frontiers | Research Topics

Recent advances in the molecular genetics and precision medicine of lung carcinoma

Edited by
Mehdi Pirooznia, Dragana Jovanovic, Jonathan N. Bella, Xiaogang Wu and Fan Zhang

Published in
Frontiers in Genetics
Frontiers in Oncology

ISSN 1664-8714
ISBN 978-2-8325-4296-5
DOI 10.3389/978-2-8325-4296-5

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

Recent advances in the molecular genetics and precision medicine of lung carcinoma

Topic editors

Mehdi Pirooznia – Johnson & Johnson, United States
Dragana Jovanovic – University of Belgrade, Serbia
Jonathan N. Bella – BronxCare Health System, United States
Xiaogang Wu – University of Texas MD Anderson Cancer Center, United States
Fan Zhang – University of North Texas Health Science Center, United States

Citation

Pirooznia, M., Jovanovic, D., Bella, J. N., Wu, X., Zhang, F., eds. (2024). *Recent advances in the molecular genetics and precision medicine of lung carcinoma*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4296-5

Table of contents

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 22 January 2024
DOI 10.3389/fgene.2024.1369247

OPEN ACCESS

EDITED AND REVIEWED BY
Anton A. Buzdin,
European Organisation for Research and Treatment of Cancer, Belgium

*CORRESPONDENCE
Mehdi Pirooznia,
pirooznia@gmail.com

RECEIVED 11 January 2024
ACCEPTED 15 January 2024
PUBLISHED 22 January 2024

CITATION
Pirooznia M, Wu X, Bella JN, Zhang F and Jovanovic D (2024), Editorial: Recent advances in the molecular genetics and precision medicine of lung carcinoma.
Front. Genet. 15:1369247.
doi: 10.3389/fgene.2024.1369247

Editorial: Recent advances in the molecular genetics and precision medicine of lung carcinoma

Mehdi Pirooznia[1]*, Xiaogang Wu[2], Jonathan N. Bella[3], Fan Zhang[4] and Dragana Jovanovic[5]

[1]Johnson & Johnson, Washington, DC, United States, [2]The University of Texas MD Anderson Cancer Center, Houston, TX, United States, [3]Icahn School of Medicine at Mount Sinai, New York, NY, United States, [4]University of North Texas Health Science Center, Fort Worth, TX, United States, [5]Policlinic Akta Clinica, Belgrade, Serbia

KEYWORDS

lung cancer, adenocarcinaoma, NSCLC, LUSC (lung squamous cell carcinoma), LUAD (lung adenocarcinoma)

Editorial on the Research Topic
Recent advances in the molecular genetics and precision medicine of lung carcinoma

The goal of this Research Topic is to elucidate the latest research in lung cancer genomics and clinical advances. Lung Carcinoma is the second most common cancer worldwide, with more than 2.2 million cases recorded globally in the year 2020, as well as 1.8 million deaths. It is the most common cancer in men and the second most common cancer in women, with an estimated lifetime risk of developing the disease of 1 in 15 for men and 1 in 17 for women. Whilst being the most prevalent cancer, it is important to note that smoking is the single biggest risk factor for developing lung cancer, with over 70% of cases being estimated to have been caused by tobacco smoke, and excessive smoking potentially leading to as much as a 25-fold increase in the likelihood of the disease developing. Recent advances in sequencing technology, computational approaches, and our biological understanding of lung cancer have revolutionized how we diagnose, prognosticate, and treat lung carcinoma. Genetic studies into this disease have revealed a plethora of information which can be used to combat the cancer, such as novel biomarkers and gene signatures, as well as opening the door to more of a "personalized medicine" approach.

We hope that this Research Topic will inform, inspire, and provide guidance to researchers in the field. Here is an overview of the issues presented in this Research Topic.

The clinical significance of ferroptosis in lung adenocarcinoma (LUAD) was investigated by Ren et al. They constructed a 15-gene prognostic signature predicting overall survival based on ferroptosis-related in The Cancer Genome Atlas (TCGA)-LUAD cohort. Functional analysis revealed that ferroptosis was closely related to cell cycle, cell metabolism, and immune pathways. This signature also acts as a regulator modulating drug resistance, tumor microenvironment infiltration, and cancer stemness. Furthermore, genes in this signature were screened as potential novel immunotherapy biomarkers targets for further research as their biological functions in ferroptosis were consistent with their prognostic significance. In another related study, Chai et al. studied ferroptosis antitumor immune responses in LUAD. They identified 38 ferroptosis-related (FRGs) and 429 immune-related (IRGs) differentially expressed genes between tumor and normal

samples and developed Lasso-penalized Cox regression risk score formulas for them. The correlation between FRGs and IRGs was evaluated using the TIMER database. The outcome indicated that the development of ferroptosis was synergistic with that of anti-tumor immunity which may be useful for future investigation of prognostic value and therapeutic potential related to ferroptosis and tumor immunity in LUAD. Zhang et al. also studied Fanconi anemia group D2 (FANCD2) which is a ferroptosis-related gene crucial for DNA damage repair and negative ferroptosis regulation. They evaluated FANCD2's association with ferroptosis and immune infiltration in LUAD using transcriptome sequencing data, clinical information, and immunohistochemistry data were collected from the TCGA, GEO, and HPA databases, respectively, for three independent cohorts. Univariate and multivariate analyses were used to assess the correlations between FANCD2 expression and overall survival or clinicopathological parameters. Their analysis revealed that FANCD2 expression levels were significantly related to tumor-infiltrating immune cells and their matching gene signatures, including $CD8^+$ T cells, natural killer (NK) cells, dendritic cells (DC), and Th2 cells in cases of LUAD and that FANCD2 is a crucial molecule underlying the synergistic effects of ferroptosis and immunotherapy for Patients with LUAD.

Chen et al. investigated the diagnosis and prognosis roles of DNASE1L3 gene in LUAD. They reported that low DNASE1L3 expression was significantly associated with higher pathological stages, T stages, and poor prognosis in LUAD cohorts and indicated that the mRNA level of DNASE1L3 was positively correlated with the infiltration of various immune cells, immune checkpoints in LUAD, especially with some m6A methylation regulators. Moreover, they reported that DNASE1L3 was positively related to G protein-coupled receptor ligand biding and G alpha (i) signaling events. Another study conducted by Zhao et al. aimed to identify the key intercellular communication-associated genes (ICAGs) in LUAD. They identified prognosis-related ICAGs and developed a risk score by using survival analysis in TCGA-LUAD. Machine learning models were then trained to predict LUAD recurrence based on the selected ICAGs and clinical information as well as comprehensive analyses on ICAGs and tumor microenvironment. The model achieved a remarkable area under receiver operator characteristic curves of 0. 841. DNA replication and cell cycle were significantly enriched by the differentially expressed genes between the high- and the low-risk groups according to their risk scores. The study identified eight key ICAGs in LUAD, which could contribute to patient stratification and act as novel therapeutic targets.

The PD-1/PD-L1 inhibitors, widely used in clinical trial, showed a low response rate and were effective for only a small number of cancer patients. Mao et al. investigated the low response rate of PD-1/PD-L1 inhibitors immunotherapy. They performed ssGSEA and unsupervised clustering analysis to identify clusters according to different immune cell infiltration status, prognosis, and biological action. One cluster showed a better survival rate, higher immune cell infiltration, and immunotherapy effect, with enrichment of a variety of immune active pathways including T and B cell signal receptors. Their analysis suggested that there were six genes with KLRC3 as the core which can efficiently improve immunotherapy responses with greater efficacy and better prognosis.

There is accumulating evidence that long noncoding RNAs (lncRNAs) are playing critical role in predicting the prognosis and immune response in carcinoma including LUAD. Using cooexpression analysis in LUAD patients, Peng et al. identified mitochondrial homeostasis–related lncRNAs (MHRlncRNAs) and mitochondrial homeostasis–related lncRNA signature (MHLncSig) as an independent predictive factor of prognosis. The study further investigated the underlying tumor microenvironment, tumor mutation burden, and immune landscape behind different risk groups and suggested that MHLncSig may be a promising tool for predicting the prognosis and guiding individualized treatment in LUAD. The lncRNAs are also investigated in lung squamous cell carcinoma (LUSC) by Huan et al. They screened the TCGA-LUSC samples and constructed a 5-lncRNA-based signature, combining lncRNA and traditional clinical indicators for prognosis prediction. The signature was significantly related to chemotherapy response, especially in cisplatin, vinorelbine, and paclitaxel. The enrichment analysis indicated that co-expression mRNAs of the 5 lncRNAs were mainly focused on RNA splicing, DNA replication, and protein serine/threonine kinase activity. Functional assays also demonstrated regulated invasion, migration, proliferation, and programmed death *in vitro*. They concluded that the 5-lncRNA-based signature has a good performance in predicting immune characteristics and prognosis of LUSC patients. On another related study, Zhang et al. investigated the interactions between ferroptosis and lncRNAs for LUSC, and its impact on tumor immune microenvironment. They identified a ferroptosis-related lncRNAs signature (FerRLSig) for LUSC prognosis and evaluated its correlation to tumor immune evasion. Based on the FerRLSig stratification, the high-risk group demonstrated significantly higher immune infiltration, as well as more severe T cell dysfunction and immune evasion, which might ultimately lead to the resistance to current immune checkpoint inhibitors. Tetik Vardarlı et al. evaluated whether lncRNAs obtained from exhaled breath condensate (EBC) samples play a role in the occurrence of metastasis in the diagnosis and follow-up of patients with advanced LUAD. They observed lncRNAs HOTAIR, PVT1, NEAT1, and MALAT1 expression levels were significantly higher than those in healthy controls. They proposed that EBC can be used as an innovative and easily reproducible approach for predicting the development of metastases, molecular diagnosis, and follow-up of lung cancer.

Tang and Guo studied dysregulation of the ubiquitin-proteasome system (UPS) that can lead to instability in the cell cycle and may act as a crucial factor in both tumorigenesis and tumor progression. They retrospectively evaluated a total of 703 LUAD patients through multivariate Cox and Lasso regression analyses, and developed an eight-UPSG signature, including ARIH2, FBXO9, KRT8, MYLIP, PSMD2, RNF180, TRIM28, and UBE2V2 as a novel prognostic predictor for LUAD. On another study, Wang et al. employed ssGSA, WGCNA, univariable and LASSO Cox regression analyses of LUAD patients to build a fatty acid-related risk score (FARS) model. They established a nomogram based on the FARS and other clinicopathological features and used ROC and calibration plots were used to validate the prediction accuracy. The tumor microenvironment (TME) of patients with high and low FARS was compared. And a total of 38 genes were identified to be

independently related to the survival outcome and put into a FARS model. High FARS patients exhibited significantly worse OS. Patients with high FARS can potentially benefit more from anti-PD1/PDL1 immunotherapy.

Zhao and Lu investigated clinical significance and prognostic and immunological function of 26S proteasome non-ATPase regulatory subunit 2 (PSMD2) in TCGA-LUAD. They conducted PSMD2-related immune infiltration analysis and tumor-Immune system interaction database (TISIDB) to verify the correlation between PSMD2 expression and tumor-infiltrating lymphocytes (TILs). They observed that both mRNA and protein expression of PSMD2 were significantly elevated in LUAD. High expression of PSMD2 was significantly correlated with high T stage, lymph node metastases, and TNM stage. The genetic mutation of PSMD2 was also correlated with poor overall survival, disease-specific survival, and progression-free survival in LUAD. Functional enrichment suggested PSMD2 expression was involved in the cell cycle, RNA transport, and cellular senescence suggesting that PSMD2 is a potential biomarker for poor prognosis and immune therapeutic target in LUAD.

Abnormal chromosome segregation regulators (CSRs) is a common hallmark of cancer. Li et al. studied the specific predictive value of it in LUAD. They performed unsupervised clustering to identify the distinguishing clusters in LUAD patients based on the expression of CSRs. The immune environment estimation, including immune cell infiltration, HLA family genes, immune checkpoint genes, and tumor immune dysfunction and exclusion (TIDE), was then assessed between the clusters. Cell cycle and chromosome segregation regulated genes were enriched in cluster 1, while metabolism regulated genes were enriched in cluster 2. Patients in cluster 2 illustrated a higher score of immune, stroma, and HLA family components, while those in cluster 1 had higher scores of TIDES and immune checkpoint genes. They concluded that the CSRs were correlated with the poor prognosis and the possible immunotherapy resistance in LUAD.

Pyroptosis is a highly inflammatory mode of regulated cell death and essential for the remodeling of tumor immune microenvironment and suppression of tumor development. Zhang and Liu studied pyroptosis-related gene polymorphisms in non-small cell lung cancer (NSCLC). They genotyped six SNPs in the GSDMB, GSDMC, and AIM2 in 650 NSCLC cases and 650 healthy controls using a MassARRAY platform. Their findings showed associations with risk of NSCLC and provided new insights into the roles of pyroptosis-related genes as new factors for assessing the risk of development of NSCLC. In another study, Zhou et al. aimed to elucidate the relationship between genomic alteration and pseudoprogression (PsPD)/hyperprogressive disease (HPD) in immunotherapy-treated advanced NSCLC and to provide clinical evidence for distinguishing between PsPD and HPD. They selected patients with advanced NSCLC who were treated with anti-PD1 and performed Whole blood next-generation sequencing analysis (NGS). They reported the gene mutation profiling of PsPD and HPD differed before treatment. They indicated true disease progression in patients with HPD and suggesting dynamic whole-genome mutation profiling can distinguish PsPD from HPD more effectively and it can potentially be used as a novel method for guiding clinical immune treatment.

Author contributions

MP: Writing–original draft, Writing–review and editing. XW: Writing–original draft, Writing–review and editing. JB: Writing–original draft, Writing–review and editing. FZ: Writing–original draft, Writing–review and editing. DJ: Writing–original draft, Writing–review and editing.

Acknowledgments

We thank the reviewers and editorial team for support in developing this Research Topic.

Conflict of interest

Author MP was employed by the company Johnson & Johnson.

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

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

frontiers in Genetics

ORIGINAL RESEARCH
published: 09 August 2021
doi: 10.3389/fgene.2021.672904

Ferroptosis-Related Genes in Lung Adenocarcinoma: Prognostic Signature and Immune, Drug Resistance, Mutation Analysis

Ziyuan Ren[1†], Minghui Hu[2†], Zhonglin Wang[3], Junpeng Ge[4], Xiaoyan Zhou[2], Guoming Zhang[1] and Hongying Zheng[2]*

[1] Basic Medical School, Cheeloo College of Medicine, Shandong University, Jinan, China, [2] Clinical Lab, The Affiliated Hospital of Qingdao University, Qingdao, China, [3] Department of Mathematics, University of California, Irvine, Irvine, CA, United States, [4] Department of Biology Engineering, Shandong Jianzhu University, Jinan, China

OPEN ACCESS

Edited by:
Raghunath Chatterjee, Indian Statistical Institute, India

Reviewed by:
Arindam Datta, National Institute on Aging, National Institutes of Health (NIH), United States
Nitai Bhattacharyya, Institute of Post Graduate Medical Education and Research (IPGMER), India

***Correspondence:**
Hongying Zheng
zhy20210101@163.com

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

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

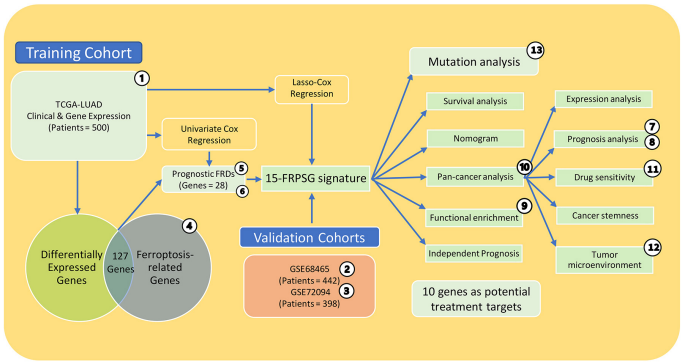
Received: 01 March 2021
Accepted: 20 May 2021
Published: 09 August 2021

Citation:
Ren Z, Hu M, Wang Z, Ge J, Zhou X, Zhang G and Zheng H (2021) Ferroptosis-Related Genes in Lung Adenocarcinoma: Prognostic Signature and Immune, Drug Resistance, Mutation Analysis. Front. Genet. 12:672904. doi: 10.3389/fgene.2021.672904

It is reported that ferroptosis has close relation with tumorigenesis and drug resistance. However, the clinical significance of ferroptosis in lung adenocarcinoma (LUAD) remains elusive, and the potential targets for ferroptosis-based treatment are limited. In this study, we constructed a 15-gene prognostic signature predicting overall survival based on ferroptosis-related genes (ferroptosis driver genes *VDAC2*, *GLS2*, *FLT3*, *TLR4*, *PHKG2*, phosphogluconate dehydrogenase (*PGD*), *PANX1*, *KRAS*, *PEBP1*, *ALOX15*, and *ALOX12B*, and suppressor genes *ACSL3*, *CISD1*, *FANCD2*, and *SLC3A2*) in The Cancer Genome Atlas (TCGA)-LUAD cohort. The signature's predictive ability was validated in the GSE68465 and GSE72094 cohorts by survival analysis and independent prognostic analysis with clinical features. Nomograms were provided for clinical reference. Functional analysis revealed that ferroptosis was closely related to cell cycle, cell metabolism, and immune pathways. Pan-cancer analysis comprehensively analyzed these 15 genes in 33 cancer types, indicating that the heterogeneity of 15 genes was evident across different cancer types. Besides, these genes were critical regulators modulating drug resistance, tumor microenvironment infiltration, and cancer stemness. Then, we screened 10 genes (*TLR4*, *PHKG2*, *PEBP1*, *GLS2*, *FLT3*, *ALOX15*, *ACSL3*, *CISD1*, *FANCD2*, and *SLC3A2*) as potential targets for further research because their biological functions in ferroptosis were consistent with their prognostic significance. Somatic mutation and copy number variation analysis revealed that the alteration rates of *KRAS*, *PGD*, and *ALOX15* were more than 1% and significantly associated with overall survival in LUAD. Moreover, the expression of *KRAS* and *PGD* was positively related to tumor mutation burden, indicating that *KRAS* and *PGD* could serve as novel biomarkers for predicting immunotherapy response rate. Our study identified and validated a ferroptosis-related gene signature for LUAD, provided a 10-gene set for future research, and screened *KRAS* and *PGD* as potential novel immunotherapy biomarkers.

Keywords: ferroptosis, lung adenocarcinoma, prognosis, tumor immune microenvironment, drug resistance, somatic mutation, copy number variation, somatic mutation

Ferroptosis-related Genes in Lung Adenocarcinoma: Prognostic signature and immune, drug resistance, mutation analysis



Authors

Ziyuan Ren, Minghui Hu, Zhonglin Wang, Junpeng Ge
Xiaoyan Zhou, Guoming Zhang, **Hongying Zheng**

Correspondence

Prof. Hongying Zheng
Email: zhy20210101@163.com

Conclusion

To conclude, our study established a ferroptosis-related gene signature for LUAD, and identified 15 FRPSGs. These genes had a close relationship with tumor immunity, drug resistance, tumor microenvironment (TME) infiltration, and cancer stemness. With further checking the mutation status of these genes, we provided new insights for ferroptosis in tumor behavior.

1. **Training Cohort, TCGA-LUAD**. RNA-sequencing. Including 535 tumor samples and 59 normal samples (https://portal.gdc.cancer.gov/repository)
2. **Validation Cohort1, GSE68465**. Microarray. Including 443 tumor and 19 normal samples, were retrieved from the GEO database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE68465)
3. **Validation Cohort2, GSE72094**. Microarray. including 443 tumor and 19 normal samples, were retrieved from the GEO database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE68465)
4. **Ferroptosis database, FerrDb**. Download the FRGs. 173 FRGs (108 driven genes and 69 suppressors genes, 4 were overlapped) (http://www.zhounan.org/ferrdb/index.html)
5. **Online PPI database, STRING** (Search Tool for the Retrieval of Interacting Genes) (version 11.0), was used to plot the PPI network based on prognostic FRDEGs (https://string-db.org/)
6. **Open-source software platform, Cytoscape**, with its plug-in "cytohubba," for visualizing complex networks, was utilized to extract "hub genes" by calculating the hub degree of each prognostic FRDEGs
7. **Online bioinformatics tool, GEPIA** (Gene Expression Profiling Interactive Analysis). was used to perform survival analysis of these FRPSGs based on TCGA and GTEx data (http://gepia.cancer-pku.cn/)
8. **Online database, Prognoscan**. Used to validate the survival analysis (http://dna00.bio.kyutech.ac.jp/PrognoScan/index.html)
9. **Functional analysis.** R packages "limma" was utilized to extract DEGs between high- and low-risk groups.
   GO and KEGGs enrichment were performed by R package "clusterProfiler".
   ssGSEA was utilized to evaluate immune score by R package "gsva."
   Mann-Whitney test was used to detect the difference between 2 groups.
10. **TCGA pan-cancer** data was conducted, which was downloaded from Xena platform (https://xenabrowser.net/datapages/)
11. **Open-access database, NCI** (National Cancer Institute) was also performed Drug sensitivity analysis of the FRPSGs (https://discover.nci.nih.gov/cellminer/)
12. **ESTIMATE** data (Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression) was performed in TME analysis to calculate immune and stromal score (http://bioinformatics.mdanderson.org/estimate/)
13. **R package "RCircos"** to visualize the location on a chromosome and the most common CNV status of the 15 FRPSGs.
    Online bioinformatics tool, cBioPortal to explore the alteration rate (including somatic mutation and CNV), type, and site. Survival analysis by Kaplan-Meier curve and log-rank test. TMB status visualized by radar plots using the R package "ggplot". (https://www.cbioportal.org/)

GRAPHICAL ABSTRACT | Ferroptosis genes in lung adenocarcinoma.

INTRODUCTION

Lung cancer would cause 2.1 million new cases and 1.8 million deaths in 2018, as estimated (Bray et al., 2018). Lung adenocarcinoma (LUAD), one of the non-small cell lung carcinomas (NSCLCs), comprises ∼50% of lung cancer reported cases and shows a continuous upward trend (Imielinski et al., 2012; Kinoshita et al., 2016). Because of the occult nature of early LUAD, most patients are diagnosed with advanced stage or distant metastasis has occurred, losing the opportunity for surgery (Kocher et al., 2015; Siegel et al., 2018). Although some novel regimens such as immune therapy were developed, the 5-year survival rate varies from 4 to 17% (Hirsch et al., 2017). Therefore, exploring novel prognostic signatures and treatment targets will be beneficial.

Ferroptosis is a newly discovered programmed cell death form, promoting reactive oxygen species (ROS) accumulation by glutathione (GSH) consumption and glutathione peroxidase 4 (GPX4) inactivation, which results in subsequent cell death (Xie et al., 2016; Stockwell et al., 2017). Latest studies have demonstrated that ferroptosis impacts cell metabolism, redox state, degenerative diseases, and ischemic reperfusion injury (Stockwell et al., 2017). As for cancers, first, ferroptosis has close relation with the malignant progression of various cancer types (Torti et al., 2018; Mou et al., 2019). The genes that regulate iron metabolisms, such as *NFS1* (NFS1 cysteine desulfurase) (Alvarez et al., 2017) and *CISD1* (CDGSH iron–sulfur domain 1) (Yuan et al., 2016), may become novel targets for tumor-targeted therapy. Second, ferroptosis can reverse the patients' chemoradiotherapy resistance. Roh et al. found that cisplatin can promote head and neck cancer cell ferroptosis without significant necrosis and apoptosis changes (Roh et al., 2016). Lei et al. (2020) reported that cancer cells were more sensitive to radiotherapy after the inhibition of SLC7A11 and GPX4. Third, tumor immunotherapy has been extensively developed, especially the appearance of immune checkpoint inhibitors. Wang et al. (2019b) discovered that after implementing the immune checkpoint inhibitor, the ferroptosis-specific lipid peroxidation was significantly increased. After blocking the ferroptosis process, the sensitivity to immunotherapy was significantly reduced, indicating that the combination of immunotherapy and inducing ferroptosis treatment is promising (Wang et al., 2019b).

The inhibition of ferroptosis in lung cancer has gained meaningful evidence. Compared with other tissues, lung tissue is in an environment with high oxygen concentration (Valavanidis et al., 2013). This unique environment makes lung tissues need to withstand much oxidative pressure (Sunnetcioglu et al., 2016; Zalewska-Ziob et al., 2019). Therefore, to avoid the facilitated and enhanced ferroptosis during the transformation process, lung cancer cells take various measures to increase the induction threshold of ferroptosis, which leads to the low occurrence of ferroptosis, including upregulation of GSH synthesis (Ji et al., 2018), reductions in iron (Kukulj et al., 2010), inhibition of

lipid synthesis (Jiang et al., 2017), compensating the missing enzymatic system GPX4 (Bersuker et al., 2019), and some others (Wang et al., 2019a). Novel regimens for lung cancer based on promoting ferroptosis have also been proposed. Villalpando-Rodriguez confirmed combined use of siramesine and lapatinib could induce ferroptosis in lung cancer cells by inhibiting heme oxygenase-1 (*HO-1*) expression (Villalpando-Rodriguez et al., 2019). However, the clinical significance of ferroptosis in LUAD remains elusive, and the potential targets for ferroptosis treatment are limited.

We established a novel ferroptosis-related 15-gene prognostic signature in The Cancer Genome Atlas (TCGA)-LUAD cohort and validated it in the GSE68465 and GSE72094. Survival analysis and independent prognostic analysis acted mutually to examine the reliability of our signature. Then, we constructed our prognostic signature nomogram to provide a clinical reference for clinicians. Furthermore, functional analysis revealed the relationship between ferroptosis and cell cycle, cell metabolism, and immune response. The pan-cancer analysis provided a comprehensive and systematic characterization of these 15 genes in all 33 cancer types. The results showed that these genes had a close relationship with tumor immunity, drug resistance, tumor microenvironment (TME) infiltration, and cancer stemness, providing new insights for ferroptosis in tumor behavior. Among these 15 genes, 10 genes could be potential treatment targets for basic researches in LUAD because their biological functions in the process of ferroptosis were consistent with their prognostic significance. We performed mutation analysis to explore the alteration and function of these 15 genes in tumor development at the level of genomics. To conclude, our study identified and validated a ferroptosis-related gene (FRG) signature for LUAD, provided a 10-gene set for future research, and screened *KRAS* and phosphogluconate dehydrogenase (*PGD*) as potential novel immunotherapy biomarkers.

MATERIALS AND METHODS

Data Collection

Preprocessed RNA-sequencing (RNA-seq) by fragments per kilobase of exon model per million mapped fragments (FPKM) normalization was downloaded from the TCGA database[1] LUAD project, including 535 tumor samples and 59 normal samples. The corresponding clinical data were also retrieved from the LUAD project, including 500 patients [13 excluded due to lack of overall survival (OS) and 2 excluded due to lack of RNA-seq] as train cohort. The normalized series matrixes of two total-RNA microarrays, GSE68465 (443 tumor samples and 19 normal samples)[2] and GSE72094 (442 tumor samples),[3] were retrieved from the Gene Expression Omnibus (GEO) database (Shedden et al., 2008; Schabath et al., 2016). The corresponding clinical data were also retrieved (GSE68465: 442 patients; GSE72094: 398 patients). The RNA-seq of TCGA and series matrixes of GSE68465 and GSE72094 were processed by "log2(normalized gene expression + 1)." All data analyzed in this study were publicly accessible. No ethics committee approval was required. The policies and publication guidelines of the TCGA and GEO databases were strictly followed.

Then, we searched the ferroptosis database, FerrDb,[4] to download the FRGs (Zhou and Bao, 2020). A total of 173 FRGs (108 driven genes and 69 suppressors genes, 4 were overlapped) were retrieved and provided in **Supplementary Table 1**.

Construction and Validation of the Model

We utilized the function "Wilcox.test" of R package "limma" to extract the ferroptosis-related differentially expressed genes (FRDEGs) between tumor and normal tissues by Mann–Whitney test. Then, we executed the univariate Cox regression to detect the prognostic significance of all FRGs based on OS in the TCGA cohort. Then, the FRDEGs and prognostic FRGs were intersected. The overlapped genes were named prognostic FRDEGs. An online protein–protein interaction (PPI) database, STRING (Search Tool for the Retrieval of Interacting Genes) (version 11.0), was used to plot the PPI network based on prognostic FRDEGs (Szklarczyk et al., 2019). The settings defaulted (network type: full network; active interaction sources: Textmining, Experiments, Database, Co-expression, Neighborhood, Gene Fusion, and Co-occurrence; minimum required interaction score: 0.4). An open-source software platform for visualizing complex networks, Cytoscape with "cytohubba," was utilized to extract "hub genes" by calculating the hub degree of each prognostic FRDEGs (Shannon et al., 2003). "Cytohubba" is a novel Cytoscape plugin, measuring nodes by their positions in the network to predict their significance. Besides, "cytohubba" identified the crucial elements of biological networks. The topological analysis method was "Degree" (Degree(node) = | the collections of its neighbors|) (Chin et al., 2014). We extracted the nodes with a degree higher than three as "hub genes." We calculated the Pearson correlation between each two of the prognostic FRDEGs to explore the intrinsic relationship among these genes (the cutoff Pearson correlation coefficient is 0.2). Among these prognostic FRDEGs, only protein-coding genes remained for further study. The R package "glmnet" was utilized to extract prognostic FRDEGs for further multivariate Cox regression by the LASSO (least absolute shrinkage and selection operator) algorithm to minimize overfitting. The selected prognostic FRDEGs were considered as ferroptosis-related prognostic signature genes (FRPSGs). A multivariate prognostic model was constructed by multivariate Cox regression. An online bioinformatics tool, GEPIA (Gene Expression Profiling Interactive Analysis), was used to perform survival analysis of these FRPSGs based on TCGA and GTEx data (Tang et al., 2017). Another online database, Prognoscan, was used to validate the survival analysis (Mizuno et al., 2009). The risk score was calculated with the formula: Risk score = sum {gene expression [log2(normalized gene expression + 1)] × coefficient}. Then, this model was

[1]https://portal.gdc.cancer.gov/repository, November 5, 2020

[2]https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE68465, December 27, 2020

[3]https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE72094, April 27, 2021

[4]http://www.zhounan.org/ferrdb/index.html

applied in the GSE68465 and GSE72094 cohorts. We stratified patients into high- and low-risk groups by median risk score, respectively. Survival analysis between the two groups was performed by R package "surviminer" using Kaplan–Meier curve and log-rank test. Receiver operator characteristics (ROC) curve was plotted by R package "timeROC" to test the model's predictive ability. Principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) were performed to detect internal characteristics in these two groups by R packages "stats" and "Rtsne." Independent analysis for risk score was performed by uni- and multivariate Cox regression. Gender, age, stage, T (tumor volume stage), N (lymph node stage), smoking history, and risk score were included in the TCGA cohort. M (metastasis stage) was excluded due to so many missing values (NA) > 30%. Gender, age, T, N, and smoking history were included in GSE68465. Gender, age, smoking history, stage, *KRAS*, *EGFR*, and *TP53* mutation were included in GSE72094.

Construction of the Nomogram

We also constructed our prognostic gene signature nomogram using the R language. All the prognostic factors included in the multivariate Cox regression of the independent prognosis analysis were included in the nomograms. The calibration curve was plotted to evaluate the fitting and predictive ability of our prognostic model.

Functional Analysis

We hypothesized that the ferroptosis level was different between the high- and low-risk groups. We compared the 15 FRPSG expressions between the two groups to prove this point. To explore the potential biological processes influenced by ferroptosis between the high- and low-risk groups, we utilized the R packages "limma" to extract DEGs between high- and low-risk groups in the three cohorts for functional analysis. For TCGA cohort, function "Wilcox.test" was utilized (Mann–Whitney test) and $|\log2FC| \geq 1$ and adj.P (adjusted P-value) < 0.05 were set as cutoff value. In the GSE68465 cohort, functions "lmFit" and "eBayes" were utilized, and $|\log2FC| \geq 0.5$ and adj.$P < 0.05$ were set as cutoff value. In the GSE72094 cohort, functions "lmFit" and "eBayes" were utilized, and $|\log2FC| \geq 0.9$ and adj.$P < 0.05$ were set as cutoff value. Gene Ontology (GO; Ashburner et al., 2000) and Kyoto Encyclopedia of Genes and Genomes (KEGGs; Kanehisa and Goto, 2000) enrichment were performed by R package "clusterProfiler." Single-sample gene set enrichment analysis (ssGSEA) was utilized to evaluate each group's immune score by R package "gsva." Mann–Whitney test was used to detect the difference between the three groups. The detailed annotation table of ssGSEA is presented in **Supplementary Table 2**.

Pan-Cancer Analysis of the FRPSGs

To comprehensively analyze the expression and function of FRPSGs in all cancer types, pan-cancer analysis based on TCGA pan-cancer data was conducted, which was downloaded from the Xena platform,[5] including RNA-seq, clinical data, and stemness scores (Goldman et al., 2020). A total of 33 cancer types were included (the full names and abbreviations are provided in **Supplementary Table 3**). Expression analysis (Mann–Whitney test) and survival analysis (univariate Cox regression) of the FRPSGs were performed based on TCGA pan-cancer data. We utilized the National Cancer Institute (NCI)-60[6] database to perform the drug sensitivity analysis of the FRPSGs. NCI-60 is an open-access database based on nine cancer types and 60 cancer cell lines, consisting of mRNA expression level and corresponding z scores of cell sensitivity data (GI50) after drug treatment. We calculated the Pearson correlation between each FRPSG expression and the GI50 to explore the association between FRPSGs and drug sensitivity. We selected 262 Food and Drug Administration-approved drugs or drugs that are currently in clinical trials in this drug sensitivity analysis (Zhang et al., 2020). TME analysis was performed by the Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data (ESTIMATE) immune and stromal score downloaded from ESTIMATE[7] (Yoshihara et al., 2013). Furthermore, to validate the FRPSGs' immune function, the six immune subtypes were utilized to test the relation between FRPSGs and immune subtypes by analysis of variance (Thorsson et al., 2018). Malta et al. (2018) established a novel index to evaluate the cancer stemness based on one-class logistic regression machine learning algorithm (OCLR). The OCLR extracts transcriptomic (mRNA expression) and epigenetic (DNA methylation pattern) features originated from non-transformed pluripotent stem cells and their differentiated progeny. This cancer stemness feature has been widely used in the computational biological analysis (Liu et al., 2021; Wang et al., 2021). Given that ferroptosis is a crucial regulator of tumor heterogeneity (Li et al., 2021), we calculated the Spearman correlation between the cancer stemness indexes (RNA stemness score, RNAss; DNA stemness score, DNAss) and each FRPSG to explore whether FRGs influence cancer stemness ($P < 0.05$). Furthermore, we utilized the online bioinformatics tool, GEPIA, to explore the correlation between the ferroptosis suppressor genes in FRPSGs and two previously discovered cancer stemness cell biomarkers [CD133 (Bertolini et al., 2009) and CD44 (Leung et al., 2010)] to validate the cancer stemness analysis. The gene expression was normalized by a housekeeper gene (GAPDH) (Schmittgen and Livak, 2008).

Somatic Mutation and Copy Number Variation Analysis

To fundamentally understand the ferroptosis process during tumor development, we performed somatic mutation and copy number variation (CNV) analysis to detect the alterations of the 15 FRPSGs in LUAD. We utilized the R package "RCircos" to visualize the location on a chromosome and the most common CNV status of the 15 FRPSGs. We utilized an online bioinformatics tool, cBioPortal,[8] to explore the alteration rate (including somatic mutation and CNV), type, and site in LUAD datasets. Genes with an alteration rate of more than 1% remained

[5]https://xenabrowser.net/datapages/, December 24, 2020

[6]https://discover.nci.nih.gov/cellminer/, April 24, 2021

[7]http://bioinformatics.mdanderson.org/estimate/

[8]https://www.cbioportal.org/

for further research. Survival analysis between the unaltered group and altered group was performed using cBioPortal by Kaplan–Meier curve and log-rank test. Genes with a statistically significant prognostic value remained for further research. We explored the association between gene expression and alteration status using cBioPortal. Tumor mutation burden (TMB) has been discovered by a previous retrospective study that showed a high correlation with response to immune checkpoint blockade therapy (Marabelle et al., 2020). We calculated the Spearman correlation between TMB status and gene expression in pan-cancer data to explore novel biomarkers for predicting immune therapy response rate. TMB status was retrieved from cBioPortal (February 6, 2021). A radar plot was presented to visualize the results using the R package "ggplot."

Statistical Analysis

The descriptive analysis of each cohort was performed by Chi-square test, Somer's d test, and independent-samples *t*-test in SPSS 26.0 (International Business Machines Corporation, Armonk, NY, United States). All the statistics, except descriptive analysis, were conducted by R language (version 4.0.3) (R Core Team, 2013). All the adj.*P* [or false rate discovery (FDR)] were adjusted by Benjamini–Hochberg (BH). Adj.$P < 0.05$ was considered as statistically significant.

RESULTS

Identification of the Prognostic FRDEGs

The research flowchart and corresponding online resources were presented in Graphical Abstract. The three cohorts' basic characteristics are summarized in **Supplementary Table 4**. Among all 173 FRGs, 127 (74.0%) were FRDEGs between 535 tumor samples and 59 normal samples (80 were upregulated in tumor samples; 47 were downregulated in tumor samples). The heatmap of the FRDEGs is shown in **Figure 1A**. We performed the univariate Cox regression for the 173 FRGs and extracted 31 prognostic FRGs based on OS in the TCGA cohort (**Figure 1B**). Twenty-eight overlapping prognostic FRDEGs in FRDEGs and prognostic FRGs were acquired (**Figure 1C**). We utilized STRING and Cytoscape to detect the intrinsic interactions between these 28 genes. The settings defaulted (network type: full network; active interaction sources: Textmining, Experiments, Database, Co-expression, Neighborhood, Gene Fusion, and Co-occurrence; minimum required interaction score: 0.4). The PPI enrichment *P*-value was 8.09e-07, indicating the network has significantly more interactions than a random protein set. The PPI network (**Figure 1D**) selected six genes, *KRAS*, *NRAS*, *FLT3*, *SLC7A11*, *TLR4*, *GCLC*, as "hub genes" because their "Degrees" calculated by "cytohubba" were higher than 3, indicating these genes were crucial elements of this biological network. We calculated the Pearson correlation analysis between each two of the 28 prognostic FRDEGs to explore the potential interactions among these genes (**Figure 1E**) (the cutoff Pearson correlation coefficient was 0.2). The interrelationship between these 28 genes was complicated. For example, *ACSL3* was positively related to seven genes (*KRAS*, *FLT3*, *NRAS*, *VDAC2*, *GCLC*, *PGD*, and *SLC7A11*) and negatively related to *DPP4*. These findings revealed many interactions between the gene pairs, and the regulatory mechanisms of ferroptosis in LUAD were complex.

Construction of the Prognostic Model

Among the 28 FRDEGs, *LINC00336* was excluded because it is a long non-coding RNA. Multivariate Cox regression of the 27 FRDEGs established a 15-gene signature. The risk score was calculated using the formula: Risk score = 0.063*ACSL3* + 0.224*ALOX12B*-0.011*ALOX15* + 0.155*CISD1* + 0.104*FANCD2*-0.308*FLT3*-0.569*GLS2* + 0.066*KRAS* + 0.118*PANX1*-0.358*PEBP1* + 0.010*PGD*-0.251*PHKG2* + 0.093*SLC3A2*-0.176*TLR4* + 0.012*VDAC2* {the gene symbol means the expression of the gene [log2(normalized gene expression + 1)]}. The median value of risk score (−1.58) divided the 500 patients into two groups (250 in the high-risk group/250 in the low-risk group) (**Figure 2A**). The parameter selection of LASSO is presented in **Supplementary Figure 1**. The basic characteristics of each group are displayed in **Supplementary Table 5**. The high-risk group contained more male cases and a higher tumor stage. The results of the PCA and t-SNE implied these two groups had diverse trends (**Figure 2B**). Survival analysis between the two groups is presented in **Figure 2G**. Notably, the OS in the high-risk group was lower ($P < 0.001$). The ROC curve showed the area under the curve (AUC) reached 0.718, 0.730, and 0.721, respectively, which was acceptable (**Figure 2J**). The survival analysis of these 15 FRPSGs using GEPIA based on OS is presented in **Supplementary Figure 2**. *TLR4*, *PHKG2*, *PEBP1*, *GLS2*, *FLT3*, and *ALOX15* were positively associated with the prognosis, and *VDAC2*, *PGD*, *PANX1*, *KRAS*, *ALOX12B*, *ACSL3*, *CISD1*, *FANCD2*, and *SLC3A2* were negatively associated with the prognosis, among which *VDAC2*, *GLS2*, *FLT3*, *TLR4*, *PGD*, *PANX1*, *PEBP1*, *ACSL3*, *CISD1*, *FANCD2*, and *SLC3A2* were of statistical significance. These findings were validated using the Prognoscan database (**Supplementary Figure 3**). A brief introduction of the 15 FRPSGs is presented in **Supplementary Table 6** (Dixon et al., 2012; Kang et al., 2014; Gao et al., 2015; Song et al., 2016; Yang et al., 2016; Yuan et al., 2016; Wenzel et al., 2017; Chen et al., 2019; Magtanong et al., 2019; Su et al., 2019; Wang et al., 2019b). The expression of 15 FRPSGs in different LUAD pathological subtypes is presented in **Supplementary Figure 4**. The expression of 15 FRPSGs was similar in each pathological subtype, except *FANCD2*. *FANCD2* was lowly expressed in micropapillary carcinoma and highly expressed in solid carcinoma, indicating *FANCD2* could participate in pathological changes in LUAD. We also compared the expression of 15 FRPSGs in different LUAD stages (**Supplementary Figure 5**). We found that the three ferroptosis suppressor genes (*ACSL3*, *CISD1*, and *FANCD2*) were increased, and two ferroptosis driver genes (*FLT3* and *PHKG2*) were decreased in higher-stage LUAD. However, some ferroptosis driver genes were increased in higher-stage LUAD. We hypothesized that these genes might involve other regulatory pathways of tumor behavior. For example, PGD is highly expressed in stage IV LUAD. In addition, many previous

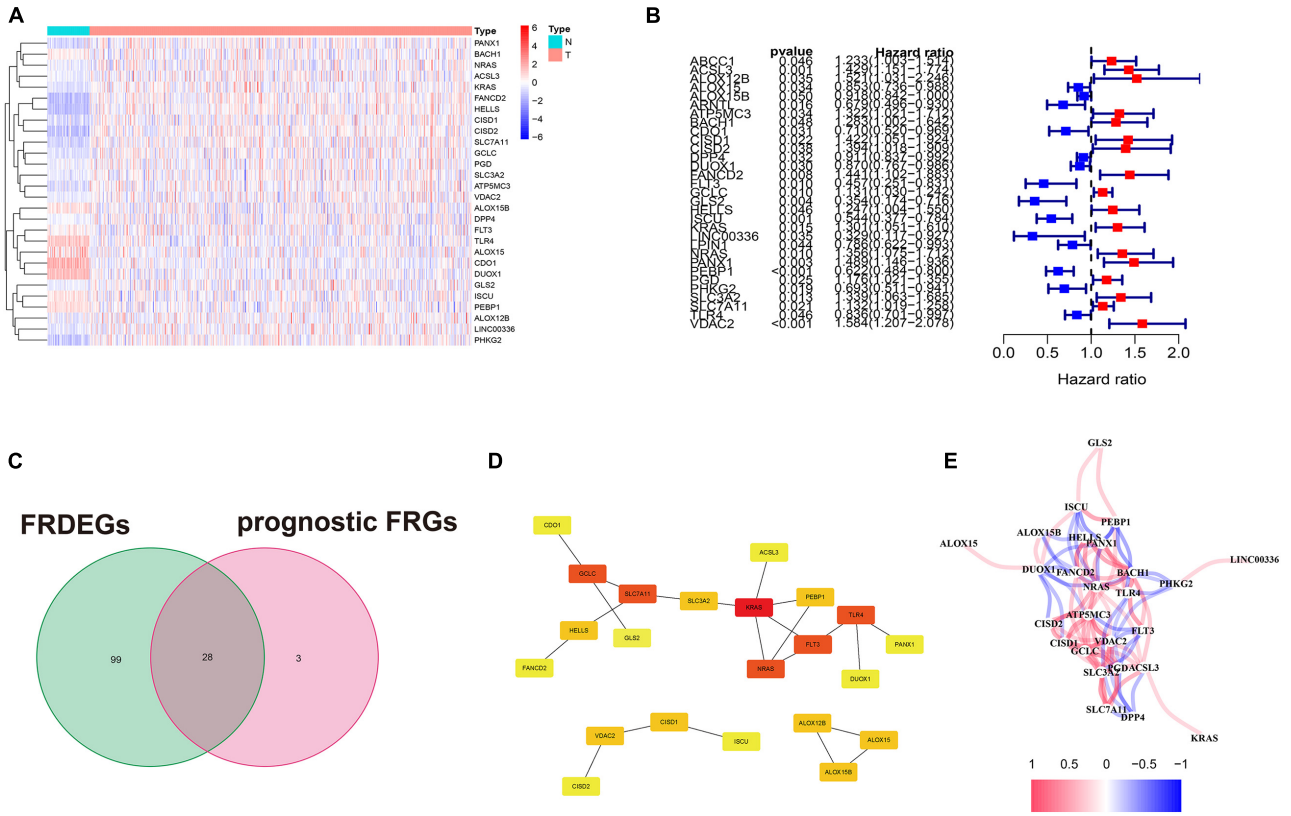


FIGURE 1 | Extraction of prognostic FRDEGs. **(A)** The heatmap of FRDEGs between normal and tumor samples. **(B)** The forest plot of univariate Cox regression. **(C)** The intersection of FRDEGs and prognostic FRGs by univariate Cox regression. **(D)** PPI network of the 28 prognostic FRDEGs; the shade indicates the "hub" degree. **(E)** The correlation among the 28 prognostic FRDEGs.

studies have demonstrated that overexpression of PGD could lead to cancer metastasis and poor prognosis through diverse mechanisms (Bechard et al., 2018; Sarfraz et al., 2020).

Validation on the GSE68465 and GSE72094 Cohorts

The acquired model was applied to the GSE68465 and GSE72094 cohorts. In GSE68465, 442 patients were divided into two groups (-6.88) (**Figure 2C**). The basic characteristics are displayed in **Supplementary Table 5**. The high-risk group had more males, more smokers, and a higher T/N stage. PCA and t-SNE analyses also implied these two groups have diverse trends (**Figures 2D,F**). OS in the high-risk group was still higher ($P < 0.001$) (**Figure 2H**). AUCs reached 0.630, 0.623, and 0.621 at 1, 2, and 3 years (**Figure 2K**), which were not very prominent, maybe due to the heterogeneity between the TCGA cohort and the GSE68465 cohort (the mean value of OS was 654.5 days in the TCGA and was 1410 days in the GSE68465). In the GSE72094, 398 patients were divided into two groups (-0.735) (**Figure 2E**). OS in the high-risk group was higher ($P < 0.001$) (**Figure 2I**). AUCs reached 0.703, 0.692, and 0.738 at 1, 2, and 3 years, which were very prominent for validation cohorts and almost the same as that in the TCGA cohort, indicating our model is robust and had an excellent predictive ability in similar cohorts [the mean value of OS was 824 days, close to that of the TCGA cohort (654.5 days)] (**Figure 2L**). We attributed the difference among the median risk score of the TCGA, GSE68465, and GSE72094 cohorts to the diversity between the RNA-seqs and microarrays. The median risk scores of GSE68465 (-0.688) and GSE72094 (-0.735) were similar but far from that of the TCGA cohort (-0.158), which supported our hypothesis.

Independent Prognostic Analysis

Because of the clinical data difference among the TCGA, GSE68465, and GSE72094 cohorts, the included prognostic factors for OS in the independent prognostic analysis were diverse. In the TCGA cohort, stage, T, N, age, gender, smoking history, and risk score were included. Univariate Cox regression showed stage, T, N, and the risk score are of significance (**Figure 3A**). Multivariate analysis selected T, N, and the risk score as prognostic factors (**Figure 3B**). In the GSE68465 cohort, T, N, age, gender, smoking history, and the risk score were included. The uni- and multivariate Cox regression showed that T, N, age, and risk score were statically significant (**Figures 3C,D**). In the GSE78094 cohort, gender, age, stage, smoking history, *KRAS*, *EGFR*, and *TP53* mutation were included. The uni- and multivariate Cox regression showed that gender, stage, and risk score

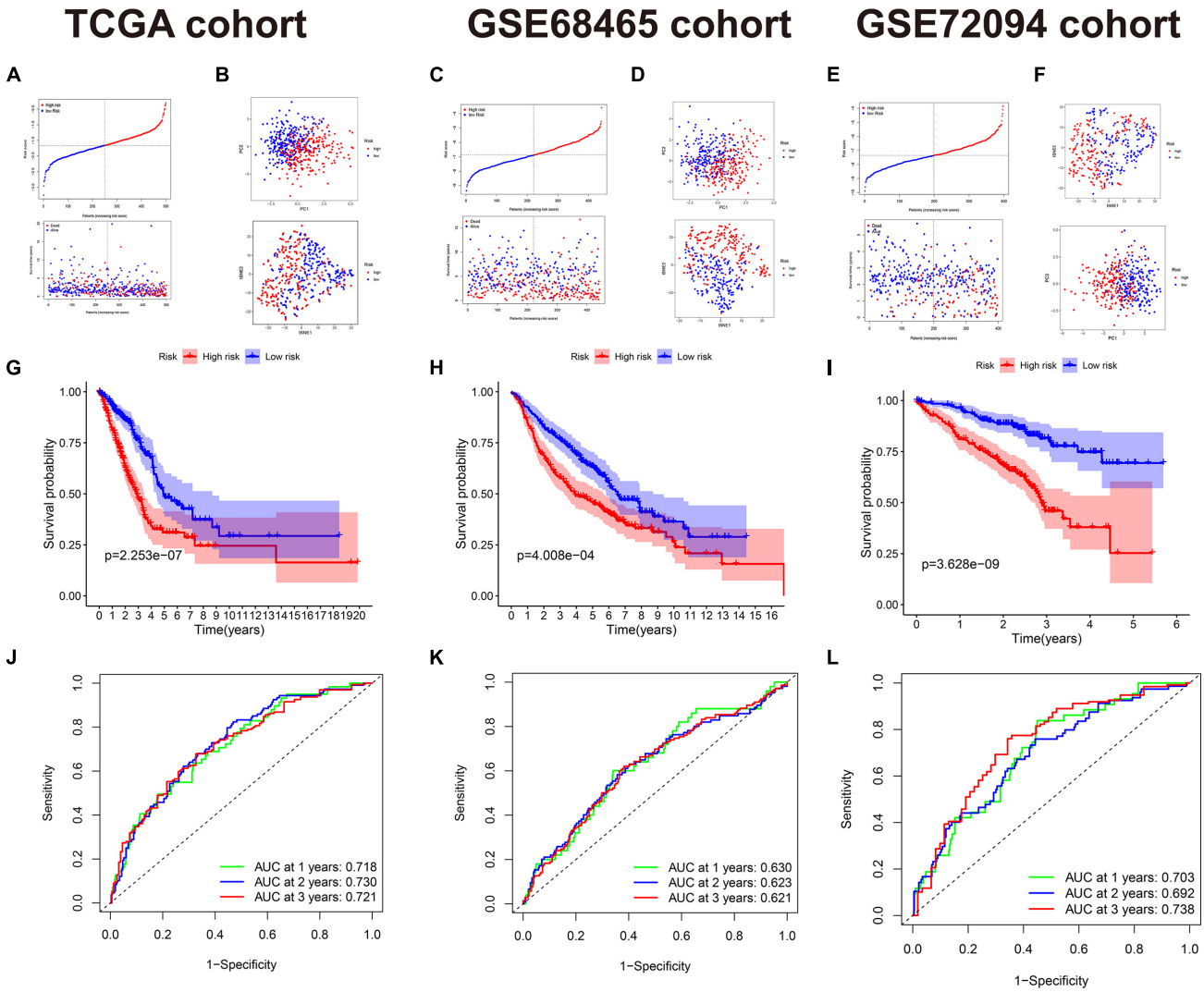


FIGURE 2 | The 15-gene signature in the TCGA, GSE68465, and GSE72094 cohorts. **(A,C,E)** The high- and low-risk groups divided by the median value of risk score and the distribution of the OS and OS status according to the risk score. **(B,D,F)** PCA and t-SNE analyses. PCA and t-SNE indicate the internal diversity between the two groups is obvious. **(G–I)** Survival analysis between the high- and low-risk groups based on OS. **(J–L)** ROC curve of the 15-gene signature model.

were statistically significant independent prognostic factors (**Figures 3E,F**).

Construction of the Nomogram

We also constructed the nomogram of our 15-gene prognostic signature for clinical reference in the TCGA cohort because RNA-seq is the most utilized gene detection method in the clinic. All the prognostic factors selected by univariate Cox regression of the independent analysis were included in the nomograms. The nomograms and calibration curves are presented in **Figures 3G,H**. The calibration curve indicates that the fitting and predictive ability of our model were preferable.

Functional Analyses

Given that ferroptosis is closely related to the malignant progression of various cancer types, we suspected that the ferroptosis level in the high-risk groups was lower than that in the low-risk groups. As shown in **Figures 4A,B**, according to the Mann–Whitney test ($P < 0.05$), all of the four ferroptosis suppressor genes were significantly upregulated in the high-risk groups, and most ferroptosis driver genes (6/9, 66.7%) were significantly downregulated in the high-risk groups (TCGA cohort). In addition, only three ferroptosis driver genes (*KRAS*, *PANX1*, and *PGD*) were significantly upregulated in the high-risk groups. Furthermore, the ferroptosis suppressor genes were also upregulated in the high-risk groups in the GSE68465 (2/2 suppressor genes were upregulated; 5/9 driver genes were downregulated) and GSE72094 (3/3 suppressor genes were upregulated; 5/8 driver genes were downloaded) cohorts (**Supplementary Figures 6A,B**). The detailed list of differentially expressed FRPSGs is provided in **Supplementary Table 7**. Therefore, we suggested that

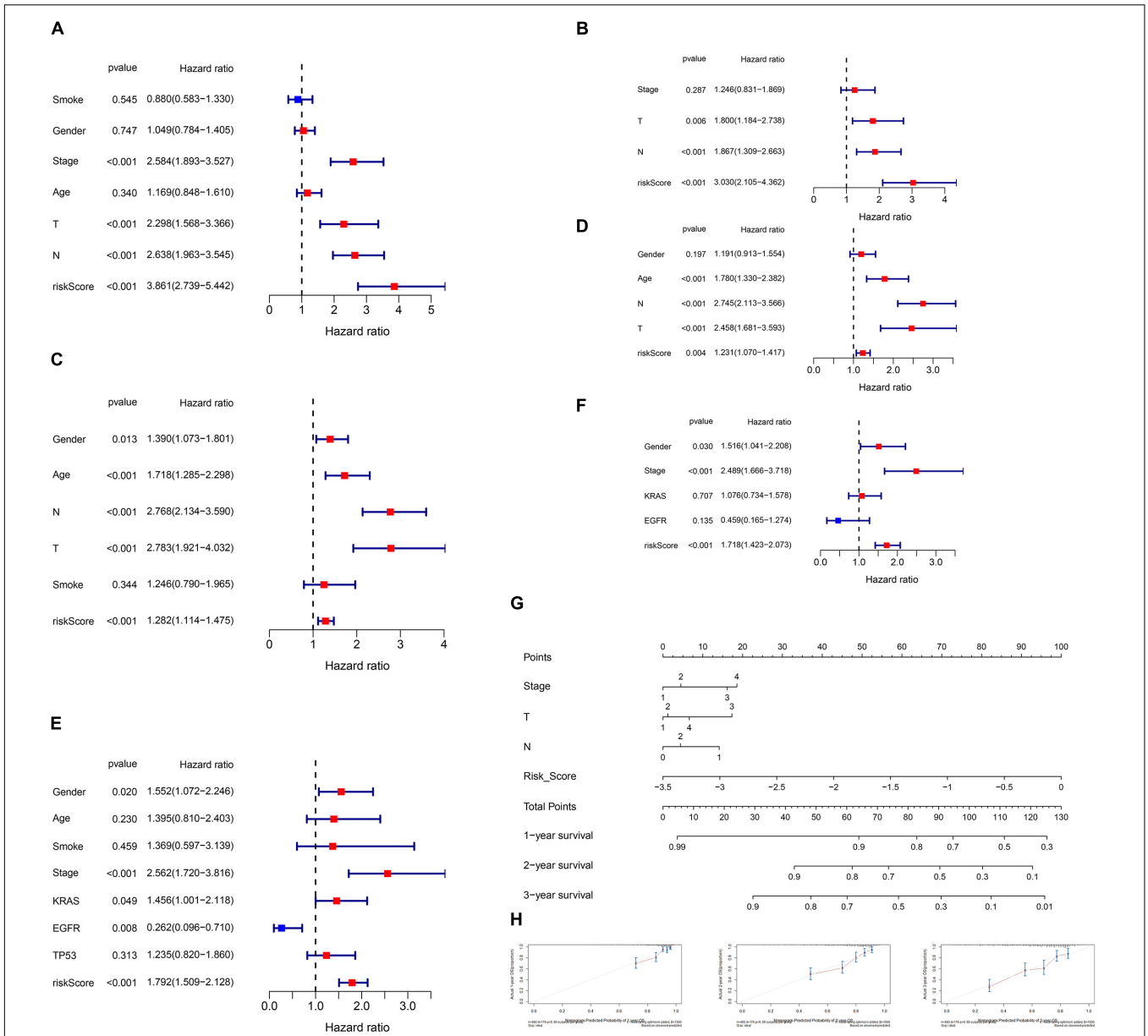


FIGURE 3 | Independent prognostic analysis of the risk score and nomogram with calibration curves of the prognostic factor screened by multivariate Cox regression. **(A)** Univariate Cox regression of the TCGA cohort. **(B)** Multivariate Cox regression of the TCGA cohort. **(C)** Univariate Cox regression of the GSE68465 cohort. **(D)** Multivariate Cox regression of the GSE68465 cohort. Gender, male vs. female; smoke, all smoked vs. never smoked; Stage, high (III/IV) vs. low (I/II); T, high (3/4) vs. low (1/2); N, involved (1/2/3) vs. not involved (0); age, high (>60) vs. low (≤60); risk score, as a continuous variable. **(E)** Univariate Cox regression of the GSE72094 cohort. **(F)** Multivariate Cox regression of the GSE72094 cohort. **(G)** Nomogram of the TCGA cohort [due to only two patients with N3, N3 was considered as NA (missing values)]. **(H)** Calibration curve of the TCGA cohort.

the ferroptosis level was lower in the high-risk groups. To explore the potential biological processes influenced by the ferroptosis process, we utilized the R package "limma" to extract DEGs between the high- and low-risk groups in the three cohorts. A total of 391 DEGs (147 upregulated and 244 downregulated in the high-risk group) in the TCGA cohort, 438 DEGs (285 upregulated and 153 downregulated in the high-risk group) in GSE68465, and 377 DEGs (116 upregulated and 261 downregulated in the high-risk group) between the high- and low-risk groups were extracted. Notably, 20 genes were upregulated and 20 genes were downregulated in the high-risk groups in all three cohorts. A detailed list of DEGs is provided in **Supplementary Table 8**. The GO and KEGG pathway enrichment indicated that the three cohorts' DEGs were massively enriched on cell metabolism and cell cycle, which revealed the association between ferroptosis and these fundamental biological processes (**Figures 4C,D** and **Supplementary Figures 6E–H**). Given that ferroptosis

is associated with tumor immunity and can influence tumor immunotherapy outcomes, it is noteworthy that several immune pathways were enriched. For example, in the TCGA cohort, the enrichment appeared on immune-related pathways (humoral immune response, etc.) in the GO database (**Figure 4C**) and three immune-related pathways (IL-17 signaling pathway, etc.) in the KEGG database (**Figure 4D**). These enriched immune-related pathways indicated that the ferroptosis process involves the development of tumor immune evasion through changing TME. Similarly, in the GSE68465 and GSE72094 cohorts, many immune-related pathways were enriched (**Supplementary Figures 6E–H**).

To elucidate the impact of ferroptosis on TME, we performed ssGSEA to compare the two groups' immune scores. The results showed that dendritic cells (DCs), activated DCs (aDCs), B cells, CD8 + T cells, iDCs, neutrophils, pDCs, T helper cells, Tfh, mast cells, and TIL were significantly reduced in the high-risk group (TCGA) (**Figure 4E**). As for immune function pathways, HLA, T cell coinhibition, checkpoint, cytolytic activity, T cell costimulation, and type II interferon (IFN) response were reduced (TCGA) (**Figure 4F**). These findings strongly suggested that the low-risk group patients' TME had a more potent antitumor effect than the high-risk group, indicating immunotherapy outcomes were better in high-risk group patients, which was consistent with previous reports (Xu et al., 2021).

Pan-Cancer Analysis

The expression of 15 FRPSGs in pan-cancer was different (**Figure 5A**). The expression of *ALOX12B*, *ALOX15*, *FLT3*, and *GLS2* was deficient, and the expression of *PEBP1* was the highest, indicating *PEBP1* could be a promising treatment target. Spearman correlation showed that *PEBP1* and *FANCD2* had the highest correlation. There were some potential interactions between these two genes. Gene differential expression analysis showed that the 15 FRPSGs were differently expressed in tumor tissues compared with corresponding normal tissues. It is noteworthy that the ferroptosis suppressor genes, *SLC3A2*, *FANCD2*, *CISD1*, and *ACSL3*, were upregulated in most cancer types, indicating that ferroptosis was generally inhibited in cancers (**Figure 5B**). As for survival analysis, the same gene had different prognostic significance in distinct cancer types. For example, *VDAC2* was negatively associated with OS in low-grade glioma (LGG) and positively in pheochromocytoma and paraganglioma (PCPG) (**Figure 5C**). It is reported that ferroptosis is a critical regulator in drug resistance. Here, we utilized the NCI-60 database to explore the association between FRPSGs and drug sensitivity using Pearson correlation. The top 16 gene–drug pairs ranked by Pearson correlation coefficient are displayed in **Figure 5D**. The whole list of the drug sensitivity analysis results is in **Supplementary Table 9**. We found that the ferroptosis driver gene, *PEBP1*, was positively related to chemotherapy sensitivity, consistent with previous reports. Interestingly, the ferroptosis suppressor gene, *ACSL3*, was negatively related to three drug sensitivity, among which AP-26113 is a potent and selective ALK and ROS1 inhibitor for lymphoma, everolimus is an mechanistic target of rapamycin (mTOR) kinase inhibitor for renal cell cancer and some other tumors, and nelfinavir is a protease inhibitor for HIV infection. These findings demonstrated that ferroptosis was involved in some tumor-targeted therapy.

Immune analysis of the 15 FRPSGs was performed. Six immune subtypes were defined as C1 (wound healing), C2 (INF-r dominant), C3 (inflammatory), C4 (lymphocyte depleted), C5 (immunologically quiet), and C6 (TGFβ dominant). As shown in **Figure 6A**, all the 15 FRPSGs expressed differently in diverse immune subtypes ($P < 0.001$) for each cancer type. Moreover, TME analysis showed the 15 FRPSGs were related to immune and stromal cell infiltration in most cancer types (**Figure 6B**). These findings were consistent with functional analysis, which further proved that the ferroptosis process involves the change of TME and tumor immunity. Finally, we performed the cancer stemness feature analysis. A stem cell approximative appearance can be gained via cancer growth progress; simultaneously, a differentiated phenotype is eliminated. Based on RNAss (mRNA expression) and DNAss (DNA-methylation pattern), the four ferroptosis suppressor genes, *ACSL3*, *CISD1*, *FANCD2*, and *SLC3A2*, increased the tumor's stem cell-like features, which indicated that the ferroptosis could cause tumor heterogeneity, leading to tumor progression in most cancer type (**Figure 6C**). Specifically, in LUAD, the four genes were positively associated with RNAss and DNAss ($P < 0.05$), especially *FANCD2* (**Supplementary Figure 7**). The Spearman correlation between *FANCD2* expression and RNAss reached 0.56 ($P < 2.2e-16$), indicating *FANCD2* plays a critical role in promoting tumor heterogeneity. We further validated these findings by exploring the association between the four ferroptosis suppressor genes with two previously discovered LUAD stemness cell biomarkers, CD133 and CD44. As shown in **Supplementary Figure 8**, the results show that the four ferroptosis suppressor genes were all positively associated with CD133 and CD44 ($P < 0.05$), which were consistent with the results of cancer stemness feature analysis. In summary, we identified that the four ferroptosis suppressor genes, *ACSL3*, *CISD1*, *FANCD2*, and *SLC3A2*, could increase the cancer stemness, indicating that the lower the ferroptosis level in lung cancer cells, the higher the tumor heterogeneity.

Somatic Mutation and CNV Analysis

To fundamentally understand the ferroptosis process during tumor development, we performed mutation analysis of the 15 FRPSGs. The location on chromosome and CNV status of the 15 FRPSGs are presented in **Figure 7J**. Four ferroptosis driver genes *FLT3*, *ALOX12B*, *ALOX15*, and *VDAC2*, had a low CNV, and four ferroptosis suppressor genes *ACSL3*, *CISD1*, *FANCD2*, and *SLC3A2*, had a high CNV, indicating that ferroptosis was inhibited in tumor development. Besides, the generally accepted oncogene *KRAS* had a high CNV. Among the 15 FRPSGs, the alteration status of three genes, *ALOX15*, *KRAS*, and *PGD*, was significantly related to OS analysis (**Figures 7A–C**, $P = 1.026e-3$, 6.699e-3, 0.0211), with an alteration rate of 1.2, 27, and 1.2% (**Figures 7G–I**). Besides, the alterations of *AOX15*, *KRAS*, and *PGD* mainly focused on domain lipoxygenase, Ras, and 6PGD, indicating that these domains played a crucial role in ferroptosis

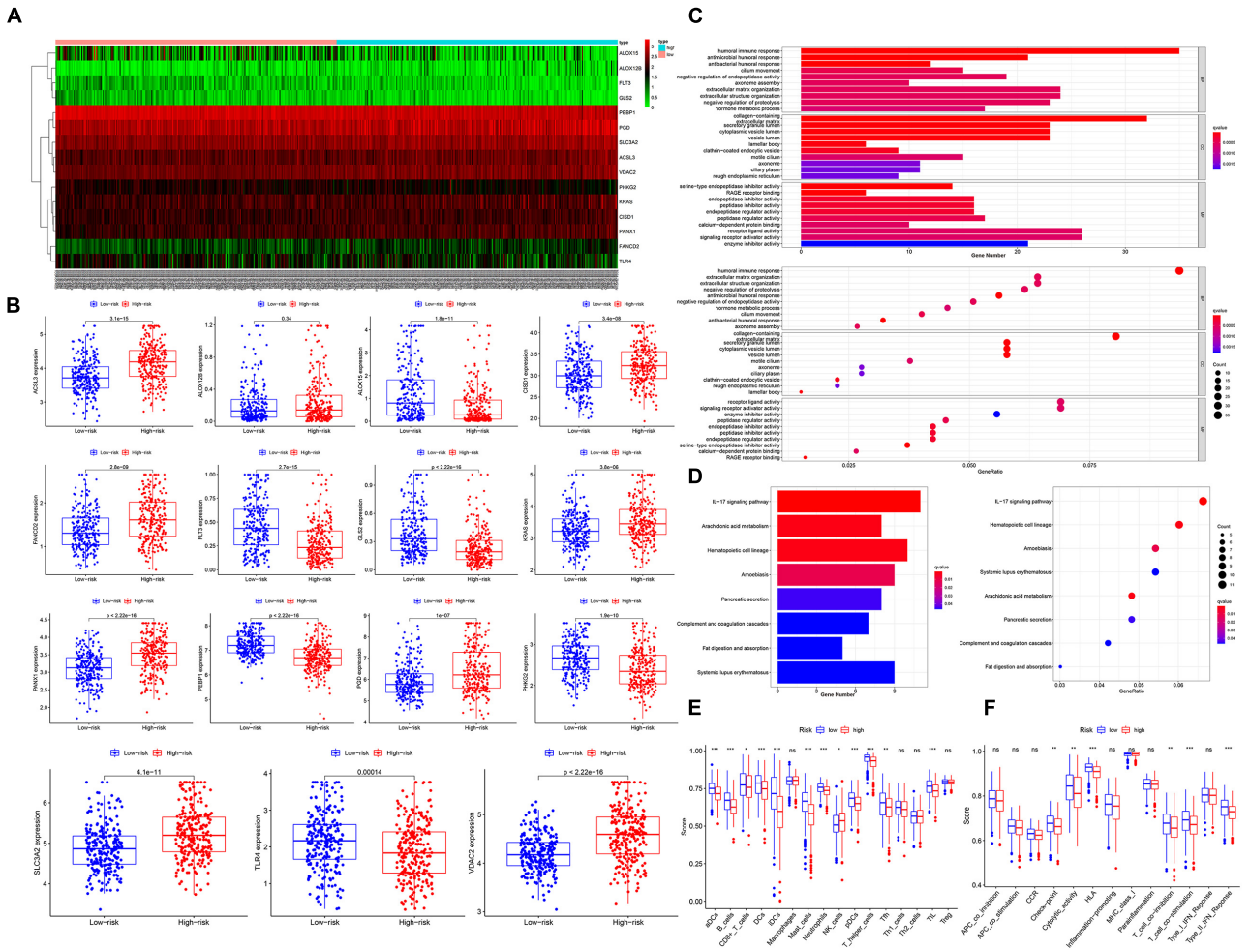


FIGURE 4 | Functional analysis between the high- and low-risk groups. **(A)** The heatmap of differentially expressed genes between the high- and low-risk groups. **(B)** The box plot of comparison of differentially expressed genes in high and low risk. **(C)** The bar plot and bubble plot of GO and KEGG in the TCGA cohort, GSE68465 cohort, and GSE72094 cohort. **(D)** The immune cell infiltrating score and immune function score in the TCGA cohort **(E)** and GSE68465 cohort **(F)**. q-value, adj.*P*-value; ns, not significant; $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

and tumor progression (**Figures 7D–F**). The alteration type of the three genes is shown in **Figures 7G–I**. Missense mutation was the most common mutation type. Splice mutation occurred in *ALOX15* alteration, fusion mutation occurred in *KRAS* alteration, and truncating mutation occurred in *PGD* mutation. The rate of these three mutation types was low. CNV was rare in *ALOX15* alteration but common in *KRAS* and *PGD* alteration, mainly consisting of amplification and deep deletion. **Figures 8A–F** showed the association between gene expression and alteration status in adenocarcinoma using cBioPortal and TMB of *ALOX15*, *KRAS*, and *PGD* in pan-cancer. The association between gene expression and CNV for *KRAS* and *PGD* was statistically significant (by Spearman correlation, $P = 9.05e\text{-}76, 2.153e\text{-}5$) with correlations of 0.61 and 0.16 (**Figures 8B,C**). Comparison of gene expression between wild type and mutation type for *ALOX15*, *KRAS*, and *PGD* indicated mutation of these three genes could increase their expression (**Figures 8D–F**). The most common mutation type was missense for three genes. Besides, splice mutation occurred in *ALOX15,* and fusion mutation occurred in *KRAS*. The association between gene expression and TMB status in pan-cancer for *ALOX15*, *KRAS*, and *PGD* is shown in **Figures 8G–I**. The expression of *KRAS* and *PGD* was positively associated with TMB in LUAD ($P < 0.001$), and the Spearman correlation reached around 0.2, indicating that *KRAS* and *PGD* played a crucial in maintaining genomic integrity. Given that TMB is positively related to immune therapy response rate, we suppose that *KRAS* and *PGD* could serve as novel biomarkers for predicting immunotherapy response rate.

DISCUSSION

Among the 15 FRPSGs, most ferroptosis driver genes were associated with a better prognosis (6/11, 54.5%), and all the ferroptosis suppressor genes (4/4, 100%) were associated with a worse prognosis, which indicates that the ferroptosis process in

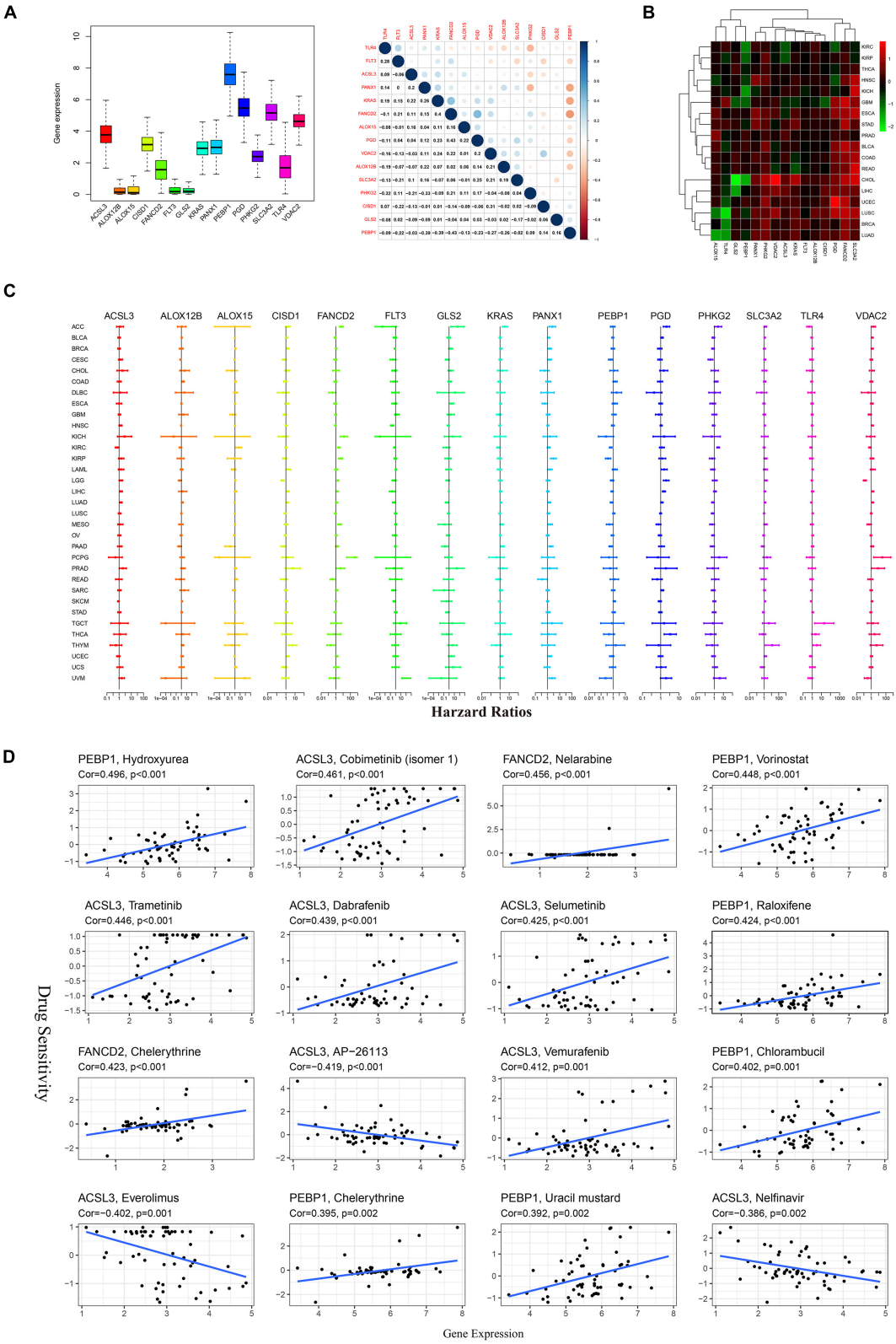


FIGURE 5 | Pan-cancer analysis of 15 FRPSGs. **(A)** The expression level of FRPSGs and correlation of gene expression among FRPSGs. **(B)** Heatmap of FRPSGs between tumor and normal samples in each cancer type. **(C)** Forest plots for hazard ratios of FRPSGs for each cancer type. **(D)** Scatter plots for the association between FRPSG expression and drug sensitivity (top 16 ranked by *P*-value).

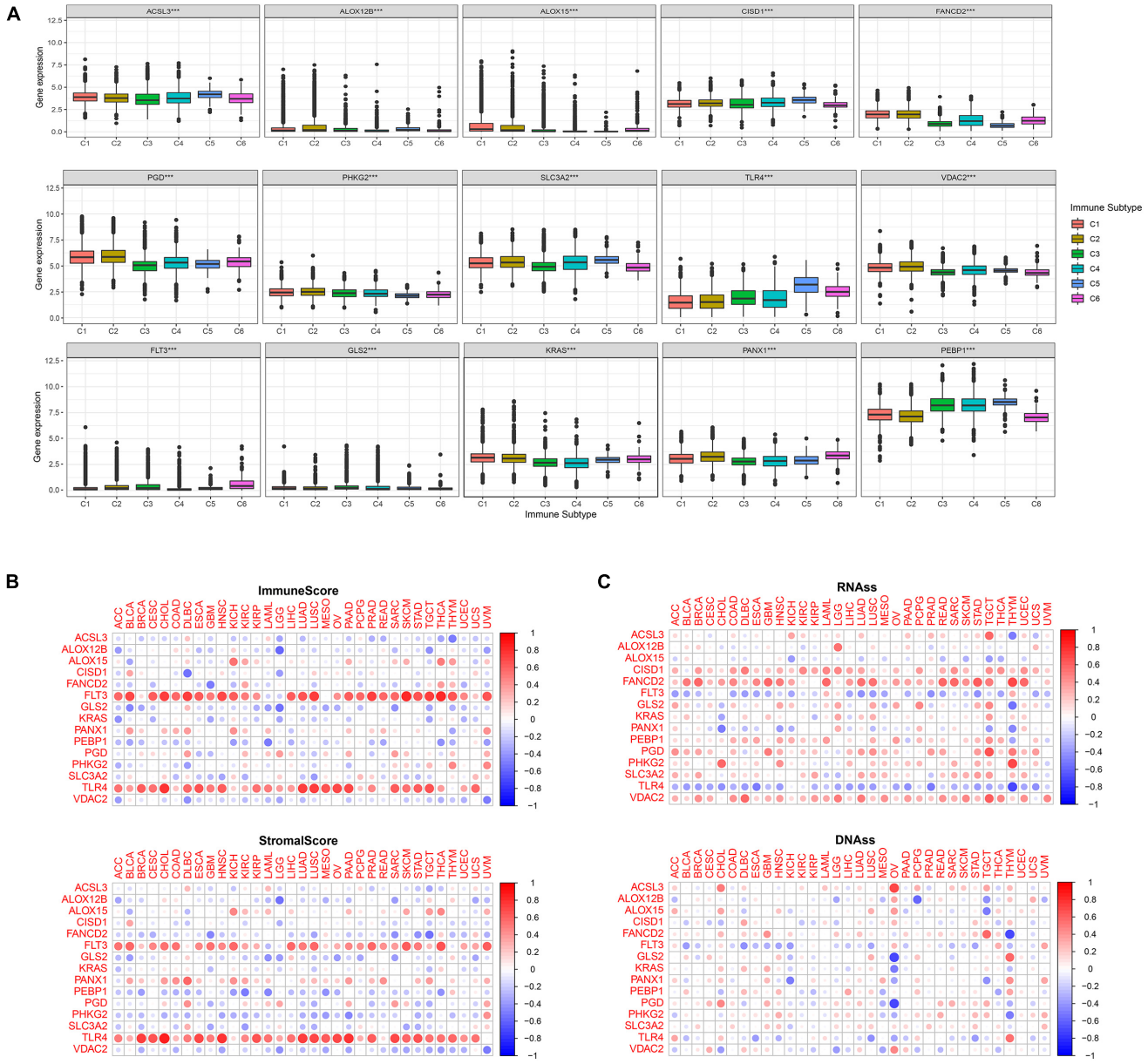


FIGURE 6 | Immune subtype analysis, TME analysis, and cancer stemness features. **(A)** The association between FRPSGs and immune subtype for each cancer type. **(B)** The association between ESTIMATE immune score, stromal score, and FRPSGs expression. **(C)** The association between cancer stemness features and FRPSGs expression based on both mRNA expression (RNAss) and DNA methylation patterns (DNAss).

LUAD inhibits the tumor development as previously reported (Stockwell et al., 2017; Liang et al., 2019). We suggested these 10 genes (*TLR4*, *PHKG2*, *PEBP1*, *GLS2*, *FLT3*, *ALOX15*, *ACSL3*, *CISD1*, *FANCD2*, and *SLC3A2*) could be potential targets for further research because their biological functions in ferroptosis were consistent with their prognostic significance.

Lipid metabolism, iron metabolism, (anti)oxidative metabolism, and energy metabolism are the main regulatory pathways of ferroptosis. In our signature, lipid metabolism involves *ACSL3*, *ALOX15*, *ALOX12B*, and *PEBP1*. ACSL3 is a principal modulator of ferroptosis by activating the exogenous monounsaturated fatty acids, which subsequently replace polyunsaturated fatty acids (PUFAs) and block lipid ROS accumulation (Tang et al., 2018; Magtanong et al., 2019). Lipoxygenase (ALOX) is the critical enzyme for lipid peroxidation and ferroptosis through the arachidonic acid (AA) oxygenation. ALOX15 and the scaffold protein PEBP1 form a complex, converting PUFAs (mainly AA) to 15-HPETEs (hydroperoxyeicosatetraenoic acids), and this leads to ferroptosis in several diseases (Anthonymuthu et al., 2018; Zhao et al., 2020). In particular, ALOX15 has long been regarded as a critical mediator and thus a prominent indicator in the seton

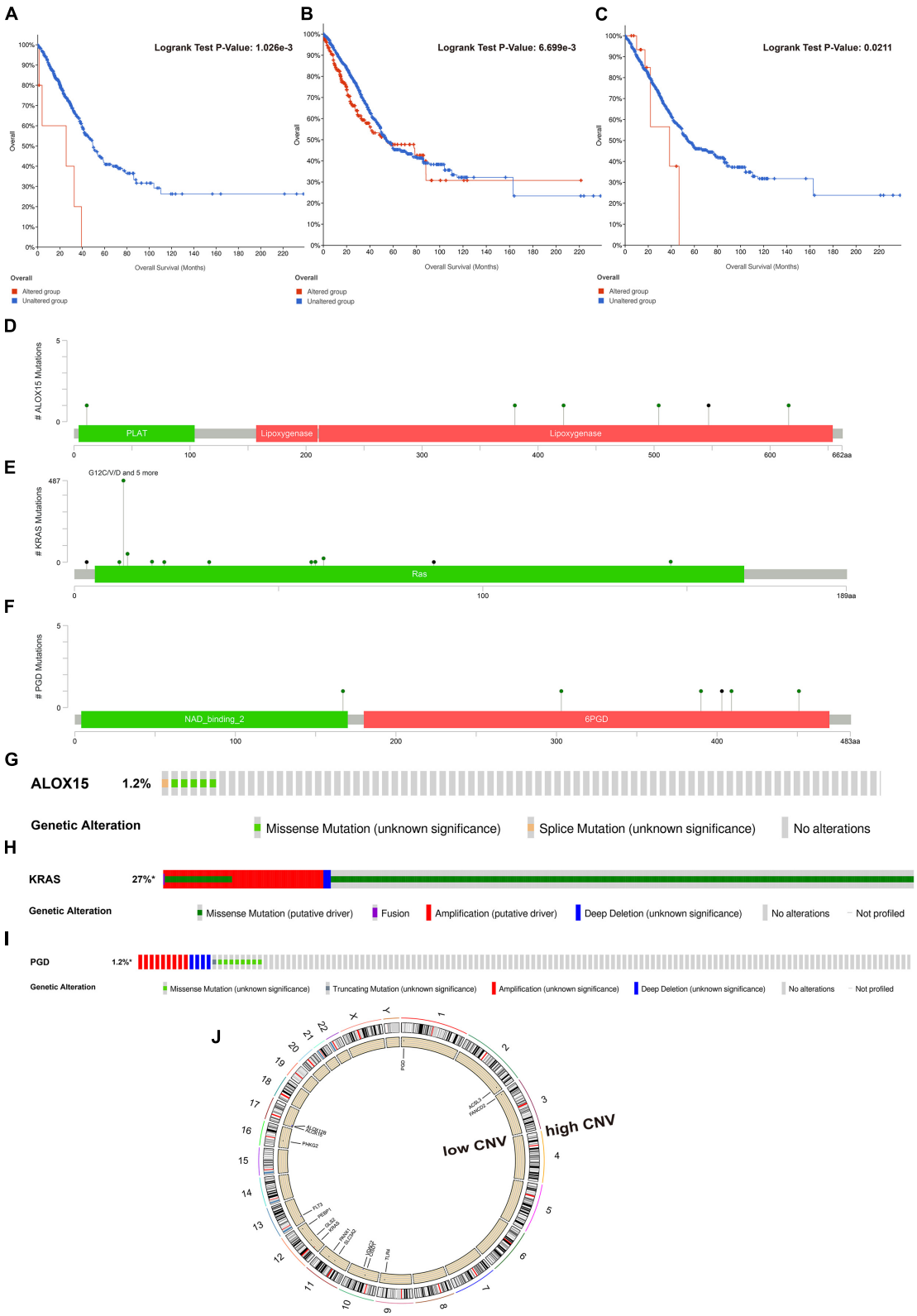


FIGURE 7 | Somatic mutation and CNV analysis of *ALOX15*, *KRAS*, and *PGD* in LUAD. Survival analysis between the altered group and unaltered group based on OS for *ALOX15* **(A)**, *KRAS* **(B)**, and *PGD* **(C)**. The alteration sites of each gene on the coding protein domain for *ALOX15* **(D)**, *KRAS* **(E)**, and *PGD* **(F)**. The alteration rate and type of *ALOX15* **(G)**, *KRAS* **(H)**, and *PGD* **(I)**. **(J)** The location on chromosomes and most common CNV status of the 15 FRPSGs.

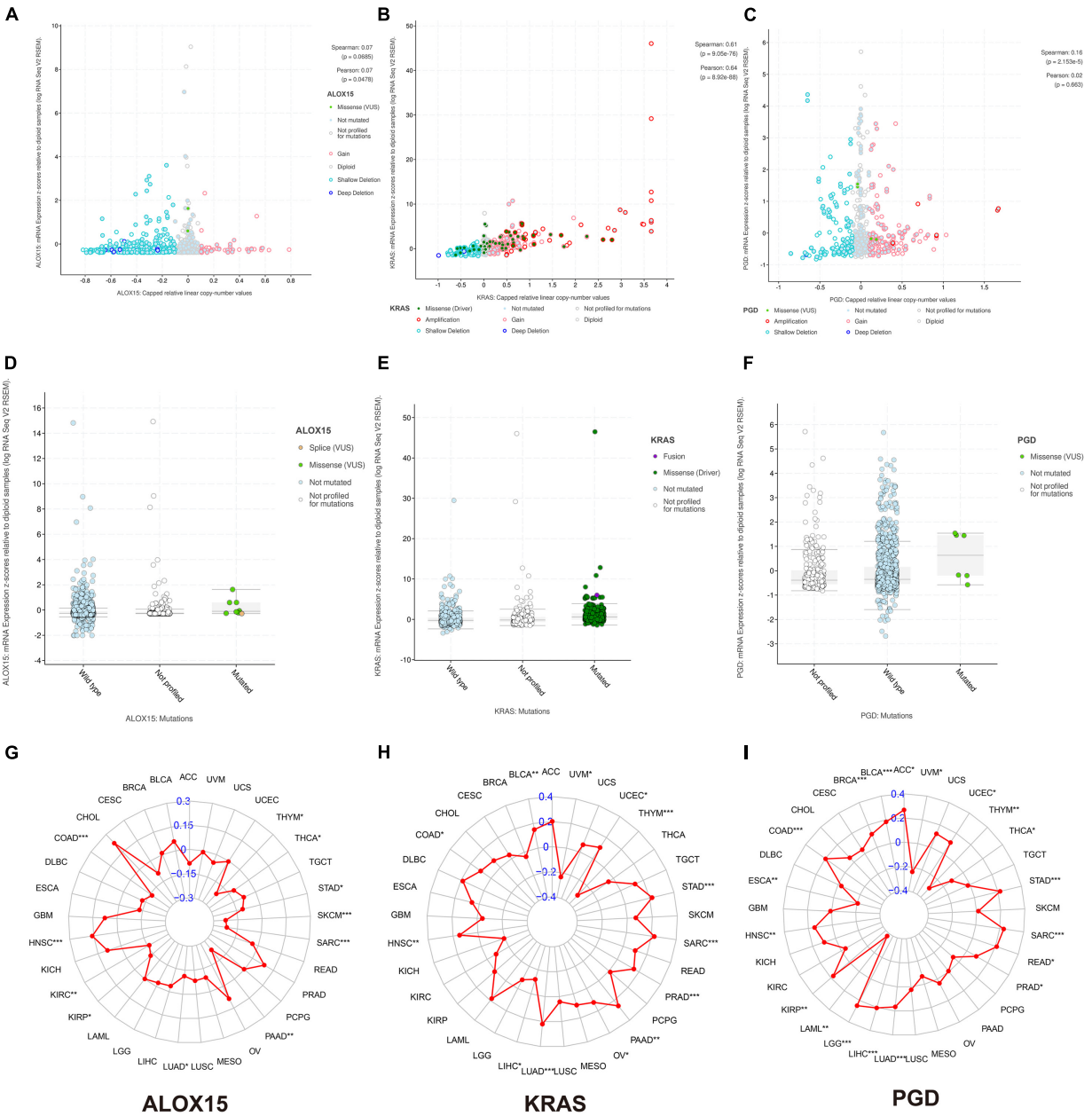


FIGURE 8 | The association between gene expression and alteration status in adenocarcinoma using cBioPortal and TMB of *ALOX15*, *KRAS*, and *PGD* in pan-cancer. The association between gene expression and CNV for *ALOX15* **(A)**, *KRAS* **(B)**, and *PGD* **(C)**. Comparison of gene expression between wild type and mutation type for *ALOX15* **(D)**, *KRAS* **(E)**, and *PGD* **(F)**. The association between gene expression and TMB status in pan-cancer for *ALOX15* **(G)**, *KRAS* **(H)**, and *PGD* **(I)**. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

session of lung cancer (Li et al., 2019a). ALOX12B, R-type of ALOX12 catalyzed by ALOX12, is necessary for p53-mediated antitumor effect via ferroptosis (Mashima and Okuyama, 2015). Jiang et al. (2020) have demonstrated that ALOX12B promotes cervical cancer progression through modulating the PI3K/ERK1 signaling pathway. ALOX12B has a negative relation to the prognosis of LUAD in our study, which indicates that ALOX12B may be involved in some other tumor regulatory mechanisms (Jiang et al., 2020).

(Anti)oxidant metabolism involves *SLC3A2*, *VDAC2*, *GLS2*, and *FLT3*, within which SLC3A2 is the heavy chain subunit of system xc-cystine/glutamate antiporter, maintaining the steady state of redox (Koppula et al., 2020). It is reported that inhibition of SLC3A2 promotes ferroptosis in both tumor cells and normal cells (Koppula et al., 2020; Liu et al., 2020). VDAC2 can interact with Mcl-1, an antiapoptotic member of the Bcl-2 family frequently upregulated in NSCLC. In A549 cells, a decrease in Mcl-1 expression level or application of VDAC-based

peptides inhibits Ca (2+) flow into the mitochondrial matrix, leading to the limitation of ROS generation (Huang et al., 2014). GLS2 plays a critical in glutaminolysis, which can provide intracellular glutamate for GSH synthesis. Zhu et al. (2017) knocked down *GLS2* in mouse embryonic fibroblasts and found that the ferroptosis is reduced. Jennis et al. (2016) explored that GLS2 may act as a transcriptional target of TP53 and promote TP53-dependent ferroptosis. FLT3, a receptor tyrosine kinase, was reported as a critical mediator in the process of glutamate oxidative stress-induced cell death. Although we have no clear understanding of what role FLT3 plays in lung cancer until today, FLT3-ligand, commonly regarded as a hematopoietic stimulator, can be used in treating lung infection via functioning in lung immune cell populations (Dessein et al., 2020).

PGD involves energy metabolism such that it functions in the pentose phosphate pathway (PPP). Although the association between energy metabolism and ferroptosis remains puzzling, Dubreuil et al. (2020) have recently discovered that disrupting PPP by PGD knockout inhibits ROS accumulation. Besides, many previous studies have demonstrated that overexpression of PGD could lead to cancer metastasis and poor prognosis (Bechard et al., 2018; Sarfraz et al., 2020). In our study, the alteration rate of *PGD* was more than 1.2% and was related to a significantly poor prognosis. It is noteworthy that the CNV was relatively common in *PGD* alteration. Shallow deletion, amplification, and deep deletion occurred, and the gene expression was significantly related to CNV. More research is needed to explore the relationship between PGD's CNV and ferroptosis.

The remaining three FRPSGs, *TLR4*, *FANCD2*, and *KRAS*, serve as specific mediators. TLR4 is a toll-like receptor responsible for activating the innate immune system and has been confirmed to be associated with inflammation, autophagy, cell migration, and adhesion/metastasis in NSCLC (Kim et al., 2020). Li et al. (2019b) constructed a heart ischemia–reperfusion model. They demonstrated that the TLR4/TRIF/type I IFN signaling pathway impacts recruiting neutrophils to injured myocardium in the ferroptosis process (Li et al., 2019b). FANCD2, a nuclear protein involved in DNA reparation, prevents bone marrow stromal cells from ferroptosis via iron accumulation and lipid peroxidation inhibition, which is fulfilled by maintaining DNA stability when treated with ferroptosis inducer such as erastin (Song et al., 2016). FANCD2 has been verified that its exhaustion would increase cancer cell proliferation in the knockout experiment on the PDX model. The overexpression of FANCD2 would bring an increased risk of the incidence of lung cancer (Wang et al., 2015; Yao et al., 2015). KRAS, one of the most frequently mutated oncogenes in human cancer, is a signal transducer protein in cell proliferation. It has been reported that erastin has exhibited greater lethality in *KRAS*-mutated cancer cells (Yagoda et al., 2007). On the contrary, Schott et al. demonstrated that *KRAS* mutation leads to increased resistance to ferroptosis. Therefore, the exact role of *KRAS* in ferroptosis remains elusive. In this study, we considered *KRAS* as a ferroptosis driver gene according to the FerrDb database. The alteration rate of *KRAS* reached 27% in LUAD patients. Besides, the high CNV status of *KRAS* means it is easy to detect at an early stage. ALL implies KRAS could be the best ferroptosis-based treatment target and biomarker.

Ferroptosis is a highly complicated cellular process involving lipid, iron, and cysteine metabolism (Stockwell et al., 2017). Discussion on specific genes cannot explain the complex mechanisms involved explicitly. Therefore, we performed a functional analysis. The results showed that the tumor cellular fundamental biological behavior within the high- and low-risk groups differ. Previous studies have demonstrated that many oncogenes and antioncogenes act as ferroptosis regulators in cancers. For example, TP53 can suppress SLC7A11 expression and promote SAT1 (spermidine/spermine N1-acetyltransferase 1), and GLS2 expression to promote ferroptosis. It also can inhibit DPP4 (dipeptidyl peptidase-4) and promote CDKN1A/p21 (cyclin-dependent kinase inhibitor 1A) expression to inhibit the ferroptosis (Xie et al., 2017; Kang et al., 2019). However, the systematic biochemical process and regulation of ferroptosis are unclear. Friedmann Angeli et al. (2019) reviewed current literature and proposed a GPX4 center view depicting the principal metabolic determinants' modulation controlling the cancer growth and persistence. As for the immune function analysis, it is recently reported that ferroptosis cancer cells also release "find me" signals to recruit antigen-presenting cells, initiating the innate immune response. Unlike apoptosis, potential signals for ferroptosis are AA oxidation products. Therefore, ALOXs also regulate immunity (Friedmann Angeli et al., 2019; Stockwell and Jiang, 2019). The reduction in DCs and HLA, two essential components for antigen-presenting, indicates that the TME in high-risk groups can inhibit signal transmission. Impaired type II IFN response also indicates the innate immunity in high-risk groups is compromised. The function of mast cells and natural killer (NK) cells in LUAD TME remains complicated and elusive. Recent studies have demonstrated that mast cells and NK cells improve the prognosis of LUAD patients (Ko et al., 2017; Aktas et al., 2018).

The pan-cancer analysis comprehensively analyzed these 15 genes. First, the expression and survival analysis indicate that the heterogeneity of 15 FRPSGs is evident among all the 33 cancer types. However, ferroptosis suppressor genes were downregulated in most cancer types, consistent with previous speculations (Sun et al., 2016; Roh et al., 2017; Seibt et al., 2019). Second, drug sensitivity analysis proved that ferroptosis is closely related to chemotherapy resistance. Interestingly, our findings also provide a new perspective that ferroptosis may involve some tumor-targeted therapy resistance. Third, immune subtype and TME analysis proved that these FRPSGs take part in the immune response, TME infiltration. Finally, we found that ferroptosis could contribute to cancer stemness based on cancer stemness indexes and previously discovered biomarkers, providing novel insights into ferroptosis.

Before us, FRG signatures had been established and applied for LUAD in recent research (Gao et al., 2021; Sun et al., 2021). Even so, our signature has better robustness with better predictive ability, providing more solid clinical reference. Our AUCs of the ROC curve were 0.718, 0.730, and 0.721 for 1, 2, and 3 years, respectively, while Sun's signature AUCs were merely 0.625, 0.588, and 0.593 for 2, 3, and 5 years, respectively; Gao's AUCs were 0.678, 0.698, and 0.697 for 1, 2, and 3 years, respectively, in the TCGA-LUAD cohort. Moreover, we emphasized the analysis

of FRPSGs, trying to elucidate the potential regulatory role of these genes through an integrative multi-omics study, including expression, prognosis, cancer stemness, drug sensitivity analysis, and mutation analysis.

There are also several limitations to our study. First, the GSE68465/GSE72094 cohorts and TCGA cohort are heterologous, especially the OS of GSE68465 (654.5 days in TCGA and 1410 days in GSE68465). Besides, the data quality and preprocessing of the two data resources (RNA-seq and microarray) were different, which lead to different data processing methods and different cutoff values. However, we could not find an appropriate RNA-seq validation cohort currently. Second, the clinical significance of our model needs to be further evaluated through prospective data in the real world. Third, experimental evidence was not acquired to prove our conclusion, including drug sensitivity analysis and cancer stemness analysis.

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

ZR and MH performed the data analysis and wrote the first draft. ZW and JG revised the manuscript and carried out data collection. XZ and GZ revised the manuscript and tables. HZ designed and supervised the study. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the Postdoctoral Science Foundation of China (No. 2020M672005), the Postdoctoral Innovation Project of Shandong Province (No. 202001016), and the Natural Science Foundation of Shandong Province (No. ZR2019PH029).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.672904/full#supplementary-material

REFERENCES

Aktas, O. N., Öztürk, A. B., Erman, B., Erus, S., Tanju, S., and Dilege, Ş (2018). Role of natural killer cells in lung cancer. *J. Cancer Res. Clin. Oncol.* 144, 997–1003. doi: 10.1007/s00432-018-2635-2633

Alvarez, S. W., Sviderskiy, V. O., Terzi, E. M., Papagiannakopoulos, T., Moreira, A. L., Adams, S., et al. (2017). NFS1 undergoes positive selection in lung tumours and protects cells from ferroptosis. *Nature* 551, 639–643. doi: 10.1038/nature24637

Anthonymuthu, T. S., Kenny, E. M., Shrivastava, I., Tyurina, Y. Y., Hier, Z. E., Ting, H. C., et al. (2018). Empowerment of 15-lipoxygenase catalytic competence in selective oxidation of membrane ETE-PE to ferroptotic death signals, HpETE-PE. *J. Am. Chem. Soc.* 140, 17835–17839. doi: 10.1021/jacs.8b09913

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Bechard, M. E., Word, A. E., Tran, A. V., Liu, X., Locasale, J. W., and McDonald, O. G. (2018). Pentose conversions support the tumorigenesis of pancreatic cancer distant metastases. *Oncogene* 37, 5248–5256. doi: 10.1038/s41388-018-0346-345

Bersuker, K., Hendricks, J. M., Li, Z., Magtanong, L., Ford, B., Tang, P. H., et al. (2019). The CoQ oxidoreductase FSP1 acts parallel to GPX4 to inhibit ferroptosis. *Nature* 575, 688–692. doi: 10.1038/s41586-019-1705-1702

Bertolini, G., Roz, L., Perego, P., Tortoreto, M., Fontanella, E., Gatti, L., et al. (2009). Highly tumorigenic lung cancer CD133+ cells display stem-like features and are spared by cisplatin treatment. *Proc. Natl. Acad. Sci. U S A.* 106, 16281–16286. doi: 10.1073/pnas.0905653106

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492

Chen, X., Xu, S., Zhao, C., and Liu, B. (2019). Role of TLR4/NADPH oxidase 4 pathway in promoting cell death through autophagy and ferroptosis during heart failure. *Biochem. Biophys. Res. Commun.* 516, 37–43. doi: 10.1016/j.bbrc.2019.06.015

Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., and Lin, C. Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Systems Biol.* 8 (Suppl. 4):S11. doi: 10.1186/1752-0509-8-s4-s11

Dessein, R., Bauduin, M., Grandjean, T., Le Guern, R., Figeac, M., Beury, D., et al. (2020). Antibiotic-related gut dysbiosis induces lung immunodepression and worsens lung infection in mice. *Crit. Care* 24:611. doi: 10.1186/s13054-020-03320-3328

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

Dubreuil, M. M., Morgens, D. W., Okumoto, K., Honsho, M., Contrepois, K., Lee-McMullen, B., et al. (2020). Systematic identification of regulators of oxidative stress reveals non-canonical roles for peroxisomal import and the pentose phosphate pathway. *Cell Rep.* 30, 1417–1433.e7. doi: 10.1016/j.celrep.2020.01.013

Friedmann Angeli, J. P., Krysko, D. V., and Conrad, M. (2019). Ferroptosis at the crossroads of cancer-acquired drug resistance and immune evasion. *Nat. Rev. Cancer* 19, 405–414. doi: 10.1038/s41568-019-0149-141

Gao, M., Monian, P., Quadri, N., Ramasamy, R., and Jiang, X. (2015). Glutaminolysis and transferrin regulate ferroptosis. *Mol. Cell.* 59, 298–308. doi: 10.1016/j.molcel.2015.06.011

Gao, X., Tang, M., Tian, S., Li, J., and Liu, W. (2021). A ferroptosis-related gene signature predicts overall survival in patients with lung adenocarcinoma. *Future Oncol.* 17, 1533–1544. doi: 10.2217/fon-2020-1113

Goldman, M. J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., et al. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. *Nat. Biotechnol.* 38, 675–678. doi: 10.1038/s41587-020-0546-548

Hirsch, F. R., Scagliotti, G. V., Mulshine, J. L., Kwon, R., Curran, W. J. Jr., Wu, Y. L., et al. (2017). Lung cancer: current therapies and new targeted treatments. *Lancet (London, England)* 389, 299–311. doi: 10.1016/s0140-6736(16)30958-30958

Huang, H., Shah, K., Bradbury, N. A., Li, C., and White, C. (2014). Mcl-1 promotes lung cancer cell migration by directly interacting with VDAC to increase mitochondrial Ca2+ uptake and reactive oxygen species generation. *Cell Death Dis.* 5:e1482. doi: 10.1038/cddis.2014.419

Imielinski, M., Berger, A. H., Hammerman, P. S., Hernandez, B., Pugh, T. J., Hodis, E., et al. (2012). Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 150, 1107–1120. doi: 10.1016/j.cell.2012.08.029

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

Ji, X., Qian, J., Rahman, S. M. J., Siska, P. J., Zou, Y., Harris, B. K., et al. (2018). xCT (SLC7A11)-mediated metabolic reprogramming promotes non-small cell lung cancer progression. *Oncogene* 37, 5007–5019. doi: 10.1038/s41388-018-0307-z

Jiang, T., Zhou, B., Li, Y. M., Yang, Q. Y., Tu, K. J., and Li, L. Y. (2020). ALOX12B promotes carcinogenesis in cervical cancer by regulating the PI3K/ERK1 signaling pathway. *Oncol. Lett.* 20, 1360–1368. doi: 10.3892/ol.2020.11641

Jiang, Y., Mao, C., Yang, R., Yan, B., Shi, Y., Liu, X., et al. (2017). EGLN1/c-Myc Induced Lymphoid-Specific Helicase Inhibits Ferroptosis through Lipid Metabolic Gene Expression Changes. *Theranostics* 7, 3293–3305. doi: 10.7150/thno.19988

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

Kang, R., Kroemer, G., and Tang, D. (2019). The tumor suppressor protein p53 and the ferroptosis network. *Free Radical Biol. Med.* 133, 162–168. doi: 10.1016/j.freeradbiomed.2018.05.074

Kang, Y., Tiziani, S., Park, G., Kaul, M., and Paternostro, G. (2014). Cellular protection using Flt3 and PI3Kα inhibitors demonstrates multiple mechanisms of oxidative glutamate toxicity. *Nat. Commun.* 5:3672. doi: 10.1038/ncomms4672

Kim, M. J., Min, Y., Son, J., Kim, J. Y., Lee, J. S., Kim, D. H., et al. (2020). AMPKα1 regulates lung and breast cancer progression by regulating TLR4-Mediated TRAF6-BECN1 signaling axis. *Cancers (Basel)* 12:3289. doi: 10.3390/cancers12113289

Kinoshita, F. L., Ito, Y., and Nakayama, T. (2016). Trends in lung cancer incidence rates by histological type in 1975-2008: a population-based study in osaka. Japan. *J. Epidemiol.* 26, 579–586. doi: 10.2188/jea.JE20150257

Ko, E. A., Sanders, K. M., and Zhou, T. (2017). A transcriptomic insight into the impacts of mast cells in lung, breast, and colon cancers. *Oncoimmunology* 6:e1360457. doi: 10.1080/2162402x.2017.1360457

Kocher, F., Hilbe, W., Seeber, A., Pircher, A., Schmid, T., Greil, R., et al. (2015). Longitudinal analysis of 2293 NSCLC patients: a comprehensive study from the TYROL registry. *Lung cancer (Amsterdam, Netherlands)* 87, 193–200. doi: 10.1016/j.lungcan.2014.12.006

Koppula, P., Zhuang, L., and Gan, B. (2020). Cystine transporter SLC7A11/xCT in cancer: ferroptosis, nutrient dependency, and cancer therapy. *Protein Cell.* doi: 10.1007/s13238-020-00789-785 [Epub ahead of print].

