000271874	Lnc-RAD1-3	5p13.2	34647370	34656270
ENSG00000240241	Lnc-ROBO2-3	3p12.3	78266893	78298888
ENSG00000267374	MIR924HG	16q22.1	39113067	*39800322*
ENSG00000276851	Lnc-PDK2-5	17q21.33	50094064	50094647

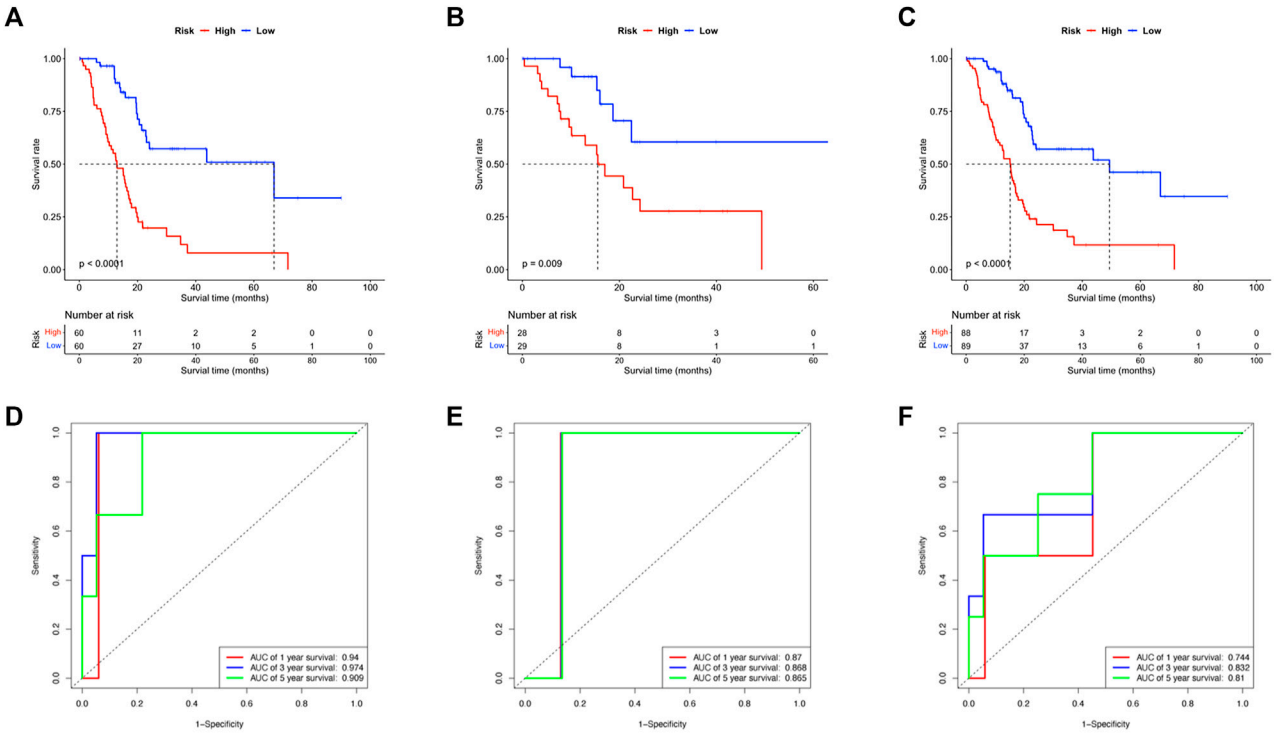


FIGURE 3
Evaluation of the efficacy of lncRNA based signature. **(A–C)** The Kaplan-Meier survival curves compared the patients in high- and low-risk subgroups of all training, testing and whole cohorts. **(D–F)** The ROC curve anticipated the 1, 3, and 5 years survival of patients in all training, testing and whole cohorts.

RAD1-3 expressed lower in the high-risk group. Therefore, more studies and experiments are in great need to explore the possible biological role of these novel lncRNAs in tumorigenesis and the development of pancreatic cancer.

For the Kaplan-Meier survival curve, high-risk and low-risk groups were compared in the training, validation, and whole cohorts respectively (Figures 3A–C). In all cohorts, patients in high-risk groups showed significantly poorer outcomes of shorter average survival time where $p < 0.01$ unanimously. These results indicated that the OS- and RFS classifiers are significantly linked with the prognosis of pancreatic cancer, which holds the potential as an effective prediction model. The results claimed that the 8-lncRNA-based signature strongly correlated to the outcome of pancreatic cancer, thus holding the possibility as a prognosis indicator. The time-dependent receiver operating characteristic (ROC) analysis was subsequently performed. The areas under the ROC curve (AUCs) of the classifier were 0.94, 0.97 and 0.90 for 1, 3, and 5 years of anticipation in the training group (Figure 3D), 0.87, 0.86 and 0.86 in the validation group (Figure 3E), 0.75, 0.83 and 0.81 for 1, 3, and 5 years in the whole group (Figure 3F). Moreover, the lncRNA-based panel signature was also assessed when the recurrence happened to predict other outcomes of the disease (Supplementary Figures S3C,D). Despite the signature that could differ the recurrence time and status in K-M regression, it failed to meet comparable accuracy in ROC analysis with less than 0.5 in 5-year anticipation.

Comprehensive analysis of lncRNA based signature and clinical characteristics

As described, the 8-lncRNAs-based molecular signature was capable of being a novel prognosis indicator with high efficacy for pancreatic cancer. Nevertheless, whether the lncRNA-based signature was relevant to conventional clinicopathologic characteristics remains unclear and requires further study. Clinical data were obtained as previously described and major clinical factors were listed with the risk score for combined analysis.