Kukulj, S., Jaganjac, M., Boranic, M., Krizanac, S., Santic, Z., and Poljak-Blazi, M. (2010). Altered iron metabolism, inflammation, transferrin receptors, and ferritin expression in non-small-cell lung cancer. *Med. Oncol. (Northwood, London, England)* 27, 268–277. doi: 10.1007/s12032-009-9203-9202

Lei, G., Zhang, Y., Koppula, P., Liu, X., Zhang, J., Lin, S. H., et al. (2020). The role of ferroptosis in ionizing radiation-induced cell death and tumor suppression. *Cell Res.* 30, 146–162. doi: 10.1038/s41422-019-0263-3

Leung, E. L., Fiscus, R. R., Tung, J. W., Tin, V. P., Cheng, L. C., Sihoe, A. D., et al. (2010). Non-small cell lung cancer cells expressing CD44 are enriched for stem cell-like properties. *PLoS One* 5:e14062. doi: 10.1371/journal.pone.0014062

Li, J., Liu, J., Xu, Y., Wu, R., Chen, X., Song, X., et al. (2021). Tumor heterogeneity in autophagy-dependent ferroptosis. *Autophagy* doi: 10.1080/15548627.2021.1872241 [Epub ahead of print].

Li, M. Y., Liu, L. Z., Li, W., Ng, C. S. H., Liu, Y., Kong, A. W. Y., et al. (2019a). Ambient fine particulate matter inhibits 15-lipoxygenases to promote lung carcinogenesis. *J. Exp. Clin. Cancer Res.* 38:359. doi: 10.1186/s13046-019-1380-z

Li, W., Feng, G., Gauthier, J. M., Lokshina, I., Higashikubo, R., Evans, S., et al. (2019b). Ferroptotic cell death and TLR4/Trif signaling initiate neutrophil recruitment after heart transplantation. *J. Clin. Invest.* 129, 2293–2304. doi: 10.1172/jci126428

Liang, C., Zhang, X., Yang, M., and Dong, X. (2019). Recent progress in ferroptosis inducers for cancer therapy. *Adv. Mater. (Deerfield Beach, Fla)* 31:e1904197. doi: 10.1002/adma.201904197

Liu, J., Xia, X., and Huang, P. (2020). xCT: a critical molecule that links cancer metabolism to redox signaling. *Mol. Therapy* 28, 2358–2366. doi: 10.1016/j.ymthe.2020.08.021

Liu, Y., Chen, P., Li, M., Fei, H., Huang, J., Zhao, T., et al. (2021). Comprehensive analysis of the control of cancer stem cell characteristics in endometrial cancer by network analysis. *Comp. Mathem. Methods Med.* 2021:6653295. doi: 10.1155/2021/6653295

Magtanong, L., Ko, P. J., To, M., Cao, J. Y., Forcina, G. C., Tarangelo, A., et al. (2019). Exogenous monounsaturated fatty acids promote a ferroptosis-resistant cell state. *Cell Chem. Biol.* 26, 420–432.e9. doi: 10.1016/j.chembiol.2018.11.016

Malta, T. M., Sokolov, A., Gentles, A. J., Burzykowski, T., Poisson, L., Weinstein, J. N., et al. (2018). Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* 173, 338–354.e15. doi: 10.1016/j.cell.2018.03.034

Marabelle, A., Fakih, M., Lopez, J., Shah, M., Shapira-Frommer, R., Nakagawa, K., et al. (2020). Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 KEYNOTE-158 study. *Lancet Oncol.* 21, 1353–1365. doi: 10.1016/s1470-2045(20)30445-30449

Mashima, R., and Okuyama, T. (2015). The role of lipoxygenases in pathophysiology; new insights and future perspectives. *Redox Biol.* 6, 297–310. doi: 10.1016/j.redox.2015.08.006

Mizuno, H., Kitada, K., Nakai, K., and Sarai, A. (2009). PrognoScan: a new database for meta-analysis of the prognostic value of genes. *BMC Med. Genom.* 2:18. doi: 10.1186/1755-8794-2-18

Mou, Y., Wang, J., Wu, J., He, D., Zhang, C., Duan, C., et al. (2019). Ferroptosis, a new form of cell death: opportunities and challenges in cancer. *J. Hematol. Oncol.* 12:34. doi: 10.1186/s13045-019-0720-y

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team.

Roh, J. L., Kim, E. H., Jang, H., and Shin, D. (2017). Nrf2 inhibition reverses the resistance of cisplatin-resistant head and neck cancer cells to artesunate-induced ferroptosis. *Redox Biol.* 11, 254–262. doi: 10.1016/j.redox.2016.12.010

Roh, J. L., Kim, E. H., Jang, H. J., Park, J. Y., and Shin, D. (2016). Induction of ferroptotic cell death for overcoming cisplatin resistance of head and neck cancer. *Cancer Lett.* 381, 96–103. doi: 10.1016/j.canlet.2016.07.035

Sarfraz, I., Rasul, A., Hussain, G., Shah, M. A., Zahoor, A. F., Asrar, M., et al. (2020). 6-Phosphogluconate dehydrogenase fuels multiple aspects of cancer cells: from cancer initiation to metastasis and chemoresistance. *BioFactors (Oxford, England)* 46, 550–562. doi: 10.1002/biof.1624

Schabath, M. B., Welsh, E. A., Fulp, W. J., Chen, L., Teer, J. K., Thompson, Z. J., et al. (2016). Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma. *Oncogene* 35, 3209–3216. doi: 10.1038/onc.2015.375

Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protocols* 3, 1101–1108. doi: 10.1038/nprot.2008.73

Seibt, T. M., Proneth, B., and Conrad, M. (2019). Role of GPX4 in ferroptosis and its pharmacological implication. *Free Radical Biol. Med.* 133, 144–152. doi: 10.1016/j.freeradbiomed.2018.09.014

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Shedden, K., Taylor, J. M., Enkemann, S. A., Tsao, M. S., Yeatman, T. J., Gerald, W. L., et al. (2008). Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat. Med.* 14, 822–827. doi: 10.1038/nm.1790

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

Song, X., Xie, Y., Kang, R., Hou, W., Sun, X., Epperly, M. W., et al. (2016). FANCD2 protects against bone marrow injury from ferroptosis. *Biochem. Biophys. Res. Commun.* 480, 443–449. doi: 10.1016/j.bbrc.2016.10.068

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

Stockwell, B. R., and Jiang, X. (2019). A physiological function for ferroptosis in tumor suppression by the immune system. *Cell Metab.* 30, 14–15. doi: 10.1016/j.cmet.2019.06.012

Su, L., Jiang, X., Yang, C., Zhang, J., Chen, B., Li, Y., et al. (2019). Pannexin 1 mediates ferroptosis that contributes to renal ischemia/reperfusion injury. *J. Biol. Chem.* 294, 19395–19404. doi: 10.1074/jbc.RA119.010949

Sun, S., Guo, W., Lv, F., Zhang, G., Wang, J., Li, R., et al. (2021). Comprehensive analysis of ferroptosis regulators in lung adenocarcinomas identifies prognostic and immunotherapy-related biomarkers. *Front. Mol. Biosci.* 8:587436. doi: 10.3389/fmolb.2021.587436

Sun, X., Niu, X., Chen, R., He, W., Chen, D., Kang, R., et al. (2016). Metallothionein-1G facilitates sorafenib resistance through inhibition of ferroptosis. *Hepatology (Baltimore, Md)* 64, 488–500. doi: 10.1002/hep.28574

Sunnetcioglu, A., Alp, H. H., Sertogulları ndan, B., Balaharoglu, R., and Gunbatar, H. (2016). Evaluation of oxidative damage and antioxidant mechanisms in COPD, lung cancer, and obstructive sleep apnea syndrome. *Respiratory Care* 61, 205–211. doi: 10.4187/respcare.04209

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. doi: 10.1093/nar/gky1131

Tang, Y., Zhou, J., Hooi, S. C., Jiang, Y. M., and Lu, G. D. (2018). Fatty acid activation in carcinogenesis and cancer development: essential roles of long-chain acyl-CoA synthetases. *Oncol. Lett.* 16, 1390–1396. doi: 10.3892/ol.2018.8843

Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 45, W98–W102. doi: 10.1093/nar/gkx247

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

Torti, S. V., Manz, D. H., Paul, B. T., Blanchette-Farra, N., and Torti, F. M. (2018). Iron and Cancer. *Annu. Rev. Nutr.* 38, 97–125. doi: 10.1146/annurev-nutr-082117-51732

Valavanidis, A., Vlachogianni, T., Fiotakis, K., and Loridas, S. (2013). Pulmonary oxidative stress, inflammation and cancer: respirable particulate matter, fibrous dusts and ozone as major causes of lung carcinogenesis through reactive oxygen species mechanisms. *Int. J. Environ. Res. Public Health* 10, 3886–3907. doi: 10.3390/ijerph10093886

Villalpando-Rodriguez, G. E., Blankstein, A. R., Konzelman, C., and Gibson, S. B. (2019). Lysosomal destabilizing drug siramesine and the dual tyrosine kinase inhibitor lapatinib induce a synergistic ferroptosis through reduced heme oxygenase-1 (HO-1) levels. *Oxid. Med. Cell. Long.* 2019:9561281. doi: 10.1155/2019/9561281

Wang, G. Z., Liu, Y. Q., Cheng, X., and Zhou, G. B. (2015). Celastrol induces proteasomal degradation of FANCD2 to sensitize lung cancer cells to DNA crosslinking agents. *Cancer Sci.* 106, 902–908. doi: 10.1111/cas.12679

Wang, M., Mao, C., Ouyang, L., Liu, Y., Lai, W., Liu, N., et al. (2019a). Long noncoding RNA LINC00336 inhibits ferroptosis in lung cancer by functioning as a competing endogenous RNA. *Cell Death Differ.* 26, 2329–2343. doi: 10.1038/s41418-019-0304-y

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

Wang, Z., Wu, D., Xia, Y., Yang, B., and Xu, T. (2021). Identification of hub genes and compounds controlling ovarian cancer stem cell characteristics via stemness indices analysis. *Ann. Trans. Med.* 9:379. doi: 10.21037/atm-20-3621

Wenzel, S. E., Tyurina, Y. Y., Zhao, J., St Croix, C. M., Dar, H. H., Mao, G., et al. (2017). PEBP1 wardens ferroptosis by enabling lipoxygenase generation of lipid death signals. *Cell* 171, 628–641.e26. doi: 10.1016/j.cell.2017.09.044

Xie, Y., Hou, W., Song, X., Yu, Y., Huang, J., Sun, X., et al. (2016). Ferroptosis: process and function. *Cell Death Differ.* 23, 369–379. doi: 10.1038/cdd.2015.158

Xie, Y., Zhu, S., Song, X., Sun, X., Fan, Y., Liu, J., et al. (2017). The tumor suppressor p53 limits ferroptosis by blocking DPP4 activity. *Cell Rep.* 20, 1692–1704. doi: 10.1016/j.celrep.2017.07.055

Xu, G., Wang, H., Li, X., Huang, R., and Luo, L. (2021). Recent progress on targeting ferroptosis for cancer therapy. *Biochem. Pharmacol.* 190:114584. doi: 10.1016/j.bcp.2021.114584

Yagoda, N., von Rechenberg, M., Zaganjor, E., Bauer, A. J., Yang, W. S., Fridman, D. J., et al. (2007). RAS-RAF-MEK-dependent oxidative cell death involving voltage-dependent anion channels. *Nature* 447, 864–868. doi: 10.1038/nature05859

Yang, W. S., Kim, K. J., Gaschler, M. M., Patel, M., Shchepinov, M. S., and Stockwell, B. R. (2016). Peroxidation of polyunsaturated fatty acids by lipoxygenases drives ferroptosis. *Proc. Natl. Acad. Sci. U S A.* 113, E4966–E4975. doi: 10.1073/pnas.1603244113

Yao, C., Du, W., Chen, H., Xiao, S., Huang, L., and Chen, F. P. (2015). Involvement of *Fanconi anemia* genes FANCD2 and FANCF in the molecular basis of drug resistance in leukemia. *Mol. Med. Rep.* 11, 4605–4610. doi: 10.3892/mmr.2015.3288

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

Yuan, H., Li, X., Zhang, X., Kang, R., and Tang, D. (2016). CISD1 inhibits ferroptosis by protection against mitochondrial lipid peroxidation. *Biochem. Biophys. Res. Commun.* 478, 838–844. doi: 10.1016/j.bbrc.2016.08.034

Zalewska-Ziob, M., Adamek, B., Kasperczyk, J., Romuk, E., Hudziec, E., Chwaliñska, E., et al. (2019). Activity of antioxidant enzymes in the tumor and adjacent noncancerous tissues of non-small-cell lung cancer. *Oxid. Med. Cell. Long.* 2019:2901840. doi: 10.1155/2019/2901840

Zhang, X., Klamer, B., Li, J., Fernandez, S., and Li, L. (2020). A pan-cancer study of class-3 semaphorins as therapeutic targets in cancer. *BMC Med. Genom.* 13(Suppl. 5):45. doi: 10.1186/s12920-020-0682-685

Zhao, J., Dar, H. H., Deng, Y., St Croix, C. M., Li, Z., Minami, Y., et al. (2020). PEBP1 acts as a rheostat between prosurvival autophagy and ferroptotic death in asthmatic epithelial cells. *Proc. Natl. Acad. Sci. U S A.* 117, 14376–14385. doi: 10.1073/pnas.1921618117

Zhou, N., and Bao, J. (2020). FerrDb: a manually curated resource for regulators and markers of ferroptosis and ferroptosis-disease associations. *Database* 2020:baaa021. doi: 10.1093/database/baaa021

Zhu, Y., Li, T., Ramos, da Silva, S., Lee, J. J., Lu, C., et al. (2017). A critical role of glutamine and asparagine γ-nitrogen in nucleotide biosynthesis in cancer cells hijacked by an oncogenic virus. *mBio* 8:e01179-17. doi: 10.1128/mBio.01179-1117

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

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 23 August 2021
doi: 10.3389/fgene.2021.702424

A Novel Intercellular Communication-Associated Gene Signature for Prognostic Prediction and Clinical Value in Patients With Lung Adenocarcinoma

Qin-Yu Zhao[1,2†], Le-Ping Liu[1†], Lu Lu[1], Rong Gui[1] and Yan-Wei Luo[1*]*

[1] Department of Blood Transfusion, The Third Xiangya Hospital of Central South University, Changsha, China, [2] College of Engineering and Computer Science, Australian National University, Canberra, ACT, Australia

OPEN ACCESS

Edited by:
Federica Calore,
The Ohio State University,
United States

Reviewed by:
Xiangqian Guo,
Henan University, China
Lei Li,
Shanghai Jiao Tong University, China

****Correspondence:***
Rong Gui
guirong@csu.edu.cn
Yan-Wei Luo
royalway@csu.edu.cn

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

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *06 May 2021*
Accepted: *04 August 2021*
Published: *23 August 2021*

Citation:
Zhao Q-Y, Liu L-P, Lu L, Gui R and Luo Y-W (2021) A Novel Intercellular Communication-Associated Gene Signature for Prognostic Prediction and Clinical Value in Patients With Lung Adenocarcinoma. Front. Genet. 12:702424. doi: 10.3389/fgene.2021.702424

Background: Lung cancer remains the leading cause of cancer death globally, with lung adenocarcinoma (LUAD) being its most prevalent subtype. This study aimed to identify the key intercellular communication-associated genes (ICAGs) in LUAD.

Methods: Eight publicly available datasets were downloaded from the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases. The prognosis-related ICAGs were identified and a risk score was developed by using survival analysis. Machine learning models were trained to predict LUAD recurrence based on the selected ICAGs and clinical information. Comprehensive analyses on ICAGs and tumor microenvironment were performed. A single-cell RNA-sequencing dataset was assessed to further elucidate aberrant changes in intercellular communication.

Results: Eight ICAGs with prognostic potential were identified in the present study, and a risk score was derived accordingly. The best machine-learning model to predict relapse was developed based on clinical information and the expression levels of these eight ICAGs. This model achieved a remarkable area under receiver operator characteristic curves of 0.841. Patients were divided into high- and low-risk groups according to their risk scores. DNA replication and cell cycle were significantly enriched by the differentially expressed genes between the high- and the low-risk groups. Infiltrating immune cells, immune functions were significantly related to ICAGs expressions and risk scores. Additionally, the changes of intercellular communication were modeled by analyzing the single-cell sequencing dataset.

Conclusion: The present study identified eight key ICAGs in LUAD, which could contribute to patient stratification and act as novel therapeutic targets.

Keywords: lung adenocarcinoma, intercellular communication, prognosis prediction, machine learning, tumor microenvironment, single-cell RNA-sequencing

INTRODUCTION

Lung cancer is the leading cause of cancer death and approximately 1.8 million deaths worldwide in 2020 had lung cancer as the primary cause (Sung et al., 2021). Non-small-cell lung cancer (NSCLC), one of the main histological subtypes, includes approximately 85% of the lung cancer cases. Lung adenocarcinoma (LUAD) is the most common subtype of NSCLC (Molina et al., 2008). Despite recent advances in various targeted therapies, LUAD is still characterized by a low 5-year survival rate (Ahluwalia et al., 2021). Therefore, it is crucial to identify a novel gene signature for LUAD patients' prognosis and for the exploration of novel therapeutic targets for LUAD.

Intercellular communication, defined as the information transfer between cells, is vital for cells to grow and function normally, and may provide a unique perspective for LUAD prognosis (Mittelbrunn and Sanchez-Madrid, 2012). Cells share information by direct and indirect signaling, and the related pathways can be regulated at the gene expression level (Mittelbrunn and Sanchez-Madrid, 2012; Brucher and Jamall, 2014). Direct intercellular communication involves self-to-self communication and adjacent communication with nearby cells, while indirect intercellular communication involves local communication *via* hormones over short or large distances, respectively. Communication, occluding, and anchoring junctions are the three essential components of intercellular communication (Brucher and Jamall, 2014).

Aberrant alterations of intercellular communication in the tumor microenvironment (TME) are related to the occurrence, invasion, metastasis, and drug resistance of cancers (Xu et al., 2018; Maacha et al., 2019). Increasing evidence suggests that versatile immune cells are infiltrated in the TME of LUAD and play an essential role in cancer progression and aggressiveness (Ma et al., 2020; Tan et al., 2020). Communication between tumor cells and tumor-infiltrating immune cells may significantly affect the functions of the immune system, potentially deteriorating the clinical outcomes (Parri et al., 2020). A better understanding of the intercellular communication in TME could thus shed light on the pathogenesis and prognosis of LUAD. As a result, intercellular communication plays a significant role in many pathways and has an important impact on TME of lung cancer.

Despite the significance of intercellular communication in LUAD, it is still considered an underexplored domain. A number of previous studies have reported gene signatures with prognostic potential for LUAD, including immune (Zhang et al., 2019), hypoxia (Mo et al., 2020), and ferroptosis-related genes (Gao et al., 2021). Nevertheless, limited work has been done so far to reveal and study the intercellular communication-associated genes (ICAGs). Besides, computational methods based on single-cell RNA sequencing data have demonstrated an outstanding potential in investigating the intercellular communication in high resolution (Wang Y. et al., 2019; Efremova et al., 2020). These methods mainly focus on ligand–receptor interactions, and therefore, less attention has been given to the prognostic potential of the ICAGs.

The present study aimed to identify the key ICAGs that could serve as prognostic markers or therapeutic targets for LUAD patients. Eight publicly available datasets were analyzed and eight LUAD prognosis ICAGs were identified. Machine learning models were then developed based on these genes and clinical information to predict the recurrence of LUAD. Comprehensive analyses on ICAGs were performed, including mutation, DNA methylation, post-transcriptional regulation, pathway activity, and drug resistance correlation analyses. Patients were divided into high- and low-risk groups according to the expression levels of these genes. Gene set enrichment analyses were performed on the differentially expressed genes (DEGs) between the high- and the low-risk groups. Tumor-infiltrating immune cells, immune functions and immune checkpoints were evaluated in different groups by using 10 different approaches. Additionally, a single-cell RNA sequencing dataset was assessed to elucidate further differences between the high- and the low-risk groups in the intercellular communication. The design of the present study was summarized in **Figure 1**.

MATERIALS AND METHODS

Expression Microarray Datasets

Systematic data mining and computerized searches in the Gene Expression Omnibus (GEO) database were conducted in our study. Seven publicly available LUAD datasets, including GSE19188, GSE30219, GSE31210, GSE31546, GSE37745, GSE50081, and GSE68465, were retrieved accessing the overall survival time. The raw data were downloaded and normalized using the same methods and parameters described in the original studies. Probes with missing gene symbols were excluded. The median expression intensity was used when there were multiple probe sets mapping to the same gene symbol. Besides, gene expression data in fragments per kilobase million (FPKM) values and clinical information of the LUAD dataset from The Cancer Genome Atlas (TCGA) were also downloaded. Expression data were then transformed into the transcripts per million (TPM) values.

An empirical Bayes method was utilized to remove the batch effects by using the functions provided in the sva R package (version 3.34.0) (Leek et al., 2012). The datasets from GEO were combined and used as the discovery set, while the TCGA-LUAD cohort was used as the validation set.

Abbreviation: NSCLC, non-small-cell lung cancer; LUAD, lung adenocarcinoma; TME, tumor microenvironment; ICAG, intercellular communication-associated gene; DEG, differentially expressed gene; GEO, Gene Expression Omnibus; FPKM, fragments per kilobase million; TCGA, The Cancer Genome Atlas; TPM, transcripts per million; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology; LASSO, the least absolute shrinkage and selection operator; PCA, principal component analysis; CatBoost, Categorical Boosting; AUROC, the areas under receiver operator characteristic curve; SHAP, SHapley Additive exPlanation; GSCA, Gene Set Cancer Analysis; SNV, single-nucleotide variant; CNV, copy number variation; miRNA, micro RNA; CTRP, the Cancer Therapeutics Response Portal; GDSC, the Genomics of Drug Sensitivity in Cancer; lncRNA, long non-coding RNA; GRCh38.p13, Genome Reference Consortium Human Build 38 patch release 13 m6A, N6-methyladenosine; ssGSEA, single-sample gene set enrichment analysis.

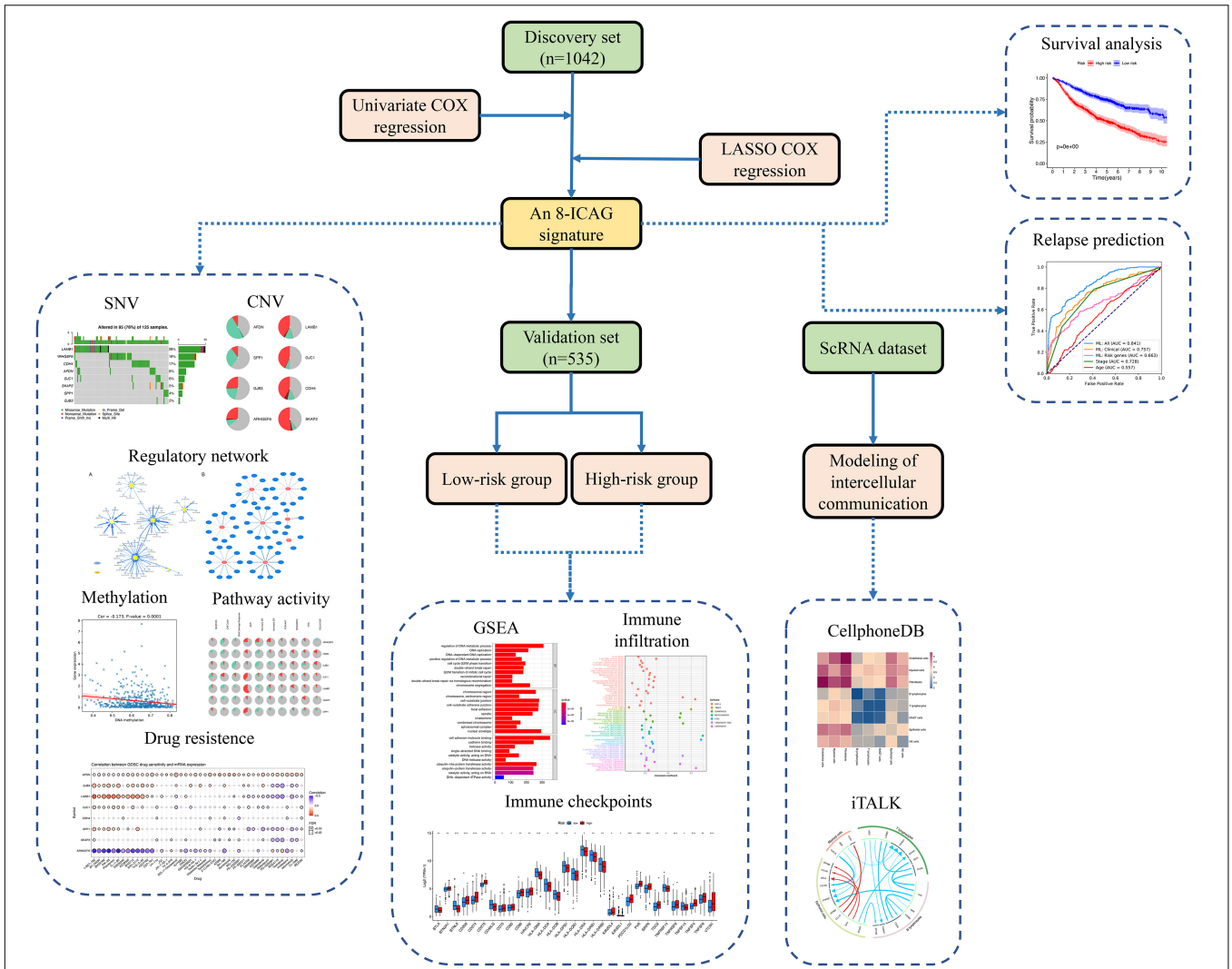


FIGURE 1 | Flow chart of the design of the present study.

Identification of Prognosis-Related ICAGs

The gene list of ICAGs was collected based on the Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), and Reactome databases (Jassal et al., 2020). A total of 440 genes were eventually downloaded from these databases. After duplicates were removed, a list of 426 ICAGs was obtained.

Univariate COX regression was used to assess ICAGs based on the discovery set, and the genes significantly correlated to the overall survival were identified with a *P*-value threshold of 0.01. The rigorous *P*-value cutoff was used to obtain a better prediction performance. Further, ICAGs with statistical significance in univariate regression were evaluated using the least absolute shrinkage and selection operator (LASSO) COX regression model. ICAGs with the best prognostic value were screened out. A risk score was then constructed according to the fitted coefficients of the LASSO COX model. The formula of the risk score was:

$$\begin{aligned}\text{RiskScore} &= \text{coef}_1 \times \text{ICAG}_1 \\ &\quad + \text{coef}_2 \times \text{ICAG}_2 + \text{coef}_3 \times \text{ICAG}_3 + \cdots\end{aligned}$$

where the ICAG represents the normalized expression of a given ICAG, and the coef represents its coefficient in the LASSO COX model. Besides, a nomogram was also developed for convenient prediction by using the R package regplot (version 1.1).

Samples in the discovery set were divided into high- and low-risk groups according to whether their risk scores exceeded the median value. The Kaplan–Meier analysis with a log-rank test was applied to assess the prognostic difference between the two risk groups. Principal component analysis (PCA) was also performed to demonstrate and visualize the differences between the two groups. Then, the risk score and clinical

variables were assessed by successively fitting the univariate and multivariate COX regression models on the GSE31210 dataset of the discovery set. It is noteworthy, that only the GSE31210 was used because it provided the most detailed clinical information. Statistical significance ($P < 0.05$) in both univariate and multivariate COX regression indicated that the risk score is an independent prognostic factor for patients with LUAD. Similarly, the validation set was also divided based on its own median value of risk scores, and the clinical significance of the risk score was also assessed on it.

Prediction of Cancer Recurrence by Using Machine Learning

A machine-learning model called Categorical Boosting (CatBoost) was developed for the tumor recurrency prediction to further elucidate the ability of ICAGs to predict clinical outcomes. CatBoost, one of the gradient boosting algorithms, iteratively trains a weak decision tree to fit residuals of previous trees (Prokhorenkova et al., 2018). It is a powerful machine-learning technique but has yet not been widely adopted in critical care research (Zhao et al., 2020). In the present study, the ability of CatBoost to predict cancer recurrency was studied and clinical potential of ICAGs was further demonstrated. Four datasets that provided relapse information were firstly selected, including GSE68465, GSE30219, GSE50081, and GSE31210 datasets. Then, three different CatBoost models were trained based on three different feature sets. The first feature set includes only ICAGs selected by LASSO COX regression, the second one includes only clinical information such as age, gender, and stage, while the third set combined the first and the second feature sets.

Ten-fold cross-validation was performed considering the limited sample size of the utilized datasets. In particular, the dataset was randomly into 10 subsets. In each iteration, nine of them were used to train the models and the last one for validation. After 10 iterations, each subset had been validated and the validation results were then combined. The areas under receiver operator characteristic curves (AUROCs) were calculated to assess the performance of the models. Finally, the SHapley Additive exPlanation (SHAP) values were calculated according to a game theory approach to illustrate the effects of each feature on the prediction results of the third model (Lundberg et al., 2020).

Gene Set Variant, Pathway Activity, and Regulatory Network Analyses

Gene Set Cancer Analysis (GSCA[1]) is an integrated genomic and immunogenomic online tool for gene set cancer research based on TCGA cohorts (Liu et al., 2018). The results of single-nucleotide variants (SNVs), copy number variations (CNVs), micro RNA (miRNA) network analyses, and pathway activity were obtained by uploading on the web-based platform the genes selected by LASSO COX regression and choosing the TCGA-LUAD cohort. Notably, SNV and CNV were analyzed based on the TCGA-LUAD cohort, while pathway activity and miRNA network analyses were performed on the 32 and 33 cancer types in TCGA, respectively. Additionally, the correlation of gene expression and drug sensitivity was assessed based on small molecules from the Cancer Therapeutics Response Portal (CTRP) (Seashore-Ludlow et al., 2015) and the Genomics of Drug Sensitivity in Cancer (GDSC) (Iorio et al., 2016).

The cBioPortal for Cancer Genomics[2] provides another web-based resource for exploring, visualizing, and analyzing multidimensional cancer genomics data (Gao et al., 2013). We also used cBioPortal to explore selected ICAGs on the TCGA-LUAD cohort. Results of variant and pathway analyses were downloaded to enhance our study.

Additionally, long non-coding RNAs (lncRNAs) in the validation set were identified according to Genome Reference Consortium Human Build 38 patch release 13 (GRCh38.p13). Co-expression analysis was conducted by assessing Pearson correlation between the selected ICAGs and lncRNAs in LUAD samples. A lncRNA regulatory network was derived according to the criteria of $|\text{Correlation Coefficient}| > 0.4$ and $P < 0.01$ using the functions provided in the stats R package (version 3.6.0). If a ICAG have more than 10 significantly correlated lncRNAs, only 10 lncRNAs with the greatest absolute value of correlation coefficients were selected. The lncRNA network was then visualized by using the Cytoscape program.

DNA Methylation and N6-Methyladenosine

In order to further analyze the selected ICAGs, DNA methylation data of LUAD (platform: Illumina HumanMethylation450 BeadChip) were downloaded from the TCGA database. The methylation level of CpGs was represented as β values (Bibikova et al., 2011). Pearson correlation coefficients and *P*-values were calculated between expression and methylation levels of ICAGs.

Besides, 12 N6-methyladenosine (m6A) regulatory genes were obtained *via* systematic review in published articles. The expression levels of these genes were compared by using the two-sample Wilcoxon test between the high- and the low-risk groups on the validation set.

Gene Set Enrichment and Immunogenomic Landscape Analyses

The DEGs between the high- and the low-risk groups in the TCGA cohort, with adjusted *P*-value < 0.01 were identified using the functions provided in the stats R package (version 3.6.0). Gene set enrichment analyses based on the GO and KEGG functional and pathway terms were conducted to assess the DEGs with adjusted *P*-values threshold of 0.05, using the clusterProfiler R package (version 3.14.3) (Yu et al., 2012).

Computational methods were used to evaluate the immune infiltration and functions, including TIMER (Li et al., 2020), quanTIseq (Finotello et al., 2019), xCell (Aran et al., 2017), MCP-counter (Becht et al., 2016), EPIC (Racle et al., 2017), CIBERSORT (Newman et al., 2015) CIBERSORTx (Newman et al., 2019), and single-sample gene set enrichment analysis

[1]http://bioinfo.life.hust.edu.cn/GSCA/#/

[2]http://cbioportal.org

TABLE 1 | Basic information of datasets included in this study.

Dataset ID	Number of samples	Brief introduction about the dataset	Number of deaths	Number of relapses
Discovery set	1042			
GSE19188	40	A genome-wide gene expression analysis on early-stage NSCLC	24	–
GSE30219	85	Identification of a group of metastatic-prone tumors in lung cancer according to "Off-context" gene expression defined by the authors	45	27 (83)
GSE31210	226	Gene expression analysis on pathological stage I–II lung adenocarcinomas	35	64 (226)
GSE31546	16	Development of an EGFR mutation gene expression signature to predict response and clinical outcome, and identification of genes associated with the EGFR-dependent phenotype	2	–
GSE37745	106	Biomarker discovery in NSCLC	77	–
GSE50081	127	Validation of a histology-independent prognostic gene signature for early-stage NSCLC, including stage IA patients	51	37 (124)
GSE68465	442	Gene expression-based survival prediction in LUAD	236	178 (178)
Validation set	535			
TCGA-LUAD	535	The LUAD cohort of TCGA, a landmark cancer genomics program, molecularly characterized over 20,000 primary cancer and matched normal samples spanning 33 cancer types.	187	–

(ssGSEA) in an attempt to comprehensively analyze the immune differences between the two groups (Rooney et al., 2015). Additionally, a list of 79 immune checkpoint genes was obtained from Hu's study (Hu et al., 2020), with most of these genes being ligands, receptors, or important molecules in the immune checkpoint pathways. The expression of these genes was compared between the high- and the low-risk groups by using the two-sample Wilcoxon test.

Modeling the Intercellular Communication Based on a Single Cell RNA Sequencing Dataset

The GSE131907 dataset, which is a LUAD single-cell RNA sequencing dataset, was downloaded from the GEO database. Raw data are not available due to patient privacy concerns, and therefore, data normalized ($\log_2$TPM) by the contributors were used in our study. Cell annotations were provided by contributors. A total of 11 LUAD and 11 distant normal lung samples were included for further analyses. The expression levels of genes included in the proposed risk score were evaluated in different cell types.

Risk scores were calculated at the cell level and averaged for each LUAD sample. Then, 11 LUAD samples were divided into high- and low-risk groups according to the median risk score. Intercellular communication was modeled by using the CellPhoneDB Python package (version 2.1.7) (Efremova et al., 2020), and the significantly differentiated between the two groups ligand–target links were summarized by using the iTALK R package (version 0.1.0) (Wang Y. et al., 2019). Specifically, CellPhoneDB integrates existing datasets of cellular communication and new manually reviewed information, including the subunit architecture for both ligands and receptors. The normalized gene expression data and the cell annotations were analyzed by the relevant Python package, with subsampling, 50 rounds of iterations and 4 calculating threads. The iTALK R package is another useful toolkit for characterizing and visualizing intercellular communication. Growth factor, cytokine, checkpoint, and other types of intercellular communication were assessed by it. The top 20 ligand–target links with the greatest differences between the high- and the low-risk groups were visualized.

RESULTS

Identification of Prognosis-Related ICAGs

Seven datasets downloaded from GEO were preprocessed as previously described, and a total of 1042 samples were eventually used as the discovery set. The results of eliminating batch effect were presented (see **Supplementary Figure 1**). Besides, 594 samples were downloaded from TCGA and used as the validation set, with 535 of them being LUAD samples and 59 of them being normal samples. The basic information of these datasets was summarized in **Table 1**. Besides, the characteristics of research subjects in each GEO dataset and the TCGA dataset were presented in **Supplementary File 2**. A total of 426 ICAGs were collected, and 354 genes, common in all datasets, were assessed in this study. Sixty-seven genes were significantly associated with the overall survival by univariate COX regression (see **Supplementary File 4**). Eight genes were finally selected because they presented non-zero coefficients in the fitted LASSO COX regression models, as it is shown in **Table 2**. The risk score was calculated as follows:

$$\begin{aligned}\text{RiskScore} &= 0.09022 \times \text{LAMB1} + 0.09287 \times \text{GJC1} + 0.12437 \\ &\times \text{CDH4} + 0.16105 \times \text{GJB3} + 0.14827 \times \text{SPP1} \\ &+ 0.14339 \times \text{AFDN} + 0.16016 \times \text{SKAP2} - 0.17981 \\ &\times \text{ARHGEF6}\end{aligned}$$

TABLE 2 | The prognostic ICAGs identified by using LASSO COX regression.

Gene	Coef	Encoded protein
ARHGEF6	−0.17981	Rho guanine nucleotide exchange factor 6
SKAP2	0.16016	Src kinase-associated phosphoprotein 2
AFDN	0.14339	Afadin, adherens junction formation factor
SPP1	0.14827	Secreted phosphoprotein 1
GJB3	0.16105	Gap junction beta-3 protein
CDH4	0.12437	Cadherin 4
GJC1	0.09287	Gap junction gamma-1 protein
LAMB1	0.09022	Laminin subunit beta-1

In addition, the nomogram we developed was presented in **Supplementary Figure 2**.

Survival Analysis on the Risk Score

Patients in the discovery and the validation sets were divided into high- and low-risk groups, according to whether their risk scores exceeded the median values. The Kaplan–Meier analysis with a log-rank test demonstrated that there were significant differences in the overall survival between the two risk groups. As shown in **Figures 2A,D**, the high-risk group has significantly worse overall survival compared to the low-risk group ($P < 0.001$). PCA also confirmed that patients in the two groups presented different patterns of gene expression and this finding is illustrated in **Figures 2B,E**. As seen, the blue points representing the low-risk group distributed together, while the red points representing the high-risk group in another part of space. This indicates significant differences in the gene expression levels between the two groups. Furthermore, as shown in **Figures 2C,F**, our risk score is significantly associated with the outcome in the univariate and the multivariate regression ($P < 0.001$), indicating the risk score was an independent predictive factor for the overall survival.

Prediction of Relapse by Using Machine Learning

Four datasets which provided relapse information were selected, including GSE68465, GSE30219, GSE50081, and GSE31210. The Kaplan–Meier analysis also proved that there were significant differences in the disease-free survival between the high- and the low-risk groups, as it is shown in **Figure 3A**.

The CatBoost algorithm was used to develop three machine-learning models based on different sets of variables to further elucidate the prognostic value of selected ICAGs. The first feature set includes only ICAGs selected by the LASSO COX regression and the model presented an AUROC of 0.663. The second includes only clinical information such as age, gender, and stage, and its AUROC was 0.757. The third set combined both clinical variables and the selected ICAGs, and a remarkable AUROC of 0.841 was achieved. The predictive performances of cancer stage and age were also assessed. The AUROCs of stage and age were 0.728 and 0.557, respectively. The machine-learning model, based on all features, outperformed all other predictive methods or models, as it is shown in **Figure 3B**.

SHapley Additive exPlanation values were assessed to evaluate the effects of each variable on the model's output. SHAP values for the model using all features were shown in **Figure 3C**. Red color represents a high value of that feature, while blue color represents a low value. A positive SHAP value means that this feature value will increase the relapse risk, while a negative one represents a protective effect. The features in **Figure 3C** were ordered from top to bottom according to their importance, which was assessed by the average absolute SHAP values. Moreover, the relationships between the SHAP values and the gene expression levels were visualized in **Figure 3D**. SKAP2, AFDN, CDH4, GJB3, and GJC1 had similar positive correlation with SHAP values, while the expression level of ARHGEF6 is negatively correlated with SHAP values. The relationships between SHAP values and SPP1 or LAMB1 are not simply linear and needs more research.

Gene Set Variant Analysis on the Validation Set

Gene Set Cancer Analysis was used for gene set variant analyses. **Supplementary Figure 3** showed the analysis on SNVs of the eight ICAGs. Missense mutations were the most common variants, and the C > A and C > T SNVs were the most frequent variants. The median of variants per sample was 1. LAMB1, ARHGEF6, and CDH4 are the top mutated genes.

Analysis of CNVs was summarized in **Supplementary Figure 4**. From this figure, it is observed that LAMB1, GJC1, CDH4, and SKAP2 had frequent heterozygous amplification, while AFDN, SPP1, and GJB3 had frequent heterozygous deletion. Homozygous variants were less frequently observed in these genes, but CDH4, SKAP2, LAMB1, and AFDN had more frequent homozygous variants than other genes.

Regulatory Networks and Methylation Modification

Micro RNA and lncRNA regulatory networks were visualized in **Supplementary Figures 5A,B**, respectively. The figures included key ICAGs, as well as the miRNAs and lncRNAs that target them. Methylation modification was summarized in **Supplementary Figure 6**. Five genes were shown in **Supplementary Figure 6A**, including CDH4, GJB3, LAMB1, SKAP2, and SPP1, of which expression levels were significantly correlated to their DNA methylation levels (P-value < 0.05). Besides, the expression levels of m6A regulatory genes were compared between the high- and the low-risk groups, as shown in **Supplementary Figure 6B**.

Pathway Activity and Gene Set Enrichment Analyses

The results of pathway activity analysis were shown in **Supplementary Figure 7**. As shown in **Supplementary Figure 7B**, these genes were significantly correlated with the EMT activation. Besides, the eight genes also presented great effects on the inhibition of cell cycle. These results confirmed that the selected genes play an important role in cancer development and metastasis.

Expressions of 25,168 genes were compared between the high- and the low-risk groups in the validation set, and

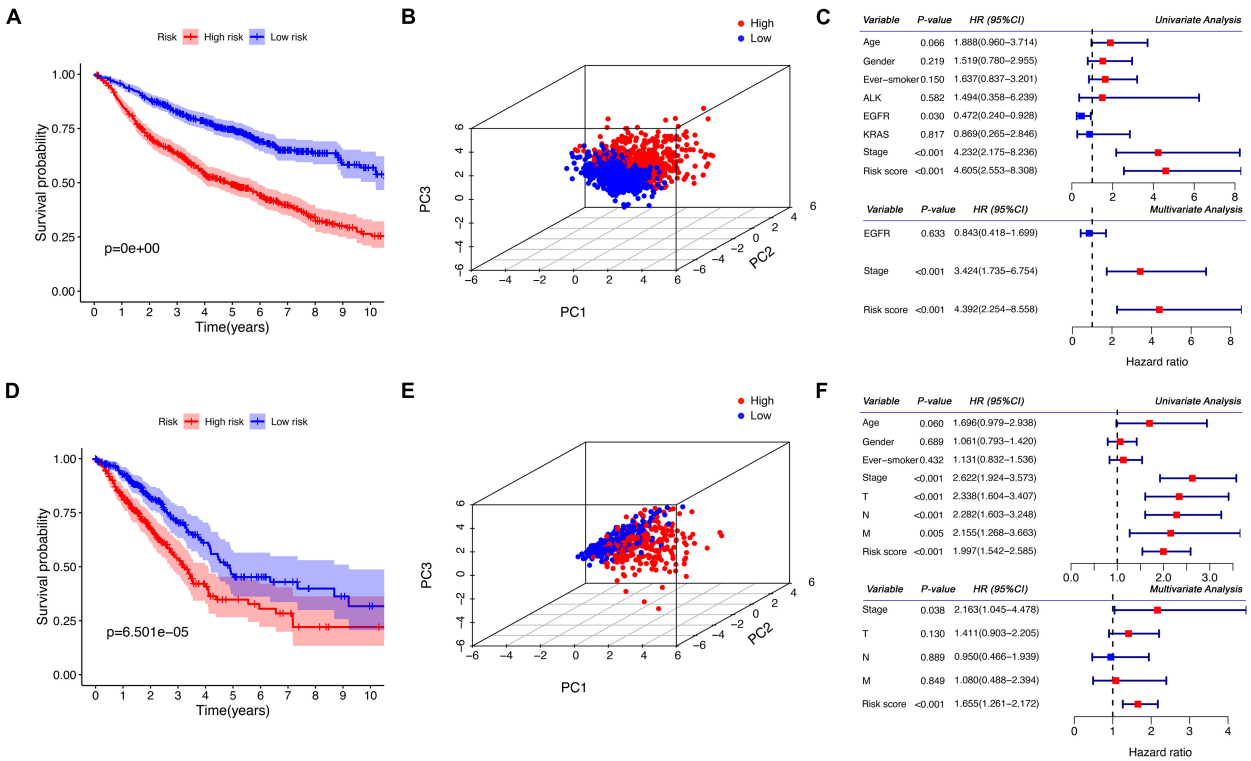


FIGURE 2 | Prognostic analysis of the risk score in the discovery and the validation sets. **(A,D)** Kaplan–Meier curves for the comparison of the overall survival between the high- and the low-risk group. **(B,E)** The three-dimensional principal component analysis on the two groups. **(C,F)** Forest plots for the results of the univariate and multivariate COX regression analyses regarding the overall survival. **(A–C)** Show the results of the discovery set, while **(D–F)** show the results of the validation set. PC, principal component; HR, hazard ratio; CI, confidence interval.

10,830 genes were found to be differentially expressed with an adjusted P-value < 0.05. Gene set enrichment analyses were performed. DNA replication, chromosomal region, cell adhesion molecule binding and cadherin binding were significantly enriched GO terms, as shown in **Supplementary Figures 8A,B**. Cell cycle, focal adhesion, spliceosome, and homologous recombination were significantly enriched pathways according to the results of the KEGG pathway analysis (**Supplementary Figures 8C,D**). Among these DEGs, 991 had absolute fold changes greater than 2. Specifically, 727 genes were upregulated and 264 were downregulated in the high-risk group, compared to the low-risk group. The heatmap and volcano plot of these genes were plotted in **Supplementary Figures 8E,F**, respectively.

Immunogenomic Landscape Analyses

Various computational approaches regarding immune infiltration were conducted and summarized in **Figure 4**. As seen, most scores of immune cells were correlated with the risk score with a negative coefficient as shown in **Figure 4B**. This finding suggests that the high-risk group had fewer infiltrated immune cells. The scores provided by CIBERSORTx were compared between the two groups, as it is shown in **Figure 4C**. As seen, the high-risk group had fewer B cells naïve, B cells memory, T cells CD8, T cells regulatory, and Mast cells resting than the low-risk group. But more T cells CD4 memory activated, Macrophages M0, and Macrophages M1 were infiltrated in the high-risk group. More Eosinophils were infiltrated in the high-risk group, but the scores in both groups were too low such that the comparison of Eosinophils was not clear in the figure. Additionally, the expressions of 79 immune checkpoint genes were compared between the high- and the low-risk groups by using the two-sample Wilcoxon test. The checkpoint genes with a P-value < 0.05 were summarized in **Figure 4D**.

Modeling of Intercellular Communication Based on a Single Cell RNA Sequencing Dataset

The Single-cell RNA sequencing dataset GSE131907 was used for further analysis in high resolution. A total of 11 LUAD and 11 distant normal lung samples were included in our study. In total, 42,679 normal cells and 45,149 tumor cells were assessed. A total of 22,977 and 22,172 tumor cells were assigned to high- and low-risk groups, respectively. The risk score and the expression levels of eight genes in different cells were assessed and compared, as demonstrated in **Figure 5A**. As seen, the expression patterns of these eight genes were different between the high- and the low-risk groups and also between the LUAD and normal groups. SPP1 was upregulated in all kinds of cells

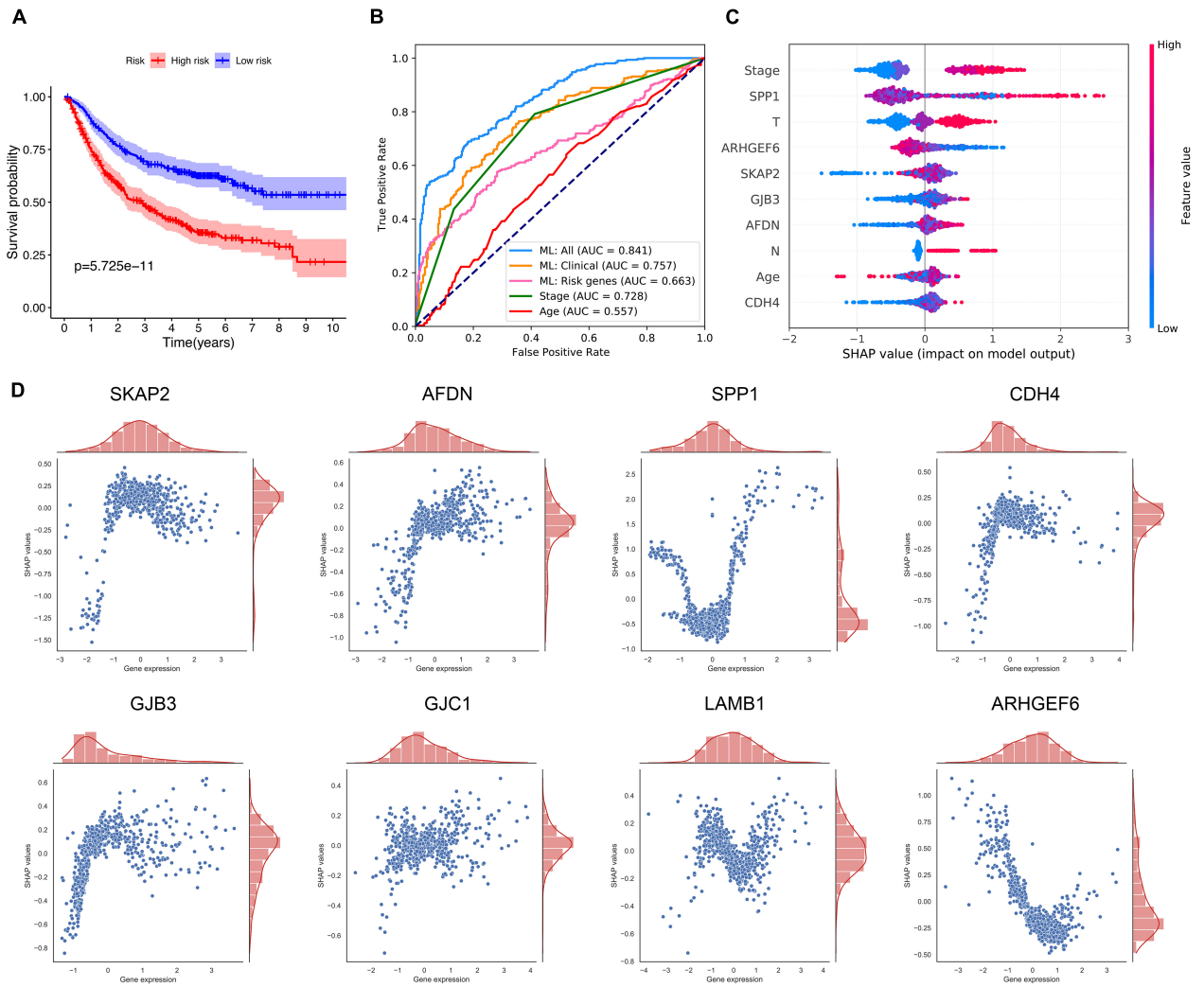


FIGURE 3 | Prediction of LUAD relapse by using machine learning. **(A)** The Kaplan–Meier curve for the comparison of the disease-free survival between the high- and the low-risk groups. **(B)** The receiver operator characteristic curves for the explored predictive models or methods. **(C)** SHAP values show the variable impacts on the prediction results of the model using all clinical information and eight ICAGs. **(D)** The distributions of the SHAP values for each ICAG on the model using all clinical information and the eight ICAGs. ML, machine learning; AUC, area under curves; SHAP, shapley additive explanation.

and especially the myeloid cells in the high-risk group compared to the low-risk group. AFDN and SKAP2 were upregulated in the epithelial cells of the high-risk group. These differences were also observed between tumor and normal samples. The correlation of risk scores between different cells was displayed in **Figure 5B**. Interestingly, the risk scores of various cells were positively correlated with each other.

Besides, intercellular communication was modeled by using CellphoneDB. The patterns of intercellular communication in the high- and low-risk groups were visualized in **Figures 5C,D**, respectively. As shown in **Figures 5C,D**, less communication was observed between B lymphocytes, T lymphocytes, NK cells, and MAST cells than between other cells, but in the high-risk group (shown in **Figure 5D**), there was more communication between epithelial cells and others in comparison with the low-risk group (shown in **Figure 5C**).

Furthermore, the differences in communication patterns between the two groups were compared by using the iTALK R package, and the results were displayed in **Figures 5E–G**. In **Figure 5G**, red color represents a gain of interaction, indicating upregulation of the ligand and the receptor genes, while blue color represents a loss of interaction, indicating downregulation of the ligand and the receptor genes. The thickness of edges indicates the expression level of the ligands, while the size of arrows indicates the expression level of receptors. Although some ligand–receptor links of autocrine were expressed more in epithelial cells, several ligand–receptor links were expressed significantly less in the high-risk group.

Drug Sensitivity Analysis

The correlation of gene expression and drug sensitivity was assessed based on small molecules from CTRP and GDSC. The

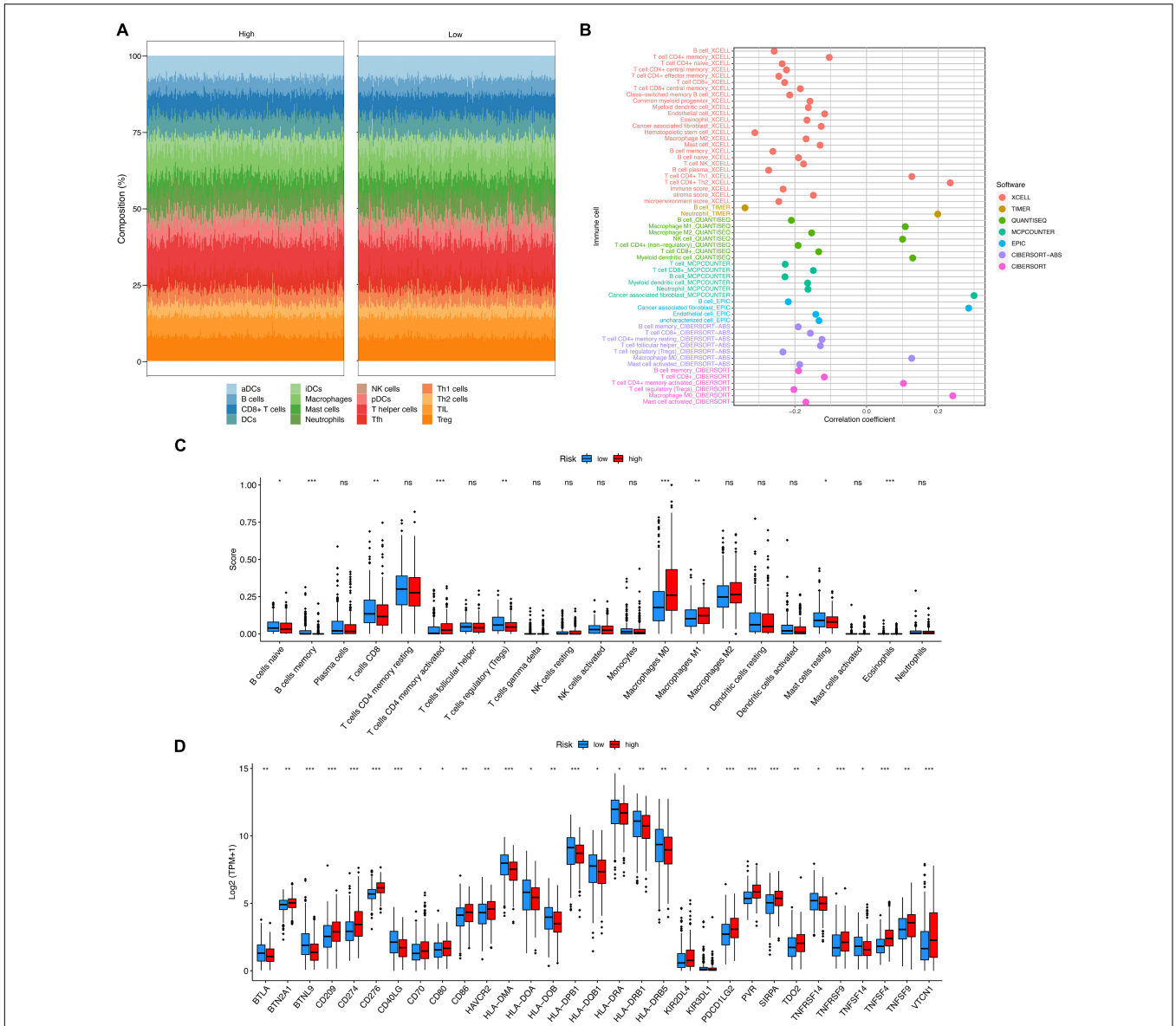


FIGURE 4 | The immunogenomic landscape analyses on the validation set. **(A)** Compositions of different infiltrated immune cells estimated by ssGSEA. **(B)** Correlation between risk score and immune cells estimated by XCELL, TIMER, quanTIseq, MCP-counter, EPIC, and CIBERSORT. **(C)** Boxplot for the immune cell scores in the high- and the low-risk groups, estimated by CIBERSORTx. **(D)** Boxplot for the expression of an immune checkpoint gene in the high- and the low-risk groups. Adjusted P-values were showed as: ns, not significant; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

correlation coefficients and adjusted *P*-values were visualized in **Supplementary Figure 9**. It is easily observed that these eight genes were significantly correlated to the sensitivity of various drugs. Specifically, the positive correlations were commonly observed on AFDN, GJB3, LAMB1, and SPP1, while ARHGEF6 had negative correlations to most drugs.

DISCUSSION

In this study, we explored the prognostic potential of ICAGs in patients with LUAD. An eight-ICAG signature was eventually identified, and a risk score was accordingly derived based on the analysis of publicly available datasets. Machine-learning models were developed to predict tumor recurrency based on clinical information and the expression levels of selected ICAGs. Comprehensive analyses were conducted, including gene set variant, regulatory network, pathway activity, gene set enrichment, immunogenomic landscape, drug sensitivity analyses as well as modeling of intercellular communication based on single-cell RNA sequencing.

A large number of studies have shown that cell-to-cell communication participates in the construction of the TME of lung cancer, promotes the formation of lung cancer blood

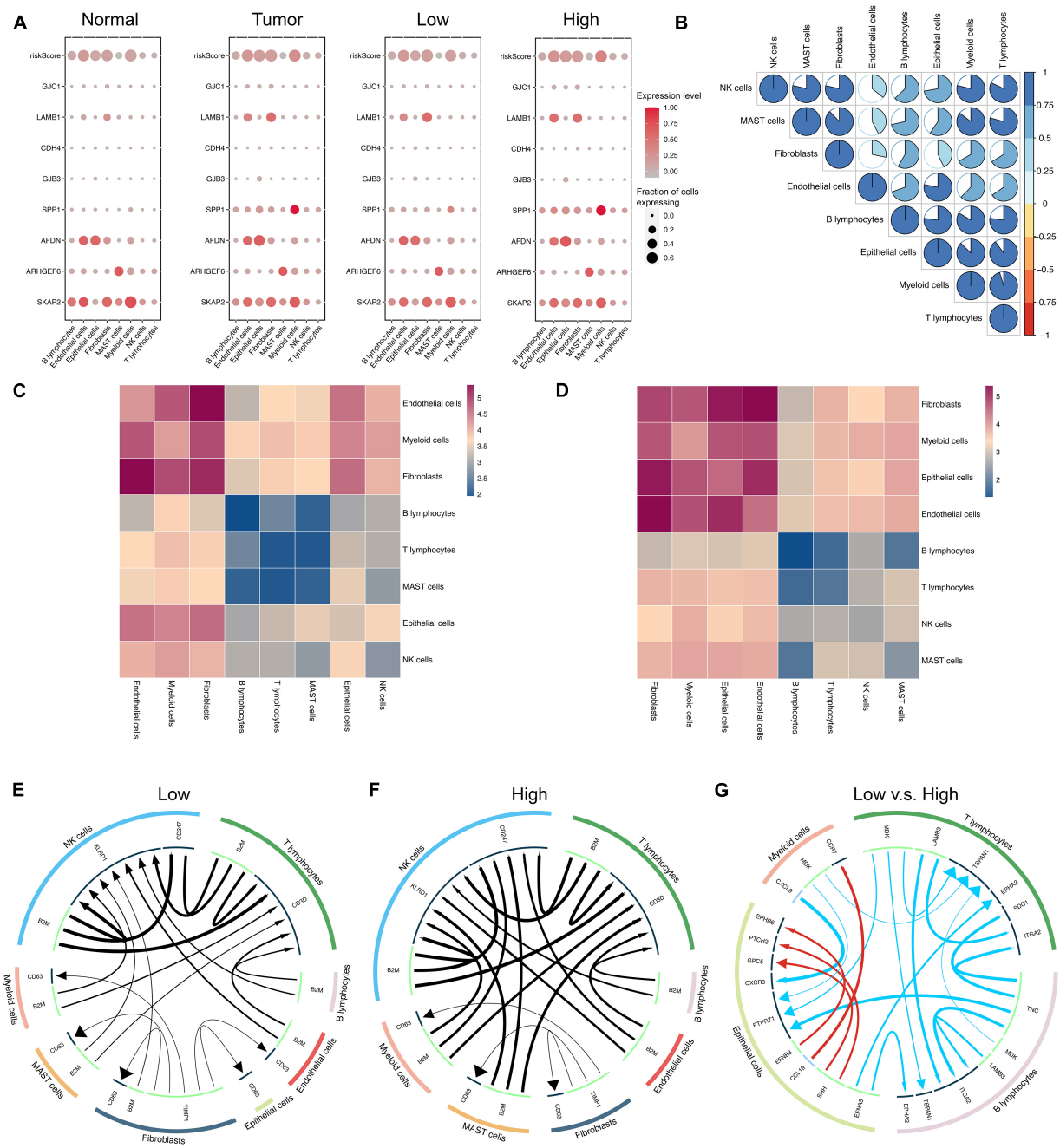


FIGURE 5 | Modeling of intercellular communication based on a single cell RNA sequencing dataset. **(A)** Heatmap for the ICAG expression and the risk score in different groups. **(B)** Correlation of the risk scores in different cells. **(C,D)** Heatmaps for the intercellular communication generated by using CellphoneDB in the low-risk group **(C)** and the high-risk group **(D)**. **(E,F)** The main ligand–target expression between the different cells in the low-risk group **(E)** and the high-risk group **(F)**. **(G)** The top differences in the expression of ligand or target genes between the low- and the high-risk groups. **(E–G)** were generated by using the iTALK R package.

vessels, and accelerates tumor invasion and metastasis. Exosomes, cytokines, etc., can be released by tumor cells into the TME and blood circulation to promote tumor progression. For example, MALAT1 derived from exosomes has been found to be highly expressed in the serum of patients with non-small cell lung cancer, which can accelerate tumor migration and promote its growth (Gutschner et al., 2013). Another study showed that the absorption of vesicles of lung cancer cells by macrophages promotes the production of M2-like phenotype by tumor-associated macrophages, which in turn produces IL-1β, which is beneficial to the survival of tumor cells (Wang et al., 2011). In short, cell-to-cell communication can regulate the progression,

metastasis, invasion, and proliferation of lung cancer through a variety of ways.

In this study, eight genes involved in intercellular communication were identified and prioritized in the present study. Previous research has reported the role some of them play in the progression of cancer. It was found that miR-135a can inhibit cancer stem cell-driven medulloblastoma development by directly repressing the expression of ARHGEF6 (Hemmesi et al., 2015) and ARHGEF6 might be a hub gene in colorectal cancer (Wang and Zheng, 2014). A high level of Afadin, which is the protein encoded by AFDN gene, was found to be associated with poor survival in breast cancer patients (Tabaries et al., 2019). GJB3, a member of the connexin gene family, was found to be a potential circulating biomarker for metastatic pancreatic cancer and might have a unique effect on cell death (Tattersall et al., 2009; Easton et al., 2019). Choi et al. (2020) detected the expression levels of GJC1 (Cx45) in HeLa cells and identified GJC1 as a major component of gap junctions. However, few studies focused on the role of these genes in LUAD. The findings in our study will contribute to a deeper understanding of the effects of these genes on the progression and relapse of LUAD.

The effects of other selected genes on LUAD or NSCLC have been previously reported by several studies. It was found that SPP1 may contribute to immune escape (Zhang et al., 2017), metastatic progression (Chiou et al., 2019), and second-generation epidermal growth factor receptor tyrosine kinase inhibitor resistance (Wang X. et al., 2019). Tanaka et al. (2016) revealed that SKAP2 is related to tumor-associated macrophage infiltration and facilitates the metastatic progression of lung cancer in mise. The prognostic value of SKAP2 was also reported in previous studies using bioinformatics analyses (Kuranami et al., 2015; Tanaka et al., 2016; Chen et al., 2019). Studies showed that CDH4 could be regulated by miR-211-5p to inhibit the proliferation, migration, and invasion of LUAD (Zhang et al., 2020). Besides, the aberrant expression of ligand–receptor pair LAMB1-ITGB1 was identified within tumor cells in LUAD (Chen et al., 2020). In our study, the selected ICAGs except ARHGEF6 presented a positive coefficient, indicating that the upregulation of the expression levels results in poor prognosis. Our study confirms the results of these prior studies and may facilitate other research on the functions of these genes.

The prognostic value of these eight ICAGs was evaluated by survival analysis. Significant differences were observed in the overall survival between the high- and the low-risk groups. The risk score derived by LASSO COX regression was proved to be an independent predictive factor for the overall survival in LUAD. Moreover, a machine-learning model based on clinical information and expression of the eight ICAGs accurately predicts LUAD recurrence better than other predictive methods or models. Our risk score and model could contribute to the determination of the severity of LUAD and to stratify patients' prognosis.

Further comparisons between the high- and the low-risk groups were performed, including gene set enrichment and immune infiltration analyses. GO and KEGG pathway analyses demonstrated that cell cycle, focal adhesion, DNA replication, and cell adhesion molecule binding were enriched by DEGs between the high- and the low-risk groups. Besides, it was shown by immune infiltration analyses that there were significant differences in the TME between the two groups. Previous studies have already reported that misleading communication within and between tumor cells and immune cells contributes to immune escape, metastatic progression, and drug resistance of LUAD (Tanaka et al., 2016; Chen et al., 2020). In addition, as shown in **Supplementary Figure 7**, these eight genes are also involved in cell growth cycle pathways such as cell cycle pathway and DNA damage response pathway. Sex hormone receptor pathways are also related to these eight genes, such as hormone AR pathway and hormone ER pathway. In addition, PI3K/AKT, RTK, RAS/MAPK, TSC/mTOR pathways can also interact to promote the occurrence and development of lung cancer. In our study, these enriched pathways and microenvironment differences were likely to result from aberrant changes in the intercellular communication. More research and experiments are required to shed light on the effects of aberrant intercellular communication in cancers.

A single-cell RNA sequencing dataset was used to further model the intercellular communication in LUAD. The ICAG expressions in different cells were assessed in the present study, evaluating the communication relationships between different cells and comparing the differences in the communication patterns between the high- and the low-risk groups. Changes in transferring information may be the key mechanism in immune escape and therapy resistance of LUAD. Our study provides insight into the potentially therapy target role of the ICAGs in LUAD.

According to drug sensitivity analysis, we found that these eight genes are related to AS605240 (PI3K inhibitor), AZD8055 (mTOR inhibitor), AZD-7762 (cell cycle checkpoint kinase), vinblastine (a lung cancer targeted drug), and other drugs. The results are consistent with the results of our previous analysis. This further supports the results of the previous pathway analysis and the conclusion of the article.

Several limitations of this study should be considered. Firstly, the analyses of this study were conducted based on public datasets, without verification or validation from *in vitro* or *in vivo* biochemical experiments. Thus, the revealed eight-ICAG signature and our machine-learning models require further validation in large-scale prospective studies to demonstrate their robustness. Secondly, various approaches to estimating immune infiltrated cells or modeling intercellular communication were used in the present study, but their results were not entirely consistent. However, any one of these computational approaches is not a "Gold Standard." In contrast, they provide different perspectives to estimate what we are interested in. That is exactly the reason why we tried as many approaches as possible in the study, instead of drawing our conclusion based on anyone of them. Lastly, other factors, such as circular RNAs and proteins, involved in intercellular communication, were not included in our study. Multi-omics data may facilitate a deeper understanding of the pathogenesis and optimize the prediction of survival in LUAD.

CONCLUSION

In this study, we comprehensively assessed the role of ICAGs in LUAD, identifying eight key ICAGs with prognostic value and developing a risk score as well as machine learning models to predict the prognosis for patients with LUAD. These genes may contribute to understanding the pathological mechanism of LUAD, and could also be considered as potential therapeutic targets.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: TCGA data were available on the project website at https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga, while the GEO datasets were available at https://www.ncbi.nlm.nih.gov/geo/.

ETHICS STATEMENT

The study was an analysis of third-party anonymized publicly available datasets with pre-existing institutional review board (IRB) approvals.

AUTHOR CONTRIBUTIONS

Q-YZ and L-PL: conception, design, data analysis, and interpretation. RG and Y-WL: administrative support. Q-YZ: collection and assembly of data. All authors: writing and final approval of manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (Nos. 81573091 and 81802668), the Natural Science Foundation of Hunan Province (Nos. 2018JJ3776 and 2017JJ3467), and the Fundamental Research Funds for the Central Universities of Central South University under Grant (No. 2020zzts892).

ACKNOWLEDGMENTS

We would like to acknowledge TCGA and GEO databases for providing their platforms and those contributors for uploading their valuable datasets. Besides, we would like to express their gratitude to EditSprings (https://www.editsprings.com/) for the expert linguistic services provided.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.702424/full#supplementary-material

Supplementary File 1 | Supplementary Figures 1–9.

Supplementary File 2 | The baseline characteristics of research subjects in each GEO dataset and the TCGA dataset.

Supplementary File 3 | List of intercellular communication-associated genes included in the study.

Supplementary File 4 | The results of univariate COX regression regarding overall survival on the discovery set.

REFERENCES

Ahluwalia, P., Ahluwalia, M., Mondal, A. K., Sahajpal, N., Kota, V., Rojiani, M. V., et al. (2021). Immunogenomic gene signature of cell-death associated genes with prognostic implications in lung cancer. *Cancers (Basel).* 13:155. doi: 10.3390/cancers13010155

Aran, D., Hu, Z., and Butte, A. J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 18:220.

Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 17:218.

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J. M., et al. (2011). High density DNA methylation array with single CpG site resolution. *Genomics* 98, 288–295. doi: 10.1016/j.ygeno.2011.07.007

Brucher, B. L., and Jamall, I. S. (2014). Cell-cell communication in the tumor microenvironment, carcinogenesis, and anticancer treatment. *Cell Physiol. Biochem.* 34, 213–243. doi: 10.1159/000362978

Chen, Y. J., Chang, W. A., Wu, L. Y., Huang, C. F., Chen, C. H., and Kuo, P. L. (2019). Identification of novel genes in osteoarthritic fibroblast-like synoviocytes using next-generation sequencing and bioinformatics approaches. *Int. J. Med. Sci.* 16, 1057–1071. doi: 10.7150/ijms.35611

Chen, Z., Yang, X., Bi, G., Liang, J., Hu, Z., Zhao, M., et al. (2020). Ligand-receptor interaction atlas within and between tumor cells and T cells in lung adenocarcinoma. *Int. J. Biol. Sci.* 16, 2205–2219. doi: 10.7150/ijbs.42080

Chiou, J., Chang, Y. C., Tsai, H. F., Lin, Y. F., Huang, M. S., Yang, C. J., et al. (2019). Follistatin-like protein 1 inhibits lung cancer metastasis by preventing proteolytic activation of osteopontin. *Cancer Res.* 79, 6113–6125. doi: 10.1158/0008-5472.can-19-0842

Choi, E. J., Palacios-Prado, N., Saez, J. C., and Lee, J. (2020). Identification of Cx45 as a major component of GJs in HeLa cells. *Biomolecules* 10:1389.

Easton, J. A., Albuloushi, A. K., Kamps, M. A. F., Brouns, G., Broers, J. L. V., Coull, B. J., et al. (2019). A rare missense mutation in GJB3 (Cx31G45E) is associated with a unique cellular phenotype resulting in necrotic cell death. *Exp. Dermatol.* 28, 1106–1113. doi: 10.1111/exd.13542

Efremova, M., Vento-Tormo, M., Teichmann, S. A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* 15, 1484–1506. doi: 10.1038/s41596-020-0292-x

Finotello, F., Mayer, C., Plattner, C., Laschober, G., Rieder, D., Hackl, H., et al. (2019). Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med.* 11:34.

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

Gao, X., Tang, M., Tian, S., Li, J., and Liu, W. (2021). A ferroptosis-related gene signature predicts overall survival in patients with lung adenocarcinoma. *Future Oncol.* 17, 1533–1544. doi: 10.2217/fon-2020-1113

Gutschner, T., Hammerle, M., and Diederichs, S. (2013). MALAT1 – a paradigm for long noncoding RNA function in cancer. *J. Mol. Med. (Berl).* 91, 791–801. doi: 10.1007/s00109-013-1028-y

Hemmesi, K., Squadrito, M. L., Mestdagh, P., Conti, V., Cominelli, M., Piras, I. S., et al. (2015). miR-135a inhibits cancer stem cell-driven medulloblastoma

development by directly repressing Arhgef6 expression. *Stem Cells* 33, 1377–1389. doi: 10.1002/stem.1958

Hu, F. F., Liu, C. J., Liu, L. L., Zhang, Q., and Guo, A. Y. (2020). Expression profile of immune checkpoint genes and their roles in predicting immunotherapy response. *Brief. Bioinform.* 22:bbaa176. doi: 10.1093/bib/bbaa176

Iorio, F., Knijnenburg, T. A., Vis, D. J., Bignell, G. R., Menden, M. P., Schubert, M., et al. (2016). A landscape of pharmacogenomic interactions in cancer. *Cell* 166, 740–754.

Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., et al. (2020). The reactome pathway knowledgebase. *Nucleic Acids Res.* 48, D498–D503.

Kuranami, S., Yokobori, T., Mogi, A., Altan, B., Yajima, T., Onozato, R., et al. (2015). Src kinase-associated phosphoprotein2 expression is associated with poor prognosis in non-small cell lung cancer. *Anticancer Res.* 35, 2411–2415.

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, T., Fu, J., Zeng, Z., Cohen, D., Li, J., Chen, Q., et al. (2020). TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* 48, W509–W514.

Liu, C. J., Hu, F. F., Xia, M. X., Han, L., Zhang, Q., and Guo, A. Y. (2018). GSCALite: a web server for gene set cancer analysis. *Bioinformatics* 34, 3771–3772. doi: 10.1093/bioinformatics/bty411

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., et al. (2020). From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* 2, 56–67. doi: 10.1038/s42256-019-0138-9

Ma, C., Luo, H., Cao, J., Zheng, X., Zhang, J., Zhang, Y., et al. (2020). Identification of a novel tumor microenvironment-associated eight-gene signature for prognosis prediction in lung adenocarcinoma. *Front. Mol. Biosci.* 7:571641. doi: 10.3389/fmolb.2020.571641

Maacha, S., Bhat, A. A., Jimenez, L., Raza, A., Haris, M., Uddin, S., et al. (2019). Extracellular vesicles-mediated intercellular communication: roles in the tumor microenvironment and anti-cancer drug resistance. *Mol. Cancer* 18:55.

Mittelbrunn, M., and Sanchez-Madrid, F. (2012). Intercellular communication: diverse structures for exchange of genetic information. *Nat. Rev. Mol. Cell Biol.* 13, 328–335. doi: 10.1038/nrm3335

Mo, Z., Yu, L., Cao, Z., Hu, H., Luo, S., and Zhang, S. (2020). Identification of a hypoxia-associated signature for lung adenocarcinoma. *Front. Genet.* 11:647. doi: 10.3389/fgene.2020.00647

Molina, J. R., Yang, P., Cassivi, S. D., Schild, S. E., and Adjei, A. A. (2008). Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin. Proc.* 83, 584–594. doi: 10.4065/83.5.584

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

Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* 37, 773–782. doi: 10.1038/s41587-019-0114-2

Parri, M., Ippolito, L., Cirri, P., Ramazzotti, M., and Chiarugi, P. (2020). Metabolic cell communication within tumour microenvironment: models, methods and perspectives. *Curr. Opin. Biotechnol.* 63, 210–219. doi: 10.1016/j.copbio.2020.03.001

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., and Gulin, A. (eds.) (2018). "CatBoost: unbiased boosting with categorical features," in *Proceedings of the Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, MontrÃľal, Canada*, 6639–6649. Available online at: https://proceedings.neurips.cc/paper/2018

Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E., and Gfeller, D. (2017). Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife* 6:e26476.

Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 160, 48–61. doi: 10.1016/j.cell.2014.12.033

Seashore-Ludlow, B., Rees, M. G., Cheah, J. H., Cokol, M., Price, E. V., Coletti, M. E., et al. (2015). Harnessing connectivity in a large-scale small-molecule sensitivity dataset. *Cancer Discov.* 5, 1210–1223. doi: 10.1158/2159-8290.cd-15-0235

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

Tabaries, S., McNulty, A., Ouellet, V., Annis, M. G., Dessureault, M., Vinette, M., et al. (2019). Afadin cooperates with Claudin-2 to promote breast cancer metastasis. *Genes Dev.* 33, 180–193. doi: 10.1101/gad.319194.118

Tan, Q., Huang, Y., Deng, K., Lu, M., Wang, L., Rong, Z., et al. (2020). Identification immunophenotyping of lung adenocarcinomas based on the tumor microenvironment. *J. Cell Biochem.* 121, 4569–4579. doi: 10.1002/jcb.29675

Tanaka, M., Shimamura, S., Kuriyama, S., Maeda, D., Goto, A., and Aiba, N. (2016). SKAP2 promotes podosome formation to facilitate tumor-associated macrophage infiltration and metastatic progression. *Cancer Res.* 76, 358–369. doi: 10.1158/0008-5472.can-15-1879

Tattersall, D., Scott, C. A., Gray, C., Zicha, D., and Kelsell, D. P. (2009). EKV mutant connexin 31 associated cell death is mediated by ER stress. *Hum. Mol. Genet.* 18, 4734–4745. doi: 10.1093/hmg/ddp436

Wang, R., Zhang, J., Chen, S., Lu, M., Luo, X., Yao, S., et al. (2011). Tumor-associated macrophages provide a suitable microenvironment for non-small lung cancer invasion and progression. *Lung Cancer* 74, 188–196. doi: 10.1016/j.lungcan.2011.04.009

Wang, X., Zhang, F., Yang, X., Xue, M., Li, X., Gao, Y., et al. (2019). Secreted phosphoprotein 1 (SPP1) contributes to second-generation EGFR tyrosine kinase inhibitor resistance in non-small cell lung cancer. *Oncol. Res.* 27, 871–877. doi: 10.3727/096504018x15426271404407

Wang, Y., Wang, R., Zhang, S., Song, S., Jiang, C., Han, G., et al. (2019). iTALK: an R package to characterize and illustrate intercellular communication. *bioRxiv[Preprint]:507871*.

Wang, Y., and Zheng, T. (2014). Screening of hub genes and pathways in colorectal cancer with microarray technology. *Pathol. Oncol. Res.* 20, 611–618. doi: 10.1007/s12253-013-9739-5

Xu, R., Rai, A., Chen, M., Suwakulsiri, W., Greening, D. W., and Simpson, R. J. (2018). Extracellular vesicles in cancer - implications for future improvements in cancer care. *Nat. Rev. Clin. Oncol.* 15, 617–638. doi: 10.1038/s41571-018-0036-9

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118

Zhang, M., Zhu, K., Pu, H., Wang, Z., Zhao, H., Zhang, J., et al. (2019). An immune-related signature predicts survival in patients with lung adenocarcinoma. *Front. Oncol.* 9:1314. doi: 10.3389/fonc.2019.01314

Zhang, S. J., Ma, J., Wu, J. C., Hao, Z. Z., Zhang, Y. N., and Zhang, Y. J. (2020). CircRNA EPB41L2 inhibits tumorigenicity of lung adenocarcinoma through regulating CDH4 by miR-211-5p. *Eur. Rev. Med. Pharmacol. Sci.* 24, 3749–3760.

Zhang, Y., Du, W., Chen, Z., and Xiang, C. (2017). Upregulation of PD-L1 by SPP1 mediates macrophage polarization and facilitates immune escape in lung adenocarcinoma. *Exp. Cell Res.* 359, 449–457. doi: 10.1016/j.yexcr.2017.08.028

Zhao, Q. Y., Liu, L. P., Luo, J. C., Luo, Y. W., Wang, H., Zhang, Y. J., et al. (2020). A machine-learning approach for dynamic prediction of sepsis-induced coagulopathy in critically Ill patients with sepsis. *Front. Med. (Lausanne).* 7:637434. doi: 10.3389/fmed.2020.637434

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

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 01 September 2021
doi: 10.3389/fgene.2021.706814

A Nomogram Integrating Ferroptosis- and Immune-Related Biomarkers for Prediction of Overall Survival in Lung Adenocarcinoma

Mengyu Chai[1†], Xiuchun Li[1†], Yaxin Zhang[1], Yemeng Tang[1], Pingping Shu[1], Jing Lin[1], Keqing Shi[2], Liangxing Wang[1] and Xiaoying Huang[1*]*

[1]Division of Pulmonary Medicine, Key Laboratory of Heart and Lung, The First Affiliated Hospital of Wenzhou Medical University, Zhejiang, China, [2]Translational Medicine Laboratory, The First Affiliated Hospital of Wenzhou Medical University, Zhejiang, China

OPEN ACCESS

Edited by:
Zhijie Xu,
Central South University, China

Reviewed by:
Lingxiao Zhou,
Shenzhen University, China
Chuan Hu,
Qingdao University Medical College, China

***Correspondence:**
Xiaoying Huang
zjwzhxy@126.com
Liangxing Wang
wzyxywlx@163.com

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

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *15 May 2021*
Accepted: *04 August 2021*
Published: *01 September 2021*

Citation:
Chai M, Li X, Zhang Y, Tang Y, Shu P, Lin J, Shi K, Wang L and Huang X (2021) A Nomogram Integrating Ferroptosis- and Immune-Related Biomarkers for Prediction of Overall Survival in Lung Adenocarcinoma. Front. Genet. 12:706814. doi: 10.3389/fgene.2021.706814

Ferroptosis plays a dual role in cancer, which is known to be affected to antitumor immune responses. However, the association between ferroptosis and antitumor immune responses is uncertain in lung adenocarcinoma (LUAD). In this work, 38 ferroptosis-related genes (FRGs) and 429 immune-related genes (IRGs) were identified as being differentially expressed between tumor and normal samples. Two risk score formulas consisting of seven FRGs and four IRGs, respectively, were developed by Lasso-penalized Cox regression and verified in the GSE13213 dataset. The CIBERSORT algorithm was used to estimate the relative abundance of immune cells in tumors. The correlation between FRGs and immune cells was evaluated using the TIMER database. The results indicated that the development of ferroptosis was synergistic with that of anti-tumor immunity in LUAD. The concordance index and calibration curves showed that the performance of a nomogram that combines clinical staging and risk scores is superior to that of models using a single prognostic factor. In conclusion, ferroptosis might be synergistic with anti-tumor immunity in LUAD. The combined nomogram could reliably predict the probability of overall survival of LUAD patients. These findings may be useful for future investigation of prognostic value and therapeutic potential related to ferroptosis and tumor immunity in LUAD.

Keywords: lung adenocarcinoma, ferroptosis, immune, bioinformatics analysis, nomogram, overall survival

INTRODUCTION

Lung cancer is the most common type of cancer and the leading cause of cancer-related death worldwide. An estimated 1.8 million deaths from lung cancer occurred in 2020 (Sung et al., 2021). Lung adenocarcinoma (LUAD) is the most prevalent subtype of lung cancer, accounting for about 40% (Ma et al., 2013). Although there are many therapeutic options for LUAD, overall prognosis remains poor (Herbst et al., 2018). Therefore, it is necessary to further explore factors related to survival time that may contribute to the development of more effective treatment methods for lung adenocarcinoma.

Ferroptosis is an iron-dependent cell death modality marked by the oxidative modification of the phospholipid membrane (Stockwell et al., 2017). The effect of ferroptosis on tumor growth is not clear and requires further investigation. On the one hand, ferroptosis inhibits tumor growth in mice (Badgley et al., 2020). For example, although pancreatic cancer cells are prone to resisting chemotherapy, they are highly sensitive to artemisinin-induced ferroptosis (Efferth, 2017). On the other hand, ferroptosis can promote tumor growth (Dai et al., 2020). Therefore, examining the association between ferroptosis marker genes and survival may increase understanding of the role played by ferroptosis in LUAD.

Moreover, there seems to be an inextricable link between ferroptosis and immunity. Levels of $CD8^+$ and $CD4^+$ T cells cannot increase if they lack glutathione peroxidase 4 (GPX4), a key regulator of ferroptosis (Matsushita et al., 2015). Induction of ferroptosis is related to release of PGE2 (Yang et al., 2014), which attenuates antitumor immunity by affecting cDC1s and NK cells (Wang and DuBois, 2015). The synergistic effect of ferroptosis and immune regulation might not only inhibit the primary tumor but also stimulate immune responses in combination with immune checkpoint blockade (Li and Rong, 2020). In a word, FRGs and immunity have a profound impact on each other, and it is necessary to do more research to understand their relationship.

In this work, we constructed two formulas for calculating ferroptosis-related and immune-related risk scores using The Cancer Genome Atlas (TCGA) database. LUAD patients from the Gene Expression Omnibus (GEO) database were used for validation. Some immune cells had the same infiltration trend as risk score increased in two prognostic multigene signatures. In addition, a nomogram combining tumor stage, ferroptosis-related risk score, and immune-related risk score was developed for more accurate prediction of patient survival. Finally, the correlation between ferroptosis-related genes (FRGs) and immune-related genes (IRGs) was assessed.

MATERIALS AND METHODS

Data Acquisition

Gene expression data, including 497 tumor samples and 54 normal samples, in the HTSeq-FPKM format were obtained from the TCGA database using the GDC tool.[1] We also downloaded clinical data at the GDC portal. Meanwhile, the microarray dataset GSE13213 consisting of 117 patients with LUAD was downloaded from the GEO database[2] as an external validation dataset. Probes were annotated based on annotation files. In total, 123 FRGs and 2,483 IRGs were obtained from the FerrDb database[3] and ImmPort Resources[4] (**Supplementary Tables S1**, **S2**), respectively. There was no ethical conflict since all data was obtained from public databases.

Construction of A Predictive Nomogram

The R/Bioconductor package "limma," which provides an integrated solution for analyzing gene expression data and is a popular choice for gene discovery through differential expression analyses, was used to identify differentially expressed FRGs and IRGs. The results were visualized in a volcano map using the popular drawing packages "ggplot2" and "ggrepel." $|\log2\ \text{foldchange}| > 1$ and $\text{FDR} < 0.05$ were considered significant. Prognosis-related genes were identified using univariate Cox proportional hazards regression analysis ($p < 0.05$). "Venn" R package is a common way to compare different datasets and identify and visualize intersections between sets. It was used to identify intersections between differentially expressed genes (DEGs) and prognosis-related genes as candidates for risk scoring.

To reduce the risk of overfitting, we applied LASSO-Cox regression to construct survival-predicting models. "Glmnet" is the most widely used package for LASSO analysis, and can generate a variety of models, including binary and multinomial logistic regression models, Poisson models, Cox proportional hazards models and SVM models. The risk score was calculated using "glmnet" and "survival" packages. The Cox coefficient and expression levels of the prognosis-related DEGs were extracted to calculate risk scores. The formula was as follows: $\text{Risk score} = \sum_{i=1}^{N}\left(\text{Coei} * \text{Expi}\right)$, where N represents gene number, Coei represents coefficient value and Expi represents gene expression level. Based on the median risk score, LUAD patients were dichotomized into low- and high-risk groups. Considering the significant batch effect between the TCGA and GEO dataset, we used the median value of each to divide the high and low risk groups. The predictive ability of the model was evaluated using the log-rank test and receiver operating characteristic (ROC) analysis. The t-distributed stochastic neighbor embedding (t-SNE) test was implemented in "Rtsne" package to visualize clustering. Principal component analysis (PCA) was completed using the prcomp function. A nomogram was constructed to predict overall survival (OS) based on the results of multivariate Cox regression. The concordance index (C-index) was calculated to assess the stability of the nomogram by 1,000 bootstrap replicates. The performance of the prognostic nomogram was evaluated by plotting calibration curves.

Functional Annotation Analysis and Evaluation of Immune Cell Infiltration

In order to identify different pathways between the two groups, DEGs among groups were analyzed using R package

[1]https://portal.gdc.cancer.gov/repository
[2]https://www.ncbi.nlm.nih.gov/geo/

[3]http://www.zhounan.org/ferrdb/
[4]https://www.immport.org

Abbreviations: LUAD, Lung adenocarcinoma; FRGs, Ferroptosis-related genes; IRGs, Immune-related genes; GPX4, Glutathione peroxidase 4; TCGA, The Cancer Genome Atlas; GEO, The Gene Expression Omnibus; NSCLC, Non-small cell lung cancer; DEGs, Differentially expressed genes; ROC, Receiver operating characteristic; t-SNE, t-Distributed stochastic neighbor embedding; PCA, Principal component analysis; OS, Overall survival; C-index, Concordance index; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; LUSC, Lung squamous cell carcinoma.

"clusterProfiler" for Gene Ontology (GO). The Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis was performed by the same procedure.

CIBERSORT is a gene-based deconvolution algorithm to estimate the abundance of any of 22 human immune cell types. We loaded "e1071" package to execute this algorithm to quantify the distribution of types of infiltrating immune cells in lung adenocarcinoma samples. The Wilcox test was conducted to identify disparities in the infiltration levels of immune cells between different risk groups, including B cells, plasma cells, T cells, natural killer cells, monocytes, macrophages, dendritic cells, mast cells, eosinophils, and neutrophils. Inter-group differences were identified by "limma" package and displayed in Violin plots using "Vioplot" package. The TIMER web server[5] is a comprehensive resource for analysis of immune infiltrates in multiple cancers. We analyzed the association between gene expression and abundance of infiltrating immune cells using gene modules.

qRT-PCR Verification

Human bronchial epithelial cells (BEAS-2B) and human LUAD cell lines (PC9 and H1299) were purchased from the American Type Culture Collection (ATCC, United States). They were cultured in RPMI-1640 medium (Gibco, China) or high glucose Dulbecco's Modified Eagle's media (DMEM; Hyclone, Logan, UT, United States) supplemented with 10% fetal bovine serum (Gibco, China) at 37°C in an atmosphere of 5% CO_2.

Total RNA was extracted using the RNeasy mini kit (Qiagen, United States) and reverse-transcribed with the iScript cDNA Synthesis Kit. qRT-PCR was performed using the CFX96 Real-Time System (Bio-Rad, Hercules, CA, United States). All samples were tested in triplicate. Primers were purchased from *Ribobio* (Guangzhou, China) and are listed in **Supplementary Table S3**.

Statistical Analysis

The Wilcoxon rank-sum test was used to compare the gene expression in tumor and normal tissues. Risk scores were validated as independent prognostic factors by the application of Cox regression. Pearson correlation analysis was implemented to assess the correlation of gene expression levels. The statistical analysis tool used in this study was R software 4.0.2. All statistical tests were two-tailed. A value of $p < 0.05$ was considered statistically significant.

RESULTS

Calculation of Risk Scores for FRGs and IRGs

The study flow chart is shown in **Figure 1**. Compared with normal samples, 38 significant ferroptosis-related DEGs (26 upregulated and 12 downregulated) and 429 significant immune-related DEGs (224 upregulated and 205 downregulated) were identified (**Figures 2A,B**; **Supplementary Tables S4, S5**). By univariate Cox regression analysis, 16 FRGs and 51 IRGs were identified as being related to prognosis (**Figures 2C,D**; **Supplementary Tables S6**, **S7**). Then the intersections of DEGs and prognostic-related genes were identified. Finally, 9 overlapping candidate FGRs and 4 overlapping candidate IRGs were obtained (**Figures 2E,F**). Heatmaps showed expression of the 13 genes (**Figures 2G,H**).

We calculated the risk score of FRGs according to the following formula: risk score = (0.048216 × SLC7A11) + (0.161996 × DDIT4) + (0.004447 × SLC7A5) + (−0.068488 × GDF15) + (−0.061102 × IL33) + (0.027458 × SLC2A1) + (0.140388 × RRM2). A total of 464 patients were divided into high-risk ($N = 232$) and low-risk ($N = 232$) groups with the median as the cutoff value. The patients in the low-risk group had significantly higher survival rates than those in the high-risk group (left panel of **Figure 3A**). The DEGs between the two groups were mainly related to cell cycle transition, mitosis, protein synthesis, and chemotaxis (**Supplementary Table S8**). The signaling pathways obtained by KEGG analysis were not only correlated with cell division and maturation, but also correlated with immunity (**Supplementary Table S9**). In the ROC analysis, the AUCs in the first, second and third year were 0.681, 0.658, and 0.684, respectively (left panel of **Figure 3B**). PCA and t-SNE analysis showed that patients were clustered into distinct groups (**Figures 3C,D**). In the univariate Cox regression analyses, the risk score was significantly associated with OS in the FRG prognostic risk model (HR = 3.617, 95% CI = 2.346–5.577, $p < 0.001$; upper panel of **Figure 3E**). Multivariate Cox regression revealed that the classifier was an independent prognostic factor (HR = 3.150, 95% CI = 2.037–4.872, $p < 0.001$; upper panel of **Figure 3F**). After applying the risk score formula to the GSE13213 dataset, significant differences in OS between the two groups still existed (right panel of **Figures 3A,B**). Moreover, the risk score remained an independent prognostic factor in the GSE13213 dataset (bottom panel of **Figures 3E,F**).

The prognostic risk score of IRGs = (−0.043721 × HLA−DRB5) + (−0.055446 × SFTPD) + (−0.053414 × PTGDS) + (−0.003929 × S100B). As shown in the left panel of **Figure 4A**, patients with high-risk scores tended to have shorter survival times than those with low-risk scores. AUCs in the first, second, and third-year reached 0.679, 0.603, and 0.610, respectively (left panel of **Figure 4B**). Patients were also distributed in 2 regions (**Figures 4C,D**). In addition, in the univariate Cox regression analysis, the IRGs prognostic risk score was a variable closely related to prognosis (HR = 3.859, 95% CI = 2.002–7.440, $p < 0.001$; upper panel of **Figure 4E**). After multivariate Cox regression analysis of multiple factors, the IRGs prognostic risk model remained a reliable independent prognostic factor (HR = 3.818, 95% CI = 1.928–7.563, $p < 0.001$; upper panel of **Figure 4F**). The validity of the model was also verified in the GSE13213 dataset (right panel of **Figures 4A,B**; bottom panel of **Figures 4E,F**).

[5]https://cistrome.shinyapps.io/timer/

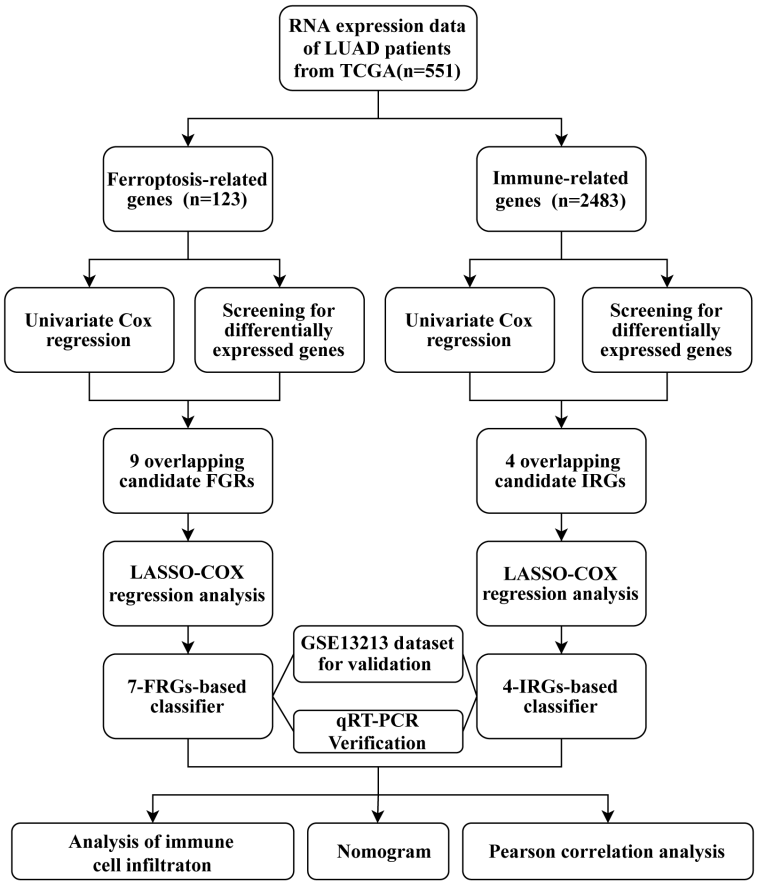


FIGURE 1 | Workflow chart.

Correlation of Immune Cell Infiltration With Risk Scores and Gene Expression

To better understand the relationship between risk score and immune response, we calculated the proportions of 22 immune cells (**Figure 5A**). In addition, the correlation matrix of immune cells showed that the infiltration level of activated memory $CD8^+$ T cells was highly correlated with $CD4^+$ T cells (**Figure 5B**). In addition, the level of resting mast cells highly correlated with that of monocytes was also high. Subsequently, immune cell infiltration was compared between high and low-risk groups. There were similar trends of differences in infiltration levels for memory B cells, activated memory $CD4^+$ T cells, resting memory $CD4^+$ T cells, monocytes, M0 macrophages, resting mast cells, and resting dendritic cells in the 2 models (**Figure 5C**). Next, the relationship between gene expression and immune cell subtype infiltration was further analyzed in the TIMER database (**Figures 5D,E**). SLC7A11 and IL33 showed a strong correlation with immune cell infiltration. In contrast to SLC7A11, IL33 promoted the infiltration of nearly all types of immune cell. All IRGs were positively correlated with the level of immune cell infiltration (**Supplementary Figure S1**).

Construction of the Nomogram

The above results above showed that tumor stage, FRGs risk score, and IRGs risk score were independent prognostic factors in LUAD. The c-indices of the stage, ferroptosis, immune and combined models were 0.663, 0.644, 0.616, and 0.712, respectively (**Figure 6A**). Therefore, the combined model was selected for prediction of 1-, 3-, and 5-year OS rates (**Figure 6B**). Calibration plots showed that the combined model performed well in predicting 1- and 3-year survival but not 5-year survival (**Figure 6C**). Taken together, compared with models established using a single prognostic factor, the combined model was superior for short-term survival prediction, which might be beneficial to diagnosis and treatment.

Relationship Between the FRGs and the IRGs in the Prognostic Risk Models

To identify interactions between FRGs and IRGs, we performed a correlation analysis of gene expression (**Figure 7A**). Interleukin 33 (IL33) was synergistically co-expressed with HLA-DRB5, SFTPD, PTGDS, and S100B, and their high expression levels helped to prolong patient survival (**Figure 7B**). The association

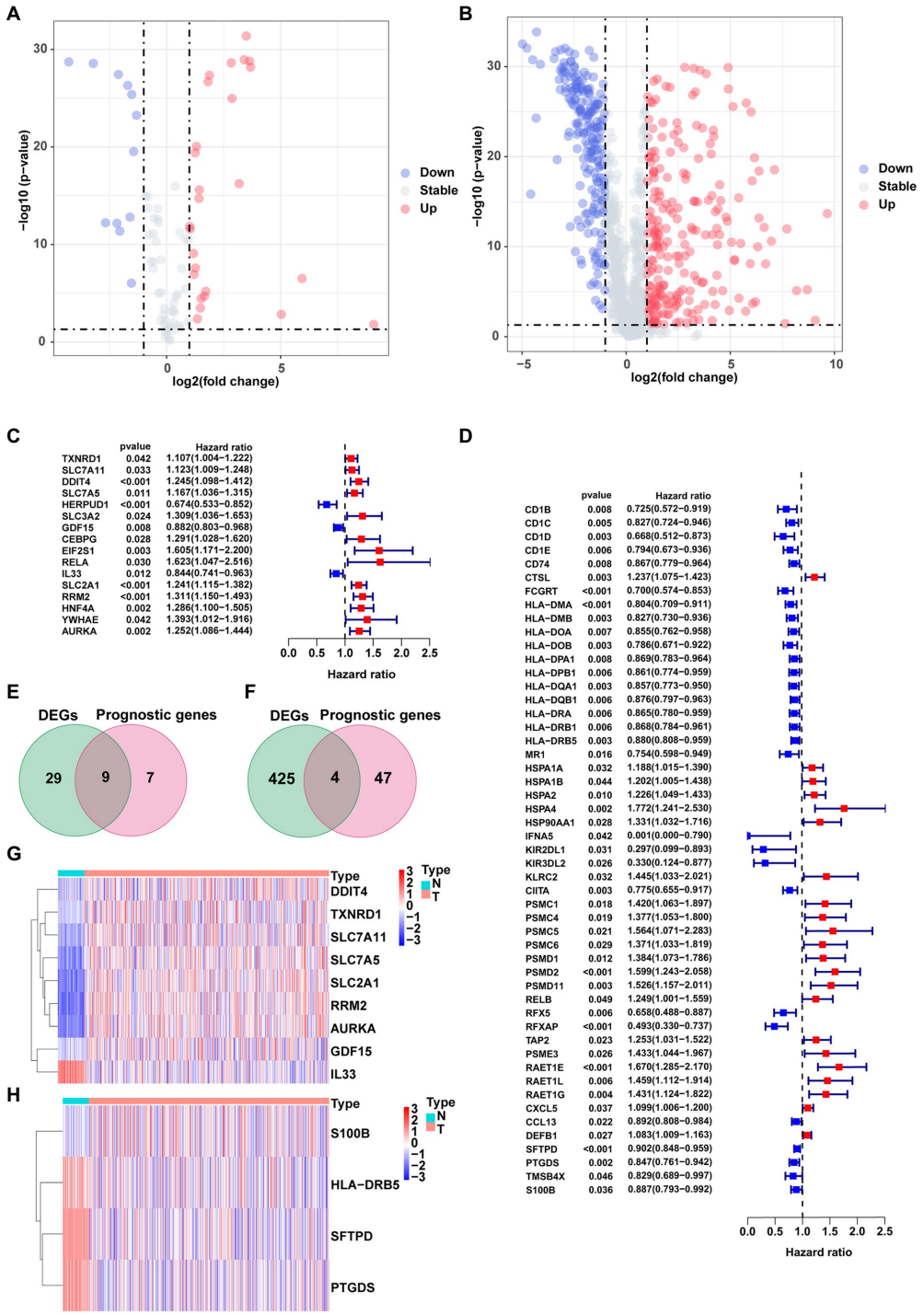


FIGURE 2 | Identification of prognostic ferroptosis-related DEGs and immune-related DEGs. **(A)** Volcano plot of ferroptosis-related DEGs. **(B)** Volcano plot of Immune-related DEGs. **(C)** Forest plot from the univariate Cox proportional hazards regression analysis between FRG expression and survival time. **(D)** Forest plot from the univariate Cox proportional hazards regression analysis between IRG expression and survival time. **(E,F)** Venn diagrams of overlapping candidate genes. **(G,H)** Heatmap plots of overlapping candidate genes.

between IL33 and prostaglandin D2 synthase (PTGDS) was the strongest among all genes, followed by IL33 and surfactant protein D (SFTPD). All the other FRGs were negatively correlated with IRGs. This result was consistent with their opposite effects on survival time. Similar results were observed in the TIMER databases (**Figure 7C**).