First, the extent of tumor stage and histological grade were found positively associated with the risk score in Pearson's chi-square analysis, supporting the conclusion that a higher risk

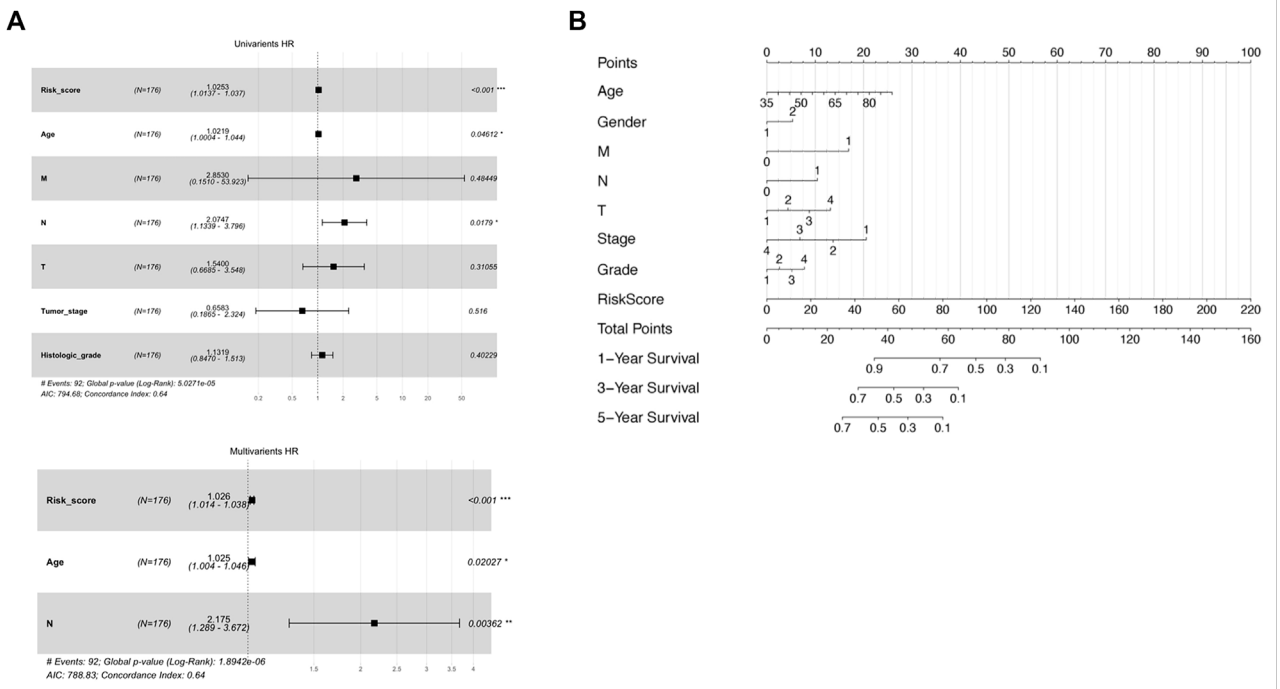


FIGURE 4
Assessment of prognostic ability of both separate and combined usage of risk score and clinical factors. **(A)** Uni-and multivariate Cox regression by risk score and other clinical factors as OS predictor. **(B)** Nomogram including risk score determined by the lncRNA-based signature and other clinical factors for OS prognostic assessment.

score represents advanced tumor progression and worse outcome (Supplementary Figure S3A,B). Next, major clinicopathological characteristics including the risk score were jointly assessed in a two-step Cox regression analysis. As a result, the age of patients, number of malignant lymph nodes, and the risk score were the three factors significantly associated with survival time in a univariate Cox test and thus further went through a multivariate Cox analysis (Figure 4A). Interestingly, all the three variables remained significant in the following multivariate Cox survival test (Figure 4B). Therefore, the 8-lncRNA-based signature and its risk scoring could be considered an independent factor for the prognosis prediction of pancreatic cancer.

To develop a practical, comprehensive model for outcome prediction in pancreatic cancer, a nomogram involving the risk scoring and other clinical characteristics was established (Figure 4C). In the very method, each clinical feature received one certain point according to its statistical weight in prognosis prediction. And the total points reflected the probability of 1, 3, and 5 years of survival. To note, the risk score of the lncRNA-based signature was of most importance and predominance while factors such as gender and histological grade of the tumor weighted minimally in the model.

Functional enrichment analysis between risk subgroups

To gain a deeper understanding of the novel mechanism underlying the lncRNA-based molecular classifier, Gene set enrichment analysis (GSEA) was applied to investigate distinguished genes, pathways, and biological processes between subgroups with a different risk scores. In specific, 1755 differentially expressed genes (DEGs) were identified between the high-risk and low-risk subgroups at the condition of absolute log2FC > 1 and $p < 0.05$. For signaling pathways described in Kyoto Encyclopedia of Genes and Genomes (KEGG), these DEGs were significantly enriched in 27 pathways in the high-risk group and 16 pathways in the low-risk group, respectively. Among them, cell cycle, DNA replication, retinol metabolism, spliceosome, and systemic lupus erythematosus were the top 5 KEGG pathways enriched in the high-risk group according to normalized enrichment score (NES), while calcium signaling pathway, hematopoietic cell lineage, neuroactive ligand-receptor interaction primary immunodeficiency and renin-angiotensin system were the most relevant pathways in low-risk group (Figures 5A,B). In addition, biological process (BP) was also evaluated as a major aspect of gene annotation (GO) analysis, where chromatin