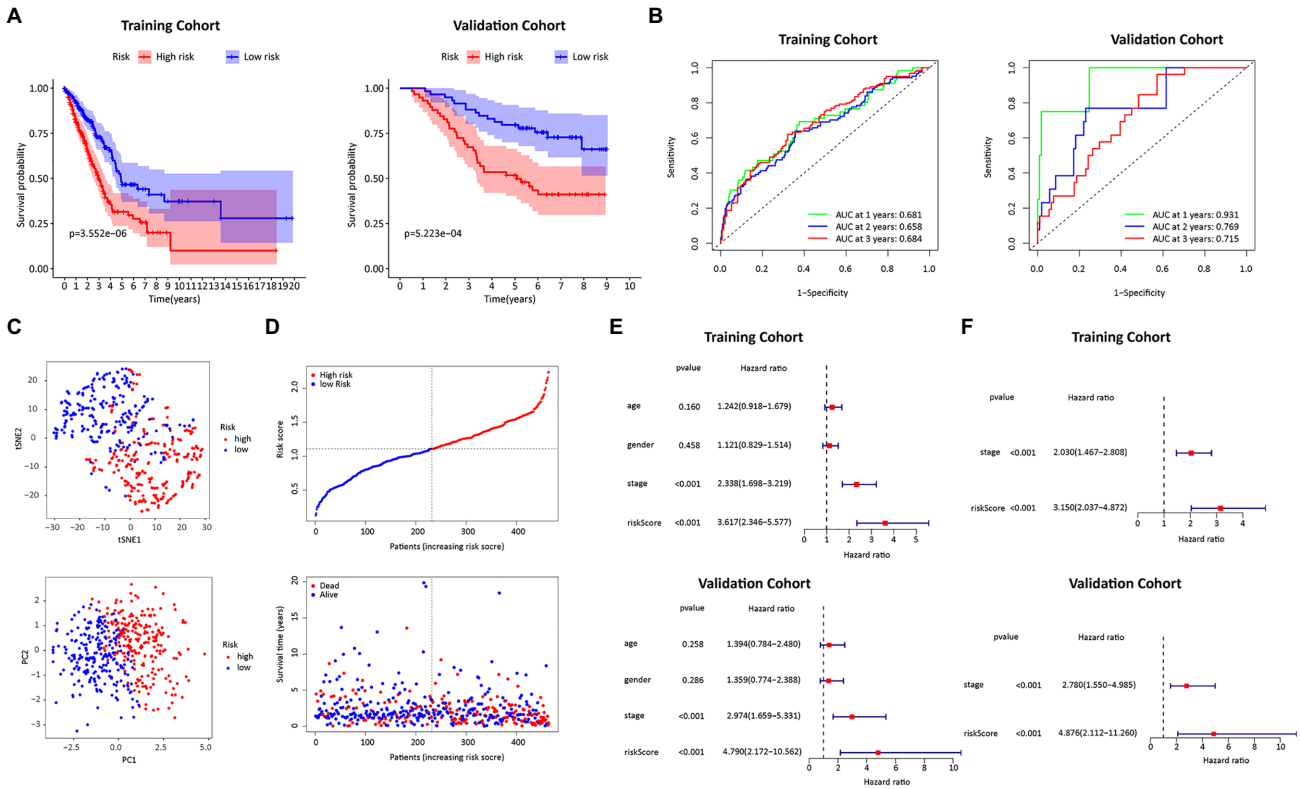


FIGURE 3 | Construction and validation of the FRGs risk score formula. **(A)** Kaplan Meier survival curves for the TCGA training cohort and the GEO validation cohort. **(B)** ROC curves of the FRGs risk score formula in the training and validation cohorts. **(C)** T-SNE analysis and PCA plot of the training cohort. **(D)** The risk and survival status plots in the training cohort. **(E)** Forest plot from the univariate Cox regression analysis in the training and validation cohort. **(F)** The results of multivariate analysis regarding OS in the training and validation cohorts.

Validation of Gene Expression by qRT-PCR

We detected the levels of 7 FRGs and 4 IRGs in the BEAS-2B, PC9, and H1299 cell lines by using qRT-PCR (**Figure 8**). SLC7A11, DDIT4, SLC7A5, GDF15 SLC2A1, and HLA-DRB5 were highly expressed in LUAD cell lines. The mRNA levels of IL33, RRM2, and SFTPD were downregulated in LUAD cell lines. PTGDS was highly expressed in the H1299 cell line but low in PC9 cell line. The above results were generally consistent with the TCGA results, which indicates that our bioinformatics analysis was credible.

DISCUSSION

Non-small cell lung cancer (NSCLC) is one of the deadliest form of cancers. LUAD and lung squamous cell carcinoma (LUSC) are the predominant histological phenotypes of NSCLC, and they exhibit significant differences in morphologic differentiation, underlying drivers, and response to various therapies (Wilkerson et al., 2010; Cancer Genome Atlas Research Network, 2014). In recent years, the incidence of LUAD has significantly increased compared with LUSC (Kinoshita et al., 2016). The possible benefits and potential harm of inducing ferroptosis during treatment are receiving increasing attention. Cell death in LUAD in response to treatment with siramesine and lapatinib has been reported to be mediated by ferroptosis (Villalpando-Rodriguez et al., 2019). Additionally, triggering ferroptosis increased sensitivity to radiotherapy in human patient-derived models of LUAD (Ye et al., 2020). This suggests that new therapies that combine immunotherapy with regulation of ferroptosis may lead to improved outcomes. Therefore, studies exploring the interaction between ferroptosis and immunity in LUAD may have far-reaching implications.

The ferroptosis-related risk score calculation formulas used here contained seven FRGs. The qRT-PCR results showed that five genes were upregulated and two genes were downregulated. Except for RRM2, the other results were consistent with TCGA results. SLC7A11 is overexpressed in multiple types of cancer (Shi et al., 2019), and suppressing transcription and protein expression of SLC7A11 can effectively induce ferroptosis in cancer cells (Chang et al., 2018). DDIT4 has potential not only as a prognostic biomarker (Ho et al., 2020) but also as a therapeutic target (Wang et al., 2015). Paradoxically, upregulation of DDIT4 may promote cellular ferroptosis (Dixon et al., 2014), and therefore its specific role in cancer merits further exploration. Inhibition of SLC7A5 can regulate amino acid transport and affect ferroptosis (Dixon et al., 2014).

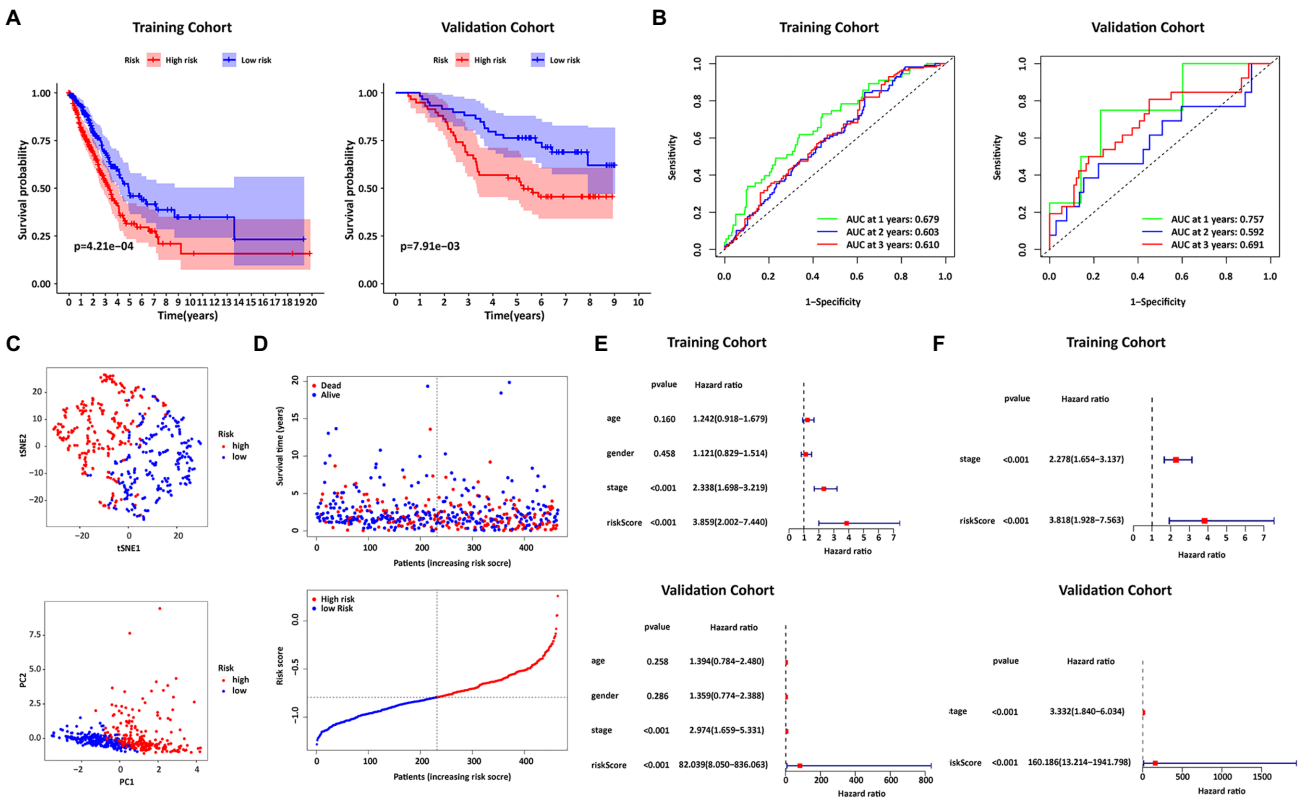


FIGURE 4 | Construction and Validation of the IRGs risk score formula. **(A)** Kaplan Meier survival curves for the TCGA training cohort and the GEO validation cohort. **(B)** ROC curves of the IRGs risk score formula in the training and validation cohorts. **(C)** T-SNE analysis and PCA plot of the training cohort. **(D)** The risk and survival status plots in the training cohort. **(E)** Forest plot from the univariate Cox regression analysis in the training and validation cohort. **(F)** The results of multivariate analysis regarding OS in the training and validation cohorts.

Ansari et al. (2020) found that there is a significant correlation between overexpression of SLC7A5 and specific immune cell subtypes. Their study not only provided clinical evidence that SLC7A5 in breast cancer can aid the personalization of anti-PD1/PDL1 inhibition therapies but also suggested that targeting SLC7A5 may enhance the efficacy of anti-PDL1 immunotherapy. GDF15 may promote ferroptosis (Dixon et al., 2014), and some studies indicate that GDF15 plays an anti-cancer role, but other data suggest that it may promote tumor progression and metastasis (Tsui et al., 2015; Wang et al., 2019b). In a study of acute kidney injury, Martin-Sanchez et al. (2017) found that IL-33 release was associated with ferroptosis, leading them to hypothesize that ferroptosis may regulate inflammation by activating IL-33 *in vivo*. Moreover, tumor-derived IL33 enhanced the antitumor effects of checkpoint inhibitors (Chen et al., 2020). SLC2A1 can be activated by lymphoid-specific helicase (LSH) to inhibit ferroptosis (Jiang et al., 2017), and high SLC2A1 expression usually correlates with poorer patient outcomes (Kawamura et al., 2001; Tohma et al., 2005). RRM2 was reported to be highly expressed in liver cancer tissues and to prevent ferroptosis (Zhang et al., 2019), and upregulation of RRM2 in NSCLC cells promoted proliferation and chemotherapeutic resistance (Huang et al., 2019).

The immune-related risk score calculation formulas used here contained four IRGs. The HLA-DRB5 alleles were associated with cervical neoplasia through a linkage disequilibrium with amino acid variations and HLA-DRB1 alleles (Bao et al., 2018). HLA-DRB5 has also been considered a possible prognostic factor for gastric cancer (Hang et al., 2018). High expression of SFTPD might function to prevent progression of lung cancer (Yamaguchi et al., 2011). Increased SFTPD levels have been shown to be associated with fewer distant metastases and progression-free survival in LUAD that harbors EGFR mutations (Umeda et al., 2017). The low expression of SFTPD in H1299 cells and TCGA samples indicates that SFTPD may be a protective factor in lung adenocarcinoma patients but is inhibited in tumor tissue. Prostaglandin H2 is converted to prostaglandin D2 (PGD2) under the catalytic action of PTGDS (Fukuhara et al., 2012). PGD2 has previously been shown to inhibit migration of cancer cells (Shyu et al., 2013). This effect can be achieved by influencing immune responses (Fagerberg et al., 2014). S100B is significantly downregulated in esophageal squamous cell carcinoma, and may cause cell growth stagnation and apoptosis through synergistic action with p53 (Ji et al., 2004). CacyBP/SIP is a target protein of S100B and an inhibitor of gastric cancer (Ning et al., 2007).

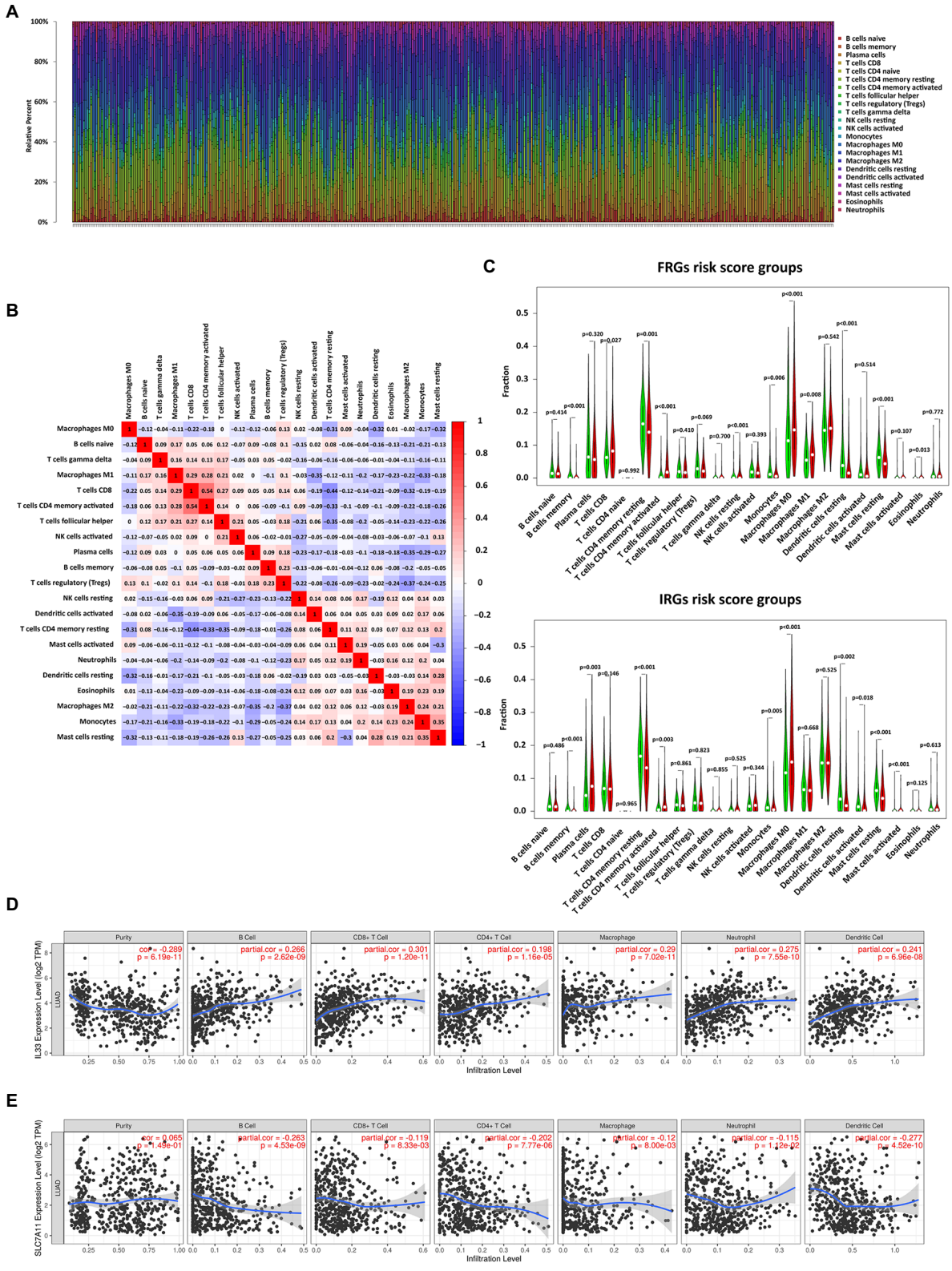


FIGURE 5 | Correlation analysis of immune cell infiltration with risk scores and expression of FRGs. **(A)** The bar chart shows the distribution of 22 types of immune cells in each sample. **(B)** Correlation matrix of immune cell proportions. **(C)** Violin plots of the differentially infiltrated immune cells between the low- and high-risk groups. **(D)** The correlation between IL33 expression and infiltration of immune cell subtypes. **(E)** The correlation between SLC7A11 expression and infiltration of immune cell subtypes.

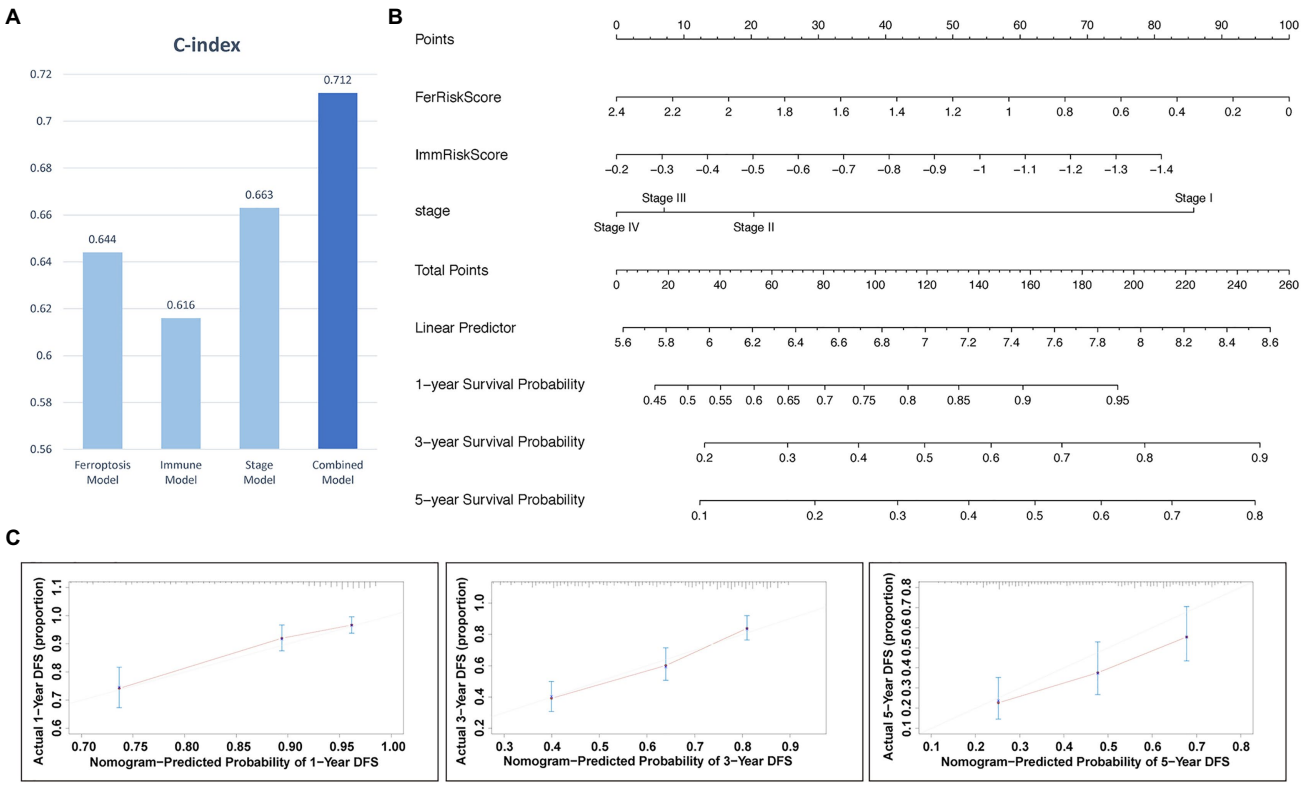


FIGURE 6 | The nomogram for predicting the 1-, 3-, and 5-year survival probability of patients with LUAD. **(A)** Comparison of C index among models. **(B)** Prognostic nomogram for predicting 1-, 3-, and 5-years OS probability of LUAD. **(C)** Calibration curves of the nomogram for assessing predicted OS at 1-, 3-, and 5- years in the TCGA cohort.

Ferroptosis seems to be involved in the immune response to tumors. We observed that expression of the FRGs that inhibit ferroptosis negatively correlated with expression of IRGs. IL33 expression was consistent with the presence of ferroptosis and positively correlated with expression of IRGs. These results might suggest that ferroptosis had a positive relationship with anti-tumor immunity in LUAD. It has been shown that reduced uptake of cystine results in increased ferroptosis of tumor cells after tumor immunotherapy, which helps improve anti-tumor efficacy (Wang et al., 2019a). This means that the immune system can drive ferroptosis to mediate inhibition of tumor growth. By calculating the abundance of immune cells, it can be shown that the two systems overlap in differences in the levels of multiple types of immune cells. Xu et al. (2020) found that IFN- γ secreted in tumor tissues by infiltrating lymphocytes can help downregulate expression of SLC7A11, thereby promoting tumor cell ferroptosis and decreasing tumor volume. Efimova et al. (2020) observed that dendritic cells can be induced to mature and active by phagocytosis of early ferroptotic cancer cells. Further exploration of the clinical relevance of ferroptosis- and immune-related marker genes, and potential connection between them, would help to identify more efficient diagnostic and therapeutic approaches to LUAD.

Although many studies have attempted to elucidate an association between ferroptosis and immunity, there is no report of combining two variables for predicting OS of LUAD patients. At present, prognosis in LUAD is based on tumor stage. In the present study, the nomogram composed of tumor staging combined with FRGs and IRGs-related risk scores provided more precise prediction of patient outcomes. However, there are several limitations. Firstly, the data used in this work were all downloaded from public databases. The results should also be externally validated with other primary data. Secondly, although we have found some similarity in immune function, further understanding of functional connections between the seven FRGs and four IRGs identified here requires additional experimental study.

In conclusion, our study identified seven FRGs and four IRGs that are differentially expressed in LUAD and significantly associated with prognosis. A nomogram that combines these sets of genes is more beneficial to individualized prognosis in LUAD than the current standard that uses a single prognostic factor. Using Pearson correlation analysis, we inferred that immune response is positively correlated with ferroptosis in LUAD. These results provide valuable information for development of new therapies that combine immunotherapy with ferroptosis-related drugs.

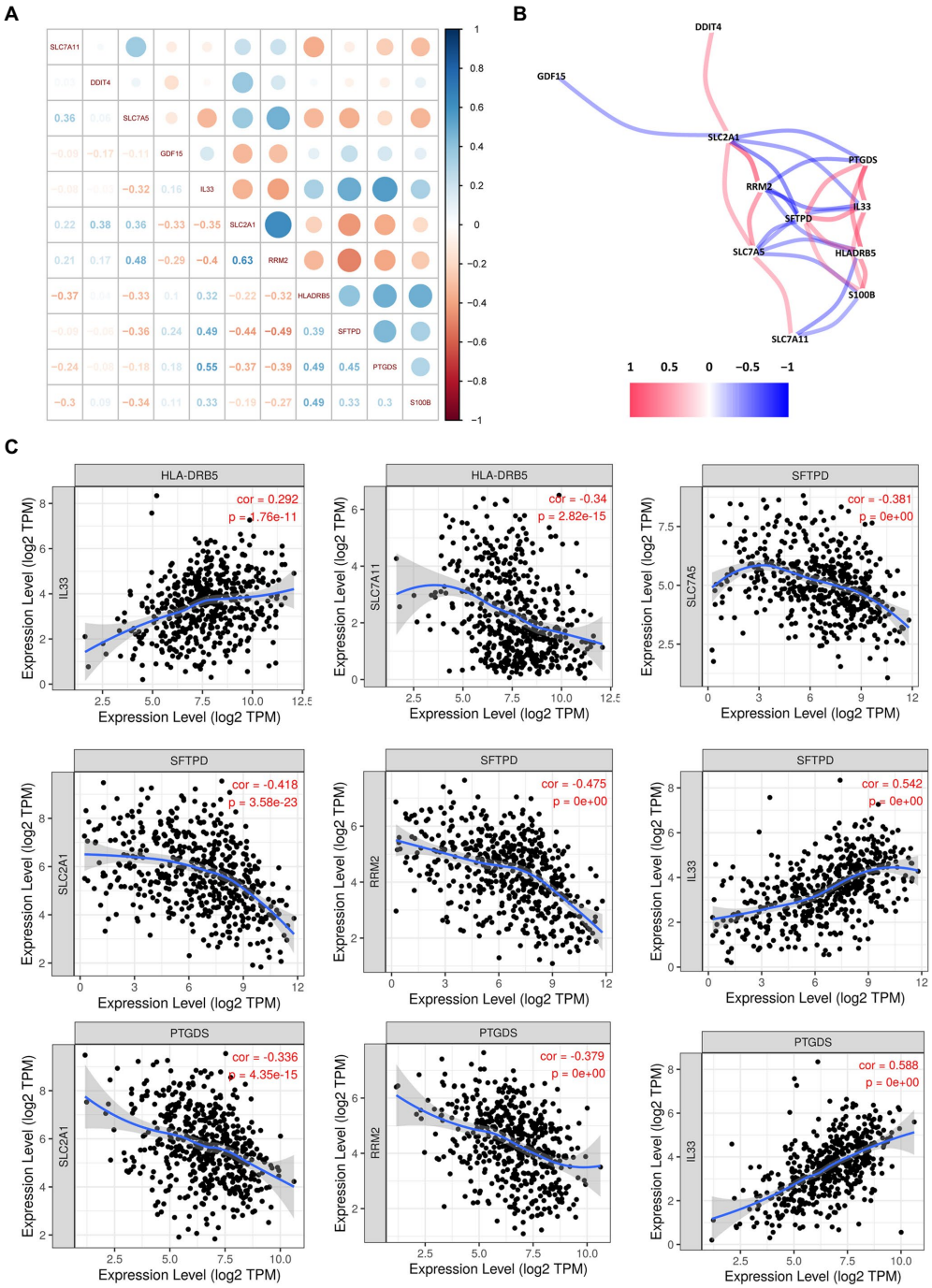


FIGURE 7 | Correlation analysis among the hub genes. **(A)** Correlation diagram of the likely relationships among the hub genes. **(B)** Correlation network of the hub genes. **(C)** Association between the expression levels of hub genes.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: TCGA database: https://portal.gdc.cancer.gov/repository. GSE13213 was from GEO database: https://www.ncbi.nlm.nih.gov/geo/.

AUTHOR CONTRIBUTIONS

MC and XL designed the study, executed the bioinformatics analysis, and prepared figures and tables. YZ and YT collected the data. MC, XL, YZ, PS, and JL wrote parts of the manuscript. XH, LW, and KS revised the manuscript.

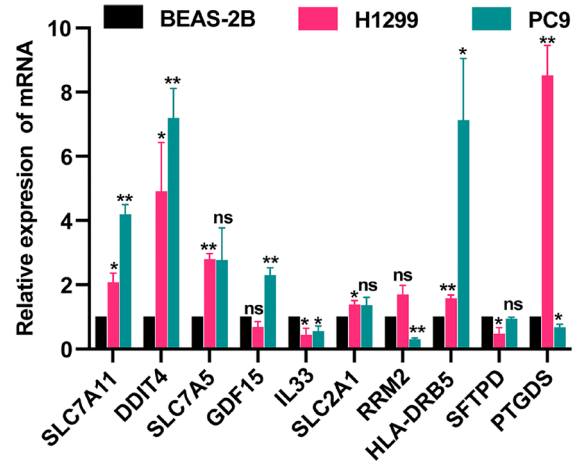


FIGURE 8 | Validation of expression of FRGs and IRGs by qRT-PCR. $^{*}p<0.05$ and $^{**}p<0.01$.

All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The generous support of the Division of Pulmonary Medicine, the First Affiliated Hospital of Wenzhou Medical University, Key Laboratory of Heart and Lung is gratefully acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.706814/full#supplementary-material

REFERENCES

Ansari, R. E., Craze, M. L., Althobiti, M., Alfarsi, L., Ellis, I. O., Rakha, E. A., et al. (2020). Enhanced glutamine uptake influences composition of immune cell infiltrates in breast cancer. *Br. J. Cancer* 122, 94–101. doi: 10.1038/s41416-019-0626-z

Badgley, M. A., Kremer, D. M., Maurer, H. C., DelGiorno, K. E., Lee, H. J., Purohit, V., et al. (2020). Cysteine depletion induces pancreatic tumor ferroptosis in mice. *Science* 368, 85–89. doi: 10.1126/science.aaw9872

Bao, X., Hanson, A. L., Madeleine, M. M., Wang, S. S., Schwartz, S. M., Newell, F., et al. (2018). HLA and KIR associations of cervical neoplasia. *J. Infect. Dis.* 218, 2006–2015. doi: 10.1093/infdis/jiy483

Cancer Genome Atlas Research Network (2014). Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511, 543–550. doi: 10.1038/nature13385

Chang, L. C., Chiang, S. K., Chen, S. E., Yu, Y. L., Chou, R. H., and Chang, W. C. (2018). Heme oxygenase-1 mediates BAY 11-7085 induced ferroptosis. *Cancer Lett.* 416, 124–137. doi: 10.1016/j.canlet.2017.12.025

Chen, L., Sun, R., Xu, J., Zhai, W., Zhang, D., Yang, M., et al. (2020). Tumor-derived IL33 promotes tissue-resident CD8(+) T cells and is required for checkpoint blockade tumor immunotherapy. *Cancer Immunol. Res.* 8, 1381–1392. doi: 10.1158/2326-6066.CIR-19-1024

Dai, E., Han, L., Liu, J., Xie, Y., Zeh, H. J., Kang, R., et al. (2020). Ferroptotic damage promotes pancreatic tumorigenesis through a TMEM173/STING-dependent DNA sensor pathway. *Nat. Commun.* 11:6339. doi: 10.1038/s41467-020-20154-8

Dixon, S. J., Patel, D. N., Welsch, M., Skouta, R., Lee, E. D., Hayano, M., et al. (2014). Pharmacological inhibition of cystine-glutamate exchange induces endoplasmic reticulum stress and ferroptosis. *elife* 3:e02523. doi: 10.7554/eLife.02523

Efferth, T. (2017). From ancient herb to modern drug: Artemisia annua and artemisinin for cancer therapy. *Semin. Cancer Biol.* 46, 65–83. doi: 10.1016/j.semcancer.2017.02.009

Efimova, I., Catanzaro, E., Van der Meeren, L., Turubanova, V. D., Hammad, H., Mishchenko, T. A., et al. (2020). Vaccination with early ferroptotic cancer cells induces efficient antitumor immunity. *J. Immunother. Cancer* 8:e001369. doi: 10.1136/jitc-2020-001369

Fagerberg, L., Hallström, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., et al. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics* 13, 397–406. doi: 10.1074/mcp.M113.035600

Fukuhara, A., Yamada, M., Fujimori, K., Miyamoto, Y., Kusumoto, T., Nakajima, H., et al. (2012). Lipocalin-type prostaglandin D synthase protects against oxidative stress-induced neuronal cell death. *Biochem. J.* 443, 75–84. doi: 10.1042/BJ20111889

Hang, X., Li, D., Wang, J., and Wang, G. (2018). Prognostic significance of microsatellite instability-associated pathways and genes in gastric cancer. *Int. J. Mol. Med.* 42, 149–160. doi: 10.3892/ijmm.2018.3643

Herbst, R. S., Morgensztern, D., and Boshoff, C. (2018). The biology and management of non-small cell lung cancer. *Nature* 553, 446–454. doi: 10.1038/nature25183

Ho, K. H., Chen, P. H., Chou, C. M., Shih, C. M., Lee, Y. T., Cheng, C. H., et al. (2020). A key role of DNA damage-inducible transcript 4 (DDIT4) connects autophagy and GLUT3-mediated stemness to desensitize temozolomide efficacy in glioblastomas. *Neurotherapeutics* 17, 1212–1227. doi: 10.1007/s13311-019-00826-0

Huang, N., Guo, W., Ren, K., Li, W., Jiang, Y., Sun, J., et al. (2019). LncRNA AFAP1-AS1 supresses miR-139-5p and promotes cell proliferation and chemotherapy resistance of non-small cell lung cancer by competitively upregulating RRM2. *Front. Oncol.* 9:1103. doi: 10.3389/fonc.2019.01103

Ji, J., Zhao, L., Wang, X., Zhou, C., Ding, F., Su, L., et al. (2004). Differential expression of S100 gene family in human esophageal squamous cell carcinoma. *J. Cancer Res. Clin. Oncol.* 130, 480–486. doi: 10.1007/s00432-004-0555-x

Jiang, Y., Mao, C., Yang, R., Yan, B., Shi, Y., Liu, X., et al. (2017). EGLN1/c-Myc induced lymphoid-specific helicase inhibits ferroptosis through lipid metabolic gene expression changes. *Theranostics* 7, 3293–3305. doi: 10.7150/thno.19988

Kawamura, T., Kusakabe, T., Sugino, T., Watanabe, K., Fukuda, T., Nashimoto, A., et al. (2001). Expression of glucose transporter-1 in human gastric carcinoma: association with tumor aggressiveness, metastasis, and patient survival. *Cancer* 92, 634–641. doi: 10.1002/1097-0142(20010801)92:3<634::AID-CNCR1364>3.0.CO;2-X

Kinoshita, F. L., Ito, Y., and Nakayama, T. (2016). Trends in lung cancer incidence rates by histological type in 1975-2008: a population-based study in Osaka, Japan. *J. Epidemiol.* 26, 579–586. doi: 10.2188/jea.JE20150257

Li, Z., and Rong, L. (2020). Cascade reaction-mediated efficient ferroptosis synergizes with immunomodulation for high-performance cancer therapy. *Biomater. Sci.* 8, 6272–6285. doi: 10.1039/D0BM01168A

Ma, J., Ward, E. M., Smith, R., and Jemal, A. (2013). Annual number of lung cancer deaths potentially avertable by screening in the United States. *Cancer* 119, 1381–1385. doi: 10.1002/cncr.27813

Martin-Sanchez, D., Ruiz-Andres, O., Poveda, J., Carrasco, S., Cannata-Ortiz, P., Sanchez-Niño, M. D., et al. (2017). Ferroptosis, but not necroptosis, is important in nephrotoxic folic acid-induced AKI. *J. Am. Soc. Nephrol.* 28, 218–229. doi: 10.1681/asn.2015121376

Matsushita, M., Freigang, S., Schneider, C., Conrad, M., Bornkamm, G. W., and Kopf, M. (2015). T cell lipid peroxidation induces ferroptosis and prevents immunity to infection. *J. Exp. Med.* 212, 555–568. doi: 10.1084/jem.20140857

Ning, X., Sun, S., Hong, L., Liang, J., Liu, L., Han, S., et al. (2007). Calcyclin-binding protein inhibits proliferation, tumorigenicity, and invasion of gastric cancer. *Mol. Cancer Res.* 5, 1254–1262. doi: 10.1158/1541-7786.MCR-06-0426

Shi, Z. Z., Fan, Z. W., Chen, Y. X., Xie, X. F., Jiang, W., Wang, W. J., et al. (2019). Ferroptosis in carcinoma: regulatory mechanisms and new method for cancer therapy. *Onco Targets Ther.* 12, 11291–11304. doi: 10.2147/ott.S232852

Shyu, R. Y., Wu, C. C., Wang, C. H., Tsai, T. C., Wang, L. K., Chen, M. L., et al. (2013). H-rev107 regulates prostaglandin D2 synthase-mediated suppression of cellular invasion in testicular cancer cells. *J. Biomed. Sci.* 20:30. doi: 10.1186/1423-0127-20-30

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

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

Tohma, T., Okazumi, S., Makino, H., Cho, A., Mochizuki, R., Shuto, K., et al. (2005). Overexpression of glucose transporter 1 in esophageal squamous cell carcinomas: a marker for poor prognosis. *Dis. Esophagus* 18, 185–189. doi: 10.1111/j.1442-2050.2005.00489.x

Tsui, K. H., Hsu, S. Y., Chung, L. C., Lin, Y. H., Feng, T. H., Lee, T. Y., et al. (2015). Growth differentiation factor-15: a p53- and demethylation-upregulating gene represses cell proliferation, invasion, and tumorigenesis in bladder carcinoma cells. *Sci. Rep.* 5:12870. doi: 10.1038/srep12870

Umeda, Y., Hasegawa, Y., Otsuka, M., Ariki, S., Takamiya, R., Saito, A., et al. (2017). Surfactant protein D inhibits activation of non-small cell lung cancer-associated mutant EGFR and affects clinical outcomes of patients. *Oncogene* 36, 6432–6445. doi: 10.1038/onc.2017.253

Villalpando-Rodriguez, G. E., Blankstein, A. R., Konzelman, C., and Gibson, S. B. (2019). Lysosomal destabilizing drug siramesine and the dual tyrosine kinase inhibitor lapatinib induce a synergistic ferroptosis through reduced heme oxygenase-1 (HO-1) levels. *Oxidative Med. Cell. Longev.* 2019:9561281. doi: 10.1155/2019/9561281

Wang, D., and DuBois, R. N. (2015). Immunosuppression associated with chronic inflammation in the tumor microenvironment. *Carcinogenesis* 36, 1085–1093. doi: 10.1093/carcin/bgv123

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

Wang, Y., Han, E., Xing, Q., Yan, J., Arrington, A., Wang, C., et al. (2015). Baicalein upregulates DDIT4 expression which mediates mTOR inhibition and growth inhibition in cancer cells. *Cancer Lett.* 358, 170–179. doi: 10.1016/j.canlet.2014.12.033

Wang, W., Yang, X., Dai, J., Lu, Y., Zhang, J., and Keller, E. T. (2019b). Prostate cancer promotes a vicious cycle of bone metastasis progression through inducing osteocytes to secrete GDF15 that stimulates prostate cancer growth and invasion. *Oncogene* 38, 4540–4559. doi: 10.1038/s41388-019-0736-3

Wilkerson, M. D., Yin, X., Hoadley, K. A., Liu, Y., Hayward, M. C., Cabanski, C. R., et al. (2010). Lung squamous cell carcinoma mRNA expression subtypes are reproducible, clinically important, and correspond to normal cell types. *Clin. Cancer Res.* 16, 4864–4875. doi: 10.1158/1078-0432.CCR-10-0199

Xu, T., Ma, Y., Yuan, Q., Hu, H., Hu, X., Qian, Z., et al. (2020). Enhanced ferroptosis by oxygen-boosted phototherapy based on a 2-in-1 nanoplatform of ferrous hemoglobin for tumor synergistic therapy. *ACS Nano* 14, 3414–3425. doi: 10.1021/acsnano.9b09426

Yamaguchi, H., Soda, H., Nakamura, Y., Takasu, M., Tomonaga, N., Nakano, H., et al. (2011). Serum levels of surfactant protein D predict the anti-tumor activity of gefitinib in patients with advanced non-small cell lung cancer. *Cancer Chemother. Pharmacol.* 67, 331–338. doi: 10.1007/s00280-010-1325-x

Yang, W. S., SriRamaratnam, R., Welsch, M. E., Shimada, K., Skouta, R., Viswanathan, V. S., et al. (2014). Regulation of ferroptotic cancer cell death by GPX4. *Cell* 156, 317–331. doi: 10.1016/j.cell.2013.12.010

Ye, L. F., Chaudhary, K. R., Zandkarimi, F., Harken, A. D., Kinslow, C. J., Upadhyayula, P. S., et al. (2020). Radiation-induced lipid peroxidation triggers ferroptosis and synergizes with ferroptosis inducers. *ACS Chem. Biol.* 15, 469–484. doi: 10.1021/acschembio.9b00939

Zhang, X., Du, L., Qiao, Y., Zhang, X., Zheng, W., Wu, Q., et al. (2019). Ferroptosis is governed by differential regulation of transcription in liver cancer. *Redox Biol.* 24:101211. doi: 10.1016/j.redox.2019.101211

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

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 15 November 2021
doi: 10.3389/fgene.2021.699242

DNASE1L3 as a Novel Diagnostic and Prognostic Biomarker for Lung Adenocarcinoma Based on Data Mining

Jianlin Chen[1], Junping Ding[2], Wenjie Huang[1], Lin Sun[1], Jinping Chen[3], Yangyang Liu[1], Qianmei Zhan[1], Gan Gao[4], Xiaoling He[5], Guowen Qiu[6], Peiying Long[5], Lishu Wei[1], Zhenni Lu[1] and Yifan Sun[1]**

[1]Departments of Clinical Laboratory of Affiliated Liutie Central Hospital of Guangxi Medical University, Liuzhou, China, [2]Departments of General surgery of Affiliated Liutie Central Hospital of Guangxi Medical University, Liuzhou, China, [3]Departments of Respiratory Medicine of Affiliated Liutie Central Hospital of Guangxi Medical University, Liuzhou, China, [4]Departments of Clinical Laboratory of Liuzhou Maternity and Child Healthcare Hospital, Liuzhou, China, [5]Department of Clinical Laboratory of People's Hospital Rong an County, Liuzhou, China, [6]Departments of Orthopedics of Affiliated Liutie Central Hospital of Guangxi Medical University, Liuzhou, China

OPEN ACCESS

Edited by:
Tania Lee Slatter,
University of Otago, New Zealand

Reviewed by:
Emil Bulatov,
Kazan Federal University, Russia
Sohini Chakraborty,
NYU Grossman School of Medicine, United States

***Correspondence:**
Jianlin Chen
cjl801900@163.com
Yifan Sun
sunyifan13@126.com

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: 23 April 2021
Accepted: 18 October 2021
Published: 15 November 2021

Citation:
Chen J, Ding J, Huang W, Sun L, Chen J, Liu Y, Zhan Q, Gao G, He X, Qiu G, Long P, Wei L, Lu Z and Sun Y (2021) DNASE1L3 as a Novel Diagnostic and Prognostic Biomarker for Lung Adenocarcinoma Based on Data Mining. Front. Genet. 12:699242. doi: 10.3389/fgene.2021.699242

Previous researches have highlighted that low-expressing deoxyribonuclease1-like 3 (DNASE1L3) may play a role as a potential prognostic biomarker in several cancers. However, the diagnosis and prognosis roles of DNASE1L3 gene in lung adenocarcinoma (LUAD) remain largely unknown. This research aimed to explore the diagnosis value, prognostic value, and potential oncogenic roles of DNASE1L3 in LUAD. We performed bioinformatics analysis on LUAD datasets downloaded from TCGA (The Cancer Genome Atlas) and GEO (Gene Expression Omnibus), and jointly analyzed with various online databases. We found that both the mRNA and protein levels of DNASE1L3 in patients with LUAD were noticeably lower than that in normal tissues. Low DNASE1L3 expression was significantly associated with higher pathological stages, T stages, and poor prognosis in LUAD cohorts. Multivariate analysis revealed that DNASE1L3 was an independent factor affecting overall survival (HR = 0.680, p = 0.027). Moreover, decreased DNASE1L3 showed strong diagnostic efficiency for LUAD. Results indicated that the mRNA level of DNASE1L3 was positively correlated with the infiltration of various immune cells, immune checkpoints in LUAD, especially with some m6A methylation regulators. In addition, enrichment function analysis revealed that the co-expressed genes may participate in the process of intercellular signal transduction and transmission. GSEA indicated that DNASE1L3 was positively related to G protein-coupled receptor ligand biding (NES = 1.738; P adjust = 0.044; FDR = 0.033) and G alpha (i) signaling events (NES = 1.635; P adjust = 0.044; FDR = 0.033). Our results demonstrated that decreased DNASE1L3 may serve as a novel diagnostic and prognostic biomarker associating with immune infiltrates in lung adenocarcinoma.

Keywords: LUAD, DNASE1L3, diagnostic, prognostic, TCGA database

INTRODUCTION

Lung cancer is the leading cause of cancer-related mortality worldwide. Lung adenocarcinoma (LUAD), also known as pulmonary adenocarcinoma, accounts for 60% of all lung cancers (Denisenko et al., 2018). LUAD mainly occurs in the distal lung and alveoli and is not easily reached by bronchoscopy, which poses a huge challenge in diagnosis and treatment of LUAD (Liu X. et al., 2021). However, activation of proto-oncogenes is known to play vital role in the formation of the vast majority of cancers, and gaining insight into its expression levels and disease-related prognosis may contribute to the development of effective diagnosis and prevention of lung adenocarcinoma.

Deoxyribonuclease 1-like 3 (DNASE1L3), a member of the deoxyribonuclease 1 family, is a DNASE1-like nuclease expressed in the endoplasmic reticulum. As one of the major serum nucleases (Rodriguez et al., 1997; Shiokawa and Tanuma, 2001), DNASE1L3 is predominantly produced by dendritic cells (DCs), macrophages (Mph), and neutrophils (NEU). It has been proved that DNASE1L3 plays vital role in DNA catabolism and cell apoptosis (Errami et al., 2012; Han et al., 2020). Increasing evidence has demonstrated that dysfunction of DNASE1L3 may cause immune responses against DNA and autoimmune diseases in mice and humans (Wilber et al., 2003; Al-Mayouf et al., 2011; Zochling et al., 2014; Sisirak et al., 2016; Sun et al., 2020). Study has also reported that DNASE1L3 was involved in breast cancer signal transduction (Sjoblom et al., 2006). Malecki et al. found that hyper expressive of DNASE1L3 gene can degrade the genome of ovarian cancer cells and lead to cell death (Malecki et al., 2013). In addition, investigators announced that the expression level of DNASE1L3 is closely align with the stage of clear cell renal cell carcinoma (Bhalla et al., 2017a). Recently, it has been reported that patients with high DNASE1L3 expression achieved significantly longer overall survival (OS) of HCC (Xu et al., 2019; Wang et al., 2020a). By the above, DNASE1L3 is highly likely to be contribute to the cancer genesis and progression. However, the role of DNASE1L3 in LUAD has not been well elucidated, which is the aim of our study.

In current study, we comprehensively analyzed DNASE1L3 expression data of LUAD patients in TCGA (The Cancer Genome Atlas), GEO (Gene Expression Omnibus), CPTAC dataset, and Human Protein Atlas (HPA) database. Using multi-omics analysis, we assessed the potential diagnosis or clinical prognosis efficiency of DNASE1L3 in lung adenocarcinoma. Moreover, we investigated the correlation of DNASE1L3 with tumor-infiltrating immune cells *via* R (version 3.6.3) GSVA package and ssGSEA (GSVA pack-built algorithm) package. Various bioinformatics tools were used to explore the genetic alteration of DNASE1L3, potential biological functions of DNASE1L3, the correlation with immune checkpoints, and m6A RNA methylation regulators of DNASE1L3 in LUAD. Our results confirmed that low DNASE1L3 expression might be a useful diagnostic and poor prognostic biomarker for lung adenocarcinoma.

TABLE 1 | Clinical characteristics of the patients with lung adenocarcinoma.

Characteristic	Levels	Overall
Age, n (%)	≤65	255 (49.4%)
	>65	261 (50.6%)
Gender, n (%)	Female	286 (53.5%)
	Male	249 (46.5%)
T stage, n (%)	T1	175 (32.9%)
	T2	289 (54.3%)
	T3	49 (9.2%)
	T4	19 (3.6%)
N stage, n (%)	N0	348 (67.1%)
	N1	95 (18.3%)
	N2	74 (14.3%)
	N3	2 (0.4%)
M stage, n (%)	M0	361 (93.5%)
	M1	25 (6.5%)
Pathologic stage, n (%)	Stage I	294 (55.8%)
	Stage II	123 (23.3%)
	Stage III	84 (15.9%)
	Stage IV	26 (4.9%)
Residual tumor, n (%)	R0	355 (95.4%)
	R1	13 (3.5%)
	R2	4 (1.1%)
Smoking, n (%)	No	75 (14.4%)
	Yes	446 (85.6%)
DNASE1L3, n (%)	High	268 (50.1%)
	Low	267 (49.9%)
OS event, n (%)	Alive	343 (64.1%)
	Dead	192 (35.9%)
DSS event, n (%)	Alive	379 (76%)
	Dead	120 (24%)

Abbreviations: OS, overall survival; DSS, disease-free survival.

DATA AND METHODS

Expression and Transcription Analysis

The TIMER2.0 was employed to observe the expression difference of DNASE1L3 in 33 cancers of the TCGA project. A total of 526 lung adenocarcinoma patients and 57 normal including clinical data and RNA-seq of DNASE1L3 ($p \leq 0.05$, $|\log2FC| \geq 2$) were downloaded from TCGA database by UCSC Xena website (https://xena.ucsc.edu/). Then, the gene expression of DNASE1L3 was analyzed basing on different clinical groups. The log2 [TPM (Transcripts per million) +1] or log2 [FPKM (Fragments Per Kilo base per Million) +1] transformed expression data was applied for data analysis. The CPTAC dataset (http://ualcan.path.uab.edu/analysis-prot.html) and Human Protein Atlas (HPA) database (http://proteinatlas.org) were utilized to compare protein level of DNASE1L3 in normal and LUAD tissues (Chen et al.). Two sets of LUAD chip datasets GSE40791 and GSE10072 were downloaded from the GEO (https://www.ncbi.nlm.nih.gov/geo/) database (Li et al., 2017) and utilized as validation sets to study the differential expression and diagnostic efficiency of DNASE1L3 in the study.

Diagnostic and Survival Analysis

The diagnostic values of DNASE1L3 was calculated by pROC package (Robin et al., 2011) of R version 3.6.3, and the ROC

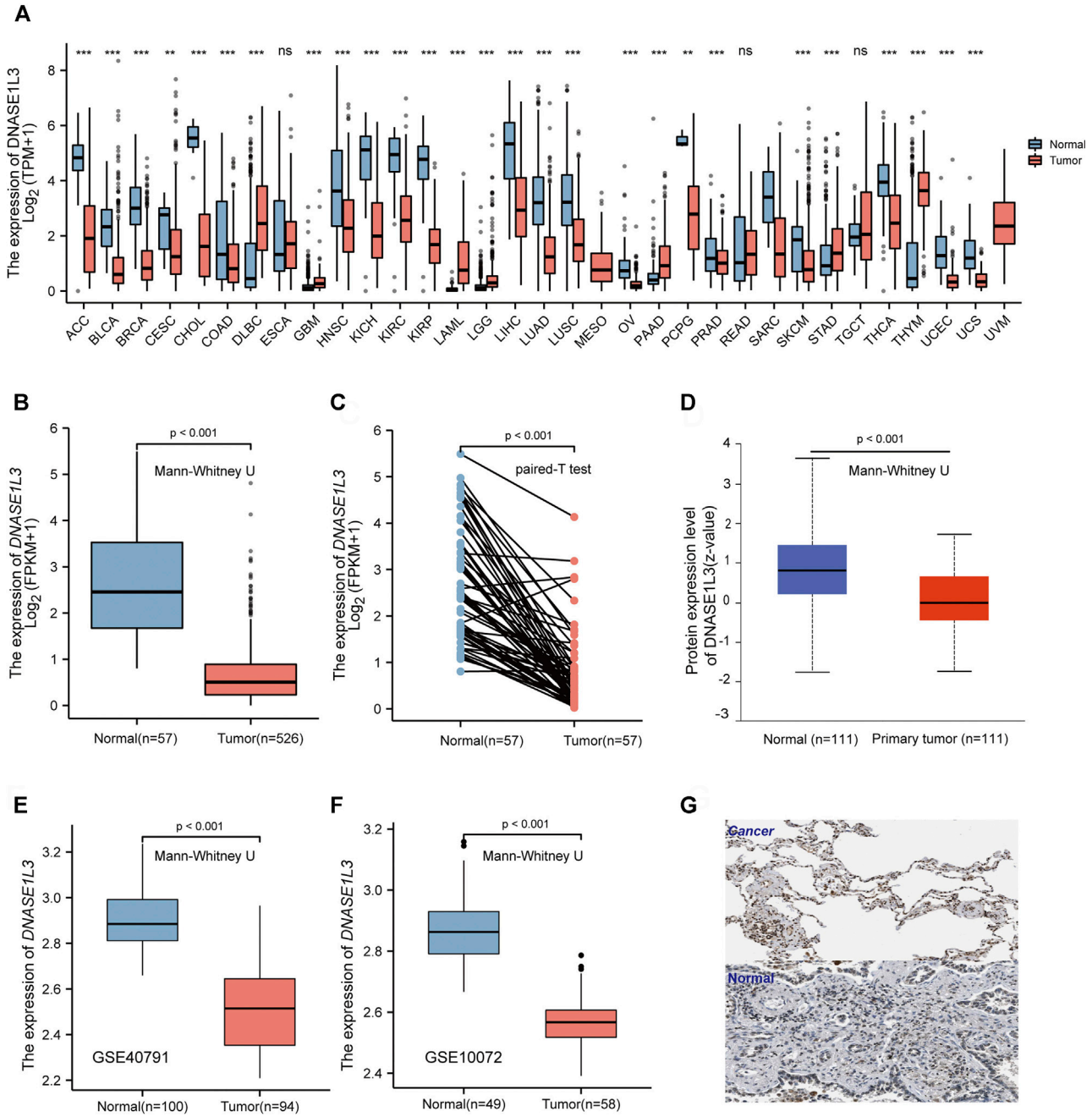


FIGURE 1 | mRNA and protein expression level of DNASE1L3 in lung adenocarcinoma. **(A)** Human DNASE1L3 expression levels in different tumor types from TCGA database were determined by TIMER ($^{ns}p > 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$). Based on the TCGA data, DNASE1L3 expression was significantly down-regulated in LUAD tissues compared to normal tissues **(B)** and noncancerous adjacent tissues (all $p < 0.001$) **(C)**. **(D)** Based on the CPTAC dataset, protein expression level of DNASE1L3 between normal tissue and primary tissue of LUAD ($p < 0.001$). **(E–F)** DNASE1L3 was expressed at a lower level in LUAD tissues than normal tissues ($p < 0.05$) from two different LUAD datasets. **(G)** Representative immunohistochemistry images of DNASE1L3 in LUAD and normal tissues derived from the HPA database.

curves were visualized by ggplot2 package (Wickham, 2016). The RNA-seq data in FPKM format were analyzed after log2 conversion. After that, LUAD cohort was clustered into low and high expression groups by median value of mRNA expression of DNASE1L3. The correlations between gene expression and OS, disease specific survival (DSS), and progress free interval (PFI) were analyzed *via* the Kaplan–Meier Plotter by R package (Liu et al., 2018). Then, the survival curves between subgroups (grouped

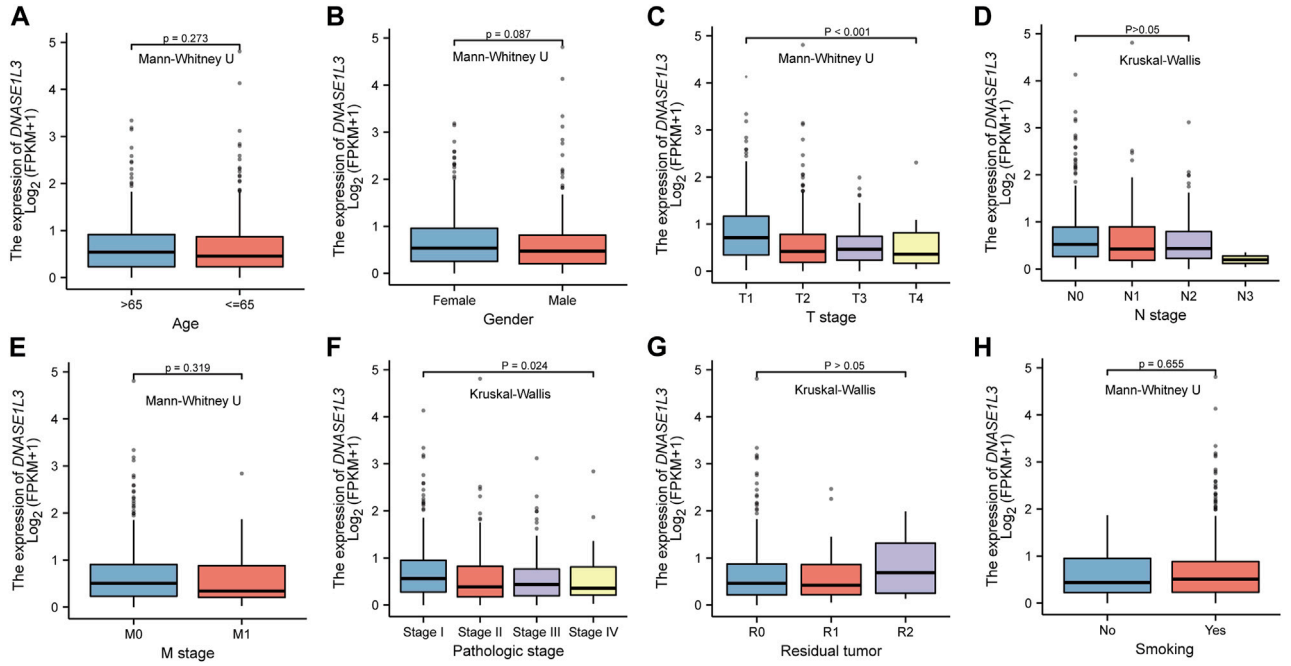


FIGURE 2 | Box plot evaluating DNASE1L3 expression of patients with LUAD according to different clinical characteristics. **(A)** Age; **(B)** Gender; **(C)** T stage; **(D)** N stage; **(E)** M stage; **(F)** pathologic stage; **(G)** residual tumor; **(H)** smoking status.

TABLE 2 | Relationship between the clinical features and DNASE1L3 expression in patients with lung adenocarcinoma.

Characteristic	Variable	Expression of DNASE1L3		p	χ^2
		Low, n (%)	High, n (%)		
T stage	T1	60 (11.3%)	115 (21.6%)	< 0.001	25.15
	T2	166 (31.2%)	123 (23.1%)		
	T3	28 (5.3%)	21 (3.9%)		
	T4	11 (2.1%)	8 (1.5%)		
N stage	N0	166 (32%)	182 (35.1%)	0.214	4.48
	N1	53 (10.2%)	42 (8.1%)		
	N2	40 (7.7%)	34 (6.6%)		
	N3	2 (0.4%)	0 (0%)		
M stage	M0	180 (46.6%)	181 (46.9%)	0.122	2.4
	M1	17 (4.4%)	8 (2.1%)		
Pathologic stage	I	131 (24.9%)	163 (30.9%)	0.013	10.82
	II	70 (13.3%)	53 (10.1%)		
	III	47 (8.9%)	37 (7%)		
	IV	18 (3.4%)	8 (1.5%)		
Gender	Female	134 (25%)	152 (28.4%)	0.154	2.04
	Male	133 (24.9%)	116 (21.7%)		
Age	≤65	134 (26%)	121 (23.4%)	0.253	1.31
	>65	123 (23.8%)	138 (26.7%)		
Residual tumor	R0	186 (50%)	169 (45.4%)	0.806	0.43
	R1	8 (2.2%)	5 (1.3%)		
	R2	2 (0.5%)	2 (0.5%)		
Smoker	No	39 (7.5%)	36 (6.9%)	0.762	0.09
	Yes	220 (42.2%)	226 (43.4%)		
OS event	Alive	148 (27.7%)	195 (36.4%)	< 0.001	16.71
	Dead	119 (22.2%)	73 (13.6%)		

by age, gender, pathological type, TNM stage, and smoking history) were plotted by the Kaplan–Meier analysis log-rank test. The HR with 95% CI was marked.

Correlations Between DNASE1L3 Expression and Immune Characteristics

Data on immune cell infiltration were obtained from TCGA. R (version 3.6.3), GSVA package, and ssGSEA package were used for statistical analysis and visualization of the data (Hänzelmann et al., 2013). The relationship between the expression of DNASE1L3 and immune infiltration was analyzed (Liu et al., 2018). Spearman correlation of DNASE1L3 gene in lung adenocarcinoma was generated by correlation module, and its statistical significance was estimated. In addition, the relationship between DNASE1L3 expression and immune checkpoint marker expression levels by Spearman correlation analyses is shown.

Tumor Mutation Profiles Analysis

Somatic mutation data were downloaded from the TCGA database and processed to identify the somatic variants and display somatic landscape by R package Maftools (Anand et al., 2018). The R package Somatic Signatures was utilize to describe the mutant signatures of lung adenocarcinoma samples (Gehring et al., 2015).

Correlations Between DNASE1L3 Expression and m6A RNA Methylation Regulators

Differentially expressed m6A RNA methylation regulators (LUADs vs. normal tissues, high- and low-DNASE1L3 LUADs) were analyzed by the Mann–Whitney U test method in R (version R 3.6.3). $p < 0.05$ and Log2|FC| > 1 were used as the significance criteria. Subsequently, expression of m6A-related

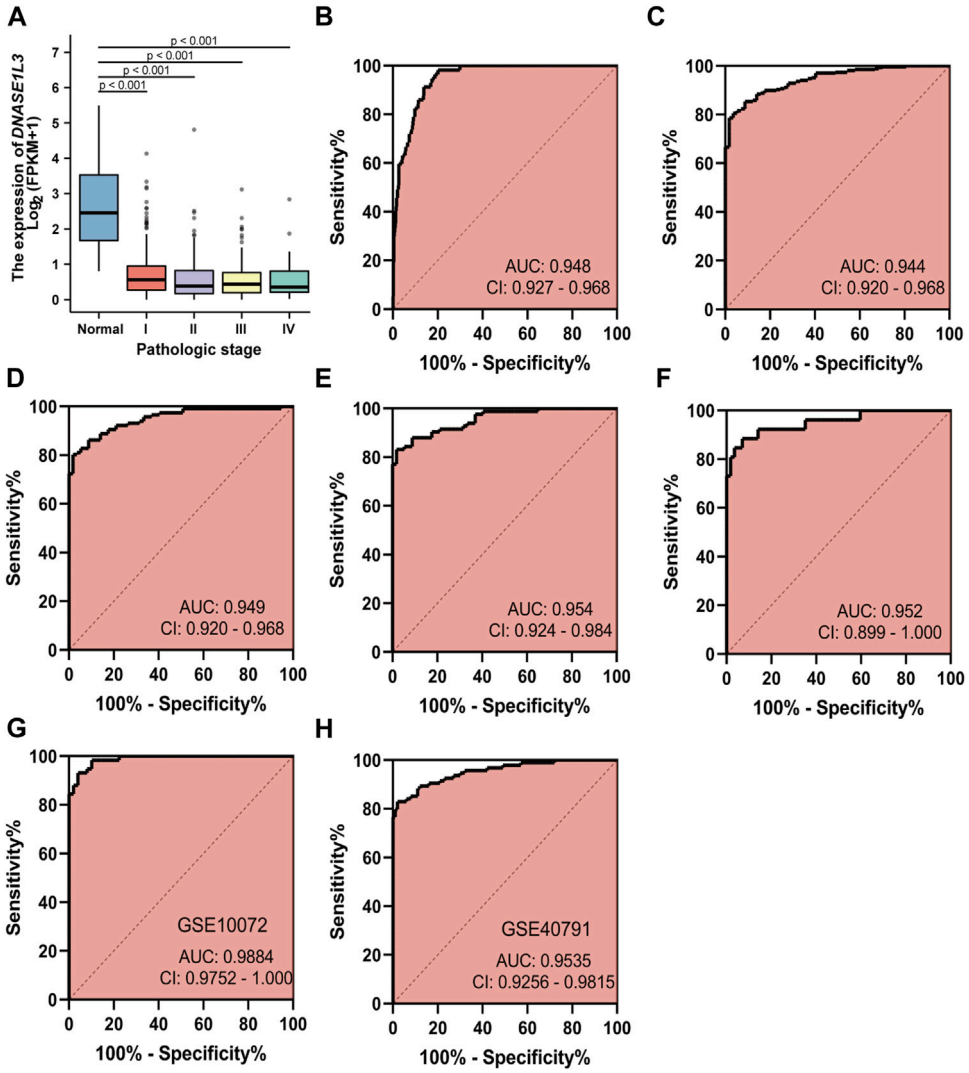


FIGURE 3 | Diagnostic value of DNASE1L3 expression in patients with lung adenocarcinoma. The results of the TGCA dataset showed higher expression of DNASE1L3 in normal tissues than in pathologic stages of LUAD (all $p < 0.001$) **(A)**; ROC curves of DNASE1L3 expression in patients with LUAD, including normal vs. overall tumor **(B)**; normal vs. stage I tumor **(C)**; normal vs. stage II tumor **(D)**; normal vs. stage III tumor **(E)**; normal vs. stage IV tumor **(F)** based on the TGCA dataset; ROC curve of DNASE1L3 for LUAD through the GEO datasets **(G–H)**. AUC, area under the curve.

genes in 526 LUAD patients and 57 normal samples was visualized.

Pathway and Enrichment Analysis

A total of 526 LUAD patients were separated into high- and low-DNASE1L3 expression groups according to DNASE1L3 median value. Limma package was used to identify differentially expressed genes (DEGs). The |log 2 (Fold Change) | larger than 1.5 and an adjusted *p*-value less than 0.05 were set as thresholds. The "pheatmap" and "EnhancedVolcano" R packages were employed to supply the heatmaps and volcano plots of the selected genes. In addition, we applied "ClusterProfiler" R package to conduct and visualize the Kyoto encyclopedia of genes and genomes (KEGG) and Gene Ontology (GO) analysis, including molecular function (MF), biological process (BP), and cell composition (CC), and adjusted $p < 0.05$ is considered to be enriched to a meaningful pathway (Yu et al., 2012). Further, we also used the R package "ClusterProfiler" to explore the Gene Set Enrichment Analysis (GSEA) of the DEGs in the two groups (Subramanian et al., 2005).

Statistical Analysis

R software (version 3.6.3) was utilized for all statistical analyses. Mann–Whitney U test was used to compare the two groups. The Kruskal–Wallis test compares three or more differences group. Chi-square test was performed and corrected to pass Fisher's exact test. $p < 0.05$ was considered statistically significant.

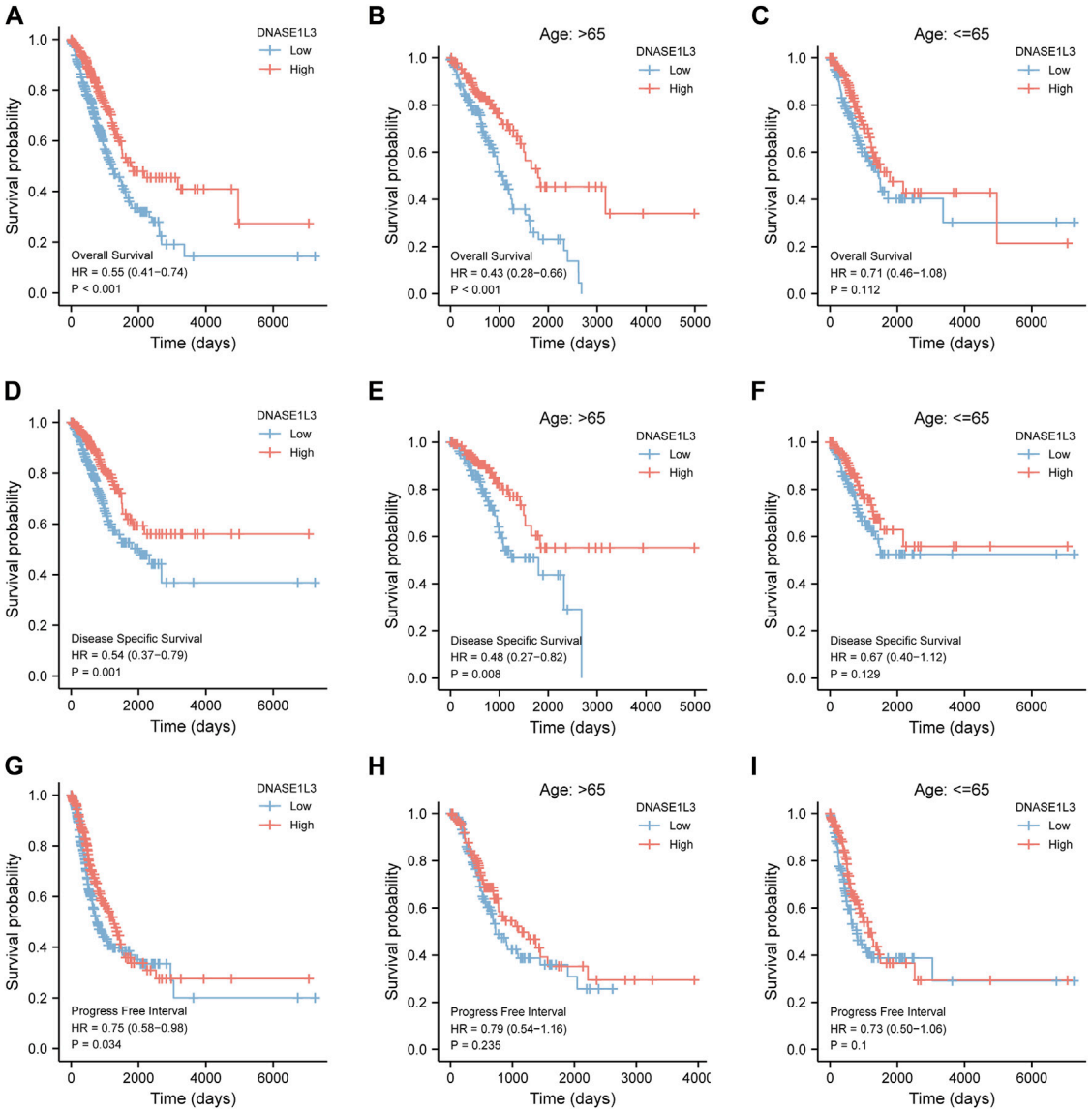


FIGURE 4 | Survival analysis of DNASE1L3 in lung adenocarcinoma by OS, DSS, and PFI as well as stratified analysis by age. **(A–C)** Survival analysis of DNASE1L3 in lung adenocarcinoma by OS as well as stratified analysis by elderly patients and young patients, respectively. **(D–F)** Survival analysis of DNASE1L3 in lung adenocarcinoma by DSS as well as stratified analysis by elderly patients and young patients, respectively. **(G,H)** Survival analysis of DNASE1L3 in lung adenocarcinoma by PFI as well as stratified analysis by elderly patients and young patients, respectively. Notes: elderly patients: age > 65; young patients: age ≤ 65.

RESULTS

Baseline Characteristics of Patients

A total of 526 LUAD patients containing the required clinical features were downloaded from the TCGA dataset in April 2021. Clinical information including age, gender, TNM stage, pathologic stage, residual tumor, DNASE1L3 expression, smoking history, overall survival, and disease-free survival are listed in **Table 1**. Among the 526 participants, 249 were male (46.5%) and 286 were female (53.5%). The residual tumor type of most patients was R0 (95.4%), and 93.5% of the patients at M0 stage. In terms of pathologic stage, 294 patients were stage I (55.8%), 123 patients were stage II (23.3%), 84 patients were stage III (15.9%), and 26 patients were stage IV (4.9%). Most patients were smokers (85.6%). The median age of all LUAD patients was 65 years.

Decreased Expression of DNASE1L3 in Lung Adenocarcinoma

To evaluate the DNASE1L3 mRNA levels in different tumors and normal tissues of multiple cancer types, we firstly examined DNASE1L3 expression using the RNA-seq data of multiple malignancies in TCGA. As shown in **Figure 1A**, DNASE1L3 expression shows significantly lower in the pancancers. Especially, the DNASE1L3 mRNA level was found to lowly

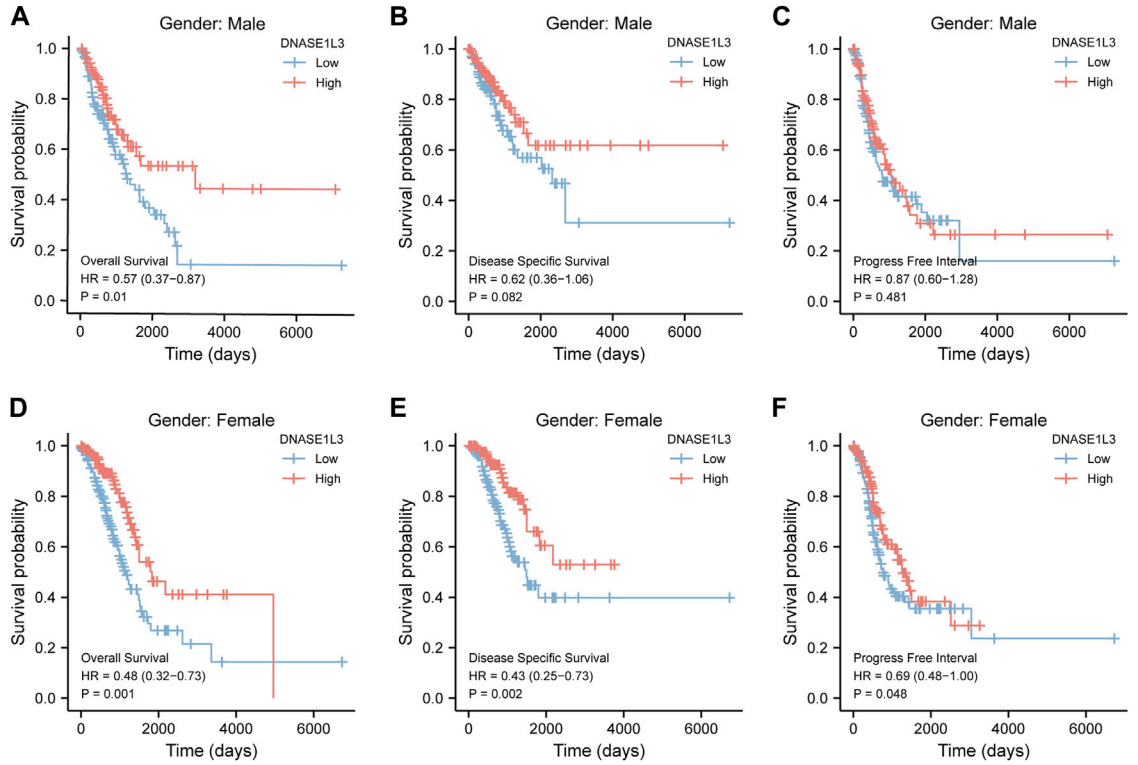


FIGURE 5 | Stratified survival analysis of DNASE1L3 in lung adenocarcinoma by gender of OS, DSS, and PFI. **(A,D)** Stratified survival analysis of DNASE1L3 in lung adenocarcinoma of male and female of OS, respectively. **(B,E)** Stratified survival analysis of DNASE1L3 in lung adenocarcinoma of male and female of DSS, respectively. **(C,F)** Stratified survival analysis of DNASE1L3 in lung adenocarcinoma of male and female of PFI, respectively.

expressed in LUAD comparing with normal tissue (**Figure 1B**, $p < 0.001$) and paired adjacent normal tissues (**Figure 1C**, $p < 0.001$). The differences in protein levels of DNASE1L3 were then analyzed by the CPTAC dataset and HPA database. As shown in **Figure 1D**, the protein expression level of DNASE1L3 was significantly down-regulated in LUAD tissues than in normal tissues (**Figure 1D**, $p < 0.001$). When showing the protein expression by immunohistochemistry images, the DNASE1L3 was still significantly down-regulated in LUAD tissues compared with normal samples (**Figure 1G**). To further verify the above conclusions, we further investigated DNASE1L3 expression in GSE40791 and GSE10072 datasets. We found that DNASE1L3 was indeed obviously expressed at a lower level in the LUAD tissues (**Figures 1E,F**, all $p < 0.001$). Taken together, all these results indicated that DNASE1L3 was remarkably down-regulated in LUAD patients at both mRNA and protein levels, implying a potential role in LUAD development and progression.

Correlation of DNASE1L3 Expression With Clinical Features

Secondly, we explored the roles of DNASE1L3 in LUAD progression. The Mann–Whitney U test or Kruskal–Wallis rank sum test method was applied to analyze the correlation between the expression level of DNASE1L3 and age, gender, TNM stage, pathologic stage, residual tumor, and smoking status (**Figures 2A–H**). The results indicated that low expression of DNASE1L3 was significantly correlated with higher T stages (**Figure 2C**, $p < 0.001$) and pathologic stages (**Figure 2F**, $p = 0.024$). Moreover, as shown in **Table 2**. The expression of DNASE1L3 was associated with OS event ($p < 0.001$), pathologic stage ($p = 0.013$), and T stage ($p < 0.001$) but was not related to N stage ($p = 0.214$), M stage ($p = 0.122$), gender ($p = 0.154$), age ($p = 0.253$), residual tumor ($p = 0.806$), and smoking history ($p = 0.762$). Collectively, these results clarified that the low expression of DNASE1L3 gene was significantly correlated with advanced T stage and pathologic stage in LUAD.

Diagnostic Value of DNASE1L3 for Lung Adenocarcinoma

As decreased expression of DNASE1L3 had a trend to be associated with pathologic stage in LUAD patients, we hypothesized that it could be a better early diagnostic parameter for LUAD. Compared with the normal group, the expression of DNASE1L3 was significantly decreased at all pathological stages of LUAD ($p < 0.0001$, **Figure 3A**). The result showed that AUC value of DNASE1L3 for LUAD were 0.948 (95% CI, 0.927–0.968, sensitivity = 79.6% and specificity = 98.3%, **Figure 3B**). Then, the diagnosis at

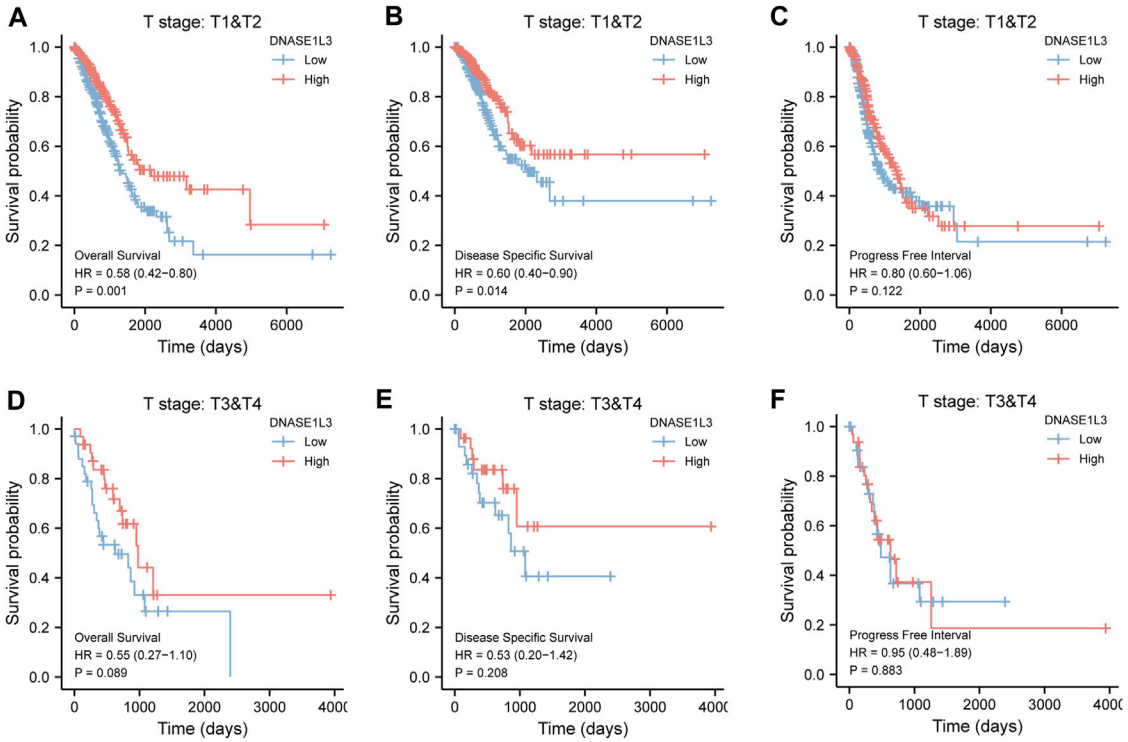


FIGURE 6 | Stratified survival analysis of DNASE1L3 in lung adenocarcinoma by T stage of OS, DSS, and PFI. **(A,D)** Stratified survival analysis of DNASE1L3 in lung adenocarcinoma of T1&T2 and T3&T4 of OS, respectively. **(B,E)** Stratified survival analysis of DNASE1L3 in lung adenocarcinoma of T1&T2 and T3&T4 of DSS, respectively. **(C,F)** Stratified survival analysis of DNASE1L3 in lung adenocarcinoma of T1&T2 and T3&T4 of PFI, respectively.

different pathological stages of LUAD was further analyzed. ROC analysis indicated that the AUC was calculated as 0.944 (strong) for stage I, 0.949 (strong) for stage II, 0.954 (strong) for stage III, and 0.952 (strong) for stage IV, respectively (**Figures 3C–F**). The diagnostic efficacy of DNASE1L3 for LUAD was also verified in GSE40791 and GSE10072 datasets. The AUC values for DNASE1L3 were 0.9984 (95% CI, 0.9752–1.000) and 0.9535 (95% CI, 0.9256–0.9815), respectively. These results indicated that DNASE1L3 could be invoked as a useful early diagnostic biomarker for lung adenocarcinoma.

Prognostic Value of DNASE1L3 in Lung Adenocarcinoma

Next, we evaluated the prognostic values of DNASE1L3 in LUAD *via* the Kaplan–Meier plotter. Significant results were shown in **Figure 4**: lower mRNA expression of DNASE1L3 associated with poorer OS, DSS, and PFI in LUAD patients, respectively (**Figure 4A** HR = 0.55, 95% CI: 0.41–0.74, and $p < 0.001$; **Figure 4D** HR = 0.54, 95% CI: 0.37–0.79, and $p = 0.001$; **Figure 4G** HR = 0.75, 95% CI: 0.58–0.98, and $p = 0.034$). In addition, we performed stratified analyses of DNASE1L3 for OS, DSS, and PFI, respectively. The stratification was conducted by age, gender, T stage, N stage, M stage, pathological stage, and smoking status.

As shown in **Figure 4**, the OS and DSS of the low expression group of elderly patients (patients older than 65 years) was poorer than that of the high expression group (**Figure 4B** HR = 0.43 (0.28–0.66) $p < 0.001$; **Figure 4E** HR = 0.48 (0.27–0.82) $p = 0.008$). But there was no difference between the two groups in young patients (age ≤ 65) (**Figures 4C,F**). Also, as for PFI, there was no difference between the two groups of young patients or elderly patients (**Figures 4H,I**). Survival analysis stratified by gender is shown in **Figure 5**. As for male patients, only the OS of the low expression group was poorer than that of the high expression group (**Figure 5A** HR = 0.57 (0.37–0.87) $p = 0.01$), whereas the OS, DSS, and PFI of the low expression group of female patients were poorer than that of the high expression group, respectively (**Figure 5D** HR = 0.48 (0.32–0.73) $p = 0.001$; **Figure 5E** HR = 0.43 (0.25–0.73) $p = 0.002$; **Figure 5F** HR = 0.69 (0.48–1.00) $p = 0.048$). To more accurately assess survival outcomes, we performed a stratified survival score for T, N, and M stages. As shown in **Figure 6**, the OS and DSS of the low expression group of T1&T2 patients were poorer than that of the high expression group (**Figure 6A** HR = 0.58 (0.42–0.80) $p = 0.001$; **Figure 6B** HR = 0.60 (0.40–0.90) $p = 0.014$). However, there was no difference between the low expression group and high expression group of T3&T4 patients (**Figures 6D,E**). There was also no statistically significant difference in PFS between the two groups of T1&T2 and T3&T4 patients (**Figures 6C,F**). Similar results were obtained when analysis was stratified by N

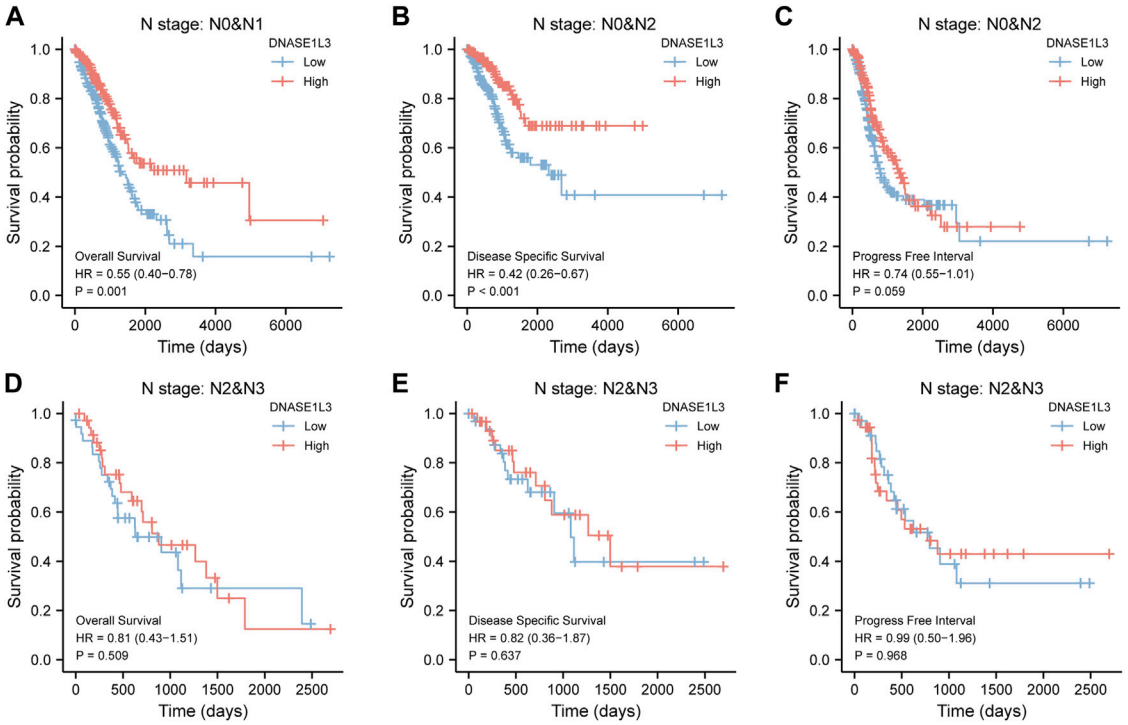


FIGURE 7 | Stratified survival analysis of DNASE1L3 in lung adenocarcinoma by N stage of OS, DSS, and PFI. **(A,D)** Stratified survival analysis of DNASE1L3 in lung adenocarcinoma of N0&N1 and N2&N3 of OS, respectively. **(B,E)** Stratified survival analysis of DNASE1L3 in lung adenocarcinoma of N0&N1 and N2&N3 of DSS, respectively. **(C,F)** Stratified survival analysis of DNASE1L3 in lung adenocarcinoma of N0&N1 and N2&N3 of PFI, respectively.

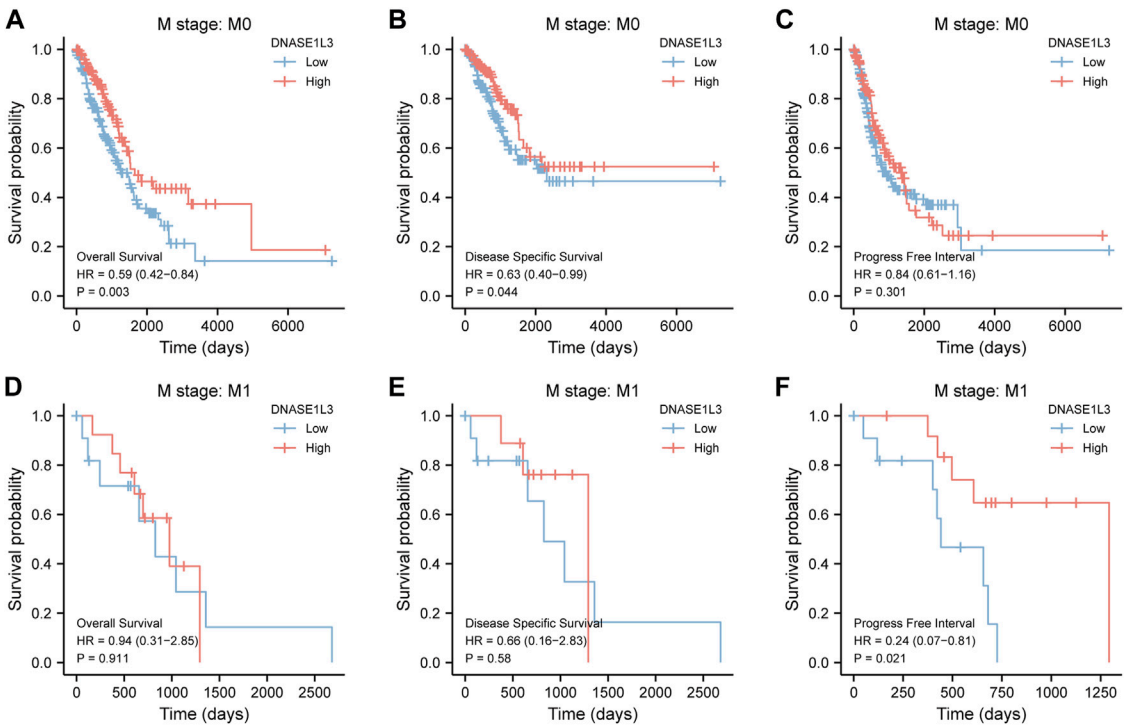


FIGURE 8 | Stratified survival analysis of DNASE1L3 in lung adenocarcinoma by M stage of OS, DSS, and PFI. **(A,D)** Stratified survival analysis of DNASE1L3 in lung adenocarcinoma of M0 and M1 of OS, respectively. **(B,E)** Stratified survival analysis of DNASE1L3 in lung adenocarcinoma of M0 and M1 of DSS, respectively. **(C,F)** Stratified survival analysis of DNASE1L3 in lung adenocarcinoma of M0 and M1 of PFI, respectively.

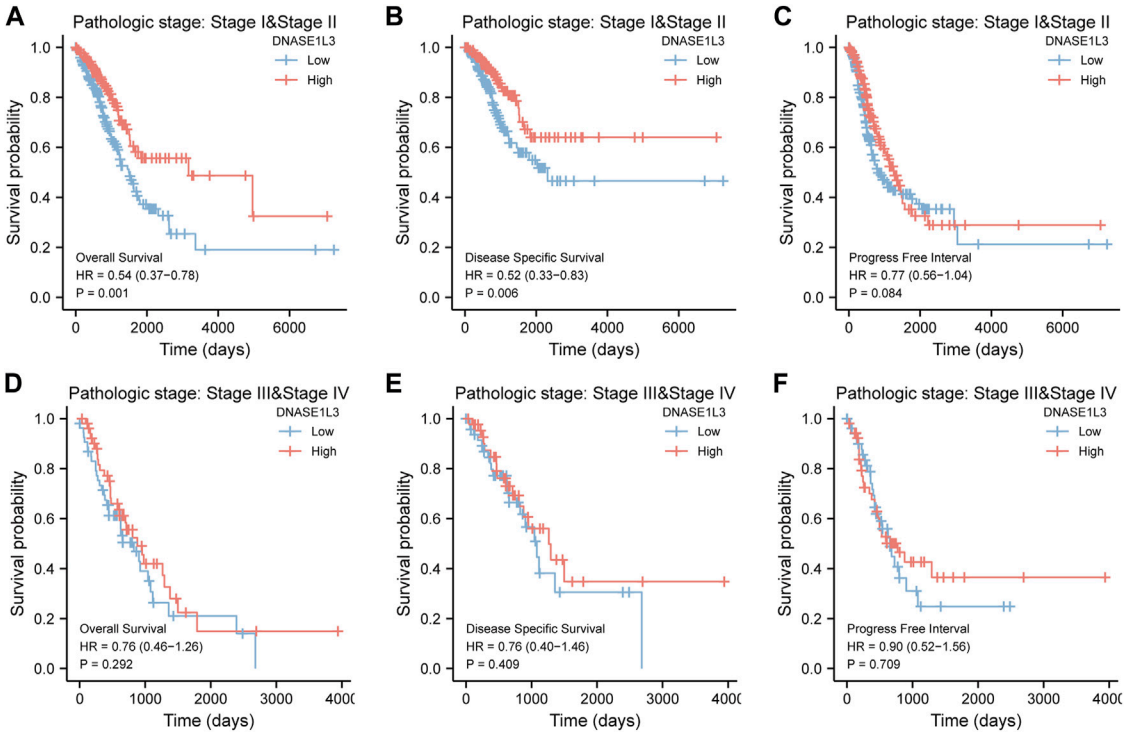


FIGURE 9 | Stratified survival analysis of DNASE1L3 in lung adenocarcinoma by pathological stage of OS, DSS, and PFI. **(A,D)** Stratified survival analysis of DNASE1L3 in lung adenocarcinoma of early stage (I &II) and late stage (III &IV) of OS, respectively. **(B,E)** Stratified survival analysis of DNASE1L3 in lung adenocarcinoma of early stage (I &II) and late stage (III &IV) of DSS, respectively. **(C,F)** Stratified survival analysis of DNASE1L3 in lung adenocarcinoma of early stage (I &II) and late stage (III &IV) of PFI, respectively.

stage (N0&N1 vs. N2&N3) (**Figures 7A–F**). The results from stratified analysis by M stage were presented in **Figure 8**. The OS and DSS of the low expression group of M0 patients were poorer than that of the high expression group (**Figure 8A** HR = 0.59 (0.42–0.84) p = 0.003; **Figure 8B** HR = 0.63 (0.40–0.99) p = 0.044). But there was no difference between the two group of M1 patients (**Figures 8D,E**). There was also no statistically significant difference in PFS between the two groups of M0 patients (**Figure 8C**). However, the PFI of the low expression group of M1patients was poorer than that of the high expression group (**Figure 8F** HR = 0.24 (0.07–0.81) p = 0.021). **Figure 9** shows results of the analysis stratified by pathological stage early stage (I & II) vs. late stage (III &IV). The result demonstrated that analysis stratified by pathological stage (**Figures 9A–F**) showed similar results with T stage (T1&T2 vs. T3&T4) and N stage (N0&N1 vs. N2&N3). The results by stratified survival analysis of DNASE1L3 in lung adenocarcinoma by smoking status of OS, DSS, and PFI are shown in **Figure 10**. The OS and DSS of the low expression group of smoking patients was poorer than that of the high expression group [**Figure 10A** HR = 0.61 (0.44–0.8) p = 0.003; **Figure 10B** HR = 0.58 (0.38–0.87) p = 0.01]. In addition, the OS of the low expression group of non-smoking patients was poorer than that of the high expression group [**Figure 10D** HR = 0.35 (0.15–0.80) p = 0.013]. However, the PFI of the low expression group of smoking patients was not significant with the high expression group (**Figure 10C**), as well as the DSS and PFI of the two groups of non-smoking patients (**Figures 10E,F**).

The prognostic significance of the DNASE1L3 level was further confirmed in a Cox regression analysis. In the univariate analysis, DNASE1L3 expression, T stage, N stage, M stage, and pathologic stage were all associated with the OS (**Table 3**, all $p < 0.05$). In the multivariate analysis, DNASE1L3 expression, along with T stage, was also associated with the OS, and the DNASE1L3 was an independent factor for the overall survival of LUAD patients (HR = 0.680; 95% CI: 0.484–0.956; p = 0.027, HR = 1.653; 95% CI: 1.020–2.680; p = 0.041) (**Table 3** and **Figure 11**). Furthermore, according to the above results, a nomogram was performed to predict the1-, 3-, and 5-years survival probability of patients by combining the expression level of DNASE1L3 with clinical parameters (**Figure 12**). Taken together, the above results suggested that DNASE1L3 could function as a potential prognostic factor in LUAD.

DNASE1L3 Correlates With Immune Infiltration in Lung Adenocarcinoma

The correlations between DNASE1L3 expression and immune infiltration were assessed with GSVA package and SSGSEA package in R. The Spearman rank correlation analysis revealed

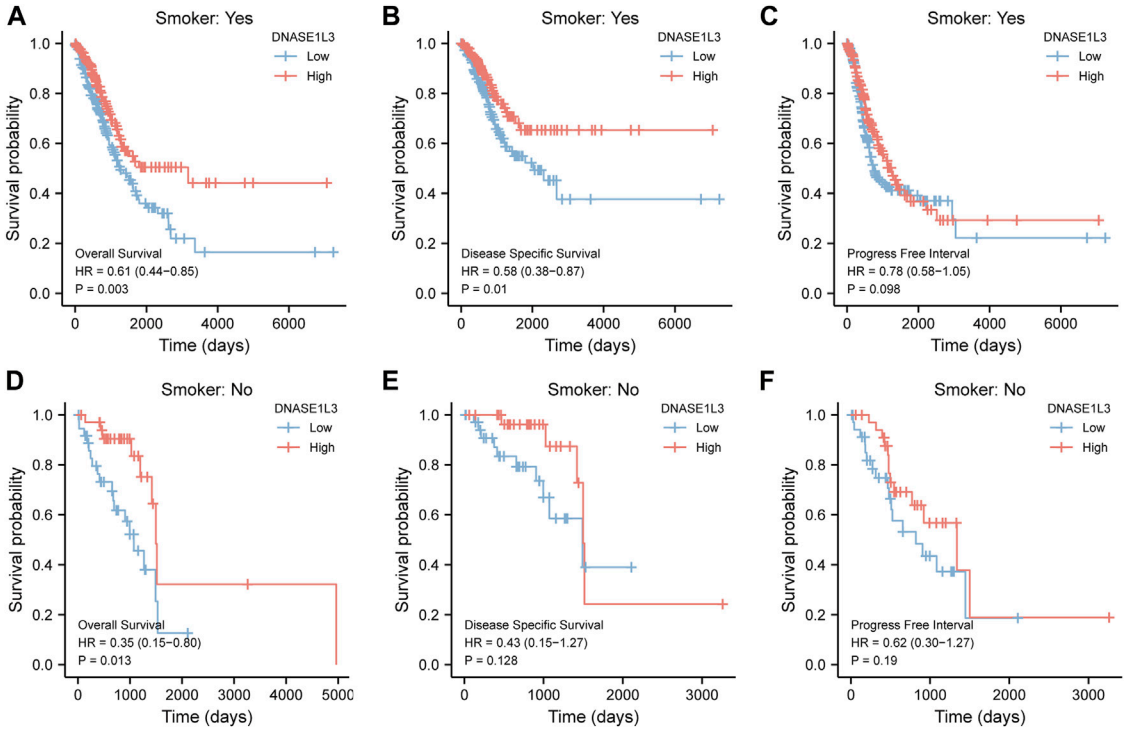


FIGURE 10 | Stratified survival analysis of DNASE1L3 in lung adenocarcinoma by smoking status of OS, DSS, and PFI. **(A,D)** Stratified survival analysis of DNASE1L3 in lung adenocarcinoma of smoking and non-smoking of OS, respectively. **(B,E)** Stratified survival analysis of DNASE1L3 in lung adenocarcinoma of smoking and non-smoking of DSS, respectively. **(C,F)** Stratified survival analysis of DNASE1L3 in lung adenocarcinoma of smoking and non-smoking of PFI, respectively.

TABLE 3 | Univariate and multivariate analysis of overall survival in patients with lung adenocarcinoma.

Characteristics	Total (N)	Univariate analysis		Multivariate analysis	
		HR (95% CI)	*p* Value	Hazard ratio (95% CI)	*p* Value
DNASE1L3	526	0.550 (0.410–0.739)	**<0.001**	0.680 (0.484–0.956)	**0.027**
T stage	523	2.317 (1.591–3.375)	**<0.001**	1.653 (1.020–2.680)	**0.041**
N stage	510	2.321 (1.631–3.303)	**<0.001**	1.313 (0.631–2.732)	0.467
M stage	377	2.136 (1.248–3.653)	**0.006**	1.094 (0.483–2.476)	0.830
Pathologic stage	518	2.664 (1.960–3.621)	**<0.001**	1.879 (0.858–4.113)	0.115
Gender	526	1.070 (0.803–1.426)	0.642	—	—
Age	516	1.223 (0.916–1.635)	0.172	—	—
Smoker	512	0.894 (0.592–1.348)	0.591	—	—

Abbreviations: HR, hazard ratio; CI, confidence interval. Bold values indicate $p < 0.05$.

significant positive correlations between DNASE1L3 expression and CD8+T cells, cytotoxic cells, DCs, neutrophils, macrophages, NK cells, Treg cells, Th1 cells, Th17 cells, B cells, and T cells (all $p < 0.05$) (**Figure 13**, **Figures 14A,B**). However, DNASE1L3 indicated a negative correlation with Th2 cell in LUAD (**Figure 13J**, **Figure 14B**). To broaden our understanding of correlations between DNASE1L3 expression and immune infiltration, correlation analysis between DNASE1L3 and several immune checkpoints LUAD were performed. As shown in the **Figure 14C**, the gene expression levels of potential immune checkpoints, including CD47, CTLA4, VSIR, CD274, HAVCR2, LAG3, PDCD1, PDCD1LG2, TIGIT, SIGLEC1, CD70, CD27, and ICOS were positively correlated with the expression of DNASE1L3 (**Figure 14C**, all $p < 0.05$). It is highly likely that DNASE1L3 is involved in LUAD immune infiltration.

Relationship Between DNASE1L3 Expression and Somatic Mutation Load

We further analyzed the association between somatic mutation load and DNASE1L3 expression in 567 LUAD cases, which

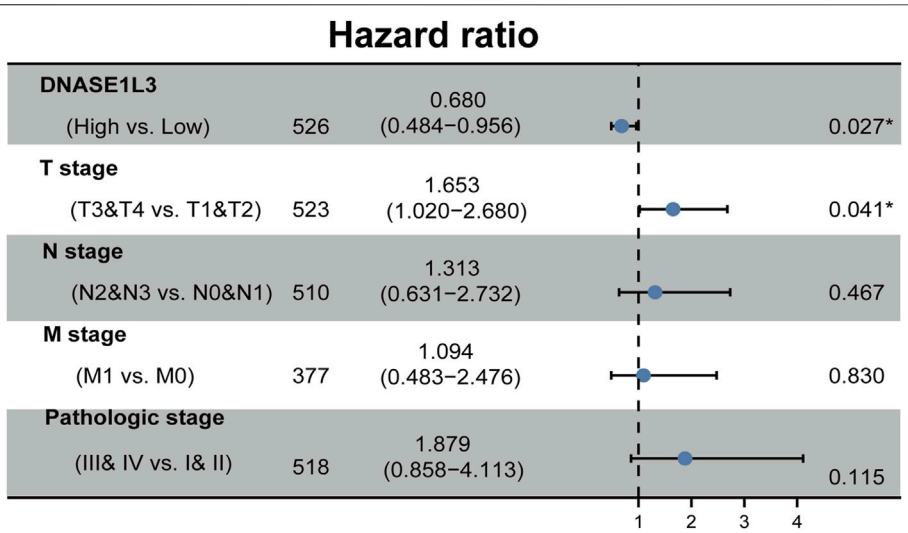


FIGURE 11 | Forest plot of the multivariate Cox regression analysis in LUAD.

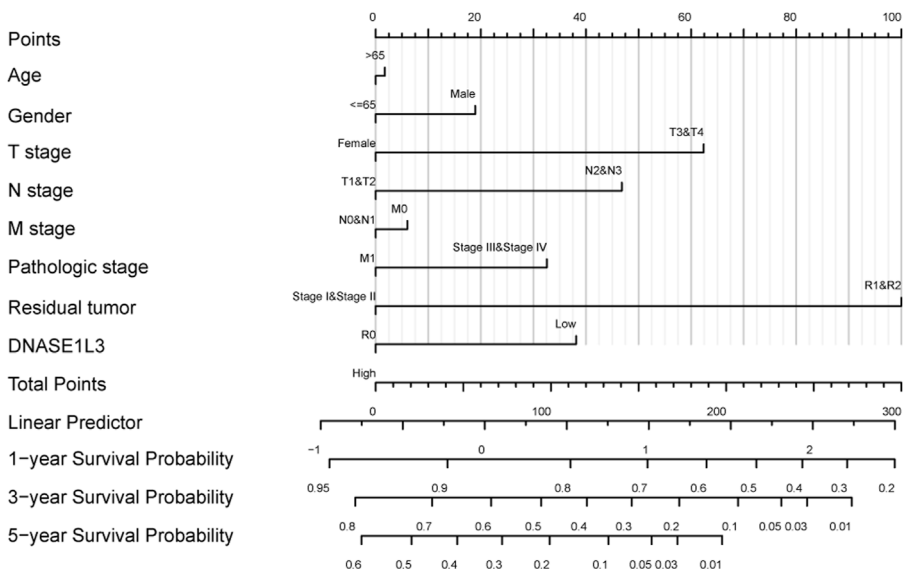


FIGURE 12 | Nomogram for predicting probability of patients with 1-, 3-, and 5-years OS.

showed missense mutation as the predominant type of variant classification (**Figures 15A,B**). Meanwhile, detailed comparisons revealed that single-nucleotide variant (SNV) in LUADs occurred more frequently than insertion or deletion (**Figure 15C**). Moreover, it was shown that C > A and C > T were found to be the main SNV class in LUADs (**Figure 15C**). In addition, the top 10 mutated genes in LUADs with ranked percentages were calculated. The gene mutation status was also shown with a waterfall plot. Waterfall plots illustrate the mutation profiles of genes in LUADs with high/low-DNASE1L3, and the mutation types are represented using various colors at the bottom of the map (Figures). Boxplots were used to exhibit the mutation frequency of each gene in high- and low-DNASE1L3 subgroups (**Figure 15B**).