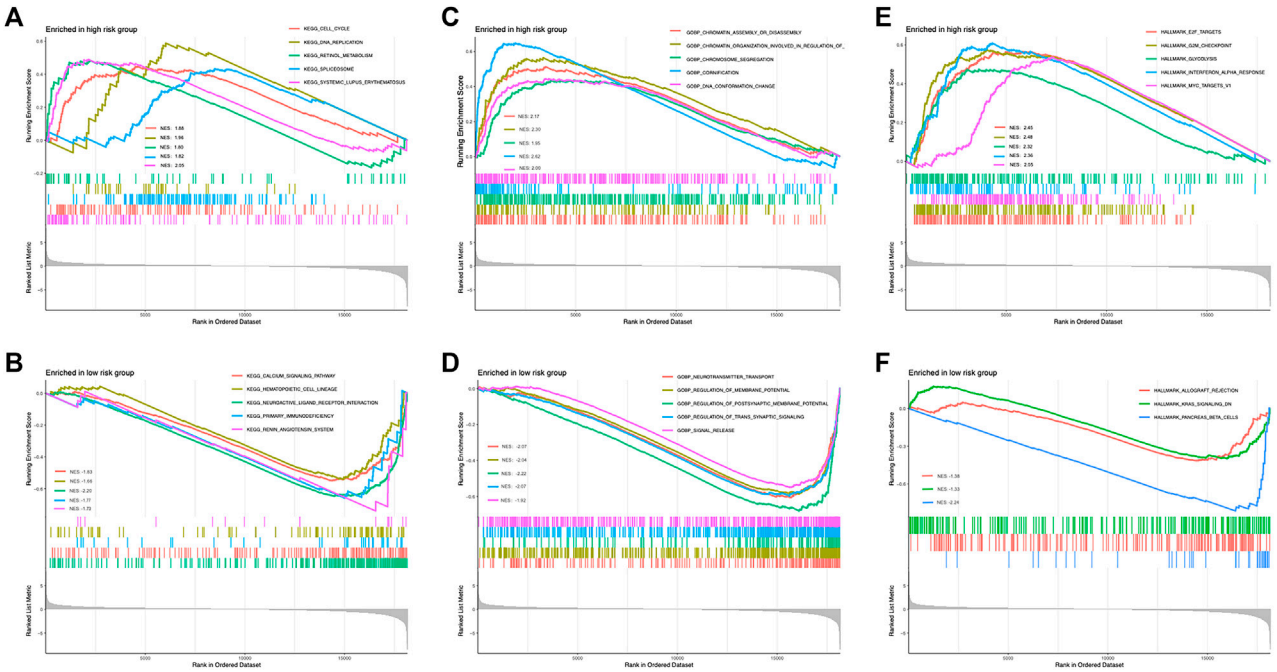


FIGURE 5
Enrichment of distinguished genes, pathways, and biological processes between low- and high-risk subgroups. **(A)** Top enriched KEGG signaling pathways in high risk group. **(B)** Top enriched KEGG signaling pathways in low risk group. **(C)** Top enriched GO biological processes in high risk group. **(D)** Top enriched GO biological processes in low risk group. **(E)** Top enriched hallmarks in high risk group. **(F)** Top enriched hallmarks in low risk group.

assembly or disassembly, chromatin organization involved in the regulation of transcription, chromosome segregation, cornification, and DNA conformation change were the most associated BP in the high-risk group, and neurotransmitter transport, regulation of membrane potential, regulation of postsynaptic membrane potential, regulation of trans synaptic signaling and signal release were enriched in the low-risk group (Figures 5C,D).

Last but not least, hallmark gene sets defined by Molecular Signature Database (MSigDB, http://software.broadinstitute.org/gsea/misigdb) were also checked. Interestingly, gene clusters related to E2F targets, G2M checkpoint, glycolysis, interferon α response, and MYC targets were the top hallmarks positively correlated to the high-risk group, but only 3 clusters, allograft rejection, KRAS signaling, and pancreas β cells were relevant to the low-risk group in our whole cohort (Figures 5E,F).

Estimation of the tumor immune microenvironment

With the increasing application of immunotherapy for pancreatic cancer in recent years, it is intriguing to investigate the possible variation of immune microenvironment among risk subgroups. To date, the infiltration and enrichment of 22 main immune cells in the subgroups were analyzed *via* the CIBERSORT algorithm (Figure 6A; Supplementary Figure S4A). The result suggested that patients with a high-risk level tended to have decreased B naïve cells but elevated M1 and M2 subtypes of macrophages and mast cells, while other immune cells remained indifferent. Moreover, Kaplan-Meier survival regression was performed to investigate the effect on patient outcomes imposed by specific immune cells according to CIBERSORT scoring (Supplementary Figure S4B–G). And in agreement with their infiltration profile, B naïve cell, M1/M2 Macrophage and mast cells were also observed to impose a significant effect on the prognosis of patients in the cohort.

Besides CIBERSORT, ssGSEA was also used to compare the activity of some major immune cells and functions between different risk subgroups (Figure 6B). Noticeably, antitumoral immune cells including B cells, $CD8^+$ T cells, T helper cells and Neutrophils, accompanied by immune functions such as T cell inhibition, T cell stimulation and type II IFN response were discovered a reduced activation in high-risk rather than low-risk patients.

Since the immune microenvironment was altered among different risk subgroups, it is reasonable to speculate the effect

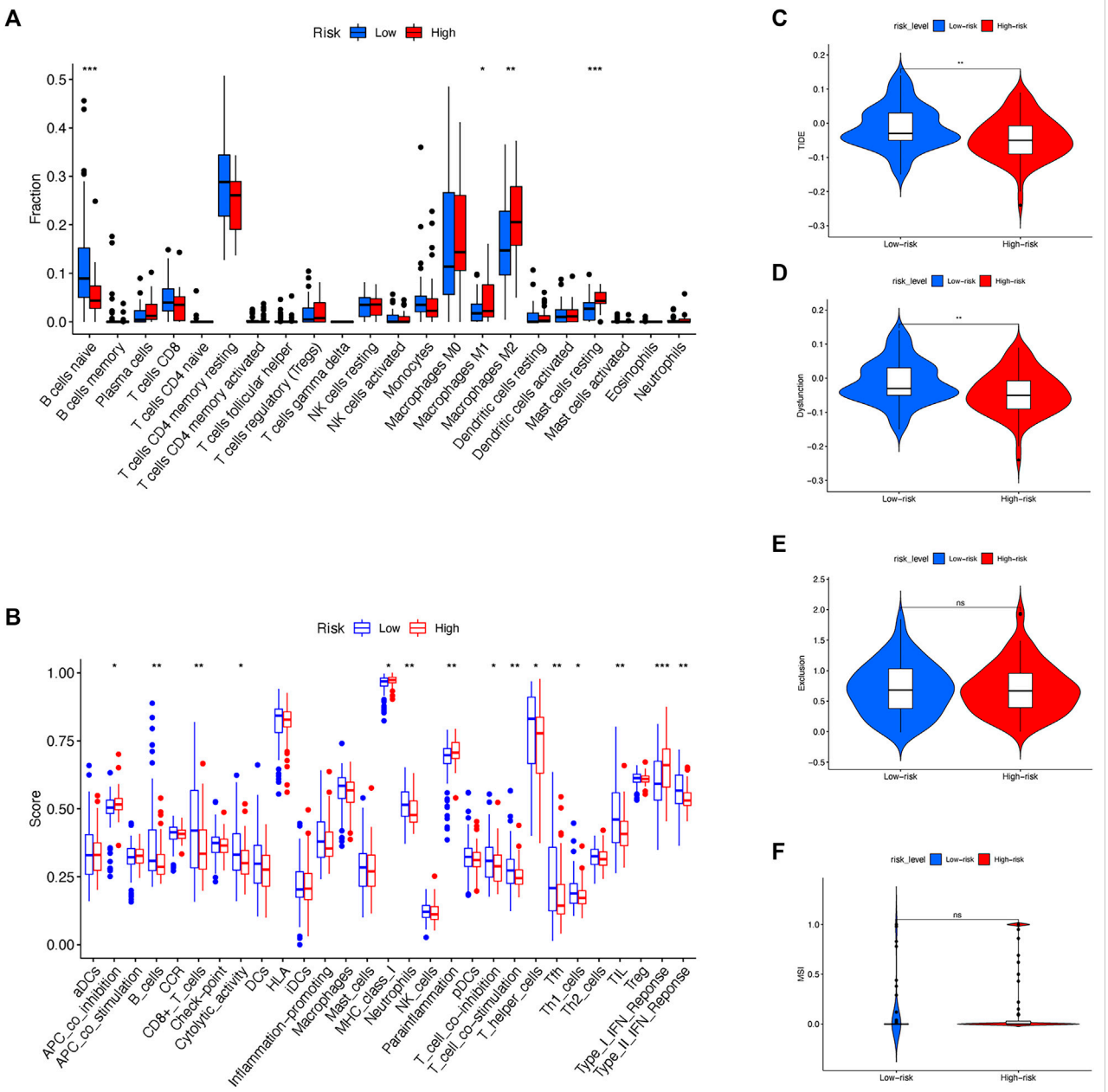


FIGURE 6
Exploration of immune landscape between low- and high-risk subgroups. **(A)** CIBERSORT algorithm evaluated the level of 22 major immune cells between risk subgroups. **(B)** ssGSEA assessed the extent of main immune cells and activations. **(C)** TIDE scoring was compared between risk sub groups. **(D)** Immune dysfunction was compared between risk sub groups. **(E)** Immune exclusion was compared between risk sub groups. **(F)** MSI was compared between risk sub groups.

of immunotherapy in risk subgroups might differ simultaneously. Therefore, the expression and risk profile in the cohort were assessed by the Tumor Immune Dysfunction and Exclusion (TIDE, http://tide.dfci.harvard.edu) and compared between risk subgroups (Figures 6C,D). To note, the high-risk group of the cohort received decreasing TIDE score as well as the extent of immune dysfunction, implying a favorable effect of immunotherapy might occur among patients with a high-risk score according to the molecular classifier. Nonetheless, the level of immune exclusion and MSI were equal between high- and low-risk subgroups (Figures 6E,F).

10.3389/fgene.2022.973444

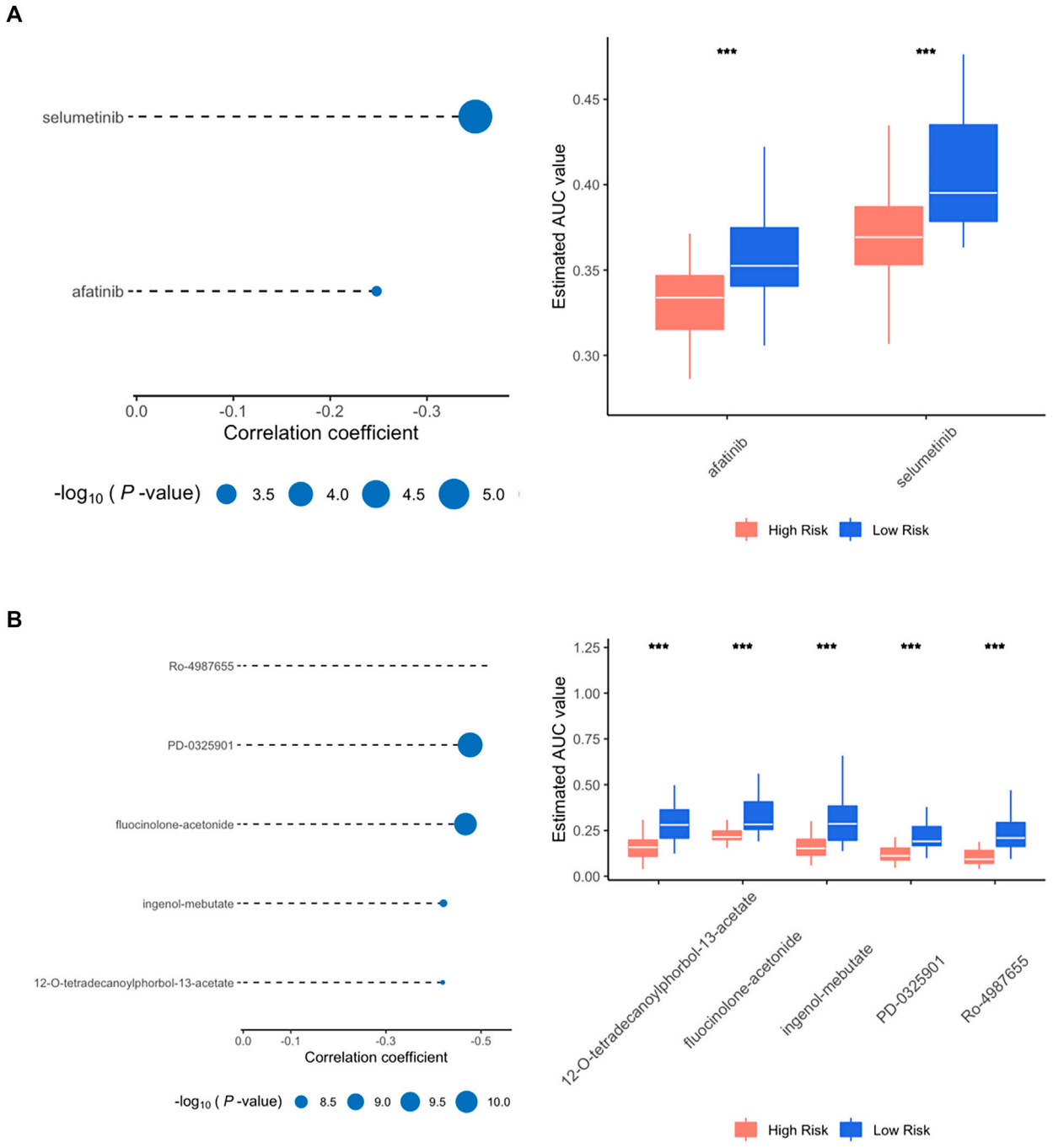


FIGURE 7
Spearman correlation and differential drug response analysis for high risk group. **(A)** Positive drugs identified in CTRP database. **(B)** Positive drugs identified in PRISM database.

Identification of potential therapeutic molecules for high-risk score patients

CTRP and PRISM databases, containing large drug sensitivity profiles in thousands of CCLs, are widely used for estimating drug response. CTRP involves 481 compounds from 860 CCLs while PRISM contains 1,448 compounds from 499 CCLs. According to the results above, patients with a high-risk score from our signature were more likely to have a deteriorative outcome. Thus, potential therapeutic agents with

TABLE 2 Information of candidate agents identified by CTRP, PRISM and CMap databases. Candidate molecules identified by public databases.