Relationship Between DNASE1L3 Expression and m6A Regulatory Genes

Finally, we analyzed the expression of 20 m6A regulatory genes in LUADs (n = 535) and normal tissues (n = 59). Heat maps showed that 15 m6A-related genes were differentially expressed between LUADs and control tissues (**Supplementary Figure S1A**) and 9 m6A-related genes were differentially expressed between high-

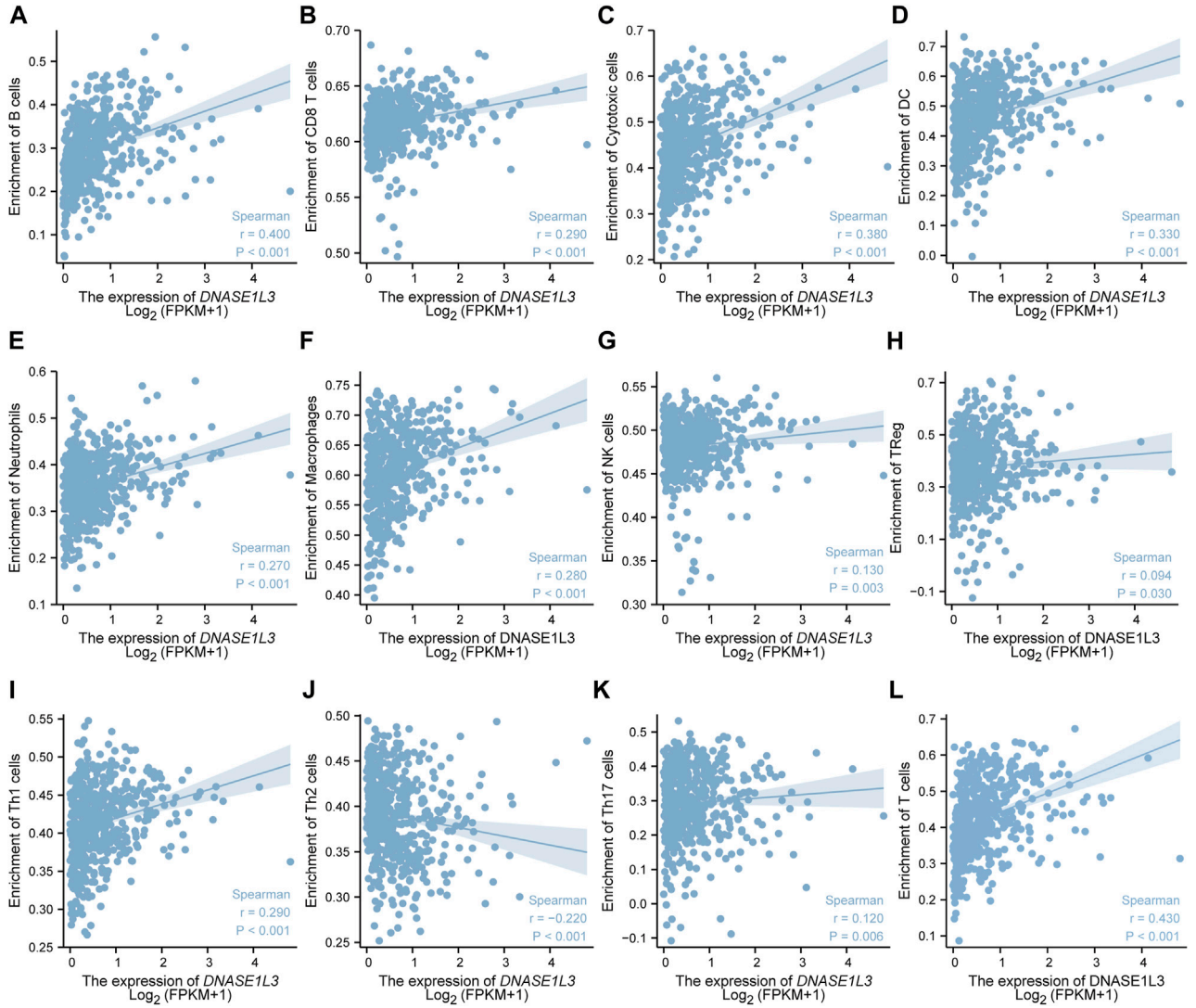


FIGURE 13 | The relationship between DNASE1L3 expression and immune infiltration level in lung adenocarcinoma. The relationship between DNASE1L3 expression and B lymphocyte infiltration **(A)**, CD8 T cell infiltration **(B)**, cytotoxic cell infiltration **(C)**, DC infiltration **(D)**, neutrophil cell infiltration **(E)**, macrophage infiltration **(F)**, NK cell infiltration **(G)**, Treg cell infiltration **(H)**, Th1 cell infiltration **(I)**, Th2 cell infiltration **(J)**, Th17 cell infiltration **(K)**, and T cell infiltration **(I)**.

and low-DNASE1L3 expression of tumor samples (**Supplementary Figure S1B**). Specifically, the expression levels of IGF2BP1, IGF2BP3, VIRMA, HNRNPC, RBMX, YTHDF1, YTHDF2, METTL3, RBM15, and HNRNPA2B1were remarkably higher in LUADs than those in normal tissues (**Supplementary Figure S1C**, all $p < 0.001$).

Enrichment Analysis of DNASE1L3-Related Partners

To further explore the potential biological function of DNASE1L3 gene in LUAD, enrichment analysist based on the DNASE1L3-relating differentially expressed genes (DEGs) was performed. According to **Supplementary Figure S2A**, 195 DEGs were obtained including 167 upregulated genes and 28 downregulated genes. Corresponding hierarchical clustering analysis of the top 5 up-regulated and down-regulated DNASE1L3 expression-correlated DEGs was displayed (**Supplementary Figure S2B**). Enrichment analyses including GO and KEGG analyses were applied to further determine the underlying molecular mechanisms of the DNASE1L3 in tumorigenesis. The results of GO analyses revealed that most of upregulated DEGs were linked to the events such as regulation of membrane potential, regulation of postsynaptic membrane potential, neurotransmitter receptor activity, and extracellular ligand-gated ion channel activity. Moreover, the results of KEGG pathway enrichment analysis of upregulated DEGs were mainly involved in neuroactive ligand-receptor interaction (**Supplementary Figure S2C,D** and **Table 4**). Furthermore, the GSEA showed DNASE1L3-associated DEGs significantly

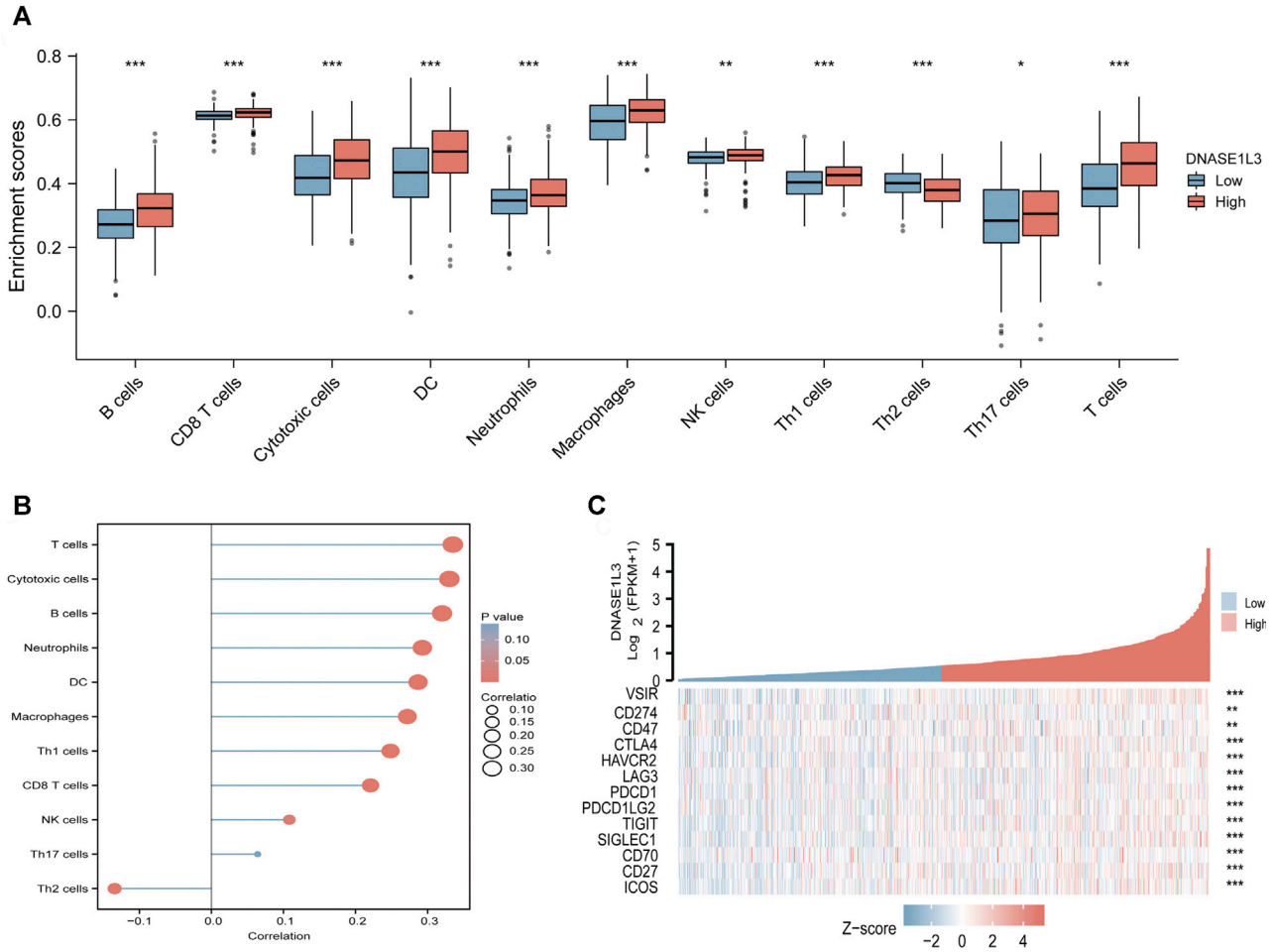


FIGURE 14 | Immune characteristics of the TCGA-LUAD subtype. **(A)** Different infiltrating abundances of 11 immune cell types estimated by between high and low expression of DNASE1L3. **(B)** The relationship between DNASE1L3 expression and immune infiltration level. **(C)** Different expression of available emerging immunotherapeutic targets between high and low expression of DNASE1L3. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

enriched in G protein-coupled receptor ligand biding (NES = 1.738; P adjust = 0.044; FDR = 0.033) and G alpha (i) signaling events (NES = 1.635; P adjust = 0.044; FDR = 0.033) (**Supplementary Figure S2E,F**). All these results demonstrated that DNASE1L3 might regulate the process of intracellular signal transduction and transmission, which could provide a new direction to the research on the crosstalk between tumor cells.

DISCUSSION

DNASE1L3 can cleave both single and double stranded DNA, generating DNA fragments with 3-OH ends (Serpas et al., 2019) and involve in intranuclear DNA fragmentation during apoptosis and necrosis. DNASE1L3 gene expression and functions in carcinoma have been recently reported (Wang et al., 2020a; Liu J. et al., 2021; Deng et al., 2021). However, the significance of its expression in prognosis and diagnosis in patients with LUAD is largely unclear.

In this study, we firstly analyzed the DNASE1L3 gene expression profile in pan-cancer *via* TIMER20. Then, we analyzed transcriptional and protein expression levels of DNASE1L3 in LUAD. We also confirmed the expression of DNASE1L3 in LUAD through the GSE40791 and GSE10072 datasets. The results demonstrated that DNASE1L3 was indeed obviously expressed at a lower level in the LUAD tissues. DNASE1L3 gene expression and its potential prognostic impact on patients with LUAD have not been evaluated. Wang et al. (Wang et al., 2020a) reported that the expression level of DNASE1L3 was significantly decreased and associated with poor overall survival in hepatocellular carcinoma. Deng et al. (Deng et al., 2021) investigated the expression levels of DNASE1L3, and the results revealed the DNASE1L3 was a prognostic biomarker in cancer of the breast, kidney, liver, stomach, lung adenocarcinoma, and sarcoma *via* bioinformatics analysis using TCGA database. This is the first comprehensive study to evaluate DNASE1L3 gene expression in the prognosis of

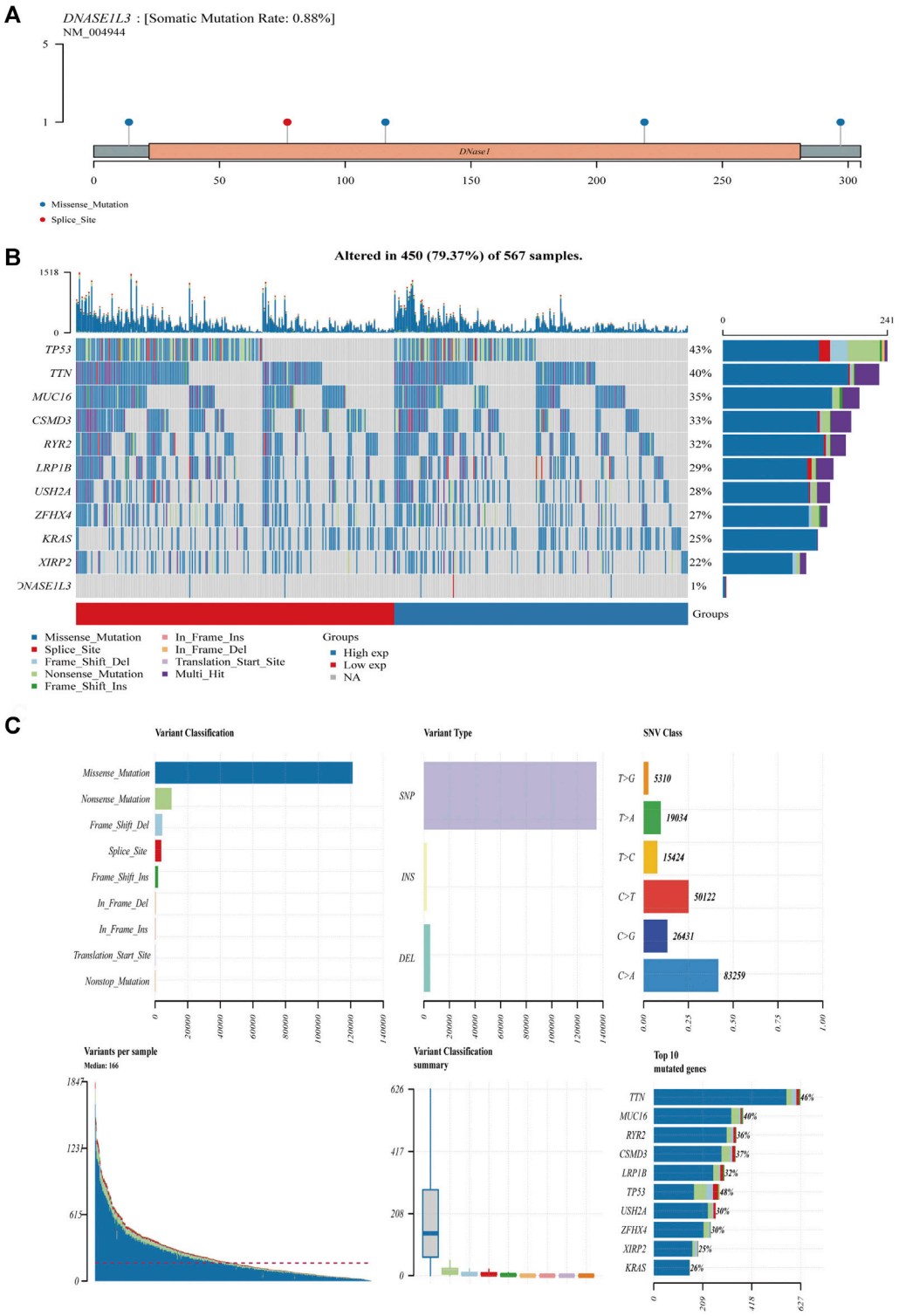


FIGURE 15 | Mutation feature of DNASE1L3 in LUAD. The alteration frequency with mutation type and mutation site of DNASE1L3 **(A)**. Distribution of frequently mutated genes in different TCGA-LUAD subgroups. The upper bar plot shows the tumor mutation burden (TMB) for each patient, whereas the left bar plot indicates the gene mutation frequency in different groups **(B)**. The top 10 LUAD-correlated mutation genes in TCGA projects and the mutation frequency, variant classification, variant type, and SNV class of the mutated genes in high- and low-DNASE1L3 subgroups **(C)**.

TABLE 4 | Results of GO enrichment and KEGG pathway analysis.

Ontology	ID	Description	Gene ratio	*p* value	*p*.adjust	*q* value
BP	GO: 0003341	Cilium movement	7/127	3.92e-07	7.81e-04	7.15e-04
BP	GO: 0007586	Digestion	8/127	4.91e-06	0.005	0.004
BP	GO: 0042391	Regulation of membrane potential	13/127	8.33e-06	0.006	0.005
BP	GO: 0060078	Regulation of postsynaptic membrane potential	7/127	4.85e-05	0.016	0.015
BP	GO: 0001976	Neurological system process involved in regulation of systemic arterial blood pressure	3/127	4.87e-05	0.016	0.015
CC	GO: 0098889	Intrinsic component of presynaptic membrane	10/137	2.63e-10	6.19e-08	4.77e-08
CC	GO: 0099240	Intrinsic component of synaptic membrane	12/137	1.66e-09	1.51e-07	1.16e-07
CC	GO: 0099056	Integral component of presynaptic membrane	9/137	1.93e-09	1.51e-07	1.16e-07
CC	GO: 0099699	Integral component of synaptic membrane	11/137	9.14e-09	5.37e-07	4.14e-07
CC	GO: 0042734	Presynaptic membrane	10/137	1.88e-07	8.82e-06	6.79e-06
MF	GO: 0030594	Neurotransmitter receptor activity	6/126	1.90e-04	0.033	0.029
MF	GO: 0005230	Extracellular ligand-gated ion channel activity	5/126	1.96e-04	0.033	0.029
MF	GO: 0099529	Neurotransmitter receptor activity involved in regulation of postsynaptic membrane potential	4/126	4.38e-04	0.038	0.033
MF	GO: 0098960	Postsynaptic neurotransmitter receptor activity	4/126	5.09e-04	0.038	0.033
MF	GO: 0015464	Acetylcholine receptor activity	3/126	5.62e-04	0.038	0.033
KEGG	hsa04080	Neuroactive ligand-receptor interaction	12/62	8.49e-06	9.09e-04	8.49e-04

Note: BP, biological process; CC, Cellular Component; MF: Molecular Function. P adjust value < 0.05 and q-value <0.25 were considered as significantly enriched.

patients with LUAD. Analysis revealed that DNASE1L3 gene expression was negatively associated with OS event, pathologic, stage and T stage. Multivariate Cox analysis further confirmed that low DNASE1L3 expression was an independent prognostic factor for OS in patients with LUAD; T stage was associated with worse prognosis in LUAD, as suggested by the forest plot. Additionally, Kaplan–Meier plotter results suggested that the down-regulation of DNASE1L3 indicated poor prognosis of LUAD, which specifically reflected on the advanced clinical characteristics of tumor pathological stages and the depth of the primary tumor invasion. These results suggest that DNASE1L3 plays a tumor suppressive role in LUAD. In this paper, we noted that the M stage represented the state of distant metastasis, but had no effect on the expression of DNASE1L3. Combined with subsequent analysis, it was found that DNASE1L3 may affect tumor progression mainly by regulating the immune microenvironment, but may not promote distant metastasis of tumor in tumor tissues.

Gender is a key factor affecting individual cancer progression. There are significant gender differences in the incidence, aggressiveness, prognosis, and treatment response of various tumors. However, as shown in **Figure 1C**, there was no difference in the expression of DNASE1L3 between male and female cancer patients. As the body ages, humans suffer many diseases, including tumors. In this study, the patients were divided into two groups by age of 65; it was found that age was not related to the expression of DNASE1L3, nor was it direct factor affecting the prognosis of LUAD patients in the TCGA-LUAD dataset, as can be seen in **Table 2**. Smoking is the leading risk factor for lung cancer. However, in this paper, smoking or not is not related to the expression of DNASE1L3, and the specific reasons need to be further explored.

Tumor immune cell infiltration plays an important role in tumor prognosis and influences the response to immunotherapy (Zhang et al., 2016; Cavnar et al., 2017; Titov et al., 2021). Currently, several studies have demonstrated the importance of tumor-infiltrating immune cells and other immune molecules (including tumor-associated macrophages, natural killer cells, and dendritic cells) in the prognosis of lung adenocarcinoma (Liu W. et al., 2021; Mu et al., 2021; Shen et al., 2021). Therefore, we further analyzed the infiltration rates of various immune cells with the low and high expression levels of DNASE1L3 gene. We observed that the level of DNASE1L3 expression was positively linked to immune infiltration, especially for DCs, Mphs, and NEUs, which is similar to previous researches. In addition, in order to gain deeper insights into the immune landscape of LUAD, the expression of immune checkpoint molecules was investigated. By assessing the relationship between immune checkpoint molecule expression and DNASE1L3 profile, we found heterogeneity in the expression of

immune checkpoint proteins in the immune microenvironment of LUAD. From the above results, we infer that the expression of DNASE1L3 involved in the immune infiltration may potentially impact the occurrence and development of lung adenocarcinoma.

As observed by Bhalla et al., DNASE1L3 was overexpressed in early stage of clear cell renal cancer and has some diagnostic ability to distinguish between early and late stage tumors (Bhalla et al., 2017a). However, its diagnostic significance in lung adenocarcinoma has not been reported. Our research first suggests and validates the diagnostic (strong) value of DNASE1L3 for lung adenocarcinoma.

Tumor mutation burden (TMB) is considered essential factors impacting on the occurrence and progression of tumor. Previously, Yue et al. (Yue et al., 2019) reported that a TMB relating gene has predictive accuracy for OS in LUAD patients. In our study, the mutation status of several cancer-related genes with high mutation probability in LUAD significantly affected the expression of DNASE1L3. The correlation between these mutated genes and the expression of DNASE1L3 suggests that differential expression of DNASE1L3 may play a regulatory role in LUAD. Nevertheless, further studies are needed to verify this hypothesis.

As a new dimension of gene expression control, RNA N6-methyladenosine (m6A) modification has attracted great academic interest in recent years. As the most abundant mRNA modification, m6A is involved in the regulation of the occurrence and development of tumors by controlling the expression of key genes, which has become the focus of research in recent years (Zhao et al., 2020). Previous studies have reported that abnormal m6A methylation modification may impact the development of lung cancer (Cheng et al., 2021). Our result showed that there are nine m6A-related regulators that were differentially expressed between high- and low-DNASE1L3 expression of LUAD. We speculate that the down expression of DNASE1L3 in lung adenocarcinoma is highly likely to be related to the abnormal regulation of m6A-related regulators; of course, the clear mechanism needs to be further studied.

In addition, we found that the up-expression phenotype of DNASE1L3 was associated with the events such as regulation of membrane potential, regulation of postsynaptic membrane potential, neurotransmitter receptor activity, and extracellular ligand-gated ion channel activity. Moreover, KEGG analysis revealed that DNASE1L3-associated DEGs were mainly involved in neuroactive ligand-receptor interaction. Furthermore, the GSEA showed significantly enriched in G protein-coupled receptor ligand biding and G alpha (i) signaling events. All these results indicated that DNASE1L3 might regulate the process of intracellular signal transduction and transmission, which could provide a new direction to the research on the crosstalk between tumor cells. However, such mechanisms require further investigation.

In a word, this is the first study to comprehensively analyze the relationship between DNASE1L3 expression level and early diagnosis and prognosis in patients with lung adenocarcinoma. Inevitably, this research has several limitations that need to be addressed. First, because the prognosis of DNASE1L3 in this research was based on data onto the TCGA datasets, additional clinical data are needed to verify it. Second, due to the small sample size, the effect of gene mutation and methylation of DNASE1L3 on prognosis cannot be carried out, and further supplement is needed. Thirdly, our study results are only limited to mRNA level, so it needs to be verified at protein level. In addition, there are still many questions to be solved, such as whether DNASE1L3 is related to chemotherapy resistance of LUAD, and how DNASE1L3 expression changes after chemotherapy? Meanwhile, what is the specific mechanism of DNASE1L3 in LUAD? Therefore, further molecular mechanism research is needed.

CONCLUSION

In summary, this study provides novel evidences for the clinical and biological significance of DNASE1L3 in LUAD. Our results demonstrated that both the mRNA and protein levels of DNASE1L3 were noticeably downregulated in LUADs compared with normal tissues. The low expression level of DNASE1L3 was significantly associated with higher pathological stages, higher T stages, and poor prognosis in LUAD, and the DNASE1L3 expression level might be an independent prognostic factor of LUAD. Furthermore, from a series of bioinformatics analysis, we found mRNA level of DNASE1L3 was positively linked to the degree of infiltration of various tumor-infiltrating immune cells, the gene expression levels of potential immune checkpoint molecules, and some m6A methylation regulators in LUAD. Finally, DNASE1L3 showed strong early diagnostic value for LUAD. We conclude that the low expression level of DNASE1L3 may serve as a clinically useful diagnostic and prognostic biomarker, and potentially as a therapeutic target in LUAD.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga.

AUTHOR CONTRIBUTIONS

1) Conception and design: JC and YS; 2) Administrative support: YS; 3) Provision of study materials or patients: WH, LS, JC, and YL; 4) Collection and assembly of data: JC; 5) Data analysis and interpretation: JC and JD; 6) Manuscript writing: All authors; 7) Final approval of manuscript: All authors.

FUNDING

The following grant information was disclosed by the authors: Liuzhou Scientific Research and Technological Development Programs: 2014JC010. Guangxi Zhuang Autonomous Region Health and Family Planning Commission: Z2014613. Guangxi Zhuang Autonomous Region Health and Family Planning Commission:

Z20200378. Guangxi Zhuang Autonomous Region Health and Family Planning Commission: Z20200841.

ACKNOWLEDGMENTS

We gratefully appreciate the efforts and contributions of all members of the research group.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.699242/full#supplementary-material

Supplementary Figure S1 | Expression of m6A modification regulators in LUAD. (A) The heat map visualizes the expression levels of m6A RNA modification regulators in each tumor samples and normal control samples. (B) The expression levels of m6A RNA modification regulators in each high- and low-DNASE1L3 expression of tumor samples. (C) The box figure shows the differentially m6A RNA modification regulators in LUAD. The significance of the two groups of samples passed the Mann-Whitney U test. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{**}p < 0.001$. **Abbreviations:** m6A: N6-methyladenosine.

Supplementary Figure S2 | Enrichment analysis of DNASE1L3 expression-correlated DEGs in LUAD. (A) Volcano plot of DEGs between samples with high DNASE1L3 expression and low DNASE1L3 expression. (B) Heat maps of the top 5 up-regulated and down-regulated DNASE1L3 expression-correlated DEGs. (C) KEGG enrichment and GO enrichment analysis by DNASE1L3 expression-correlated upregulated DEGs. (D) Interaction networks of enriched biological processes. (E-F) Functional annotation of DEGs in LUAD patients with distinct DNASE1L3 levels. NES: normalized enrichment score; FDR: false discovery rate. FDR<0.25 and P adjust<0.05 is generally considered as significant enrichment.

REFERENCES

Al-Mayouf, S. M., Sunker, A., Abdwani, R., Abrawi, S. A., Almurshedi, F., Alhashmi, N., et al. (2011). Loss-of-function Variant in DNASE1L3 Causes a Familial Form of Systemic Lupus Erythematosus. *Nat. Genet.* 43 (12), 1186–1188. doi:10.1038/ng.975

Anand, M., De-Chen, L., Yassen, A., Plass, C., and Koeffler, H. P. (2018). Maftools: Efficient and Comprehensive Analysis of Somatic Variants in Cancer. *Genome Res.* 28 (11), 1747–1756. doi:10.1101/gr.239244.118

Bhalla, S., Chaudhary, K., Kumar, R., Sehgal, M., Kaur, H., Sharma, S., et al. (2017a). Gene Expression-Based Biomarkers for Discriminating Early and Late Stage of clear Cell Renal Cancer. *Sci. Rep.* 7, 44997. doi:10.1038/srep44997

Cavnar, M. J., Turcotte, S., Katz, S. C., Kuk, D., Gönen, M., Shia, J., et al. (2017). Tumor-Associated Macrophage Infiltration in Colorectal Cancer Liver Metastases Is Associated with Better Outcome. *Ann. Surg. Oncol.* 24 (7), 1835–1842. doi:10.1245/s10434-017-5812-8

Chen, F., Chandrashekar, D. S., Varambally, S., and Creighton, C. J. (2019). Pan-cancer Molecular Subtypes Revealed by Mass-Spectrometry-Based Proteomic Characterization of More Than 500 Human Cancers. *Nat. Commun.* 10, 5679. doi:10.1038/s41467-019-13528-0

Cheng, Y., Wang, M., Zhou, J., Dong, H., Wang, S., and Xu, H. (2021). The Important Role of N6-Methyladenosine RNA Modification in Non-small Cell Lung Cancer. *Genes* 12 (3), 440. doi:10.3390/genes12030440

Deng, Z., Xiao, M., Du, D., Luo, N., Liu, D., Liu, T., et al. (2021). DNASE1L3 as a Prognostic Biomarker Associated with Immune Cell Infiltration in Cancer. *Ott* 14, 2003–2017. doi:10.2147/ott.S294332

Denisenko, T. V., Budkevich, I. N., and Zhivotovsky, B. (2018). Cell Death-Based Treatment of Lung Adenocarcinoma. *Cell Death Dis* 9 (2), 117. doi:10.1038/s41419-017-0063-y

Errami, Y., Naura, A. S., Kim, H., Ju, J., Suzuki, Y., Elbahrawy, A. H., et al. (2012). Apoptotic DNA Fragmentation May Be a Cooperative Activity between Caspase-Activated DNASE and the PARP-Regulated DNAS1L3, an ER-Localized Endonuclease that Translocates to the Nucleus during Apoptosis, *J. Biol. Chem.* 288 (5), 3460 doi:10.1074/jbc.M112.423061

Gehring, J. S., Bernd, F., Michael, L., and Wolfgang, H. J. B. (2015). SomaticSignatures: Inferring Mutational Signatures from Single-Nucleotide Variants. *Bioinformatics* 31 (22), 3673–3675.

Han, D. S. C., Ni, M., Chan, R. W. Y., Chan, V. W. H., Lui, K. O., Chiu, R. W. K., et al. (2020). The Biology of Cell-free DNA Fragmentation and the Roles of DNASE1, DNASE1L3, and DFFB. *Am. J. Hum. Genet.* 106 (2), 202–214. doi:10.1016/j.ajhg.2020.01.008

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC Bioinformatics* 14 (1), 7. doi:10.1186/1471-2105-14-7

Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. S., et al. (2017). TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* 77 (21), e108–e110. doi:10.1158/0008-5472.Can-17-0307

Liu, J., Lichtenberg, T., Hoadley, K. A., Poisson, L. M., Lazar, A. J., Cherniack, A. D., et al. (2018). An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* 173 (2), 400. doi:10.1016/j.cell.2018.02.052

Liu, J., Yi, J., Zhang, Z., Cao, D., Li, L., and Yao, Y. (2021a). Deoxyribonuclease 1-like 3 May Be a Potential Prognostic Biomarker Associated with Immune Infiltration in colon Cancer. *Aging* 13 (12), 16513–16526. doi:10.18632/aging.203173

Liu, W., Jiang, K., Wang, J., Mei, T., Zhao, M., and Huang, D. (2021b). Upregulation of GNPNAT1 Predicts Poor Prognosis and Correlates with Immune Infiltration in Lung Adenocarcinoma. *Front. Mol. Biosci.* 8, 605754. doi:10.3389/fmolb.2021.605754

Liu, X., Xiang, D., Xu, C., and Chai, R. (2021c). EIF3m Promotes the Malignant Phenotype of Lung Adenocarcinoma by the Up-Regulation of Oncogene CAPRIN1. *Am. J. Cancer Res.* 11 (3), 979–996.

Melissa Haig, J. D., Dahlke, J., Haig, M., Wohlwend, L., and Malecki, R. (2013). Eradication of Human Ovarian Cancer Cells by Transgenic Expression of Recombinant DNASE1, DNASE1L3, DNASE2, and DFFB Controlled by EGFR Promoter: Novel Strategy for Targeted Therapy of Cancer§. *J. Genet. Syndr. Gene Ther.* 04 (6), 152. doi:10.4172/2157-7412.1000152

Mu, L., Ding, K., Tu, R., and Yang, W. (2021). Identification of 4 Immune Cells and a 5-lncRNA Risk Signature with Prognosis for Early-Stage Lung Adenocarcinoma. *J. Transl Med.* 19 (1), 127. doi:10.1186/s12967-021-02800-x

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 (1), 77. doi:10.1186/1471-2105-12-77

Rodriguez, A. M., Rodin, D., Nomura, H., Morton, C. C., Weremowicz, S., and Schneider, M. C. (1997). Identification, Localization, and Expression of Two Novel Human Genes Similar to Deoxyribonuclease I. *Genomics* 42 (3), 507–513. doi:10.1006/geno.1997.4748

Serpas, L., Chan, R. W. Y., Jiang, P., Ni, M., Sun, K., Rashidfarrokhi, A., et al. (2019). Dnase1l3 Deletion Causes Aberrations in Length and End-Motif Frequencies in Plasma DNA. *Proc. Natl. Acad. Sci. USA* 116 (2), 641–649. doi:10.1073/pnas.1815031116

Shen, E., Han, Y., Cai, C., Liu, P., Chen, Y., Gao, L., et al. (2021). Low Expression of NLRP1 Is Associated with a Poor Prognosis and Immune Infiltration in Lung Adenocarcinoma Patients. *Aging* 13 (5), 7570–7588. doi:10.18632/aging.202620

Shiokawa, D., and Tanuma, S.-i. (2001). Characterization of Human DNase I Family Endonucleases and Activation of DNase γ during Apoptosis. *Biochemistry* 40 (1), 143–152. doi:10.1021/bi001041a

Sisirak, V., Sally, B., D'Agati, V., Martinez-Ortiz, W., Özçakar, Z. B., David, J., et al. (2016). Digestion of Chromatin in Apoptotic Cell Microparticles Prevents Autoimmunity. *Cell* 166 (1), 88–101. doi:10.1016/j.cell.2016.05.034

Sjöblom, T., Jones, S., Wood, L. D., Parsons, D. W., Lin, J., Barber, T. D., et al. (2006). The Consensus Coding Sequences of Human Breast and Colorectal Cancers. *Science* 314 (5797), 268–274. doi:10.1126/science.1133427

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 (43), 15545–15550. doi:10.1073/pnas.0506580102

Sun, Y., Ouyang, B., Xie, Q., Wang, L., Zhu, S., and Jia, Y. (2020). Serum Deoxyribonuclease 1-like 3 Is a Potential Biomarker for Diagnosis of Ankylosing Spondylitis. *Clin. Chim. Acta* 503, 197–202.

Titov, A., Zmievskaya, E., Ganeeva, I., Valiullina, A., Petukhov, A., Rakhmatullina, A., et al. (2021). Adoptive Immunotherapy beyond CAR T-Cells. *Cancers* 13 (4), 743. doi:10.3390/cancers13040743

Wang, S., Ma, H., Li, X., Mo, X., Zhang, H., Yang, L., et al. (2020a). DNASE1L3 as an Indicator of Favorable Survival in Hepatocellular Carcinoma Patients Following Resection. *Aging* 12 (2), 1171–1185. doi:10.18632/aging.102675

Wickham, H. (2016). *ggplot2 Egant Graphics for Data Analysis*. Cham: Springer International Publishing. doi:10.1007/978-3-319-24277-4

Wilber, A., O'Connor, T. P., Lu, M. L., Karimi, A., and Schneider, M. C. (2003). Dnase1l3 Deficiency in Lupus-Prone MRL and NZB/W F1 Mice. *Clin. Exp. Immunol.* 134 (1), 46–52. doi:10.1046/j.1365-2249.2003.02267.x

Xu, B., Lv, W., Li, X., Zhang, L., and Lin, J. (2019). Prognostic Genes of Hepatocellular Carcinoma Based on Gene Coexpression Network Analysis. *J. Cel Biochem* 120 (7), 11616–11623. doi:10.1002/jcb.28441

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

Yue, C., Ma, H., and Zhou, Y. (2019). Identification of Prognostic Gene Signature Associated with Microenvironment of Lung Adenocarcinoma. *PeerJ* 7(3), e8128. doi:10.7717/peerj.8128

Zhang, H., Liu, H., Shen, Z., Lin, C., Wang, X., Qin, J., et al. (2016). Tumor-infiltrating Neutrophils Is Prognostic and Predictive for Postoperative Adjuvant Chemotherapy Benefit in Patients with Gastric Cancer. *Ann. Surg.* 267 (2), 311–318. doi:10.1097/SLA.0000000000002058

Zhao, Z., Yang, L., Fang, S., Zheng, L., Wu, F., Chen, W., et al. (2020). The Effect of m6A Methylation Regulatory Factors on the Malignant Progression and Clinical Prognosis of Hepatocellular Carcinoma. *Front. Oncol.* 10, 1435. doi:10.3389/fonc.2020.01435

Zochling, J., Newell, F., Charlesworth, J. C., Leo, P., Stankovich, J., Cortes, A., et al. (2014). An Immunochip-Based Interrogation of Scleroderma Susceptibility Variants Identifies a Novel Association at DNASE1L3. *Arthritis Res. Ther.* 16 (5), 438. doi:10.1186/s13075-014-0438-8

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

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 31 March 2022
doi: 10.3389/fgene.2022.810193

Integrative Analysis of Multi-Omics Data-Identified Key Genes With KLRC3 as the Core in a Gene Regulatory Network Related to Immune Phenotypes in Lung Adenocarcinoma

Kai Mao[1,2†], Yunxi Zhao[3†], Bo Ding[1,2], Peng Feng[4], Zhenqing Li[1,2], You Lang Zhou[2] and Qun Xue[1,2]**

[1]Cardiovascular Surgery Department, Affiliated Hospital of Nantong University, Nantong, China, [2]Research Center of Clinical Medicine, Affiliated Hospital of Nantong University, Nantong, China, [3]Department of Neurology, Rugao People's Hospital, Nantong, China, [4]Neurosurgery, Shaanxi Provincial People's Hospital, Shaanxi, China

In a recent study, the PD-1 inhibitor has been widely used in clinical trials and shown to improve various cancers. However, PD-1/PD-L1 inhibitors showed a low response rate and were effective for only a small number of cancer patients. Thus, it is important to figure out the issue about the low response rate of immunotherapy. Here, we performed ssGSEA and unsupervised clustering analysis to identify three clusters (clusters A, B, and C) according to different immune cell infiltration status, prognosis, and biological action. Of them, cluster C showed a better survival rate, higher immune cell infiltration, and immunotherapy effect, with enrichment of a variety of immune active pathways including T and B cell signal receptors. In addition, it showed more significant features associated with immune subtypes C2 and C3. Furthermore, we used WGCNA analysis to confirm the cluster C-associated genes. The immune-activated module highly correlated with 111 genes in cluster C. To pick candidate genes in SD/PD and CR/PR patients, we used the least absolute shrinkage (LASSO) and SVM-RFE algorithms to identify the targets with better prognosis, activated immune-related pathways, and better immunotherapy. Finally, our analysis suggested that there were six genes with KLRC3 as the core which can efficiently improve immunotherapy responses with greater efficacy and better prognosis, and our study provided clues for further investigation about target genes associated with the higher response rate of immunotherapy.

Keywords: lung adenocarcinoma, PD-1 inhibitor, LASSO analysis and SVM-RFE, immune cell infiltration, TCGA

OPEN ACCESS

Edited by:
Anton A. Buzdin, I.M. Sechenov First Moscow State Medical University, Russia

Reviewed by:
Alessandro Rizzo, National Cancer Institute Foundation (IRCCS), Italy
Farzam Vaziri, University of California, Davis, United States

****Correspondence:***
You Lang Zhou
zhouyoulang@ntu.edu.cn
Qun Xue
lee17706073991@163.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *01 December 2021*
Accepted: *04 March 2022*
Published: *31 March 2022*

Citation:
Mao K, Zhao Y, Ding B, Feng P, Li Z, Zhou YL and Xue Q (2022) Integrative Analysis of Multi-Omics Data-Identified Key Genes With KLRC3 as the Core in a Gene Regulatory Network Related to Immune Phenotypes in Lung Adenocarcinoma. Front. Genet. 13:810193. doi: 10.3389/fgene.2022.810193

INTRODUCTION

According to recent research, lung cancer is a highly malignant type of cancer with poor prognosis. It ranks among the top cancers in terms of morbidity and mortality (Jemal et al., 2011), wherein about 60% of the patients in the early stages of lung cancer underwent combined treatment of surgery, chemotherapy, and radiotherapy but did not achieve satisfactory therapeutic effect, and cancer-driver

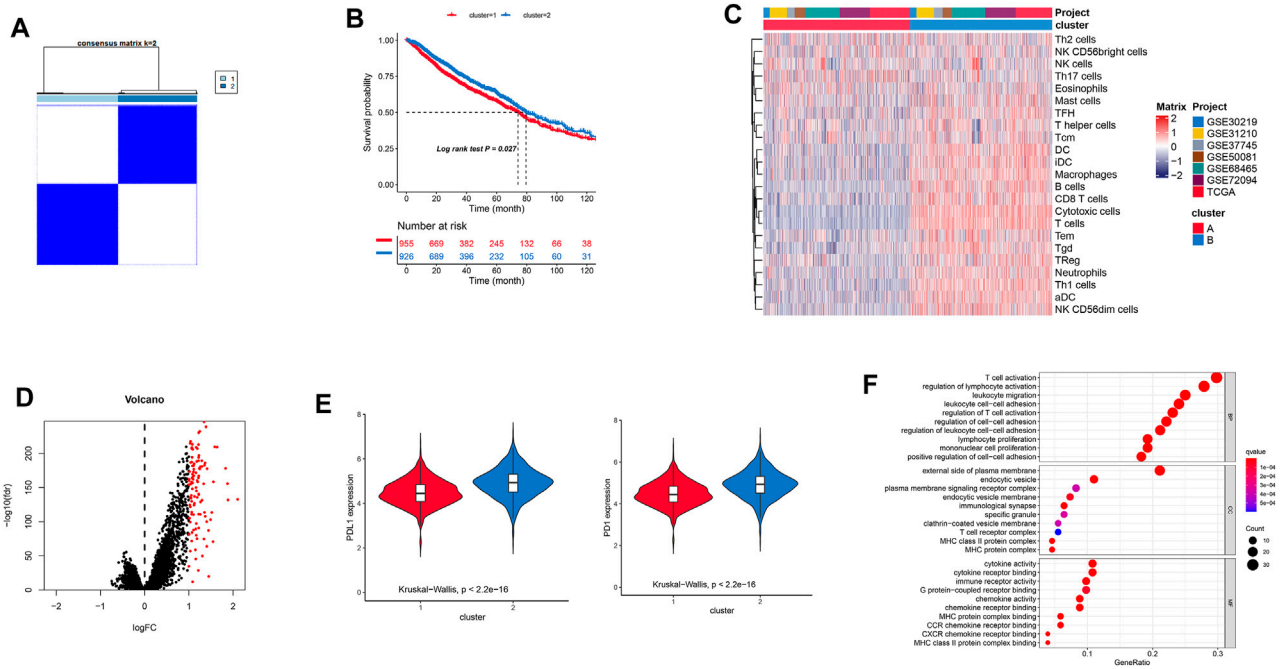


FIGURE 1 | (A) Consensus clustering-based identification of two clusters (n = 1,881). Sample consensus is displayed by heatmaps shown in white (consensus value = 0) for samples that never aggregated jointly and blue color (consensus value = 1) for samples that always aggregated jointly. **(B)** Survival analysis of patients with two clusters. **(C)** Landscape immune cell infiltration in two clusters. **(D)** Differential gene expressions are imaged by a volcano plot. **(E)** Different expressions of PD-1/PD-L1 in the two clusters. **(F)** GO analysis showing differential gene expressions.

gene-targeted therapy also encounters the problem of drug resistance [27]. The emergence of immunotherapy brings hope to lung cancer patients. Also, the interactions of PD-1 and its ligands constitute a major immunosuppressive pathway in tumors (Brahmer et al., 2012; Robert et al., 2015). Since 2010, the PD-1 or PD-L1 antibodies have been showing significant antitumor activity, including NSCLC (Ohe et al., 2007; Herbst et al., 2015). In addition to activating T-cells by binding to its ligands (PD-L1 and PD-L2) to enhance T-cell antitumor functions, the PD-1 effect was also seen in other immune cells (Barber et al., 2006; Sharpe et al., 2007).

Nowadays, several therapeutic antibodies with PD-1- or PD-L1-suppressing properties have been formulated for managing malignancies such as NSCLC clinically. However, these immunotherapies show a low remission rate in patients. Thus, a great deal of effort has been undertaken to find predictive biomarkers for patients with optimal response to these inhibitions, and there were also studies devoted to finding biological targets to improve the response rate of immunotherapy. Hu et al. analyzed the differences in the gene expression profiles of patients with high and low expression of PD-1 and PD-L1 and found that more than one hundred genes, including IL-21, KLRC3, and KLRC4, were significantly upregulated in the high-expression group, compared with the low-expression group [26]. In the present, it is urgent to confirm the targets that can be applied to improving the response rate of immunotherapy.

Here, we performed the ssGSEA and unsupervised clustering analysis to identify targets with better prognosis, immunotherapy, and that are activated in immune-related pathways based on the red module for 111 genes that are highly correlated with cluster C based on the LASSO and SVM-RFE analyses. The six genes with KLRC3 as the core were identified as the key genes with a better prognosis and correlated with immunotherapy.

MATERIALS AND METHODS

Datasets and Samples

The gene expressions of a total of 1881 patients with detailed survival information obtained from TCGA-LUAD (https://portal.gdc.cancer.gov/) and GEO datasets of GSE31210, GSE30219, GSE68465, GSE37745, GSE50081, and GSE72094 were generated. The expression values were log-transformed, and the "ComBat" algorithm was used for reducing probable batch effects resulting from the inter-dataset biases (non-biotech) (Johnson et al., 2007).

Gene Signature and Single-Sample Gene Set Enrichment Analysis

A set of marker genes for types of immune cells was selected based on Bindea et al. (2013). For enrichment computation in the individual sample gene set, the absolute enrichment fractions were derived via the GSEA program for traits that have been validated by prior experimentation. For confirming the immune cell populations, the ssGSEA analysis of each sample was accomplished using the immune

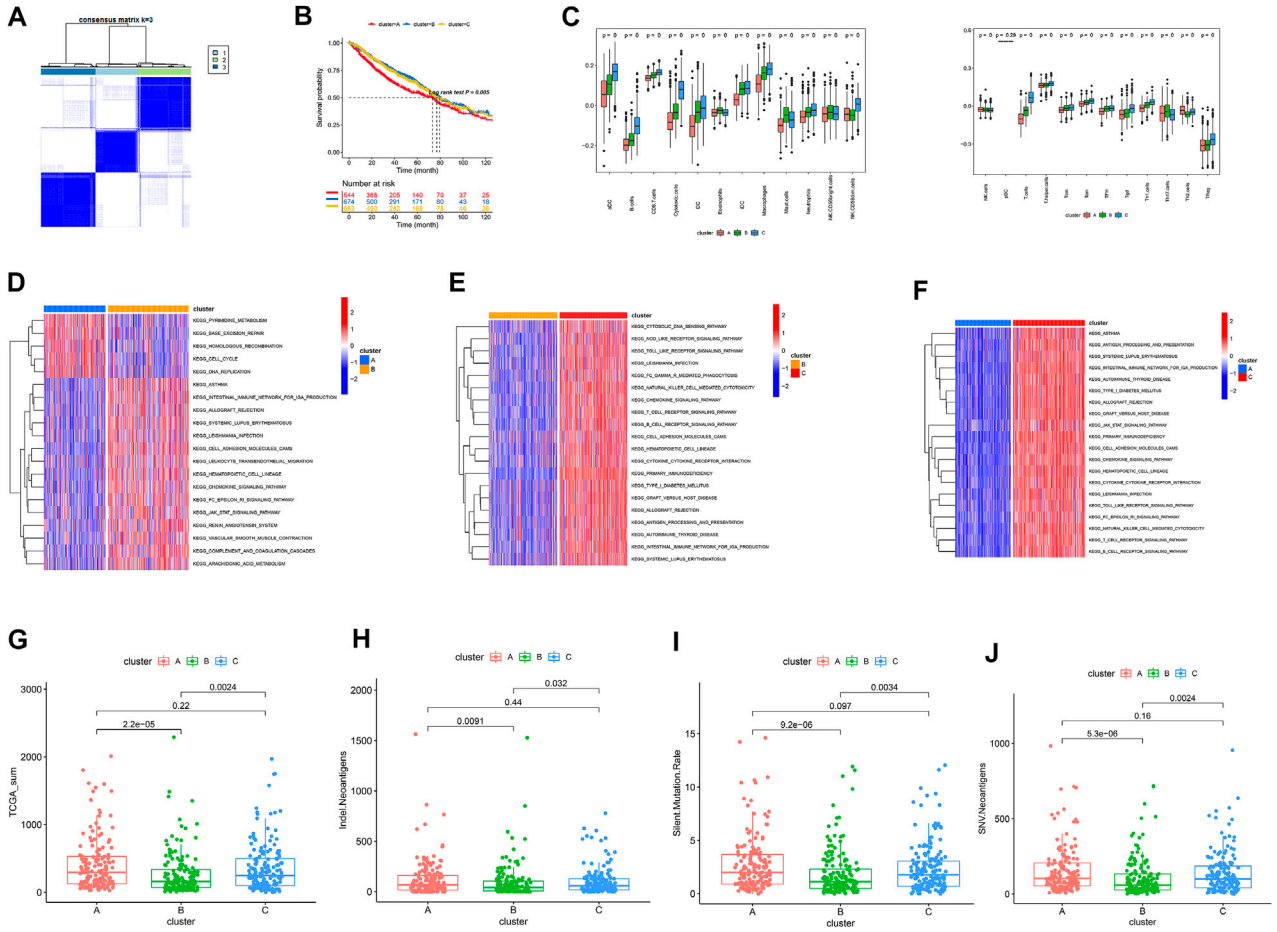


FIGURE 2 | (A) Identification of three clusters by consensus clustering. **(B)** Patient survival analysis. **(C)** Landscapes of immune cell infiltration. **(D–F)** KEGG analysis. **(G–J)** Distribution of TMB, neo-antigens (Indel, SNV), and silent mutation rate in the three clusters.

cell signature gene predictions (**Supplementary Table S1**) (Bindea et al., 2013).

Gene Set Variation Analysis and Functional Annotation

To explore the biological event differences between clusters A, B, and C, we used the "GSVA" R packages to conduct the GSVA enrichment analysis (**Figure 2A**). The "c2.cp.kegg.v7.2.symbols.gmt" was obtained from the MSig DB dataset, and the *p*-adjust < 0.5 was considered as statistically significant. The subtypes also were seen to correlate with the immune studies and prognosis by Thorsson et al. (2018), aggressive subtype by Dama et al. (2017), and luminal and basal subtypes by Zhao et al. (2019) to analyze the overlap with our study cluster C (**Figure 3A**). To confirm the events in the different subgroups, the distribution of mutant gene frequencies affected by SNVs and CNVs was investigated across various subtypes (**Figure 3B**). The frequency of mutations in each gene was significantly different across subtypes (Fisher exact test with the BH test correction, adjusted $p < 0.05$).

Consensus Clustering for Tumor-Infiltrating Immune Cells and Differential Expression Genes

For each sample, hierarchical agglomerative clustering was implemented for LUAD depending on a specific pattern. In this procedure, the "Consensus Cluster Plus" R package was used to perform "PAM" analysis, which is a Euclidean distance and Ward's linkage-based unsupervised clustering approach. To ensure clustering stability, the aforementioned process was repeated about 100 times.

Differential Expression Genes Associated With the Two Clusters

Depending on the infiltration of immune cells, patients were classified into high and low immune-cell infiltration subtypes. To determine DEGs between two clusters, the limma R package was

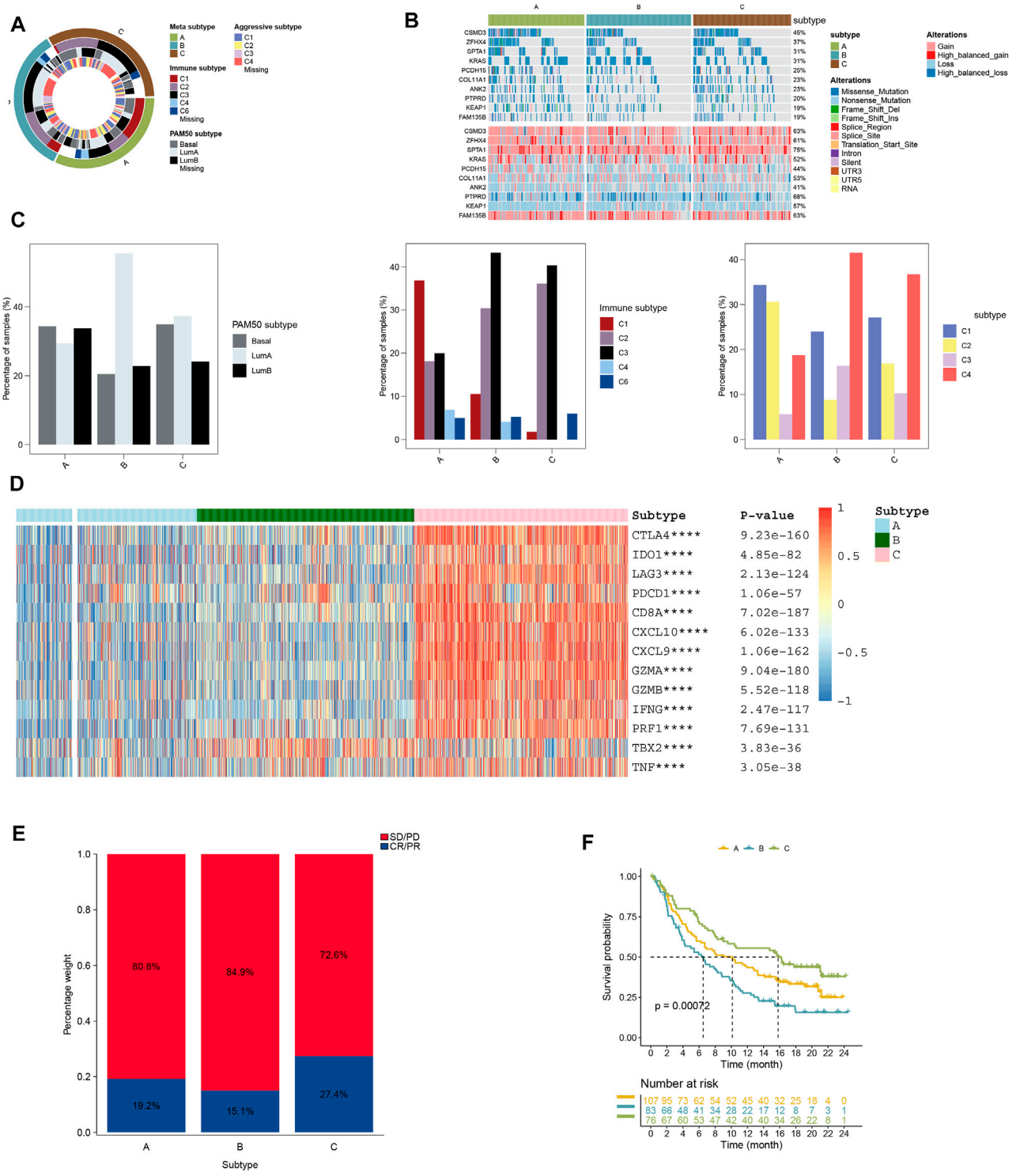


FIGURE 3 | (A) Overlay of different clusters (inner ring) with LUAD expression subtypes (outer ring). **(B)** Oncoprint distributions of somatic mutation (SNV/indel) and copy number variation (CNV) events in different clusters. **(C)** Distribution of immune, aggressive luminal, and basal subtypes in different clusters. **(D)** Distribution of immune-related gene expression in different clusters. **(E)** Rate of clinical response to anti-PD-L1 immunotherapy in different clusters, and **(F)** Kaplan–Meier curves for samples with different clusters in the IMvigor210 cohort.

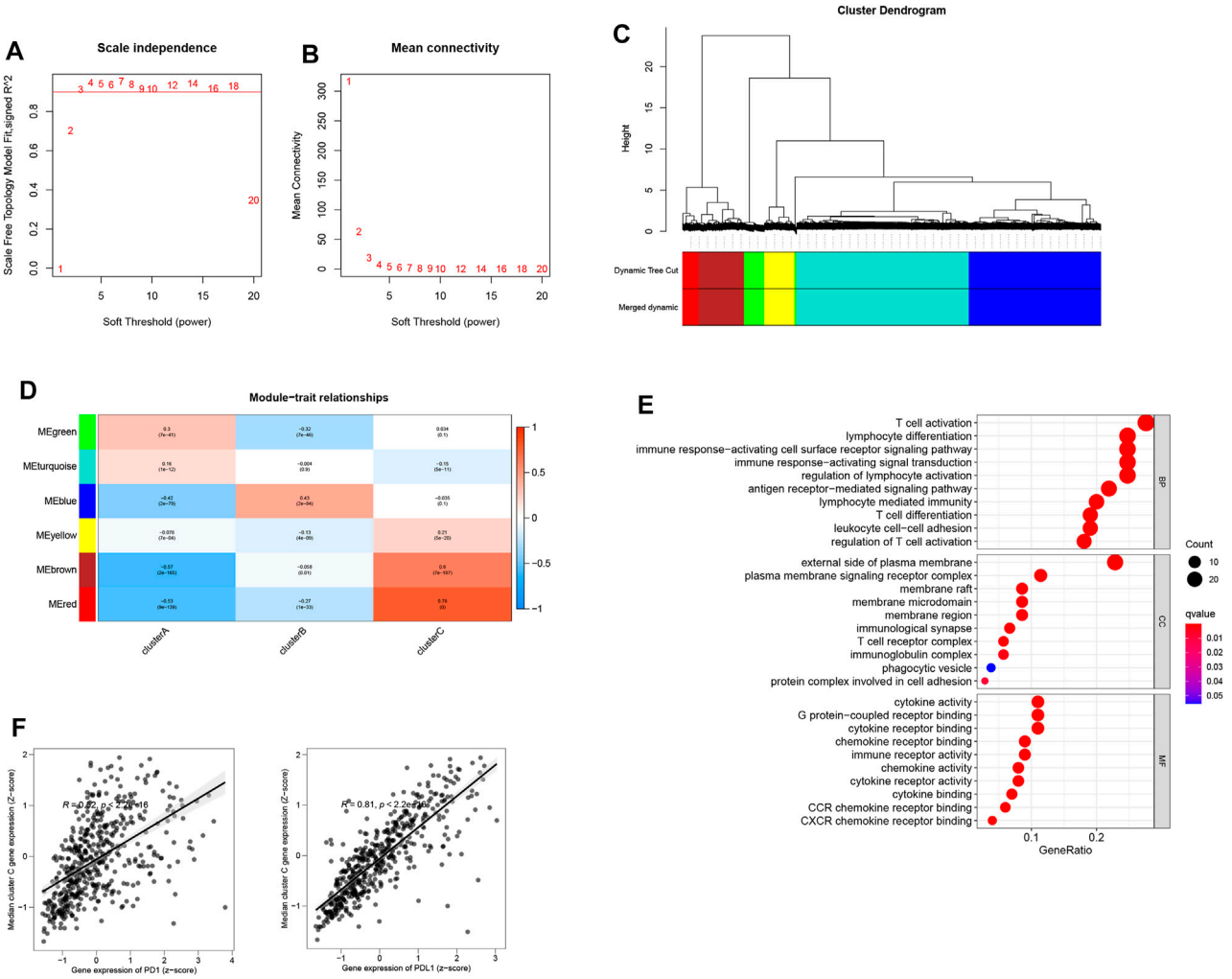


FIGURE 4 | (A) Scale-free fit index analysis of 1–20 soft threshold power β. **(B)** Mean connectivity analysis of 1–20 soft threshold power. **(C)** Genes are hierarchically clustered into various modules indicated by different colors. **(D)** Heatmap displaying correlations among module eigengenes. **(E)** GO analysis of the red module genes. **(F)** Correlation of the median expression of genes (red module) with the PD-1/PD-L1 level.

utilized, and absolute fold change was designated to >1, and significance criterion adjusted to $p < 0.05$.

Construction of Signature Gene of Lung Adenocarcinoma

For immunotherapy response assessment of lung cancer patients presenting newly defined immunophenotypes, the gene expression profiles and clinical outcome data of 348 patients from the IMvigor210 (a clinical response trial dealing with PD-L1 blockade by atezolizumab) were collected. The responses to anti-PD-L1 therapy constituted the observed endpoints, which were complete response (CR), partial response (PR), progression of disease (PD), and stability of disease (SD). Regarding the objective response rate (ORR) and disease control rate (DCR), they, respectively, involved patients with CR and PR (for ORR) and patients with CR, PR, and SD (for DCR). Based on the top 10 marked genes of metabolic subtypes, we separated the IMvigor210 cohort into three subtypes (cluster A, cluster B, and cluster C).

For candidate gene selection, we used a least absolute shrinkage and selector operation (LASSO) algorithm, whose penalty parameter was adjusted by setting a cross-validation (10-folds) approach. Meanwhile, we used another algorithm, support vector machine–recursive feature elimination (SVM-RFE), to accomplish gene selection for the CR/PR and SD/PD patients. For further narrowing on the gene among the training cohort, L1-penalized Cox analysis was eventually carried out through gene integrations from either of the aforementioned two algorithms.

Gene Expression Data With Immunotherapy

Under the Creative Commons 3.0 license (http://research-pub.gene.com/IMvigor210CoreBiologies), the IMvigor210 dataset

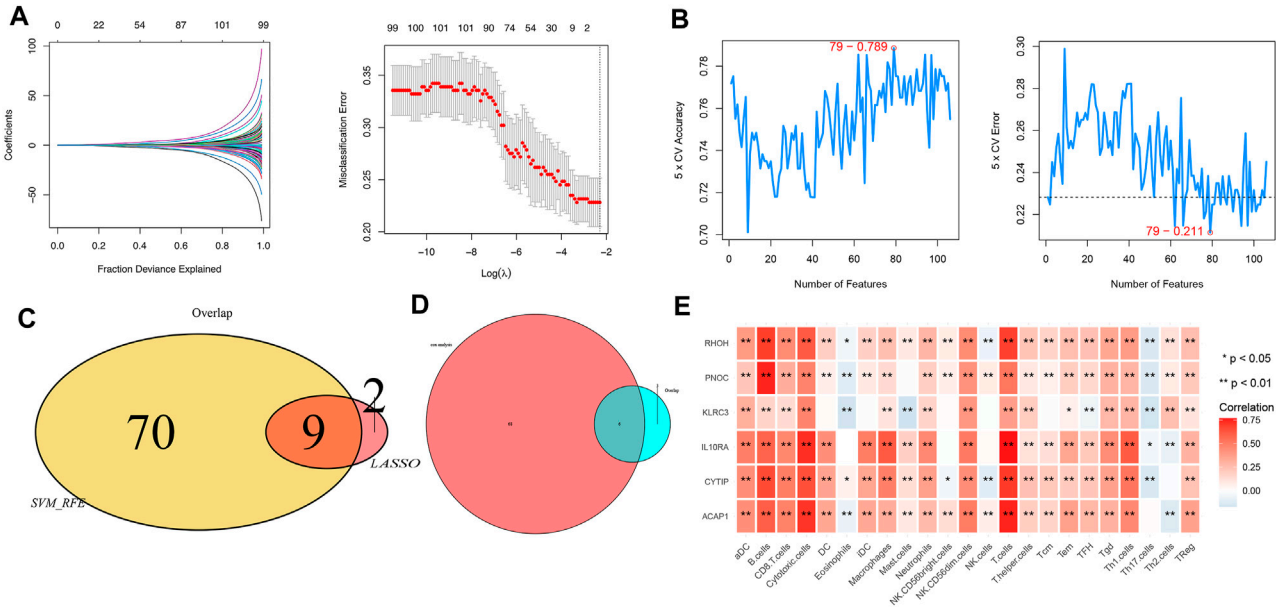


FIGURE 5 | (A) LASSO and **(B)** SVM-RFE algorithms in the detection cohort. **(C)** Overlap of incorporated genes selected from two algorithms in the detection cohort. **(D)** Intersection of characteristic genes with PD/RD and Cox analysis genes. **(E)** Correlation between immune cell infiltration and selected genes.

was obtained from accessible, well-documented software and data package. To determine the status of binary response in various clusters, 298 urothelial cancer patients and 80 immunotherapy recipients with cutaneous melanoma having complete clinical records were analyzed.

Statistical Analysis

GraphPad and R 4.0.0. were used for all statistical analyses. The Wilcoxon test was conducted for pairwise comparison analysis, and the Kruskal–Wallis test was adopted for comparison among more than two groups. The FDRs in limma and GSEA were adjusted by the Benjamini–Hochberg approach with a significance level of $p < 0.05$. The correlation of categorical clinical information with defined clusters was statistically examined by Fisher's exact test. All statistical differences were considered significant when p-value < 0.05.

RESULTS

Identification of Different Subtypes

In this study, the immune cell infiltration matrix was used to identify two clusters with different survival rates (**Figures 1A,B**), and cluster one showed high immune cell infiltration (**Figure 1C**) such as the DC, B-cells, CD8 T-cells, cytotoxic cells, DC, iDC, macrophages, mast cells, neutrophils, NK, CD-56 dim cells, T-cells, T-helper cells, Tcm, Tem, TFH, Tgd, Th1-cells, and T-Reg. Furthermore, the PD-1 and PD-L1 showed high expression in cluster 2 (**Figure 1E**). The differential expressions of 110 genes between the two clusters (**Figure 1D**) were enriched in various immune-related pathways such as T-cell activation and cytokine activity (**Figure 1F**).

Identification of Gene Subtypes and Association With Known Subtypes

With the aid of the limma package, the differential expression genes (DEGs) analyses for transcriptome evolution investigation among these clusters performed to identify the biological function of different clusters showed 110 differential gene expressions. For the elimination of redundant genes, the Cox analysis was performed to collect significantly correlated prognosis genes.

The survival records used to assess the prognostic implication of the clusters (**Figure 2B**) showed clusters B and C to have a better survival rate than cluster A ($p = 0.005$). Meanwhile, cluster C showed greater immune cell infiltrations such as aDC, B cells, CD8 T cells, cytotoxic cells, DC, eosinophils, iDC, macrophages, mast cells, neutrophils, NK CD56-dim cells, T cells, Tcm, Tem, TFH, Tgd, and Th1 cells (**Figure 2C**). Furthermore, enrichment of cluster C showed multiple immune-associated pathways, including those for signaling T- and B-cell receptors (**Figures 2D–F**). Additionally, cluster C showed high tumor mutational burden, neo-antigen (indel and SNV), and PD-1 expression than other subtypes (**Figures 2H–K**).

Our findings showed clusters B and C to have high-immune subtypes C3 and C2, high frequency of LumB subtype, and aggressive subtype C4 (**Figure 3C**). Of these subtypes, the immune subtype C3, LumA subtype, and aggressive subtype C4 found in our study (**Figure 2B**) were consistent with the findings of others for a better prognosis. In

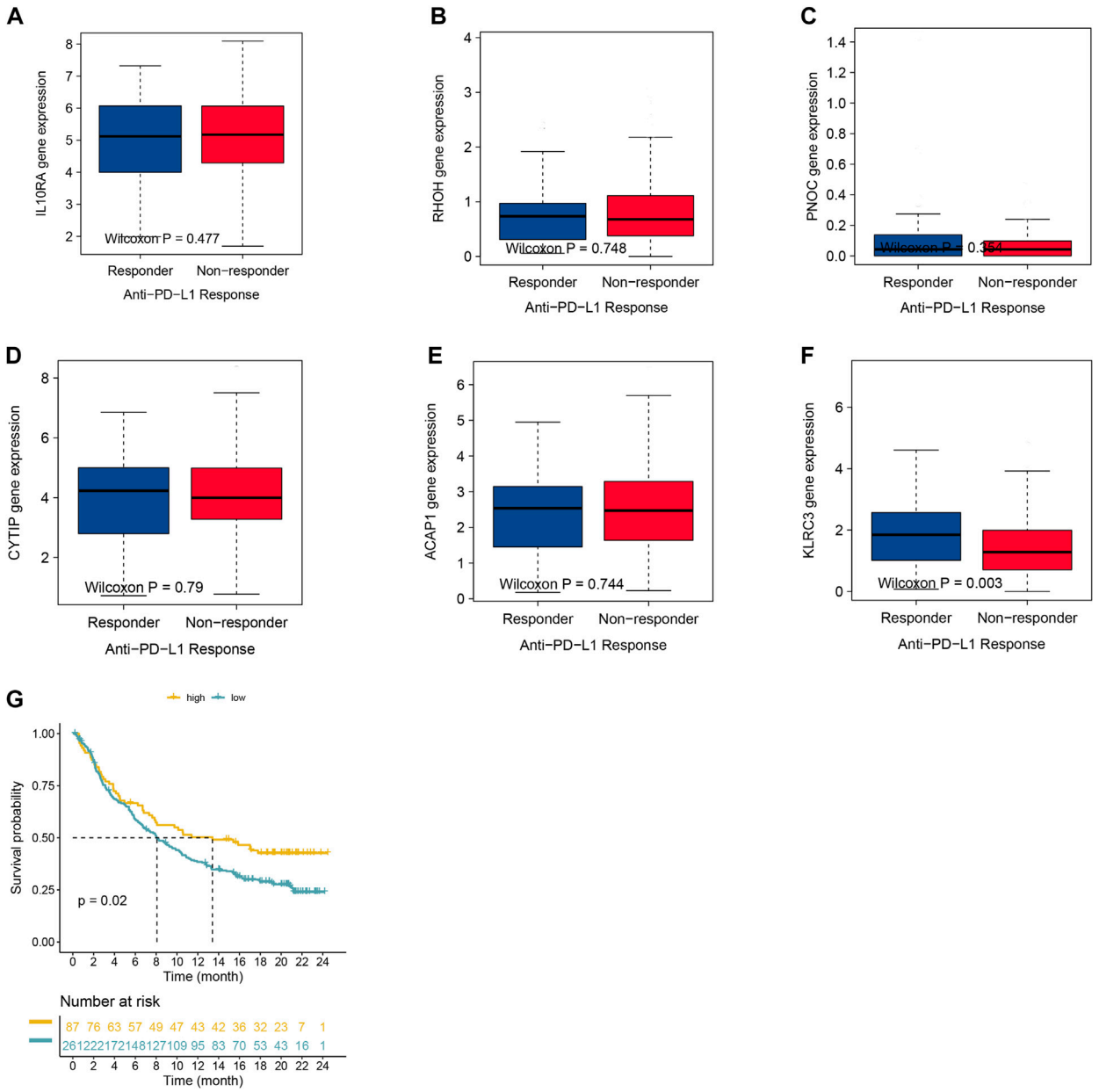


FIGURE 6 | (A–F) Expression of selected genes with varying anti-PD-1 responses. **(G)** Kaplan–Meier graphs of KLRC3 expression in the IMvigor210 cohort.

addition to this, a majority of immune-checkpoint-relevant signatures (CD274, CTLA4, HAVCR2, IDO1, LAG3, and PDCD1) and immune-activity-relevant signatures (CD8A, CXCL9, CXCL10, GZMA, GZMB, IFNG, PRF1, TBX2, and TNF) exhibited significant high expression in cluster C (**Figure 3D**).

Immunotherapy Response

Our findings also showed that the cluster A subtype had high SD/PD with 72.6% and low CR/PR with 27.4%; however, the cluster A subtype had higher CR/PR with 19.2% than the cluster C subtype (**Figure 3F**). In the IMvigor210 cohort, the cluster C subtype had a higher survival rate than other subtypes from anti-PD-1 treatment (**Figure 3E**).

WGCNA Analysis

Aided by the R package "WGCNA", the co-expression network was created from the expression levels of 11,518 genes (Langfelder and Horvath, 2008). Clustering of 1,881 samples was performed by calculating the mean linkage and Pearson's correlation coefficient. The soft threshold power was set at $\beta = 3$ and scale-free $R2 = 0.96$ to ensure that the scale-free network was constructed (**Figures 4A,B**). A dynamic mixing and cutting technique was employed to establish the hierarchical clustering

tree, each of whose leaves was used to refer to a gene. Meanwhile, a tree branch constituted gene assemblies resembling expression data that were used to refer to a gene module. In this study, a total of six modules were produced (**Figures 4C,D**), of which the red module showed a high correlation with cluster C from 111 genes enriched by T-cell activation, cytokine activity, and regulation of T-cell activation (**Figure 4E**). The median expressions of red module genes showed a high correlation with PD-1/PD-L1 expression (**Figure 4F**).

Identification of Predictive Signature

Furthermore, the most significant genes were selected via two algorithms from the CR/PR and SD/PD patients, and the LASSO algorithms were used to identify the prognosis gene. A total of 106 gene candidates were identified after the integration of LASSO- and SVM-RFE-selected genes, of which six genes were selected by both algorithms (**Figures 5A–C**). The correlations were seen between the overlapping genes, including the PNOC, RHOH, ACAP1, CYTIP, IL10RA, and KLRC3 along with the immune cell infiltrations (**Figure 5D**). The expression of KLRC3 showed significant response/nonresponse of PD-L1 (**Figures 6A–F**.) and better survival (**Figure 6G**).

DISCUSSION

With the development of immunotherapy, PD-1/CTLA4 inhibitors have been widely used in clinical applications and improved prognosis of various cancers. However, the effective response rate of PD-1 inhibitors in cancer patients, including lung cancer, is low, which is thought to be mainly affected by tumor mutational burden (Goodman et al. (2017), microsatellite instability (MSI), efficient DNA mismatch repair (DMMR) (Xiao and Freeman, 2015), and other factors. In order to improve the response rate of immunotherapy, researchers are constantly working on the study of marker genes and potential biological therapeutic targets related to immunotherapy response. It has been reported that IFN-γ-related gene expression contributes to immunotherapy prediction, such as CCR5, CXCL9, IFNG, STAT1, and PRF1 (Ayers et al., 2017), and these genes are related to tumor antigen presentation, T cytotoxic activity, and immune cell infiltration. These IFN-γ-related genetic signatures are necessary for immunotherapy response (Sharma and Allison, 2015). However, the immune response marker genes for lung adenocarcinoma need to be further studied.

Here, by analyzing the gene expression data of a large number of lung adenocarcinoma patients, we searched for genes associated with better prognosis and immune cell infiltration, hoping to be further used in immunotherapy. Based on differential gene expression, we identified three subtypes showing different survival and immunotherapy effects. Among them, immune-related genes include CD274, CTLA4, HAVCR2, IDO1, LAG3, and PDCD1 as immune checkpoint-related markers, and CD8A, CXCL10, CXCL9, GZMA, GZMB, IFNG, PRF1, TBX2, and TNF (immune activity-related signatures) expressions in cluster C were higher than other subtypes, indicating that cluster C has a high level of immunoreactivity, that is, subtype C has a high potential for immune activation, which can elicit an effective immune response.

Ultimately, six genes including PNOC, RHOH, ACAP1, CYTIP, IL10RA, and KLRC3 were selected by both algorithms of LASSO and SVM-RFE analysis. Prepronociceptin (PNOC) is a preprotein of a series of intracellular products involved in pain and inflammation signaling (Rodriguez-Romaguera et al., 2020). In cancer research, PNOC is highly expressed in glioma cells, epithelial ovarian cancer, and other tumors and has been reported as a prognostic biomarker (Chan et al., 2012; Jin et al., 2018). RHOH, a member of the RhoE/Rnd3 subfamily of GTPases, is highly expressed in B cells, suggesting that it is involved in the development of B-cell malignant leukemia (Dallery-Prudhomme et al., 1997) and is closely related to the immune microenvironment in chronic leukemia (Troeger et al., 2012). ACAP1 is involved in cell membrane transport and cell migration, is an Arf6 GAP, that is, important for immune cell migration and infiltration, and is associated with tumor immune infiltration and prognosis in breast cancer (Zhang et al., 2020). Cytohesin-interacting protein (Cytip) is associated with dendritic cell (DC) maturation and T cell activation and may function in the tumor immune microenvironment (Heib et al., 2012). IL10RA encodes a receptor molecule for the inflammatory factor IL10 and is associated with IL10 expression and STAT3 phosphorylation in colorectal cancer. The expression of IL10RA is also considered to be associated with the clinical stage of colorectal cancer (Zadka et al., 2018).

Most importantly, patients with high expression of KLRC3 had significantly higher response rates to immunotherapy and better prognosis, and KLRC3 can be regarded as the core gene of these six genes. KLRC3 is a natural killer cell receptor gene, and previous studies have shown that KLRC3 affects the stemness and proliferative potential of glioma cells and is involved in glioblastoma tumorigenesis and progression (Cheray et al., 2017). This gene is also upregulated in patients with high expression of PD1 and PDL1 (Hu et al., 2020; Lamberti et al., 2020), suggesting that it may be involved in the regulation of PD1 and PD-L1 expression.

We used WGCNA analysis to confirm cluster C-associated red modules with 111 genes. LASSO and SVM-RFE algorithms confirmed that the key gene KLRC3 was significantly associated with CR/PR status with better prognosis. Our study also has some limitations, such as the lack of experimental and clinical validation data, although the omics data analysis was performed; so, our study results will stimulate further verification or falsification in the scientific community.

Based on our findings in this study, specific changes in the landscape of immune cell infiltration and mutations and transcriptome profiles may have a dramatic impact on improving immunotherapy. KLRC3 may be a core gene suitable for prognosis in lung adenocarcinoma and is associated with SD/PD and CR/PR patients, which will improve reference for future immunotherapy-related studies.

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

KM conceived and designed the study, obtained funding, and drafted the manuscript. YZ and KM acquired the data and drafted the manuscript. YZ critically revised the manuscript. PF and ZL performed statistical analysis and technical support. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.810193/full#supplementary-material

REFERENCES

Ayers, M., Lunceford, J., Nebozhyn, M., Murphy, E., Loboda, A., Kaufman, D. R., et al. (2017). IFN-γ-related mRNA Profile Predicts Clinical Response to PD-1 Blockade. *J. Clin. Invest.* 127 (8), 2930–2940. doi:10.1172/jci91190

Barber, D. L., Wherry, E. J., Masopust, D., Zhu, B., Allison, J. P., Sharpe, A. H., et al. (2006). Restoring Function in Exhausted CD8 T Cells during Chronic Viral Infection. *Nature* 439, 682–687. doi:10.1038/nature04444

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., et al. (2013). Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the Immune Landscape in Human Cancer. *Immunity* 39, 782–795. doi:10.1016/j.immuni.2013.10.003

Brahmer, J. R., Tykodi, S. S., Chow, L. Q. M., Hwu, W.-J., Topalian, S. L., Hwu, P., et al. (2012). Safety and Activity of Anti-PD-L1 Antibody in Patients with Advanced Cancer. *N. Engl. J. Med.* 366, 2455–2465. doi:10.1056/nejmoa1200694

Chan, M. H., Kleinschmidt-Demasters, B. K., Donson, A. M., Birks, D. K., Foreman, N. K., and Rush, S. Z. (2012). Pediatric Brainstem Gangliogliomas Show Overexpression of Neuropeptide Prepronociceptin (PNOC) by Microarray and Immunohistochemistry. *Pediatr. Blood Cancer* 59 (7), 1173–1179. doi:10.1002/pbc.24232

Cheray, M., Bessette, B., Lacroix, A., Mélin, C., Jawhari, S., Pinet, S., et al. (2017). KLRC 3, a Natural Killer Receptor Gene, Is a Key Factor Involved in Glioblastoma Tumourigenesis and Aggressiveness. *J. Cel. Mol. Med.* 21 (2), 244–253. doi:10.1111/jcmm.12960

Dallery-Prudhomme, E., Roumier, C., Denis, C., Preudhomme, C., Kerckaert, J.-P., and Galiegue-Zouitina, S. (1997). Genomic Structure and Assignment of theRhoH/TTFSmall GTPase Gene (ARHH) to 4p13 Byin SituHybridization. *Genomics* 43 (1), 89–94. doi:10.1006/geno.1997.4788

Dama, E., Melocchi, V., Dezi, F., Pirroni, S., Carletti, R. M., Brambilla, D., et al. (2017). An Aggressive Subtype of Stage I Lung Adenocarcinoma with Molecular and Prognostic Characteristics Typical of Advanced Lung Cancers. *Clin. Cancer Res.* 23 (1), 62–72. doi:10.1158/1078-0432.ccr-15-3005

Goodman, A. M., Kato, S., Bazhenova, L., Patel, S. P., Frampton, G. M., Miller, V., et al. (2017). Tumor Mutational burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. *Mol. Cancer Ther.* 16 (11), 2598–2608. doi:10.1158/1535-7163.mct-17-0386

Heib, V., Sparber, F., Tripp, C. H., Ortner, D., Stoitzner, P., and Heufler, C. (2012). Cytip Regulates Dendritic-Cell Function in Contact Hypersensitivity. *Eur. J. Immunol.* 42 (3), 589–597. doi:10.1002/eji.201041286

Herbst, R. S., Baas, P., Kim, D. W., Felip, E., Perez-Gracia, J. L., Han, J. Y., et al. (2015). Pembrolizumab versus Docetaxel for Previously Treated, PD-L1-Positive, Advanced Non-small-cell Lung Cancer (KEYNOTE-010): a Randomised Controlled Trial. *Lancet* 387 (10027), 1540–1550. doi:10.1016/S0140-6736(15)01281-7

Hu, Z., Bi, G., Sui, Q., Bian, Y., Du, Y., Liang, J., et al. (2020). Analyses of Multi-Omics Differences between Patients with High and Low PD1/PDL1 Expression in Lung Squamous Cell Carcinoma. *Int. Immunopharmacology* 88, 106910. doi:10.1016/j.intimp.2020.106910

Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., and Forman, D. (2011). Global Cancer Statistics. *CA: A Cancer J. Clinicians* 61 (2), 69–90. doi:10.3322/caac.20107

Jin, S., Zhu, W., and Li, J. (2018). Identification of Key Genes Related to High-Risk Gastrointestinal Stromal Tumors Using Bioinformatics Analysis. *J. Can. Res. Ther.* 14 (Suppl. ment), 243–s7. doi:10.4103/0973-1482.207068

Johnson, W. E., Li, C., and Rabinovic, A. (2007). Adjusting Batch Effects in Microarray Expression Data Using Empirical Bayes Methods. *Biostatistics* 8, 118–127. doi:10.1093/biostatistics/kxj037

Lamberti, G., Andrini, E., Sisi, M., Rizzo, A., Parisi, C., Di Federico, A., et al. (2020). Beyond EGFR, ALK and ROS1: Current Evidence and Future Perspectives on Newly Targetable Oncogenic Drivers in Lung Adenocarcinoma. *Crit. Rev. Oncology/Hematology* 156, 103119. doi:10.1016/j.critrevonc.2020.103119

Langfelder, P., and Horvath, S. (2008). WGCNA: an R Package for Weighted Correlation Network Analysis. *BMC Bioinformatics* 9, 559. doi:10.1186/1471–2105–9–55910.1186/1471-2105-9-559

Ohe, Y., Ohashi, Y., Kubota, K., Tamura, T., Nakagawa, K., Negoro, S., et al. (2007). Randomized Phase III Study of Cisplatin Plus Irinotecan versus Carboplatin Plus Paclitaxel, Cisplatin Plus Gemcitabine, and Cisplatin Plus Vinorelbine for Advanced Non-small-cell Lung Cancer: Four-Arm Cooperative Study in Japan. *Ann. Oncol.* 18, 317–323. doi:10.1093/annonc/mdl377

Robert, C., Long, G. V., Brady, B., Dutriaux, C., Maio, M., Mortier, L., et al. (2015). Nivolumab in Previously Untreated Melanoma withoutBRAFMutation. *N. Engl. J. Med.* 372, 320–330. doi:10.1056/nejmoa1412082

Rodriguez-Romaguera, J., Ung, R. L., Nomura, H., Otis, J. M., Basiri, M. L., Namboodiri, V. M. K., et al. (2020). Prepronociceptin-Expressing Neurons in the Extended Amygdala Encode and Promote Rapid Arousal Responses to Motivationally Salient Stimuli. *Cel Rep.* 33 (6), 108362. doi:10.1016/j.celrep.2020.108362

Sharma, P., and Allison, J. P. (2015). The Future of Immune Checkpoint Therapy. *Science* 348 (6230), 56–61. doi:10.1126/science.aaa8172

Sharpe, A. H., Wherry, E. J., Ahmed, R., and Freeman, G. J. (2007). The Function of Programmed Cell Death 1 and its Ligands in Regulating Autoimmunity and Infection. *Nat. Immunol.* 8, 239–245. doi:10.1038/ni1443

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–e14. doi:10.1016/j.immuni.2018.03.023

Troeger, A., Johnson, A. J., Wood, J., Blum, W. G., Andritsos, L. A., Byrd, J. C., et al. (2012). RhoH Is Critical for Cell-Microenvironment Interactions in Chronic Lymphocytic Leukemia in Mice and Humans. *Blood* 119 (20), 4708–4718. doi:10.1182/blood-2011-12-395939

Xiao, Y., and Freeman, G. J. (2015). The Microsatellite Instable Subset of Colorectal Cancer Is a Particularly Good Candidate for Checkpoint Blockade Immunotherapy. *Cancer Discov.* 5 (1), 16–18. doi:10.1158/2159-8290.cd-14-1397

Zadka, Ł., Kulus, M. J., Kurnol, K., Piotrowska, A., Glatzel-Plucińska, N., Jurek, T., et al. (2018). The Expression of IL10RA in Colorectal Cancer and its Correlation with the Proliferation index and the Clinical Stage of the Disease. *Cytokine* 110, 116–125. doi:10.1016/j.cyto.2018.04.030

Zhang, J., Zhang, Q., Zhang, J., and Wang, Q. (2020). Expression of ACAP1 Is Associated with Tumor Immune Infiltration and Clinical Outcome of Ovarian Cancer. *DNA Cel Biol.* 39 (9), 1545–1557. doi:10.1089/dna.2020.5596

Zhao, S. G., Chen, W. S., Das, R., Chang, S. L., Tomlins, S. A., Chou, J., et al. (2019). Clinical and Genomic Implications of Luminal and Basal Subtypes across Carcinomas. *Clin. Cancer Res.* 25 (8), 2450–2457. doi:10.1158/1078-0432.ccr-18-3121

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: 11 May 2022
doi: 10.3389/fgene.2022.825685

Upregulation of Ferroptosis-Related Fanconi Anemia Group D2 is a Poor Prognostic Factor and an Indicator of Tumor Immune Cell Infiltration in Lung Adenocarcinoma

Jingtao Zhang[1], Dongli Wang[2], Xiubao Chen[3], Lingyun Ji[4], Minmin Yu[5], Minghao Guo[3], Dexin Zhang[4], Weida Chen[3] and Fei Xu[3]**

[1]College of Traditional Chinese Medicine, Shandong University of Traditional Chinese Medicine, Jinan, China, [2]Digestive Department, Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan, China, [3]Department of Geriatric Medicine, Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan, China, [4]First Clinical Medical College, Shandong University of Traditional Chinese Medicine, Jinan, China, [5]Department of Pathology, Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan, China

OPEN ACCESS

Edited by:
Mehdi Pirooznia,
Johnson & Johnson, United States

Reviewed by:
Tienan Feng,
Shanghai Jiao Tong University, China
Song Xu,
Tianjin Medical University General Hospital, China
Li Jiancheng,
Fujian Provincial Cancer Hospital, China

***Correspondence:**
Fei Xu
Fei_Xu1012@126.com
Weida Chen
chenweida_2000@126.com

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: 14 February 2022
Accepted: 22 April 2022
Published: 11 May 2022

Citation:
Zhang J, Wang D, Chen X, Ji L, Yu M, Guo M, Zhang D, Chen W and Xu F (2022) Upregulation of Ferroptosis-Related Fanconi Anemia Group D2 is a Poor Prognostic Factor and an Indicator of Tumor Immune Cell Infiltration in Lung Adenocarcinoma. Front. Genet. 13:825685. doi: 10.3389/fgene.2022.825685

Fanconi anemia (FA) group D2 (*FANCD2*) is a ferroptosis-related gene crucial for DNA damage repair and negative ferroptosis regulation. Our study aimed to evaluate its prognostic value as well as its association with ferroptosis and immune infiltration in lung adenocarcinoma (LUAD). Transcriptome sequencing data, clinical information, and immunohistochemistry data were collected from the TCGA, GEO, and HPA databases, respectively, for three independent cohorts. Univariate and multivariate analyses were used to assess the correlations between FANCD2 expression and overall survival or clinicopathological parameters. cBioPortal was utilized to investigate the *FANCD2* alteration status. Gene and protein networks based on FANCD2 interactions were generated using GeneMANIA and STRING, respectively. Based on the CancerSEA database, the function of FANCD2 was explored at the single-cell level. The relationships between *FANCD2* expression levels and tumor-infiltrating immune cells and their equivalent gene signatures were analyzed using TIMER, GEPIA, TISIDB, and ssGSEA databases. CIBERSORT was used to analyze the relevance of the infiltration of 24 types of immune cells. The results revealed that FANCD2 expression was significantly upregulated in LUAD and lung squamous cell carcinoma (LUSC) tissues than that in normal tissues. Further, the overexpression of FANCD2 was closely associated with poor survival for Patients with LUAD but not for patients with LUSC. FANCD2 expression levels were related to tumor-infiltrating immune cells and their matching gene signatures, including $CD8^+$ T cells, natural killer (NK) cells, dendritic cells (DC), and Th2 cells in cases of LUAD. Therefore, FANCD2 was identified as a crucial molecule underlying the synergistic effects of ferroptosis and immunotherapy for Patients with LUAD.

Keywords: FANCD2, lung adenocarcinoma, ferroptosis, tumor-infiltrating immune cells, biomarker, prognosis

INTRODUCTION

In the past decades, lung cancer has remained the major contributing factor to cancer-related deaths worldwide. According to available data, 2.2 million new cases of lung cancer are diagnosed each year, with 1.8 million people dying from the disease annually (Sung et al., 2021). Non-small cell lung cancer (NSCLC), which includes lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC), contributes to approximately 85% of all lung cancer cases. However, the majority of patients are in an advanced, unresectable stage of the disease at the time of diagnosis (Brahmer et al., 2018), which is associated with a low overall median 5-year survival rate (Sung et al., 2021). The incidence and mortality of LUAD are increasing; it has now surpassed squamous cell carcinoma to become the most common histological subtype of NSCLC (Jemal et al., 2011; Siegel et al., 2017). Despite the active treatment measures available for LUAD, it has the highest mortality rate among all cancers; this might be associated with its tendency to metastasize at an early stage (Deng et al., 2019; Zhang D. et al., 2019; Li et al., 2021). Thus, it is of critical importance to develop novel and more effective therapeutic strategies for LUAD. For this, it is vital to more intensively probe the molecular pathology of the disease.

Ferroptosis is an iron-dependent form of regulated cell death induced by iron-dependent lipid peroxidation owing to metabolic dysfunction; it is distinct from apoptosis, cell necrosis, and autophagy (Stockwell et al., 2017; Gao and Jiang, 2018). The ferroptosis pathway can restrain tumor growth and induce cancer cell death. Its induction has thus received widespread attention as a potential novel anti-tumor treatment strategy (Hassannia et al., 2019; Liang et al., 2019; Bebber et al., 2020). Inducing ferroptosis has been reported to suppress LUAD by regulating lipid peroxidation to promote tumor cell death (Ma et al., 2021; Wang et al., 2021; Zhang Y. et al., 2021). In addition, ferroptosis is also connected to cell immunity and may have applications in cancer immunotherapy (Wang et al., 2019).

With the rapid advances in high-throughput sequencing technologies and transcriptome sequencing [RNA sequencing (RNA-seq)], an increasing number of key driver oncogenes are being discovered. However, it is necessary to identify additional key driver genes, particularly those affecting the tumor immune microenvironment (TIME) in LUAD. The Fanconi anemia (FA) pathway plays an important role in DNA damage repair by blocking DNA replication, for instance, *via* interstrand cross-links. FANCD2, a member of the FA family of proteins, forms a FANCD2-FANCI heterodimer with FANCI and participates in DNA damage repair via the FA pathway (Nalepa and Clapp, 2018). Some studies have demonstrated that FANCD2 depletion enhances interstrand crosslink (ICL) agent-induced DNA damage sensitivity and promotes apoptosis in lung cancer cells by inhibiting the FA pathway (Wang et al., 2015; Fan et al., 2021). In addition, FANCD2 negatively regulates ferroptosis by regulating iron metabolism-related genes and/or protein expression and lipid peroxidation (Song et al., 2016). However, the molecular mechanisms of FANCD2 governing the regulation of the immune response in LUAD are still unclear.

This study aimed to explore the association between *FANCD2* expression and clinical information and prognosis in LUAD. The results show that FANCD2 expression regulates the level of tumor-infiltrating immune cells through multiple pathways, which contributes to the formation of the immunosuppressive microenvironment. Therefore, this study anticipates promising potential therapeutic strategies for LUAD based on FANCD2.

MATERIALS AND METHODS

Identification of Ferroptosis- and NSCLC-Related Targets

Ferroptosis-related targets were identified from FerrDb (http://www.zhounan.org/ferrdb), which has data on 253 regulators, 111 markers, and 95 ferroptosis-associated diseases. NSCLC-related targets were identified from the GEO database, which contains many bioinformatics datasets from the National Center of Biotechnology Information (https://www.ncbi.nlm.nih.gov/geo/). Three gene expression datasets (GSE75037, GSE19188, and GSE116959) derived from human NSCLC tissues and adjacent normal tissues were included. Genes with an adjusted $p < 0.01$ and $|\log_2(\text{fold-change})|>2$ were defined as differentially expressed genes (DEGs), which were considered to play essential roles in NSCLC progression and defined as key targets for inducing ferroptosis to treat NSCLC.

Data Acquisition and Analysis

TCGA (https://portal.gdc.cancer.gov/) is a large-scale and open-access cancer genomic database. All transcriptome RNA-seq data (n = 1,145) and equivalent clinical data related to NSCLC were downloaded. Based on pathological characteristics, all patients were divided into LUAD (n = 516) and LUSC (n = 493) groups. Subgroup analysis was performed to investigate the effect of FANCD2 on the pathological stage and outcome. Three independent cohorts including tumor tissues and control samples of NSCLC patients from the GEO databases GSE75037 (n = 166), GSE19188 (n = 156), and GSE116959 (n = 68) were used to confirm the findings from TCGA datasets. HPA (https://www.proteinatlas.org/) is a comprehensive resource database of the human proteome, and the protein levels of FANCD2 in normal lung tissues and LUAD and LUSC tissues compared according to immunohistochemistry (IHC) results.

Survival and Statistical Analyses

Based on the median expression of *FANCD2*, patients with LUAD and LUSC were divided into a high and a low group. To identify whether the FANCD2 expression level influenced LUAD and LUSC patient clinical survival, Kaplan–Meier (KM) survival curves were generated to estimate its prognostic significance.

Univariate and Multivariate Logistic Regression Analyses

To further identify the prognostic value of FANCD2 in LUAD, Cox analyses were adopted to evaluate the relationships between

various clinical characteristics and prognosis. Univariate Cox analysis was conducted for every variable comparing the expression level of FANCD2 and patient overall survival (OS) in each cohort to confirm their association with LUAD prognosis. Subsequently, multivariate Cox analysis, including all variables, was used to evaluate whether FANCD2 was an independent prognostic factor for LUAD patient outcome.

Genetic Alteration and Interaction Network Analyses

Based on cBioPortal (http://cbioportal.org), an open-access multidimensional cancer genomics resource, FANCD2 alterations were analyzed in LUAD samples collected from TCGA. GeneMANIA (http://genemania.org/) and STRING (https://string-preview.org/) are source websites used to construct gene–gene and protein–protein interaction networks, respectively. Both were applied to investigate FANCD2-related genes and proteins.

Single-Cell Analysis

The CancerSEA (http://biocc.hrbmu.edu.cn/CancerSEA/home.jsp) database provides a cancer single-cell functional state atlas. This study used CancerSEA to explore the function of FANCD2-regulated genes, as well as the correlation between FANCD2 expression levels and these functions.

TISIDB Database Analysis

The TISIDB (http://cis.hku.hk/TISIDB/) web portal was used to probe tumor-immune interactions, which was applied to evaluate correlations between FANCD2 and immune-suppressive genes and immune-activating genes in this study.

TIMER Database Analysis

TIMER (https://cistrome.shinyapps.io/timer/) is a web server for the systematic analysis of six tumor-infiltrating immune subsets across diverse cancer types. Here, the correlation between FANCD2 expression and the infiltration of B cells, $CD8^+$ T cells, $CD4^+$ T cells, dendritic cells (DCs), macrophages, and neutrophils was assessed in Patients with LUAD. Additionally, the association between FANCD2 expression and tumor purity was tested, according to the "Correlation" module of TIMER and the tumor purity-corrected partial.

ssGSEA and CIBERSORT Analysis

In total, 24 types of FANCD2-related immune cells were acquired from ssGSEA using the GSVA package in R (4.0.3). Pearson correction analysis was performed to further assess the relevance and enrichment scores between FANCD2 and the different immune cells. CIBERSORT (http://cibersort.stanford.edu/), a deconvolution algorithm based on gene expression, was used to reveal the relevance of 24 types of immune cells in LUAD.

GEPIA Database Analysis

GEPIA (http://gepia.cancer-pku.cn/index.html) covers thousands of tumors and normal samples from TCGA and the Genotype-Tissue Expression Project (GTEx, http://www.gtexportal.org/home/index.html). It focuses on the analysis of RNA-seq data. The association between FANCD2 levels and multiple markers for various immune cells was evaluated according to the GEPIA database.

Statistical Analysis

Student's *t*-tests and Wilcoxon tests were performed for assessing differences between two groups, and a Kruskal-Wallis test was used for more than two groups. To evaluate patient survival, KM curves, as well as univariate and multivariate logistic regression analyses, were conducted. Spearmen or Pearson correlations were applied to calculate the relationship between FANCD2 and immune infiltration. Data analyses were based on R (v4.0.3), and $p < 0.05$ was considered to indicate statistical significance.

RESULTS

Hub Genes Associated With Ferroptosis and NSCLC

In total, 173 genes in the FerrDb and 2,414, 3,322, and 1,830 DEGs in GSE19188, GSE75037, and GSE116959, respectively, were identified. As shown in **Figure 1**, 17 candidate genes comprised the intersection of the four datasets. Of these, the top five targets were nicotinamide adenine dinucleotide phosphate (NAD[P]H) dehydrogenase (quinone 1) (*NQO1*), heme oxygenase 1 (*HMOX1*), *FANCD2*, *helicase*, lymphoid specific (*HELLS*), and cluster of differentiation 44 (*CD44*). Notably, in addition to participating in ferroptosis, FANCD2 is involved in repairing DNA lesions and ensuring accurate DNA replication (Wang et al., 2015; Yang et al., 2016; Li et al., 2020). Owing to numerous pathways that FANCD2 could regulate in the occurrence and development of cancer, it deserves deeper investigation. Therefore, *FANCD2* was chosen for further analysis.

Patient Characteristics

The RNA-seq expression data and clinical prognostic information of 535 LUAD patients and 493 LUSC patients with were obtained from the TCGA database. The clinical information of patients with LUAD or LUSC with high or low FANCD2 expression levels is summarized in **Table 1**, including age, sex, smoking status, pathologic stage (T, N, or M), histologic grade, OS, disease-specific survival (DSS), and the progression-free interval (PFI).

FANCD2 Expression Is Higher in Tumor Samples Than in Normal Tissues

According to the TCGA and GTEx databases, *FANCD2* exhibited higher expression in 19 types of tumor tissues, including LUAD and LUSC, than in adjacent normal samples in the TCGA database (**Figure 2A**). Correlation analysis showed that *FANCD2* mRNA expression was significantly higher in older patients (age>65 years, $p < 0.05$), in males ($p < 0.01$), and in smoking patients ($p < 0.05$), as well as in patients with higher M stage ($p < 0.05$,

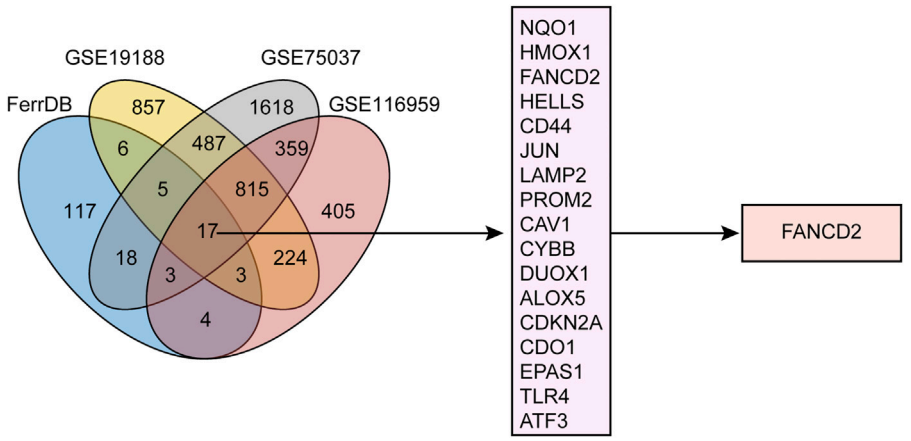


FIGURE 1 | Workflow of target screening. Venn diagram showing differentially expressed genes (DEGs) in NSCLC from three GEO cohorts alongside ferroptosis-related genes.

TABLE 1 | Clinical characteristics of lung adenocarcinoma (LUAD) and lung squamous carcinoma (LUSC) patients with FANCD2 high or low expression based on TCGA database.

Clinical factors		LUAD			LUSC		
		Low	High	*p* Value	Low	High	*p* Value
Age, n (%)	Median (IQR)	67 (59, 74)	64 (59, 72)	0.036	69 (63, 74)	67 (60, 73)	0.033
	≤65	115 (22.3%)	140 (27.1%)	0.043	87 (17.6%)	104 (21.1%)	0.170
	>65	142 (27.5%)	119 (23.1%)		158 (32%)	144 (29.2%)	
Gender, n (%)	Female	158 (29.5%)	128 (23.9%)		79 (15.7%)	52 (10.4%)	0.008
	Male	109 (20.4%)	140 (26.2%)		172 (34.3%)	199 (39.6%)	
Smoker, n (%)	No	47 (9%)	28 (5.4%)	0.018	8 (1.6%)	10 (2%)	0.797
	Yes	210 (40.3%)	236 (45.3%)		238 (48.6%)	234 (47.8%)	
T stage, n (%)	T1	98 (18.4%)	77 (14.5%)	0.031	61 (12.2%)	53 (10.6%)	0.416
	T2	127 (23.9%)	162 (30.5%)		138 (27.5%)	156 (31.1%)	
	T3	29 (5.5%)	20 (3.8%)		40 (8%)	31 (6.2%)	
	T4	11 (2.1%)	8 (1.5%)		12 (2.4%)	11 (2.2%)	
N stage, n (%)	N0	175 (33.7%)	173 (33.3%)	0.580	167 (33.7%)	153 (30.8%)	0.321
	N1	44 (8.5%)	51 (9.8%)		58 (11.7%)	73 (14.7%)	
	N2	38 (7.3%)	36 (6.9%)		17 (3.4%)	23 (4.6%)	
	N3	0 (0%)	2 (0.4%)		3 (0.6%)	2 (0.4%)	
M stage, n (%)	M0	187 (48.4%)	174 (45.1%)	0.088	197 (47%)	215 (51.3%)	0.270
	M1	8 (2.1%)	17 (4.4%)		5 (1.2%)	2 (0.5%)	
Pathologic stage, n (%)	Stage I	147 (27.9%)	147 (27.9%)	0.455	128 (25.7%)	117 (23.5%)	0.491
	Stage II	61 (11.6%)	62 (11.8%)		76 (15.3%)	86 (17.3%)	
	Stage III	44 (8.3%)	40 (7.6%)		41 (8.2%)	43 (8.6%)	
	Stage IV	9 (1.7%)	17 (3.2%)		5 (1%)	2 (0.4%)	
OS event, n (%)	Alive	180 (33.6%)	163 (30.5%)	0.134	144 (28.7%)	142 (28.3%)	0.928
	Dead	87 (16.3%)	105 (19.6%)		107 (21.3%)	109 (21.7%)	0.777
DSS event, n (%)	Alive	198 (39.7%)	181 (36.3%)	0.079	183 (40.7%)	178 (39.6%)	
	Dead	51 (10.2%)	69 (13.8%)		43 (9.6%)	46 (10.2%)	
PFI event, n (%)	Alive	162 (30.3%)	147 (27.5%)		177 (35.3%)	177 (35.3%)	1.000
	Dead	105 (19.6%)	121 (22.6%)		74 (14.7%)	74 (14.7%)	

Figure 2B). Meanwhile, *FANCD2* expression was not related to pathological stage, T stage, or N stage (**Figure 2C**). *FANCD2* expression exhibited similar tendencies concerning the age and gender of patients with LUSC, but there were no obvious differences between *FANCD2* expression and other variables.