Name	Description	Status	CMap score
Selumetinib	MEK inhibitor	Approved	3.28
Afatinib	EGFR inhibitor	Approved	−34.5
Ro-4987655	MEK inhibitor	Phase I	NA
PD-0325901	MEK inhibitor	Phase I/II	2.64
Fluocinolone-acetonide	Glucocorticoid receptor agonist	Approved	0
Ingenol-mebutate	PKC activator	Approved	85.49
12-O-tetradecanoylphorbol-13-acetate	PKC activator	Phase I/II	NA
LY-303511	Casein kinase inhibitor	No data	−98.38
RO-90-7501	Beta amyloid inhibitor	No data	−98.17
TG-101348	FLT3 inhibitor	Phase I/II	−97.64
Baeomycesic-acid	Lipoxygenase inhibitor	No data	−97.24
Pirarubicin	Topoisomerase inhibitor	Phase II	−97.22
PIK-75	DNA protein kinase inhibitor	No data	−96.65

higher drug sensitivity were searched particularly for high-risk patients in two different approaches.

First, drug response profiles were obtained from CTRP and PRISM databases separately. The differential drug response was compared between high- (top 20%) and low-risk (bottom 20%) patients to screen out agents with lower estimated AUC values in high-risk patients (log2FC > 0.05). Subsequently, spearman correlation analysis was applied between the AUC value of specific candidates and the risk score of the patients, in which correlation coefficient $r < -0.3$ for CTRP and $r < -0.5$ for PRISM were considered potentially effective. Altogether, 2 compounds (selumetinib and afatinib) from CTRP and 5 compounds (Ro-4987655, PD-0325901, fluocinolone-acetonide, ingenol-mebutate, 12-O-tetradecanoylphorbol-13-acetate) from PRISM were identified in which all these molecules had a significantly lower value of AUC in high-risk patients compared to low-risk ones (Figures 7A,B).

To further confirm the effectiveness and mechanism of these drug candidates in the cohort, the CMap mode-of-action (MoA) database including nearly 3000 small-molecule compounds was applied. The CMap algorithm compares the expression profile of DElncRNAs in different risk subgroups with the existing response pattern of gene expression for thousands of drugs and molecules in the library. In specific, the CMap score of less than −95 will be considered potentially effective. Positive agents selected *via* CTRP and PRISM were evaluated in the CMap library, respectively. And all agents with CMap score less than −95 were also listed below (Table 2).

Discussion

As a lethal malignancy that causes the second most cancer-related death, pancreatic cancer remains a critical global health challenge. Despite tremendous progress has been made during past decades in understanding the genesis and development of this fatal disease, only a fraction of patients have the opportunity to receive radical or surgical resection. Currently, an increasing number of curative approaches including chemotherapy, targeted therapy, or immunotherapy are available for patients with unresectable or metastatic disease, but with merely reluctant effects. To improve the process of clinical decision-making, physicians and surgeons dedicated years to looking for novel strategies for better diagnosing and prognosis guiding for pancreatic cancer. At the moment, clinicians depend largely on pathological factors such as TNM classification, AJCC tumor staging, or histological grade of the tumor to select proper therapy and forecast the outcome for a certain patient. Nonetheless, novel biomarker panels with high accuracy and specificity are widely accepted as a promising approach that could shed light on improving clinical surveillance and management of pancreatic cancer.

In recent years, the roles of non-coding RNA including micro RNA (miRNA), circular RNA (circRNA) and lncRNA have been increasingly emphasized in tumor biology. Numerous studies have revealed that dysregulated lncRNA participates in processes of carcinogenesis and progression of pancreatic cancer. For instance, Liu has reported that lncRNA NR2F1-AS1 promotes proliferation and invasion of pancreatic cancer by regulating the neighboring NR2F1 gene and activating AKT/mTOR signaling pathway (Liu et al., 2022). Huang has demonstrated that lncRNA LNC00842 prompts the malignancy of pancreatic cancer by preventing acetylate PGC-1α from deacetylation and remodeling the metabolic status of cancer cells (Huang et al., 2021). Additionally, Zheng has announced that lncRNA LINC00673 serves as a tumor suppressor by accelerating the ubiquitination of oncogene PTPN11 *via* binding to miRNA-1231 and competing for the endogenous RNA (ceRNA) mechanism (Zheng et al., 2016). As our knowledge of translating molecular profiling and genetic

defects into novel biomarkers and targets soars, the potentiality of lncRNA in pancreatic cancer is not to be neglected.

The major pursuit of the present study is to construct a lncRNA-based molecular signature that hosts high efficacy in predicting the outcomes of pancreatic cancer. We have noticed that all patients with pancreatic cancer are now treated with one universal norm mainly based on imaging experiments. The lack of corresponding biomarkers and specified measures possibly result in unsatisfactory therapeutic effects. Therefore, another purpose of the study is to develop a tailored treatment for a certain proportion of patients with possible high risk and worse outcomes.

To achieve these goals, the TCGA-PAAD cohort was randomly split and applied for prognostic analysis. After screening out the differentially expressed lncRNAs and prognostic lncRNAs that were Cox-positive, an RSF model based on minimal depth were established to finally identify target lncRNAs and to form a lncRNA based classifier with the highest efficacy and effectiveness. Hence, the cohorts were subsequentially divided into high-risk and low-risk groups according to the median value of the risk scores by the signature. The survival status of the high-risk and low-risk groups was then compared *via* Kaplan-Meier survival analysis. The result indicates a huge gap between the subgroups where individual samples with high-risk scores reveal remarkably worse prognoses than their counterpart in low-risk groups. In addition, the signature exhibits superior efficacy in prognosis anticipation, where the AUC value in ROC analysis exceeds over 0.90 for all 1, 3, and 5 years prediction of OS in the training group and reaches 0.75 to 0.83 overall. Furthermore, the signature was approved by the uni- and multivariate Cox as one independent predictive factor with marked significance. Last but not least, a combined prediction model that includes multiple associated factors was carried out using a nomogram to put the prognosis prediction of pancreatic cancer into practice.

Despite our analysis suggesting that the lncRNA-based signature is of both expressional and prognostic significance, most component lncRNAs consisting it remains unknown in tumor biology. As the champion with the highest coefficient value among all member lncRNAs, the association between CASC8 and the malignant tumor has been marked recently. It was reported that CASC8 promoted the proliferation of retinoblastoma cells *via* manipulating the methylation of miRNA-34a (Yang B. et al., 2020). Besides, inhibiting CASC8 led to decreased development in non-small cell lung cancer and enhanced sensitivity against chemotherapy, implying CASC8 might be a novel target for cancer treatment in the future (Jiang et al., 2021). MIR600HG is another elemental lncRNA with a high coefficient but rather negative than positive. Pieces of the literature suggested that MIR600HG suppressed metastasis and development by targeting oncogenic ALDH1A3 in colorectal cancer (Yao and Li, 2020). Nevertheless, other studies also indicated that MIR600HG induced but not hindered the progression of the same disease, reflecting its complex nature in tumor biology (Huang et al., 2022). Furthermore, the finding that elemental lncRNAs had considerable mutual correlation in expression suggests some of them might have relevant mechanisms. More functional study is in great need to gain a deeper understanding of these lncRNAs which might unravel novel mechanisms in pancreatic development.