Similarly, *FANCD2* was also significantly elevated in NSCLC tissues compared to levels in normal tissues based on GSE19188, GSE75037, and GSE116959 datasets (all $p < 0.001$, **Figure 3A**). In addition, the IHC results revealed that FANCD2 was overexpressed in LUSC and LUAD tissue in comparison with that in normal tissue according to HPA databases (**Figure 3B**). These results indicated the gene and protein expression levels of FANCD2 are significantly higher in LUAD and LUSC tissues.

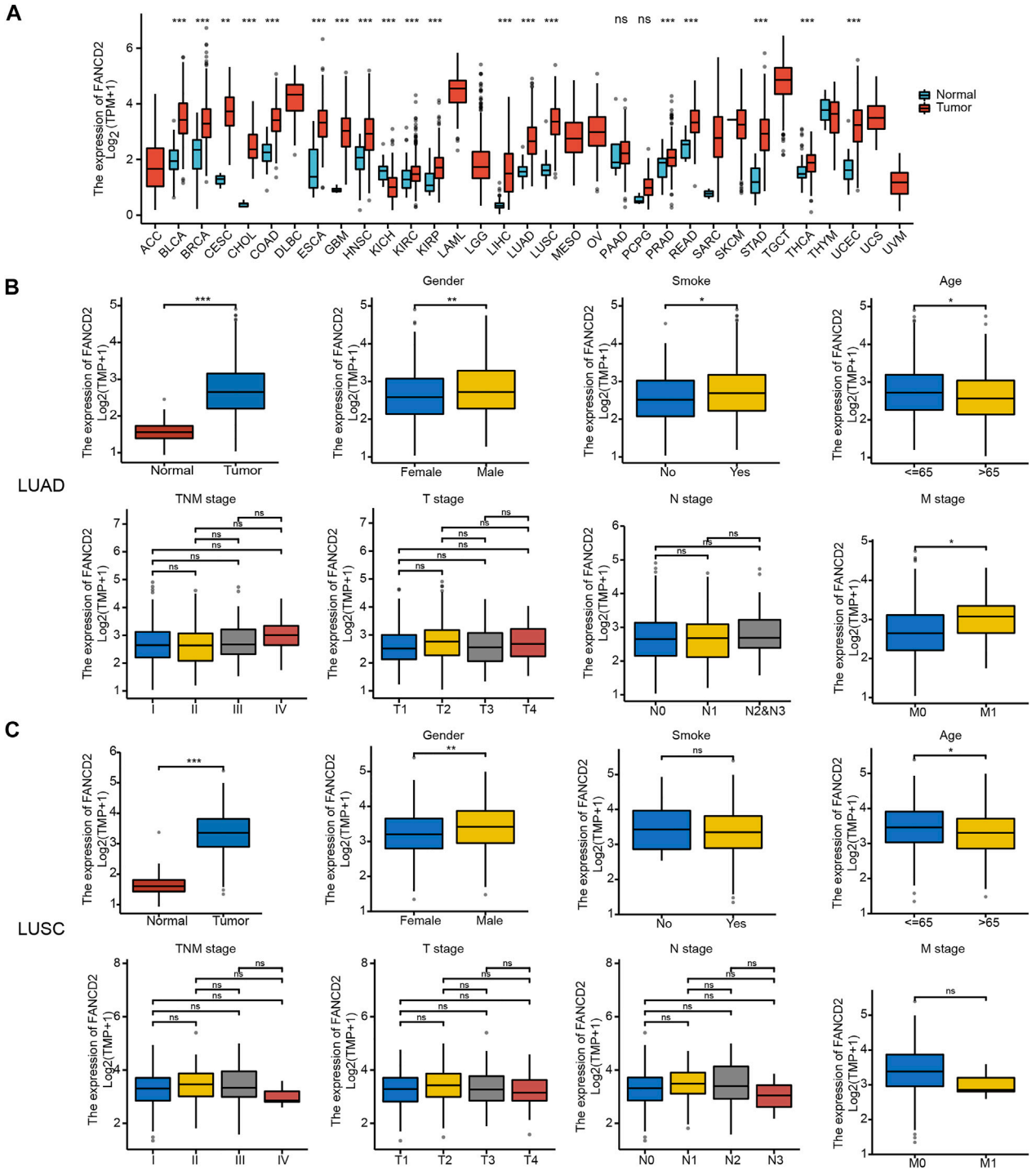


FIGURE 2 | Pan-cancer FANCD2 expression status and clinical characteristics in the LUAD and LUSC sub-groups. **(A)** FANCD2 expression status in different tumor tissues and adjacent normal tissues. **(B,C)** Comparison of FANCD2 expression levels with different clinical characteristics. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

High FANCD2 mRNA Expression Is Related to Short OS in Patients With LUAD

As shown in **Figure 4A**, the 20-year OS, DSS, and PFI rates of Patients with LUAD were remarkably higher with low *FANCD2* expression compared to those with high *FANCD2* expression (p = 0.04, 0.03, and 0.09 respectively). However, there was no significant difference between *FANCD2* expression and OS (p = 0.639), DSS (p = 0.68), and PFI (p = 0.492) in LUSC

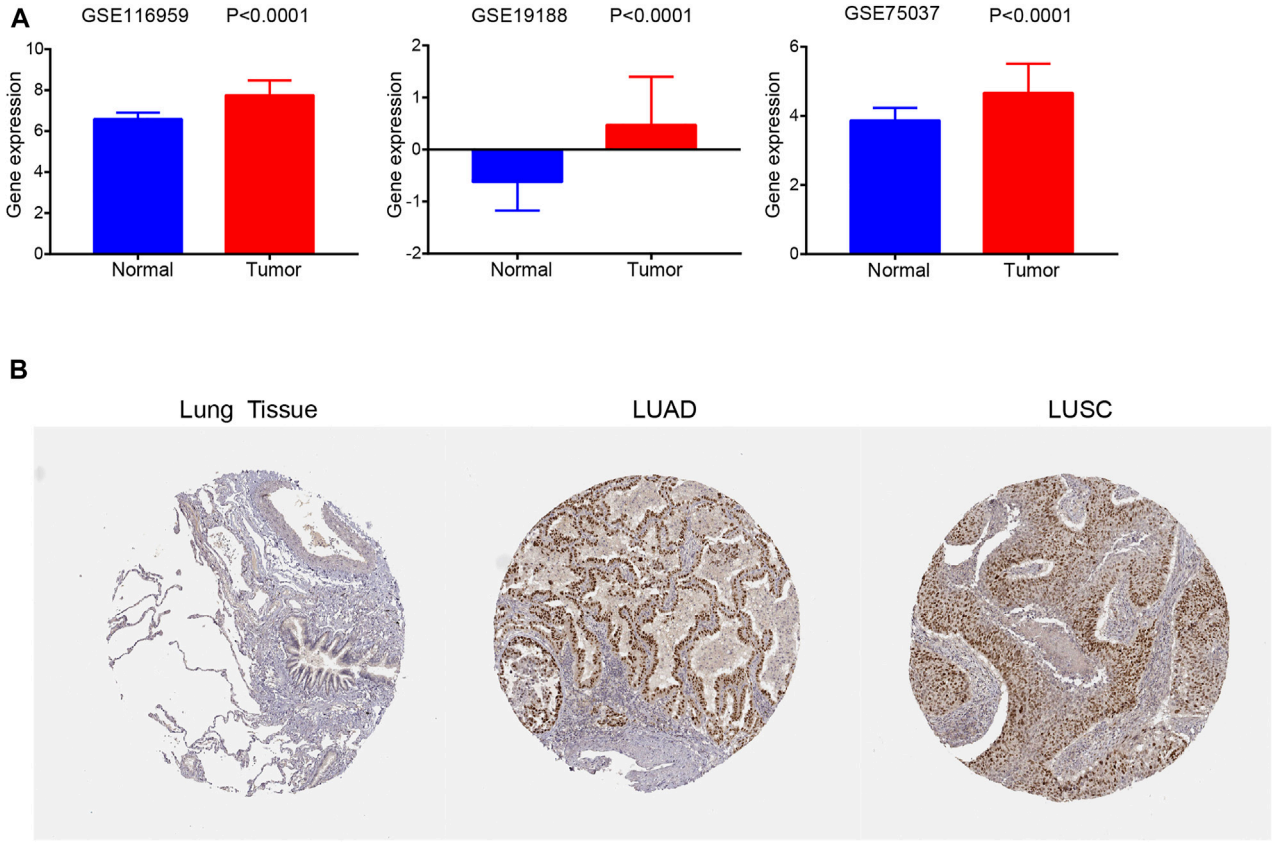


FIGURE 3 | FANCD2 expression in GEO datasets and the HPA database. **(A)** *FANCD2* mRNA expression levels were higher in tumors than in normal tissues based on GSE19188 and GSE116959 datasets; $p < 0.01$. **(B)** FANCD2 protein levels in LUAD and LUSC tissues and normal samples based on the HPA database.

patients (**Figure 4B**). Univariate analysis showed that both high *FANCD2* expression and high pathological grade and stage (TNM) were related to poor OS in Patients with LUAD (**Figure 5A**). The results of multivariate Cox regression analysis suggested that high *FANCD2* expression was an independent predictor of OS (HR = 1.716, 95% CI = 1.195–2.465, $p < 0.01$, **Figure 5B**). Therefore, *FANCD2* was considered a risk factor in predicting a worse prognosis.

Genetic Alteration and Interaction Network Analyses of FANCD2

The cBioportal online tool was then performed to explore the types and frequencies of *FANCD2* alterations in the patients with LUAD. Results revealed that *FANCD2* was highly conserved (only a 1.4% frequency of genomic alterations; **Figure 6A**). Subsequently, the interaction networks showed 20 genes (**Figure 6B**) and 10 proteins (**Figure 6C**) with the highest relevance to FANCD2. FANCI, FANCL, FAN1, FANCE, FANCC, and USP1 appeared in two networks, for which correlation scores were 0.999, 0.999, 0.999, 0.999, 0.998, and 0.998, respectively, in PPI network (**Figure 6C**). These genes/proteins belong to the FA family. FANCD2 and FANCI, via their heterodimer, serve a function in the FA pathway, in which the dimer, monoubiquitylated by the FA core complex and an E2-E3 ubiquitin ligase, participates in the recruitment of DNA repair effectors to chromatin lesions to resolve DNA damage (Nalepa and Clapp, 2018). In this process, FANCE and FANCC form part of the FA core complex; FANCE encodes an E3 ubiquitin ligase that monoubiquitylates FANCD2 and FANCI; USP1 and UAF1 regulate the deubiquitination of the FANCD2-FANCI heterodimer (Liu et al., 2010; Nalepa and Clapp, 2018; Lemonidis et al., 2021).

Functions of FANCD2 in LUAD

To better understand the relevance of FANCD2 expression in LUAD and potential mechanism, single-cell analysis was utilized to explore the associated functional states based on the CancerSEA database. The results suggested that FANCD2 expression was correlated with 14 functional states, including angiogenesis, apoptosis, cell cycle, differentiation, DNA damage, DNA repair, EMT, hypoxia, inflammation, invasion, metastasis, proliferation, quiescence, and stemness (**Figure 7A**). In addition, FANCD2 was found to be mainly positively associated with cell cycle, DNA repair, DNA damage, and proliferation but negatively correlated with angiogenesis, quiescence, inflammation, metastasis, and differentiation (all $p < 0.001$, **Figure 7B**).

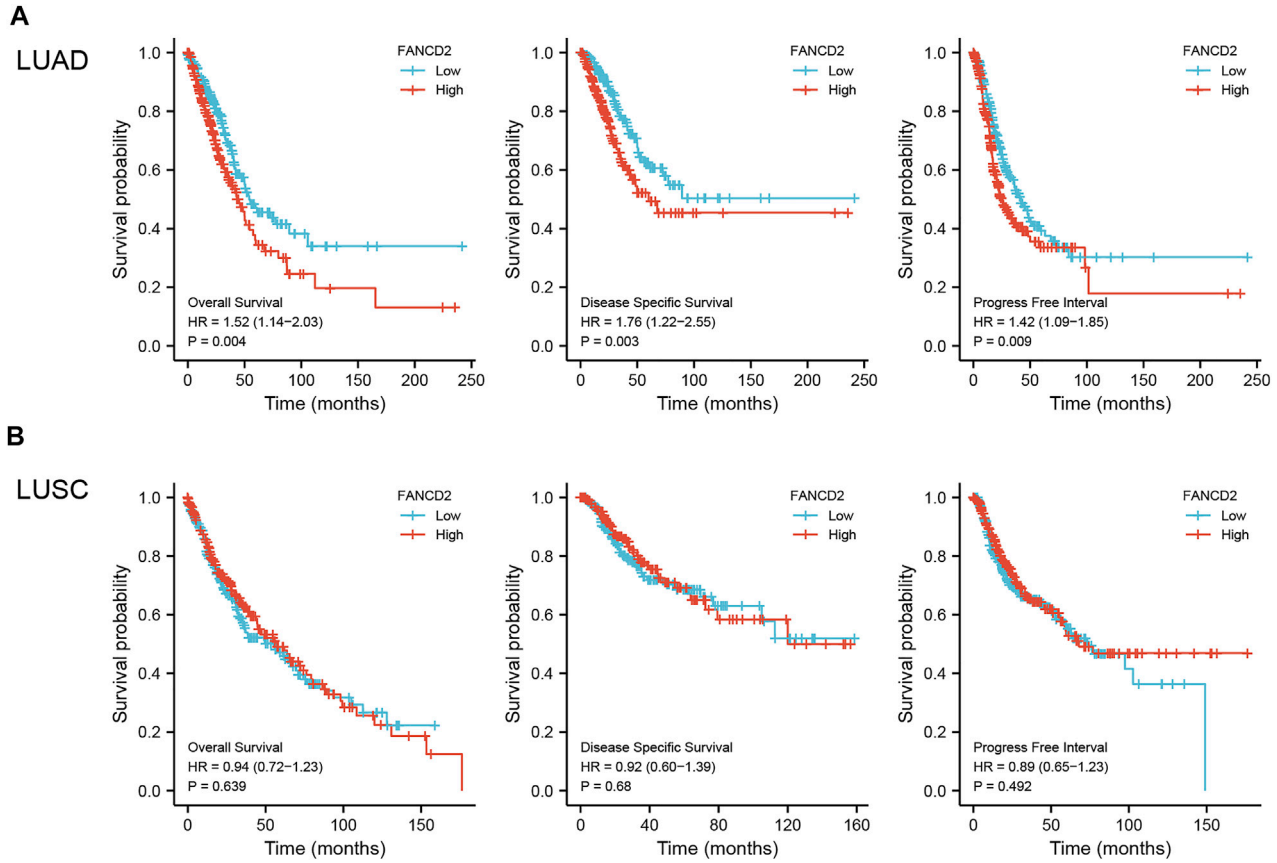


FIGURE 4 | Kaplan–Meier (K-M) survival curves showing the association between FANCD2 expression levels and overall survival (OS), disease-specific survival (DSS), and the progression-free interval (PFI) for LUAD **(A)** and LUSC **(B)** patients; $p < 0.05$ was used to assess differences.

Moreover, over-representation analysis (ORA) illustrated that FANCD2 participated in the interleukin signaling pathway, *de novo* pyrimidine deoxyribonucleotide biosynthesis, DNA replication, *de novo* purine biosynthesis, arginine biosynthesis, angiotensin II-stimulated signaling through G proteins and beta-arrestin, the circadian clock system, and the EGF receptor signaling pathway (all $p < 0.05$, **Figure 7C**). Notably, the interleukin signaling pathway had the highest correlation with FANCD2. It is well-known that interleukin family members play important roles in the immune response and inflammation. These results indicate that FANCD2 might participate in immune response and inflammation *via* the interleukin signaling pathway.

Association Between FANCD2 Expression and Immune-Inhibitory and Immune-Stimulatory Functions

A co-expression study was performed to investigate the relationship between FANCD2 expression and immunomodulators based on TISIDB, including immune inhibitors and immunostimulators. Heatmaps illustrated the association between FANCD2 expression and 12 immunoinhibitors and 28 immunostimulators across 30 tumors (**Figure 8**). Among these immunoinhibitors, FANCD2 expression had positive associations with CD274 (Cor = 0.191, p = 1.22e-05), CTLA4 (Cor = 0.157, p = 3.52e-04), LAG3 (Cor=(Cor = 0.206, p = 2.47e-06), and PDCD1 (Cor = 0.137, p = 1.8e-03) in LUAD (**Figure 8A**). Regarding various immunostimulators, FANCD2 had negative associations with TNFSF13 (Cor = −0.466, p < 2.2e-16), TMEM173 (Cor = −0.439, p < 2.2e-16), CD40LG (Cor = −0.26, p = 2.41e-09), HHLA2 (Cor = −0.254, p = 4.97e-09), and IL6R (Cor = −0.288, p = 3.23e-11) (**Figure 8B**). The above results suggest that FANCD2 is involved in regulating these immunomodulators.

Correlation Between FANCD2 Expression and Infiltrating Immune Cells

Tumor-infiltrating lymphocytes are associated with the prognosis of patients with multiple cancers (Salgado and Loi, 2018). As shown in **Figure 9A**, FANCD2 expression levels were strongly associated with levels of infiltrating B cells (Cor = −0.157, p = 4.84e-04), $CD8^+$ T cells (Cor = 0.127, p = 4.74e-03), macrophages (Cor = 0.105, p = 1.97e-02), and neutrophils (Cor = 0.232, p =

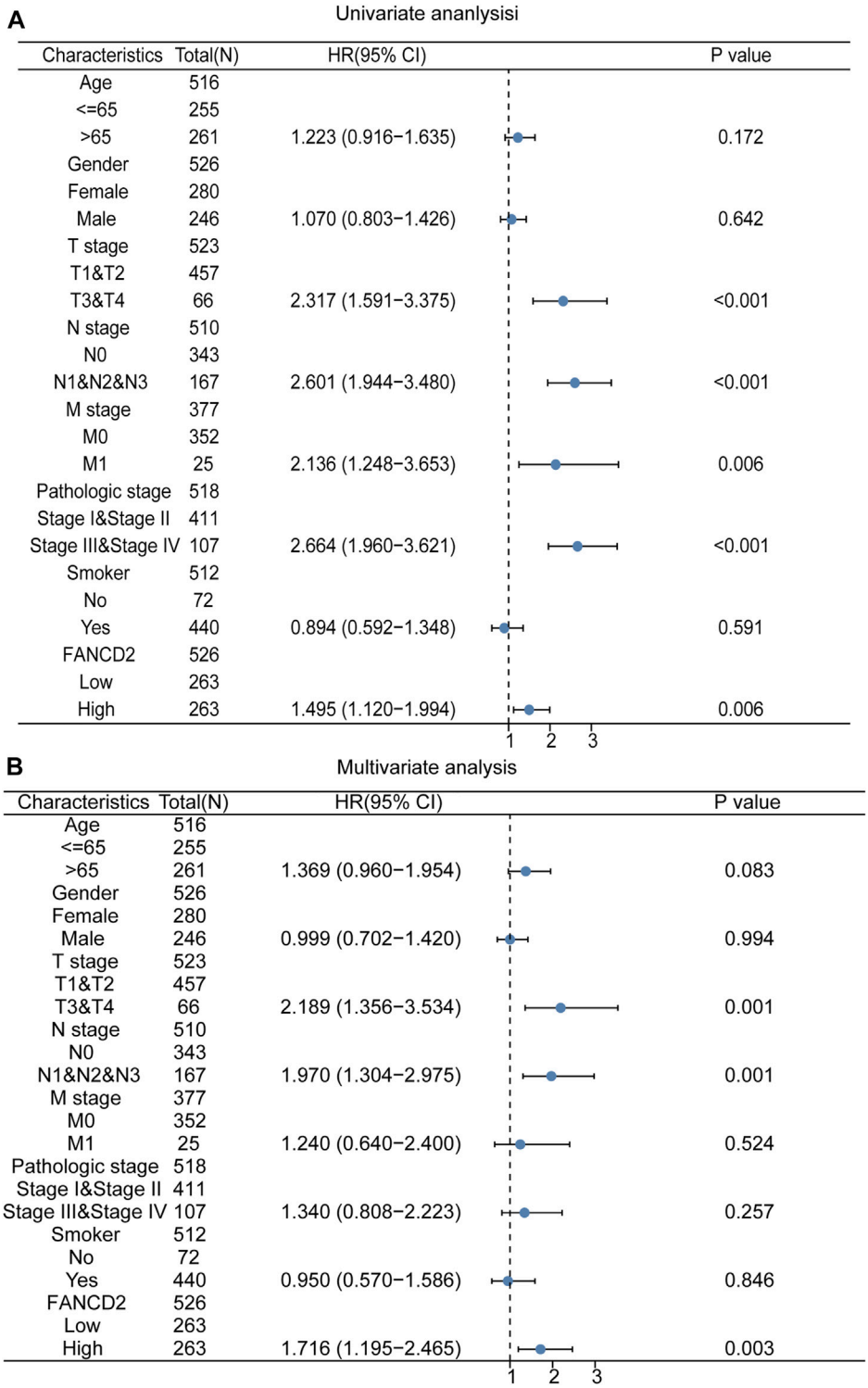
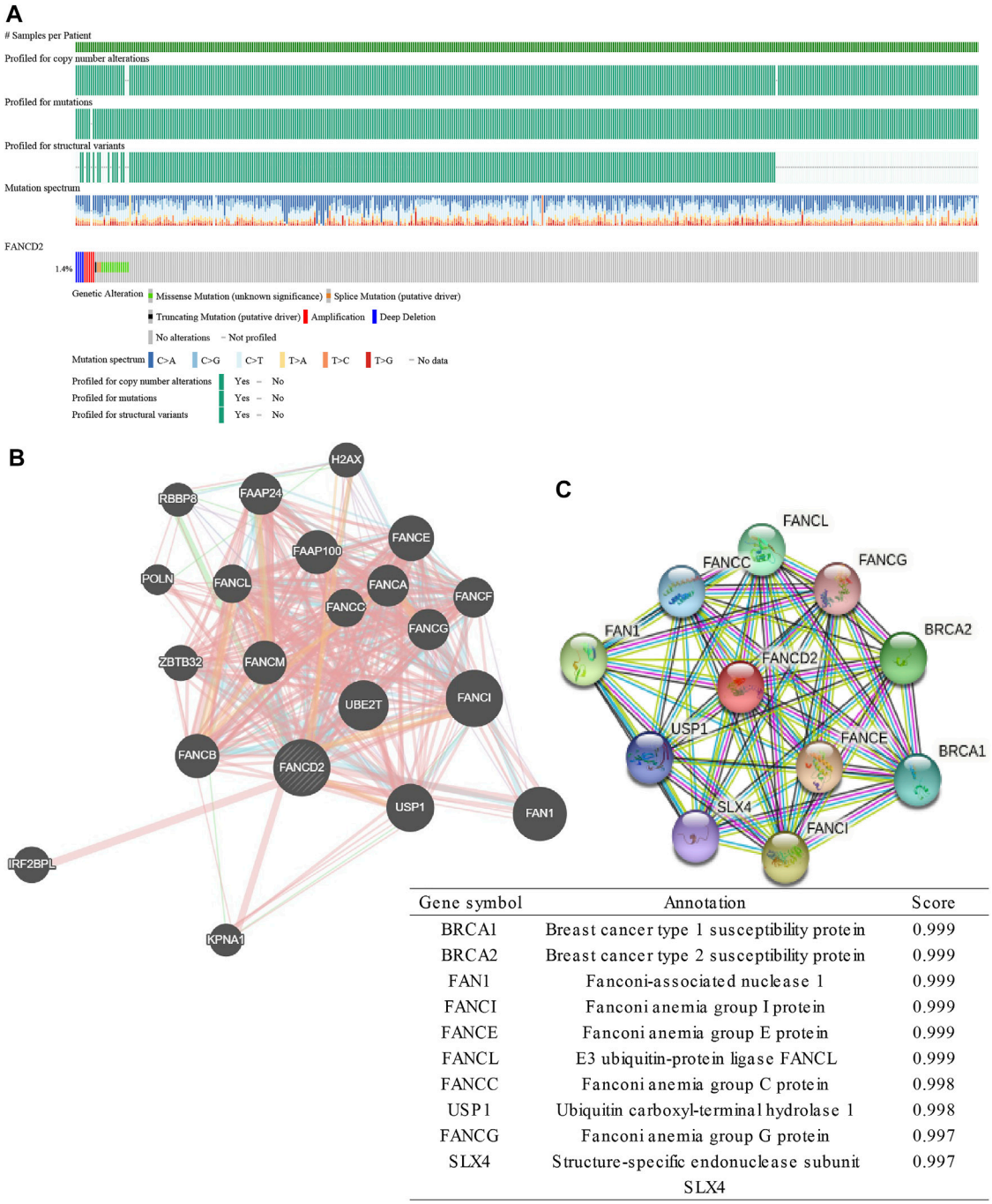


FIGURE 5 | Univariate **(A)** and multivariate **(B)** analyses of FANCD2 expression and important clinicopathological parameters concerning prognosis among Patients with LUAD.

1.82e-07) in LUAD cases. However, there was no association between FANCD2 expression and tumor purity, DCs, or $CD4^+$ T cells. To gain more insight into the relationship between FANCD2 expression and immune infiltration, this study assessed subjects based on 24 types of infiltrating immune cells using the ssGSEA database (**Figure 9B**). Specifically,



Gene symbol	Annotation	Score
BRCA1	Breast cancer type 1 susceptibility protein	0.999
BRCA2	Breast cancer type 2 susceptibility protein	0.999
FAN1	Fanconi-associated nuclease 1	0.999
FANCI	Fanconi anemia group I protein	0.999
FANCE	Fanconi anemia group E protein	0.999
FANCL	E3 ubiquitin-protein ligase FANCL	0.999
FANCC	Fanconi anemia group C protein	0.998
USP1	Ubiquitin carboxyl-terminal hydrolase 1	0.998
FANCG	Fanconi anemia group G protein	0.997
SLX4	Structure-specific endonuclease subunit SLX4	0.997

FIGURE 6 | Genomic alterations of FANCD2 and the interaction network and protein interaction network of FANCD2. **(A)** Genomic alterations of FANCD2 based on cBioPortal. **(B)** The gene–gene interaction network of FANCD2, as constructed with GeneMANIA. **(C)** The protein-protein network of FANCD2 is derived from STRING.

FANCD2 was negatively related to B cells, $CD8^+$ T cells, DC cells, macrophages, eosinophils, Idc (interdendritic) cells, mast cells, neutrophils, NK CD56 bright cells, pDC (plasmacytoid dendritic) cells, TFH cells, TH17 cells, and NK cells but was positively related to Th2 cells, T helper cells, and Tcm cells (all $p < 0.001$). Moreover, the heat map showed that most subpopulations among the 24 types of immune cells had moderate to strong relationships (**Figure 9C**). These findings evealed that FANCD2 plays an important role in immune infiltration in LUAD.

GEPIA and TIMER databases were used to further assess the correlation between FANCD2 and the marker sets of diverse immune cells in LUAD. **Table 2** shows that multiple markers of immune cells were significantly related to FANCD2 expression, including Th2 (GATA3), Th9 (TGFBR2), Th17 (IL-21R), Treg

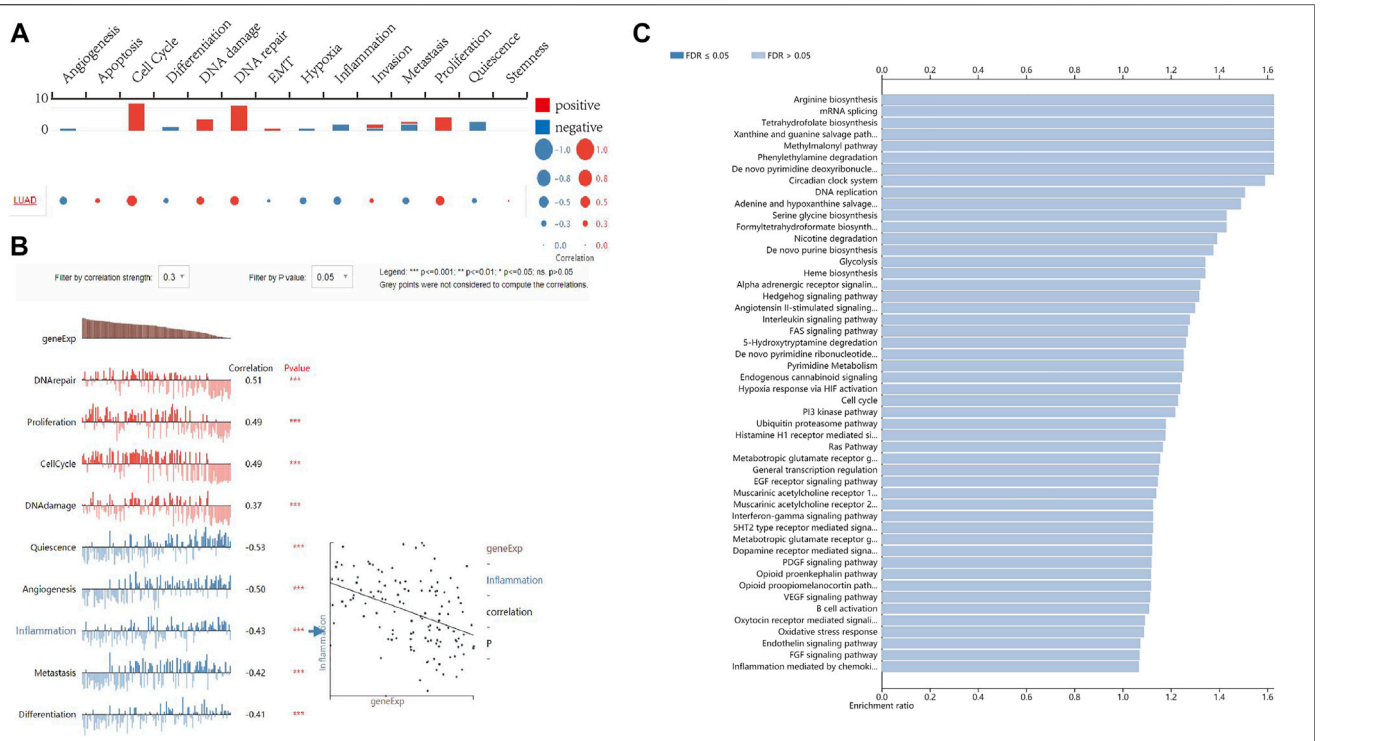


FIGURE 7 | The function of FANCD2 in LUAD. **(A,B)** The single-cell analysis shows that FANCD2 is involved in the cell cycle, DNA repair, proliferation, angiogenesis, quiescence, inflammation, metastasis, and differentiation. **(C)** The biological processes related to FANCD2 are based on ORA.

(FOXP3, CD25, CCR8), T cell exhaustion (PD-1, CTLA4, LAG3), tumor-associated macrophages (TAMs) (CD80, CCR5), and DCs (CD1C, CD141). These results implied that FANCD2 might affect the function of immune cells by modulating marker gene expression.

DISCUSSION

LUAD is the most common subtype of lung cancer. Currently, the efficacy of surgery, radiotherapy, chemotherapy, and targeted therapy is not satisfying. Notably, ferroptosis is a novel form of cell death, and an increasing body of research has confirmed that it plays a crucial role in anti-tumor treatment, especially in LUAD (Ma et al., 2021; Zhang X. et al., 2021). Moreover, a connection between ferroptosis and cell immunity and cancer immunotherapy has been shown, but the underlying mechanism is not clear (Wang et al., 2019). The latest study showed that 76.9% of ferroptosis-related genes are differentially expressed between LUAD tumor tissues and adjacent normal tissues, and some of these DEGs were determined to be remarkably associated with OS (Gao et al., 2021). Thus, ferroptosis-related genes are valuable prognostic markers for LUAD.

FANCD2, a member of the FA protein family, participates in the maintenance of genomic stability *via* the FA pathway. The interaction network derived from GeneMANIA and STRING shows a close connection between FANCD2 and other FA family genes/proteins. Until now, at least twenty-two FA proteins that form part of the FA core complex have been identified. The FA core complex participates in the recruitment and monoubiquitination of the heterodimer FANCD2-FANCI (Tsui and Crismani, 2019). Ubiquitylation of the FANCD2-FANCI heterodimer enables the recruitment of DNA repair effectors (Nalepa and Clapp, 2018), which participate in three classic DNA repair pathways, including nucleotide excision repair, homologous recombination, and mutagenic translesion synthesis (Moldovan and D'Andrea, 2009). As a nuclear protein, FANCD2 supports the maintenance of a stable genome, but it also has a negative regulatory role in ferroptosis, which is mainly involved in two biological pathways: iron accumulation and lipid peroxidation. Tumor cells with low FANCD2 expression undergo ferroptosis easily. Specifically, FANCD2 deficiency contributes to lipid peroxidation through a decrease in glutathione peroxidase 4 (GPX4), as well as the accumulation of iron through an increase in the expression of transferrin (TF), and a decrease in ferritin heavy chain 1 (FTH1) and SLC40A1 (Song et al., 2016).

This study investigated the role of FANCD2 in LUAD progression and prognosis, as well as its relationship with immune cell infiltration (**Figure 10**). We observed that the mRNA and protein expression of FANCD2 were upregulated in LUAD samples compared to levels in normal tissues in the TCGA, GEO, and HPA databases, and patients with higher FANCD2 had shorter OS, poor DSS, and worse PFI based on KM plots in LUAD. Moreover, univariate and multivariate Cox analysis further confirmed that high expression of FANCD2 was

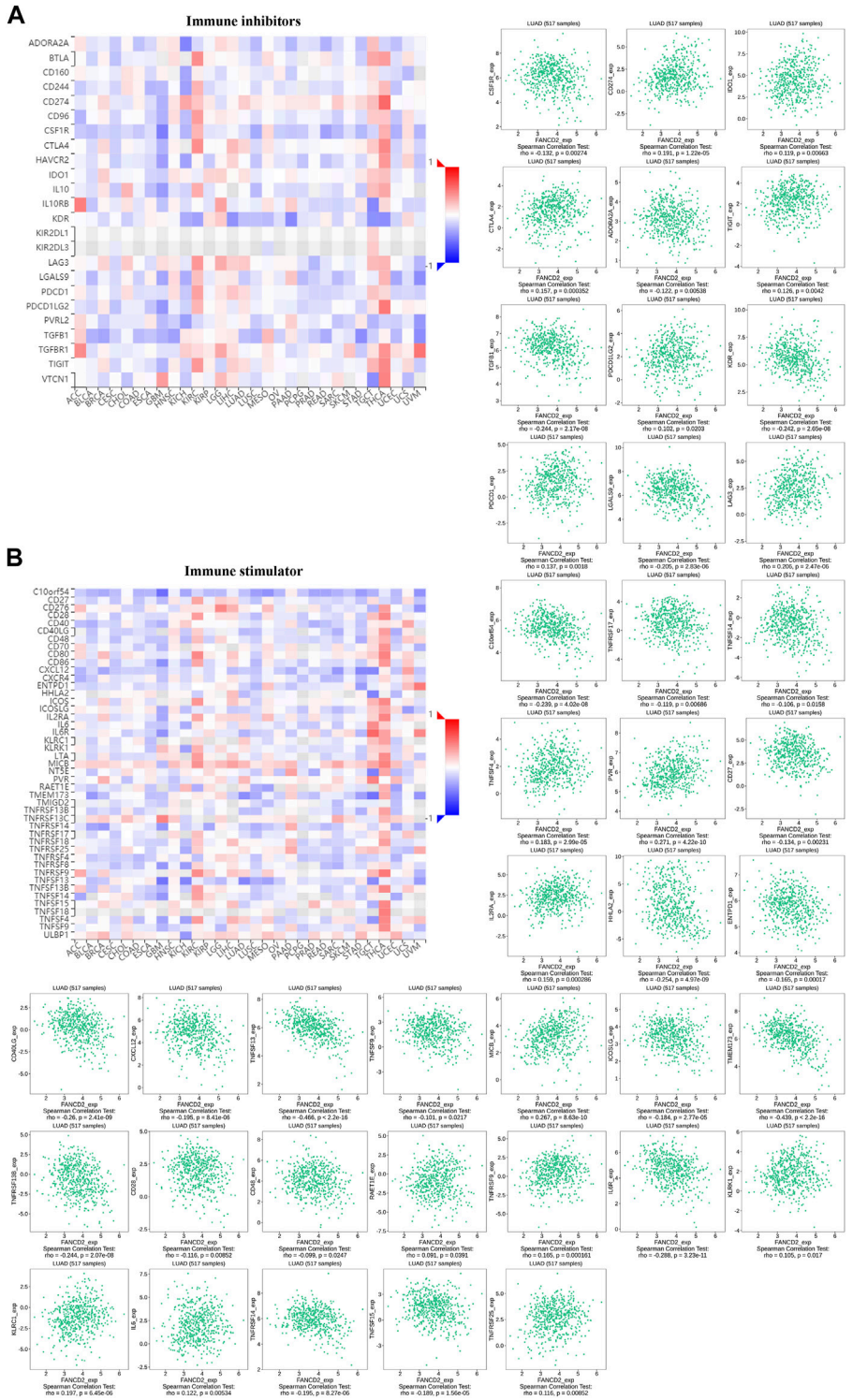
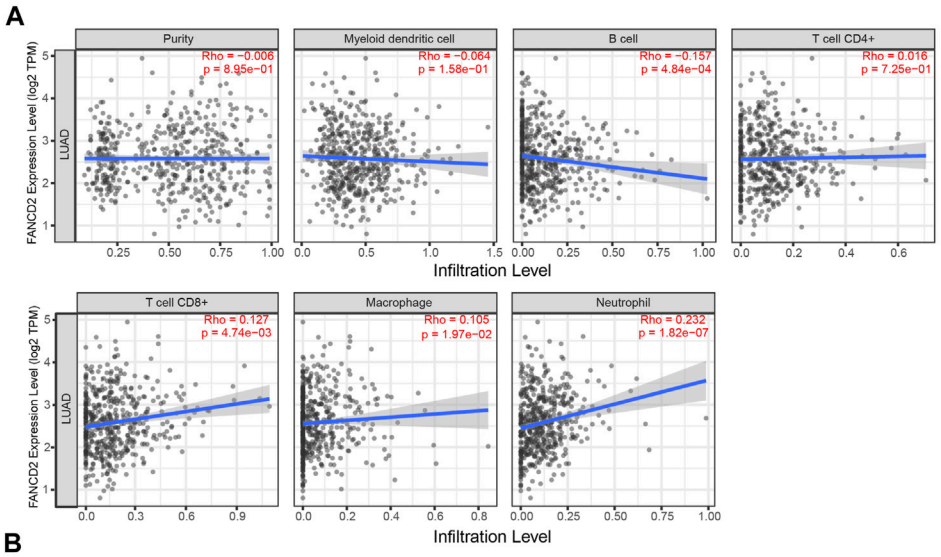
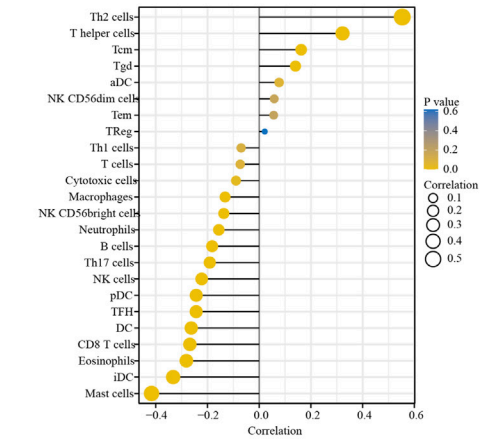


FIGURE 8 | Co-expression analysis of FANCD2 with immunoinhibitory **(A)** and immunostimulatory genes **(B)** in the pan-cancer database, as well as detailed information on LUAD.

an independent adverse prognostic factor, which was consistent with the results of Lei et al. (2020). Some research shows that inhibiting FANCD2 function or decreasing the expression of FANCD2 enhances the sensitivity of patients to chemotherapy, such as cisplatin, and the efficacy of treatments such as ICL agent and ionizing radiation (IR) (Wang et al., 2015; Yang et al., 2016).





Cell Type	Pearson	P value
aDC	0.030	0.486
B cells	-0.189	<0.001
CD8 T cells	-0.266	<0.001
Cytotoxic cells	-0.104	0.017
DC	-0.274	<0.001
Eosinophils	-0.274	<0.001
iDC	-0.336	<0.001
Macrophages	-0.161	<0.001
Mast cells	-0.419	<0.001
Neutrophils	-0.182	<0.001
NK CD56bright cells	-0.191	<0.001
NK CD56dim cells	0.019	0.667
NK cells	-0.229	<0.001
pDC	-0.264	<0.001
T cells	-0.089	0.039
T helper cells	0.307	<0.001
Tcm	0.193	<0.001
Tem	0.067	0.121
TFH	-0.255	<0.001
Tgd	0.140	0.001
Th1 cells	-0.092	0.033
Th17 cells	-0.224	<0.001
Th2 cells	0.542	<0.001
Treg	-0.016	0.720

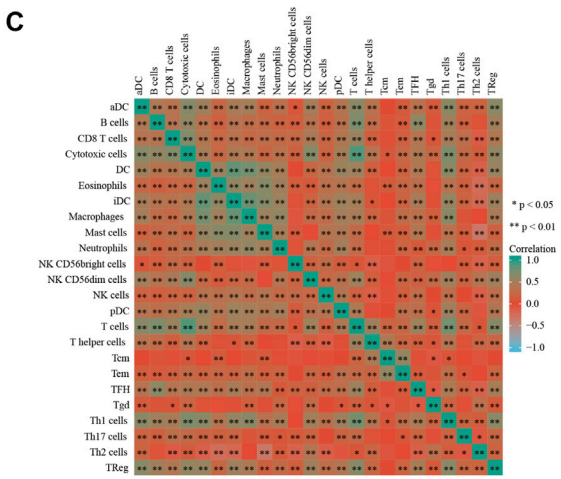


FIGURE 9 | Correlation between FANCD2 expression and immune infiltration levels in LUAD. **(A)** Correlation analysis of FANCD2 expression and the infiltration of six types of immune cells based on TIMER. **(B)** Forest plots show that FANCD2 expression was positively correlated with 8 types of immune cells and negatively correlated with 16 subsets of immune cells. The sizes of dots represent the absolute value of Pearson r. **(C)** Heatmap showing the relationship among 24 types of immune cells in LUAD. *$p < 0.05$, **$p < 0.01$.

TABLE 2 | Correlation between FANCD2 levels and markers of immune cells based on TIMER and GEPIA databases.

Cell type	Gene marker	None		Purity		Tumor		Normal	
		Cor	P	Cor	P	R	P	R	P
B cell	CD19	−0.003	0.9410	0.004	0.9210	−0.085	0.063	0.41	0.0011**
	CD20(KRT20)	0.027	0.5440	0.021	0.6440	0.071	0.12	−0.039	0.77
	CD38	0.049	0.2720	0.064	0.1560	−0.0043	0.92	0.14	0.29
$CD8^+$ T cell	CD8A	0.201	***	0.231	***	0.078	0.085	0.11	0.41
	CD8B	0.192	***	0.202	***	0.12	0.0082**	0.085	0.52
	BCL6	0.019	0.6600	0.024	0.5960	0.056	0.22	0.13	0.31
	ICOS	0.162	0.0002***	0.202	***	0.069	0.13	0.32	0.014**
	CXCR5	−0.013	0.7760	−0.005	0.9150	−0.098	0.032	0.4	0.0017**
Th1	T-bet (TBX21)	0.156	0.0004***	0.184	***	0.035	0.44	−0.033	0.8
	STAT4	0.081	0.0646	0.098	0.0291	0.033	0.47	0.31	0.18
	IL12RB2	0.457	***	0.483	***	0.22	***	0.14	0.29
	WSX1(IL27RA)	−0.047	0.2900	−0.043	0.3410	−0.058	0.21	0.24	0.071
	STAT1	0.44	***	0.472	***	0.3	***	0.12	0.36
	IFN-γ(IFNG)	0.311	***	0.337	***	0.19	0.000026***	0.075	0.57
	TNF-α(TNF)	0.103	0.0194*	0.124	0.0056**	0.055	0.23	0.33	0.011*
Th2	GATA3	0.174	0.0001***	0.204	***	0.19	0.000026***	0.03	0.82
	CCR3	−0.02	0.6430	0.001	0.9890	−0.015	0.74	0.11	0.41
	STAT6	−0.07	0.1120	−0.077	0.0858	−0.028	0.55	0.27	0.042*
	STAT5A	0.055	0.2140	0.076	0.0925	−0.0088	0.85	0.44	0.00056**
Th9	TGFBR2	−0.13	0.0032**	−0.123	0.0064***	−0.12	0.0077***	−0.19	0.14
	IRF4	0.062	0.1620	0.077	0.0858	−0.068	0.13	0.44	0.00053**
	PU.1(SPI1)	−0.039	0.3770	−0.038	0.4040	−0.11	0.015*	0.24	0.071
Th17	STAT3	0.023	0.5980	0.021	0.6350	0.021	0.64	0.12	0.38
	IL-21R	0.149	0.0007***	0.188	***	0.037	0.41	0.26	0.05
	IL-23R	0.002	0.9580	0.006	0.9010	−0.033	0.47	0.11	0.39
	IL-17A	0.089	0.0436*	0.089	0.0494	0.073	0.11	0.17	0.19
Th22	CCR10	0.094	0.0321*	0.08	0.0774	0.11	0.013*	−0.086	0.52
	AHR	−0.014	0.7550	−0.014	0.7500	−0.012	0.79	0.003	0.98
Treg	FOXP3	0.157	0.0004***	0.183	***	0.04	0.38	0.42	0.001**
	CD25(IL2RA)	0.24	***	0.271	***	0.15	0.00068	0.11	0.39
	CCR8	0.183	***	0.214	***	0.064	0.16	0.21	0.11
T cell exhaustion	PD-1(PDCD1)	0.219	***	0.259	***	0.083	0.068	0.38	0.0029**
	CTLA4	0.24	***	0.29	***	0.14	0.0029	0.27	0.04
	LAG3	0.265	***	0.286	***	0.12	0.0086	0.43	0.00071***
	TIM-3(HAVCR2)	0.074	0.0951	0.088	0.0516	−0.033	0.46	0.097	0.46
Macrophage	CD68	0.034	0.4350	0.044	0.3340	−0.032	0.49	0.15	0.26
	CD11b (ITGAM)	0.006	0.9000	0.019	0.6670	−0.026	0.57	0.23	0.074
M1	INOS(NOS2)	0.042	0.3420	0.029	0.5260	0.048	0.29	−0.25	0.058
	IRF5	0.125	0.0045**	0.134	0.0028	0.087	0.057	0.4	0.0016**
	COX2(PTGS2)	0.016	0.7150	0	0.9920	0.02	0.66	−0.048	0.72
M2	CD16	0.042	0.3440	0.055	0.2230	−0.0086	0.85	0.065	0.62
	ARG1	−0.001	0.9830	−0.011	0.8150	0.059	0.19	−0.038	0.77
	MRC1	−0.083	0.0588	−0.079	0.0786	−0.075	0.1	0.38	0.0031**
	MS4A4A	−0.001	0.9860	0.007	0.8760	−0.083	0.069	0.25	0.06
TAM	CCL2	0.074	0.0921	0.081	0.0739	0.0014	0.97	−0.015	0.91
	CD80	0.129	0.0033**	0.153	0.0006	0.079	0.082	0.15	0.25
	CD86	0.078	0.0755	0.095	0.0355	−0.028	0.54	0.22	0.1
	CCR5	0.14	0.0014	0.177	0.0001	0.056	0.22	0.27	0.04*
Monocyte	CD14	0.041	0.3530	0.058	0.1960	−0.06	0.19	−0.12	0.35
	CD16(FCGR3B)	0.042	0.3440	0.055	0.2230	−.0086	0.85	0.065	0.62
	CD115(CSF1R)	0.011	0.8050	0.026	0.5690	−0.047	0.3	0.34	0.00088**
Neutrophil	CD66b (CEACAM8)	−0.23	***	−0.229	***	−0.1	0.025	−0.013	0.92
	CD15(FUT4)	0.058	0.1890	0.047	0.2940	0.15	0.0013	0.15	0.26
	CD11b (ITGAM)	0.006	0.9000	0.019	0.6670	−0.026	0.57	0.23	0.074
Natural killer cell	XCL1	0.225	***	0.261	***	0.2	0.0000058	−0.17	0.19
	CD7	0.191	***	0.219	***	0.063	0.17	0.16	0.22
	KIR3DL1	0.035	0.4270	0.042	0.3560	0.00052	0.99	−0.19	0.16
Dendritic cell	CD1C(BDCA-1)	−0.339	***	−0.344	***	−0.23	0.00000028	0.11	0.43
	CD141 (THBD)	−0.183	***	−0.188	***	−0.075	0.1	−0.12	0.38
	CD11c	0.133	0.0026**	0.154	0.0006	0.011	0.81	0.35	0.0071**

*correlation without adjustment, Purity: correlation adjusted by purity, Cor, correlation coefficient. *p < 0.05, **p < 0.01, and ***p < 0.001.*

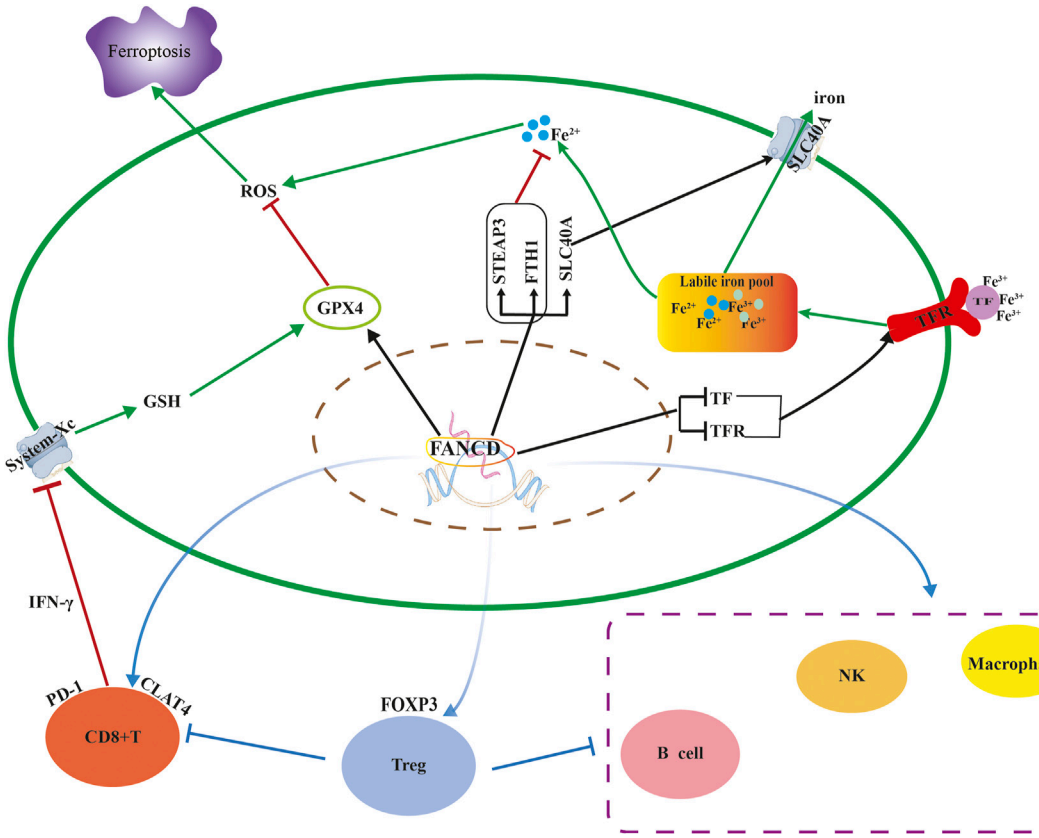


FIGURE 10 | The mechanism of FANCD2 regulating ferroptosis and immune cell infiltration.

In addition, downregulated FANCD2 significantly inhibits tumor growth in nude mice (Fan et al., 2021). Therefore, high expression of FANCD2 was considered a poor prognostic biomarker for Patients with LUAD.

To be better able to elaborate on the molecular mechanisms of the highly conserved gene FANCD2 in LUAD, CancerSEA, and ORA were conducted to further investigate its function. CancerSEA at a single-cell level illustrated that FANCD2 participates in inflammation, and its intensity decreases with the expression of FANCD2; ORA results showed that FANCD2 participates in the interleukin signaling pathway. Numerous studies have confirmed that factors of the interleukin family, such as IL8, IL10, and IL17, are involved in immune responses and are associated with the outcome of patients (Schalper et al., 2020; Zhang H. et al., 2020; Zhang Y. et al., 2020). Therefore, we further investigated the relationship between FANCD2 expression and tumor immunity in LUAD.

We found that FANCD2 was closely associated with immunomodulating factors. A high level of FANCD2 expression upregulated the immune inhibitor expression of CTLA4, PDCD1, and LAG3 while downregulating immunostimulators, such as IL6R, TMEM173, TNFSF13, CD40LG, and HHLA2, which indicates that FANCD2 contributes to tumor immune escape by modulating the immunosuppressive microenvironment. This finding is consistent with previous studies (Hong et al., 2021; Yang et al., 2021). Moreover, higher FANCD2 expression led to a remarkable reduction in the infiltration of $CD8^+$ T cells, NK cells, and DC cells, but it also recruited Th2 cells, and it is closely connected to the corresponding marker genes of tumor infiltrates immune cells (TIICs) based on the ssGSEA, CIBERSORT, and GEPIA databases. These results validate the role of FANCD2 in the LUAD immune system.

$CD8^+$ T cells, as preferred cancer-targeting immunotherapy cells, through exocytosis and the release of perforin-granzyme and activation of caspases *via* the release of cytochrome c in cancer cells, contribute to tumor cell apoptosis (Farhood et al., 2019). NK cells are the first line of defense against tumors, and they not only release perforin and granzymes but also excrete various cytokines (IFN-γ, TNF), chemokines (IL10), or growth factors (GM-CSF) to play a crucial role in antitumor effects and antiviral infection. Notably, IFN-γ enhances the function of antigen-presenting cells, inhibits angiogenesis, induces Th1 cells, and promotes M1 macrophage polarization, which remarkably increases the effect of immune surveillance and immune elimination in the TME (Morvan and Lanier, 2016). In addition, $CD141^+$ DCs express lymphotoxin beta transcripts to contribute to lymphocyte recruitment, the priming and proliferation of cytotoxic T cells, and $CD1C^+$ subpopulations of DCs that promote the maintenance of immune memory

(Nizzoli et al., 2016; Lavin et al., 2017). In contrast, Th2 cells secrete anti-inflammatory factors such as IL-4, IL-5, and IL-10 to weaken the anti-tumor immune response. GATA3, as the genetic marker of Th2 cells, was positively correlated with the FANCD2 level, and this not only promotes Th2 differentiation but also inhibits Th1 differentiation (Yagi et al., 2011). Hence, the Th2 shift in the TME is considered to promote tumor relapse, metastasis, and poor prognosis (Liu et al., 2019).

TAMs participate in angiogenesis and lymphangiogenesis, contributing to the progression of NSCLC (Hwang et al., 2020). However, this study showed that the FANCD2 expression level had no significant association with the infiltration of TAMs, but it could modulate the expression of CD80 and CCR5 to enhance the immunosuppressive function of TAMs. A previous study suggested that CD80 binds to the CTLA-4 receptor, inhibiting T-cell activation (Chikuma, 2017). TAMs could independently stimulate tumor cell growth and migration *via* the CCL5/CCR1/CCR5 axis (Pham et al., 2020). These findings revealed that FANCD2 plays a crucial role in recruiting different TIICs and regulating anti-tumor immunity.

In addition, Treg cell markers, such as FOXP3 and CD25, which have a crucial function in suppressing the antitumor immune response (Litwin et al., 2021), were strongly correlated with FANCD2 expression. Several studies have documented that live tumor-infiltrating Tregs are related to poor prognosis in NSCLC patients (Shimizu et al., 2010). However, apoptotic Treg cells also mediate immunosuppression *via* the adenosine and A2A pathways (Maj et al., 2017). Intriguingly, Treg cells that undergo ferroptosis caused by GPX4 deficiency potentiate antitumor immunity, characterized by high ratios of cytotoxic $CD8^+$ T cells to $CD4^+$ T cells in the TME (Xu et al., 2021). Therefore, inducing cell ferroptosis of Treg cells is also an antitumor treatment.

Based on the results of previous studies, there is a synergism between ferroptosis and immunomodulation (Hong et al., 2021; Xu et al., 2021; Yang et al., 2021). In TME, macrophages can be converted from M2 to M1, making more H_2O_2 available in the Fenton reaction, resulting in the ferroptosis of tumor cells (Zanganeh et al., 2016). An additional study demonstrated that activated $CD8^+$T cells release IFN-γ to restrain system xc- uptake cystine, promoting tumor cell lipid peroxidation and subsequently contributing to ferroptosis (Wang et al., 2019). When tumor cells undergo ferroptosis, tumor antigens are released, which creates an immunogenic TME, thus enhancing the response to immunomodulation (Zhang F. et al., 2019).

Given the function of FANCD2 in ferroptosis regulation and TME and the infiltration level of TIICs, FANCD2 is a crucial molecule for synergetic ferroptosis-induction treatment and immunotherapy. Therefore, FANCD2 might be a powerful predictor of patient outcomes, and its expression level is a potential novel standard to select treatment options for Patients with LUAD clinically. Patients with a high level of FANCD2 are more suitable for ferroptosis-induction treatment and/or immunotherapy.

This study explored the predictive value of FANCD2 and uncovered a potential mechanism of activity in LUAD tumorigenesis. However, it has several limitations. First, this study was not a prospective study, and all data analyzed were obtained from public databases. Second, due to the inability to receive more detailed patient information, the baseline of the survival curves is unadjusted, and the results might be biased. Third, the functions of FANCD2 in ferroptosis and tumor immunity, as well as their mechanism, have not been clarified *in vitro or in vivo*. Clearly, clinical studies are needed to validate its prognostic value, and more in-depth experimental studies are required to reveal the mechanisms.

CONCLUSIONS

This study systematically analyzed the role of FANCD2 in tumor progression, prognosis, and therapy for patients with LUAD. These results demonstrated that upregulated FANCD2 contributes to immune escape and was associated with worse outcomes for Patients with LUAD. It might be associated with FANCD2 participating in maintaining a stable tumor cell genome, protecting cells from ferroptosis, and constructing an immunosuppressive microenvironment. Furthermore, FANCD2 recruits immunosuppressive cells into the TME and regulates the expression of corresponding immune markers to weaken the anti-tumor immune response (**Figure 10**). Hence, FANCD2 is a biomarker for predicting human LUAD prognoses and may be a novel potential bio-target for identifying patients who may benefit from ferroptosis-induction treatment and/or immunotherapy.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

JZ wrote the original draft, FX and MG prepared the figures and tables, DW and XC analyzed the data, LJ and MY downloaded the raw data from TCGA and GEO databases, DZ reviewed the relevant literature, WC proofread the manuscript, and FX edited the draft and made revisions.

FUNDINGS

The present study was supported by the National Natural Science Foundation of China (grant No. 82004281), the China Postdoctoral Science Foundation (grant No. 2021T140427 and 2021M691986), and the Development Plan of Shandong Medical and Health Technology (grant No. 2019WS581).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.825685/full#supplementary-material

REFERENCES

Bebber, C. M., Müller, F., Prieto Clemente, L., Weber, J., and von Karstedt, S. (2020). Ferroptosis in Cancer Cell Biology. *Cancers* 12 (1), 164. doi:10.3390/cancers12010164

Brahmer, J. R., Govindan, R., Anders, R. A., Antonia, S. J., Sagorsky, S., Davies, M. J., et al. (2018). The Society for Immunotherapy of Cancer Consensus Statement on Immunotherapy for the Treatment of Non-small Cell Lung Cancer (NSCLC). *J. Immunother. cancer* 6 (1), 75. doi:10.1186/s40425-018-0382-2

Chikuma, S. (2017). CTLA-4, an Essential Immune-Checkpoint for T-Cell Activation. *Curr. Top. Microbiol. Immunol.* 410, 99–126. doi:10.1007/82_2017_61

Deng, H.-Y., Zeng, M., Li, G., Alai, G., Luo, J., Liu, L.-X., et al. (2019). Lung Adenocarcinoma Has a Higher Risk of Lymph Node Metastasis Than Squamous Cell Carcinoma: A Propensity Score-Matched Analysis. *World J. Surg.* 43 (3), 955–962. doi:10.1007/s00268-018-4848-7

Fan, X.-Z., Chen, Y.-F., Zhang, S.-B., He, D.-H., Wei, S.-F., Wang, Q., et al. (2021). Centipeda Minima Extract Sensitizes Lung Cancer Cells to DNA-Crosslinking Agents via Targeting Fanconi Anemia Pathway. *Phytomedicine* 91, 153689. doi:10.1016/j.phymed.2021.153689

Farhood, B., Najafi, M., and Mortezaee, K. (2019). CD8 + Cytotoxic T Lymphocytes in Cancer Immunotherapy: A Review. *J. Cell. Physiology* 234 (6), 8509–8521. doi:10.1002/jcp.27782

Gao, M., and Jiang, X. (2018). To Eat or Not to Eat - the Metabolic Flavor of Ferroptosis. *Curr. Opin. Cell. Biol.* 51, 58–64. doi:10.1016/j.ceb.2017.11.001

Gao, X., Tang, M., Tian, S., Li, J., and Liu, W. (2021). A Ferroptosis-Related Gene Signature Predicts Overall Survival in Patients with Lung Adenocarcinoma. *Future Oncol.* 17 (12), 1533–1544. doi:10.2217/fon-2020-1113

Hassannia, B., Vandenabeele, P., and Vanden Berghe, T. (2019). Targeting Ferroptosis to Iron Out Cancer. *Cancer Cell.* 35 (6), 830–849. doi:10.1016/j.ccell.2019.04.002

Hong, Y., Lin, M., Ou, D., Huang, Z., and Shen, P. (2021). A Novel Ferroptosis-Related 12-gene Signature Predicts Clinical Prognosis and Reveals Immune Relevancy in Clear Cell Renal Cell Carcinoma. *BMC Cancer* 21 (1), 831. doi:10.1186/s12885-021-08559-0

Hwang, I., Kim, J. W., Ylaya, K., Chung, E. J., Kitano, H., Perry, C., et al. (2020). Tumor-associated Macrophage, Angiogenesis and Lymphangiogenesis Markers Predict Prognosis of Non-small Cell Lung Cancer Patients. *J. Transl. Med.* 18 (1), 443. doi:10.1186/s12967-020-02618-z

Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., and Forman, D. (2011). Global Cancer Statistics. *CA A Cancer J. Clin.* 61 (2), 69–90. doi:10.3322/caac.20107

Lavin, Y., Kobayashi, S., Leader, A., Amir, E.-a. D., Elefant, N., Bigenwald, C., et al. (2017). Innate Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell Analyses. *Cell.* 169 (4), 750–765.e717. doi:10.1016/j.cell.2017.04.014

Lei, L. C., Yu, V. Z., Ko, J. M. Y., Ning, L., and Lung, M. L. (2020). FANCD2 Confers a Malignant Phenotype in Esophageal Squamous Cell Carcinoma by Regulating Cell Cycle Progression. *Cancers* 12 (9), 2545. doi:10.3390/cancers12092545

Lemonidis, K., Arkinson, C., Rennie, M. L., and Walden, H. (2021). Mechanism, Specificity, and Function of FANCD2-FANCI Ubiquitination and Deubiquitination. *Febs J.* doi:10.1111/febs.16077

Li, C., Tian, C., Liu, Y., Liang, J., Zeng, Y., Yang, Q., et al. (2021). Comprehensive Profiling Reveals Distinct Microenvironment and Metabolism Characterization of Lung Adenocarcinoma. *Front. Genet.* 12, 619821. doi:10.3389/fgene.2021.619821

Li, X., Liu, J., Wang, K., Zhou, J., Zhang, H., Zhang, M., et al. (2020). Polymorphisms and Rare Variants Identified by Next-Generation Sequencing Confer Risk for Lung Cancer in Han Chinese Population. *Pathology - Res. Pract.* 216 (4), 152873. doi:10.1016/j.prp.2020.152873

Liang, C., Zhang, X., Yang, M., and Dong, X. (2019). Recent Progress in Ferroptosis Inducers for Cancer Therapy. *Adv. Mat.* 31 (51), 1904197. doi:10.1002/adma.201904197

Litwin, T. R., Irvin, S. R., Chornock, R. L., Sahasrabuddhe, V. V., Stanley, M., and Wentzensen, N. (2021). Infiltrating T-Cell Markers in Cervical Carcinogenesis: a Systematic Review and Meta-Analysis. *Br. J. Cancer* 124 (4), 831–841. doi:10.1038/s41416-020-01184-x

Liu, T., Ghosal, G., Yuan, J., Chen, J., and Huang, J. (2010). FAN1 Acts with FANCI-FANCD2 to Promote DNA Interstrand Cross-Link Repair. *Science* 329 (5992), 693–696. doi:10.1126/science.1192656

Liu, X.-S., Lin, X.-K., Mei, Y., Ahmad, S., Yan, C.-X., Jin, H.-L., et al. (2019). Regulatory T Cells Promote Overexpression of Lgr5 on Gastric Cancer Cells via TGF-Beta1 and Confer Poor Prognosis in Gastric Cancer. *Front. Immunol.* 10, 1741. doi:10.3389/fimmu.2019.01741

Ma, L., Zhang, X., Yu, K., Xu, X., Chen, T., Shi, Y., et al. (2021). Targeting SLC3A2 Subunit of System XC− Is Essential for m6A Reader YTHDC2 to Be an Endogenous Ferroptosis Inducer in Lung Adenocarcinoma. *Free Radic. Biol. Med.* 168, 25–43. doi:10.1016/j.freeradbiomed.2021.03.023

Maj, T., Wang, W., Crespo, J., Zhang, H., Wang, W., Wei, S., et al. (2017). Oxidative Stress Controls Regulatory T Cell Apoptosis and Suppressor Activity and PD-L1-Blockade Resistance in Tumor. *Nat. Immunol.* 18 (12), 1332–1341. doi:10.1038/ni.3868

Moldovan, G.-L., and D'Andrea, A. D. (2009). How the Fanconi Anemia Pathway Guards the Genome. *Annu. Rev. Genet.* 43, 223–249. doi:10.1146/annurev-genet-102108-134222

Morvan, M. G., and Lanier, L. L. (2016). NK Cells and Cancer: You Can Teach Innate Cells New Tricks. *Nat. Rev. Cancer* 16 (1), 7–19. doi:10.1038/nrc.2015.5

Nalepa, G., and Clapp, D. W. (2018). Fanconi Anaemia and Cancer: an Intricate Relationship. *Nat. Rev. Cancer* 18 (3), 168–185. doi:10.1038/nrc.2017.116

Nizzoli, G., Larghi, P., Paroni, M., Crosti, M. C., Moro, M., Neddermann, P., et al. (2016). IL-10 Promotes Homeostatic Proliferation of Human CD8+memory T Cells and, when Produced by CD1c+DCs, Shapes Naive CD8+T-Cell Priming. *Eur. J. Immunol.* 46 (7), 1622–1632. doi:10.1002/eji.201546136

Pham, K., Huynh, D., Le, L., Delitto, D., Yang, L., Huang, J., et al. (2020). E-cigarette Promotes Breast Carcinoma Progression and Lung Metastasis: Macrophage-Tumor Cells Crosstalk and the Role of CCL5 and VCAM-1. *Cancer Lett.* 491, 132–145. doi:10.1016/j.canlet.2020.08.010

Salgado, R., and Loi, S. (2018). Tumour Infiltrating Lymphocytes in Breast Cancer: Increasing Clinical Relevance. *Lancet Oncol.* 19 (1), 3–5. doi:10.1016/s1470-2045(17)30905-1

Schalper, K. A., Carleton, M., Zhou, M., Chen, T., Feng, Y., Huang, S.-P., et al. (2020). Elevated Serum Interleukin-8 Is Associated with Enhanced Intratumor Neutrophils and Reduced Clinical Benefit of Immune-Checkpoint Inhibitors. *Nat. Med.* 26 (5), 688–692. doi:10.1038/s41591-020-0856-x

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 (5), 585–590. doi:10.1097/JTO.0b013e3181d60fd7

Siegel, R. L., Miller, K. D., and Jemal, A. (2017). Cancer Statistics, 2017. *CA A Cancer J. Clin.* 67 (1), 7–30. doi:10.3322/caac.21387

Song, X., Xie, Y., Kang, R., Hou, W., Sun, X., Epperly, M. W., et al. (2016). FANCD2 Protects against Bone Marrow Injury from Ferroptosis. *Biochem. Biophysical Res. Commun.* 480 (3), 443–449. doi:10.1016/j.bbrc.2016.10.068

Stockwell, B. R., Friedmann Angeli, J. P., Bayir, H., Bush, A. I., Conrad, M., Dixon, S. J., et al. (2017). Ferroptosis: A Regulated Cell Death Nexus Linking Metabolism, Redox Biology, and Disease. *Cell.* 171 (2), 273–285. doi:10.1016/j.cell.2017.09.021

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

Tsui, V., and Crismani, W. (2019). The Fanconi Anemia Pathway and Fertility. *Trends Genet.* 35 (3), 199–214. doi:10.1016/j.tig.2018.12.007

Wang, G. Z., Liu, Y. Q., Cheng, X., and Zhou, G. B. (2015). Celastrol Induces Proteasomal Degradation of FANCD 2 to Sensitize Lung Cancer Cells to DNA Crosslinking Agents. *Cancer Sci.* 106 (7), 902–908. doi:10.1111/cas.12679

Wang, W., Green, M., Choi, J. E., Gijón, M., Kennedy, P. D., Johnson, J. K., et al. (2019). CD8+ T Cells Regulate Tumour Ferroptosis during Cancer Immunotherapy. *Nature* 569 (7755), 270–274. doi:10.1038/s41586-019-1170-y

Wang, Z., Zhang, X., Tian, X., Yang, Y., Ma, L., Wang, J., et al. (2021). CREB Stimulates GPX4 Transcription to Inhibit Ferroptosis in Lung Adenocarcinoma. *Oncol. Rep.* 45 (6). doi:10.3892/or.2021.8039

Xu, C., Sun, S., Johnson, T., Qi, R., Zhang, S., Zhang, J., et al. (2021). The Glutathione Peroxidase Gpx4 Prevents Lipid Peroxidation and Ferroptosis to Sustain Treg Cell Activation and Suppression of Antitumor Immunity. *Cell. Rep.* 35 (11), 109235. doi:10.1016/j.celrep.2021.109235

Yagi, R., Zhu, J., and Paul, W. E. (2011). An Updated View on Transcription Factor GATA3-Mediated Regulation of Th1 and Th2 Cell Differentiation. *Int. Immunol.* 23 (7), 415–420. doi:10.1093/intimm/dxr029

Yang, L., Li, C., Qin, Y., Zhang, G., Zhao, B., Wang, Z., et al. (2021). A Novel Prognostic Model Based on Ferroptosis-Related Gene Signature for Bladder Cancer. *Front. Oncol.* 11, 686044. doi:10.3389/fonc.2021.686044

Yang, S.-Y., Hsiung, C.-N., Li, Y.-J., Chang, G.-C., Tsai, Y.-H., Chen, K.-Y., et al. (2016). Fanconi Anemia Genes in Lung Adenocarcinoma- a Pathway-wide Study on Cancer Susceptibility. *J. Biomed. Sci.* 23, 23. doi:10.1186/s12929-016-0240-9

Zanganeh, S., Hutter, G., Spitler, R., Lenkov, O., Mahmoudi, M., Shaw, A., et al. (2016). Iron Oxide Nanoparticles Inhibit Tumour Growth by Inducing Pro-inflammatory Macrophage Polarization in Tumour Tissues. *Nat. Nanotech* 11 (11), 986–994. doi:10.1038/nnano.2016.168

Zhang, F., Li, F., Lu, G.-H., Nie, W., Zhang, L., Lv, Y., et al. (2019). Engineering Magnetosomes for Ferroptosis/Immunomodulation Synergism in Cancer. *ACS Nano* 13 (5), 5662–5673. doi:10.1021/acsnano.9b00892

Zhang, H., Li, R., Cao, Y., Gu, Y., Lin, C., Liu, X., et al. (2020). Poor Clinical Outcomes and Immunoevasive Contexture in Intratumoral IL-10-Producing Macrophages Enriched Gastric Cancer Patients. *Ann. Surg.* 275, e626–e635. doi:10.1097/sla.0000000000004037

Zhang, X., Yu, K., Ma, L., Qian, Z., Tian, X., Miao, Y., et al. (2021). Endogenous Glutamate Determines Ferroptosis Sensitivity via ADCY10-dependent YAP Suppression in Lung Adenocarcinoma. *Theranostics* 11 (12), 5650–5674. doi:10.7150/thno.55482

Zhang, Y., Chandra, V., Riquelme Sanchez, E., Dutta, P., Quesada, P. R., Rakoski, A., et al. (2020). Interleukin-17-induced Neutrophil Extracellular Traps Mediate Resistance to Checkpoint Blockade in Pancreatic Cancer. *J. Exp. Med.* 217 (12). doi:10.1084/jem.20190354

Zhang, Y., Li, S., Li, F., Lv, C., and Yang, Q.-k. (2021). High-fat Diet Impairs Ferroptosis and Promotes Cancer Invasiveness via Downregulating Tumor Suppressor ACSL4 in Lung Adenocarcinoma. *Biol. Direct* 16 (1), 10. doi:10.1186/s13062-021-00294-7

Zhang, D.Chen, X., Zhu, D., Qin, C., Dong, J., Qiu, X., et al. (2019). Intrapulmonary Lymph Node Metastasis Is Common in Clinically Staged IA Adenocarcinoma of the Lung. *Thorac. Cancer* 10 (2), 123–127. doi:10.1111/1759-7714.12908

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: 31 May 2022
doi: 10.3389/fgene.2022.893511

A Ubiquitin-Proteasome Gene Signature for Predicting Prognosis in Patients With Lung Adenocarcinoma

Yunliang Tang[1] *and Yinhong Guo*[2]*

[1]Department of Rehabilitation Medicine, First Affiliated Hospital of Nanchang University, Nanchang, China, [2]Department of Oncology, Zhuji People's Hospital of Zhejiang Province, Zhuji, China

OPEN ACCESS

Edited by:
Mehdi Pirooznia,
Johnson & Johnson, United States

Reviewed by:
Piotr Bragoszewski,
Nencki Institute of Experimental Biology (PAS), Poland
Panagiotis J. Vlachostergios,
Cornell University, United States
Jerry Vriend,
University of Manitoba, Canada

***Correspondence:**
Yinhong Guo
guoyinhongzj@outlook.com

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: 18 March 2022
Accepted: 05 May 2022
Published: 31 May 2022

Citation:
Tang Y and Guo Y (2022) A Ubiquitin-Proteasome Gene Signature for Predicting Prognosis in Patients With Lung Adenocarcinoma. Front. Genet. 13:893511. doi: 10.3389/fgene.2022.893511

Background: Dysregulation of the ubiquitin-proteasome system (UPS) can lead to instability in the cell cycle and may act as a crucial factor in both tumorigenesis and tumor progression. However, there is no established prognostic signature based on UPS genes (UPSGs) for lung adenocarcinoma (LUAD) despite their value in other cancers.

Methods: We retrospectively evaluated a total of 703 LUAD patients through multivariate Cox and Lasso regression analyses from two datasets, the Cancer Genome Atlas ($n = 477$) and GSE31210 ($n = 226$). An independent dataset (GSE50081) containing 128 LUAD samples were used for validation.

Results: An eight-UPSG signature, including *ARIH2*, *FBXO9*, *KRT8*, *MYLIP*, *PSMD2*, *RNF180*, *TRIM28*, and *UBE2V2*, was established. Kaplan-Meier survival analysis and time-receiver operating characteristic curves for the training and validation datasets revealed that this risk signature presented with good performance in predicting overall and relapsed-free survival. Based on the signature and its associated clinical features, a nomogram and corresponding web-based calculator for predicting survival were established. Calibration plot and decision curve analyses showed that this model was clinically useful for both the training and validation datasets. Finally, a web-based calculator (https://ostool.shinyapps.io/lungcancer) was built to facilitate convenient clinical application of the signature.

Conclusion: An UPSG based model was developed and validated in this study, which may be useful as a novel prognostic predictor for LUAD.

Keywords: lung adenocarcinoma, ubiquitin-proteasome, prognosis, predictive modeling, gene signature

Abbreviations: LUAD, lung adenocarcinoma; TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus; OS, overall survival; RFS, relapsed-free survival; UPS, ubiquitin-proteasome system; UPSG, ubiquitin-proteasome system related gene; AUC, area under the curve; ROC, receiver operating characteristic; DCA, decision curve analysis.

INTRODUCTION

Non-small cell lung cancer (NSCLC) constitutes approximately 40%–50% of total lung cancers, which are the leading cause of tumor-associated mortality worldwide (Wu et al., 2019; Siegel et al., 2020). NSCLC is still difficult to diagnose in the early stages due to its insidious progression, and more than two thirds of patients are only diagnosed in its advanced stages. Despite the existence of various chemo/physical/immunological therapies, the 5-year survival rate for patients with any type of lung cancer remains very low (Bade and Dela Cruz, 2020). Histologically, lung adenocarcinoma (LUAD) is the most common form of NSCLC, and comprises nearly 40% of all cases (Hua et al., 2020). Therefore, it is necessary to explore the specific molecular pathogenesis of LUAD and develop novel therapeutic targets for its treatment (Wang J. et al., 2020; Tang et al., 2020).

The ubiquitin-proteasome system (UPS) is a major intracellular protein degradation pathway in eukaryotic organisms, which controls a wide range of physiological processes and disease conditions, such as transcription, translation, DNA repair, and the cell cycle (Chen and Chen, 2016). The UPS primarily comprises the E1, E2, and E3 ubiquitin ligases, the proteasome, and various deubiquitinating enzymes. Dysregulation of the UPS can contribute to the inhibition of cell cycle regulation and disorder of cancer cell metabolism, ultimately leading to carcinogenesis (Dang et al., 2021). In fact, the UPS is already an important target in various cancer treatments, including the use of proteasome and ubiquitin E3 ligases as therapeutic targets for various chemotherapies (Manasanch and Orlowski, 2017). Abnormal expression of different UPS gene families alters proteolysis, inhibiting the proliferation and metastasis of LUAD (He et al., 2021). Thus, evaluating the relationship between UPS and tumorigenesis might provide novel insights into LUAD pathology and facilitate better patient prognosis. However, the prognostic values of UPS-related genes (UPSGs) and their molecular function in LUAD remain poorly understood.

Here, we investigated changes in UPSGs expression in the data downloaded from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases and then built a scoring model by classifying LUAD patients based on a multi-UPSGs signature, in combination with other clinicopathological factors, to improve our ability to predict the prognosis of LUAD patients, thereby helping to guide clinical treatment. This model might be meaningful for the development of comprehensive therapeutic approaches for LUAD patients.

METHODS

Data Collection

The transcriptomic (HTSeq-FPKM), demographic, and clinical information of patients with LUAD was collected from TCGA (training set) and GEO databases (validation set) for evaluations. First, TCGA-LUAD dataset was curated to remove any cases with incomplete survival data or a follow-up time of less than <30 days, leaving a total of 477 patients with LUAD who were then included in this study. GSE50081 were applied as validation datasets (Okayama et al., 2012; Der et al., 2014), where GSE31210 comprised a total of 226 primary LUAD patients at stage I–II, with a median age of 61 years and an age range of 30–76 years; the cohort included 105 male and 121 female patients. GSE50081 comprised a total of 128 LUAD cases with a median age of 70.38 years and an age range of 40.16–85.91 years; this cohort consisted of 98 male and 83 female patients. The clinical characteristics of all three datasets analyzed in this study are summarized in **Table 1**.

The mutation profiles for each of the UPSGs identified in LUAD patients were downloaded from the cBioPortal database (https://www.cbioportal.org/) (Gao et al., 2013). A total of 804 UPSGs were identified in a previous study and used as the basis of our evaluations in this study (**Supplementary Table S1**) (Wang et al., 2021).

Establishing a Prognostic UPSG Signature

To narrow the number of candidate UPSGs, we conducted univariate Cox regression analysis on TCGA-LUAD and

TABLE 1 | Clinical information analyzed in this study.

Features	TCGA-LUAD	GSE31210	GSE50081
Sample	477	226	128
Mean age (years; range)	66 (33–88)	61 (30–76)	70.38 (40.16–85.91)
Gender			
Male	215	105	65
Female	254	121	63
Stage			
I	253	168	92
II	113	58	36
III	78	-	-
IV	25	-	-
Status			
Alive	314	191	76
Dead	115	35	52
Platform	Illumina HiSeqV2	HG-U133_Plus_2	HG-U133_Plus_2

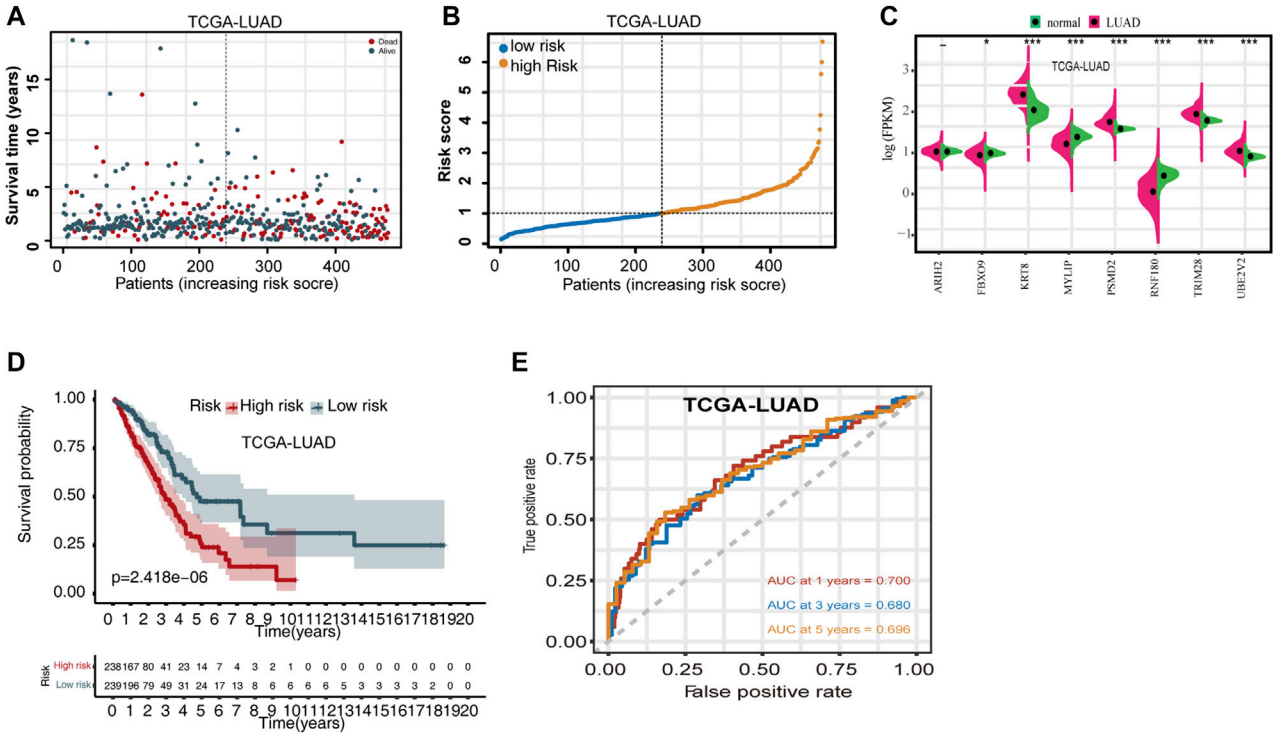


FIGURE 1 | Evaluating the predictive power of the UPSG signature in TCGA-LUAD training set. **(A)** Distribution of overall survival status ranked by risk score. **(B)** Risk score distribution across OS prediction. **(C)** Comparison of the eight-UPSG mRNA expression levels between normal and tumor tissues. **(D)** Kaplan-Meier survival analysis of patients in the high- and low-risk cohorts. **(E)** Time-dependent receiver operating characteristic curves of the signature for predicting 1-, 3-, and 5-year overall survival rates. $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$ represent difference between normal and tumor tissues.

GSE31210 datasets. The overlapping overall survival (OS)-related genes were selected for further study. These targets were then cross-validated using least absolute shrinkage and selection operator (LASSO) regression, which then produced a list of potential predictors with nonzero coefficients using the R packages "glmnet" and "survival." Finally, we performed multivariate Cox regression analysis on TCGA-LUAD dataset to confirm the identity of highly correlated genes and construct the OS gene signature using the following risk score model:

$$\text{Risk score} = \sum_{i=0}^{N} (\beta_i \times Exp_i),$$

where N represents the number of UPSGs included in the signature, Expi represents the mRNA level of these genes, and βi represents the regression coefficient obtained using the Cox regression analyses.

The patients in each dataset were classified as high- or low-risk, using the median risk score as the cutoff value, and the receiver operating characteristic (ROC) curves were created using the "survivalROC" package in R. The AUC values were calculated to assess the predictive potential of the UPSG signature.

TABLE 2 | The prognostic ubiquitin-proteasome system-related genes identified by using LASSO COX regression.

Symbol	Coefficient	Name
ARIH2	−0.075232	Ariadne RBR E3 ubiquitin protein ligase 2
FBXO9	−0.055219	F-box protein 9
KRT8	0.000590	Keratin 8
MYLIP	−0.034945	Myosin regulatory light chain interacting protein
PSMD2	0.006608	Proteasome 26S subunit, non-ATPase 2
RNF180	−0.123541	Ring finger protein 180
TRIM28	0.004102	Tripartite motif containing 28
UBE2V2	0.031981	Ubiquitin conjugating enzyme E2 V2

Validation of the Prognostic Signature

The prognostic UPSG model was then applied to the GSE50081 validation dataset. Validation dataset was split into high- and low-risk groups for evaluation using Kaplan-Meier curve analysis, time-dependent ROC analysis, and patient outcome distribution to assess the categorization of the UPSG signature.

Subgroup Analyses of the UPSG Signature

The clinical usefulness of the prognostic UPSG signature was evaluated through stratification analysis that was performed to investigate the differences in the prognosis of LUAD patients presenting with different clinical characteristics. Based on both characteristics and risk score, the LUAD cases were divided into low- and high-risk groups, and Cox regression was applied to analyze these subgroups.

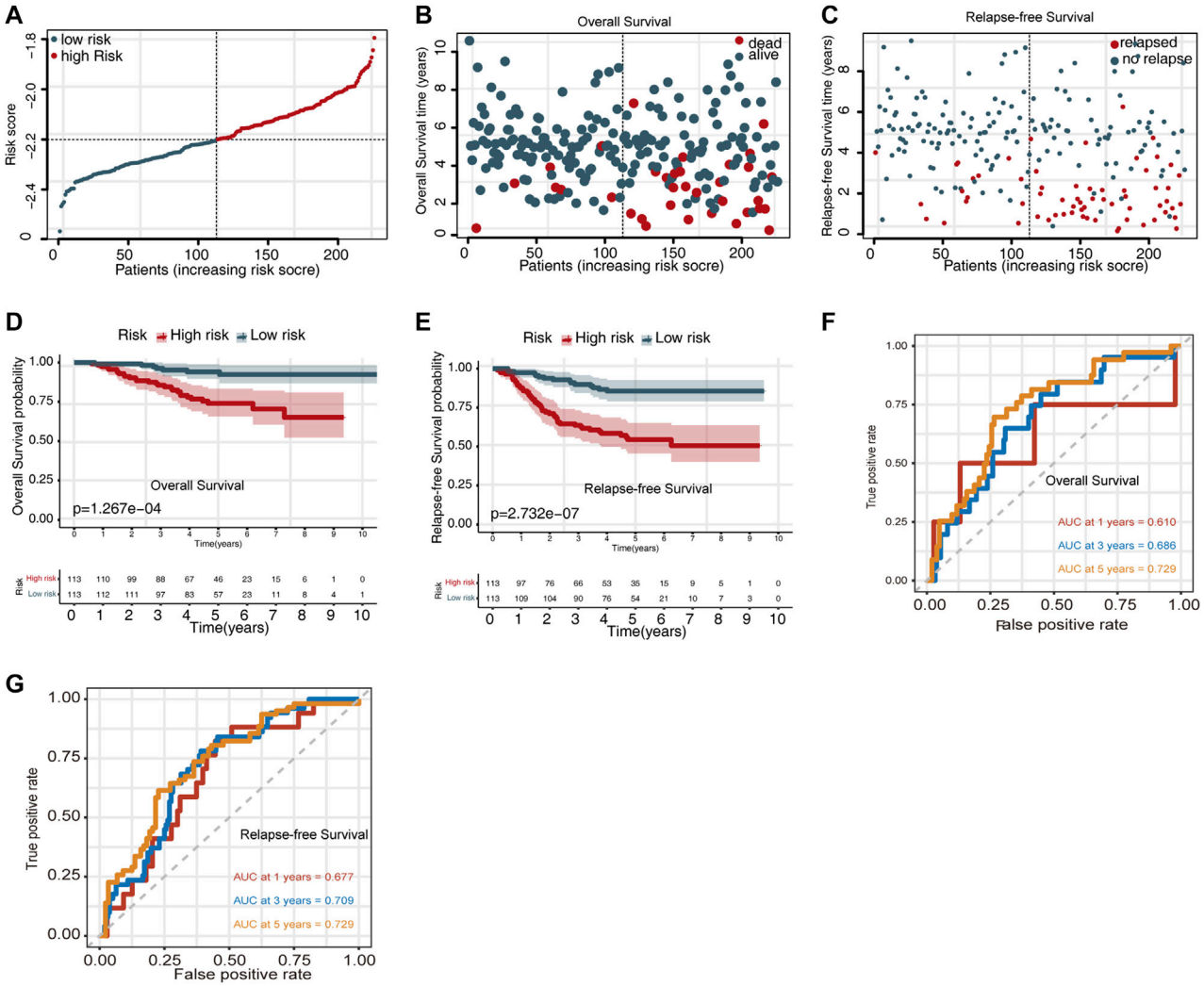


FIGURE 2 | Evaluating the predictive power of the UPSG signature in GSE31210 dataset. **(A)** Risk score distribution for survival prediction. **(B)** Distribution of the overall survival status ranked by risk score. **(C)** Distribution of the relapse-free survival status ranked by risk score. **(D)** Overall survival analysis of patients in the high- and low-risk groups. **(E)** Relapse-free survival analysis of patients in the high- and low-risk groups. **(F)** Time-dependent receiver operating characteristic curves of the signature for predicting 1-, 3-, and 5-year overall survival. **(G)** Time-dependent receiver operating characteristic curves of the signature for predicting 1-, 3-, and 5-year relapse-free survival.

Estimation of Immune Cell Infiltration

To determine whether and how UPSGs affected the tumor immune microenvironment, Cell type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT) was used to predict the fractions of immune cell types between the high- and low-risk cohorts. Eventually, 22 types of differentially expressed immune cells associated with specific clinicopathologic characteristics in the high- and low-risk cohorts were identified and depicted as a landscape map.

Construction and Evaluation of the Nomogram

The UPSG signature and clinicopathological predictors were combined to construct a nomogram for TCGA training set. A calibration curve was prepared to assess the agreement between the nomogram model and actual observation. The time-ROC curves and calibration plots were generated, and decision curve analyses (DCA) was used to assess the predictive accuracy of the signature.

RESULTS

Establishing a UPSG Signature

We first defined OS-related UPSGs through a univariate Cox regression analysis from TCGA-LUAD and GSE31210 datasets. The two datasets comprised 95 and 157 OS-related UPSGs, respectively. Then, we reduced the number of OS-related UPSGs by taking the intersection of TCGA-LUAD and GSE31210 datasets. Eventually, 33 overlapping OS-related

TABLE 3 | Stratified survival analyses and clinical characteristics with UPSGs prognostic signature in the TCGA-LUAD cohort (n = 477).

Characteristics	No.		%	Overall survival	
	High-risk	High-risk		HR (95% CI)	*p* Value
Age (years)					
<65	109	103	44.4	2.058 (1.235–3.429)	0.006
≥65	129	136	55.6	2.241 (1.448–3.469)	0.000
Sex					
Male	116	104	46.1	2.828 (1.711–4.675)	0.000
Female	122	135	53.9	1.686 (1.076–2.644)	0.023
Stage					
I	106	147	53.0	1.695 (0.973–2.950)	0.062
II	68	45	23.7	1.594 (0.847–2.998)	0.148
III	46	32	16.4	2.385 (1.218–4.673)	0.011
IV	16	9	5.2	2.683 (0.744–9.673)	0.132
Unknow	2	6	1.7	--	-
T stage					
T1	65	94	33.3	1.578 (0.826–3.016)	0.167
T2	137	117	53.2	2.085 (1.345–3.230)	0.001
T3	27	16	9.0	5.539 (1.257–24.408)	0.024
T4	8	10	3.8	2.823 (0.690–11.549)	0.149
Unknow	1	2	0.6	-	-
M stage					
M0	156	157	65.6	2.181 (1.447–3.286)	0.000
M1	16	8	5.0	3.726 (0.821–16.902)	0.088
Unknow	66	74	29.4	-	-
N stage					
N0	133	174	64.4	1.984 (1.228–3.205)	0.005
N1	59	31	18.9	1.556 (0.816–2.965)	0.179
N2	40	27	14.0	2.133 (1.044–4.358)	0.038
N3	1	1	0.4	-	-
Unknow	5	6	2.3	-	-

UPSGs were identified (**Supplementary Figure S1**). Subsequently, the 33 UPSGs were used in a Lasso-Cox proportional hazards regression and ten-fold cross-validation analyses designed to construct the best gene signature; eventually, 13 UPSGs were identified for downstream analysis (**Supplementary Figure S1**). Furthermore, a multivariate Cox regression was used to evaluate the Lasso results (**Supplementary Figure S1**). The prognostic risk score according to the expression of the eight UPSGs was determined as follows: Risk score = ARIH2 × (−0.075232) + FBXO9 × (−0.055219) + KRT8 × 0.00059 + MYLIP × (−0.034945) + PSMD2 × 0.006608 + RNF180 × (−0.123541) + TRIM28 × 0.004102 + UBE2V2 × 0.031981 (**Table 2**).

Moreover, mutation analysis of the eight UPSGs included in the prognostic signature revealed that *PSMD2*, *RNF180*, *UBE2V2*, *KRT8*, and *MYLIP* were the most frequently mutated genes. Notably, amplification was the most common type of mutation, and *UBE2V2*, *PSMD2*, and *FBXO9* were frequently amplified in LUAD (**Supplementary Figure S1**).

Evaluating the Prognostic Significance of the UPSG Signature

Using the risk score calculation formula derived from the multivariate Cox regression, the risk score for each TCGA-LUAD patient was calculated. Based on the median risk score, patients were separated into high- and low-risk cohorts. The distribution of OS status ranked by risk score is presented in **Figure 1A**, and the risk score distribution for OS prediction in TCGA dataset is shown in **Figure 1B**. Differences in expression of the eight UPSGs between normal and tumor tissues is shown in **Figure 1C**. The results show that *KRT8*, *PSMD2*, *TRIM28*, and *UBE2V2* are significantly overexpressed in tumor tissues, while the expression level of *FBXO9, MYLIP*, and *RNF180* is significantly reduced in LUAD ($p < 0.05$).

We then evaluated the differences in the survival rates of the high- and low-risk groups. The results revealed that the high-risk group exhibited a significantly poorer OS rate than the low-risk group (**Figure 1D**). In addition, we then used time-dependent ROC curves and their AUC values to evaluate the prognostic significance of this stratification. AUC values were estimated to be 0.700, 0.680, and 0.696 for 1-, 3-, and 5-year OS, respectively (**Figure 1E**). Moreover, subset analyses using stratification *via* clinicopathologic feature revealed that this UPSG signature exhibited significant predictive value in most subgroups (**Table 3**).

Moreover, we also evaluated the prognostic significance of the UPSG signature in the GSE31210 dataset. The distribution of the risk scores for the GSE31210 dataset is presented in **Figure 2A**. The risk score distribution for OS prediction is shown in **Figure 2B**, and the risk score distribution for relapsed-free survival (RFS) prediction is shown in **Figure 2C**. Survival analysis indicated that patients in the high-risk group had worse OS and RFS compared with the low-risk group (**Figures 2D,E**). Using the time-dependent ROC curves, the AUC values for 1-, 3-, and 5-year OS prediction were estimated to be 0.610,

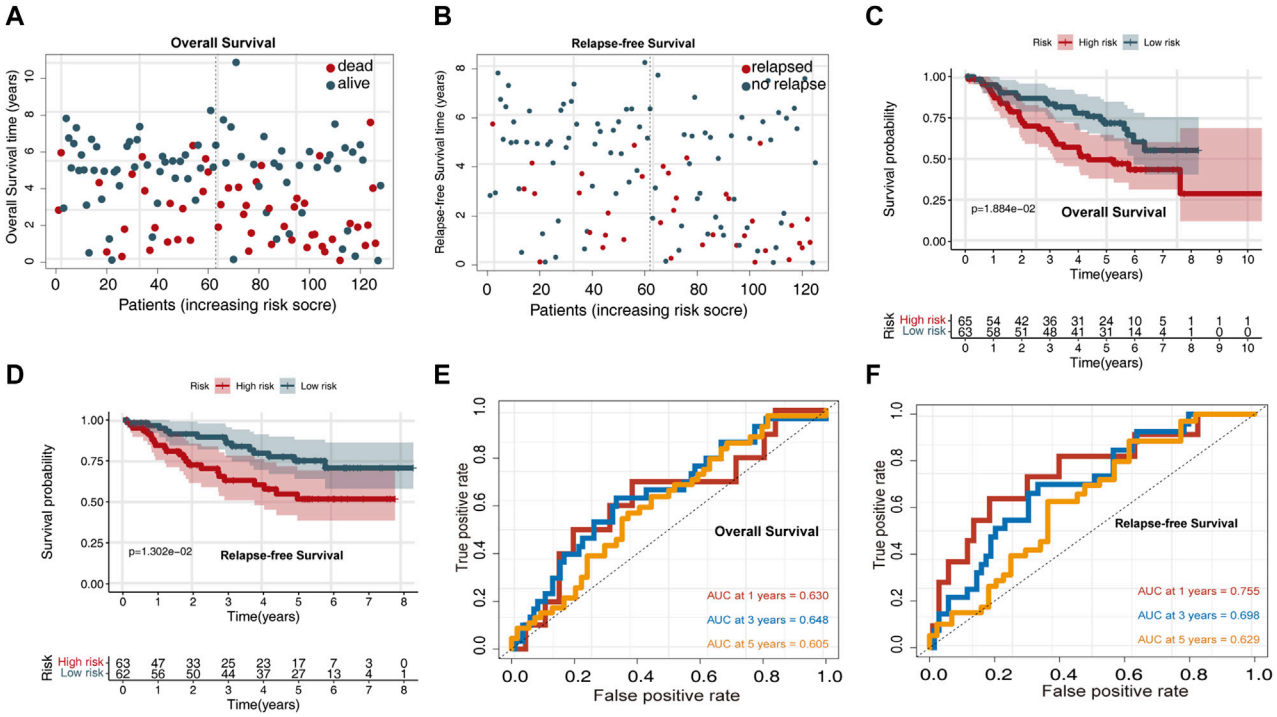


FIGURE 3 | Validating the predictive power of the UPSG signature in the GSE50081 validation set. **(A)** Distribution of overall survival status ranked by risk score. **(B)** Distribution of relapse-free survival status ranked by risk score. **(C)** Overall survival analysis of patients in the high- and low-risk groups. **(D)** Relapse-free survival analysis of patients in the high- and low-risk groups. **(E)** Time-dependent receiver operating characteristic curves of the signature for predicting 1-, 3-, and 5-year overall survival. **(F)** Time-dependent receiver operating characteristic curves of the signature for predicting 1-, 3-, and 5-year relapse-free survival.

0.686, and 0.729, respectively (**Figure 2F**), while those for RFS prediction were 0.677, 0.709, and 0.729, respectively (**Figure 2G**).

Validating the Prognostic Value of the UPSG Signature

To validate the prognostic value of the UPSG signature, we verified its prediction performance in the GSE50081 datasets.

The risk score distribution for OS prediction in the GSE50081 dataset is displayed in **Figure 3A**, and the risk score distribution for RFS prediction is shown in **Figure 4B**. Survival analysis indicated that the high-risk group exhibited a worse OS and RFS than the low-risk group (**Figures 3C,D**). Time-dependent ROC curves for OS prediction revealed AUC values for 1-, 3-, and 5-year OS of 0.63, 0.648, and 0.605, respectively (**Figure 3E**). Time-dependent ROC curves for RFS prediction revealed AUC values for 1-, 3-, and 5-year RFS of 0.755, 0.698, and 0.629, respectively (**Figure 3F**).

We then used both univariate and multivariate Cox regression analyses to determine whether this signature and other clinicopathological features were independent predictors of LUAD survival. The results indicated that the UPSG signature may serve as an independent prognostic indicator in both TCGA-LUAD (**Figures 4A,B**), GSE31210 (**Figures 4C,D**) datasets, and GSE50081 (**Figures 4E,F**) dataset.

Correlation Between UPSGs and Clinicopathological Factors

We analyzed the relevance of the signature genes in relation to various clinicopathological parameters of LUAD using the data from TCGA-LUAD training set. These analyses revealed that *TRIM28*, *UBE2V2*, *KRT8*, *MYLIP*, *FBXO9*, *ARIH2*, and *RNF180* were all significantly correlated with specific clinicopathological features (**Figures 5A–P**).