Pancreatic cancer is well known for its feature of immune suppression due to the oncogenic drivers (Bear et al., 2020). By far no single-agent immune therapy was proven clinically effective. And immune modulators are jointly applied with other treatments. After decades of dedication, scientists gradually unraveled the pivotal role of the classical oncogene KRAS and the activation of its mutation in pancreatic cancer. Not only as the trigger of carcinogenesis, but the inception of mutant KRAS signaling also orchestrates a complex network of immunosuppression by manipulating the tumor microenvironment (TME) in pancreatic cancer. Evidence shows that the hyperactivation of KRAS prevents both the innate and adaptive immunity by regulating the expression of immune checkpoint CD47 and PD-L1, activating immune suppressive cells, modulating the level of major histocompatibility complex class I (MHC), forming an inconvenient stromal microenvironment and so forth (cancer cell p2). Despite all the disadvantages, developing novel strategies of immunotherapy for pancreatic cancer is still able to catch the public interest. Several approaches have been proposed for future combinatorial treatments such as stimulating the antigen specificity of T cell immunity, increasing the function of effector T cells, and diminishing the immunosuppressive myeloid and stromal cells.

In this study, the assumption was made that risk subgroups might relate to different immune landscapes. Thus, activation of major immune cells and functions were compared by CIBERSORT and ssGSEA. The result revealed that the activation of a series of major immune cells including B naïve cell, $CD8^+$ T cell, T helper cell and neutrophils were modestly down-regulated in the high-risk group, suggesting that the activation of immune cells in tumors of the high-risk patients are possibly paralyzed. Consistently, crucial immune functions such as T cell inhibition, T cell stimulation and IFN response were also found to decline in the high-risk group. Taking together, the enhanced immunosuppression in the high-risk group might be one possible explanation for its notorious outcome.

Of note, as one pivotal component in the tumor microenvironment of pancreatic cancer, the activation of M2 macrophage was found elevated in high-risk patients. The polarization of tumor-associated macrophage (TAM) has been widely accepted as one symbolic event in early and advanced tumorigenesis. Also, several studies confirmed increased M2 deviation of macrophages that promotes tumor behaviors including tumor proliferation, metastasis, and immune escape in pancreatic cancer (Yang S. et al., 2020). Moreover, a meta-analysis containing 1,699 patients with pancreatic cancer concludes that the activity of M2 macrophage is not only closely associated with carcinogenesis, but also has a clear impact on the OS of pancreatic cancer, and thus might be considered a diagnostic and therapeutic target in the future (Yu et al., 2019).

Besides, the gene enrichment and tumor immune analysis unanimously observed an increased interferon-alpha response in the high-risk subgroup. As one of the oldest immune-based therapeutic options for cancer treatment, interferon-alpha is widely used to suppress tumor growth in melanoma, lymphoma, renal carcinoma and so forth. In pancreatic cancer, some clinical trials suggested that chemotherapy based on interferon might improve the overall outcome after surgical resection (Jensen et al., 2014; Ohman et al., 2017). Nonetheless, more efforts are still required to assess the potential of interferon as a first-class approach in a certain population of pancreatic cancer.

In addition, the potential responsibility for the immunotherapy was evaluated by TIDE analysis. Paradoxically, the result indicated that the high-risk group positively correlated with the declined level of TIDE score and immune dysfunction to their low-risk peer, suggesting promising anticipation of therapeutic effect in the high-risk group. To our knowledge, the TIDE algorithm was built on specific tumor types of melanoma and non-small cell lung cancer (NSCLC). The unique characteristic of heterogeneity and immunosuppression in pancreatic cancer could leave the result debatable. In sum, further studies are drastically needed to gain a deeper understanding of immune activity in pancreatic cancer before novel promising therapeutic strategies are to be developed.

Last but not least, potential small molecules that might have therapeutic effects, particularly for high-risk patients were searched *via* CTRP, PRISM and CMap databases. As the only agent identified by CTRP and PRISM with a negative CMap score, Afatinib belongs to the tyrosine kinase inhibitor family and is mainly effective for epidermal growth factor receptor (EGFR) and human epidermal growth factor receptor 2 (HER2). Afatinib, under the commercial name of Gilotrif, has received approval as a first-line treatment for NSCLC. Other indication includes advanced breast cancer with HER2 positive. On-going and complete clinical trials are revealing the potential efficacy of Afatinib in lung cancer other than NSCLC, Head, and Neck squamous cell carcinoma, glioma, and prostate cancer (Molife et al., 2014; Reardon et al., 2015; Hayashi et al., 2022; Kao et al., 2022). Noticeably, two studies exploring the effect of Afatinib on pancreatic cancer acquired only negative results (Haas et al., 2021; van Brummelen et al., 2021). In a phase I study, Afatinib, together with another agent selected from the CTRP database, Selumetinib, was administrated on KRAS-mutated pancreatic cancer (van Brummelen et al., 2021). The result suggested that despite the combination can be used on KRAS-mutated tumors without severe complications, the clinical efficacy was also limited. The other phase II trial concluded that the combination therapy of Afatinib plus gemcitabine did not exhibit a synergistic effect and failed to surpass gemcitabine application alone (Haas et al., 2021). Yet it is still not clear if Afatinib might be more beneficial to certain portions of pancreatic cancer patients.

According to CMap analysis, 5 novel molecules exhibited strong therapeutic potentiality (CMap < −95). LY303511 inhibits the activity of casein kinase 2, which is known to prompt the translation from the G1 to G2 phase and therefore down-regulated cellular proliferation in A549 cells (Kristof et al., 2005). Afterward, LY303511 also increases apoptosis in tumor cells *via* sensitizing TRAIL signaling in HeLa cells (Tucker-Kellogg et al., 2012). However, little is known about the function of LY303511 in pancreatic cancer, and no clinical trial using the agent were carried out so far. TG-101348, a selective JAK antagonist, is one rising star in the antitumor pharmacy. It has been extensively studied in hematology by both *in vitro* and *in vivo* models (Lasho et al., 2008; Wernig et al., 2008; Lasho et al., 2010). Running clinical studies are investigating the role of this promising molecule in leukemia and myeloproliferative neoplasm. Interestingly, a recent study showed that TG-101348 ameliorates hepatic fibro-genesis by inhibiting the TGF-β relied upon activation of hepatic fibroblasts, indicating a broad future prospective of this novel molecule (Akcora et al., 2019).

Conclusion

To. conclude, this study generated a novel lncRNA-based signature based on a random forest model to predict the overall survival of pancreatic cancer. The efficacy and effectiveness of the signature were evaluated individually and combined with other clinical characteristics. This lncRNA panel, either alone or in combined efforts with other clinical factors, can provide a novel strategy for prognosis anticipation and clinical decision of pancreatic cancer. High-risk patients entitled to the signature tend to have considerably worse outcomes than their low-risk counterparts. And enhanced immunosuppression might be one reasonable explanation. Last but not least, potential therapeutic molecules were excavated from public databases. The result turns out Afatinib, LY-303511, TG-101348 and Pirarubicin could be candidates that are particularly effective for patients with a high-risk score. But before understanding all this, more efforts on validation and mechanistic exploration of these genes and drugs are still in great demand.

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

DM and YY collected the data, performed informatic analysis, and interpreted the results. XD and BS designed the study. DM wrote the draft of the manuscript. QC and FY provided help in analyzing data and revising the manuscript.

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.

Supplementary Material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.973444/full#supplementary-material

References

Akcora, B. O., Dathathri, E., Ortiz-Perez, A., Gabriel, A. V., Storm, G., Prakash, J., et al. (2019). TG101348, a selective JAK2 antagonist, ameliorates hepatic fibrogenesis *in vivo*. *FASEB J.* 33, 9466–9475. doi:10.1096/fj.201900215RR

Bear, A. S., Vonderheide, R. H., and O'Hara, M. H. (2020). Challenges and opportunities for pancreatic cancer immunotherapy. *Cancer Cell* 38, 788–802. doi:10.1016/j.ccell.2020.08.004

Christenson, E. S., Jaffee, E., and Azad, N. S. (2020). Current and emerging therapies for patients with advanced pancreatic ductal adenocarcinoma: a bright future. *Lancet. Oncol.* 21, e135–e145. doi:10.1016/S1470-2045(19)30795-8

Haas, M., Waldschmidt, D. T., Stahl, M., Reinacher-Schick, A., Freiberg-Richter, J., Fischer von Weikersthal, L., et al. (2021). Afatinib plus gemcitabine versus gemcitabine alone as first-line treatment of metastatic pancreatic cancer: the randomised, open-label phase II ACCEPT study of the Arbeitsgemeinschaft Internistische Onkologie with an integrated analysis of the 'burden of therapy' method. *Eur. J. Cancer* 146, 95–106. doi:10.1016/j.ejca.2020.12.029

Hayashi, H., Yonesaka, K., Nakamura, A., Fujimoto, D., Azuma, K., Sakata, S., et al. (2022). Alternating therapy with osimertinib and afatinib for treatment-naive patients with EGFR-mutated advanced non-small cell lung cancer: a single-group, open-label phase 2 trial (WJOG10818L). *Lung Cancer* 168, 38–45. doi:10.1016/j.lungcan.2022.04.004

Huang, X., Pan, L., Zuo, Z., Li, M., Zeng, L., Li, R., et al. (2021). LINC00842 inactivates transcription co-regulator PGC-1α to promote pancreatic cancer malignancy through metabolic remodelling. *Nat. Commun.* 12, 3830. doi:10.1038/s41467-021-23904-4

Huang, Y., Wang, L., and Liu, D. (2022). lncRNA MIR600HG induces the proliferation and invasion of colorectal cancer cells via regulating miR-144-3p/KIF3A. *Int. Immunopharmacol.* 108, 108686. doi:10.1016/j.intimp.2022.108686

Jensen, E. H., Armstrong, L., Lee, C., Tuttle, T. M., Vickers, S. M., Sielaff, T., et al. (2014). Neoadjuvant interferon-based chemoradiation for borderline resectable and locally advanced pancreas cancer: a phase II pilot study. *HPB Oxf.* 16, 131–139. doi:10.1111/hpb.12086

Jiang, X., Guan, J., Xu, Y., Ren, H., Jiang, J., Wudu, M., et al. (2021). Silencing of CASC8 inhibits non-small cell lung cancer cells function and promotes sensitivity to osimertinib via FOXM1. *J. Cancer* 12, 387–396. doi:10.7150/jca.47863

Kao, H. F., Liao, B. C., Huang, Y. L., Huang, H. C., Chen, C. N., Chen, T. C., et al. (2022). Afatinib and pembrolizumab for recurrent or metastatic Head and Neck squamous cell carcinoma (ALPHA study): a phase II study with biomarker analysis. *Clin. Cancer Res.* 28, 1560–1571. doi:10.1158/1078-0432.CCR-21-3025

Kristof, A. S., Pacheco-Rodriguez, G., Schremmer, B., and Moss, J. (2005). LY303511 (2-piperazinyl-8-phenyl-4H-1-benzopyran-4-one) acts via phosphatidylinositol 3-kinase-independent pathways to inhibit cell proliferation via mammalian target of rapamycin (mTOR)- and non-mTOR-dependent mechanisms. *J. Pharmacol. Exp. Ther.* 314, 1134–1143. doi:10.1124/jpet.105.083550

Lasho, T., Tefferi, A., and Pardanani, A. (2010). Inhibition of JAK-STAT signaling by TG101348: a novel mechanism for inhibition of KITD816V-dependent growth in mast cell leukemia cells. *Leukemia* 24, 1378–1380. doi:10.1038/leu.2010.109

Lasho, T. L., Tefferi, A., Hood, J. D., Verstovsek, S., Gilliland, D. G., and Pardanani, A. (2008). TG101348, a JAK2-selective antagonist, inhibits primary hematopoietic cells derived from myeloproliferative disorder patients with JAK2V617F, MPLW515K or JAK2 exon 12 mutations as well as mutation negative patients. *Leukemia* 22, 1790–1792. doi:10.1038/leu.2008.56

Liu, Y., Chen, S., Cai, K., Zheng, D., Zhu, C., Li, L., et al. (2022). Hypoxia-induced long noncoding RNA NR2F1-AS1 maintains pancreatic cancer proliferation, migration, and invasion by activating the NR2F1/AKT/mTOR axis. *Cell Death Dis.* 13, 232. doi:10.1038/s41419-022-04669-0

Mizrahi, J. D., Surana, R., Valle, J. W., and Shroff, R. T. (2020). Pancreatic cancer. *Lancet* 395, 2008–2020. doi:10.1016/S0140-6736(20)30974-0