Evaluating the Infiltrating Immune Cell Ratios in High- and Low-Risk LUAD Patients

TCGA-LUAD gene expression data were uploaded to the CIBERSORT platform to evaluate any changes in the overall proportions of the 22 immune cell types in response to their allocation into high- and low-risk groups. Compared to low-risk patients, high-risk patients exhibited an increased proportion of activated CD4 memory T cells, resting NK cells, M0 macrophages, M1 macrophages, and activated mast cells and a reduced proportion of naïve B cells, plasma cells, resting CD4

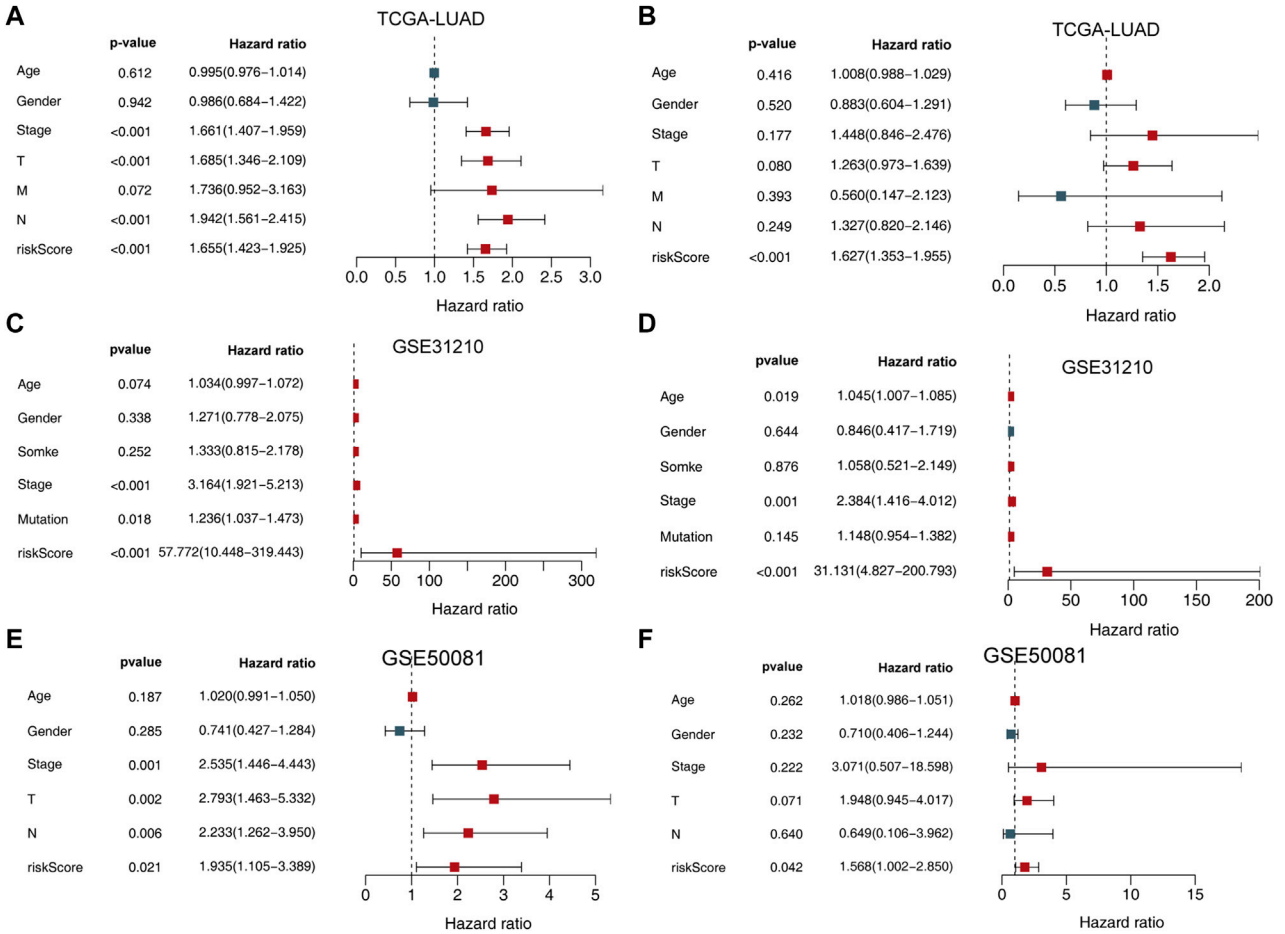


FIGURE 4 | Univariate and multivariate analysis of the influence of specific clinical characteristics in the outcome of LUAD patients. **(A)** Univariate analysis for TCGA-LUAD training set. **(B)** multivariate analysis of TCGA-LUAD training set. **(C)** Univariate analysis for the GSE31210 dataset. **(D)** Multivariate analysis for the GSE31210 dataset. **(E)** Univariate analysis for the GSE50081 validation set. **(F)** Multivariate analysis for the GSE50081 validation set.

memory T cells, activated NK cells, resting dendritic cells, and resting mast cells (**Figure 6**).

Establishing and Evaluating a Novel LUAD Nomogram

Given these results, we established a novel nomogram for accurate clinical prediction of LUAD outcomes, using age, gender, stage, and signature risk score in TCGA-LUAD (**Figure 7A**), and then evaluated its precision using a calibration plot to predict the 1-, 3-, and 5-year OS rates for these samples, all of which exhibited strong correlation with the nomogram findings (**Figure 7B**). The AUC values for the nomogram at 1-, 3-, and 5-year OS were 0.776, 0.758, and 0.762, respectively (**Figure 7C**). Furthermore, results of DCA for predicting 1-, 3-, and 5-year survival probability confirmed our findings. As shown in **Figures 7D–F**, the nomogram presented higher net benefit than stage in predicting OS.

Then, we evaluated this nomogram in the GSE31210 dataset. The AUC values of the nomogram at 1-, 3-, and 5-year OS in the ROC curves were 0.851, 0.792, and 0.735, respectively (**Figure 8A**). The AUC values of the nomogram at 1-, 3-, and 5-year RFS in the ROC curves were 0.769, 0.761, and 0.764, respectively (**Figure 8B**). Furthermore, DCA for predicting 1-, 3-, and 5-year OS and RFS probability confirmed these results. As shown in **Figures 8C,D**.

Moreover, we also evaluated this nomogram in the GSE50081 dataset. The AUC values of the nomogram at 1-, 3-, and 5-year OS in the ROC curves were 0.652, 0.693, and 0.676, respectively (**Supplementary Figure S1A**). The AUC values of the nomogram at 1-, 3-, and 5-year RFS in the ROC curves were 0.728, 0.688, and 0.703, respectively (**Supplementary Figure S1B**). Furthermore, DCA for predicting 1-, 3-, and 5-year OS and RFS probability confirmed these results. As shown in **Supplementary Figures S1C,D** the nomogram based on the best model exhibited higher net benefit than stage during OS and RFS prediction.

Finally, we established a web-based tool (https://ostool.shinyapps.io/lungcancer/) for estimating the OS of patients with LUAD (**Figures 9A,B**) for clinical application of our nomogram and gene signature.

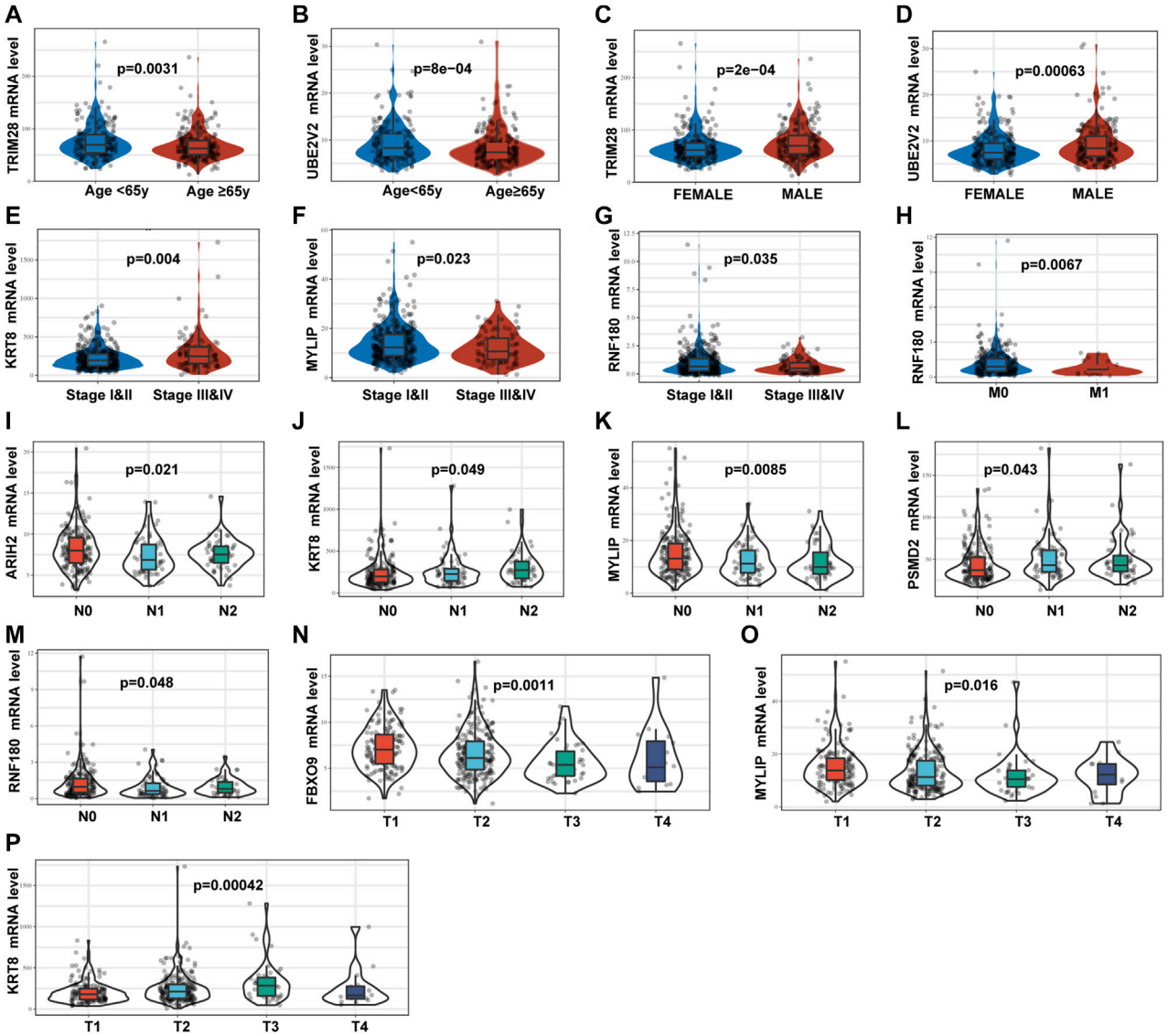


FIGURE 5 | Relationships between the mRNA expression of identified UPSGs and clinicopathological factors in patients with LUAD based on the TCGA-LUAD cohort. **(A)** TRIM28 expression and age. **(B)** UBE2V2 expression and age. **(C)** TRIM28 expression and gender. **(D)** UBE2V2 expression and gender. **(E)** KRT8 expression and tumor stage. **(F)** MYLIP expression and tumor stage. **(G)**RNF180 expression and tumor stage. **(H)**RNF180 expression and M stage. **(I)** ARIH2 expression and N stage. **(J)** KRT8 expression and N stage. **(K)** MYLIP expression and N stage. **(L)** PSMD2 expression and N stage. **(M)** RNF180 expression and N stage. **(N)** FBXO9 expression and T stage. **(O)** MYLIP expression and T stage. **(P)** KRT8 expression and T stage.

For convenient clinical use and visualization of the prognostic model, we developed a simple-to-operate web-based model (https://ostool.shinyapps.io/lungcancer/) to predict the prognosis of LUAD based on the established nomogram (**Figures 9A,B**). Estimated LUAD survival probabilities can be obtained by drawing a perpendicular line from the total point axis to the outcome axis.

DISCUSSION

Our understanding of the ubiquitin-proteasome system has opened up a new era of anticancer treatment through the maintenance and protection of protein quality control and homeostasis (Li X. et al., 2018; Narayanan et al., 2020). Various UPS inhibitors are being evaluated in clinical trials as novel cancer therapeutics (Johnson, 2015). However, the prognostic value of UPSGs in LUAD patients remains poorly understood. Therefore, complete comprehensive analysis of UPSGs is urgently required to explore their clinical significance in LUAD.

Here, we used RNA-seq data from both TCGA and GEO databases to establish a UPSG prognostic risk signature incorporating *ARIH2*, *FBXO9*, *KRT8*, *MYLIP*, *PSMD2*, *RNF180*, *TRIM28*, and *UBE2V2* for LUAD patients. Kaplan-

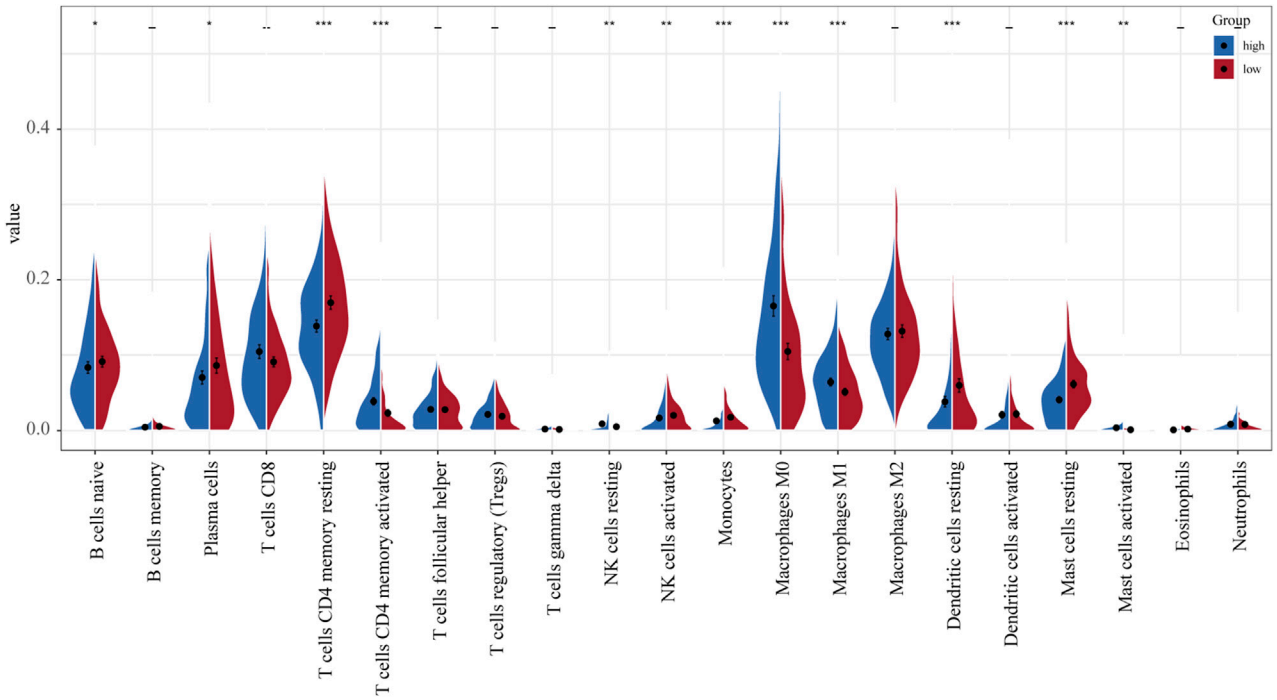


FIGURE 6 | Differences in the distribution of immune cell infiltration between normal and tumor tissues (relative proportion of immune cells in each group).

Meier survival and ROC curve analyses based on both the training and validation sets indicated that this risk signature exhibited excellent fitting and predictive capacity. Thus, these results could help in the development of novel prognostic biomarkers and provide clinical research ideas for LUAD.

Among the eight UPSGs included in the signature, several genes have previously been explored in terms of their association with LUAD. ARIH2 (Ariadne RBR E3 ubiquitin ligase 2), an E3 ligase whose expression levels are regulated by PABPN1, is one of the most important factors involved in regulating biological function (Raz et al., 2014). A previous study has reported that loss of *ARIH2* significantly accelerates the development of resistant HCC827 cells, but the specific mechanism of action needs further investigation (Zeng et al., 2019). FBXO9 (F-box only protein), an SCF (Skp1-Cul-F-box)-type ubiquitin ligase for Neurog2 (Liu J. A. et al., 2020), has been studied in various malignancies. Liu et al. reported that inhibition of FBXO9 increased proteasome activity and sensitivity to bortezomib, suggesting that FBXO9 has an important role in bortezomib susceptibility (Hynes-Smith et al., 2019). However, FBXO9 does not have E3 ligase activity by itself and it is a substrate recognition subunit of an E3 ligase complex. KRT8 (keratin 8), a type II basic intermediate filament protein, is essential for the development and metastasis of various cancers, including LUAD (Xie et al., 2019). Wang et al. explored the expression of KRT8 in LUAD and found that it was overexpressed in tumor tissues and associated with poorer prognosis (Wang W. et al., 2020); these results are consistent with those of our study. MYLIP (myosin regulatory light chain interacting protein), an E3 ubiquitin ligase, which is associated with the regulation of cell motility and migration (Chen et al., 2017), has been found to play a critical role in malignant tumors (Zhao et al., 2020). Ni et al. reported that MYLIP regulates the growth and metastasis-related phenotypes of cervical cancer cells Ni et al. (2021), while the knockdown of *PSMD2* (proteasome 26S subunits, non-ATPase 2) has been reported to suppress tumor cell proliferation (Matsuyama et al., 2011; Li Y. et al., 2018; Okumura et al., 2018). PSMD2 has also been shown to be associated with the acquisition of metastatic phenotypes and poor prognosis in various lung cancers (Matsuyama et al., 2011). RNF180 (ring finger protein 180), an important member of the E3 ubiquitin ligase family, act as a tumor suppressor by inhibiting the proliferation, invasion, and migration of gastric cancer cells (Wu et al., 2020). Previous studies have also indicated that reduced RNF180 expression levels were associated with poor biological behavior in A549 and HCC827 cells (Liu H. et al., 2020). TRIM28 (tripartite motif-containing protein 28) is involved in several cellular processes including gene transcription, cell growth and differentiation, genome stability, immunity, development, and carcinogenesis (Hatakeyama, 2011; Iyengar and Farnham, 2011; Messerschmidt et al., 2012; Sio et al., 2012). Chen et al. reported that TRIM28 is a tumor suppressor in the early transformation process of lung cancer but acts as an oncogene in advanced stages of this disease Chen et al. (2014). UBE2V2 (ubiquitin-conjugating enzyme E2 variant 2) has sequence similarity to other ubiquitin-conjugating enzymes but lacks the conserved cysteine residues that are essential for the catalytic activity of E2 (Dikshit et al., 2018; Zhao et al., 2018). Hua et al.

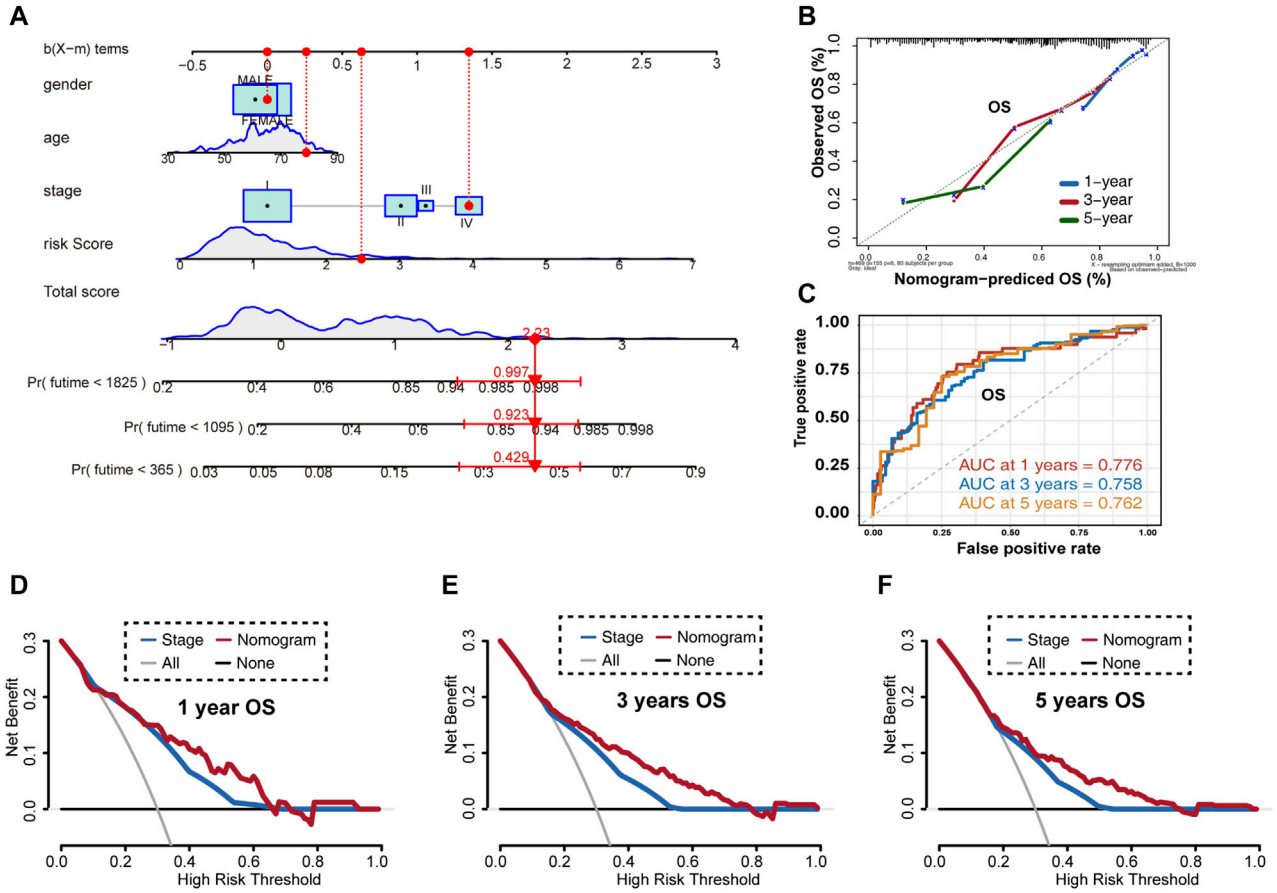


FIGURE 7 | Overall survival prediction using the LUAD nomogram based on TCGA-LUAD training set. **(A)** Nomogram predictions for 1-, 3-, and 5-year overall survival in LUAD patients. **(B)** Calibration plot of the predictive model. **(C)** Time-dependent receiver operating characteristic curves for nomogram-mediated prediction of 1-, 3-, and 5-year overall survival in LUAD patients. **(D)** Decision curve analysis for 1-year overall survival rates. **(E)** Decision curve analysis for 3-year overall survival rates. **(F)** Decision curve analysis for 5-year overall survival rates. OS, overall survival.

reported that UBE2V2 was identified as an independent prognostic indicator for LUAD and might function as a therapeutic target Hua et al. (2021).

The UPS is also regarded as an important player in the regulation of immune cell function and immune response (Çetin et al., 2021). Here, we evaluated the differences in the proportions of 22 immune cell types between low- and high-risk LUAD groups. Our results showed that compared to low-risk patients, high-risk patients have a higher proportion of activated CD4 memory T cells, resting NK cells, M0 macrophages, M1 macrophages, and activated mast cells and a reduced proportion of naïve B cells, plasma cells, resting CD4 memory T cells, activated NK cells, resting dendritic cells, and resting mast cells. These data indicate that this novel gene signature represents differences in immune infiltration, suggesting that immune status may have a critical effect on prognosis.

Given these promising results, we constructed a more reliable and individualized clinical prediction model, using our eight-UPSG signature and several other clinical features to establish a nomogram. The AUC values of the nomogram for TCGA-LUAD data at 1-, 3-, and 5-year OS were 0.776, 0.758, and 0.762, respectively, while those for the GSE31210 dataset were 0.851, 0.792, and 0.735, respectively (**Figure 8A**). In addition, the AUC values of the nomogram at 1-, 3-, and 5-year RFS in this dataset were 0.769, 0.761, and 0.764, respectively (**Figure 8B**). DCA results also suggested that the nomogram based on the best model exhibited a larger net prognostic benefit than stage. For the convenience of clinical application, we established a web-based tool for estimating the OS of patients with LUAD. Taken together, these results may enable us to deliver higher accuracy in prognosis prediction for patients with LUAD.

To the best of our knowledge, this is the first study to establish a UPSG signature for predicting the survival of patients with LUAD. However, this study had some limitations. First, we did not investigate the detailed mechanisms of each of the eight UPSGs identified in this study; these underlying mechanisms warrant further investigation through both *in vitro* and *in vivo* studies. Second, the inclusion of clinical indicators in the model seems to be insufficient; other factors such as, radiotherapy, surgery type, and targeted drug use might affect the prognosis

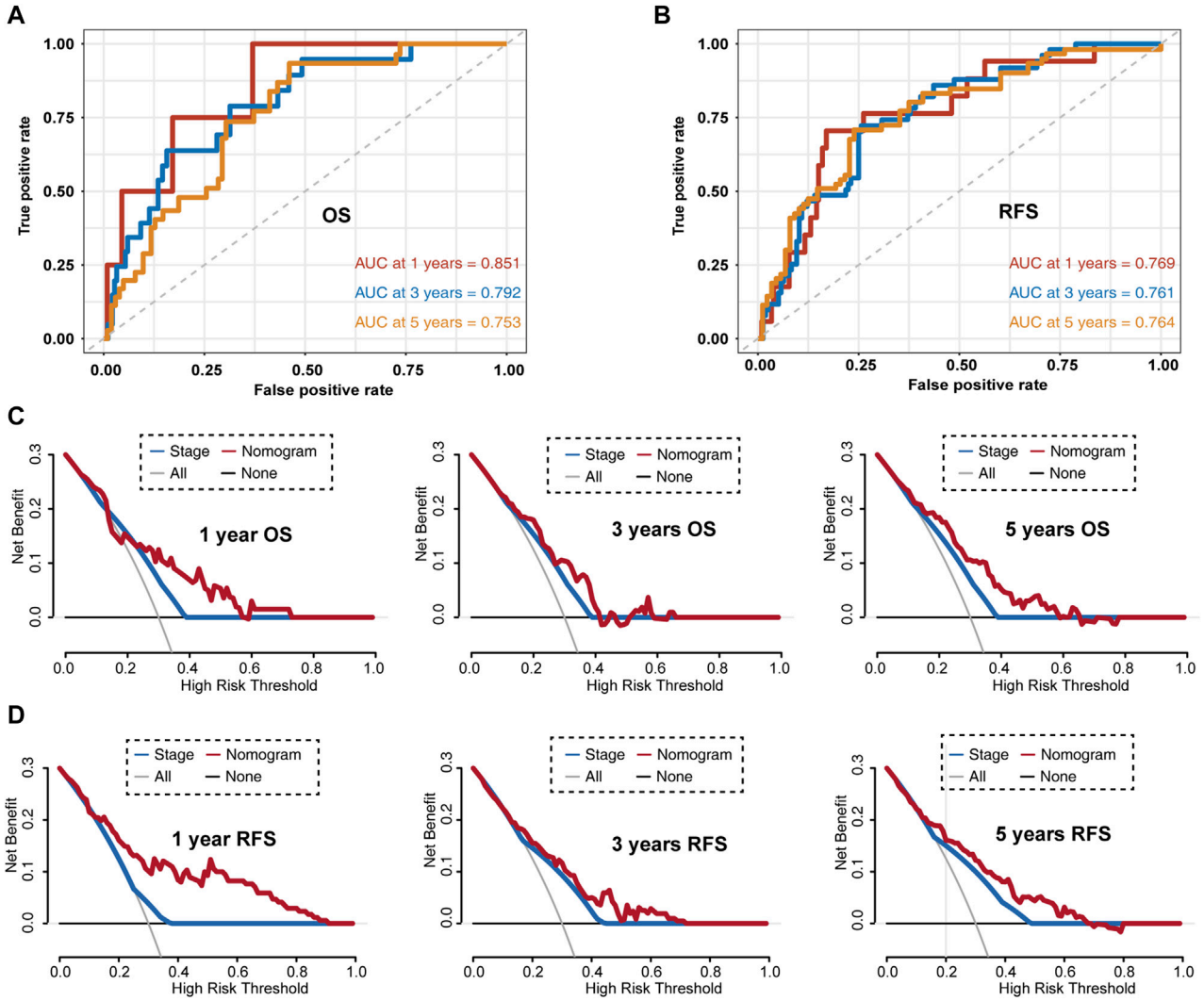


FIGURE 8 | Evaluating LUAD survival prediction nomogram based on the GSE31210 dataset. **(A)** Time-dependent receiver operating characteristic curves for nomogram prediction of 1-, 3-, and 5-year overall survival rates. **(B)** Time-dependent receiver operating characteristic curves for nomogram mediated prediction of 1-, 3-, and 5-year relapse-free survival in LUAD patients. **(C)** Decision curve analysis for overall survival. **(D)** Decision curve analysis for relapse-free survival. OS, overall survival; RFS, relapsed-free survival.

of LUAD. Third, only two GEO datasets were applied for validation; however, further validation using larger datasets is required.

Compared to a signature developed on a complete set of genes, our UPSG signature has a major advantage. We made a comprehensive exploration of the prognostic role of UPSGs in LUAD from a bioinformatic perspective. Compared to the large and complex genome signature, our risk had a better clinical operability which was calculated by several simple UPSGs. However, our research also has a major disadvantage. As our research was based on UPSGs, other prognosis genes were ignored and not included in this study.

In summary, we established a novel signature model for LUAD prognosis based on an eight-UPSG signature, which accurately predicted prognosis of LUAD patients. This signature may serve as an effective tool for designing personalized therapies and guiding medical decisions for LUAD patients in the future. Moreover, these UPSGs in the signature are promising targets for pharmacological and gene therapy in the future.

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.

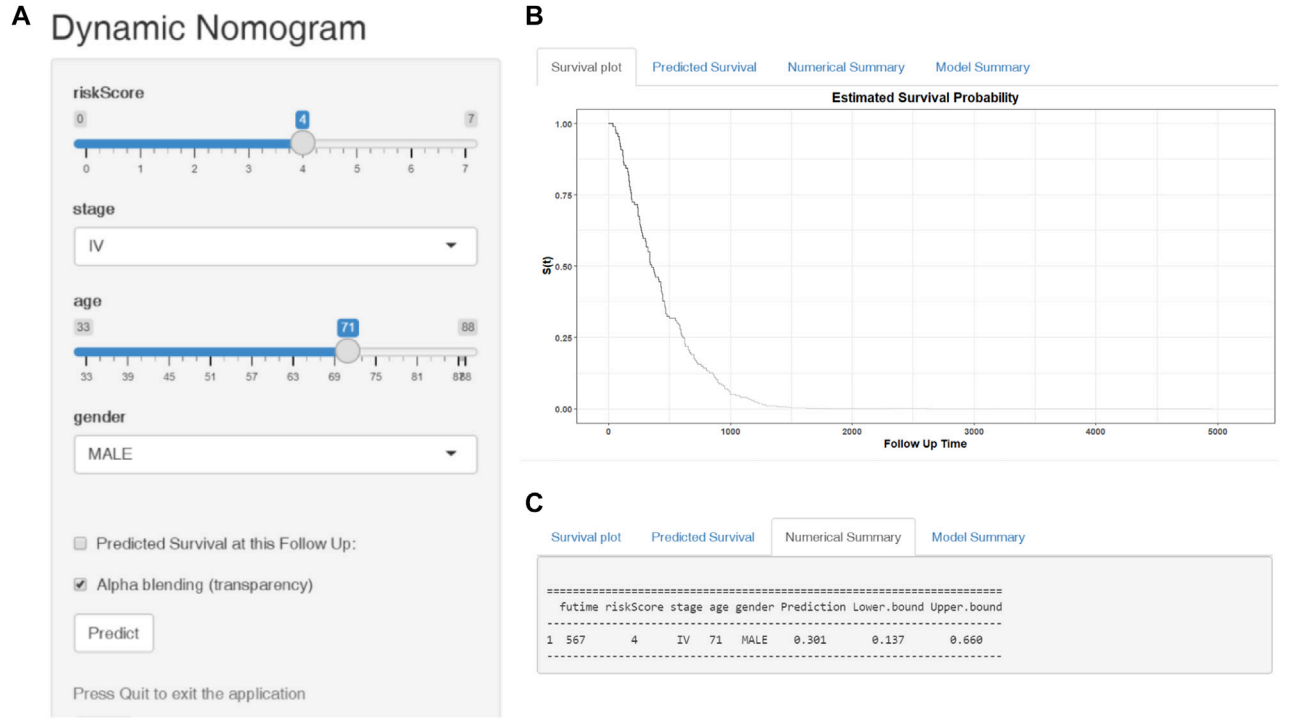


FIGURE 9 | Web-based dynamic nomogram for predicting the overall survival of LUAD patients (https://ostool.shinyapps.io/lungcancer/). **(A)** Input interface: You can enter the related variables of participants in this interface. **(B)** Survival plot: It represents participants' survival probability with different follow up time. **(C)** Numerical summary: It shows the actual values of probability and 95% confidence intervals.

AUTHOR CONTRIBUTIONS

YT and YG designed the experiment. YG analyzed the data. YG and YT interpreted the data. YG and YT wrote the manuscript. YG carefully reviewed the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by Youth Talent Cultivation Project of First Affiliated Hospital of Nanchang University (no. YFYPY202116), and the Science and technology Project of Jiangxi Provincial Health Commission (no.202210362).

ACKNOWLEDGMENTS

We would like to acknowledge TCGA and GEO databases for providing their platforms and those 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.2022.893511/full#supplementary-material

REFERENCES

Bade, B. C., and Dela Cruz, C. S. (2020). Lung Cancer 2020. *Clin. Chest Med.* 41, 1–24. doi:10.1016/j.ccm.2019.10.001

Çetin, G., Klafack, S., Studencka-Turski, M., Krüger, E., and Ebstein, F. (2021). The Ubiquitin-Proteasome System in Immune Cells. *Biomolecules* 11, 60. doi:10.3390/biom11010060

Chen, H.-Y., and Chen, R.-H. (2016). Cullin 3 Ubiquitin Ligases in Cancer Biology: Functions and Therapeutic Implications. *Front. Oncol.* 6, 113. doi:10.3389/fonc.2016.00113

Chen, L., Muñoz-Antonia, T., and Cress, W. D. (2014). Trim28 Contributes to EMT via Regulation of E-Cadherin and N-Cadherin in Lung Cancer Cell Lines. *PLoS One* 9, e101040. doi:10.1371/journal.pone.0101040

Chen, S.-F., Chen, P.-Y., Hsu, H.-J., Wu, M.-J., and Yen, J.-H. (2017). Xanthohumol Suppresses Mylip/Idol Gene Expression and Modulates LDLR Abundance and Activity in HepG2 Cells. *J. Agric. Food Chem.* 65, 7908–7918. doi:10.1021/acs.jafc.7b02282

Dang, F., Nie, L., and Wei, W. (2021). Ubiquitin Signaling in Cell Cycle Control and Tumorigenesis. *Cell Death Differ.* 28, 427–438. doi:10.1038/s41418-020-00648-0

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

Dikshit, A., Jin, Y. J., Degan, S., Hwang, J., Foster, M. W., Li, C.-Y., et al. (2018). UBE2N Promotes Melanoma Growth via MEK/FRA1/SOX10 Signaling. *Cancer Res.* 78, 6462–6472. doi:10.1158/0008-5472.can-18-1040

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., et al. (2013). Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci. Signal* 6, pl1. doi:10.1126/scisignal.2004088

Hatakeyama, S. (2011). TRIM Proteins and Cancer. *Nat. Rev. Cancer* 11, 792–804. doi:10.1038/nrc3139

He, D. H., Chen, Y. F., Zhou, Y. L., Zhang, S. B., Hong, M., Yu, X., et al. (2021). Phytochemical Library Screening Reveals Betulinic Acid as a Novel Skp2-SCF E3 Ligase Inhibitor in Non-small Cell Lung Cancer. *Cancer Sci.* 112, 3218–3232. doi:10.1111/cas.15005

Hua, X., Zhao, W., Pesatori, A. C., Consonni, D., Caporaso, N. E., Zhang, T., et al. (2020). Genetic and Epigenetic Intratumor Heterogeneity Impacts Prognosis of Lung Adenocarcinoma. *Nat. Commun.* 11, 2459. doi:10.1038/s41467-020-16295-5

Hua, Z.-D., Liu, X.-B., Sheng, J.-H., Li, C., Li, P., Cai, X.-Q., et al. (2021). UBE2V2 Positively Correlates with PD-L1 Expression and Confers Poor Patient Survival in Lung Adenocarcinoma. *Appl. Immunohistochem. Mol. Morphol.* 29, 585–591. doi:10.1097/pai.0000000000000928

Hynes-Smith, R. W., Swenson, S. A., Vahle, H., Wittorf, K. J., Caplan, M., Amador, C., et al. (2019). Loss of FBXO9 Enhances Proteasome Activity and Promotes Aggressiveness in Acute Myeloid Leukemia. *Cancers (Basel)* 11, 1717. doi:10.3390/cancers11111717

Iyengar, S., and Farnham, P. J. (2011). KAP1 Protein: an Enigmatic Master Regulator of the Genome. *J. Biol. Chem.* 286, 26267–26276. doi:10.1074/jbc.r111.252569

Johnson, D. E. (2015). The Ubiquitin-Proteasome System: Opportunities for Therapeutic Intervention in Solid Tumors. *Endocr. Relat. Cancer* 22, T1–T17. doi:10.1530/erc-14-0005

Li, X., Elmira, E., Rohondia, S., Wang, J., Liu, J., and Dou, Q. P. (2018). A Patent Review of the Ubiquitin Ligase System: 2015-2018. *Expert Opin. Ther. Pat.* 28, 919–937. doi:10.1080/13543776.2018.1549229

Li, Y., Huang, J., Zeng, B., Yang, D., Sun, J., Yin, X., et al. (2018). PSMD2 Regulates Breast Cancer Cell Proliferation and Cell Cycle Progression by Modulating P21 and P27 Proteasomal Degradation. *Cancer Lett.* 430, 109–122. doi:10.1016/j.canlet.2018.05.018

Liu, H., Yang, P., Li, X., and Jia, Y. (2020). Ring Finger Protein 180 Is Associated with Biological Behavior and Prognosis in Patients with Non-small Cell Lung Cancer. *Oncol. Lett.* 20, 35. doi:10.3892/ol.2020.11898

Liu, J. A., Tai, A., Hong, J., Cheung, M. P. L., Sham, M. H., Cheah, K. S. E., et al. (2020). Fbxo9 Functions Downstream of Sox10 to Determine Neuron-Glial Fate Choice in the Dorsal Root Ganglia through Neurog2 Destabilization. *Proc. Natl. Acad. Sci. U.S.A.* 117, 4199–4210. doi:10.1073/pnas.1916164117

Manasanch, E. E., and Orlowski, R. Z. (2017). Proteasome Inhibitors in Cancer Therapy. *Nat. Rev. Clin. Oncol.* 14, 417–433. doi:10.1038/nrclinonc.2016.206

Matsuyama, Y., Suzuki, M., Arima, C., Huang, Q. M., Tomida, S., Takeuchi, T., et al. (2011). Proteasomal Non-catalytic Subunit PSMD2 as a Potential Therapeutic Target in Association with Various Clinicopathologic Features in Lung Adenocarcinomas. *Mol. Carcinog.* 50, 301–309. doi:10.1002/mc.20632

Messerschmidt, D. M., de Vries, W., Ito, M., Solter, D., Ferguson-Smith, A., and Knowles, B. B. (2012). Trim28 Is Required for Epigenetic Stability during Mouse Oocyte to Embryo Transition. *Science* 335, 1499–1502. doi:10.1126/science.1216154

Narayanan, S., Cai, C.-Y., Assaraf, Y. G., Guo, H.-Q., Cui, Q., Wei, L., et al. (2020). Targeting the Ubiquitin-Proteasome Pathway to Overcome Anti-cancer Drug Resistance. *Drug Resist. Updat.* 48, 100663. doi:10.1016/j.drup.2019.100663

Ni, M., Yan, Q., Xue, H., Du, Y., Zhao, S., and Zhao, Z. (2021). Identification of MYLIP Gene and miRNA-802 Involved in the Growth and Metastasis of Cervical Cancer Cells. *Cancer Biomark.* 30, 287–298. doi:10.3233/cbm-201523

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

Okumura, T., Ikeda, K., Ujihira, T., Okamoto, K., Horie-Inoue, K., Takeda, S., et al. (2018). Proteasome 26S Subunit PSMD1 Regulates Breast Cancer Cell Growth through P53 Protein Degradation. *J. Biochem.* 163, 19–29. doi:10.1093/jb/mvx053

Raz, V., Buijze, H., Raz, Y., Verwey, N., Anvar, S. Y., Aartsma-Rus, A., et al. (2014). A Novel Feed-Forward Loop between ARIH2 E3-Ligase and PABPN1 Regulates Aging-Associated Muscle Degeneration. *Am. J. Pathology* 184, 1119–1131. doi:10.1016/j.ajpath.2013.12.011

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

Sio, F. R. S., Barde, I., Offner, S., Kapopoulou, A., Corsinotti, A., Bojkowska, K., et al. (2012). KAP1 Regulates Gene Networks Controlling T-cell Development and Responsiveness. *FASEB J.* 26, 4561–4575. doi:10.1096/fj.12-206177

Tang, Y., Jiang, Y., Qing, C., Wang, J., and Zeng, Z. (2020). Systematic Construction and Validation of an Epithelial-Mesenchymal Transition Risk Model to Predict Prognosis of Lung Adenocarcinoma. *Aging* 13, 794–812. doi:10.18632/aging.202186

Wang, J., He, L., Tang, Y., Li, D., Yang, Y., and Zeng, Z. (2020). Development and Validation of a Nomogram with an Epigenetic Signature for Predicting Survival in Patients with Lung Adenocarcinoma. *Aging (Albany NY)* 12, 23200–23216. doi:10.18632/aging.104090

Wang, J., Li, J., Zhang, L., Qin, Y., Zhang, F., Hu, R., et al. (2021). Comprehensive Analysis of Ubiquitin-Proteasome System Genes Related to Prognosis and Immunosuppression in Head and Neck Squamous Cell Carcinoma. *Aging* 13, 20277–20301. doi:10.18632/aging.203411

Wang, W., He, J., Lu, H., Kong, Q., and Lin, S. (2020). KRT8 and KRT19, Associated with EMT, Are Hypomethylated and Overexpressed in Lung Adenocarcinoma and Link to Unfavorable Prognosis. *Biosci. Rep.* 40, BSR20193468. doi:10.1042/BSR20193468

Wu, C., Li, M., Meng, H., Liu, Y., Niu, W., Zhou, Y., et al. (2019). Analysis of Status and Countermeasures of Cancer Incidence and Mortality in China. *Sci. China Life Sci.* 62, 640–647. doi:10.1007/s11427-018-9461-5

Wu, Z., Liu, H., Sun, W., Du, Y., He, W., Guo, S., et al. (2020). RNF180 Mediates STAT3 Activity by Regulating the Expression of RhoC via the Proteasomal Pathway in Gastric Cancer Cells. *Cell Death Dis.* 11, 881. doi:10.1038/s41419-020-03096-3

Xie, L., Dang, Y., Guo, J., Sun, X., Xie, T., Zhang, L., et al. (2019). High KRT8 Expression Independently Predicts Poor Prognosis for Lung Adenocarcinoma Patients. *Genes (Basel)* 10, 36. doi:10.3390/genes10010036

Zeng, H., Castillo-Cabrera, J., Manser, M., Lu, B., Yang, Z., Strande, V., et al. (2019). Genome-wide CRISPR Screening Reveals Genetic Modifiers of Mutant EGFR Dependence in Human NSCLC. *Elife* 8, e50223. doi:10.7554/eLife.50223

Zhao, L., Zhou, Y., Zhao, Y., Li, Q., Zhou, J., and Mao, Y. (2020). Long Non-coding RNA TUSC8 Inhibits Breast Cancer Growth and Metastasis via miR-190b-5p/MYLIP axis. *Aging* 12, 2974–2991. doi:10.18632/aging.102791

Zhao, Y., Long, M. J. C., Wang, Y., Zhang, S., and Aye, Y. (2018). Ube2V2 Is a Rosetta Stone Bridging Redox and Ubiquitin Codes, Coordinating DNA Damage Responses. *ACS Cent. Sci.* 4, 246–259. doi:10.1021/acscentsci.7b00556

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: 08 June 2022
doi: 10.3389/fgene.2022.905581

Prognostic Implication and Immunological Role of PSMD2 in Lung Adenocarcinoma

Huihui Zhao[1] *and Guojun Lu*[2]*

[1]Department of Oncology, The Second Hospital of Nanjing, Nanjing University of Chinese Medicine, Nanjing, China, [2]Department of Respiratory Medicine, Nanjing Chest Hospital, Affiliated Nanjing Brain Hospital, Nanjing Medical University, Nanjing, China

Background: Although previous studies reported that 26S proteasome non-ATPase regulatory subunit 2 (*PSMD2*) is involved in many human cancers. However, its clinical significance and function in lung adenocarcinoma remain unclear. Here, we examined the prognostic and immunological role of *PSMD2* in lung adenocarcinoma.

Methods: The Cancer Genome Atlas (TCGA) was conducted to analyze *PSMD2* expression and verified using UALCAN. PrognoScan and Kaplan-Meier curves were utilized to assess the effect of *PSMD2* on survival. cBioPortal database was conducted to identify the mutation characteristics of *PSMD2*. Functional enrichment was performed to determine *PSMD2*-related function. Cancer Single-cell State Atlas (CancerSEA) was used to explore the cancer functional status of *PSMD2* at single-cell resolution. *PSMD2*-related immune infiltration analysis was conducted. Tumor-Immune system interaction database (TISIDB) was performed to verify the correlation between *PSMD2* expression and tumor-infiltrating lymphocytes (TILs).

Results: Both mRNA and protein expression of *PSMD2* were significantly elevated in lung adenocarcinoma. High expression of *PSMD2* was significantly correlated with high T stage ($p = 0.014$), lymph node metastases ($p < 0.001$), and TNM stage $p = 0.005$). Kaplan-Meier curves indicated that high expression of *PSMD2* was correlated with poor overall survival (38.2 vs. 59.7 months, $p < 0.001$) and disease-specific survival (59.9 months vs. not available, $p = 0.004$). Multivariate analysis suggested that *PSMD2* was an independent biomarker for poor overall survival (HR 1.471, 95%CI, 1.024–2.114, $p = 0.037$). *PSMD2* had a high mutation frequency of 14% in lung adenocarcinoma. The genetic mutation of *PSMD2* was also correlated with poor overall survival, disease-specific survival, and progression-free survival in lung adenocarcinoma. Functional enrichment suggested *PSMD2* expression was involved in the cell cycle, RNA transport, and cellular senescence. CancerSEA analysis indicated *PSMD2* expression was positively correlated with cell cycle, DNA damage, and DNA repair. Immune infiltration analysis suggested that *PSMD2* expression was correlated with immune cell infiltration levels and abundance of TILs.

Conclusion: The upregulation of *PSMD2* is significantly correlated with poor prognosis and immune infiltration levels in lung adenocarcinoma. Our findings suggest that *PSMD2* is

OPEN ACCESS

Edited by:
Wei Zhao,
City University of Hong Kong, Hong Kong SAR, China

Reviewed by:
Jin Bai,
Xuzhou Medical University, China
Shouyu Wang,
School of Medicine, Nanjing University, China

***Correspondence:**
Guojun Lu
guojunlu@njmu.edu.cn

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *27 March 2022*
Accepted: *23 May 2022*
Published: *08 June 2022*

Citation:
Zhao H and Lu G (2022) Prognostic Implication and Immunological Role of PSMD2 in Lung Adenocarcinoma. Front. Genet. 13:905581. doi: 10.3389/fgene.2022.905581

a potential biomarker for poor prognosis and immune therapeutic target in lung adenocarcinoma.

Keywords: PSMD2, lung adenocarcinoma, prognosis, biomarker, immune infiltration

INTRODUCTION

Lung cancer is still the most common cancer in China and the leading cause of cancer-related death in China and the United States of America (USA) (Lee et al., 2019; Siegel et al., 2022; Xia et al., 2022). Lung adenocarcinoma is the most common histological subtype of lung cancer and accounts for about 40% of diagnoses (Hua et al., 2020). The research to date indicates that targeted therapy and immune-checkpoint inhibitors plus chemotherapy have become standard therapy for lung adenocarcinoma and bring survival benefits (Hirsch et al., 2017; Gandhi et al., 2018; Mok et al., 2019). However, up to now, the 5-year survival rate of lung adenocarcinoma has remained a struggle at 16% (Huo et al., 2019). Furthermore, previous studies have reported that tumor-infiltrating lymphocytes (TILs) including tumor-associated macrophages and tumor-infiltrating neutrophils are correlated with the prognosis and sensitivity of chemotherapy and immunotherapy (Waniczek et al., 2017; Pan et al., 2019). As a result, it is urgent to explore the immune phenotype of lung adenocarcinoma and identify novel prognostic biomarkers and therapeutic strategies for lung adenocarcinoma.

The Proteasome 26S Subunit, Non-ATPase (*PSMD*) gene family, is composed of the *PSMD1* to *PSMD14*. Previous studies have shown that *PSMD* family genes can play important role in the progress of circulation and tumorigenesis by regulating ubiquitinated protein breakdown. *PSMD1-3* and *PSMD7* were elevated in breast cancer and correlated with poor prognosis. They can promote cell proliferation and cell cycle progression prognosis in breast cancer cell lines (Oguro et al., 2015; Okumura et al., 2018; Fararjeh et al., 2019; Zhao et al., 2020). In esophageal cancer, upregulation of *PSMD4* can reduce endoplasmic reticulum stress-induced cell apoptosis to promote tumor progression (Ma et al., 2019). It was also reported that *PSMD9* was correlated with recurrence after radiotherapy in cervical cancer and can predict radiotherapy benefits in breast cancer (Langlands et al., 2014; Köster et al., 2020). A previous study has shown that *PSMD2* was overexpressed in lung cancer and patients with higher expression of *PSMD2* were correlated with poorer prognosis (Matsuyama et al., 2011). However, the number of lung cancer patients enrolled in this study was relatively small. Moreover, *PSMD* family genes are reported to correlate with immune infiltration profiles in breast cancer (Xuan et al., 2021). Up to now, there is no studies have designed to investigate the association between *PSMD2* expression and immune cell infiltration in lung adenocarcinoma.

In this study, we first examined the expression of *PSMD2* in lung adenocarcinoma in the TCGA and UALCAN databases. Then we conducted PrognoScan and Kaplan-Meier curves to assess the correlation between *PSMD2* expression and prognosis. cBioPortal database was conducted to identify the mutation characteristics and prognostic significance of *PSMD2*. Furthermore, functional enrichment analysis, cancer functional status at single-cell resolution and immune cell infiltration analysis were also conducted. Our findings link the upregulation of *PSMD2* with the poor prognosis and propose a therapeutic immunological target for lung adenocarcinoma.

MATERIALS AND METHODS

The Cancer Genome Atlas

TCGA (https://portal.gdc.cancer.gov/) is a landmark cancer genomics program (Tomczak et al., 2015). In the present study, the expression level of *PSMD2* in different cancer types and related clinical data in lung adenocarcinoma were downloaded from TCGA for further analysis.

UALCAN and Clinical Proteomic Tumor Analysis Consortium Database

The UALCAN (http://ualcan.path.uab.edu/) is an online web resource to analyze cancer omics data (Chandrashekar et al., 2017). It can be used to perform protein expression analysis of the CPTAC dataset (Edwards et al., 2015). In this study, we input *PSMD2* in the "scan by gene" module of UALCAN to explore the total protein expression between primary tumor and normal tissues with the CPTAC dataset of lung adenocarcinoma.

The Human Protein Atlas

The HPA database (https://proteinatlas.org/) contains human protein expression profiles in tumor tissues and normal tissues (Uhlén et al., 2015; Uhlen et al., 2017). In the present study, HPA was conducted to confirm the protein expression of PSMD2 in lung adenocarcinoma.

PrognoScan

PrognoScan database (http://dna00.bio.kyutech.ac.jp/PrognoScan/index.html) is an online database for prognosis analysis (Mizuno et al., 2009). In the present study, we input *PSMD2* in the "Enter gene identifier" module of PrognoScan and validated the prognostic importance of *PSMD2* in lung adenocarcinoma with two datasets (GSE31210, GSE13213).

cBioPortal Database

The cBioPortal for Cancer Genomics (https://www.cbioportal.org/) is a Web resource to explore, visualize, and analyze multidimensional cancer genomics data (Cerami et al., 2012; Gao et al., 2013). In this study, we utilized the cBioPortal database to assess mutation data of *PSMD2*, capture its prognostic value in altered lung adenocarcinoma patients, and acquire co-expressed genes of *PSMD2*.

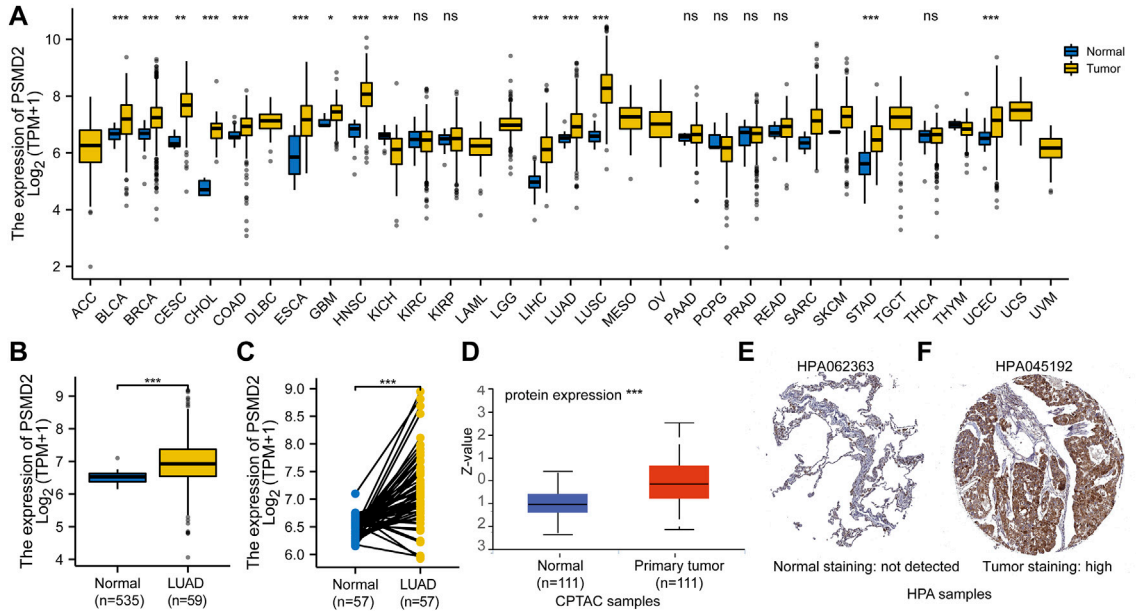


FIGURE 1 | Expression pattern of *PSMD2* in Pan-cancer perspective and lung adenocarcinoma. **(A)** *PSMD2* is increased in 13 cancer types and decreased only in Kidney Chromophobe from TCGA. **(B,C)** Both the Mann-Whitney U test and paired t-test indicate mRNA expression of *PSMD2* is elevated in lung adenocarcinoma. **(D)** The protein expression of PSMD2 is upregulated in CPTAC. **(E)** The PSMD2 protein expression in normal tissue from HPA. **(F)** The PSMD2 protein expression in lung adenocarcinoma tissue from HPA. ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, ns, no significant).

Functional Enrichment Analysis

After acquiring co-expressed genes of *PSMD2* from cBioPortal, we performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis with R package clusterProfiler to further quantify the functional annotations of these co-expressed genes (Yu et al., 2012). R package ggplot2 was adopted to visualize the results of GO and KEGG.

Cancer Single-Cell State Atlas

CancerSEA is an online database to comprehensively decode distinct functional states of cancer cells from different cancer types at single-cell resolution (Yuan et al., 2019). In this study, the relevant data of *PSMD2* across different tumor functional states based on single-cell sequencing data were downloaded from CancerSEA and a heatmap was drawn. The t-SNE diagram was also downloaded from CancerSEA to describe the distribution of all individual cells.

Immune Infiltration Analysis

The methods for analyzing immune infiltration were described as before (Li et al., 2021; Zhang et al., 2022). Tumor infiltration expression of 24 immune cell types was downloaded from published literature (Bindea et al., 2013) and quantified by the ssGSEA method with R package GSVA (Hänzelmann et al., 2013). Furthermore, Spearman correlation and Mann-Whitney U test were conducted to determine the correlation between low/high expression of *PSMD2* and immune cell infiltration in lung adenocarcinoma.

Tumor-Immune System Interaction Database

TISIDB (http://cis.hku.hk/TISIDB/) is an integrated repository portal for tumor and immune system interaction (Ru et al., 2019). In the present study, TISIDB was employed to capture the relations between *PSMD2* expression and the abundance of TILs in lung adenocarcinoma. The relations between *PSMD2* expression and TILs were determined by Spearman's test.

Statistical Analyses

R (V 3.6.3, https://www.r-project.org/) was utilized for statistical analyses and R package ggplot2 was employed for visualization of expression differences. Paired t-test and Mann-Whitney U test were proposed to explore the expression differences between lung adenocarcinoma tissues and adjacent normal tissues. Kaplan-Meier curves and Cox regression were conducted with R package survminer and survival to assess the effect of *PSMD2* on survival. $p < 0.05$ was considered statistically significant.

RESULTS

Assessment of PSMD2 mRNA Expression in Pan-Cancer Perspective and Lung Adenocarcinoma

To assess the transcription level of *PSMD2* in multiple tumors and normal samples, we conducted analyses on the TCGA database. As

TABLE 1 | The basic characteristics of clinicopathological factors in lung adenocarcinoma patients.

Characteristics	Total	Low expression	High expression	*p*-value
	N (%)	N (%)	N (%)	
T stage				0.078
T1	175 (32.9)	94 (17.7)	81 (15.2)	
T2	289 (54.3)	138 (25.9)	151 (28.4)	
T3	49 (9.2)	28 (5.3)	21 (3.9)	
T4	19 (3.6)	5 (0.9)	14 (2.6)	
N stage				0.038*
N0	348 (67.0)	186 (35.8)	162 (31.2)	
N1	95 (18.3)	40 (7.7)	55 (10.6)	
N2	74 (14.3)	29 (5.6)	45 (8.7)	
N3	2 (0.4)	1 (0.2)	1 (0.2)	
M stage				1.000
M0	361 (93.5)	174 (45.1)	187 (48.4)	
M1	25 (6.5)	12 (3.1)	13 (3.4)	
Pathologic stage				0.131
Stage I	294 (55.8)	155 (29.4)	139 (26.4)	
Stage II	123 (23.3)	62 (11.8)	61 (11.6)	
Stage III	84 (16.0)	32 (6.1)	52 (9.9)	
Stage IV	26 (4.9)	13 (2.5)	13 (2.5)	
Gender				0.759
Female	286 (53.5)	145 (27.1)	141 (26.4)	
Male	249 (46.5)	122 (22.8)	127 (23.7)	
Age				1.000
<=65	255 (49.4)	127 (24.6)	128 (24.8)	
>65	261 (50.6)	131 (25.4)	130 (25.2)	
Smoker				0.034d*
No	75 (14.4)	46 (8.8)	29 (5.6)	
Yes	446 (85.6)	211 (40.5)	235 (45.1)	
Anatomic neoplasm subdivision				0.518
Left	205 (39.4)	107 (20.6)	98 (18.8)	
Right	315 (60.6)	154 (29.6)	161 (Zhang et al., 2022)	
Anatomic neoplasm subdivision2				0.844
Central Lung	62 (32.8)	30 (15.9)	32 (16.9)	
Peripheral Lung	127 (67.2)	58 (30.7)	69 (36.5)	

*, $p < 0.05$.

shown in **Figure 1A**, compared with normal tissue controls, the expression of *PSMD2* was upregulated in 13 cancer types and only downregulated in Kidney Chromophobe. Looking at **Figure 1B**, the Mann-Whitney U test indicated that the *PSMD2* mRNA expression level in lung adenocarcinoma ($n = 535$) was significantly upregulated relative to normal lung tissues ($n = 59$) (6.964 ± 0.698 vs. 6.510 ± 0.177, $p < 0.001$). As shown in **Figure 1C**, paired t-test analysis showed that *PSMD2* mRNA expression level in lung adenocarcinoma ($n = 57$) were significantly upregulated relative to normal lung tissues ($n = 57$) (7.156 ± 0.662 vs. 6.512 ± 0.177, $p < 0.001$). Taken together, these results suggest that mRNA expression of *PSMD2* is elevated in lung adenocarcinoma tissues.

Assessment of PSMD2 Protein Expression in Lung Adenocarcinoma

To explore the total protein expression of PSMD2 between lung adenocarcinoma and normal tissues, we analyzed CPTAC with the UALCAN dataset. The result in **Figure 1D** also showed that the protein expression of PSMD2 in lung adenocarcinoma was significantly elevated than in normal tissues. Protein expression from HPA also suggested that PSMD2 in lung adenocarcinoma tissue was higher than that in normal lung tissue (**Figures 1E,F**). Overall, these results indicate that protein expression of PSMD2 is increased in lung adenocarcinoma tissues.

Relationships Between PSMD2 Expression and Clinicopathological Factors of Lung Adenocarcinoma Patients

The basic clinicopathological factors of lung adenocarcinoma patients was listed in **Table 1**. To examine the clinical relevance of *PSMD2* expression, we conducted the expression between different clinicopathological factors of lung adenocarcinoma patients with the Mann-Whitney U test. As listed in **Figure 2**, the high expression of *PSMD2* was significantly correlated with T stage (T1 vs. T2-4, $p = 0.014$), N stage (N0 vs. N1-N3, $p < 0.001$), and TNM stage (Stage I-II vs. Stage III-IV, $p = 0.005$). However, *PSMD2* expression had no significant correlation with other clinicopathological factors, such as M stage ($p = 0.401$), gender ($p = 0.174$), age ($p = 0.884$), smoking condition ($p = 0.225$), and anatomic subdivision (right vs. left, $p = 0.999$; peripheral vs. central, $p = 0.149$). On the whole, these results indicate that

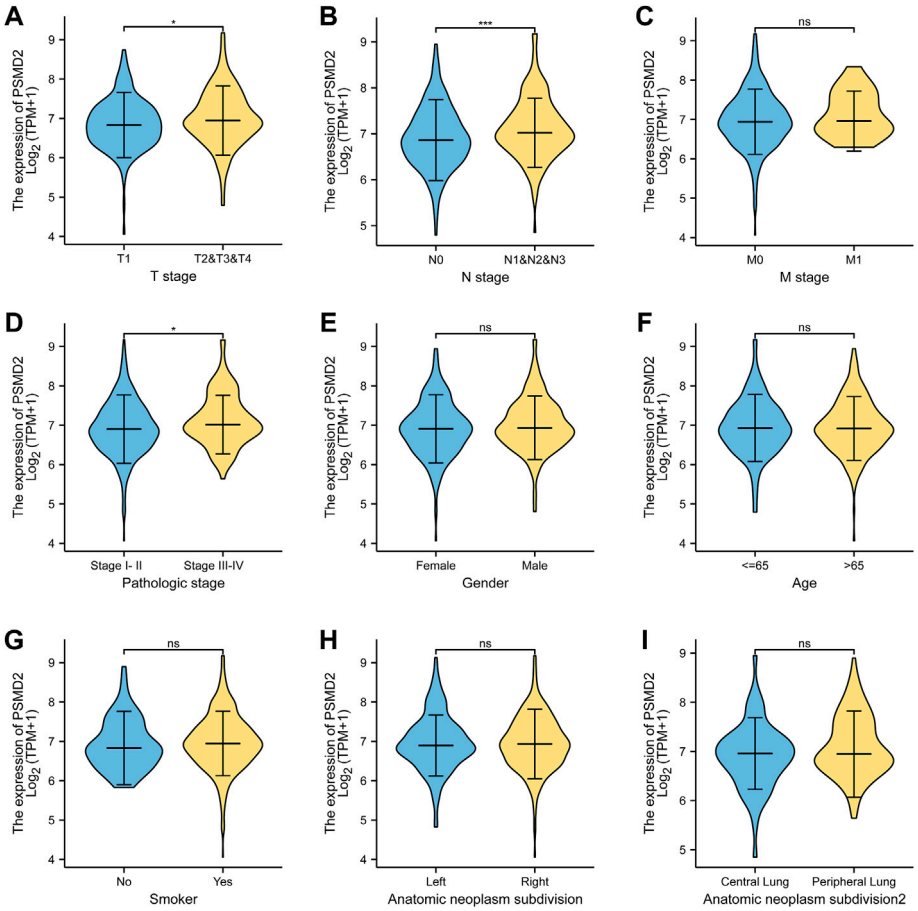


FIGURE 2 | The clinical relevance of *PSMD2* expression and clinicopathological factors in lung adenocarcinoma patients. *PSMD2* mRNA expression was significantly increased with high T stage **(A)**, N stage **(B)**, and TNM stage **(D)**. No observed differences in M stage **(C)**, gender **(E)**, age **(F)**, smoke condition **(G)**, and anatomic subdivision (right vs. left, central vs. peripheral) **(H,I)**. (ns, no significance, $^{*}p < 0.05$, $^{***}p < 0.001$).

PSMD2 had a significant correlation with T stage, lymph node metastases, and high TNM stage.

Correlation Between PSMD2 Expression and Prognosis

To investigate the correlation between *PSMD2* mRNA expression and prognosis of lung adenocarcinoma patients, Kaplan-Meier curves with R package survminer and survival were used. As can be seen in **Figure 3A**, the overall survival of lung adenocarcinoma patients with higher *PSMD2* expression was significantly shorter than those with lower *PSMD2* expression (38.2 vs. 59.7 months, $p < 0.001$). As shown in **Figure 3B**, the disease-specific survival of lung adenocarcinoma patients with higher *PSMD2* expression was also significantly shorter than those with lower *PSMD2* expression (59.9 months vs. not available, $p = 0.004$). However, as shown in **Figure 3C**, no significant difference was observed between higher and lower *PSMD2* expression and progression-free interval (28.8 vs. 40.1 months, $p = 0.089$). Furthermore, we validated the prognostic importance of *PSMD2* in PrognoScan. The result from two datasets in **Figures 3D,E** suggested that high *PSMD2* expression was correlated with poor overall survival in lung adenocarcinoma. As a result, by associating *PSMD2* and other clinical characteristics in **Figures 4A,B**, we established a nomogram and calibration plot for predicting the overall survival probability of lung adenocarcinoma patients.

Prognostic Role of PSMD2 in Lung Adenocarcinoma Patients

To further explore the prognostic role of *PSMD2* for overall survival and disease-specific survival in lung adenocarcinoma patients, we performed Cox univariate and multivariate analyses. As shown in **Figure 5A**, univariate analysis showed that high expression of *PSMD2* was significantly correlated with poor overall survival of lung adenocarcinoma patients (HR 1.694, 95%CI, 1.264–2.269, $p < 0.001$). Moreover, we conducted a multivariate analysis with the Cox proportional hazards model. The multivariate analysis in **Figure 5B** suggested that T stage (HR 1.767, 95%CI, 1.098–2.842, $p = 0.019$), N stage (HR 1.791, 95%CI, 1.181–2.716, $p = 0.006$), Pathologic stage (HR 1.693, 95%CI, 1.056–2.717, $p = 0.029$), and *PSMD2* (HR 1.471, 95%CI, 1.024–2.114, $p = 0.037$) were independent factors for

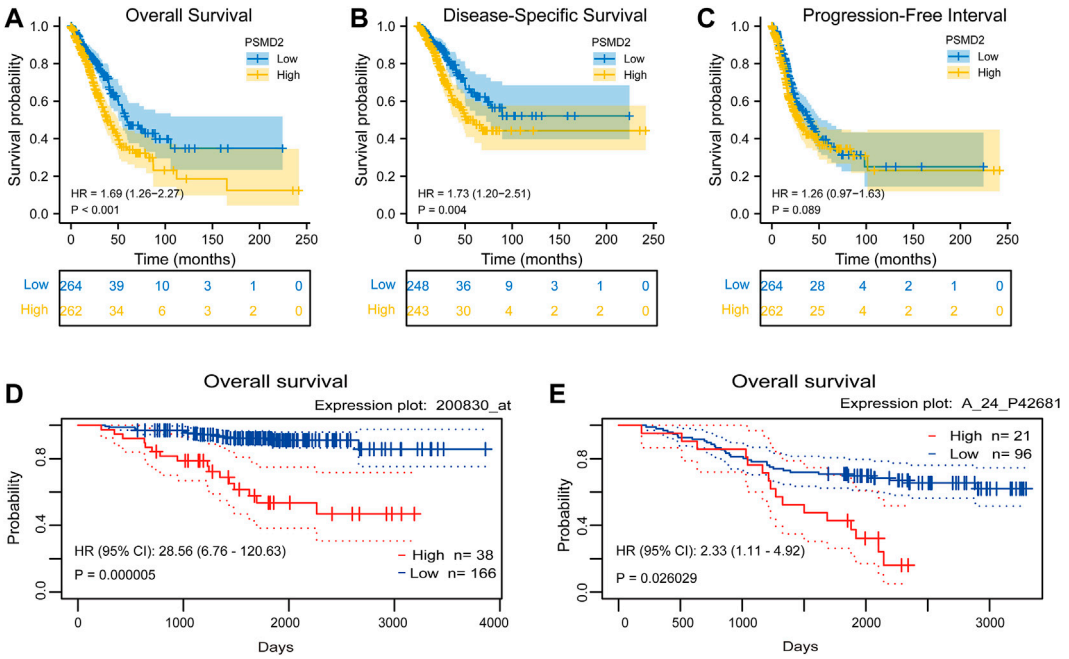


FIGURE 3 | Correlation between *PSMD2* expression and prognosis in lung adenocarcinoma. **(A)** Higher *PSMD2* expression had a shorter overall survival. **(B)** Higher *PSMD2* expression had a shorter disease-specific survival. **(C)** No significant difference was observed between *PSMD2* expression and progression-free interval. **(D,E)** Two datasets (GSE31210, GSE13213) in PrognoScan suggested that high *PSMD2* expression was correlated with poor overall survival in lung adenocarcinoma.

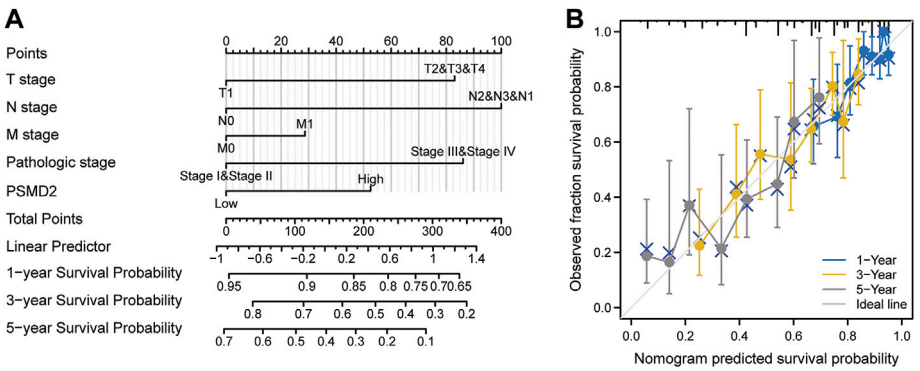


FIGURE 4 | A nomogram and calibration plot for predicting the 1-, 3-, and 5-year overall survival probability of lung adenocarcinoma patients. **(A)** To predict the survival probability, first draw a vertical line from each clinical factor up to the points axis and get a point. Repeat until points for all clinical factors are decided. Sum all the points and then locate the total point on the total points axis. Draw a line straight down to the risk axis and finally obtain survival probability. **(B)** Calibration plot of the nomogram.

overall survival in lung adenocarcinoma patients. Furthermore, as shown in **Supplementary Table S1**, univariate analysis showed that high expression of *PSMD2* was significantly correlated with poor disease-specific survival of lung adenocarcinoma patients (HR 1.734, 95%CI, 1.198–2.512, p = 0.004). However, the multivariate analysis suggested *PSMD2* (HR 1.615, 95%CI, 0.998–2.614, p = 0.051) was not an independent factor for disease-specific survival in lung adenocarcinoma. Our data indicate that *PSMD2* is an independent factor for overall survival in lung adenocarcinoma patients.

Genetic Mutation of PSMD2 was Correlated With Poor Prognosis

To identify the mutation characteristics of *PSMD2* and its prognostic value in altered lung adenocarcinoma patients, we utilized an analysis on the cBioPortal database. It can be seen from **Figure 6A** that *PSMD2* had a high mutation frequency of 14% in lung adenocarcinoma (TCGA, PanCancer Atlas). The main alteration frequency of *PSMD2* was mRNA upregulation and amplification (**Figure 6B**). Moreover,

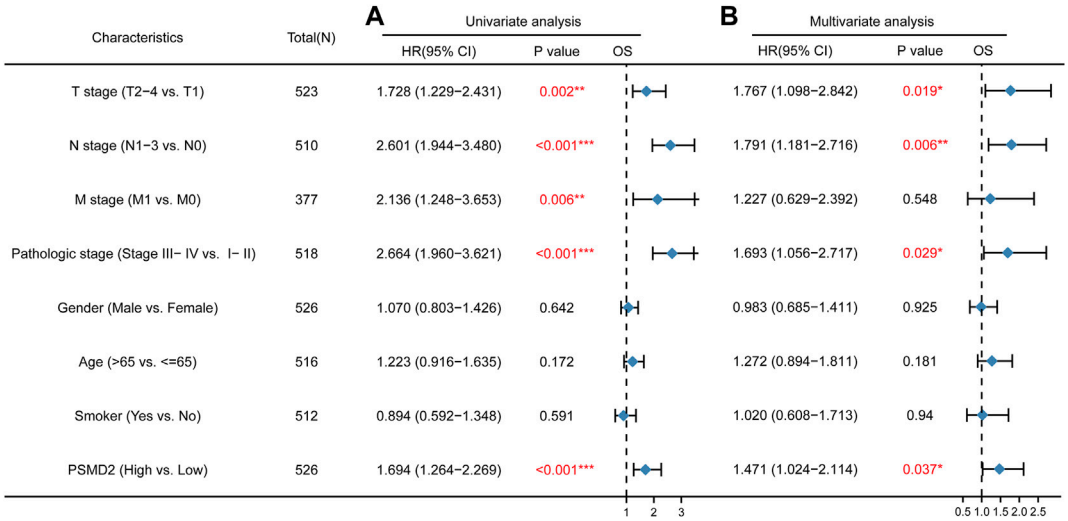


FIGURE 5 | Cox regression analyses and forest plots of prognostic factors for overall survival. **(A)** Univariate Cox analysis and forest plot results of *PSMD2* for OS. **(B)** Multivariate Cox analysis and forest plot indicated *PSMD2* was an independent prognostic biomarker for OS in lung adenocarcinoma. (Red colors mean significant results. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$).

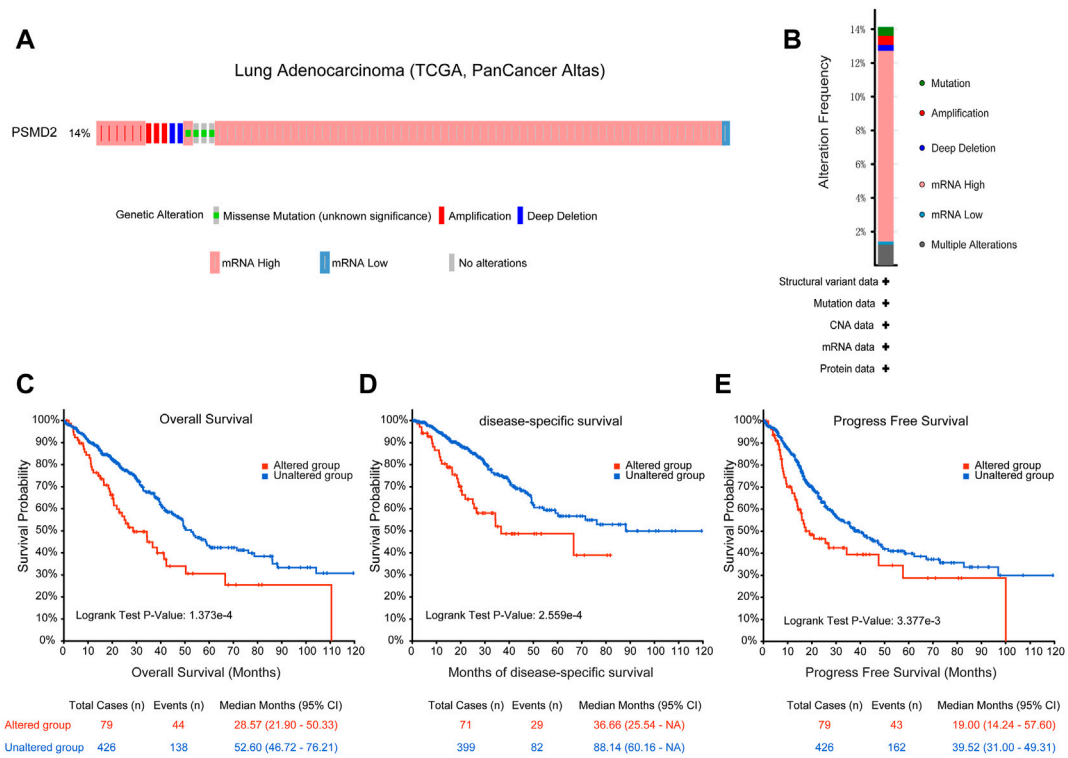


FIGURE 6 | Genomic mutation of *PSMD2* in lung adenocarcinoma. **(A)** OncoPrint presented the different mutation types and proportions of *PSMD2*. **(B)** Cancer types summary showed the alteration frequency. **(C–E)** Genomic mutation of *PSMD2* was correlated with poor overall survival, disease-specific survival, and progression-free survival.

compared with the unaltered group, the result of Kaplan-Meier curves found that the altered group was associated with poor prognosis in overall survival (**Figure 6C**, 28.57 vs. 52.60 months, p = 1.737e-4), disease-specific survival (**Figure 6D**, 36.66 vs. 88.14 months, p = 2.559e-4), and progression-free survival (**Figure 6E**, 19.00 vs. 39.52

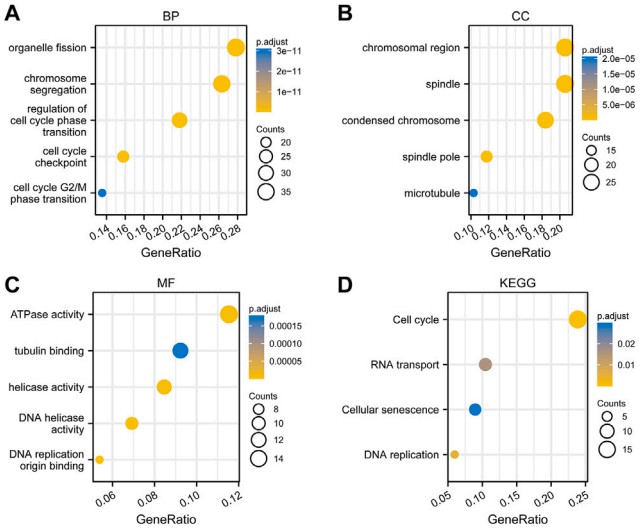


FIGURE 7 | Functional enrichment analysis of co-expressed genes of *PSMD2*. **(A)** Biological progress of organelle fission, chromosome segregation, and regulation of cell cycle phase transition. **(B)** Structural constituents in the condensed chromosome, spindle, and chromosomal region. **(C)** The molecular function of ATPase activity, tubulin binding, and DNA helicase activity. **(D)** KEGG pathway analysis indicated enrichment function in the cell cycle, RNA transport, and cellular senescence.

months, p = 3.377e-4). Overall, these results suggest that genetic mutation of *PSMD2* was correlated with poor prognosis in lung adenocarcinoma.

Functional Enrichment Analysis of Co-expressed Genes of PSMD2

To perform GO and KEGG analysis, we first downloaded co-expressed genes of *PSMD2* from cBioPortal. As shown in **Supplementary Table S2**, with |cor Spearman| > 0.5 and p < 0.05, *PSMD2* had 139 co-expressed genes, including 137 positive and two negative co-expressed genes. R package clusterProfiler was utilized for GO and KEGG analysis. As shown in **Figure 7A**, GO analysis revealed that these co-expressed genes of *PSMD2* were involved in the biological progress of organelle fission, chromosome segregation, and regulation of cell cycle phase transition. They acted as structural constituents in the condensed chromosome, spindle, and chromosomal region (**Figure 7B**), and played an important part in the molecular function of ATPase activity, tubulin binding, and DNA helicase activity (**Figure 7C**). KEGG pathway analysis in **Figure 7D** indicated enrichment function in cell cycle, RNA transport, and cellular senescence.

The Correlation Between PSMD2 Expression and Cancer Functional States

To capture the expression of *PSMD2* at single-cell resolution and its correlation with cancer functional states, we conducted an analysis on CancerSEA. The results listed in **Figure 8A** suggested that *PSMD2* expression was significantly positively correlated with angiogenesis and DNA damage, while negatively correlated with proliferation in acute lymphoblastic leukemia (ALL). *PSMD2* expression was positively correlated with cell cycle, DNA damage, and DNA repair, while negatively correlated with quiescence in lung adenocarcinoma (**Figure 8B**). The t-SNE diagram in **Figure 8C** described the *PSMD2* expression profile in single cells of lung adenocarcinoma.

PSMD2 Was Correlated With Immune Infiltration in Lung Adenocarcinoma

To analyze the correlation between the expression of *PSMD2* and immune cell infiltration, we conducted ssGSEA method with R package GSVA. The results of Spearman correlation analysis were listed in **Figures 9A–G** and **Table 2**. It was apparent that *PSMD2* expression was negatively correlated with the immune cell infiltration levels of CD8 T cells (r = −0.238, p < 0.001), B cells (r = −0.225, p < 0.001), pDC (r = −0.194, p < 0.001), mast cells (r = −0.191, p < 0.001), TFH (r = −0.190, p < 0.001), Tcells (r = −0.160, p < 0.001), Th17 cells (r = −0.143, p < 0.001), NK CD56bright cells (r = −0.128, p < 0.001), cytotoxic cells (r = −0.122, p < 0.001), and positively correlated with the immune cell infiltration levels of Th2 cells (r = 0.438, p < 0.001), Tgd (r = 0.146, p < 0.001), aDC (r = 0.142, p < 0.001), NK CD56dim cells (r = 0.121, p < 0.001), and NK cells (r = 0.142, p < 0.001). On the whole, these results suggest that *PSMD2* may regulate the level of tumor-infiltrating immune cells to affect lung adenocarcinoma progression.

The Relations Between PSMD2 Expression and Abundance of Tumor-Infiltrating Lymphocytes in Tumor-Immune system interaction database

To further confirm the relations between *PSMD2* expression and TILs, we performed an analysis on TISIDB. As shown in **Figure 10A**, we found that *PSMD2* expression was negatively correlated with abundance of eosinophil cells (r = −0.39, p < 2.2e-16), mast cells (r = −0.267, p = 7.49e-10), activated B cells (r = −0.262, p = 1.79e-09), Th17 (r = −0.216, p = 7.23e-07), immature B cells (r = −0.204, p = 3.03e-06), tem CT8 T cells (r = −0.191, p = 1.28e-05), pDC (r = −0.169, p = 1.19e-04), macrophage cells (r = −0.163, p = 2.09e-04), and neutrophil cells (r = −0.152, p = 5.34e-04), NK cells (r = −0.136, p = 1.89e-03), TFH (r = −0.117, p = 7.59e-03), Th1 (r = −0.106, p = 0.016). *PSMD2* expression was negatively correlated with abundance of activated CD4 cells (r = 0.284, p = 6.14e-11), CD56dim cells (r = 0.107, p = 1.46e-02). The results were similar to **Figure 9**. Scatter diagrams for correlation between *PSMD2* expression and abundance of eosinophil cells, mast cells, activated CD4 cells and CD56dim cells were listed in **Figures 10B–E**. Taken together, these results indicate that *PSMD2* expression is correlated with most TILs in lung adenocarcinoma, further suggesting that *PSMD2* may play an important role in the lung adenocarcinoma microenvironment.

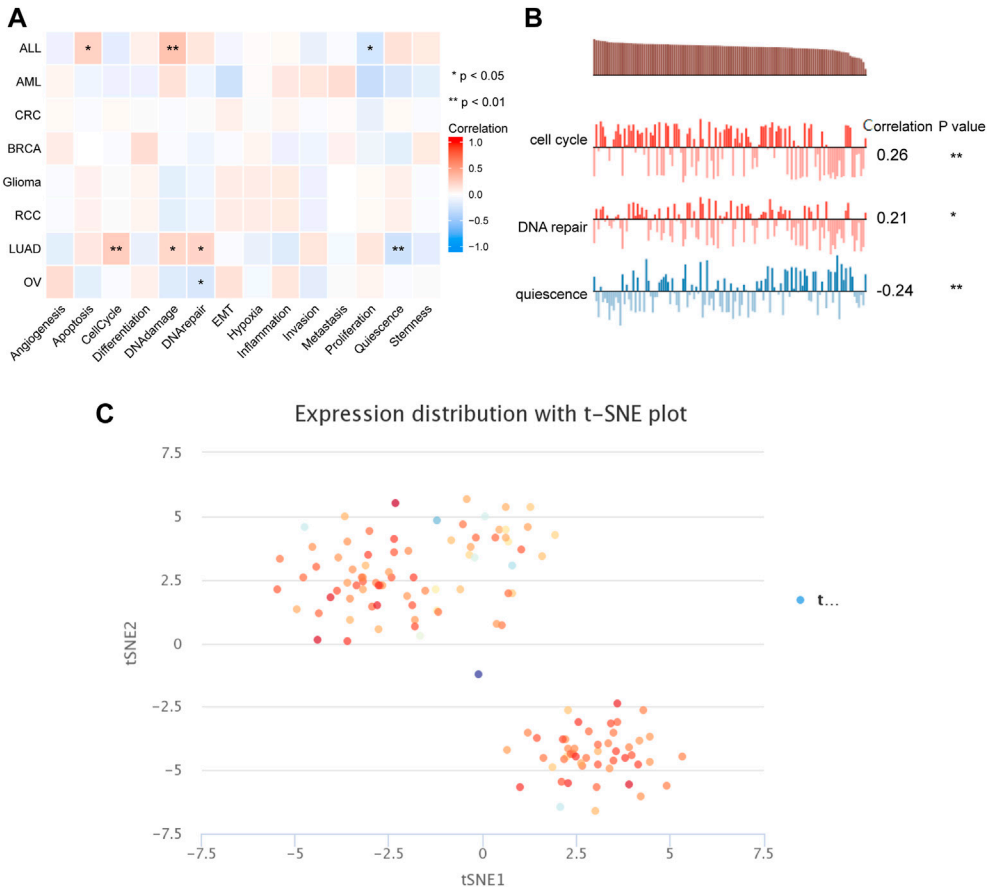


FIGURE 8 | The correlation between *PSMD2* expression and cancer functional states. **(A)** A heatmap indicated the relationship between *PSMD2* expression and cancer functional states from CancerSEA. **(B)** *PSMD2* expression was correlated with cell cycle, DNA repair, and quiescence in lung adenocarcinoma. **(C)** The t-SNE diagram described the *PSMD2* expression profile in single cells of lung adenocarcinoma.

DISCUSSION

In this study, we reported that the expression level of *PSMD2* is significantly elevated in lung adenocarcinoma and its upregulation is a reliable predictor of high T stage, lymph node metastases, and high TNM stage. In light of the PrognoScan database, Kaplan-Meier curves, and multivariate Cox analysis, our results confirmed that high expression of *PSMD2* is correlated with poor prognosis and *PSMD2* is an independent prognostic biomarker for overall survival of lung adenocarcinoma patients. Our results further indicated that the genetic mutation of *PSMD2* was also correlated with poor overall survival, disease-specific survival, and progression-free survival in lung adenocarcinoma. The genetic mutation of *PSMD2* was also correlated with poor overall survival, disease-specific survival, and progression-free survival in lung adenocarcinoma. Moreover, immune infiltration analysis suggested that *PSMD2* expression had a significant correlation with the level of tumor-infiltrating immune cells, further suggesting a specific role for *PSMD2* in the immunological interactions in lung adenocarcinoma.

As a member of The *PSMD* gene family, *PSMD2* has been characterized as an important non-ATPase regulatory subunit of the 19S proteasome (Li et al., 2019). Previous studies elucidated that proteasome can induce ubiquitination and degradation of protein, which in turn play an important role in the development of cell proliferation, apoptosis, and cell cycle (Chesnel et al., 2006). Functioning as a component of the proteasome, *PSMD2* was reported to play a vital role in tumor progression. In colorectal cancer, *PSMD2* can facilitate the degradation of diverse Ras-related GTPase and then inhibit cell proliferation and affect the expression of cell-cycle protein *via* blocking NF-kappaB signaling (Ying et al., 2022). In breast cancer, a paper from Li *et al.* reported that *PSMD2* can interact with p21 and p27, mediate the degradation of their ubiquitin-proteasome, and then promote cell proliferation and cell cycle progression in breast cancer (Oguro et al., 2015). In hepatocellular carcinoma, *PSMD2* can modulate cellular lipid metabolism to regulate HepG2 cell proliferation *via* p38-JNK and AKT signaling (Tan et al., 2019). In the present study, our results showed that the mRNA and protein expression of *PSMD2* is elevated in lung adenocarcinoma tissues. Moreover, we analyzed the correlation between *PSMD2* expression and the clinicopathological factors of lung adenocarcinoma patients. The current study suggests that high expression of *PSMD2* is significantly correlated with high T stage, N stage, and TNM stage. Given our results, it is likely that *PSMD2* is involved in tumorigenesis and metastasis of lung adenocarcinoma. Furthermore, both functional

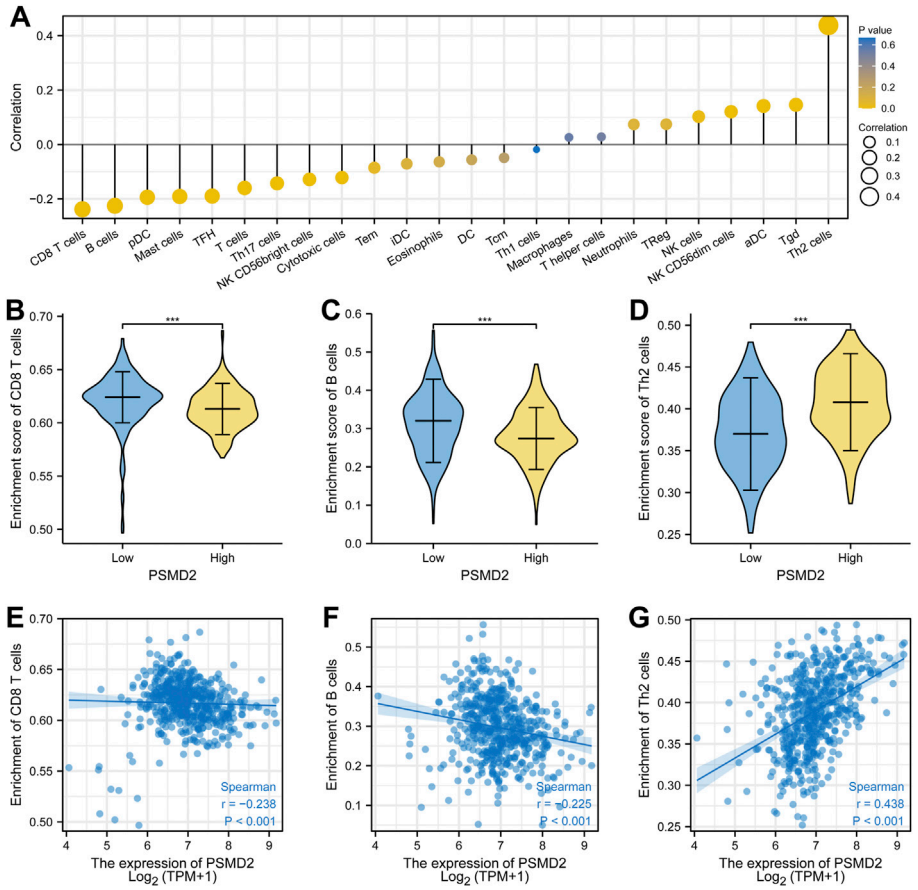


FIGURE 9 | Spearman correlation between *PSMD2* expression and immune cell infiltration. **(A)** The correlation between *PSMD2* expression and the relative abundances of 24 immune cells. **(B–D)** Mann-Whitney U test indicated there were differences in immune cells (CD8 T cells, B cells, Th2 cells) between high/low expression of *PSMD2*. **(E–G)** Scatter diagrams for correlation between immune cells (CD8 T cells, B cells, Th2 cells) and *PSMD2* expression.

enrichment analysis and CancerSEA results indicate *PSMD2* expression is correlated with cell cycle, further suggesting that *PSMD2* can regulate cell cycle to promote cancer progression. However, this should be tested in other experiments.

Previous studies demonstrated that upregulation of *PSMD2* is correlated with poor prognosis in many cancers. A paper from established that patients with high *PSMD2* expression have poor overall survival and progression-free survival in bladder urothelial carcinoma Salah Fararjeh et al. (2021). Based on the result of univariate and multivariate analysis, *PSMD2* has been identified as an independent prognostic biomarker for overall survival in bladder urothelial carcinoma (Salah Fararjeh et al., 2021). In breast cancer, reported that upregulation of *PSMD2* is associated with shorter overall survival and distant-metastasis-free survival, further suggesting that *PSMD2* could act as a factor for an unfavorable prognosis in breast cancer Li et al. (2018). Our findings on the prognostic value of *PSMD2* in lung adenocarcinoma are consistent with these reports. In this study, our results indicate that lung adenocarcinoma patients with high *PSMD2* expression have poor overall survival and progression-free survival. Univariate and multivariate analysis found that *PSMD2* is an independent prognostic factor for overall survival in lung adenocarcinoma patients. Furthermore, mutation characteristics of *PSMD2* from the cBioPortal database also suggest that the altered group is associated with poor prognosis in overall survival, disease-specific survival, and progression-free survival in lung adenocarcinoma. Based on our data, we conclude that *PSMD2* is a potential biomarker for poor prognosis in lung adenocarcinoma.

Many studies have described that tumor-infiltrating immune cells are the representative cellular components and play an important part in host antitumor immune responses (Oguro et al., 2015). There is also some evidence that infiltrating immune cells is closely associated with the efficacy of immunotherapy (Zhang et al., 2020). The findings indicate that compared with high immune cell infiltration, patients with low immune cell infiltration levels may appear poor outcomes to conventional therapy in some solid tumors (Tobin et al., 2019). A paper from reported that the expression of *PSMD* member genes including *PSMD2* is related to markers of six tumor-infiltrating immune cell types in breast cancer Xuan et al. (2021). In this study, we analyzed the correlation between the expression of *PSMD2* and immune cell infiltration. By ssGSEA method, we found that *PSMD2*

TABLE 2 | The correlation between PSMD2 expression and immune infiltration.

Gene	Immune cells	Pearson correlation		Spearman correlation	
		R	*p*-value	R	*p*-value
PSMD2	aDC	0.174	<0.001***	0.142	0.001**
PSMD2	B cells	−0.189	<0.001***	−0.225	<0.001***
PSMD2	CD8 T cells	−0.032	0.466	−0.238	<0.001***
PSMD2	Cytotoxic cells	−0.057	0.185	−0.122	0.005**
PSMD2	DC	0.035	0.423	−0.056	0.194
PSMD2	Eosinophils	−0.018	0.684	−0.064	0.140
PSMD2	iDC	0.007	0.868	−0.071	0.100
PSMD2	Macrophages	0.038	0.384	0.026	0.547
PSMD2	Mast cells	−0.143	<0.001***	−0.191	<0.001***
PSMD2	Neutrophils	0.125	0.004**	0.074	0.088
PSMD2	NK CD56bright cells	0.015	0.733	−0.128	0.003**
PSMD2	NK CD56dim cells	0.178	<0.001***	0.121	0.005**
PSMD2	NK cells	0.275	<0.001***	0.102	0.018*
PSMD2	pDC	−0.142	0.001**	−0.194	<0.001***
PSMD2	T cells	−0.139	0.001**	−0.160	<0.001***
PSMD2	T helper cells	−0.019	0.664	0.028	0.518
PSMD2	Tcm	−0.171	<0.001***	−0.049	0.258
PSMD2	Tem	−0.103	0.018*	−0.086	0.048*
PSMD2	TFH	−0.121	0.005**	−0.190	<0.001***
PSMD2	Tgd	0.172	<0.001***	0.146	<0.001***
PSMD2	Th1 cells	0.014	0.742	−0.019	0.668
PSMD2	Th17 cells	−0.158	<0.001***	−0.143	<0.001***
PSMD2	Th2 cells	0.431	<0.001***	0.438	<0.001***
PSMD2	TReg	0.152	<0.001***	0.075	0.085

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

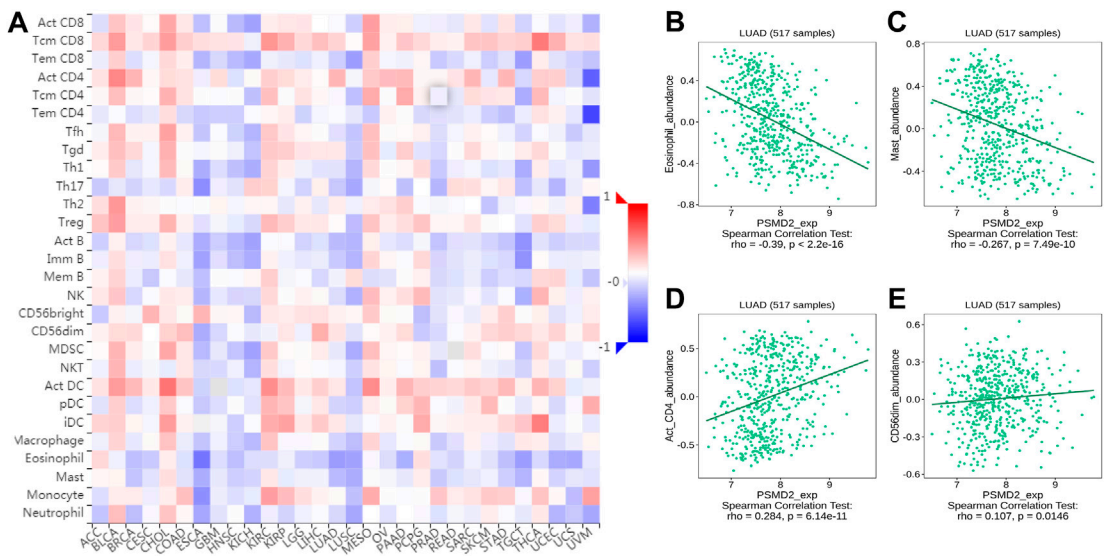


FIGURE 10 | The relations between *PSMD2* expression and abundance of TILs in TISIDB. **(A)** Relations between *PSMD2* expression and 28 types of TILs across human cancers. **(B–E)** Scatter diagrams for correlation between *PSMD2* expression and abundance of eosinophil cells, mast cells, activated CD4 cells, and CD56dim cells.

expression was negatively correlated with the immune cell infiltration levels of CD8 T cells, B cells, pDC, mast cells, TFH, T cells, Th17 cells, NK CD56bright cells, and cytotoxic cells. Moreover, *PSMD2* expression is positively correlated with the immune cell infiltration levels of Th2 cells, Tgd, aDC, NK CD56dim cells, and NK cells. To further confirm the relations between *PSMD2* expression and TILs, we performed an analysis on TISIDB. The results from TISIDB were similar to the ssGSEA method. Our findings suggest that *PSMD2* expression is correlated with immune cell infiltration and

raise the possibility that *PSMD2* can be a potential immunotherapy target in lung adenocarcinoma. However, this hypothesis should be tested with further research.

In conclusion, our findings showed that *PSMD2* expression is significantly elevated in lung adenocarcinoma and its upregulation is a reliable predictor of high T stage, lymph node metastases, and high TNM stage. Our results confirmed that high expression of *PSMD2* is correlated with poor prognosis and *PSMD2* is an independent prognostic biomarker for lung adenocarcinoma patients. Moreover, *PSMD2* expression had a significant correlation with the level of tumor-infiltrating immune cells.

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

GL designed the study and contributed analysis tools. HZ performed data analysis and wrote the manuscript. All authors reviewed the manuscript.