Molife, L. R., Omlin, A., Jones, R. J., Karavasilis, V., Bloomfield, D., Lumsden, G., et al. (2014). Randomized Phase II trial of nintedanib, afatinib and sequential combination in castration-resistant prostate cancer. *Future Oncol.* 10, 219–231. doi:10.2217/fon.13.250

Ohman, K. A., Liu, J., Linehan, D. C., Tan, M. C., Tan, B. R., Fields, R. C., et al. (2017). Interferon-based chemoradiation followed by gemcitabine for resected pancreatic adenocarcinoma: long-term follow-up. *HPB Oxf.* 19, 449–457. doi:10.1016/j.hpb.2017.01.012

Reardon, D. A., Nabors, L. B., Mason, W. P., Perry, J. R., Shapiro, W., Kavan, P., et al. (2015). Phase I/randomized phase II study of afatinib, an irreversible ErbB family blocker, with or without protracted temozolomide in adults with recurrent glioblastoma. *Neuro. Oncol.* 17, 430–439. doi:10.1093/neuonc/nou160

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

Tucker-Kellogg, L., Shi, Y., White, J. K., and Pervaiz, S. (2012). Reactive oxygen species (ROS) and sensitization to TRAIL-induced apoptosis, in Bayesian network modelling of HeLa cell response to LY303511. *Biochem. Pharmacol.* 84, 1307–1317. doi:10.1016/j.bcp.2012.08.028

van Brummelen, E. M. J., Huijberts, S., van Herpen, C., Desar, I., Opdam, F., van Geel, R., et al. (2021). Phase I study of afatinib and selumetinib in patients with KRAS-mutated colorectal, non-small cell lung, and pancreatic cancer. *Oncologist* 26, 290–e545. doi:10.1002/onco.13631

Wernig, G., Kharas, M. G., Okabe, R., Moore, S. A., Leeman, D. S., Cullen, D. E., et al. (2008). Efficacy of TG101348, a selective JAK2 inhibitor, in treatment of a murine model of JAK2V617F-induced polycythemia vera. *Cancer Cell* 13, 311–320. doi:10.1016/j.ccr.2008.02.009

Wu, M., Li, X., Zhang, T., Liu, Z., and Zhao, Y. (2019). Identification of a nine-gene signature and establishment of a prognostic nomogram predicting overall survival of pancreatic cancer. *Front. Oncol.* 9, 996. doi:10.3389/fonc.2019.00996

Xiao, M., Liang, X., Yan, Z., Chen, J., Zhu, Y., Xie, Y., et al. (2022). A DNA-methylation-driven genes based prognostic signature reveals immune microenvironment in pancreatic cancer. *Front. Immunol.* 13, 803962. doi:10.3389/fimmu.2022.803962

Yang, B., Gu, B., Zhang, J., Xu, L., and Sun, Y. (2020). CASC8 lncRNA promotes the proliferation of retinoblastoma cells through downregulating miR34a methylation. *Cancer Manag. Res.* 12, 13461–13467. doi:10.2147/CMAR.S268380

Yang, S., Liu, Q., and Liao, Q. (2020). Tumor-associated macrophages in pancreatic ductal adenocarcinoma: Origin, polarization, function, and reprogramming. *Front. Cell Dev. Biol.* 8, 607209. doi:10.3389/fcell.2020.607209

Yao, Y., and Li, N. (2020). MIR600HG suppresses metastasis and enhances oxaliplatin chemosensitivity by targeting ALDH1A3 in colorectal cancer. *Biosci. Rep.* 40 (4), BSR20200390. doi:10.1042/BSR20200390

Yu, J., Lang, Q., Zhong, C., Wang, S., and Tian, Y. (2021). Genome-wide identification of autophagy prognostic signature in pancreatic cancer. *Dose. Response.* 19, 15593258211023260. doi:10.1177/15593258211023260

Yu, M., Guan, R., Hong, W., Zhou, Y., Lin, Y., Jin, H., et al. (2019). Prognostic value of tumor-associated macrophages in pancreatic cancer: a meta-analysis. *Cancer Manag. Res.* 11, 4041–4058. doi:10.2147/CMAR.S196951

Zhang, J., Liu, Z., Zhang, Z., Tang, R., Zeng, Y., and Chen, P. (2022). Identification of a glycolysis-related gene signature for predicting pancreatic cancer survival. *J. Gastrointest. Oncol.* 13, 380–399. doi:10.21037/jgo-22-17

Zheng, J., Huang, X., Tan, W., Yu, D., Du, Z., Chang, J., et al. (2016). Pancreatic cancer risk variant in LINC00673 creates a miR-1231 binding site and interferes with PTPN11 degradation. *Nat. Genet.* 48, 747–757. doi:10.1038/ng.3568

Zhu, B., Wu, X., Guo, T., Guan, N., and Liu, Y. (2021). Epidemiological characteristics of pancreatic cancer in China from 1990 to 2019. *Cancer Control.* 28, 10732748211051536. doi:10.1177/10732748211051536

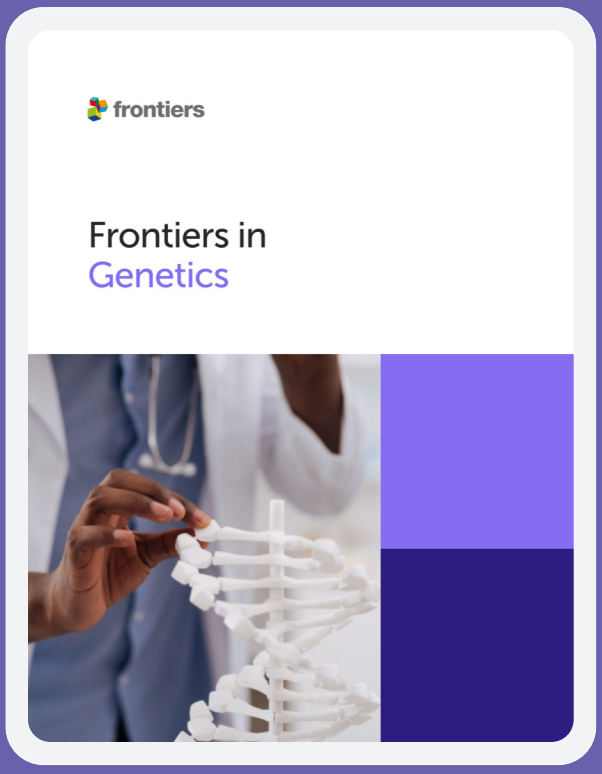
Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

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