FUNDING

This study was funded by Wu JiePing Medical Foundation (320.6750.19059).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.905581/full#supplementary-material

REFERENCES

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., et al. (2013). Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the Immune Landscape in Human Cancer. *Immunity* 39 (4), 782–795. doi:10.1016/j.immuni.2013.10.003

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer Discov.* 2 (5), 401–404. doi:10.1158/2159-8290.CD-12-0095

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

Chesnel, F., Bazile, F., Pascal, A., and Kubiak, J. Z. (2006). Cyclin B Dissociation from CDK1 Precedes its Degradation upon MPF Inactivation in Mitotic Extracts of *Xenopus laevis* Embryos. *Cell. Cycle* 5 (15), 1687–1698. doi:10.4161/cc.5.15.3123

Edwards, N. J., Oberti, M., Thangudu, R. R., Cai, S., McGarvey, P. B., Jacob, S., et al. (2015). The CPTAC Data Portal: A Resource for Cancer Proteomics Research. *J. Proteome Res.* 14 (6), 2707–2713. doi:10.1021/pr501254j

Fararjeh, A. S., Chen, L. C., Ho, Y. S., Cheng, T. C., Liu, Y. R., Chang, H. L., et al. (2019). Proteasome 26S Subunit, Non-ATPase 3 (PSMD3) Regulates Breast Cancer by Stabilizing HER2 from Degradation. *Cancers* 11 (4), 527. doi:10.3390/cancers11040527

Gandhi, L., Rodríguez-Abreu, D., Gadgeel, S., Esteban, E., Felip, E., De Angelis, F., et al. (2018). Pembrolizumab Plus Chemotherapy in Metastatic Non-small-cell Lung Cancer. *N. Engl. J. Med.* 378 (22), 2078–2092. doi:10.1056/NEJMoa1801005

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., et al. (2013). Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci. Signal.* 6 (269), pl1. doi:10.1126/scisignal.2004088

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

Hirsch, F. R., Scagliotti, G. V., Mulshine, J. L., Kwon, R., Curran, W. J., Jr, Wu, Y.-L., et al. (2017). Lung Cancer: Current Therapies and New Targeted Treatments. *Lancet* 389 (10066), 299–311. doi:10.1016/S0140-6736(16)30958-8

Hua, X., Zhao, W., Pesatori, A. C., Consonni, D., Caporaso, N. E., Zhang, T., et al. (2020). Genetic and Epigenetic Intratumor Heterogeneity Impacts Prognosis of Lung Adenocarcinoma. *Nat. Commun.* 11 (1), 2459. doi:10.1038/s41467-020-16295-5

Huo, J., Xu, Y., Sheu, T., Volk, R. J., and Shih, Y.-C. T. (2019). Complication Rates and Downstream Medical Costs Associated with Invasive Diagnostic Procedures for Lung Abnormalities in the Community Setting. *JAMA Intern Med.* 179 (3), 324–332. doi:10.1001/jamainternmed

Köster, F., Sauer, L., Hoellen, F., Ribbat-Idel, J., Bräutigam, K., Rody, A., et al. (2020). PSMD9 Expression Correlates with Recurrence after Radiotherapy in Patients with Cervical Cancer. *Oncol. Lett.* 20 (1), 581–588. doi:10.3892/ol.2020.11622

Langlands, F. E., Dodwell, D., Hanby, A. M., Horgan, K., Millican-Slater, R. A., Speirs, V., et al. (2014). PSMD9 Expression Predicts Radiotherapy Response in Breast Cancer. *Mol. Cancer* 13, 73. doi:10.1186/1476-4598-13-73

Lee, J. J.-K., Park, S., Park, H., Kim, S., Lee, J., Lee, J., et al. (2019). Tracing Oncogene Rearrangements in the Mutational History of Lung Adenocarcinoma. *Cell.* 177 (7), 1842–1857. doi:10.1016/j.cell.2019.05.013

Li, C., Hu, J., Hu, X., Zhao, C., Mo, M., Zu, X., et al. (2021). LncRNA SNHG9 Is a Prognostic Biomarker and Correlated with Immune Infiltrates in Prostate Cancer. *Transl. Androl. Urol.* 10 (1), 215–226. doi:10.21037/tau-20-1134

Li, P., Li, H., Zhao, Y., Guo, Q., Yu, Y., Zhu, S., et al. (2019). Asporin Promotes Cell Proliferation via Interacting with PSMD2 in Gastric Cancer. *Front. Biosci.* 24 (6), 1178–1189. doi:10.2741/4774

Li, Y., Huang, J., Zeng, B., Yang, D., Sun, J., Yin, X., et al. (2018). PSMD2 Regulates Breast Cancer Cell Proliferation and Cell Cycle Progression by Modulating P21 and P27 Proteasomal Degradation. *Cancer Lett.* 430, 109–122. doi:10.1016/j.canlet.2018.05.018

Ma, A. G., Yu, L. M., Zhao, H., Qin, C. W., Tian, X. Y., and Wang, Q. (2019). PSMD4 Regulates the Malignancy of Esophageal Cancer Cells by Suppressing Endoplasmic Reticulum Stress. *Kaohsiung J. Med. Sci.* 35 (10), 591–597. doi:10.1002/kjm2.12093

Matsuyama, Y., Suzuki, M., Arima, C., Huang, Q. M., Tomida, S., Takeuchi, T., et al. (2011). Proteasomal Non-catalytic Subunit PSMD2 as a Potential Therapeutic Target in Association with Various Clinicopathologic Features in Lung Adenocarcinomas. *Mol. Carcinog.* 50 (4), 301–309. doi:10.1002/mc.20632

Mizuno, H., Kitada, K., Nakai, K., and Sarai, A. (2009). PrognoScan: a New Database for Meta-Analysis of the Prognostic Value of Genes. *BMC Med. Genomics* 2, 18. doi:10.1186/1755-8794-2-18

Mok, T. S. K., Wu, Y. L., Kudaba, I., Kowalski, D. M., Cho, B. C., Turna, H. Z., et al. (2019). Pembrolizumab versus Chemotherapy for Previously Untreated, PD-L1-Expressing, Locally Advanced or Metastatic Non-small-cell Lung Cancer (KEYNOTE-042): a Randomised, Open-Label, Controlled, Phase 3 Trial. *Lancet* 393 (10183), 1819–1830. doi:10.1016/S0140-6736(18)32409-7

Oguro, S., Ino, Y., Shimada, K., Hatanaka, Y., Matsuno, Y., Esaki, M., et al. (2015). Clinical Significance of Tumor-infiltrating Immune Cells Focusing on BTLA and Cbl-b in Patients with Gallbladder Cancer. *Cancer Sci.* 106 (12), 1750–1760. doi:10.1111/cas.12825

Okumura, T., Ikeda, K., Ujihira, T., Okamoto, K., Horie-Inoue, K., Takeda, S., et al. (2018). Proteasome 26S Subunit PSMD1 Regulates Breast Cancer Cell Growth through P53 Protein Degradation. *J. Biochem.* 163 (1), 19–29. doi:10.1093/jb/mvx053

Pan, J.-h., Zhou, H., Cooper, L., Huang, J.-l., Zhu, S.-b., Zhao, X.-x., et al. (2019). LAYN Is a Prognostic Biomarker and Correlated with Immune Infiltrates in Gastric and Colon Cancers. *Front. Immunol.* 10, 6. doi:10.3389/fimmu.2019.00006

Ru, B., Wong, C. N., Tong, Y., Zhong, J. Y., Zhong, S. S. W., Wu, W. C., et al. (2019). TISIDB: an Integrated Repository Portal for Tumor-Immune System Interactions. *Bioinformatics* 35 (20), 4200–4202. doi:10.1093/bioinformatics/btz210

Salah Fararjeh, A., Al-Khader, A., Al-Saleem, M., and Abu Qauod, R. (2021). The Prognostic Significance of Proteasome 26S Subunit, Non-ATPase (PSMD) Genes for Bladder Urothelial Carcinoma Patients. *Cancer Inf.* 20, 117693512110676. doi:10.1177/11769351211067692

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

Tan, Y., Jin, Y., Wu, X., and Ren, Z. (2019). PSMD1 and PSMD2 Regulate HepG2 Cell Proliferation and Apoptosis via Modulating Cellular Lipid Droplet Metabolism. *BMC Mol. Biol.* 20 (1), 24. doi:10.1186/s12867-019-0141-z

Tobin, J. W. D., Keane, C., Gunawardana, J., Mollee, P., Birch, S., Hoang, T., et al. (2019). Progression of Disease within 24 Months in Follicular Lymphoma Is Associated with Reduced Intratumoral Immune Infiltration. *J. Clin. Oncol.* 37 (34), 3300–3309. doi:10.1200/JCO.18.02365

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

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2015). Proteomics. Tissue-based Map of the Human Proteome. *Science* 347 (6220), 1260419. doi:10.1126/science.1260419

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

Waniczek, D., Lorenc, Z., Śnietura, M., Wesecki, M., Kopec, A., and Muc-Wierzgoń, M. (2017). Tumor-Associated Macrophages and Regulatory T Cells Infiltration and the Clinical Outcome in Colorectal Cancer. *Arch. Immunol. Ther. Exp.* 65 (5), 445–454. doi:10.1007/s00005-017-0463-9

Xia, C., Dong, X., Li, H., Cao, M., Sun, D., He, S., et al. (2022). Cancer Statistics in China and United States, 2022: Profiles, Trends, and Determinants. *Chin. Med. J. Engl.* 135 (5), 584–590. doi:10.1097/CM9.0000000000002108

Xuan, D. T. M., Wu, C.-C., Kao, T.-J., Ta, H. D. K., Anuraga, G., Andriani, V., et al. (2021). Prognostic and Immune Infiltration Signatures of Proteasome 26S Subunit, Non-ATPase (PSMD) Family Genes in Breast Cancer Patients. *Aging* 13 (22), 24882–24913. doi:10.18632/aging.203722

Ying, K., Wang, C., Liu, S., Kuang, Y., Tao, Q., and Hu, X. (2022). Diverse Ras-Related GTPase DIRAS2, Downregulated by PSMD2 in a Proteasome-Mediated Way, Inhibits Colorectal Cancer Proliferation by Blocking NF-κB Signaling. *Int. J. Biol. Sci.* 18 (3), 1039–1050. doi:10.7150/ijbs.68312

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

Yuan, H., Yan, M., Zhang, G., Liu, W., Deng, C., Liao, G., et al. (2019). CancerSEA: a Cancer Single-Cell State Atlas. *Nucleic Acids Researc* 47 (D1), D900–D908. doi:10.1093/nar/gky939

Zhang, S., Yang, H., Xiang, X., Liu, L., Huang, H., and Tang, G. (2022). THBS2 Is Closely Related to the Poor Prognosis and Immune Cell Infiltration of Gastric Cancer. *Front. Genet.* 13, 803460. doi:10.3389/fgene.2022.803460

Zhang, W., Shen, Y., Huang, H., Pan, S., Jiang, J., Chen, W., et al. (2020). A Rosetta Stone for Breast Cancer: Prognostic Value and Dynamic Regulation of Neutrophil in Tumor Microenvironment. *Front. Immunol.* 11, 1779. doi:10.3389/fimmu.2020.01779

Zhao, Y., Yang, X., Xu, X., Zhang, J., Zhang, L., Xu, H., et al. (2020). Deubiquitinase PSMD7 Regulates Cell Fate and Is Associated with Disease Progression in Breast Cancer. *Am. J. Transl. Res.* 12 (9), 5433

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 June 2022
doi: 10.3389/fgene.2022.870302

Mitochondrial Homeostasis–Related lncRNAs are Potential Biomarkers for Predicting Prognosis and Immune Response in Lung Adenocarcinoma

Bo Peng[1], Han Lou[2], Chen Chen[2], Lei Wang[2], Huawei Li[1], Tong Lu[1], Ruisi Na[3], Ran Xu[1], Tong Xin[4], Lingqi Yao[1], Henghui Xu[2], Kaiyu Wang[1], Xin Liu[2] and Linyou Zhang[1]**

[1]Department of Thoracic Surgery, The Second Affiliated Hospital of Harbin Medical University, Harbin, China, [2]Department of Pharmacology (State-Province Key Laboratories of Biomedicine-Pharmaceutics of China and Key Laboratory of Cardiovascular Medicine Research, Ministry of Education), College of Pharmacy, Harbin Medical University, Harbin, China, [3]Department of Gastrointestinal Medical Oncology, Harbin Medical University Cancer Hospital, Harbin, China, [4]The Fourth Department of Medical Oncology, Harbin Medical University Cancer Hospital, Harbin, China

OPEN ACCESS

Edited by:
Matthew B. Schabath,
Moffitt Cancer Center, United States

Reviewed by:
Xingguo Liu,
Guangzhou Institutes of Biomedicine and Health (CAS), China
Yuvabharath Kondaveeti,
University of North Carolina at Chapel Hill, United States

***Correspondence:**
Xin Liu
freyaliuxin@163.com
Linyou Zhang
lyzhang@hrbmu.edu.cn

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: 11 February 2022
Accepted: 11 May 2022
Published: 13 June 2022

Citation:
Peng B, Lou H, Chen C, Wang L, Li H, Lu T, Na R, Xu R, Xin T, Yao L, Xu H, Wang K, Liu X and Zhang L (2022) Mitochondrial Homeostasis–Related lncRNAs are Potential Biomarkers for Predicting Prognosis and Immune Response in Lung Adenocarcinoma. Front. Genet. 13:870302. doi: 10.3389/fgene.2022.870302

The prognosis of the most common histological subtype of lung cancer, lung adenocarcinoma (LUAD), is relatively poor. Mitochondrial homeostasis depends to a great extent on the coordination between mitophagy and mitochondrial biogenesis, the deregulation of which causes various human diseases, including cancer. There is accumulating evidence that long noncoding RNAs (lncRNAs) are critical in predicting the prognosis and immune response in carcinoma. Therefore, it is critical to discern lncRNAs related to mitochondrial homeostasis in LUAD patients. In this study, we identified mitochondrial homeostasis–related lncRNAs (MHRlncRNAs) by coexpression analysis. In order to construct a prognostic signature composed of three MHRlncRNAs, univariate and multivariate Cox regression analyses were performed. Kaplan–Meier analysis, stratification analysis, principal component analysis (PCA), receiver operating characteristic (ROC) curve, gene set enrichment analysis (GSEA), and nomogram were applied to evaluate and optimize the risk model. Subsequently, we identified the mitochondrial homeostasis–related lncRNA signature (MHLncSig) as an independent predictive factor of prognosis. Based on the LUAD subtypes regrouped by this risk model, we further investigated the underlying tumor microenvironment, tumor mutation burden, and immune landscape behind different risk groups. Likewise, individualized immunotherapeutic strategies and candidate compounds were screened to aim at different risk subtypes of LUAD patients. Finally, we validated the expression trends of lncRNAs included in the risk model using quantitative real-time polymerase chain reaction (qRT-PCR) assays. The established MHLncSig may be a promising tool for predicting the prognosis and guiding individualized treatment in LUAD.

Keywords: mitochondrial homeostasis, long noncoding RNAs, lung adenocarcinoma, nomogram, prognosis, immunotherapy

INTRODUCTION

As the most commonly diagnosed pathological subtype of lung cancer, the incidence of lung adenocarcinoma (LUAD) is increasing globally every year. LUAD patients have poor prognosis which commonly results from late detection and individual treatment differences. Compared with the conventional treatments including surgery and chemotherapy, the development of multiple agents targeting driver gene mutations showed appreciable promise in LUAD treatment due to the advances of cancer genomics (Denisenko et al., 2018; Joseph et al., 2018). Unfortunately, secondary mutations in tumors usually contribute to targeted therapeutic resistance (Liu et al., 2018). In recent years, advance of cancer immunology makes immunotherapy a hotspot in the clinical treatment of LUAD. Classical immune checkpoint inhibitors (ICIs) including antiprogrammed cell death 1 (PD-1) and antiprogrammed cell death-ligand 1 (PD-L1) agents exert a persisting and powerful antitumor effect in LUAD patients (Forde et al., 2018). However, only a proportion of patients could benefit from immunotherapy because of the relatively low overall response rate of ICI (Li et al., 2018a). For the aforementioned reasons, the 5-year survival rate of LUAD remains inadequate (Wang et al., 2021; Zhu et al., 2022). Therefore, novel molecular biomarkers are urgently required to predict the prognosis and therapeutic response for LUAD patients.

Mitochondria are essential organelles that regulate ATP production and energy transformation *via* oxidative phosphorylation (OXPHOS) and the tricarboxylic acid cycle (TCA), which also modulate iron metabolism, Ca^{2+} signaling, innate immunity, and apoptotic cell death in mammalian cells (Zong et al., 2016; Pathak and Trebak, 2018). Mitochondria, generally defined as highly motile and plastic organelles, constantly undergo processes of fusion and fission and update through mitophagy and mitochondrial biogenesis to maintain homeostasis (Ma et al., 2020). In most cases, mitochondrial DNA (mtDNA) mutations, deleted, or damaged DNA replication induce the dysbiosis of mitochondrial homeostasis and consequent mitochondrial dysfunction (Butow and Avadhani, 2004; Wallace, 2012). Mitochondrial dysfunction caused by significant abnormalities in the mtDNA copy number is intimately associated with many diseases such as age-related pathologies, mtDNA depletion syndrome, and several carcinomas (Park et al., 2009; Imanishi et al., 2011; Greaves et al., 2012). The results from previous studies indicated that low copy numbers of mtDNA are widely observed in multiple cancers including colon, breast, hepatocellular carcinomas, prostate cancer, and astrocytoma (Horton et al., 1996; Lee et al., 2005; Petros et al., 2005; Tseng et al., 2006; Correia et al., 2011). Moreover, mtDNA depletion induced by experimental methods promotes aggressive phenotype in prostate and colorectal cancer cells (Moro et al., 2009; Guo et al., 2011).

Long noncoding RNAs (lncRNAs), a subclass of noncoding RNAs with a length of >200 nucleotides, have been confirmed to be involved in the tumorigenesis and progression of various tumors, including LUAD (Zheng et al., 2021). Recent studies have revealed that dysregulation of specific lncRNAs was inextricably associated with tumor proliferation, metastasis, and drug resistance in lung cancer (Mao et al., 2018; Wang et al., 2019a; Yang et al., 2020). Nonetheless, the specific function of lncRNAs in mitochondrial homeostasis remains to be clarified. Therefore, investigating the potential mechanism of mitochondrial homeostasis–related lncRNAs (MHRlncRNAs) in LUAD may be valuable for prognostic biomarkers.

In our study, we extracted the expression profiles of 1,499 mitochondrial homeostasis-related genes (MHRGs) from the publicly available dataset: The Cancer Genome Atlas (TCGA) dataset. Using Pearson's correlation analysis, we identified 2,850 lncRNAs coexpressed with MHRGs. Next, mitochondrial homeostasis–related lncRNA signature (MHLncSig) was constructed to forecast the survival of LUAD patients utilizing differential expression analysis, univariate and multivariate Cox regression analyses. This personalized and robust prognostic signature is not only an independent indicator of overall survival, but also significantly related to tumor microenvironment, tumor mutation burden (TMB), immune infiltration, immunotherapeutic efficacy, and drug sensitivity.

MATERIALS AND METHODS

Data Acquisition and Study Design

We downloaded the RNA transcriptome sequence data, corresponding clinical features, and mutation data for the patients with LUAD from The Cancer Genome Atlas (TCGA) (https://cancergenome.nih.gov/) database. The patients with missing survival information were removed from this study. The study flowchart is presented in **Figure 1**.

Selection of Mitochondrial Homeostasis-Related Genes and lncRNAs

1,499 MHRGs with the relevance score >4 were obtained from the GeneCards (https://www.genecards.org/) database, including mitochondrial DNA-encoded genes and genomic DNA-encoded genes that regulate mitochondrial homeostasis *via* multiple pathways. We extracted the expression profile of MHRGs from the downloaded transcriptome data and performed differential expression analysis between the normal samples and LUAD samples. 294 differentially expressed MHRGs (|Log fold change (FC)| > 1 and false discovery rate (FDR) adjusted $p < 0.05$) were considered to be closely related to the pathogenesis and progression of LUAD. Next, we used Pearson's correlation analysis and identified 2,850 MHRlncRNAs (|Pearson R| > 0.4 and $p < 0.001$). Further differential expression analysis based on 2,850 MHRlncRNAs was conducted to discern candidate lncRNAs for model building.

Establishment of the Risk Signature

We identified 131 differentially expressed MHRlncRNAs as candidate lncRNAs from the 2,850 MHRlncRNAs (|log FC| > 1 and FDR <0.05). The TCGA set was randomly separated into training (252 patients) and testing sets (252 patients), and no

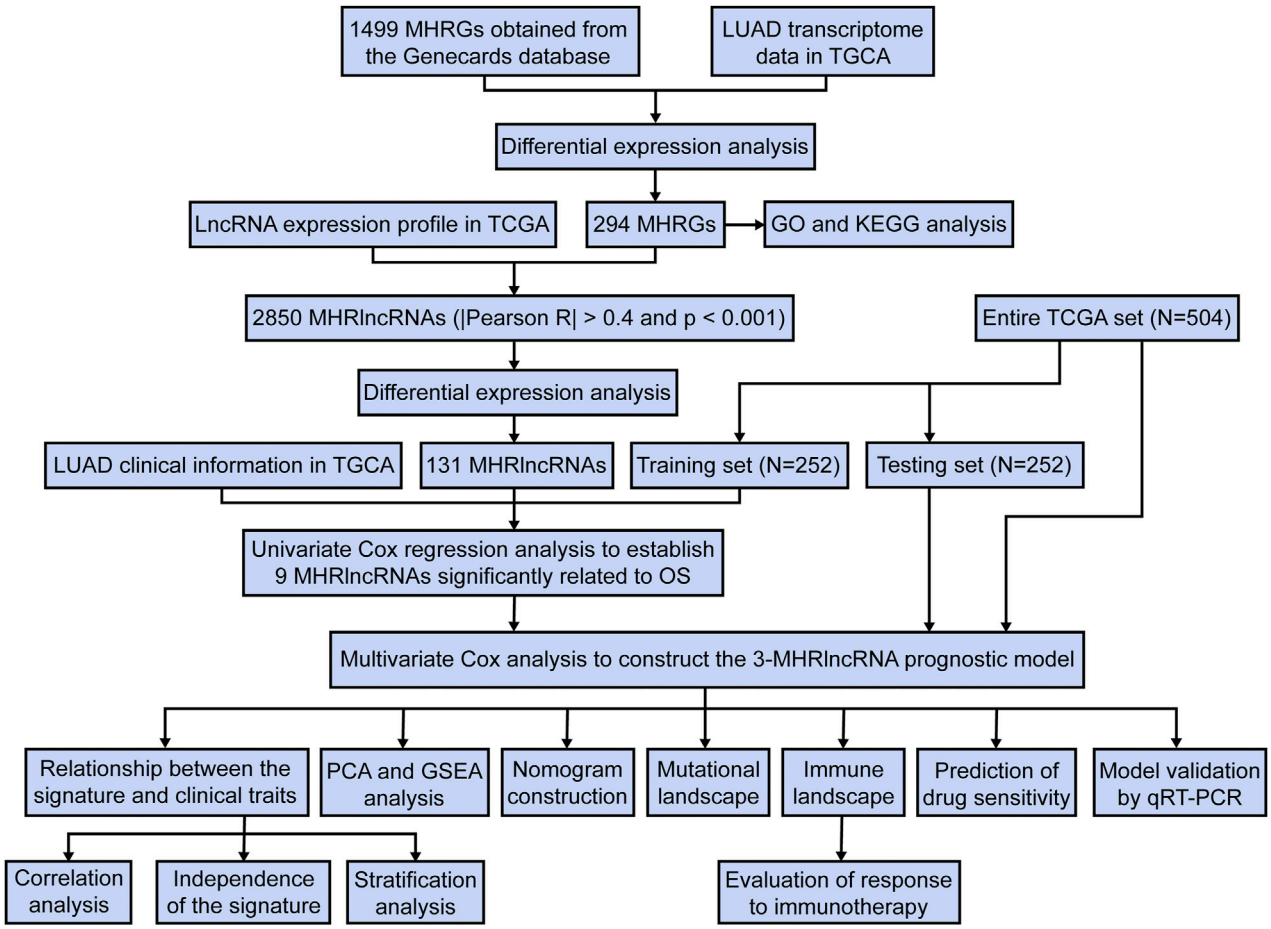


FIGURE 1 | A concise flow diagram of this study.

TABLE 1 | Clinical features of the three LUAD patients sets.

Covariate		Training set (*n* = 252)	Testing set (*n* = 252)	TCGA set (*n* = 504)	*p*-value
Age, no (%)	≤65	117 (46.43)	121 (48.02)	238 (47.22)	0.721
	>65	131 (51.98)	125 (49.6)	256 (50.79)	
	unknown	4 (1.59)	6 (2.38)	10 (1.98)	
Gender, no (%)	female	129 (51.19)	141 (55.95)	270 (53.57)	0.326
	male	123 (48.81)	111 (44.05)	234 (46.43)	
T stage, no (%)	T1–2	217 (86.11)	220 (87.3)	437 (86.71)	0.701
	T3–4	34 (13.49)	30 (11.9)	64 (12.7)	
	unknown	1 (0.4)	2 (0.79)	3 (0.6)	
N stage, no (%)	N0	161 (63.89)	164 (65.08)	325 (64.48)	0.849
	N1–3	85 (33.73)	82 (32.54)	167 (33.13)	
	unknown	6 (2.38)	6 (2.38)	12 (2.38)	
M stage, no (%)	M0	169 (67.06)	166 (65.87)	335 (66.47)	1
	M1	13 (5.16)	12 (4.76)	25 (4.96)	
	unknown	70 (27.78)	74 (29.37)	144 (28.57)	
Pathologic stage, no (%)	I–II	191 (75.79)	198 (78.57)	389 (77.18)	0.629
	III–IV	56 (22.22)	51 (20.24)	107 (21.23)	
	unknown	5 (1.98)	3 (1.19)	8 (1.59)	

LUAD, lung adenocarcinoma.

significant differences in the clinical properties was observed. **Table 1** summarized the clinical features of these three sets. We used the training set to construct the prognostic signature The testing set and the entire TCGA set were utilized to validate the established signature. Univariate Cox regression analysis was performed to screen nine lncRNAs with significant prognostic value from 131 differentially expressed MHRlncRNAs. Multivariate Cox regression analysis was utilized to analyze and further confirm the prognostic significance from the previous steps, and three lncRNAs were retained during multiple computing. Thereafter, according to a linear combination of expression levels weighted with the regression coefficients calculated by multivariate Cox regression analysis, we established a risk formula that can calculate the risk score for each patient with LUAD

$$MHLncSig\,(risk\ score) = \sum_{i=1}^{n} coef\,ficient\,(\text{lncRNAi}) * expression\,level\,(\text{lncRNAi})\ ,$$

where n indicates the amount of MHRlncRNAs included in the prognostic model, and the multivariate Cox regression analysis provided the coefficient for lncRNAi. According to the median risk score, we further distinguished the high- and low-risk subgroups of LUAD.

Function Enrichment Analysis

To discern the potential biological roles and metabolic pathways in differentially expressed MHRGs, the Gene Ontology (GO) categories and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were identified in the R program. We also used the Gene Set Enrichment Analysis (GSEA) (version 4.1.0) to investigate the biological processes and metabolic pathways involved in the different risk subgroups of LUAD stratified by MHLncSig. In terms of the reference file, c2.cp.kegg.v7.4.symbols.gmt was used and FDR <0.05 was considered significant.

Tumor Microenvironment and Tumor Mutation Burden (TMB)

R "estimate" package was utilized to calculate the stromal score, immune score, and ESTIMATE score for each LUAD patient, which represents the infiltration of both the stromal and immune cells in tumor tissues. We downloaded the somatic mutation data from the TCGA database and calculated the TMB score for each LUAD sample using a Perl script.

Immune Landscape and Immunotherapeutic Response Based on the MHLncSig

In order to calculate the relative abundance of 22 kinds of tumor-infiltrating immune cells (TICs) for each sample in the TCGA cohort, we used the CIBERSORT method. Subsequently, the Wilcoxon test was employed to verify the differentiation of 22 types of TICs between the low- and high-risk groups. Next, we converted the gene expression profile of tumor samples into immune function–related score and performed differential analysis of immune function between different risk subgroups using R "limma," "GSVA," "GSEABase," "ggpubr," and "reshape2" packages. Finally, we performed the Tumor Immune Dysfunction and Exclusion (TIDE) algorithm to access the response to immunotherapy.

Screening the Potential Compounds for Clinical Treatment of LUAD Risk Subgroups

We computed the IC50 of antitumor drugs that are commonly recommended for LUAD treatment based on gene expression profile of LUAD patients using the "pRRophetic" package in R software. The Wilcoxon test was adopted to evaluate the difference in the IC50 levels between distinct subgroups.

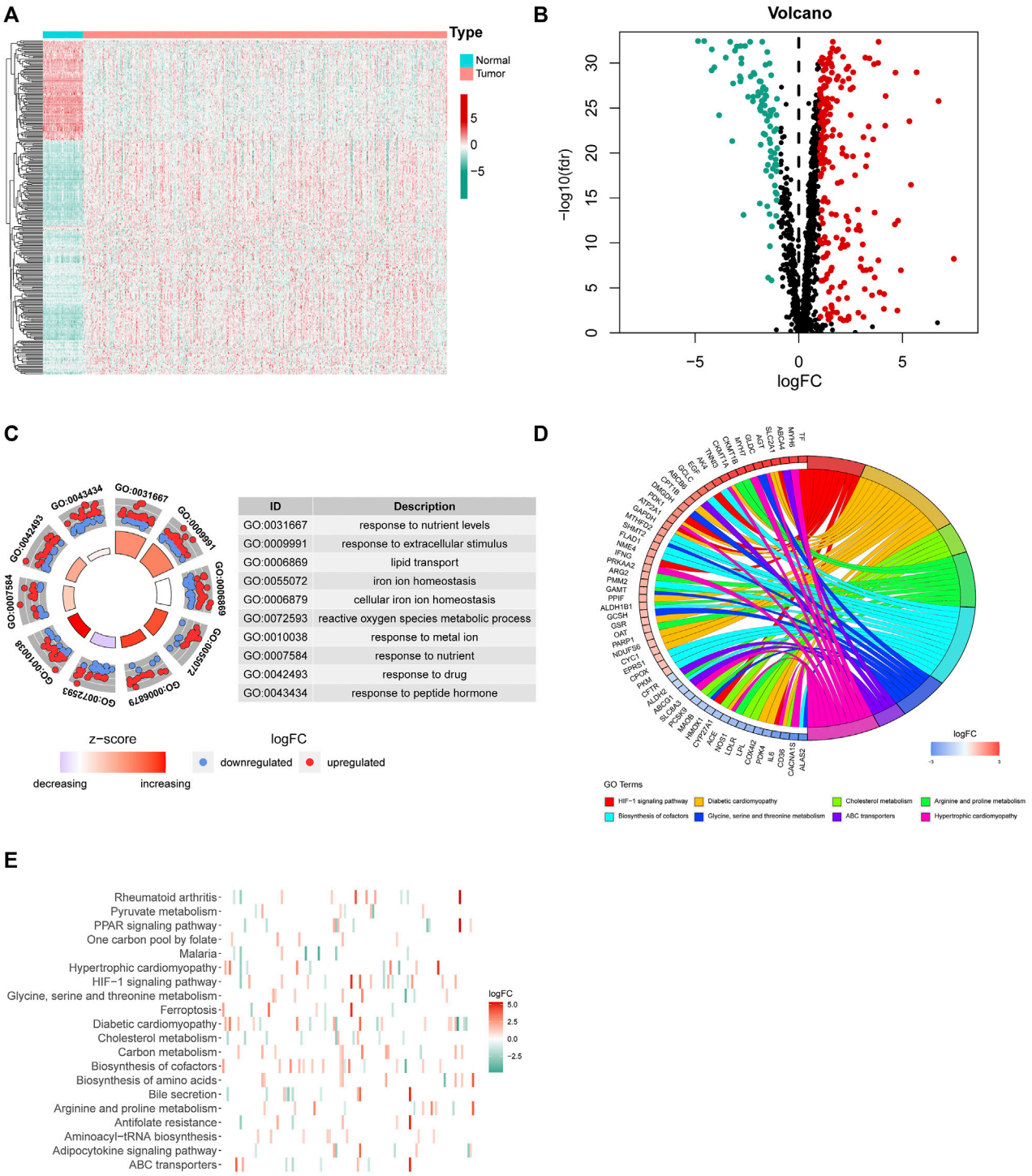
RNA Isolation and qRT-PCR

Patient samples comprising a cohort of 15 paired LUAD and adjacent normal tissues were collected between October 2021 and December 2021 from the Second Affiliated Hospital of Harbin Medical University. With the approval for experiments from the Ethics and Scientific Committees of the Second Affiliated Hospital of Harbin Medical University (Approval Number: KY2021-375), written informed consent was provided by all enrolled patients. The total RNA was extracted using the TRIzol reagent (Invitrogen, CA, United States). 1 μg total RNA and ReverTra Ace qPCR RT Master Mix (TOYOBO) were used for reverse transcriptase reaction. Next, 1 μl synthesized cDNA was used in PCR amplification. The levels of three lncRNAs were measured quantitatively by the SYBR Green Master Mix Kit (TOYOBO). GAPDH was selected as an internal reference. The relative expression was calculated based on the comparative Ct ($2^{-\Delta\Delta Ct}$) method, and Student's t-test (two-tailed) was utilized to assess the significance of lncRNA expression differences in GraphPad Prism (version 8.0). The qRT-PCR primers in this study were as follows:

FENDRR
Forward: 5′-GCCTCAGAGTGGGCTAGATT-3′
Reverse: 5′-TAACGATCCCACCAACACCA-3′
AL590666.2
Forward: 5′-ACAGAATGATCCAGGCACCA-3′
Reverse: 5′-AGGACAAGATGGACGCAGAT-3′
AC090559.1
Forward: 5′-TGCTAGGCAATTCTGGAAGC-3′
Reverse: 5′-TTGCTGTTGCCACAAAGTGA-3′
GAPDH
Forward: 5′-CATGTTCGTCATGGGTGTGAA-3′
Reverse: 5′-GGCATGGACTGTGGTCATGAG-3′

Statistical Analyses

We performed all bioinformatics analyses in R-version 4.1.1. Survival curves were plotted using Kaplan–Meier analysis to estimate the difference of survival for both the subgroups of LUAD in the R "survival" and "survminer" packages. Receiver operating characteristic (ROC) curves and the area under the curve (AUC) values were applied to evaluate the degree of robustness and accuracy of our MHLncSig. Principal component analysis (PCA)



ID	Description
GO:0031667	response to nutrient levels
GO:0009991	response to extracellular stimulus
GO:0006869	lipid transport
GO:0055072	iron ion homeostasis
GO:0006879	cellular iron ion homeostasis
GO:0072593	reactive oxygen species metabolic process
GO:0010038	response to metal ion
GO:0007584	response to nutrient
GO:0042493	response to drug
GO:0043434	response to peptide hormone

FIGURE 2 | Differential expression analysis and corresponding functional enrichment of the 1,499 MHRGs. **(A)** A heatmap of the 294 differentially expressed genes. **(B)** A volcano plot based on the 205 upregulated and 89 downregulated genes. **(C)** The most highly significant 10 results in the GO analysis. **(D,E)** The corresponding genes and the distribution of enriched biological pathways in the KEGG analysis.

was utilized to reduce the dimension of high-dimensional data and visualize the differentiation between the risk subgroups in terms of entire genes expression in TCGA, 2,850 MHRlncRNAs, and the risk model of the 3 MHRlncRNAs. The R "limma" and "scatterplot3d" packages were employed to enable this process. In addition, we constructed a nomogram for predicting the overall

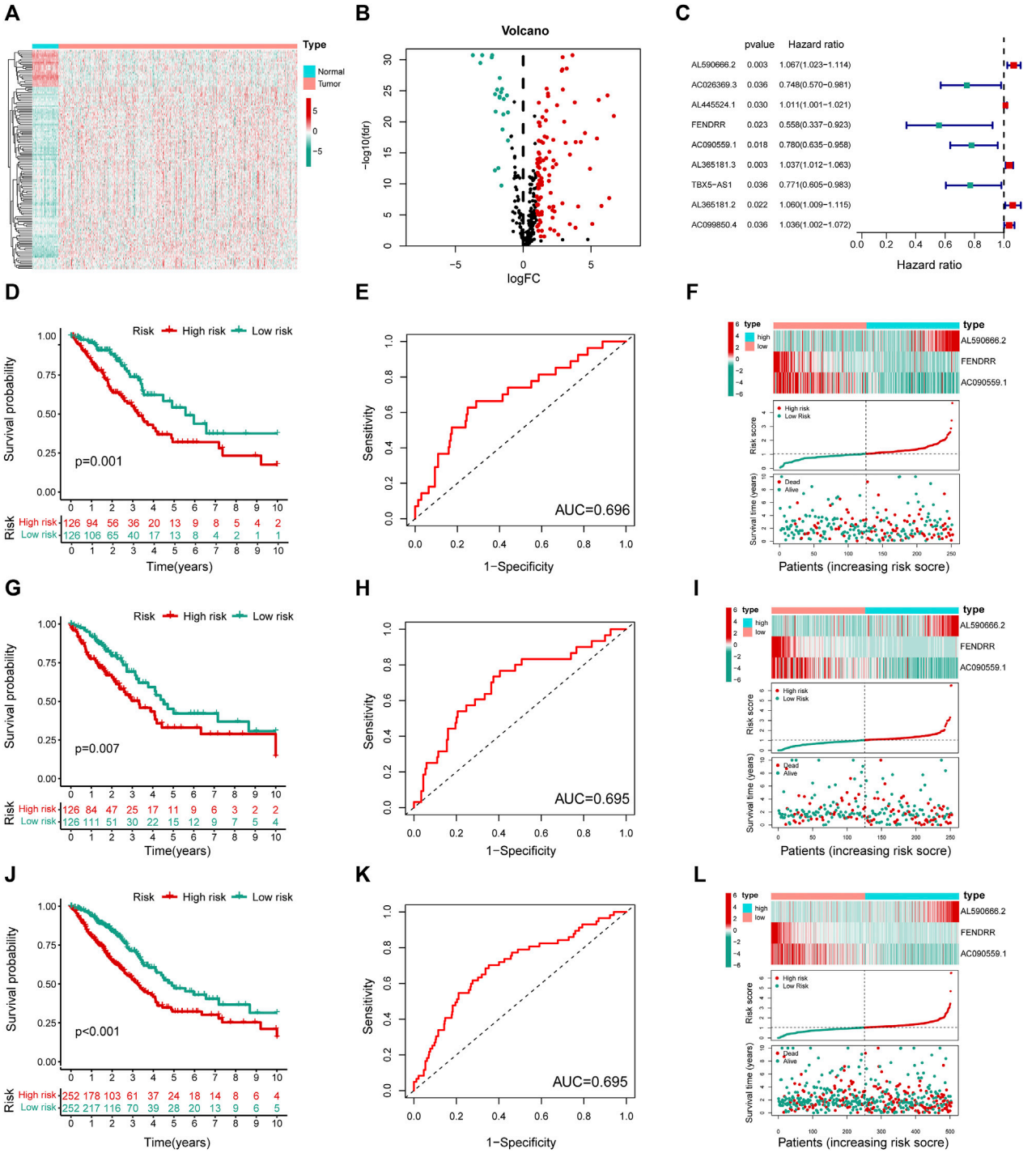


FIGURE 3 | Construction and validation of the MHLncSig. **(A)** A heatmap of differential expression analysis based on 2,850 MHRlncRNAs. **(B)** A volcano plot based on the 109 upregulated MHRlncRNAs and 22 downregulated MHRlncRNAs. **(C)** A forest plot of nine MHRlncRNAs significantly related to survival in LUAD. **(D,G,J)** Different Kaplan–Meier curves in the two risk groups stratified by MHLncSig in the training set, testing set, and the entire TCGA set. **(E,H,K)** ROC curve of the MHLncSig at 3 years in the training set, testing set, and the entire TCGA set. **(F,I,L)** The expression patterns of MHLncSig, the risk grade distribution, and the survival status between the low- and high-risk groups in the training set, testing set, and the entire TCGA set.

survival of 1, 3, and 5 years, and a correction curve was applied to assess the consistency between the model prediction outcome and practical outcome by using the R "survival," "survminer," "rms," and "regplot" packages, respectively. We set the statistically significant threshold as p-value <0.05 (*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$).

TABLE 2 | Multivariate Cox regression analysis of three prognostic lncRNAs.

LncRNA	Coefficient	HR	HR.95L	HR.95H	*p*-value
AL590666.2	0.063	1.066	1.019	1.114	0.005
FENDRR	−0.527	0.590	0.348	1.002	0.051
AC090559.1	−0.158	0.854	0.691	1.055	0.144

HR, hazard ratio; CI, confidence interval.

RESULTS

Identification of Mitochondrial Homeostasis–Related lncRNAs in LUAD Patients

First, we downloaded the coding gene list associated with mitochondrial homeostasis from the GeneCards database. Based on the expression profile of LUAD and normal subjects in TCGA, differential expression analysis of 1,499 MHRGs was conducted to discern the genes that may serve a role in tumorigenesis. A heatmap and a volcano map were plotted to illustrate the 294 differentially expressed genes (205 upregulated genes and 89 downregulated genes in LUAD) (**Figures 2A,B**), and the detailed data are summarized in **Supplementary Table S1**. Furthermore, GO and KEGG analyses were utilized to explore the molecular functions and pathways behind these 294 genes. **Figure 2C** visualized the most highly significant cellular components, biological processes, and molecular functions. Moreover, biological pathways were mainly enriched in processes associated with mitochondrial synthesis and metabolism (**Figures 2D,E**).

Next, 2,850 lncRNAs coexpressed with the 294 differentially expressed genes were identified as MHRlncRNAs by Pearson's correlation analysis. Furthermore, differential expression analysis based on the 2,850 lncRNAs was utilized to screen the candidate MHRlncRNAs for model construction. The result of differential analysis is summarized in **Supplementary Table S2** and visualized in **Figures 3A,B**.

Construction and Validation of the MHLncSig in LUAD Patients

To investigate the prognostic value of 131 candidate MHRlncRNAs, we performed univariate Cox analysis in the training set (252 patients) and identified nine MHRlncRNAs which are significantly associated with the survival of LUAD (**Figure 3C**). According to the multivariate Cox proportional hazards regression analysis, AL590666.2, FENDRR, and AC090559.1 were further established as components of the risk model. The details of multivariate Cox regression analysis are shown in **Table 2**. Finally, we designed a risk-score formula for LUAD patients' survival prediction. The risk score formula is as follows: risk score = 0.063 × expression quantity of AL590666.2 + (−0.527) × expression quantity of FENDRR + (−0.158) × expression quantity of AC090559.1. AL590666.2 with a positive coefficient tended to be a detrimental factor. However, the other two lncRNAs including FENDRR and AC090559.1 tended to be protective factors. Next, grouping was conducted based on the median risk score of the training set; the Kaplan–Meier plot showed entirely different survival curves between different groups (p = 0.001, **Figure 3D**). The AUC value of the receiver operating characteristic (ROC) curve was 0.696, indicating that MHLncSig is equipped with quite accurate prediction performance for the prognosis (**Figure 3E**). In addition, the expression standards of MHLncSig for each patient, the risk grade distribution, and the survival status between the two groups are depicted in **Figure 3F**.

Next, we validated our MHLncSig in the testing set (252 patients) and the entire TCGA set (504 patients). The survival analysis indicated that the patients in the low-risk group had better overall survival (OS) than that in the high-risk group (**Figure 3G**). The AUC value of MHLncSig in the testing set reached 0.695 (**Figure 3H**). **Figure 3I** shows the expression patterns of MHLncSig, the risk grade distribution, and the survival of LUAD patients. Notably, the entire TCGA set also manifested similar results as the aforementioned findings (**Figures 3J–L**).

TABLE 3 | Univariate and multivariate Cox regression analysis of the MHLncSig and prognosis.

Variable	Univariate model			Multivariate model		
	HR	95% CI	*p*-value	HR	95% CI	*p*-value
Training set (n = 252)						
Age	1.030	1.005–1.055	0.016	1.031	1.007–1.056	0.012
Gender	0.957	0.628–1.460	0.840			
Stage	1.464	1.201–1.784	<0.001	1.289	1.044–1.592	0.018
Risk score	2.098	1.529–2.878	<0.001	1.804	1.265–2.573	0.001
Testing set (n = 252)						
Age	0.992	0.973–1.012	0.451			
Gender	1.295	0.856–1.960	0.221			
Stage	1.879	1.535–2.300	<0.001	1.869	1.526–2.289	<0.001
Risk score	1.369	1.127–1.662	0.002	1.376	1.116–1.698	0.003
TCGA set (n = 504)						
Age	1.008	0.993–1.023	0.309			
Gender	1.102	0.820–1.479	0.520			
Stage	1.629	1.417–1.872	<0.001	1.550	1.349–1.782	<0.001
Risk score	1.510	1.297–1.759	<0.001	1.415	1.188–1.684	<0.001

HR, hazard ratio; CI, confidence interval.

Relationship Between MHLncSig and Clinical Traits

We next performed correlation analysis between the risk score calculated by MHLncSig and clinical traits including age, gender, T stage, N stage, M stage, and pathological stage. The results indicated that there were high correlations between MHLncSig and all these clinical features in the entire TCGA set ($p < 0.05$, **Supplementary Figure S1**). Furthermore, we evaluated the predictive performance of our MHLncSig in different subgroups classified by clinical traits. In the vast majority of subgroups of LUAD patients, the low-risk group continued to maintain supremacy in OS (**Supplementary Figure S2**). Moreover, univariate and multivariate Cox regression analyses were performed to investigate whether the MHLncSig had independent prognostic effect for LUAD. After correction for other clinical variables, we observed that the MHLncSig retained independent significance for the prediction of OS in the training set, the testing set, and the entire TCGA set (**Table 3**).

Evaluation of the Prognostic MHLncSig in Terms of PCA and GSEA Analyses

We conducted the PCA analysis to estimate the distributions of the two different risk groups based on the total gene expression profiles, 2,850 MHRlncRNAs, and the MHLncSig categorized by the expression profiles of the three risk lncRNAs (**Figure 4A**). Compared with the relatively scattered distributions of the two different risk groups based on the total gene expression profiles and 2,850 MHRlncRNAs, the results based on our prognostic signature indicated that the MHLncSig had excellent grouping ability to a certain extent. Furthermore, we performed GSEA enrichment analysis to reveal the underlying mechanisms and pathways behind the high-risk group characterized by dismal prognosis. A total of 37 pathways were significantly enriched in the high-risk group (FDR < 0.05, **Supplementary Table S3**). The pathways involving cell cycle, cysteine and methionine metabolism, glutathione metabolism, oxidative phosphorylation, proteasome, purine metabolism, and pyrimidine metabolism are visualized in **Figure 4B**.

Establishment and Performance Evaluation of the Prognostic Nomogram

We constructed a prognostic nomogram comprising the risk signature and clinical traits to predict the survival rate of 1, 2, and 3 years (**Figure 4C**). A total score was assigned to each LUAD patient by combining six individual scores in the nomogram, where a lower total point was related to a better outcome. Furthermore, we utilized ROC curve and calibration plot analyses to evaluate the validity of the nomogram (**Figures 4D,E**). The prediction curve was very close to the actual survival curve in the calibration plot and the AUCs of 1-, 2-, and 3-year curve were 0.728, 0.71, and 0.696, respectively, suggesting that the nomogram predicted the prognosis for LUAD patients well.

Evaluation of Tumor Microenvironment and TMB Using MHLncSig

Tumor microenvironment comprising stromal and immune components together with the secreted factors provide an immunosuppressive and protumoral environment for tumor development, which also correlates closely with immunotherapy positive response (Fridman et al., 2012; Gajewski et al., 2013). We made a comparison of the levels of stromal score, immune score, and ESTIMATE score between the two risk groups. The violin plots (**Figure 5A**) showed that all these three indicators were significantly decreased in the group with high risk. Further survival analysis suggested that lower levels of stromal score, immune score, and ESTIMATE score were significantly correlated with the worse prognosis in LUAD cohorts (**Figure 5B**). Moreover, we utilized R "maftools" package to analyze and visualize the mutational landscape of LUAD patients. The top 20 genes with the highest mutation frequency between the two risk subgroups are shown in **Figures 5C,D**. It is worth noting that the TMB score in the high-risk group exceeded that in the low-risk group, suggesting that MHLncSig had a high degree of TMB relevance (**Figure 5E**).

Immune Landscape and Immunotherapeutic Response Based on the Prognostic Signature

To further explore the relationship between our prognostic signature and immune infiltration, we used the CIBERSORT algorithm to analyze the relative abundance of 22 TICs in each LUAD patient. **Figure 5F** showed that the contents of nearly half of these TICs were significantly different between the two different risk groups. Of these, the plasma cells, T-cell CD8, B-cell naive, macrophages M0, and T-cell CD4 memory activated were dramatically increased in the high-risk group, while monocytes, macrophages M2, T-cell CD4 memory resting, mast cells resting, and dendritic cells resting were significantly decreased in the high-risk group. Moreover, we observed a comprehensive suppression of immune functions in the high-risk group (**Figure 6A**). To elucidate the relationship between the MHLncSig and immune checkpoint genes, the expression of immune checkpoint genes were compared between the low-risk ($n = 252$) and high-risk ($n = 252$) groups. The results indicated that 40 types of immune checkpoint genes were significantly upregulated in the low-risk group (**Figure 6B**). Previous literatures have reported that the TIDE algorithm is a powerful tool to assess the immunotherapeutic response (Jiang et al., 2018). We discovered that LUAD patients in the high-risk group are more likely to benefit from immunotherapy, indicating that the MHLncSig may serve as a potential indicator for predicting immunotherapeutic response (**Figure 6C**).

Screening of Candidate Compounds Targeting MHLncSig and Model Validation in Human Surgical Resection Specimens

According to the pRRophetic algorithm, we calculated the IC50 of nine chemotherapeutic and targeted agents in the low- and high-risk patients, which are recommended by the National Comprehensive Cancer Network (NCCN) guidelines for LUAD treatment. Wilcoxon test analysis indicated that five of these agents (paclitaxel, docetaxel, erlotinib, pemetrexed, and

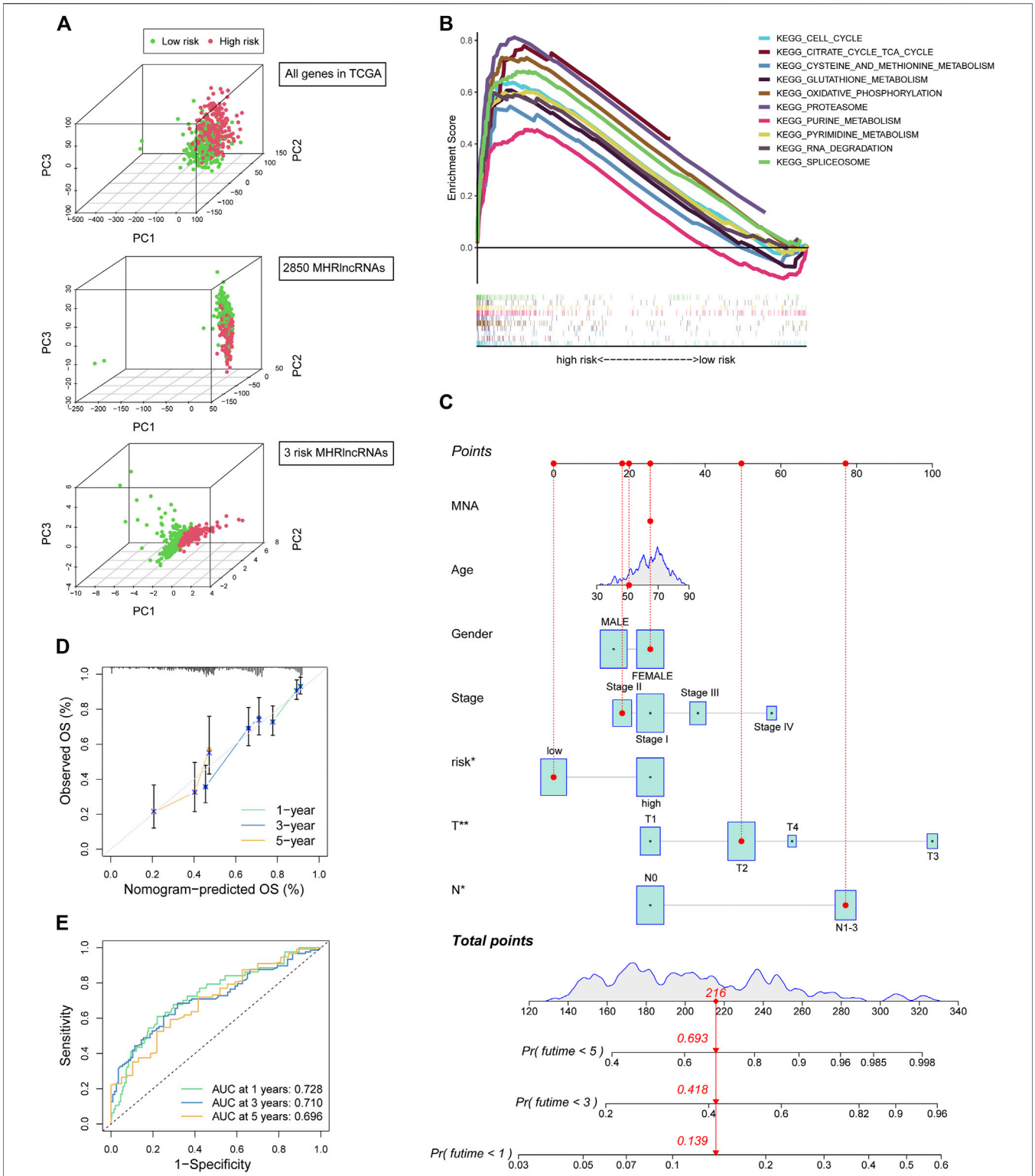


FIGURE 4 | PCA and GSEA analyses of the MHLncSig and construction of a prognostic nomogram. **(A)** PCA based on the total gene expression profiles, 2,850 MHRlncRNAs, and the MHLncSig categorized by three risk MHRlncRNAs. **(B)** GSEA analysis of the high-risk group calculated by MHLncSig. **(C)** A nomogram predicting 1-, 2-, and 3-year OS of LUAD cancer. **(D)** The calibration plot of the nomogram. **(E)** ROC curves of the prognostic nomogram at 1, 3, and 5 years.

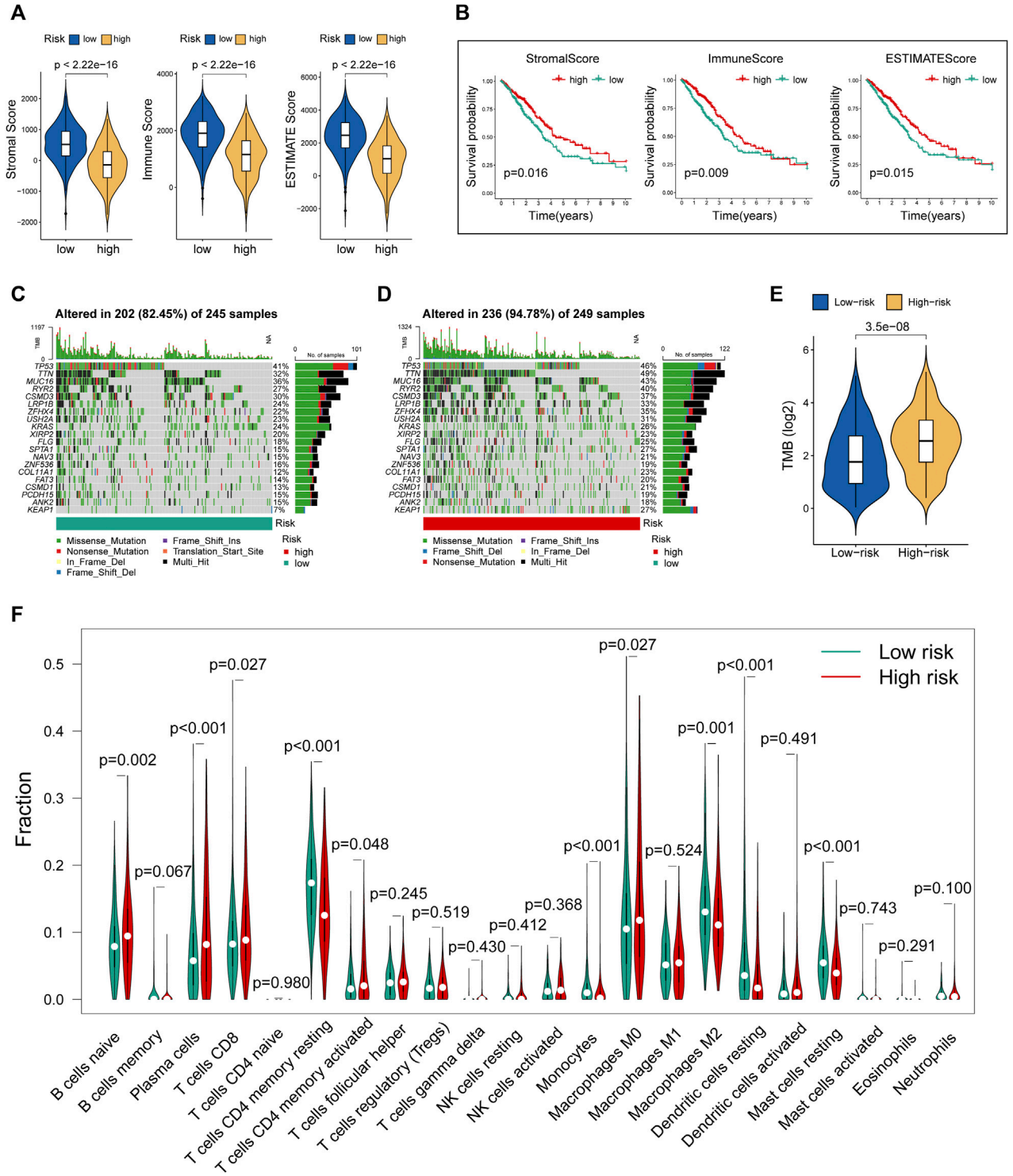


FIGURE 5 | Evaluation of tumor microenvironment, TMB, and immune cell infiltration between different risk groups. **(A)** The differences of stromal score, immune score, and ESTIMATE score between the low- and high-risk groups. **(B)** The survival differences between the LUAD subgroups classified by stromal score, immune score, and ESTIMATE score. **(C)** The mutational landscape of the top 20 genes with the highest mutation frequency in the low-risk group. **(D)** The mutational landscape of the top 20 genes with the highest mutation frequency in the high-risk group. **(E)** The differences of TMB score between the low- and high-risk groups. **(F)** Estimation of immune cell infiltration in LUAD patients using the CIBERSORT algorithm.

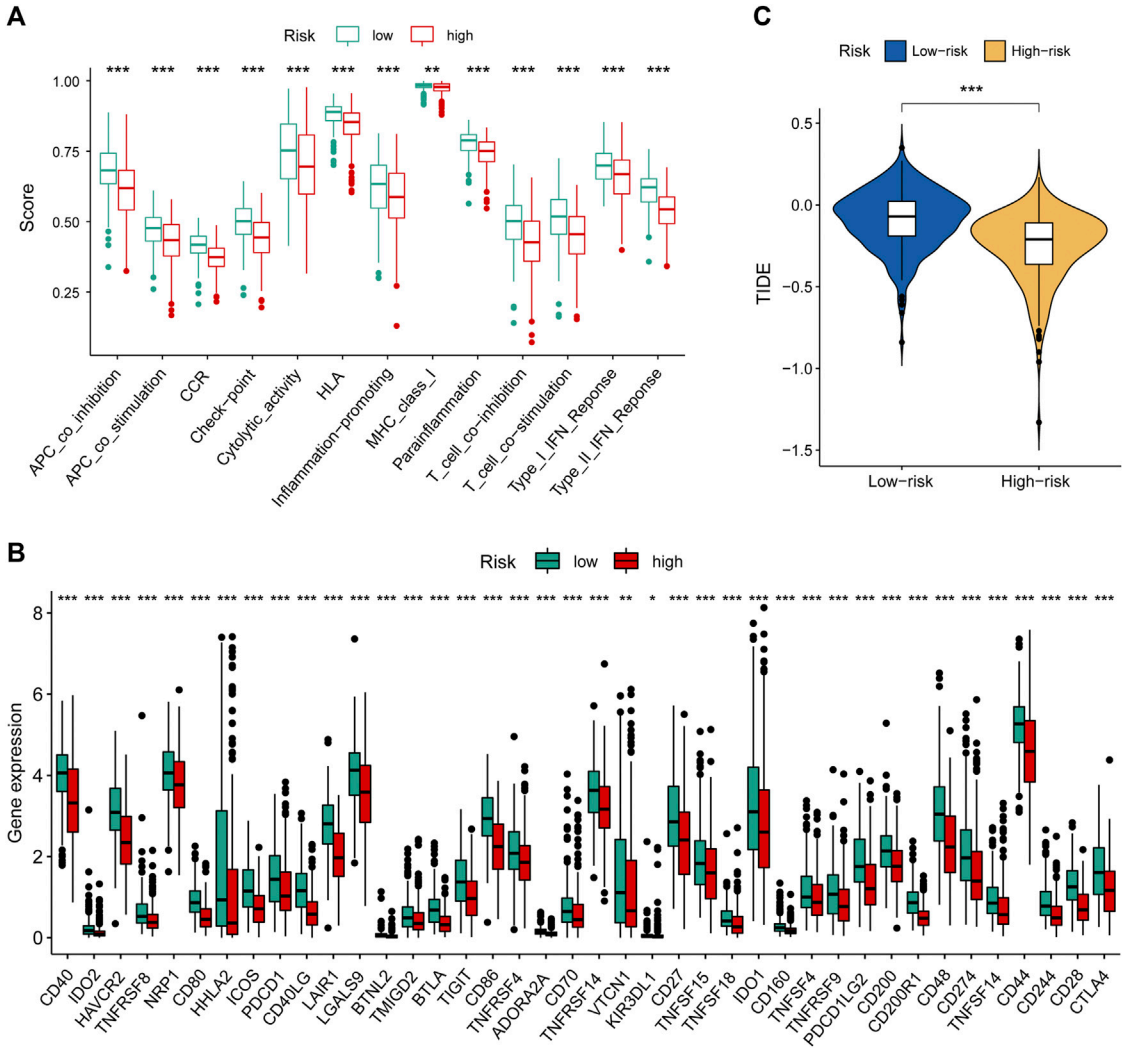


FIGURE 6 | Immune landscape and immunotherapeutic response based on the MHLncSig. **(A)** The difference of immune function between the low- and high-risk groups. **(B)** Expression levels of immune checkpoint genes in the low- and high-risk patients. **(C)** TIDE difference between the low- and high-risk groups.

carboplatin) had lower IC50 in high-risk patients ($p < 0.05$, **Figures 7A–I**), which suggested that LUAD patients with high risk were more sensitive to these five agents. Finally, based on human paired LUAD tissues obtained by surgery, we validated the differential expression of three risk lncRNAs included in the MHLncSig by qRT-PCR assays (**Figures 7J–L**). The differential analysis revealed that AL590666.2 was significantly upregulated in LUAD tissues, while FENDRR and AC090559.1 were remarkably downregulated in LUAD tissues.

DISCUSSION

In actual clinical work, pathologic staging is still the mainstream procedure to predict long-term survival and guide treatment modality for LUAD patients. However, a population with the same pathologic staging would also present with distinct clinical outcomes, indicating that the traditional staging system could not accommodate the individualized prediction and treatment. More recently, next-generation sequencing brings a revolutionary change in predicting the prognosis of cancer (Nair et al., 2018). A growing number of prognostic signatures on the basis of encoding genes and noncoding RNAs were established to forecast the survival outcome of cancer patients (Peng et al., 2021; Zhou et al., 2021). To the best of our knowledge, we are the first to identify the prognostic MHLncSig and comprehensively analyze the tumor immunity of this risk model in LUAD, which provides valuable novel options for the guidance of clinical personalized treatment.

Due to their pivotal functions in the resistance of cancer cells to regulated cell death (RCD) induced by treatment, mitochondria have drawn considerable attention for their potential value in developing novel anticancer agents. Actually, mitochondria affect immunosurveillance through both the

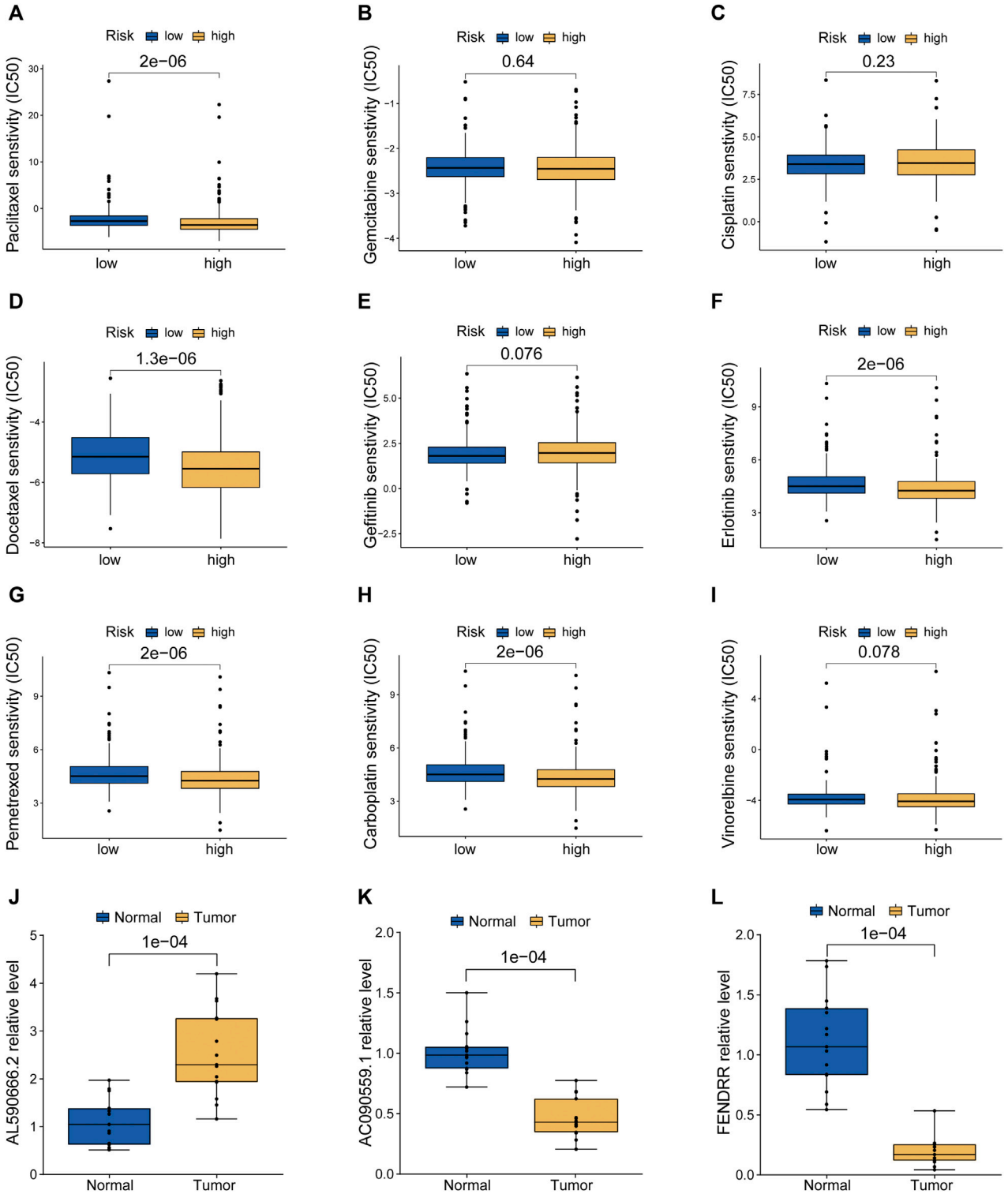


FIGURE 7 | Screening of candidate compounds and model validation in human LUAD specimens. **(A–I)** Sensitivity analysis of nine chemotherapeutic and molecular-targeted agents in the low- and high-risk patients. **(J–L)** Validation of the expression differences of three risk MHRlncRNAs in human LUAD tissues and adjacent normal tissues.

malignant cell-extrinsic and malignant cell-intrinsic mechanisms. From one side, the mitochondria of malignant cells release many danger signals when RCD occurs, and these signals are paramount for activating the dendritic cells to promote antitumor immune responses. From another aspect, mitochondrial metabolism involves a number of functions related to antitumor immunity, including the establishment of protective immunological memory, inflammasome activation, and the differentiation of macrophage subsets (Porporato et al., 2018). It is worth noting that dysregulation of mitochondrial metabolism is found to be associated with metastatic cascade and worse prognosis in the nine types of tumors (Gaude and Frezza, 2016). Moreover, the dysbiosis of mitochondrial homeostasis is related to the excessive generation of reactive oxygen species (ROS) and consequent metastatic cascade (Fu et al., 2017). Recently, accumulated evidence has suggested that lncRNAs play important roles in tumorigenesis and tumor development. The dysfunctional lncRNAs may be presumed as key factors in the active phases of tumors (Guan et al., 2019). Zheng et al. (2015) discovered novel mechanisms of lncRNA HOTAIR in maintaining mitochondrial function in the HeLa cells. In addition, nuclear-encoded lncRNA MALAT1 was reported to regulate the metabolic reprogramming through the mitophagy pathway in hepatocellular carcinoma cells (Zhao et al., 2021). Nonetheless, prognostic biomarkers of LUAD and studies on the protumor mechanisms concerning MHRlncRNAs are still lacking. Therefore, we attempted to establish an independent signature for predicting the prognosis of LUAD based on MHRlncRNAs.

In our study, we identified 2,850 MHRlncRNAs based on the TCGA data. Subsequent differential analysis and univariate Cox analysis confirmed nine MHRlncRNAs with prognostic value, and three of these were utilized to construct a MHLncSig to predict the OS of LUAD patients.

Among them, FENDRR is commonly considered to be a tumor suppressor, which has been extensively studied in multiple cancer types, including cholangiocarcinoma, malignant melanoma, non–small cell lung cancer, hepatocellular carcinoma, and breast cancer (Li et al., 2018b; Zhang et al., 2018; Wang et al., 2019b; Qin et al., 2019; Chen et al., 2020). Nevertheless, specific mechanisms of two other lncRNAs (AL590666.2 and AC090559.1) during tumorigenesis and progression have never been previously reported and our research unmasked their underlying roles in LUAD.

TMB is defined typically as the total number of somatic coding mutations, which is associated with the production of neoantigens that induce antitumor immunity (Allgäuer et al., 2018). Recent studies have indicated that the higher TMB score was related to durable clinical benefit and improved objective response in tumor immunotherapy (Rizvi et al., 2015). Notably, the LUAD patients with high risk tend to possess higher levels of TMB, indicating that they are more likely to benefit from immunotherapy. This finding was subsequently confirmed by the TIDE algorithm. Moreover, another significant finding of our study suggested that the MHLncSig was significantly associated with the immune infiltration in LUAD, which further supported the fact that mitochondrial homeostasis plays a critical role in the tumor immune microenvironment.

There are still several limitations of the present study that need to be considered. First, it would be better to validate our prognostic model with several external datasets. In fact, we were unable to find an ideal Gene Expression Omnibus (GEO) database owing to the lack of complete expression profiles of the three risk MHRlncRNAs and detailed clinical information. Second, although we have initially screened a number of candidate compounds targeting MHLncSig in our study, the sensitivity of LUAD patients to specific immunotherapy agents requires further investigation due to a lack of immunotherapy drugs for LUAD in the R "pRRophetic" package.

Finally, functional experiments should be performed *in vivo* and *in vitro* to further corroborate our findings.

In summary, using multiple bioinformatics approaches and qRT-PCR experiments, we established and validated a 3-MHRlncRNA signature to independently predict the OS of LUAD patients. In addition, the MHLncSig shows advantages in terms of guiding individualized treatment of LUAD.

CONCLUSION

In conclusion, we successfully developed an accurate prognostic model of mitochondrial homeostasis, which was closely associated with tumor microenvironment, TMB, immune infiltration, and the response to immunotherapy in LUAD. Meanwhile, the findings in our study may provide clues for further elucidating the molecular mechanism of MHRlncRNAs in the tumorigenesis and LUAD 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**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics and Scientific Committees of the Second Affiliated Hospital of Harbin Medical University. The patients/ participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

BP conceived the study, participated in the bioinformatics analysis, and wrote the manuscript. HL, CC, LW, and HX participated in the study design and performed experiments. HL, TL, RN, RX, TX, LY, and KW conducted the bioinformatics work. The manuscript was revised and

approved by LZ and XL. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We sincerely thank professor Yong Zhang (Department of Pharmacology, College of Pharmacy, Harbin Medical University, China), for his precious suggestions contributed to the revision of the manuscript. We also thank all participants for their contribution to the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.870302/full#supplementary-material

REFERENCES

Allgäuer, M., Budczies, J., Christopoulos, P., Endris, V., Lier, A., Rempel, E., et al. (2018). Implementing Tumor Mutational Burden (TMB) Analysis in Routine Diagnostics-A Primer for Molecular Pathologists and Clinicians. *Transl. Lung Cancer Res.* 7 (6), 703–715. doi:10.21037/tlcr.2018.08.14

Butow, R. A., and Avadhani, N. G. (2004). Mitochondria and cell signalling. *Mol. Cell.* 14 (1), 1–15. doi:10.1016/s1097-2765(04)00179-0

Chen, X.-e., Chen, P., Chen, S., Lu, J., Ma, T., Shi, G., et al. (2020). Long Non-coding RNA FENDRR Inhibits Migration and Invasion of Cutaneous Malignant Melanoma Cells. *Biosci. Rep.* 40 (3), BSR20191194. doi:10.1042/BSR20191194

Correia, R. L., Oba-Shinjo, S. M., Uno, M., Huang, N., and Marie, S. K. N. (2011). Mitochondrial DNA Depletion and its Correlation with TFAM, TFB1M, TFB2M and POLG in Human Diffusely Infiltrating Astrocytomas. *Mitochondrion* 11 (1), 48–53. doi:10.1016/j.mito.2010.07.001

Denisenko, T. V., Budkevich, I. N., and Zhivotovsky, B. (2018). Cell Death-Based Treatment of Lung Adenocarcinoma. *Cell. Death Dis.* 9, 117. doi:10.1038/s41419-017-0063-y

Forde, P. M., Chaft, J. E., Smith, K. N., Anagnostou, V., Cottrell, T. R., Hellmann, M. D., et al. (2018). Neoadjuvant PD-1 Blockade in Resectable Lung Cancer. *N. Engl. J. Med.* 378, 1976–1986. doi:10.1056/NEJMoa1716078

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

Fu, L., Dong, Q., He, J., Wang, X., Xing, J., Wang, E., et al. (2017). SIRT4 Inhibits Malignancy Progression of NSCLCs, through Mitochondrial Dynamics Mediated by the ERK-Drp1 Pathway. *Oncogene* 36 (19), 2724–2736. doi:10.1038/onc.2016.425

Gajewski, T. F., Schreiber, H., and Fu, Y. X. (2013). Innate and Adaptive Immune Cells in the Tumor Microenvironment. *Nat. Immunol.* 14 (10), 1014–1022. doi:10.1038/ni.2703

Gaude, E., and Frezza, C. (2016). Tissue-specific and Convergent Metabolic Transformation of Cancer Correlates with Metastatic Potential and Patient Survival. *Nat. Commun.* 7, 13041. doi:10.1038/ncomms13041

Greaves, L. C., Reeve, A. K., Taylor, R. W., and Turnbull, D. M. (2012). Mitochondrial DNA and Disease. *J. Pathol.* 226 (2), 274–286. doi:10.1002/path.3028

Guan, H., Zhu, T., Wu, S., Liu, S., Liu, B., Wu, J., et al. (2019). Long Noncoding RNA LINC00673-V4 Promotes Aggressiveness of Lung Adenocarcinoma via Activating WNT/β-catenin Signaling. *Proc. Natl. Acad. Sci. U.S.A.* 116 (28), 14019–14028. doi:10.1073/pnas.1900997116

Guo, J., Zheng, L., Liu, W., Wang, X., Wang, Z., Wang, Z., et al. (2011). Frequent Truncating Mutation of TFAM Induces Mitochondrial DNA Depletion and Apoptotic Resistance in Microsatellite-Unstable Colorectal Cancer. *Cancer Res.* 71 (8), 2978–2987. doi:10.1158/0008-5472.CAN-10-3482

Horton, T. M., Petros, J. A., Heddi, A., Shoffner, J., Kaufman, A. E., Graham, S. D., Jr, et al. (1996). Novel Mitochondrial DNA Deletion Found in a Renal Cell Carcinoma. *Genes. Chromosom. Cancer* 15 (2), 95–101. doi:10.1002/(sici)1098-2264(199602)15:2<95:aid-gcc3>3.0.co;2-z

Imanishi, H., Hattori, K., Wada, R., Ishikawa, K., Fukuda, S., Takenaga, K., et al. (2011). Mitochondrial DNA Mutations Regulate Metastasis of Human Breast Cancer Cells. *PLoS One* 6 (8), e23401. doi:10.1371/journal.pone.0023401

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T Cell Dysfunction and Exclusion Predict Cancer Immunotherapy Response. *Nat. Med.* 24 (10), 1550–1558. doi:10.1038/s41591-018-0136-1

Joseph, N. A., Chiou, S.-H., Lung, Z., Yang, C.-L., Lin, T.-Y., Chang, H.-W., et al. (2018). The Role of HGF-MET Pathway and CCDC66 cirRNA Expression in EGFR Resistance and Epithelial-To-Mesenchymal Transition of Lung Adenocarcinoma Cells. *J. Hematol. Oncol.* 11, 74. doi:10.1186/s13045-018-0557-9

Lee, H.-C., Yin, P.-H., Lin, J.-C., Wu, C.-C., Chen, C.-Y., Wu, C.-W., et al. (2005). Mitochondrial Genome Instability and mtDNA Depletion in Human Cancers. *Ann. N. Y. Acad. Sci.* 1042, 109–122. doi:10.1196/annals.1338.011

Li, X., Shao, C., Shi, Y., and Han, W. (2018). Lessons Learned from the Blockade of Immune Checkpoints in Cancer Immunotherapy. *J. Hematol. Oncol.* 11, 31. doi:10.1186/s13045-018-0578-4

Li, Y., Zhang, W., Liu, P., Xu, Y., Tang, L., Chen, W., et al. (2018). Long Non-coding RNA FENDRR Inhibits Cell Proliferation and Is Associated with Good Prognosis in Breast Cancer. *Ott* Vol. 11, 1403–1412. doi:10.2147/OTT.S149511

Liu, Q., Yu, S., Zhao, W., Qin, S., Chu, Q., and Wu, K. (2018). EGFR-TKIs Resistance via EGFR-independent Signaling Pathways. *Mol. Cancer* 17, 53. doi:10.1186/s12943-018-0793-1

Ma, K., Chen, G., Li, W., Kepp, O., Zhu, Y., and Chen, Q. (2020). Mitophagy, Mitochondrial Homeostasis, and Cell Fate. *Front. Cell. Dev. Biol.* 8, 467. doi:10.3389/fcell.2020.00467

Mao, C., Wang, X., Liu, Y., Wang, M., Yan, B., Jiang, Y., et al. (2018). A G3BP1-Interacting lncRNA Promotes Ferroptosis and Apoptosis in Cancer via Nuclear Sequestration of P53. *Cancer Res.* 78, 3454–3496. doi:10.1158/0008-5472.CAN-17-3454

Moro, L., Arbini, A. A., Yao, J. L., di Sant'Agnese, P. A., Marra, E., and Greco, M. (2009). Mitochondrial DNA Depletion in Prostate Epithelial Cells Promotes Anoikis Resistance and Invasion through Activation of PI3K/Akt2. *Cell. Death Differ.* 16 (4), 571–583. doi:10.1038/cdd.2008.178

Nair, M., Sandhu, S. S., and Sharma, A. K. (2018). Cancer Molecular Markers: A Guide to Cancer Detection and Management. *Seminars Cancer Biol.* 52 (Pt 1), 39–55. doi:10.1016/j.semcancer.2018.02.002

Park, J. S., Sharma, L. K., Li, H., Xiang, R., Holstein, D., Wu, J., et al. (2009). A Heteroplasmic, Not Homoplasmic, Mitochondrial DNA Mutation Promotes Tumorigenesis via Alteration in Reactive Oxygen Species Generation and Apoptosis. *Hum. Mol. Genet.* 18 (9), 1578–1589. doi:10.1093/hmg/ddp069

Pathak, T., and Trebak, M. (2018). Mitochondrial Ca2+ Signaling. *Pharmacol. Ther.* 192, 112–123. doi:10.1016/j.pharmthera.2018.07.001

Peng, B., Li, H., Na, R., Lu, T., Li, Y., Zhao, J., et al. (2021). Identification of a Novel Prognostic Signature of Genome Instability-Related LncRNAs in Early Stage Lung Adenocarcinoma. *Front. Cell. Dev. Biol.* 9, 706454. doi:10.3389/fcell.2021.706454

Petros, J. A., Baumann, A. K., Ruiz-Pesini, E., Amin, M. B., Sun, C. Q., Hall, J., et al. (2005). mtDNA Mutations Increase Tumorigenicity in Prostate Cancer. *Proc. Natl. Acad. Sci. U.S.A.* 102 (3), 719–724. doi:10.1073/pnas.0408894102

Porporato, P. E., Filigheddu, N., Pedro, J. M. B.-S., Kroemer, G., and Galluzzi, L. (2018). Mitochondrial Metabolism and Cancer. *Cell. Res.* 28 (3), 265–280. doi:10.1038/cr.2017.155

Qin, X., Lu, M., Zhou, Y., Li, G., and Liu, Z. (2019). LncRNA FENDRR Represses Proliferation, Migration and Invasion through Suppression of Survivin in Cholangiocarcinoma Cells. *Cell. Cycle* 18 (8), 889–897. doi:10.1080/15384101.2019.1598726

Rizvi, N. A., Hellmann, M. D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J. J., et al. (2015). Mutational landscape determines sensitivity to PD-1 blockade in non–small cell lung cancer. *Science* 348 (6230), 124–128. doi:10.1126/science.aaa1348

Tseng, L.-M., Yin, P.-H., Chi, C.-W., Hsu, C.-Y., Wu, C.-W., Lee, L.-M., et al. (2006). Mitochondrial DNA Mutations and Mitochondrial DNA Depletion in Breast Cancer. *Genes. Chromosom. Cancer* 45 (7), 629–638. doi:10.1002/gcc.20326

Wallace, D. C. (2012). Mitochondria and Cancer. *Nat. Rev. Cancer* 12 (10), 685–698. doi:10.1038/nrc3365

Wang, B., Xian, J., Zang, J., Xiao, L., Li, Y., Sha, M., et al. (2019). Long Non-coding RNA FENDRR Inhibits Proliferation and Invasion of Hepatocellular Carcinoma by Down-Regulating Glypican-3 Expression. *Biochem. Biophysical Res. Commun.* 509 (1), 143–147. doi:10.1016/j.bbrc.2018.12.091

Wang, M., Mao, C., Ouyang, L., Liu, Y., Lai, W., Liu, N., et al. (2019). Long Noncoding RNA LINC00336 Inhibits Ferroptosis in Lung Cancer by Functioning as a Competing Endogenous RNA. *Cell. Death Differ.* 26, 2329–2343. doi:10.1038/s41418-019-0304-y

Wang, Y., Tan, H., Yu, T., Chen, X., Jing, F., and Shi, H. (2021). Potential Immune Biomarker Candidates and Immune Subtypes of Lung Adenocarcinoma for Developing mRNA Vaccines. *Front. Immunol.* 12, 755401. doi:10.3389/fimmu.2021.755401

Yang, Y., Tai, W., Lu, N., Li, T., Liu, Y., Wu, W., et al. (2020). lncRNA ZFAS1 Promotes Lung Fibroblast-To-Myofibroblast Transition and Ferroptosis via Functioning as a ceRNA through miR-150-5p/SLC38A1 axis. *Aging* 12, 9085–9102. doi:10.18632/aging.103176

Zhang, M. Y., Zhang, Z. L., Cui, H. X., Wang, R. K., and Fu, L. (2018). Long Non-coding RNA FENDRR Inhibits NSCLC Cell Growth and Aggressiveness by Sponging miR-761. *Eur. Rev. Med. Pharmacol. Sci.* 22 (23), 8324–8332. doi:10.26355/eurrev_201812_16530

Zhao, Y., Zhou, L., Li, H., Sun, T., Wen, X., Li, X., et al. (2021). Nuclear-Encoded lncRNA MALAT1 Epigenetically Controls Metabolic Reprogramming in HCC Cells through the Mitophagy Pathway. *Mol. Ther. - Nucleic Acids* 23, 264–276. doi:10.1016/j.omtn.2020.09.040

Zheng, J., Zhou, Z., Qiu, Y., Wang, M., Yu, H., Wu, Z., et al. (2021). A Prognostic Ferroptosis-Related lncRNAs Signature Associated with Immune Landscape and Radiotherapy Response in Glioma. *Front. Cell. Dev. Biol.* 9, 675555. doi:10.3389/fcell.2021.675555

Zheng, P., Xiong, Q., Wu, Y., Chen, Y., Chen, Z., Fleming, J., et al. (2015). Quantitative Proteomics Analysis Reveals Novel Insights into Mechanisms of Action of Long Noncoding RNA Hox Transcript Antisense Intergenic RNA (HOTAIR) in HeLa Cells. *Mol. Cell. Proteomics* 14 (6), 1447–1463. doi:10.1074/mcp.M114.043984

Zhou, P., Lu, Y., Zhang, Y., and Wang, L. (2021). Construction of an Immune-Related Six-lncRNA Signature to Predict the Outcomes, Immune Cell Infiltration, and Immunotherapy Response in Patients with Hepatocellular Carcinoma. *Front. Oncol.* 11, 661758. doi:10.3389/fonc.2021.661758

Zhu, M., Zhang, L., Cui, H., Zhao, Q., Wang, H., Zhai, B., et al. (2022). Co-Mutation of FAT3 and LRP1B in Lung Adenocarcinoma Defines a Unique Subset Correlated with the Efficacy of Immunotherapy. *Front. Immunol.* 12, 800951. doi:10.3389/fimmu.2021.800951

Zong, W.-X., Rabinowitz, J. D., and White, E. (2016). Mitochondria and Cancer. *Mol. Cell.* 61 (5), 667–676. doi:10.1016/j.molcel.2016.02.011

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: 08 July 2022
doi: 10.3389/fgene.2022.917150

Chromatin Separation Regulators Predict the Prognosis and Immune Microenvironment Estimation in Lung Adenocarcinoma

Zhaoshui Li[1,2], Zaiqi Ma[2], Hong Xue[3], Ruxin Shen[1], Kun Qin[1], Yu Zhang[1], Xin Zheng[4] and Guodong Zhang[5]**

[1]Qingdao Medical College, Qingdao University, Qingdao, China, [2]Cardiothoracic Surgery Department, Qingdao Hiser Hospital Affiliated to Qingdao University, Qingdao, China, [3]Heart Center Department, Qingdao Hiser Hospital Affiliated to Qingdao University, Qingdao, China, [4]Cancer Center Department, Qingdao Hiser Hospital Affiliated to Qingdao University, Qingdao, China, [5]Thoracic Surgery Department, Shandong Cancer Hospital Affiliated to Shandong First Medical University, Jinan, China

OPEN ACCESS

Edited by:
Katarzyna Leszczyńska, Nencki Institute of Experimental Biology (PAS), Poland

Reviewed by:
Hiroshi Harada, Kyoto University, Japan
Monica Olcina, University of Oxford, United Kingdom

***Correspondence:**
Xin Zheng
zhengxin66999@163.com
Guodong Zhang
guodongzhang2022@163.com

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: 11 April 2022
Accepted: 23 May 2022
Published: 08 July 2022

Citation:
Li Z, Ma Z, Xue H, Shen R, Qin K, Zhang Y, Zheng X and Zhang G (2022) Chromatin Separation Regulators Predict the Prognosis and Immune Microenvironment Estimation in Lung Adenocarcinoma. Front. Genet. 13:917150. doi: 10.3389/fgene.2022.917150

Background: Abnormal chromosome segregation is identified to be a common hallmark of cancer. However, the specific predictive value of it in lung adenocarcinoma (LUAD) is unclear.

Method: The RNA sequencing and the clinical data of LUAD were acquired from The Cancer Genome Atlas (TACG) database, and the prognosis-related genes were identified. The Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) were carried out for functional enrichment analysis of the prognosis genes. The independent prognosis signature was determined to construct the nomogram Cox model. Unsupervised clustering analysis was performed to identify the distinguishing clusters in LUAD-samples based on the expression of chromosome segregation regulators (CSRs). The differentially expressed genes (DEGs) and the enriched biological processes and pathways between different clusters were identified. The immune environment estimation, including immune cell infiltration, HLA family genes, immune checkpoint genes, and tumor immune dysfunction and exclusion (TIDE), was assessed between the clusters. The potential small-molecular chemotherapeutics for the individual treatments were predicted via the connectivity map (CMap) database.

Results: A total of 2,416 genes were determined as the prognosis-related genes in LUAD. Chromosome segregation is found to be the main bioprocess enriched by the prognostic genes. A total of 48 CSRs were found to be differentially expressed in LUAD samples and were correlated with the poor outcome in LUAD. Nine CSRs were identified as the independent prognostic signatures to construct the nomogram Cox model. The LUAD-samples were divided into two distinct clusters according to the expression of the 48 CSRs. Cell cycle and chromosome segregation regulated genes were enriched in cluster 1, while metabolism regulated genes were enriched in cluster 2. Patients in cluster 2 had a higher score of immune, stroma, and HLA family components, while those in cluster 1 had higher scores of TIDES and immune checkpoint genes. According to the hub genes highly

expressed in cluster 1, 74 small-molecular chemotherapeutics were predicted to be effective for the patients at high risk.

Conclusion: Our results indicate that the CSRs were correlated with the poor prognosis and the possible immunotherapy resistance in LUAD.

Keywords: lung adenocarcinoma, chromosome segregation regulators, prognostic signature, immune environment, bioinformatics

INTRODUCTION

Lung cancer is a malignant tumor with the highest mortality rate in the world (Nooreldeen and Bach, 2021). Lung adenocarcinoma (LUAD) is now the most common histological subtype of primary lung cancer (Travis et al., 2011; Torre et al., 2016; Hutchinson et al., 2019), accounting for more than 40% of cases (Hutchinson et al., 2019). Improvements in multimodal treatment strategies (e.g., targeted therapy, radiotherapy, and immunotherapy) have markedly increased the overall survival (OS) of LUAD-patients in recent years (Moreira and Eng, 2014; Jin et al., 2020), quite a few patients eventually become resistant to these therapies, partly attributed to the malfunctioning of genes that regulate cardinal bioprocesses (Philpott et al., 2017). Thus, sufficient strategies are needed to predict prognosis and guide individual treatment in LUAD. The availability of public cohorts with RNA sequencing data and improved technology brought the opportunity to identify a more generalized prognostic signature for LUAD. For instance, an immune-related four-gene prognostic signature in LUAD was identified to regulate the innate immune response and to be a benefit for the prognosis prediction (Sun et al., 2020). Pyroptosis-related prognostic gene signature and metabolism-associated gene signature were also identified as a predictor of the prognosis in LUAD-patients (He et al., 2020; Lin et al., 2021). However, the novel biomarkers with guiding significance for therapy of LUAD still need to be explored.

Mitotic cell division is commonly thought to involve the equal distribution of duplicated genomes into the two daughter cells through appropriate chromosome segregation (Neumüller and Knoblich, 2009; Sarkar et al., 2021). Abnormal chromosome segregation at mitosis causes the aneuploidy of the daughter cells with an unequal distribution of chromosomes (Levine and Holland, 2018), this is one way by which neoplastic cells accumulate the many genetic abnormalities required for tumor development (Gisselsson, 2008; Naylor and van Deursen, 2016; Levine and Holland, 2018). Moreover, mitotic errors and aneuploidization are found during tumor evolution, and the extent of chromosomal aberrations is correlated with tumor grade and poor prognosis (Loeper et al., 2001; M'Kacher et al., 2010; Bakhoum et al., 2011; Levine and Holland, 2018; Ben-David and Amon, 2020). Cells have well-conserved mechanisms to ensure proper chromosome segregation (Tanaka and Hirota, 2009), whose dysregulation may be involved in tumorigenesis. Therefore, chromosome missegregation is becoming a critical hallmark of tumor biology.

In this study, we determined a chromosome segregation-related gene prognostic signature from The Cancer Genome Atlas (TACG) LUAD cohort. The distinct chromosome segregation regulators (CSRs)-related clusters of LUAD-samples were established, and the overall survival (OS) and the immune environment estimation were assessed in the distinct clusters. This study indicates that the CSRs were correlated with the poor prognosis and the possible immunotherapy resistance in LUAD, which might lay a theoretical foundation for the individualized treatment of LUAD-patients.

MATERIALS AND METHODS

Data Acquisition

The gene expression matrix (HTSEQ-Counts, HTSEQ-FPKM) for LUAD was acquired from the Genomic Commons Data Portal GDC (https://portal.gdc.cancer.gov/) of The Cancer Genome Atlas (TCGA) database (Tomczak et al., 2015), which contains 510 samples of patients and 58 normal samples. Clinical survival data (n = 738) and phenotype data (n = 877) of TCGA-LUAD matched patients were acquired from TCGA database. Excluding patients without information of survival, a total of 497 patients were retained for the prognostic analysis. For the nomogram model construction, 115 patients without information of the clinical index were excluded, and a total of 382 patients were retained.

The external validation sets of the Cox model were constructed using GSE3141 (n = 111) (Bild et al., 2006), GSE13213 (n = 117) (Tomida et al., 2009), GSE31210 (n = 226) (Yamauchi et al., 2012), GSE30219 (n = 278) (Rousseaux et al., 2013), and GSE50081 (n = 181) (Der et al., 2014) datasets from the Gene Expression Omnibus (GEO) database (Barrett et al., 2013). The data was acquired by the GEOquery R package (Zhou J. et al., 2021).

The Cox and Nomogram Model Construction

The gene expression matrix of the 48 CSRs in the 497 LUAD samples was used for the univariate cox regression, LASSO regression, and multivariate cox regression analyses. The Survival R package was used to calculate the correlation between the expression of each gene and overall survival (OS), and genes with a p-value < 0.05 were retained for the following LASSO regression analysis. Glmnet and the survival R package were used for the LASSO regression analysis to screen the significant variables in the univariate cox regression

analysis. In order to obtain more accurate independent prognostic factors (prognostic characteristic genes), multivariate cox regression analysis was used for the final screening. The risk score was calculated as the follows: risk score = (exp-gene1*coef-gene1) + (exp-gene n*coef-gene n). Patients were divided into high- and low-risk groups based on the median of risk score.

Time-dependent receiver operating characteristic (ROC) curves were used to assess survival predictions, and the Time ROC R package was used to calculate the area under the ROC curve (AUC) value to measure prognosis and predict accuracy. Survcomp R package was used for the C-index analysis. For the nomogram analysis, the phenotype data (n = 382) was used and the clinical indexes, including age, gender, race, TNM staging, and stage, were brought into the nomogram analysis. The calibration and decision curve analysis (DCA) were performed to assess the predictive power of the nomogram model.

The correlation of the 48 CSRs with the risk score was determined by Spearman correlation analysis. The Wilcoxon rank-sum test was used for the significant statistics.

Identification of CSRs Pattern in LUAD Patients

Unsupervised clustering analysis (Testa et al., 2021) was used to identify the distinct clusters of LUAD patients according to the expression of 48 CSRs. R "Consensus Cluster Plus" package (Wilkerson and Hayes, 2010) was used for the clustering analysis.

DEGs Determination and the Functional Enrichment Analysis

The differentially expressed genes (DEGs) were calculated using HTSEQ-FPKM of TCGA-LUAD by the Deseq 2 R package (Love et al., 2014), and visualized by the Ggplot 2 R package. The threshold is folded change>2 and $P_{.\ adjust}$ <0.05.

Gene Ontology (GO) (Ashburner et al., 2000) and pathway Kyoto Encyclopedia of Genes and Genomes (KEGG) (Ogata et al., 1999) enrichment were performed by the ClusterProfiler R package (Yu et al., 2012), and visualized by the Ggplot 2 R package.

Identification of Hub Genes

The protein-protein interaction network was constructed by STRING (https://cn.string-db.org/) (Szklarczyk et al., 2019), and visualized by Cytoscape (v3.7.2) (Shannon et al., 2003). The rank of each gene in the network was calculated by CytoHubba (Chin et al., 2014). The Top 50 hub genes were chosen for the following analysis.

TME Estimate Analysis

Stromal score, Immune score, ESTIMATE score, and Tumor purity score were calculated based on mRNA expression matrix (Count, n = 497) by an estimate R package (Yoshihara et al., 2013). Immunization checkpoint block (ICB) assessment was performed by calculating the tumor immune dysfunction and exclusion (TIDE) score. This is a kind of computing algorithm based on gene expression profiles (http://tide.dfci.harvard.edu) (Fu et al., 2020). The difference between HLA family and immune checkpoint genes between the two clusters was performed based on the TPM of these genes.

The immune cell infiltration was calculated respectively by CIBERSORT algorithm (Newman et al., 2015) and xCell algorithm (Aran et al., 2017). The gene expression matrix data (FPKM, n = 497) were uploaded to CIBERSORT, and the 22 types of immune cell infiltration matrix were obtained. The distribution of the immune cell infiltration in each sample was shown using Ggplot 2 R package. The 38 types of cells in xCell algorithm were obtained by immunedeconv R package (Sturm et al., 2020).

The correlation of the 48 CSRs with immunocyte fraction was determined by Spearman correlation analysis. The Wilcoxon rank-sum test was used for the significant statistics.

Correlation Analysis Between CSRs and TMB

For each tumor sample, the total number of somatic mutations (except silent mutations) detected in the tumor is defined as the tumor mutation burden (TMB) (Merino et al., 2020), TMB score for each sample was calculated, and the difference between the two clusters was performed. The correlation of the 48 CSRs with the TMB score was determined by Spearman correlation analysis. The Wilcoxon rank-sum test was used for the significant statistics.

The Chemotherapeutics Forecast

The 50 hub genes were used for the chemotherapeutics forecast, which was performed using the mode of action (moa) module of the connectivity map (CMap, https://clue.io/command) (Gao et al., 2019).

Statistical Analysis

The statistical analysis was calculated via Wilcoxon rank-sum test and unpaired t-text. All statistical tests were bilateral. All statistical tests and visualization were performed in R software (version 4.0.2).

RESULTS

Chromatin Separation Is the Main Bioprocess Enriched by the Prognostic Genes in LUAD

To determine the significant genes involved in the prognosis of patients with LUAD, the batch prognostic analysis of whole genes in samples of TCGA-LUAD was performed. Among the 17,430 genes, a total of 2,416 genes significantly correlated with prognosis were obtained, with a threshold of p-value < 0.05 (**Supplementary Table S1**). GO and KEGG were performed to analyze the functional enrichment of these prognosis related genes, which showed that these genes were enriched in the biological processes of chromosome segregation, organelle

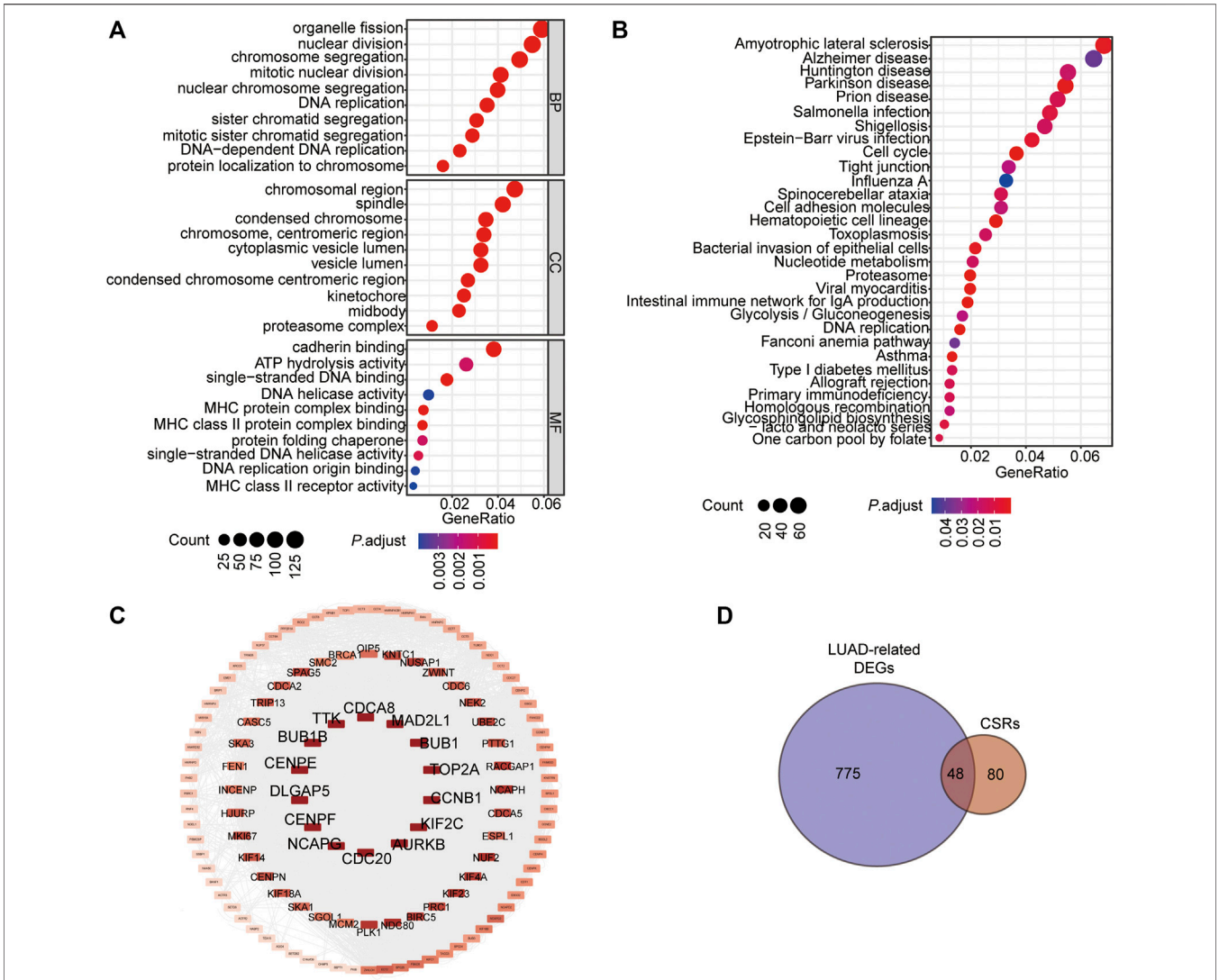


FIGURE 1 | Chromatin Separation is the main bioprocess enriched by the prognostic genes in LUAD. **(A)** GO enrichment of the prognostic genes in LUAD. **(B)** KEGG enrichment of the prognostic genes in LUAD. **(C)** PPI network of the 128 CSRs. **(D)** Venn drawing showing the co genes between the 128 CSRs and LUAD-related DEGs. The LUAD-related DEGs were genes differentially expressed in LUAD-samples compared to the normal samples. Fold change>2, *P*. adjust<0.05.

fission, and nuclear division, the cellular components of chromosomal region, condensed chromosome, centromeric region, spindle, and condensed chromosome, and molecular functions of cadherin binding, single-stranded DNA binding, and ATP hydrolysis activity (**Figure 1A**, **Supplementary Table S2**). Additionally, the KEGG pathways these prognostic genes were enriched in Amyotrophic Lateral sclerosis, Parkinson's disease, Cell cycle, and DNA replication (**Figure 1B**, **Supplementary Table S3**). We noticed that chromosome segregation was the most significant bioprocess enriched by these genes (*P*.adjust = 5.5350E-19). We selected 128 genes enriched in the terms of chromosome segregation, nuclear chromosome segregation, and sister chromatid segregation for the subsequent analysis (**Figure 1C**, **Supplementary Table S4**), and we named them chromosome segregation regulators (CSRs). According to the calculated rank in the Protein-Protein interaction (PPI) network (**Figure 1C**), the top 14 genes were selected, including centromere protein E (*CENPE*), mitotic arrest deficient 2 like 1 (*MAD2L1*), BUB1 mitotic checkpoint serine/threonine kinase (*BUB1*), BUB1 mitotic checkpoint serine/threonine kinase B (*BUB1B*), TTK protein kinase (*TTK*), cell division cycle 20 (*CDC20*), aurora kinase B (*AURKB*), aurora kinase B (*KIF2C*), DNA topoisomerase II alpha (*TOP2A*), DLG associated protein 5 (*DLGAP5*), non-SMC condensing I complex subunit G (*NCAPG*), cyclin B1 (*CCNB1*), centromere protein F (*CENPF*), and cell division cycle associated 8 (*CDCA8*) (**Figure 1C**, **Supplementary Table S5**). Among these genes, a total of 48 CSRs were found to overlap with the 823 DEGs in TCGA-LUAD (**Figure 1D**, **Supplementary Table S6**, **Supplementary Table S7**).

The Kaplan-Meier analysis showed that the 48 CSRs were all correlated with the poor OS in LUAD (**Supplementary Figure**

TABLE 1 | Correlation of CSRs with Risk_Score and TMB.

Gene Name	Risk_Score		TMB	
	Correlation (spearman)	*p*-Value	Correlation (spearman)	*p*-Value
AURKB	0.36	<2.2e-16	0.46	3.31e-27
BIRC5	0.41	<2.2e-16	0.43	2.04e-24
BUB1	0.45	<2.2e-16	0.45	5.99e-6
BUB1B	0.49	<2.2e-16	0.40	4.84e-21
CCNB1	0.48	<2.2e-16	0.38	2.92e-18
CCNE1	0.34	<4.4e-15	0.38	1.33e-18
CDC20	0.43	<2.2e-16	0.47	1.12e-28
CDC6	0.42	<2.2e-16	0.41	6.12e-22
CDCA2	0.46	<2.2e-16	0.38	8.74e-19
CDCA5	0.46	<2.2e-16	0.45	2.7e-26
CDCA8	0.41	<2.2e-16	0.45	8.53e-26
CDT1	0.38	<2.2e-16	0.37	6.67e-18
CENPE	0.45	<2.2e-16	0.39	7.37e-20
CENPF	0.43	<2.2e-16	0.40	6.8e-21
CENPK	0.36	<2.2e-16	0.30	1.53e-11
DLGAP5	0.5944	<2.2e-16	0.42	5.64e-23
EME1	0.26	2.8e-09	0.45	1.4e-26
ESPL1	0.3928	<2.2e-16	0.43	1.61e-23
FAM83D	0.43	<2.2e-16	0.36	2.19e-16
HJURP	0.5546	<2.2e-16	0.46	2.7e-28
KIF14	0.5192	<2.2e-16	0.43	8.08e-24
KIF18B	0.3933	<2.2e-16	0.43	2.17e-24
KIF23	0.5	<2.2e-16	0.44	5.55e-25
KIF2C	0.43	<2.2e-16	0.48	5.13e-30
KIF4A	0.45	<2.2e-16	0.40	4.16e-21
KIFC1	0.4	<2.2e-16	0.47	1.29e-28
KNTC1	0.21	3.2e-06	0.39	3.31e-19
MAD2L1	0.46	<2.2e-16	0.36	4.73e-17
MKI67	0.5	<2.2e-16	0.36	1.44e-16
NCAPG	0.46	<2.2e-16	0.43	1.49e-23
NCAPH	0.45	<2.2e-16	0.47	1.06e-28
NDC80	0.45	<2.2e-16	0.46	6.56e-26
NEK2	0.45	<2.2e-16	0.47	4.71e-29
NUF2	0.354	<2.2e-16	0.52	1.2e-34
NUSAP1	0.47	<2.2e-16	0.39	1.12e-19
OIP5	0.5075	<2.2e-16	0.40	5.67e-21
PLK1	0.6434	<2.2e-16	0.42	1.35e-22
PRC1	0.51	<2.2e-16	0.40	1.43e-20
SKA1	0.45	<2.2e-16	0.44	2.32e-25
SKA3	0.48	<2.2e-16	0.45	8.21e-27
SPAG5	0.38	<2.2e-16	0.46	9.98e-28
SPC24	0.35	2.3e-15	0.39	2.58e-19
SPC25	0.45	<2.2e-16	0.40	6.9e-21
TOP2A	0.38	<2.2e-16	0.45	6.78e-25
TRIP13	0.37	<2.2e-16	0.45	5.54e-27
TTK	0.41	<2.2e-16	0.47	1.48e-29
UBE2C	0.34	2.9e-15	0.47	1.04e-28
ZWINT	0.39	<2.2e-16	0.39	2.54–19

S1). In addition, the Spearman correlation analysis showed that these CSRs were positively associated with the high TMB in LUAD (**Table 1**). These indicated that the 48 CSRs may play important roles in LUAD progression.

The CSRs Are Involved in LUAD Process

The 48 CSRs we identified were used to perform the Univariate Cox regression analysis in TCGA-LUAD (n = 497), and 47 genes were eligible for screening (p-value < 0.05, **Supplementary Table S8**). The 47 genes were then chosen to perform the LASSO regression, and 12 genes were screened out to build a multivariate Cox regression analysis (**Figures 2A,B**, **Supplementary Table S9**). Finally, 9 genes, including *PLK1*, TTK, DLG associated protein 5 (*DLGAP5*), Holliday junction recognition protein (*HJURP*), kinesin family member 14 (*KIF14*), Opa interacting protein 5 (*OIP5*), extra spindle pole bodies like 1, separase (*ESPL1*), kinesin family member 18B (*KIF18B*), and NUF2 component of NDC80 kinetochore complex (*NUF2*), were identified to be independent prognostic signatures (**Figures 2C,D**, **Supplementary Table S10**, **Supplementary Table S11**). The 497 LUAD samples were divided into two subgroups with

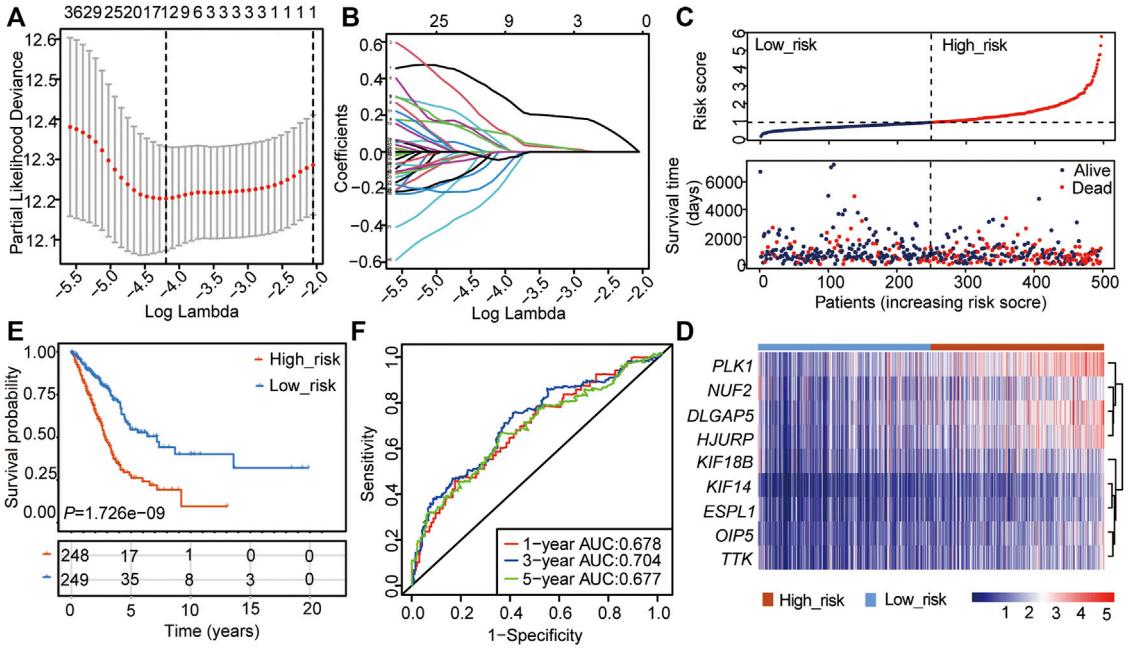


FIGURE 2 | Construction of the cox regression model by the 48 CSRs. **(A)** LASSO coefficient spectrum of prognostic gene screening. **(B)** Stepwise Cox proportional risk regression model to screen the prognostic genes. **(C)** The risk score distribution and survival status of patients in the training cohort. **(D)** The heatmap of prognostic gene distribution in the training cohort. **(E)** The overall survival of high- and low-risk groups. **(F)** ROC analyses of the cox model for 1-, 3-, and 5-years.

TABLE 2 | The clinical index of LUAD patients used in the Cox model.

Characteristic	Levels	Overall	High-Risk	Low-Risk
n (Dead/Alive)		382 (149/233)	204 (103/101)	178 (46/132)
Age, n (%)	>=65	213 (55.76%)	114 (53.52%)	98 (46.48%)
	<65	169 (44.24%)	90 (44.12%)	90 (55.88%)
Gender, n (%)	Male	170 (44.50%)	78 (45.88%)	92 (54.12%)
	Female	212 (55.50%)	126 (59.43%)	86 (40.57%)
N stage, n (%)	N0	249 (65.18%)	129 (51.80%)	120 (48.20%)
	N1	65 (17.01%)	40 (61.53%)	25 (38.47%)
	N2	56 (14.65%)	27 (48.21%)	29 (51.79%)
	N3	1 (0.26%)	1 (100%)	0 (0)
	NX	11 (2.90%)	7 (63.63%)	4 (36.37%)
M stage, n (%)	M0	241 (63.08%)	124 (51.45%)	117 (48.55%)
	M1	21 (5.49%)	8 (38.09%)	13 (61.91%)
	MX	120 (31.43%)	72 (60.00%)	48 (40.00%)
T stage, n (%)	T1	132 (34.55%)	75 (56.82%)	57 (44.18%)
	T2	196 (51.30%)	97 (49.49%)	99 (50.51%)
	T3	40 (10.47%)	22 (55.00%)	18 (45.00%)
	T4	13 (3.40%)	9 (69.23%)	4 (30.77%)
	TX	1 (0.28%)	1 (100%)	0 (0%)
Pathologic stage, n (%)	Stage I	209 (54.71%)	107 (51.20%)	102 (48.80%)
	Stage II	87 (22.77%)	53 (60.92%)	34 (39.08%)
	Stage III	59 (15.44%)	33 (55.93%)	26 (44.07%)
	Stage IV	21 (5.49%)	8 (38.10%)	13 (61.90%)
	N/A	6 (1.59%)	3 (50.00%)	3 (50.00%)

different risk scores: high- and low-risk subgroups, according to the median risk score based on the 9-gene independent prognostic signature (**Figure 2C**). The mortality of the high-risk group was higher than that of the low-risk group (**Figure 2C**), and the patients in the high group had a poor outcome compared with those in the low-risk group (p = 1.726e-09, **Figure 2E**). Meanwhile, the 48 CSRs were all positively associated with the risk score (**Figure 2D**;

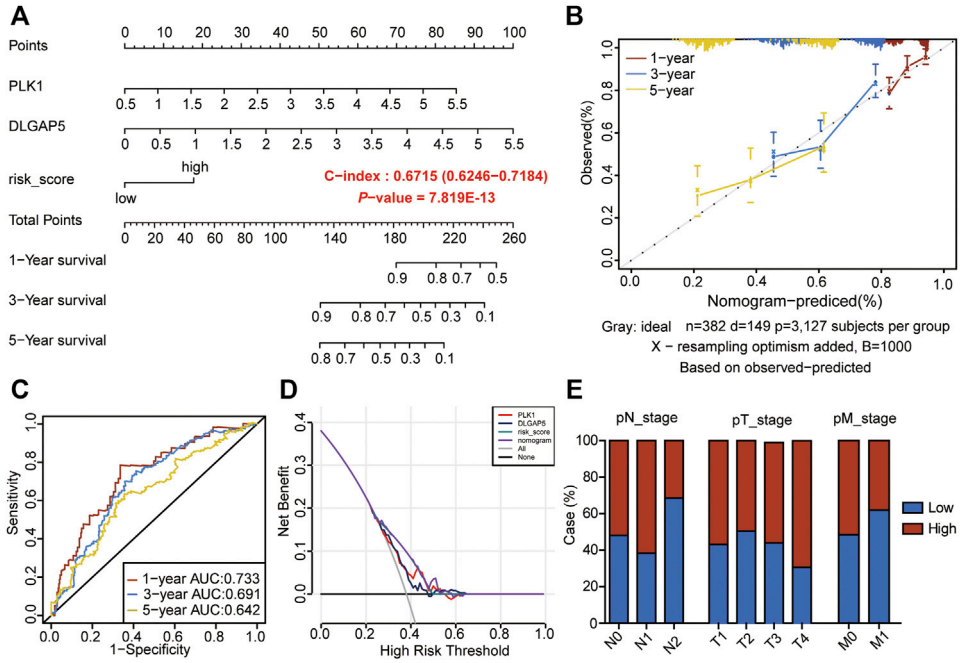


FIGURE 3 | Construction of nomogram model by the 9-gene prognostic signature. **(A)** The nomogram to predict the prognosis of LUAD-patients. **(B)** The calibration analysis of the nomogram predicted a probability of 1-, 3-, and 5-years survival. **(C)** ROC analyses of the nomogram model for 1-, 3-, and 5-years. **(D)** DCA result of the prognostic model. **(E)** Risk groups in relation to a clinical index.

TABLE 3 | Information of GEO data sets used in the validation of the Cox.

Gene Name				
	GPL6480	117 (49/68)	58 (33/25)	59 (16/43)
GSE31210	GPL570	226 (35/191)	113 (28/85)	113 (7/106)
GSE30219	GPL570	278 (188/90)	139 (115/24)	139 (72/66)
GSE3141	GPL570	111 (58/53)	55 (33/22)	56 (25/31)
GSE50081	GPL570	181 (75/106)	91 (48/42)	90 (27/64)

Table 1). The ROC curve was used to predict the prognosis for 1-year, 3-years, and 5-years, which showed that the AUC value ranged from 0.628 to 0.73 (**Figure 2F**). These indicated that the high expression of CSRs predicts the poor OS in LUAD.

Construction of Nomogram Model by the Nine-Gene Prognostic Signature

To further analyze the prognostic significance of the CSRs in LUAD patients, the 9-gene prognostic signature was used to perform the nomogram model with the combination of the clinical indexes, including pT_stage, pN_stage, pM_stage, and stage (**Table 2**). Two significant genes (*PLK1* and *DLGAP5*) were brought into the nomogram model, which showed that the expression values of the two genes and the risk score predicted the survival of LUAD patients (**Figures 3A,B**). The ROC curve showed that the nomogram Cox model could precisely predict the survival of patients, with the AUC ranging from 0.642 to 0.733 (**Figure 3C**). The decision curve analysis (DCA) and calibration analysis of nomogram predicted probability also suggested the accuracy of the Cox model (**Figures 3B,D**). However, there was no significant difference in the pTNM staging between the two risk groups (**Figure 3E**).

Validation of the CSRs-Related Prognostic Signature With GEO Datasets

We used the nine genes to conduct the Cox model in five GEO data sets (GSE3141, GSE13213, GSE30219, GSE31210, and GSE50081), which contains 913 LUAD samples (**Table 3**, **Supplementary Table S12**). The Kaplan-Meier curve showed the patients in the high-risk group had poor outcomes in the five validation sets (**Figures 4A–E**). The AUC value of 1-, 3-, and 5-years in the five validation sets were ranging from 0.612 to 0.788 (**Figures 4F–J**), which verified our previous results that the 9-gene prognostic signature was linked to predicting the OS of LUAD patients.

Identification of Different Clusters Mediated by the 48 CSRs in LUAD.

To investigate the specific function of the CSRs involved in the development of LUAD, the unsupervised consensus clustering analysis was conducted for LUAD samples based on the expression of 48 CSRs. The result showed that the 497 LUAD samples were divided into two distinct subgroups of cluster 1 (C1, n = 338) and cluster 2 (C1, n = 159) (**Figures 5A–D**, **Supplementary Table S13**). The Kaplan-Meier curve showed

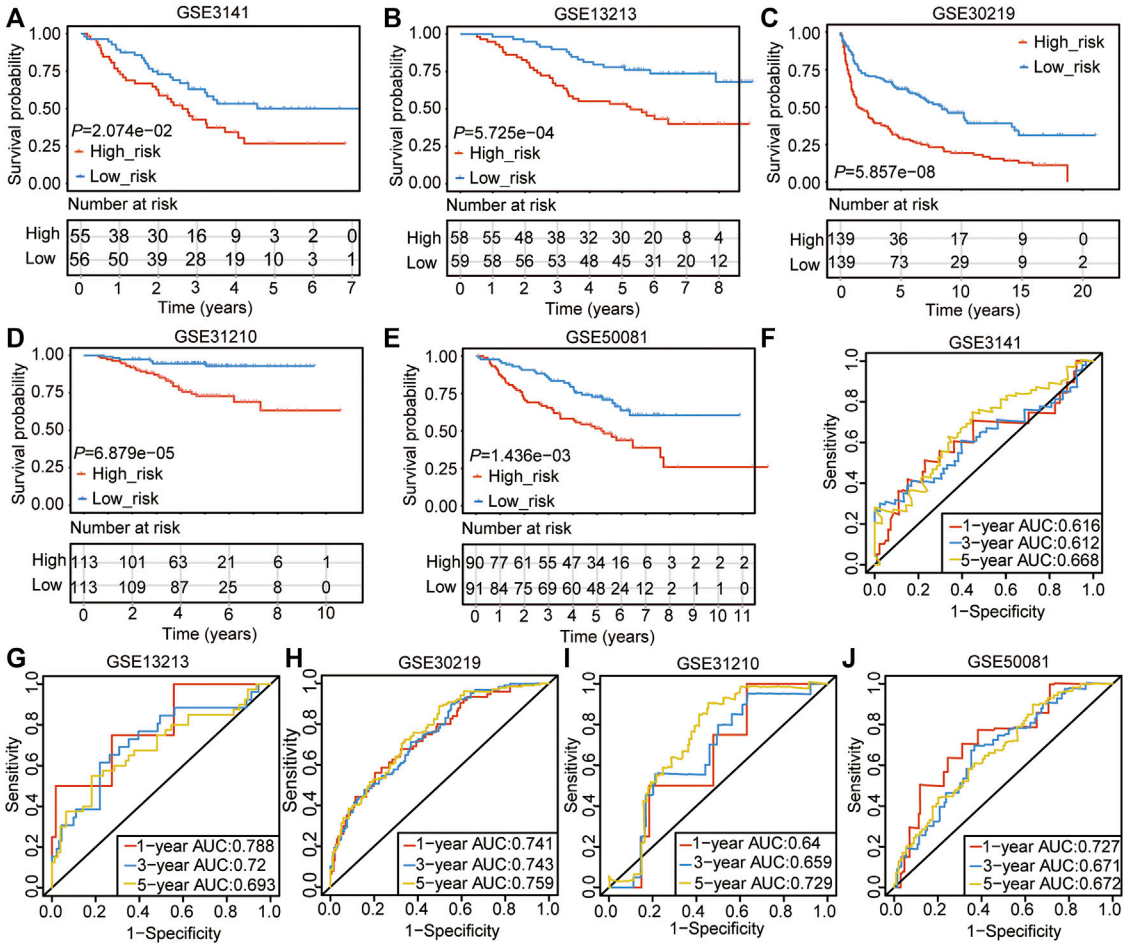


FIGURE 4 | Validation of the CSRs-related prognostic signature with GEO datasets. **(A–E)** The overall survival of the high- and the low-risk group in the five validation sets. **(F–J)** ROC analyses of the cox model for 1-, 3-, and 5-years in the five validation sets.

that patients in C1 had a worse outcome than those in C2 (p = 0.000351, **Figure 5E**). The difference in TMB score between the two clusters was assessed, and the result showed that C1 had a higher TMB score (**Figure 5F**). Moreover, the 48 CSRs were consistently highly expressed in group C1 compared with those in cluster 2 (**Figures 5G,H**, Supplementary Table 14), validating the existence of distinguishing CSRs patterns in LUAD, and the cluster 1 may represent high-risk groups.

DEGs and the Functional Analysis Between the Two CSRs-Related Clusters

Then, we identified the DEGs between the two clusters. As a result, a total of 536 genes (278 up-regulated genes and 258 down-regulated genes) were found to be significantly and differently expressed in group C1 compared to group C2 (**Figures 6A,B**, **Supplementary Table S15**). The subsequent KEGG and GO showed that the up-regulated genes were enriched in the pathways of cell cycle (**Figure 6C**, **Supplementary Table S16**), and the biological processes of chromosome segregation, organelle fission, and nuclear division, and mitotic nuclear division (**Figure 6D**, **Supplementary Table S16**). The down-regulated genes were enriched in the pathways of Complement and coagulation cascades, Arachidonic acid metabolism, and Drug metabolism–cytochrome P450 (**Figure 6E**, **Supplementary Table S16**), and biological processes of antibacterial humoral response, eicosanoid biosynthetic process, protein processing, respond to corticosteroid, respond to glucocorticoid, and eicosanoid metabolic process (**Figure 6F**, **Supplementary Table S16**).

Identification of Hub-Genes Between the Distinct Clusters

The 536 DEGs between the two CSRs-related clusters were used to perform the PPI network, which showed that there was an obvious group with a close correlation occurring in the up-regulated genes (**Figure 7A**). We calculated the rank by cytohubba in the PPI network, and the top 50 hub-genes were selected, which were all up-regulated genes in cluster 1 (**Figure 7B**, **Supplementary Table S17**). The subsequent GO showed these genes were enriched in the biological processes of the mitotic cell cycle, cell division, and regulation of chromosome segregation (**Figure 7C**).

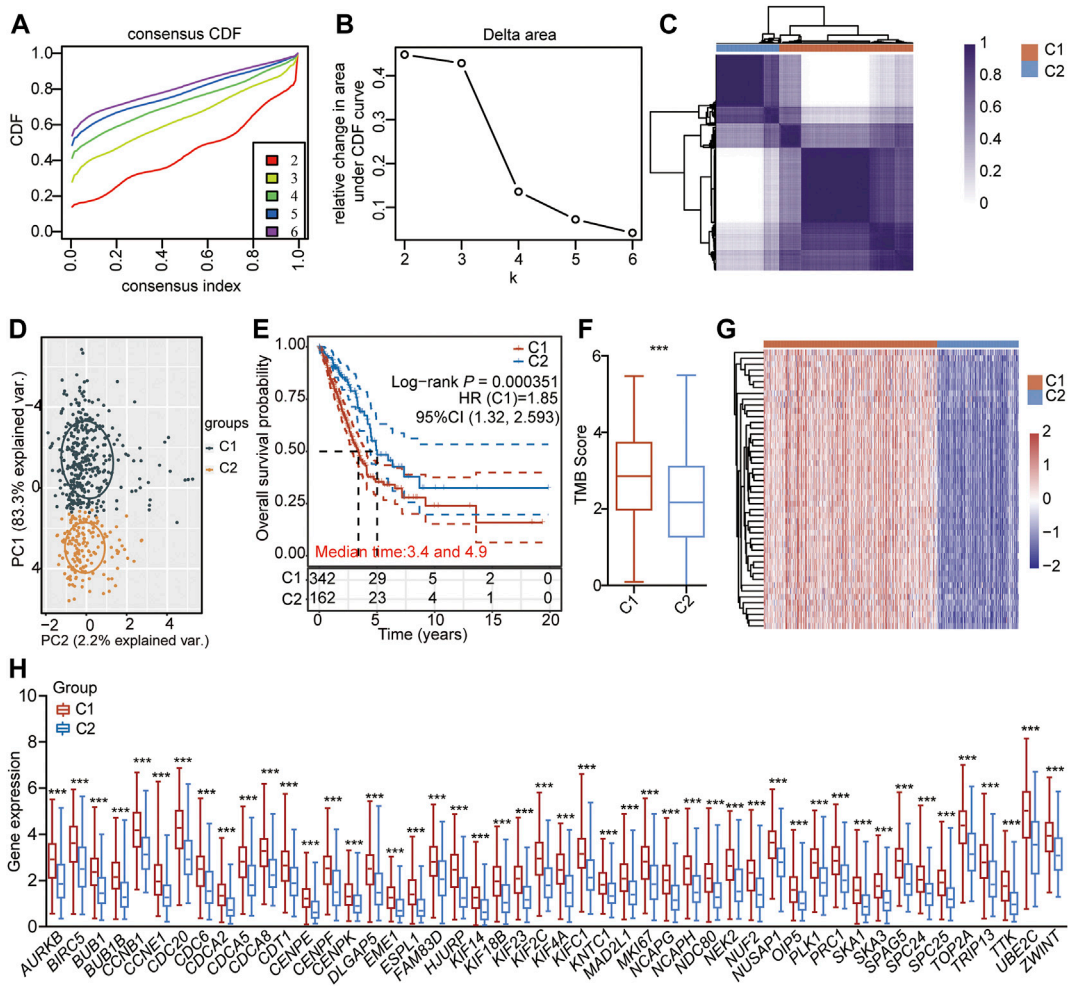


FIGURE 5 | Unsupervised consensus clustering of LUAD-samples based on the 48 CSRs. **(A)** Consistent cumulative distribution shows the cumulative distribution function with different values of k, which is used to judge the optimal value of k. **(B)** Delta area map. **(C)** The consistency matrix of all data sets with k = 2. **(D)** PCA showing the LUAD-samples were divided into two distinct clusters. **(E)** The OS of LUAD-patients in the cluster 1 and 2. **(F)** The difference in TMB between the two clusters. The significance was calculated by Wilcoxon rank-sum test. ***$p < 0.001$. **(G)** Heatmap showing the expression pattern of the 48 CSRs in the two clusters. **(H)** The 48 CSRs were all highly expressed in cluster 1. The significance was calculated by Wilcoxon rank-sum test. ***$p < 0.001$.

CSRs Are Associated With Immunization Checkpoint Block in LUAD

Cancer cells must have evaded the anti-tumor immune response to grow progressively, which relies in part on the expression on their surface of proteins with immunosuppressive functions, such as programmed cell death 1 (PD-L1) (Reisländer et al., 2020). Enhanced ability to escape immune detection always caused malignant development of cancer cells, and the following poor outcome for patients with LUAD. To find out the possible molecular mechanism that the CSRs impact on the prognosis, the difference in immunization checkpoint block (ICB) score between the two clusters, including the levels of immune checkpoint genes, tumor immune dysfunction and exclusion (TIDE) score, and HLA component expression, were evaluated. We found that the C1 had a higher level of risk score (p = 2.26e-21, **Figure 8A**), and TIDE score (p = 1.8e-11, **Figure 8B**). Additionally, the immune checkpoint genes, CD274 (PD-L1), lymphocyte activating 3 (LAG3), programmed cell death 1 (PDCD1 (PD-1)), and programmed cell death 1 ligand 2 (PDCD1LG2), were significantly expressed higher in C1 than in C2 (**Figure 8C**). Moreover, among the eight MHC-II components, seven molecules were expressed lower in C1 than in C2 (**Figure 8D**). One MHC-I component was also expressed lower in C1 (**Figure 8D**). These all indicated that the patients in cluster 1 may have a higher immunization checkpoint block (ICB).

CSRs Are Associated With Immune Characteristics of LUAD

To further explain the possible impact of the CSRs on anti-tumor immune response, the TME characteristics between the two clusters were also assessed. The result of TME showed that the

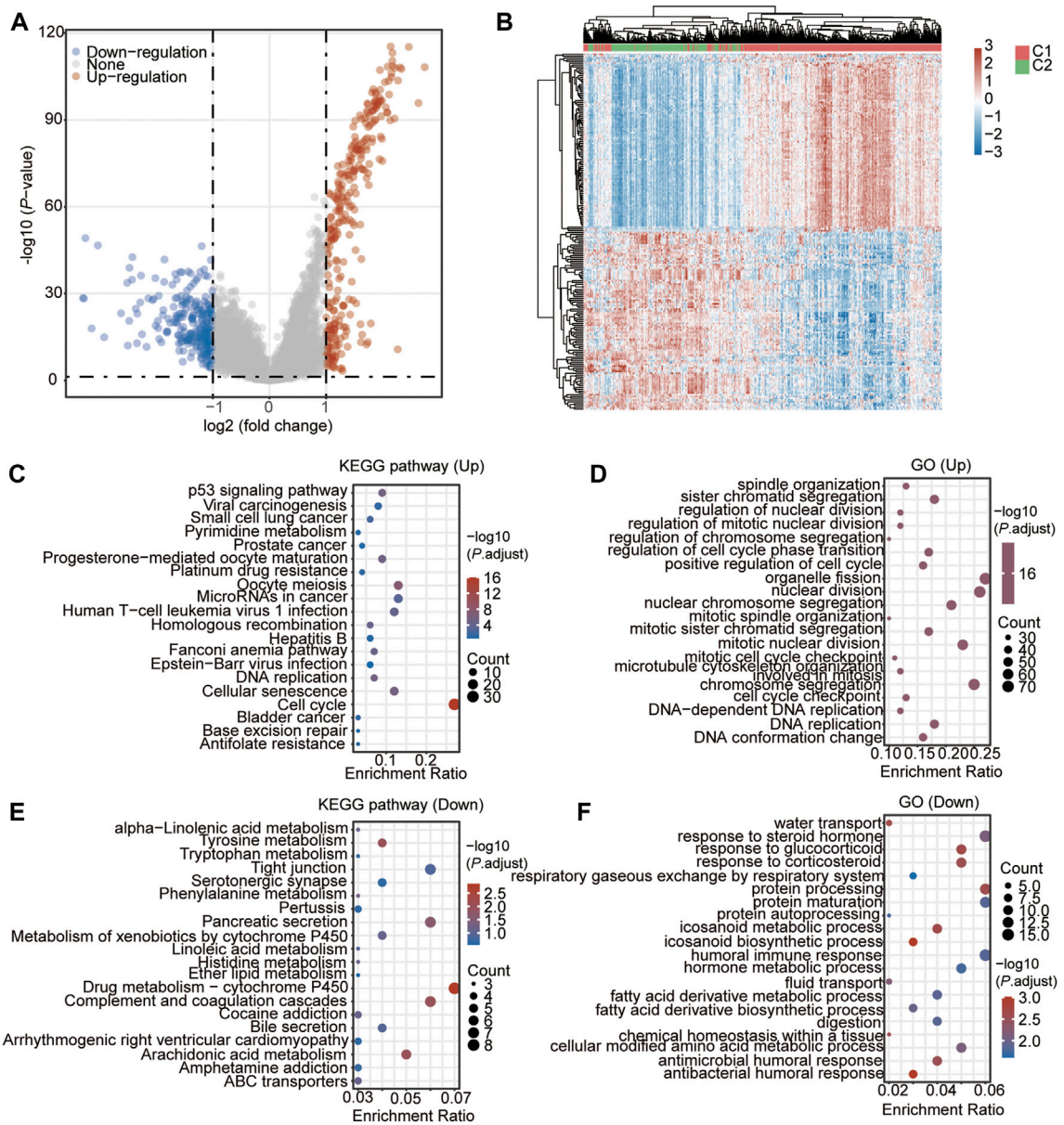


FIGURE 6 | DEGs and the functional analysis between the two CSRs-related clusters. **(A)** Volcano Plot showing the DEGs in cluster 1 vs. cluster 2. **(B)** Heatmap showing the DEGs in cluster 1 vs. cluster 2. **(C)** KEGG enrichment of up-regulated genes in cluster 1. **D** GO enrichment of up-regulated genes in cluster 1. **(E)** KEGG enrichment of down-regulated genes in cluster 1. **(F)** GO enrichment of down-regulated genes in cluster 1.

ESTIMATE score (p = 0.0042), immune score (p = 0.0196), and stromal score (p = 0.0017) were lower in cluster 1 than in cluster 2 (**Figures 9A–C**), while the tumor purity (p = 0.003) was higher in C1 than in C2 (**Figure 9D**). CIBERSORT algorithm was firstly used to calculate the 22 types of immune cells. The result showed that the filtrating proportion of resting memory CD4$^+$ T cells, activated mast cells, resting myeloid dendritic cells, and memory B cells were higher in cluster 2 than in cluster 1 (**Figure 9E**), while the filtrating proportion of follicular helper T cells, CD8$^+$ T cells, M0 macrophages, and M1 macrophages were higher in cluster 1 than in cluster 2 (**Figure 9E**). The xCELL algorithm was further used to assess another immune cell set, which contains 35 kinds of immune cells. The result showed that the proportion of T cell CD8$^+$ naïve, Common lymphoid progenitor, CD4$^+$ Th2, CD4$^+$ Th1, plasmacytoid dendritic cell, and M1 macrophage were higher in cluster 1 (**Figure 9F**), while the content of activated myeloid dendritic cell, granulocyte-monocyte progenitor, mast cell, myeloid dendritic cell, M2 macrophage, memory CD4$^+$ effector T cells, and NK T cells were higher in cluster 2 (**Figure 9F**). Besides this, immune score, microenvironment score, and stroma score were lower in cluster 1 than in cluster 2, which was consistent with our previous result (**Figure 9F**).

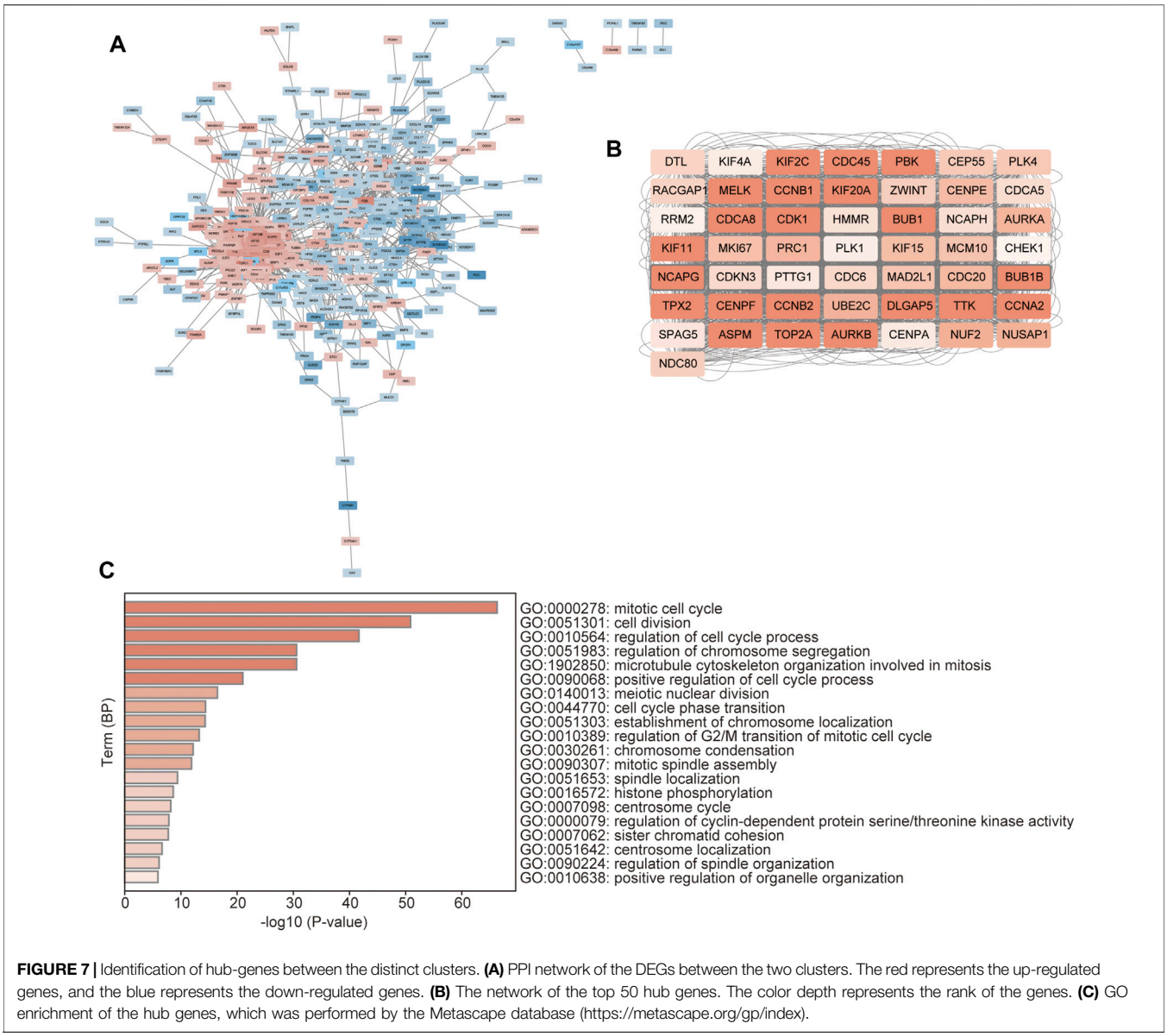


FIGURE 7 | Identification of hub-genes between the distinct clusters. **(A)** PPI network of the DEGs between the two clusters. The red represents the up-regulated genes, and the blue represents the down-regulated genes. **(B)** The network of the top 50 hub genes. The color depth represents the rank of the genes. **(C)** GO enrichment of the hub genes, which was performed by the Metascape database (https://metascape.org/gp/index).

The difference in the abundance of 22 immune microenvironment infiltrating cells calculated by CIBERSORT algorithm between high- and low-risk groups in TCGA cohort and two GEO datasets was further revealed. The result in TCGA cohort showed that the high-risk group has a very similar pattern of infiltrated immune cells as the C1 population (**Figure 9G**). In addition, the infiltrating abundance of M0 and M1 macrophages were significantly higher in the high-risk group in all three datasets (**Figure 9G**, **Supplementary Figure S2**), and the filtrating proportion of resting memory $CD4^+$ T cells was lower in the high-risk group compared to the low-risk group in all three datasets (**Figure 9G**, **Supplementary Figure S2**), indicating that the nine-gene independent prognostic signature was positively correlated with the infiltrating abundance of M0 and M1 macrophages and negatively associated with resting memory $CD4^+$ T cells. These all demonstrated that the CSRs were associated with the immune characteristics in the TME of LUAD.

The Small-Molecular Chemotherapeutics Forecast for High-Risk Patients Based on the Hub-Genes

Our previous results showed that patients in cluster 1 were the predicted high-risk population, and we then wanted to find the potential chemotherapeutics suited for these populations. According to the different biological characteristics between the two clusters, the hub genes were assumed to be the critical and potential targets for the therapy of the high-risk groups. Therefore, the adjuvant chemotherapeutics targeting the 50 hub genes were assessed through the mode of action (moa) module in the CMap database. The result showed that a total of 74 small-

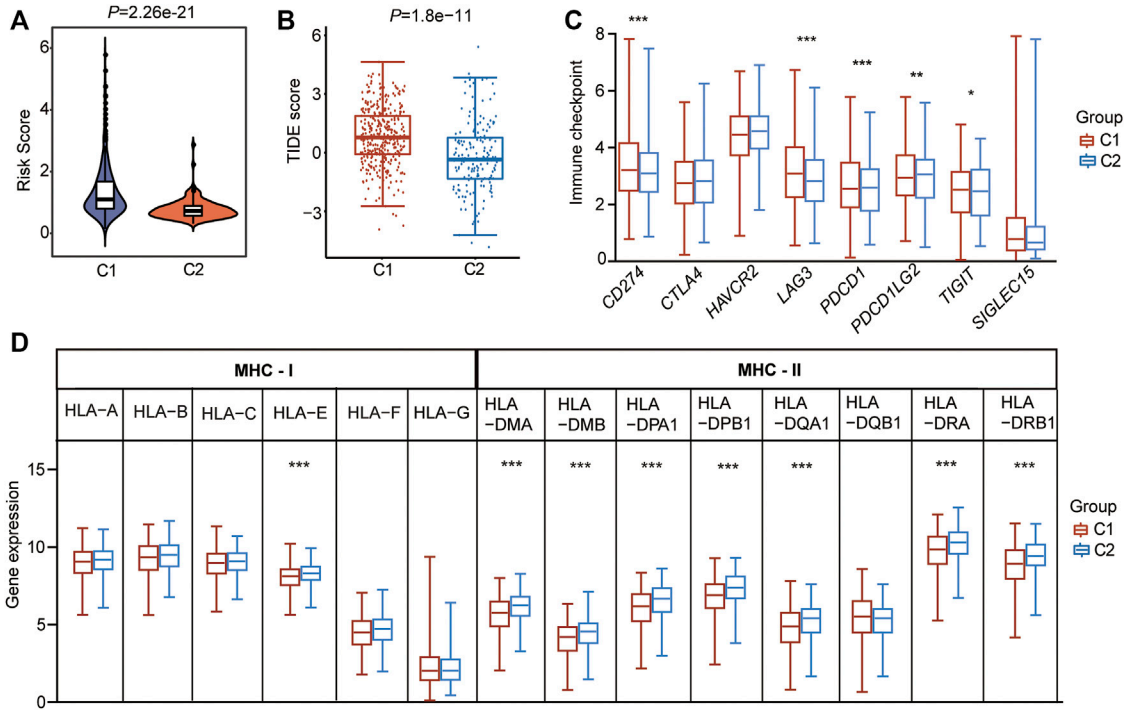


FIGURE 8 | CSRs are associated with immunization checkpoint block in LUAD. **(A)** The risk score in the two clusters. **(B)** The TIDE score in the two clusters. **(C)** The immune checkpoint genes in the two clusters. **(D)** The HLA family genes in the two clusters. The statistical significance was calculated *via* Wilcoxon rank-sum test, ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.

molecular perturbagens, targeting aurora kinase B (*AURKA*), cyclin-dependent kinase 1 (*CDK1*), *TOP2A*, *AURKB*, polo-like kinase 4 (*PLK4*), *PLK1*, ribonucleotide reductase regulatory subunit M2 (*RRM2*), *CCNB1*, *TTK*, and cyclin A2 (*CCNA2*), were predicted to be the potential chemotherapeutics for patients in cluster 1 (**Figure 10**). A total of 47 moas were predicted to be the possible pathways by which these chemotherapeutics function, such as Aurora kinase inhibitor, CDK inhibitor, and topoisomerase inhibitor (**Figure 10**). Among the 74 potential chemotherapeutics, AT-9283, indirubin, and LY-294002 were the most outstanding ones, which function as the most common pathways (marked in blue, **Figure 10**).

DISCUSSION

Lung cancer is the most common malignancy and remains the leading cause of cancer mortality worldwide (Bade and Cruz, 2020). Approximately 85% of patients have a histological subtype known as non-small cell lung cancer (NSCLC) (Herbst et al., 2018), with the main subtype of LUAD (Herbst et al., 2018; Sun and Zhao, 2019). Recently, LUAD has been the major cause of cancer-associated mortality with a poor prognosis (Song et al., 2021). Chromosomal abnormalities are reported to be a common characteristic of cancer, which is attributed to the ongoing chromosome segregation errors during mitosis (Bakhoum et al., 2018; Kou et al., 2020a). It is associated with poor outcomes and therapeutic resistance in various cancers (Potapova and Gorbsky, 2017; Bakhoum and Cantley, 2018; Kou et al., 2020a). Chromosomal segregation errors or the alteration of CSRs are usually the critical factors of genomic instability that drive tumor evolution (Hanahan and Weinberg, 2011; Bolhaqueiro et al., 2019; Tayoun et al., 2021). For example, aneuploidy is a direct consequence of chromosome segregation errors in mitosis (Janssen et al., 2011; Kawakami et al., 2019). This is always accompanied by chromosomal instability (CIN) and genomic instability and causes additional chromosome gains or losses in a significant proportion of cell divisions, which further manifests the complexity of cancer karyotypes (Dürrbaum and Storchová, 2016). Aneuploidy correlates with increased metastatic and drug resistance, indicating that Aneuploidy is more beneficial for cancerous cells than diploid cells (Dürrbaum and Storchová, 2016). The genomic diversification caused by chromosome segregation errors usually promotes tumor evolution and heterogeneity (Soto et al., 2019), which is an important reason for therapeutic resistance (Dagogo-Jack and Shaw, 2018; Lim and Ma, 2019). Although aneuploidy provides advantages for the proliferation and drug resistance of cancer cells, excessive aneuploidy beyond a critical level is lethal to cancer cells (Kawakami et al., 2019). Therefore, molecular and bioprocesses engaged in chromosome segregation should be utilized as potential therapeutic targets for cancers.

Previous studies have indicated that CIN and loss of heterozygosity (LOH) play significant roles in the development of LUAD (Ninomiya et al., 2006). In this study, chromosomal segregation during mitosis was found to be correlated with the

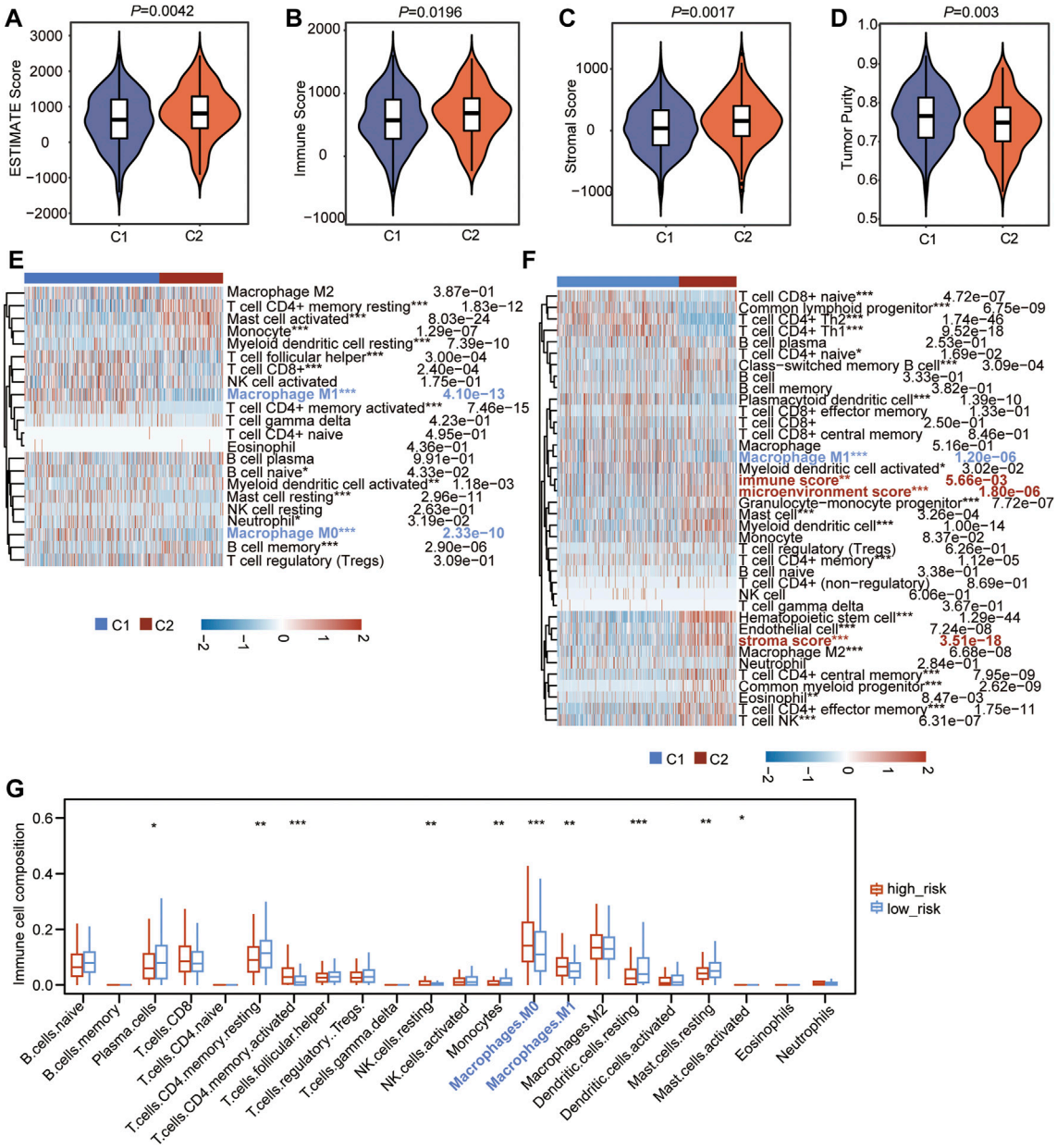


FIGURE 9 | CSRs are associated with immune characteristics of LUAD. **(A–D)** Comparison of estimate score **(A)**, immune score **(B)**, stromal score **(C)**, and tumor purity **(D)** in the two clusters. **(E)** Heatmap showing the 22 types of immune cells infiltration in the two clusters by CIBERSORT algorithm. **(F)** Heatmap showing the 27 types of immune cells infiltration in the two clusters by xCell algorithm. The red label represents the cells that were highly infiltrated in cluster 2, the blue label represents the cells that were highly infiltrated in cluster 1. The statistical significance was calculated via Wilcoxon rank-sum test, ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$. **(G)** The difference in infiltrated abundance of immune cells between the two risk groups. The infiltrated abundance of immune cells was calculated by CIBERSORT algorithm. The statistical significance was calculated via Wilcoxon rank-sum test, ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.

poor prognosis of LUAD. Among the determined 2,416 prognosis-related genes, 128 genes were found to be enriched in the biological processes of chromosomal segregation, indicating that these chromosomal segregation-related genes may play an important role in the development of LUAD. 48 of 128 genes were found to be up-regulated in LUAD compared with the normal people and were associated with the high TMB in patients with LUAD. Besides, 47 genes were identified as the prognostic signatures after suffering from univariate cox regression. By LASSO and multivariate cox regression, nine genes were finally determined as the independent prognostic signature to construct the nomogram cox model. Patients in the high-risk subgroup had significantly poor OS. Notably, *PLK1* and *DLGAP5* were found to have valuable significance for the prediction of prognosis. PLK1 plays multiple roles in the initiation, maintenance, and

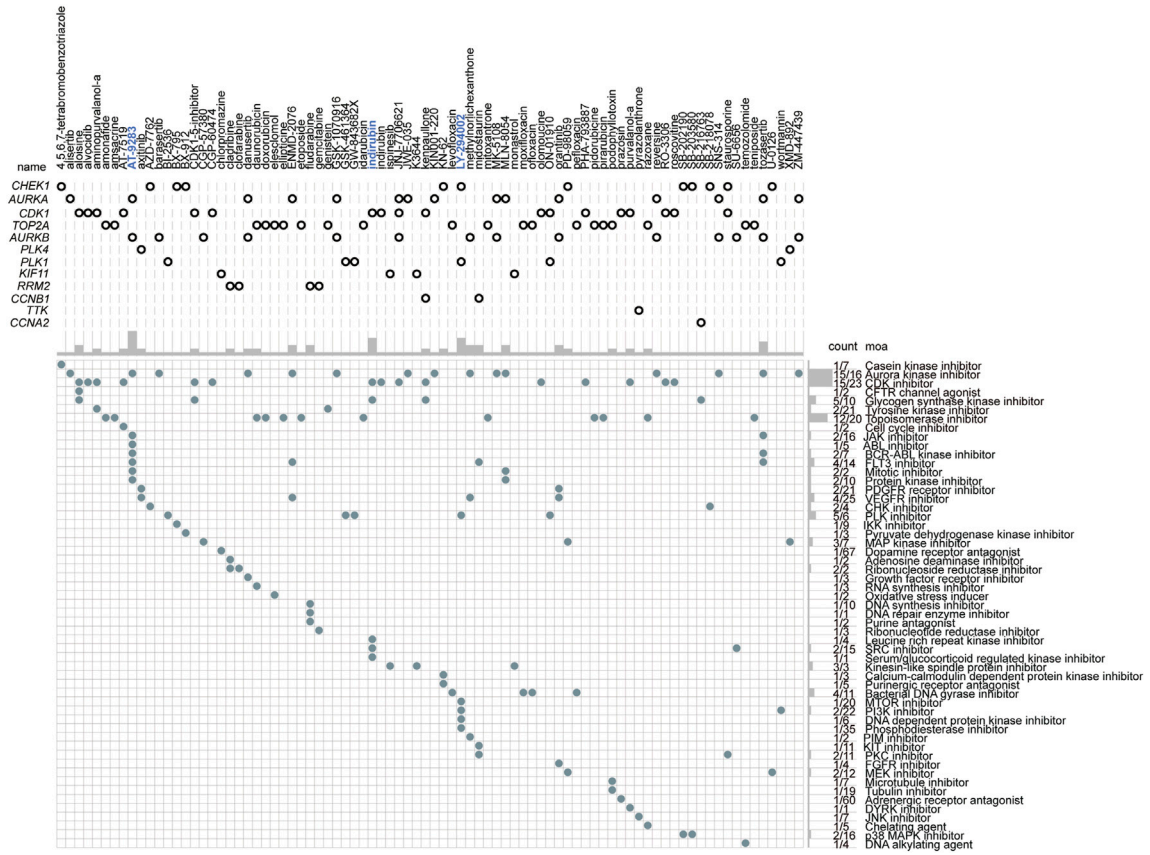


FIGURE 10 | The small-molecular chemotherapeutics forecast for patients with high-risk based on the hub genes. The blue label represents the most outstanding small-molecular perturbagens.

completion of mitosis (Liu et al., 2017), maintaining genome stability (Zhang et al., 2016; Gheghiani et al., 2021), and DNA damage response (DDR) (Peng et al., 2021). It is found to be highly expressed in most of the human cancers (Gutteridge et al., 2016), and enhanced gene expression is associated with a poor prognosis (Ramani et al., 2015; Tut et al., 2015; Zhang et al., 2015; Chen Z. et al., 2019; Montaudon et al., 2020). Besides, PLK1 has been reported to be associated with chemotherapeutic drugs resistance, including doxorubicin (Wang P. et al., 2018), paclitaxel (Gasca et al., 2016; Shin et al., 2019; Shin et al., 2020), metformin (Shao et al., 2015; Zhu et al., 2022), and gemcitabine (Song et al., 2013; Li et al., 2016; Mao et al., 2018). DLGAP5 is a microtubule-associated protein and is identified to be a prognosis biomarker in various cancers (Schneider et al., 2017; Branchi et al., 2019; Xu et al., 2020; Zheng et al., 2020; Feng Y. et al., 2021; Zhou D. et al., 2021). Inhibition of this gene suppresses cell proliferation and invasion, induces G_2/M phase arrest, and promotes apoptosis in human cancers (Liao et al., 2013; Zhang et al., 2021a). Not only these findings were confirmed, but also the predictive value of the two genes was proposed in this study. Here, *PLK1* and *DLGAP5* were found as the independent prognostic signatures combined with *HJURP*, *KIF14*, *OIP5*, *TTK*, *ESPL1*, *KIF18B*, and *NUF2* to predict the prognosis of LUAD-patients. *HJURP* is reported to be an oncogene that promotes cancer cell proliferation, migration, and invasion (Chen et al., 2018; Wei et al., 2019; Serafim et al., 2020; Li Y. et al., 2021; Lai et al., 2021). *KIF14* is confirmed to promote cancer cell proliferation and contribute to chemoresistance (Singel et al., 2014; Wang Z. Z. et al., 2018; Xiao et al., 2021). Silencing *TTK* is also found to inhibit the proliferation and invasion and increase radiosensitivity and chemosensitivity of cancer cells (Chen S. et al., 2019; Huang et al., 2020; Liu Y. et al., 2021; Zhang et al., 2021b; Qi et al., 2021). *ESPL1* is determined to be a novel prognostic biomarker and is associated with the malignant features in several cancers (Finetti et al., 2014; Wang R. et al., 2020; Liu Z. et al., 2021). *KIF18B* is also illustrated to be the oncogenesis to promote tumor progression and enhance therapeutic resistance (Li B. et al., 2020; Liu et al., 2020; Jiang J. et al., 2021). *OIP5* upregulation is observed in human cancers (Chow et al., 2021), and seems to be linked to drug resistance (Rodrigues-Junior et al., 2018). *NUF2* is found to be a prognostic biomarker and therapeutic target which is correlated with the immune infiltration in patients with cancer (Jiang X. et al., 2021; Shan et al., 2021; Xie et al., 2021). Combining the expression of the nine genes, we conducted the cox risk model. The subsequent DCA and calibration verified the accuracy of the model to predict the prognosis of LUAD. However, we found the nine-gene prognostic signature had no impact on the

clinical pNTM stage of patients. This may be attributed to the small sample size, thus, larger samples should be analyzed further.

According to the unsupervised consensus clustering by the expression of the 48 CSRs, the 497 TCGA-LUAD samples were divided into clusters 1 and 2. The 48 CSRs were all highly expressed in cluster 1, and the risk score was significantly higher in cluster 1. The Kaplan-Meier showed that patients in cluster 1 had a poor OS. These all suggest that the patients in cluster 1 have a higher risk than those in cluster 2, and the subgroup cluster 1 is deemed to be the high-risk group. A total of 536 genes (278 up-regulated genes and 258 down-regulated genes) were determined to be the DEGs in cluster 1 compared to cluster 2. The up-regulated genes were enriched in the pathways of cell cycle and the biological processes of chromosome segregation, organelle fission, and nuclear division; the down-regulated genes were enriched in the pathways of drug metabolism-cytochrome P450 and the biological process of protein processing. The main function of cytochrome P450 (CYP450) is oxidative catalysis of endogenous and exogenous substances (Mittal et al., 2015). 80% of drugs currently in use, including anti-cancer drugs, are involved in phase I metabolism of CYP450 (Mittal et al., 2015). Interestingly, the hub genes with top 50 ranks were all highly expressed in cluster 1, which were enriched in the biological processes of mitotic cell cycle, regulation of chromosome segregation, and microtubule cytoskeleton organization involved in mitosis. As we discussed previously, the dysregulation of mitotic chromosome segregation is associated with poor prognosis and therapeutic resistance in human cancers. Among the 50 hub genes, *AURKB*, *BUB1*, *BUB1B*, *CCNB1*, *CDCA8*, *CENPF*, *DLGAP5*, *KIF2C*, *NCAPG*, *TOP2A*, *TTK*, assembly factor for spindle microtubules (*ASPM*), cyclin A2 (*CCNA2*), cyclin B2 (*CCNB2*), cell division cycle 45 (*CDC45*), cyclin-dependent kinase 1 (*CDK1*), kinesin family member 11 (*KIF11*), kinesin family member 20A (*KIF20A*), maternal embryonic leucine zipper kinas (*MELK*), PDZ binding kinase (*PBK*), TPX2 microtubule nucleation factor (*TPX2*) were found to be the hub genes with the rank 1, that were all reported to have pro-metastatic effects in human cancers (Chen et al., 2015; Bertran-Alamillo et al., 2019; Gan et al., 2019; Huang et al., 2019; Kou et al., 2020b; Wang L. et al., 2020; Wang X. et al., 2020; Hu et al., 2021; Wang S. et al., 2021; Li M.-X. et al., 2021; Wu et al., 2021; Feng Z. et al., 2021; Huang et al., 2021).

The difference in ICB between the two clusters was further assessed, including TIDE score, immune checkpoint genes, HLA family components to evaluate the possible anti-tumor immune response associated with the CSRs. The cluster 1 subtype featured lower levels of HLA family genes, higher TIDE scores, and higher levels of immune checkpoint genes, indicating that patients in cluster 1 have the higher ICB and, in turn, a possible poor anti-tumor immune response. The highly expressed immune checkpoint genes, *CD274*, *LAG3*, *PDCD1*, *PDCD1LG2*, in this subgroup further confirmed the result. PDCD1 (PD-1) is a key coinhibitory receptor expressed on activated T cells (Ai et al., 2020). The engagement with its ligands, mainly PD-L1, leads to the events of inhibition of T cell proliferation, activation, cytokine production, alters metabolism and cytotoxic T lymphocytes (CTLs) killer functions, and eventual death of activated T cells (Ai et al., 2020). The overexpression of PD-L1 has been verified to contribute to the immune surveillance evasion of cancer cells and caused the invasion and migration (Iwai et al., 2002). PDCD1LG2 (PD-L2) is a second ligand for PD-1 and inhibits T cell activation (Latchman et al., 2001). LAG3 (CD223), an emerging targetable inhibitory immune checkpoint molecule, is the third inhibitory receptor pathway to be targeted in the clinic (Solinas et al., 2019). It is mainly found on activated immune cells and is involved in the exhaustion of T cells in malignant diseases (Puhr and Ilhan-Mutlu, 2019). LAG3 has been reported to play a negative regulatory role in cancer immunology by interacting with its ligands (Wang M. et al., 2021). What's more, seven MHC-II and one MHC-I molecules were found to be under-expressed in cluster 1, further supporting the result that patients in cluster 1 have a poor antitumor immune response. MHC-I molecules function to bind the encoded peptides, transport and display the antigenic information on the cell surface, and allow $CD8^+$ T cells to identify pathological cells, such as cancers that are expressing mutated proteins (Dhatchinamoorthy et al., 2021). Loss of MHC-I antigen presentation always leads to cancer immune evasion (Dhatchinamoorthy et al., 2021). MHC-II is an antigen-presenting complex, that is important for antigen presentation to $CD4^+$ T cells. Tumor-specific MHC-II is associated with favorable outcomes in patients with cancer, including those with immunotherapies (Axelrod et al., 2019). The lower level of MHC-II and one MHC-I molecule in patients of cluster 1 may help the tumor evasion of immune checkpoints and poor immunotherapies. What's more, we found a higher TMB score in cluster 1, which was positively associated with the 48 CSRs. A recent study has shown that the average copy number variation (CNVA) of chromosome fragments is a potential surrogate for tumor mutational burden in predicting responses to immunotherapy in NSCLC (Lei et al., 2021). Therefore, the increased genomic instability in tumors with dysregulated chromosome segregation alters mutational load and in turn impacts antitumoral immune responses, which further impacts the prognosis in LUAD. However, we cannot define either CIN was a cause or result of somatic mutation, and this needs more experiment assays to be determined.

Meanwhile, the infiltrating proportions of immune cells were further evaluated to analyze the difference in immune characteristics between the two clusters. We found that cluster 1 had a lower immune, stromal, and ESTIMATE score compared with cluster 2. According to the CIBERSORT and xCell algorithms, cluster 1 was shown to have a higher infiltrating proportion of M0 and M1 macrophages, while cluster 2 had a higher infiltrating proportion of M2 macrophages and mast cells. Macrophages play a critical role in cancer development and metastasis, which could be identified as 2 major subpopulations of M1 macrophages (Proinflammatory) and M2 macrophages (anti-inflammatory) (Xia et al., 2020). M0 macrophages are naïve macrophages without polarization. M1 macrophages have antimicrobial and antitumoral activity, while M2 macrophages participate in angiogenesis, immunoregulation, tumor formation, and progression (Shapouri-Moghaddam et al., 2018). Increased mast cell density is associated with the prognosis and plays a multifaced role in TME by regulating tumor biology, including cell proliferation, angiogenesis, invasiveness, and

metastasis (Aponte-López and Muñoz-Cruz, 2020). These results seem contradictory that M1 macrophages have antitumoral activity but were higher infiltrated in cluster 1, whereas M2 macrophages have protumoral activity but were lower infiltrated in cluster 1, as were the follicular helper T cells and $CD8^+$ T cells. We think this elevation may be a compensatory effect, that these antitumoral immune cells are regulated to increase against the tumor cells. However, the effective activity of these anti-tumor immune cells may be inhibited because of the high levels of the immunosuppressive proteins on their surface. Most tumor-infiltrating immune cells were functionally inactive.

Based on the 50 hub genes that were upregulated in cluster 1, a total of 74 small-molecule perturbagens were predicted to be the potential chemotherapeutics that were possibly suited for these patients at high risk. Among these, AT-9283, indirubin, and LY-294002 were the most outstanding ones, which function as the most common pathways. A total of 47 moas, including Aurora kinase inhibitor, CDK inhibitor, and topoisomerase inhibitor, were forecasted to be the possible molecular mechanism by these chemotherapeutics functions. The inhibitory role of AT-9283 in cancer cell growth and survival has been demonstrated in cell-based systems (Kimura, 2010). Although phase II clinical trials have not been completed, it showed good safety and efficacy in phase I clinical trials conducted in patients with hematological malignancies and solid tumors (Kimura, 2010). Indirubin has also been reported to exert anticancer effects in human cancers (Li Z. et al., 2020). LY-294002 enhances the chemosensitivity of liver cancer to oxaliplatin (Xu et al., 2021). We found here these small-molecule perturbagens might have favorable therapeutic effects for patients in cluster 1, who had the high expression of their target genes. This will provide an effective therapeutic regimen for the individual treatment in LUAD.

Our study is the first one to systematically analyze the relationship between chromosome segregation regulation and the immune microenvironment. This can provide a new direction for immune-related LUAD pathogenesis and therapeutic research. However, there are certain limitations in this current study. First, although the exploration of a decent number of samples and summarizing data can be helpful in the research community, this analysis may have some bias due to the small size of TCGA-LUAD cohort. The relationship between the prognostic signatures and clinical indices, such as pNTM staging, has not been well explained. Therefore, further analysis with larger samples is still needed. Second, further experimental verifications are necessary to elucidate the potential impact of these predicted genes in the immune microenvironment. Moreover, the protein expression levels of the hub CSRs in pathogenesis and progression of LUAD depend on further experimental studies to elucidate. Additionally, some of the genes we focused on in this study have been reported to be activated by post-translational modification and function as kinase. Therefore, database analyses depending on the gene expression profile at the mRNA levels have a limit. Further analysis should be focused on the protein or post-translational modification levels.

In conclusion, this study demonstrated that the CSRs were important factors to influence the development and progression of LUAD. The high expression of these regulators was correlated with the poor prognosis and the possible immunotherapeutic resistance in LUAD, which could be the potential therapeutic target for LUAD.

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

ZL, GZ, and XZ designed experiments and interpreted data. ZL, ZM, HX, RS, KQ, and YZ conducted bioinformatic and statistical analyses. ZL wrote the paper. All authors have read and approved the manuscript for publication. GZ conceived the concept, designed the manuscript, coordinated, and critically revised manuscript, and was responsible for its financial supports and the corresponding works.

FUNDING

This work was supported by grants from the Natural Science Foundation of Shandong Province (ZR2021MH274 to XZ.).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.917150/full#supplementary-material

REFERENCES

Ai, L., Xu, A., and Xu, J. (2020). Roles of PD-1/pd-L1 Pathway: Signaling, Cancer, and beyond. *Adv. Exp. Med. Biol.* 1248, 33–59. doi:10.1007/978-981-15-3266-5_3

Aponte-López, A., and Muñoz-Cruz, S. (2020). Mast Cells in the Tumor Microenvironment. *Adv. Exp. Med. Biol.* 1273, 159–173. doi:10.1007/978-3-030-49270-0_9

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

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene Ontology: Tool for the Unification of Biology. The Gene Ontology Consortium. *Nat. Genet.* 25 (1), 25–29. doi:10.1038/75556

Axelrod, M. L., Cook, R. S., Johnson, D. B., and Balko, J. M. (2019). Biological Consequences of MHC-II Expression by Tumor Cells in Cancer. *Clin. Cancer Res.* 25 (8), 2392–2402. doi:10.1158/1078-0432.ccr-18-3200

Bade, B. C., and Cruz, C. S. D. (2020). Lung Cancer 2020: Epidemiology, Etiology, and Prevention. *Clin. Chest Med.* 41 (1), 1–24. doi:10.1016/j.ccm.2019.10.001

Bakhoum, S. F., and Cantley, L. C. (2018). The Multifaceted Role of Chromosomal Instability in Cancer and its Microenvironment. *Cell* 174 (6), 1347–1360. doi:10.1016/j.cell.2018.08.027

Bakhoum, S. F., Danilova, O. V., Kaur, P., Levy, N. B., and Compton, D. A. (2011). Chromosomal Instability Substantiates Poor Prognosis in Patients with Diffuse Large B-Cell Lymphoma. *Clin. Cancer Res.* 17 (24), 7704–7711. doi:10.1158/1078-0432.ccr-11-2049

Bakhoum, S. F., Ngo, B., Laughney, A. M., Cavallo, J.-A., Murphy, C. J., Ly, P., et al. (2018). Chromosomal Instability Drives Metastasis through a Cytosolic DNA Response. *Nature* 553 (7689), 467–472. doi:10.1038/nature25432

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-Uupdate. *Nucleic Acids Res.* 41 (Database issue), D991–D995. doi:10.1093/nar/gks1193

Ben-David, U., and Amon, A. (2020). Context Is Everything: Aneuploidy in Cancer. *Nat. Rev. Genet.* 21 (1), 44–62. doi:10.1038/s41576-019-0171-x

Bertran-Alamillo, J., Cattan, V., Schoumacher, M., Codony-Servat, J., Giménez-Capitán, A., Cantero, F., et al. (2019). AURKB as a Target in Non-small Cell Lung Cancer with Acquired Resistance to Anti-EGFR Therapy. *Nat. Commun.* 10 (1), 1812. doi:10.1038/s41467-019-09734-5

Bild, A. H., Yao, G., Chang, J. T., Wang, Q., Potti, A., Chasse, D., et al. (2006). Oncogenic Pathway Signatures in Human Cancers as a Guide to Targeted Therapies. *Nature* 439 (7074), 353–357. doi:10.1038/nature04296

Bolhaqueiro, A. C. F., Ponsioen, B., Bakker, B., Klaasen, S. J., Kucukkose, E., van Jaarsveld, R. H., et al. (2019). Ongoing Chromosomal Instability and Karyotype Evolution in Human Colorectal Cancer Organoids. *Nat. Genet.* 51 (5), 824–834. doi:10.1038/s41588-019-0399-6

Branchi, V., García, S. A., Radhakrishnan, P., Győrffy, B., Hissa, B., Schneider, M., et al. (2019). Prognostic Value of DLGAP5 in Colorectal Cancer. *Int. J. Colorectal Dis.* 34 (8), 1455–1465. doi:10.1007/s00384-019-03339-6

Chen, H., Lee, J., Kljavin, N. M., Haley, B., Daemen, A., Johnson, L., et al. (2015). Requirement for BUB1B/BUBR1 in Tumor Progression of Lung Adenocarcinoma. *Genes. Cancer* 6 (3-4), 106–118. doi:10.18632/genesandcancer.53

Chen, S., Wang, J., Wang, L., Peng, H., Xiao, L., Li, C., et al. (2019). Silencing TTK Expression Inhibits the Proliferation and Progression of Prostate Cancer. *Exp. Cell Res.* 385 (1), 111669. doi:10.1016/j.yexcr.2019.111669

Chen, T., Huang, H., Zhou, Y., Geng, L., Shen, T., Yin, S., et al. (2018). HJURP Promotes Hepatocellular Carcinoma Proliferation by Destabilizing P21 via the MAPK/ERK1/2 and AKT/GSK3β Signaling Pathways. *J. Exp. Clin. Cancer Res.* 37 (1), 193. doi:10.1186/s13046-018-0866-4

Chen, Z., Chai, Y., Zhao, T., Li, P., Zhao, L., He, F., et al. (2019). Effect of PLK1 Inhibition on Cisplatin-resistant Gastric Cancer Cells. *J. Cell. Physiology* 234 (5), 5904–5914. doi:10.1002/jcp.26777

Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., and Lin, C. Y. (2014). cytoHubba: Identifying Hub Objects and Sub-networks from Complex Interactome. *BMC Syst. Biol.* 8 (Suppl. 4), S11. doi:10.1186/1752-0509-8-S4-S11

Chow, M. J., Gu, Y., He, L., Lin, X., Dong, Y., Mei, W., et al. (2021). Prognostic and Therapeutic Potential of the OIP5 Network in Papillary Renal Cell Carcinoma. *Cancers (Basel)* 13 (17), 4483. doi:10.3390/cancers13174483

Dagogo-Jack, I., and Shaw, A. T. (2018). Tumour Heterogeneity and Resistance to Cancer Therapies. *Nat. Rev. Clin. Oncol.* 15 (2), 81–94. doi:10.1038/nrclinonc.2017.166

Der, S. D., Sykes, J., Pintilie, M., Zhu, C.-Q., Strumpf, D., Liu, N., et al. (2014). Validation of a Histology-independent Prognostic Gene Signature for Early-Stage, Non-small-cell Lung Cancer Including Stage IA Patients. *J. Thorac. Oncol.* 9 (1), 59–64. doi:10.1097/jto.0000000000000042

Dhatchinamoorthy, K., Colbert, J. D., and Rock, K. L. (2021). Cancer Immune Evasion through Loss of MHC Class I Antigen Presentation. *Front. Immunol.* 12, 636568. doi:10.3389/fimmu.2021.636568

Dürrbaum, M., and Storchová, Z. (2016). Effects of Aneuploidy on Gene Expression: Implications for Cancer. *FEBS J.* 283 (5), 791–802. doi:10.1111/febs.13591

Feng, Y., Li, F., Yan, J., Guo, X., Wang, F., Shi, H., et al. (2021). Pan-cancer Analysis and Experiments with Cell Lines Reveal that the Slightly Elevated Expression of DLGAP5 Is Involved in Clear Cell Renal Cell Carcinoma Progression. *Life Sci.* 287, 120056. doi:10.1016/j.lfs.2021.120056

Feng, Z., Zhang, J., Zheng, Y., Liu, J., Duan, T., and Tian, T. (2021). Overexpression of Abnormal Spindle-like Microcephaly-Associated (ASPM) Increases Tumor Aggressiveness and Predicts Poor Outcome in Patients with Lung Adenocarcinoma. *Transl. Cancer Res. TCR* 10 (2), 983–997. doi:10.21037/tcr-20-2570

Finetti, P., Guille, A., Adelaide, J., Birnbaum, D., Chaffanet, M., and Bertucci, F. (2014). ESPL1 Is a Candidate Oncogene of Luminal B Breast Cancers. *Breast Cancer Res. Treat.* 147 (1), 51–59. doi:10.1007/s10549-014-3070-z

Fu, J., Li, K., Zhang, W., Wan, C., Zhang, J., Jiang, P., et al. (2020). Large-scale Public Data Reuse to Model Immunotherapy Response and Resistance. *Genome Med.* 12 (1), 21. doi:10.1186/s13073-020-0721-z

Gan, H., Lin, L., Hu, N., Yang, Y., Gao, Y., Pei, Y., et al. (2019). KIF2C Exerts an Oncogenic Role in Nonsmall Cell Lung Cancer and Is Negatively Regulated by miR-325-3p. *Cell Biochem. Funct.* 37 (6), 424–431. doi:10.1002/cbf.3420

Gao, Y., Kim, S., Lee, Y. I., and Lee, J. (2019). Cellular Stress-Modulating Drugs Can Potentially Be Identified by In Silico Screening with Connectivity Map (CMap). *Int. J. Mol. Sci.* 20 (22), 5601. doi:10.3390/ijms20225601

Gasca, J., Flores, M. L., Giráldez, S., Ruiz-Borrego, M., Tortolero, M., Romero, F., et al. (2016). Loss of FBXW7 and Accumulation of MCL1 and PLK1 Promote Paclitaxel Resistance in Breast Cancer. *Oncotarget* 7 (33), 52751–52765. doi:10.18632/oncotarget.10481

Gheghiani, L., Wang, L., Zhang, Y., Moore, X. T. R., Zhang, J., Smith, S. C., et al. (2021). PLK1 Induces Chromosomal Instability and Overrides Cell-Cycle Checkpoints to Drive Tumorigenesis. *Cancer Res.* 81 (5), 1293–1307. doi:10.1158/0008-5472.can-20-1377

Gisselsson, D. (2008). Classification of Chromosome Segregation Errors in Cancer. *Chromosoma* 117 (6), 511–519. doi:10.1007/s00412-008-0169-1

Gutteridge, R. E. A., Ndiaye, M. A., Liu, X., and Ahmad, N. (2016). Plk1 Inhibitors in Cancer Therapy: From Laboratory to Clinics. *Mol. Cancer Ther.* 15 (7), 1427–1435. doi:10.1158/1535-7163.mct-15-0897

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. *Cell* 144 (5), 646–674. doi:10.1016/j.cell.2011.02.013

He, L., Chen, J., Xu, F., Li, J., and Li, J. (2020). Prognostic Implication of a Metabolism-Associated Gene Signature in Lung Adenocarcinoma. *Mol. Ther. - Oncolytics* 19, 265–277. doi:10.1016/j.omto.2020.09.011

Herbst, R. S., Morgensztern, D., and Boshoff, C. (2018). The Biology and Management of Non-small Cell Lung Cancer. *Nature* 553 (7689), 446–454. doi:10.1038/nature25183

Hu, C., Wu, J., Wang, L., Liu, X., Da, B., Liu, Y., et al. (2021). miR-133b Inhibits Cell Proliferation, Migration, and Invasion of Lung Adenocarcinoma by Targeting CDCA8. *Pathology - Res. Pract.* 223, 153459. doi:10.1016/j.prp.2021.153459

Huang, H., Yang, Y., Zhang, W., Liu, X., and Yang, G. (2020). TTK Regulates Proliferation and Apoptosis of Gastric Cancer Cells through the Akt-mTOR Pathway. *FEBS Open Bio* 10 (8), 1542–1549. doi:10.1002/2211-5463.12909

Huang, J., Li, Y., Lu, Z., Che, Y., Sun, S., Mao, S., et al. (2019). Analysis of Functional Hub Genes Identifies CDC45 as an Oncogene in Non-small Cell Lung Cancer - a Short Report. *Cell Oncol.* 42 (4), 571–578. doi:10.1007/s13402-019-00438-y

Huang, Y., Zhong, L., Nie, K., Li, L., Song, S., Liu, F., et al. (2021). Identification of LINC00665-miR-Let-7b-CCNA2 Competing Endogenous RNA Network Associated with Prognosis of Lung Adenocarcinoma. *Sci. Rep.* 11 (1), 4434. doi:10.1038/s41598-020-80662-x

Hutchinson, B. D., Shroff, G. S., Truong, M. T., and Ko, J. P. (2019). Spectrum of Lung Adenocarcinoma. *Seminars Ultrasound, CT MRI* 40 (3), 255–264. doi:10.1053/j.sult.2018.11.009

Iwai, Y., Ishida, M., Tanaka, Y., Okazaki, T., Honjo, T., and Minato, N. (2002). Involvement of PD-L1 on Tumor Cells in the Escape from Host Immune System and Tumor Immunotherapy by PD-L1 Blockade. *Proc. Natl. Acad. Sci. U.S.A.* 99 (19), 12293–12297. doi:10.1073/pnas.192461099

Janssen, A., van der Burg, M., Szuhai, K., Kops, G. J. P. L., and Medema, R. H. (2011). Chromosome Segregation Errors as a Cause of DNA Damage and Structural Chromosome Aberrations. *Science* 333 (6051), 1895–1898. doi:10.1126/science.1210214

Jiang, J., Liu, T., He, X., Ma, W., Wang, J., Zhou, Q., et al. (2021). Silencing of KIF18B Restricts Proliferation and Invasion and Enhances the Chemosensitivity of Breast Cancer via Modulating Akt/GSK -3β/β-catenin Pathway. *Biofactors* 47 (5), 754–767. doi:10.1002/biof.1757

Jiang, X., Jiang, Y., Luo, S., Sekar, K., Koh, C. K. T., Deivasigamani, A., et al. (2021). Correlation of NUF2 Overexpression with Poorer Patient Survival in Multiple Cancers. *Cancer Res. Treat.* 53 (4), 944–961. doi:10.4143/crt.2020.466

Jin, C.-Y., Du, L., Nuerlan, A.-H., Wang, X.-L., Yang, Y.-W., and Guo, R. (2020). High Expression of RRM2 as an Independent Predictive Factor of Poor Prognosis in Patients with Lung Adenocarcinoma. *Aging* 13 (3), 3518–3535. doi:10.18632/aging.202292

Kawakami, M., Liu, X., and Dmitrovsky, E. (2019). New Cell Cycle Inhibitors Target Aneuploidy in Cancer Therapy. *Annu. Rev. Pharmacol. Toxicol.* 59, 361–377. doi:10.1146/annurev-pharmtox-010818-021649

Kimura, S. (2010). AT-9283, a Small-Molecule Multi-Targeted Kinase Inhibitor for the Potential Treatment of Cancer. *Curr. Opin. Investig. Drugs* 11 (12), 1442–1449. doi:10.2174/157015910793358141

Kou, F., Sun, H., Wu, L., Li, B., Zhang, B., Wang, X., et al. (2020). TOP2A Promotes Lung Adenocarcinoma Cells' Malignant Progression and Predicts Poor Prognosis in Lung Adenocarcinoma. *J. Cancer* 11 (9), 2496–2508. doi:10.7150/jca.41415

Kou, F., Wu, L., Ren, X., and Yang, L. (2020). Chromosome Abnormalities: New Insights into Their Clinical Significance in Cancer. *Mol. Ther. - Oncolytics* 17, 562–570. doi:10.1016/j.omto.2020.05.010

Lai, W., Zhu, W., Xiao, C., Li, X., Wang, Y., Han, Y., et al. (2021). HJURP Promotes Proliferation in Prostate Cancer Cells through Increasing CDKN1A Degradation via the GSK3β/JNK Signaling Pathway. *Cell Death Dis.* 12 (6), 583. doi:10.1038/s41419-021-03870-x

Latchman, Y., Wood, C. R., Chernova, T., Chaudhary, D., Borde, M., Chernova, I., et al. (2001). PD-L2 Is a Second Ligand for PD-1 and Inhibits T Cell Activation. *Nat. Immunol.* 2 (3), 261–268. doi:10.1038/85330

Lei, Y., Zhang, G., Zhang, C., Xue, L., Yang, Z., Lu, Z., et al. (2021). The Average Copy Number Variation (CNVA) of Chromosome Fragments Is a Potential Surrogate for Tumor Mutational Burden in Predicting Responses to Immunotherapy in Non-small-cell Lung Cancer. *Clin. Transl. Immunol.* 10 (1), e1231. doi:10.1002/cti2.1231

Levine, M. S., and Holland, A. J. (2018). The Impact of Mitotic Errors on Cell Proliferation and Tumorigenesis. *Genes. Dev.* 32 (9-10), 620–638. doi:10.1101/gad.314351.118

Li, B., Liu, B., Zhang, X., Liu, H., and He, L. (2020). KIF18B Promotes the Proliferation of Pancreatic Ductal Adenocarcinoma via Activating the Expression of CDCA8. *J. Cell Physiol.* 235 (5), 4227–4238. doi:10.1002/jcp.29201

Li, J., Wang, R., Schweickert, P. G., Karki, A., Yang, Y., Kong, Y., et al. (2016). Plk1 Inhibition Enhances the Efficacy of Gemcitabine in Human Pancreatic Cancer. *Cell Cycle* 15 (5), 711–719. doi:10.1080/15384101.2016.1148838

Li, M.-X., Zhang, M.-Y., Dong, H.-H., Li, A.-J., Teng, H.-F., Liu, A.-L., et al. (2021). Overexpression of CENPF Is Associated with Progression and Poor Prognosis of Lung Adenocarcinoma. *Int. J. Med. Sci.* 18 (2), 494–504. doi:10.7150/ijms.49041

Li, Y., Yi, Q., Liao, X., Han, C., Zheng, L., Li, H., et al. (2021). Hypomethylation-driven Overexpression of HJURP Promotes Progression of Hepatocellular Carcinoma and Is Associated with Poor Prognosis. *Biochem. Biophysical Res. Commun.* 566, 67–74. doi:10.1016/j.bbrc.2021.05.102

Li, Z., Wang, H., Wei, J., Han, L., and Guo, Z. (2020). Indirubin Exerts Anticancer Effects on Human Glioma Cells by Inducing Apoptosis and Autophagy. *Amb. Expr.* 10 (1), 171. doi:10.1186/s13568-020-01107-2

Liao, W., Liu, W., Yuan, Q., Liu, X., Ou, Y., He, S., et al. (2013). Silencing of DLGAP5 by siRNA Significantly Inhibits the Proliferation and Invasion of Hepatocellular Carcinoma Cells. *PLoS One* 8 (12), e80789. doi:10.1371/journal.pone.0080789

Lim, Z.-F., and Ma, P. C. (2019). Emerging Insights of Tumor Heterogeneity and Drug Resistance Mechanisms in Lung Cancer Targeted Therapy. *J. Hematol. Oncol.* 12 (1), 134. doi:10.1186/s13045-019-0818-2

Lin, W., Chen, Y., Wu, B., chen, Y., and Li, Z. (2021). Identification of the Pyroptosis-related P-rognostic G-ene S-ignature and the A-ssociated R-egulation axis in L-ung A-denocarcinoma. *Cell Death Discov.* 7 (1), 161. doi:10.1038/s41420-021-00557-2

Liu, W., Yu, Z., Tang, H., Wang, X., Zhang, B., Zhao, J., et al. (2020). Silencing KIF18B Enhances Radiosensitivity: Identification of a Promising Therapeutic Target in Sarcoma. *EBioMedicine* 61, 103056. doi:10.1016/j.ebiom.2020.103056

Liu, Y., Zhu, K., Guan, X., Xie, S., Wang, Y., Tong, Y., et al. (2021). TTK Is a Potential Therapeutic Target for Cisplatin-Resistant Ovarian Cancer. *J. Ovarian Res.* 14 (1), 128. doi:10.1186/s13048-021-00884-z

Liu, Z., Lian, X., Zhang, X., Zhu, Y., Zhang, W., Wang, J., et al. (2021). ESPL1 Is a Novel Prognostic Biomarker Associated with the Malignant Features of Glioma. *Front. Genet.* 12, 666106. doi:10.3389/fgene.2021.666106

Liu, Z., Sun, Q., and Wang, X. (2017). PLK1, A Potential Target for Cancer Therapy. *Transl. Oncol.* 10 (1), 22–32. doi:10.1016/j.tranon.2016.10.003

Loeper, S., Romeike, B. F., Heckmann, N., Jung, V., Henn, W., Feiden, W., et al. (2001). Frequent Mitotic Errors in Tumor Cells of Genetically Micro-heterogeneous Glioblastomas. *Cytogenet Cell Genet.* 94 (1-2), 1–8. doi:10.1159/000048773

Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8

M'Kacher, R., Andreoletti, L., Flamant, S., Milliat, F., Girinsky, T., Dossou, J., et al. (2010). JC Human Polyomavirus Is Associated to Chromosomal Instability in Peripheral Blood Lymphocytes of Hodgkin's Lymphoma Patients and Poor Clinical Outcome. *Ann. Oncol.* 21 (4), 826–832. doi:10.1093/annonc/mdp375

Mao, Y., Xi, L., Li, Q., Wang, S., Cai, Z., Zhang, X., et al. (2018). Combination of PI3K/Akt Pathway Inhibition and Plk1 Depletion Can Enhance Chemosensitivity to Gemcitabine in Pancreatic Carcinoma. *Transl. Oncol.* 11 (4), 852–863. doi:10.1016/j.tranon.2018.04.011

Merino, D. M., McShane, L. M., Fabrizio, D., Funari, V., Chen, S. J., White, J. R., et al. (2020). Establishing Guidelines to Harmonize Tumor Mutational Burden (TMB): In Silico Assessment of Variation in TMB Quantification across Diagnostic Platforms: Phase I of the Friends of Cancer Research TMB Harmonization Project. *J. Immunother. Cancer* 8 (1), e000147. doi:10.1136/jitc-2019-000147

Mittal, B., Tulsyan, S., Kumar, S., Mittal, R. D., and Agarwal, G. (2015). Cytochrome P450 in Cancer Susceptibility and Treatment. *Adv. Clin. Chem.* 71, 77–139. doi:10.1016/bs.acc.2015.06.003

Montaudon, E., Nikitorowicz-Buniak, J., Sourd, L., Morisset, L., El Botty, R., Huguet, L., et al. (2020). PLK1 Inhibition Exhibits Strong Anti-tumoral Activity in CCND1-Driven Breast Cancer Metastases with Acquired Palbociclib Resistance. *Nat. Commun.* 11 (1), 4053. doi:10.1038/s41467-020-17697-1

Moreira, A. L., and Eng, J. (2014). Personalized Therapy for Lung Cancer. *Chest* 146 (6), 1649–1657. doi:10.1378/chest.14-0713

Naylor, R. M., and van Deursen, J. M. (2016). Aneuploidy in Cancer and Aging. *Annu. Rev. Genet.* 50, 45–66. doi:10.1146/annurev-genet-120215-035303

Neumüller, R. A., and Knoblich, J. A. (2009). Dividing Cellular Asymmetry: Asymmetric Cell Division and its Implications for Stem Cells and Cancer. *Genes. Dev.* 23 (23), 2675–2699. doi:10.1101/gad.1850809

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

Ninomiya, H., Nomura, K., Satoh, Y., Okumura, S., Nakagawa, K., Fujiwara, M., et al. (2006). Genetic Instability in Lung Cancer: Concurrent Analysis of Chromosomal, Mini- and Microsatellite Instability and Loss of Heterozygosity. *Br. J. Cancer* 94 (10), 1485–1491. doi:10.1038/sj.bjc.6603121

Nooreldeen, R., and Bach, H. (2021). Current and Future Development in Lung Cancer Diagnosis. *Int. J. Mol. Sci.* 22 (16), 8661. doi:10.3390/ijms22168661

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 27 (1), 29–34. doi:10.1093/nar/27.1.29

Peng, B., Shi, R., Bian, J., Li, Y., Wang, P., Wang, H., et al. (2021). PARP1 and CHK1 Coordinate PLK1 Enzymatic Activity during the DNA Damage Response to Promote Homologous Recombination-Mediated Repair. *Nucleic Acids Res.* 49 (13), 7554–7570. doi:10.1093/nar/gkab584

Philpott, C., Tovell, H., Frayling, I. M., Cooper, D. N., and Upadhyaya, M. (2017). The NF1 Somatic Mutational Landscape in Sporadic Human Cancers. *Hum. Genomics* 11 (1), 13. doi:10.1186/s40246-017-0109-3

Potapova, T., and Gorbsky, G. J. (2017). The Consequences of Chromosome Segregation Errors in Mitosis and Meiosis. *Biol. (Basel)* 6 (1), 12. doi:10.3390/biology6010012

Puhr, H. C., and Ilhan-Mutlu, A. (2019). New Emerging Targets in Cancer Immunotherapy: The Role of LAG3. *ESMO Open* 4 (2), e000482. doi:10.1136/esmoopen-2018-000482

Qi, G., Ma, H., Li, Y., Peng, J., Chen, J., and Kong, B. (2021). TTK Inhibition Increases Cisplatin Sensitivity in High-Grade Serous Ovarian Carcinoma through the mTOR/autophagy Pathway. *Cell Death Dis.* 12 (12), 1135. doi:10.1038/s41419-021-04429-6

Ramani, P., Nash, R., Sowa-Avugrah, E., and Rogers, C. (2015). High Levels of Polo-like Kinase 1 and Phosphorylated Translationally Controlled Tumor

Protein Indicate Poor Prognosis in Neuroblastomas. *J. Neurooncol* 125 (1), 103–111. doi:10.1007/s11060-015-1900-4

Reisländer, T., Groelly, F. J., and Tarsounas, M. (2020). DNA Damage and Cancer Immunotherapy: A STING in the Tale. *Mol. Cell* 80 (1), 21–28. doi:10.1016/j.molcel.2020.07.026

Rodrigues-Junior, D. M., Biassi, T. P., Carlin, V., Buri, M. V., Torrecilhas, A. C., Bortoluci, K. R., et al. (2018). OIP5 Expression Sensitize Glioblastoma Cells to Lomustine Treatment. *J. Mol. Neurosci.* 66 (3), 383–389. doi:10.1007/s12031-018-1184-1

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

Sarkar, S., Sahoo, P. K., Mahata, S., Pal, R., Ghosh, D., Mistry, T., et al. (2021). Mitotic checkpoint defects: En route to cancer and drug resistance. *Chromosome Res.* 29 (2), 131–144. doi:10.1007/s10577-020-09646-x

Schneider, M. A., Christopoulos, P., Muley, T., Warth, A., Klingmueller, U., Thomas, M., et al. (2017). AURKA, DLGAP5, TPX2, KIF11 and CKAP5: Five Specific Mitosis-Associated Genes Correlate with Poor Prognosis for Non-small Cell Lung Cancer Patients. *Int. J. Oncol.* 50 (2), 365–372. doi:10.3892/ijo.2017.3834

Serafim, R. B., Cardoso, C., Di Cristofaro, L. F. M., Soares, C. P., Silva, W. A., Espreafico, E. M., et al. (2020). HJURP Knockdown Disrupts Clonogenic Capacity and Increases Radiation-Induced Cell Death of Glioblastoma Cells. *Cancer Gene Ther.* 27 (5), 319–329. doi:10.1038/s41417-019-0103-0

Shan, L., Zhu, X.-L., Zhang, Y., Gu, G.-J., and Cheng, X. (2021). Expression and Clinical Significance of NUF2 in Kidney Renal Clear Cell Carcinoma. *Transl. Androl. Urol.* 10 (9), 3628–3637. doi:10.21037/tau-21-620

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13 (11), 2498–2504. doi:10.1101/gr.1239303

Shao, C., Ahmad, N., Hodges, K., Kuang, S., Ratliff, T., and Liu, X. (2015). Inhibition of Polo-like Kinase 1 (Plk1) Enhances the Antineoplastic Activity of Metformin in Prostate Cancer. *J. Biol. Chem.* 290 (4), 2024–2033. doi:10.1074/jbc.m114.596817

Shapouri-Moghaddam, A., Mohammadian, S., Vazini, H., Taghadosi, M., Esmaeili, S. A., Mardani, F., et al. (2018). Macrophage Plasticity, Polarization, and Function in Health and Disease. *J. Cell Physiol.* 233 (9), 6425–6440. doi:10.1002/jcp.26429

Shin, S. B., Kim, C. H., Jang, H. R., and Yim, H. (2020). Combination of Inhibitors of USP7 and PLK1 Has a Strong Synergism against Paclitaxel Resistance. *Int. J. Mol. Sci.* 21 (22), 8629. doi:10.3390/ijms21228629

Shin, S. B., Woo, S. U., and Yim, H. (2019). Cotargeting Plk1 and Androgen Receptor Enhances the Therapeutic Sensitivity of Paclitaxel-Resistant Prostate Cancer. *Ther. Adv. Med. Oncol.* 11, 1758835919846375. doi:10.1177/1758835919846375

Singel, S. M., Cornelius, C., Zaganjor, E., Batten, K., Sarode, V. R., Buckley, D. L., et al. (2014). KIF14 Promotes AKT Phosphorylation and Contributes to Chemoresistance in Triple-Negative Breast Cancer. *Neoplasia* 16 (3), 247–256. doi:10.1016/j.neo.2014.03.008

Solinas, C., Migliori, E., De Silva, P., and Willard-Gallo, K. (2019). LAG3: The Biological Processes that Motivate Targeting This Immune Checkpoint Molecule in Human Cancer. *Cancers (Basel)* 11 (8), 1213. doi:10.3390/cancers11081213

Song, B., Liu, X. S., Rice, S. J., Kuang, S., Elzey, B. D., Konieczny, S. F., et al. (2013). Plk1 Phosphorylation of Orc2 and Hbo1 Contributes to Gemcitabine Resistance in Pancreatic Cancer. *Mol. Cancer Ther.* 12 (1), 58–68. doi:10.1158/1535-7163.mct-12-0632

Song, J., Sun, Y., Cao, H., Liu, Z., Xi, L., Dong, C., et al. (2021). A Novel Pyroptosis-Related lncRNA Signature for Prognostic Prediction in Patients with Lung Adenocarcinoma. *Bioengineered* 12 (1), 5932–5949. doi:10.1080/21655979.2021.1972078

Soto, M., Raaijmakers, J. A., and Medema, R. H. (2019). Consequences of Genomic Diversification Induced by Segregation Errors. *Trends Genet.* 35 (4), 279–291. doi:10.1016/j.tig.2019.01.003

Sturm, G., Finotello, F., and List, M. (2020). Immunedeconv: An R Package for Unified Access to Computational Methods for Estimating Immune Cell Fractions from Bulk RNA-Sequencing Data. *Methods Mol. Biol.* 2120, 223–232. doi:10.1007/978-1-0716-0327-7_16

Sun, G. Z., and Zhao, T. W. (2019). Lung Adenocarcinoma Pathology Stages Related Gene Identification. *Math. Biosci. Eng.* 17 (1), 737–746. doi:10.3934/mbe.2020038

Sun, S., Guo, W., Wang, Z., Wang, X., Zhang, G., Zhang, H., et al. (2020). Development and Validation of an Immune-related Prognostic Signature in Lung Adenocarcinoma. *Cancer Med.* 9 (16), 5960–5975. doi:10.1002/cam4.3240

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

Tanaka, K., and Hirota, T. (2009). Chromosome Segregation Machinery and Cancer. *Cancer Sci.* 100 (7), 1158–1165. doi:10.1111/j.1349-7006.2009.01178.x

Tayoun, T., Oulhen, M., Aberlenc, A., Farace, F., and Pawlikowska, P. (2021). Tumor Evolution and Therapeutic Choice Seen through a Prism of Circulating Tumor Cell Genomic Instability. *Cells* 10 (2), 337. doi:10.3390/cells10020337

Testa, D., Jourde-Chiche, N., Mancini, J., Varriale, P., Radoszycki, L., and Chiche, L. (2021). Unsupervised Clustering Analysis of Data from an Online Community to Identify Lupus Patient Profiles with Regards to Treatment Preferences. *Lupus* 30 (11), 1837–1843. doi:10.1177/09612033211033977

Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015). The Cancer Genome Atlas (TCGA): An Immeasurable Source of Knowledge. *Contemp. Oncol. Pozn.* 19 (1A), A68–A77. doi:10.5114/wo.2014.47136

Tomida, S., Takeuchi, T., Shimada, Y., Arima, C., Matsuo, K., Mitsudomi, T., et al. (2009). Relapse-related Molecular Signature in Lung Adenocarcinomas Identifies Patients with Dismal Prognosis. *J. Clin. Oncol.* 27 (17), 2793–2799. doi:10.1200/jco.2008.19.7053

Torre, L. A., Siegel, R. L., and Jemal, A. (2016). Lung Cancer Statistics. *Adv. Exp. Med. Biol.* 893, 1–19. doi:10.1007/978-3-319-24223-1_1

Travis, W. D., Brambilla, E., Noguchi, M., Nicholson, A. G., Geisinger, K., Yatabe, Y., et al. (2011). International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society: International Multidisciplinary Classification of Lung Adenocarcinoma: Executive Summary. *Proc. Am. Thorac. Soc.* 8 (5), 381–385. doi:10.1513/pats.201107-042st

Tut, T. G., Lim, S. H. S., Dissanayake, I. U., Descallar, J., Chua, W., Ng, W., et al. (2015). Upregulated Polo-Like Kinase 1 Expression Correlates with Inferior Survival Outcomes in Rectal Cancer. *PLoS One* 10 (6), e0129313. doi:10.1371/journal.pone.0129313

Wang, L., Yang, X., An, N., and Liu, J. (2020). Bioinformatics Analysis of BUB1 Expression and Gene Regulation Network in Lung Adenocarcinoma. *Transl. Cancer Res. TCR* 9 (8), 4820–4833. doi:10.21037/tcr-20-1045

Wang, M., Du, Q., Jin, J., Wei, Y., Lu, Y., and Li, Q. (2021). LAG3 and its Emerging Role in Cancer Immunotherapy. *Clin. Transl. Med.* 11 (3), e365. doi:10.1002/ctm2.365

Wang, P., Yu, N., Wang, Y., Sun, H., Yang, Z., and Zhou, S. (2018). Co-delivery of PLK1-specific shRNA and Doxorubicin via Core-Crosslinked pH-Sensitive and Redox Ultra-sensitive Micelles for Glioma Therapy. *J. Mat. Chem. B* 6 (1), 112–124. doi:10.1039/c7tb02160g

Wang, R., Zang, W., Hu, B., Deng, D., Ling, X., Zhou, H., et al. (2020). Serum ESPL1 Can Be Used as a Biomarker for Patients with Hepatitis B Virus-Related Liver Cancer: A Chinese Case-Control Study. *Technol. Cancer Res. Treat.* 19, 1533033820980785. doi:10.1177/1533033820980785

Wang, S., Sun, H., Zhan, X., and Wang, Q. (2021). [Retracted] MicroRNA718 Serves a Tumorsuppressive Role in Nonsmall Cell Lung Cancer by Directly Targeting CCNB1. *Int. J. Mol. Med.* 48 (3), 180. doi:10.3892/ijmm.2021.5013

Wang, X., Xiao, H., Wu, D., Zhang, D., and Zhang, Z. (2020). miR-335-5p Regulates Cell Cycle and Metastasis in Lung Adenocarcinoma by Targeting CCNB2. *Onco. Targets Ther.* 13, 6255–6263. doi:10.2147/ott.s245136

Wang, Z. Z., Yang, J., Jiang, B. H., Di, J. B., Gao, P., Peng, L., et al. (2018). KIF14 Promotes Cell Proliferation via Activation of Akt and Is Directly Targeted by miR-200c in Colorectal Cancer. *Int. J. Oncol.* 53 (5), 1939–1952. doi:10.3892/ijo.2018.4546

Wei, Y., Ouyang, G. L., Yao, W. X., Zhu, Y. J., Li, X., Huang, L. X., et al. (2019). Knockdown of HJURP Inhibits Non-small Cell Lung Cancer Cell

Proliferation, Migration, and Invasion by Repressing Wnt/β-Catenin Signaling. *Eur. Rev. Med. Pharmacol. Sci.* 23 (9), 3847–3856. doi:10.26355/eurrev_201905_17812

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: A Class Discovery Tool with Confidence Assessments and Item Tracking. *Bioinformatics* 26 (12), 1572–1573. doi:10.1093/bioinformatics/btq170

Wu, Y., Lin, Y., Pan, J., Tu, X., Xu, Y., Li, H., et al. (2021). NCAPG Promotes the Progression of Lung Adenocarcinoma via the TGF-β Signaling Pathway. *Cancer Cell Int.* 21 (1), 443. doi:10.1186/s12935-021-02138-w

Xia, Y., Rao, L., Yao, H., Wang, Z., Ning, P., and Chen, X. (2020). Engineering Macrophages for Cancer Immunotherapy and Drug Delivery. *Adv. Mater* 32 (40), e2002054. doi:10.1002/adma.202002054

Xiao, L., Zhang, S., Zheng, Q., and Zhang, S. (2021). Dysregulation of KIF14 Regulates the Cell Cycle and Predicts Poor Prognosis in Cervical Cancer: A Study Based on Integrated Approaches. *Braz J. Med. Biol. Res.* 54 (11), e11363. doi:10.1590/1414-431X2021e11363

Xie, X., Jiang, S., and Li, X. (2021). Nuf2 Is a Prognostic-Related Biomarker and Correlated with Immune Infiltrates in Hepatocellular Carcinoma. *Front. Oncol.* 11, 621373. doi:10.3389/fonc.2021.621373

Xu, R., Zhang, Y., Li, A., Ma, Y., Cai, W., Song, L., et al. (2021). LY-294002 Enhances the Chemosensitivity of Liver Cancer to Oxaliplatin by Blocking the PI3K/AKT/HIF-1α Pathway. *Mol. Med. Rep.* 24 (1), 508. doi:10.3892/mmr.2021.12147

Xu, T., Dong, M., Li, H., Zhang, R., and Li, X. (2020). Elevated mRNA Expression Levels of DLGAP5 Are Associated with Poor Prognosis in Breast Cancer. *Oncol. Lett.* 19 (6), 4053–4065. doi:10.3892/ol.2020.11533

Yamauchi, M., Yamaguchi, R., Nakata, A., Kohno, T., Nagasaki, M., Shimamura, T., et al. (2012). Epidermal Growth Factor Receptor Tyrosine Kinase Defines Critical Prognostic Genes of Stage I Lung Adenocarcinoma. *PLoS One* 7 (9), e43923. doi:10.1371/journal.pone.0043923

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

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

Zhang, H., Liu, Y., Tang, S., Qin, X., Li, L., Zhou, J., et al. (2021). Knockdown of DLGAP5 Suppresses Cell Proliferation, Induces G2/M Phase Arrest and Apoptosis in Ovarian Cancer. *Exp. Ther. Med.* 22 (5), 1245. doi:10.3892/etm.2021.10680

Zhang, H., Yao, W., Zhang, M., Lu, Y., Tang, J., Jiang, M., et al. (2021). TTK Inhibitor Promotes Radiosensitivity of Liver Cancer Cells through P21. *Biochem. Biophysical Res. Commun.* 550, 84–91. doi:10.1016/j.bbrc.2021.01.089

Zhang, R., Shi, H., Ren, F., Liu, H., Zhang, M., Deng, Y., et al. (2015). Misregulation of Polo-like Protein Kinase 1, P53 and P21WAF1 in Epithelial Ovarian Cancer Suggests Poor Prognosis. *Oncol. Rep.* 33 (3), 1235–1242. doi:10.3892/or.2015.3723

Zhang, Z., Hou, S.-Q., He, J., Gu, T., Yin, Y., and Shen, W. H. (2016). PTEN Regulates PLK1 and Controls Chromosomal Stability during Cell Division. *Cell Cycle* 15 (18), 2476–2485. doi:10.1080/15384101.2016.1203493

Zheng, R., Shi, Z., Li, W., Yu, J., Wang, Y., and Zhou, Q. (2020). Identification and Prognostic Value of DLGAP5 in Endometrial Cancer. *PeerJ* 8, e10433. doi:10.7717/peerj.10433

Zhou, D., Wang, M., Zhang, Y., Wang, K., Zhao, M., Wang, Y., et al. (2021). Screening and Identification of LMNB1 and DLGAP5, Two Key Biomarkers in Gliomas. *Biosci. Rep.* 41 (5), BSR20210231. doi:10.1042/bsr20210231

Zhou, J., Jiang, G., Xu, E., Zhou, J., Liu, L., and Yang, Q. (2021). Identification of SRXN1 and KRT6A as Key Genes in Smoking-Related Non-small-cell Lung Cancer through Bioinformatics and Functional Analyses. *Front. Oncol.* 11, 810301. doi:10.3389/fonc.2021.810301

Zhu, D., Xia, J., Liu, C., and Fang, C. (2022). Numb/Notch/PLK1 Signaling Pathway Mediated Hyperglycemic Memory in Pancreatic Cancer Cell Radioresistance and the Therapeutic Effects of Metformin. *Cell. Signal.* 93, 110268. doi:10.1016/j.cellsig.2022.110268

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.2022.905508

Characterization of Fatty Acid Metabolism in Lung Adenocarcinoma

Suyu Wang[1,2†], Aona Chen[3†], Wanli Zhu[4†], Di Feng[1], Juan Wei[1], Quanfu Li[1], Xuan Shi[1], Xin Lv[1] and Meiyun Liu[1]**

[1]Department of Anesthesiology, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai, China, [2]Department of Cardiothoracic Surgery, Changzheng Hospital, Naval Medical University, Shanghai, China, [3]Department of General Surgery, Changzheng Hospital, Naval Medical University, Shanghai, China, [4]Department of General Surgery, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai, China

Background: Lung adenocarcinoma (LUAD) is the most common subtype of non-small cell lung cancer. Fatty acid metabolism takes part in malignancy progression. However, the roles fatty acid metabolism plays in LUAD are still unclear.

Methods: The transcriptomic and clinical data of LUAD patients from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases were extracted. ssGSEA, WGCNA, univariable Cox regression, and LASSO Cox regression analyses were performed to identify the fatty acid metabolism-related genes which influenced the overall survival (OS) and build a fatty acid-related risk score (FARS) model. A nomogram was established based on the FARS and other clinicopathological features, and ROC and calibration plots were used to validate the prediction accuracy. The tumor microenvironment (TME) of patients with high and low FARS was compared.

Results: A total of 38 genes were identified to be independently related to the survival outcome and put into a FARS model. High FARS patients exhibited significantly worse OS. The nomogram included the FARS and pathological stage, and the AUC of the nomogram predicting 1-, 2-, 3-, 4-, and 5-year OS was 0.789, 0.807, 0.798, 0.809, and 0.753, respectively. Calibration plots also indicated good accuracy. Moreover, the samples of the high FARS had higher expression of PDL1.

Conclusion: We constructed a FARS model which could accurately predict the survival outcome of the LUAD patients. The genes of the FARS are related to the tumor microenvironment and patients with high FARS can potentially benefit more from anti-PD1/PDL1 immunotherapy. In addition, the mechanisms of the genes in the FARS affecting prognosis are worthy of further research to develop new gene-targeted drugs.

Keywords: fatty acid metabolism, lung adenocarcinoma, prognosis, fatty acid-related risk score, immunotherapy

OPEN ACCESS

Edited by:
Valerio Costa, National Research Council (IGB-CNR), Italy

Reviewed by:
Wei Li, Lanzhou University Second Hospital, China
Umberto Malapelle, University of Naples Federico II, Italy

***Correspondence:**
Xin Lv
xinlvg@126.com
Meiyun Liu
liumeiyun0@126.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *27 March 2022*
Accepted: *14 June 2022*
Published: *14 July 2022*

Citation:
Wang S, Chen A, Zhu W, Feng D, Wei J, Li Q, Shi X, Lv X and Liu M (2022) Characterization of Fatty Acid Metabolism in Lung Adenocarcinoma. Front. Genet. 13:905508. doi: 10.3389/fgene.2022.905508

Abbreviations: AJCC TNM, American Joint Committee on Cancer Tumor Node Metastasis; AUC, area under the curve; FARS, fatty acid-related risk score; CTLA4, cytotoxic T-lymphocyte-associated protein 4; GEO, Gene Expression Omnibus; GOBP, gene ontology biological process; LASSO, least absolute shrinkage and selection operator; LUAD, lung adenocarcinoma; NSCLC, non-small cell lung cancer; OS, overall survival; PD1, programmed cell death protein 1; PDL1, programmed death-ligand 1; pStage, pathological stage; ssGSEA, single-sample Gene Set Enrichment Analysis; TCGA, The Cancer Genome Atlas; TMEs, tumor microenvironments; tROC, time-dependent receiver operating characteristic curves; WGCNA, Weighted Gene Co-expression Network Analysis.

INTRODUCTION

Non-small cell lung cancer (NSCLC), the leading reason for cancer-related deaths, constitutes approximately 85% of malignant lung tumors (Bray et al., 2018). Lung adenocarcinoma (LUAD) is the most common subtype which accounts for nearly half of NSCLC (Chen et al., 2014). Moreover, LUAD is well-known for its heterogeneity in clinical, behavioral, cellular, and molecular features. Although most lung cancers are characterized by aggressive nature, almost 18.5% of lung cancers found by computed tomography (CT) screening are dormant and can lead patients to the hazard of overdiagnosis and overtreatment (Patz et al., 2014). Although significant effort has been made, the underlying cellular and molecular mechanisms of tumor behavior are still unclear, and long-term survival rates of lung cancer patients have been scarcely improved compared with other cancers (Siegel et al., 2016). Thus, it is important to detect new potential molecular signatures and therapeutic targets for LUAD.

Gene markers, particularly in tumor tissues, are dependable factors for predicting the long-term survival of cancer patients (Li et al., 2017; Hu et al., 2021). Therefore, detecting the molecular characterization which may cause poor outcomes can guide clinical adjuvant therapeutic strategies for a subgroup of patients who are at high risk. Moreover, this can be helpful in identifying new molecular targets for developing new medicines. The public database of gene expression of large cohorts of patients facilitates the aim of establishing a metabolic gene signature for predicting survival outcomes and analyzing the tumor microenvironments (TMEs).

Attributed to the fast proliferation of cancer cells and insufficient angiogenesis, the main characteristics of TMEs are malnutrition, hypoxia, high oxidation, and acidity. Hence, compared to normal cells, tumor cells manifest distinct metabolic features to cope with diverse deleterious microenvironments *via* metabolic recoding processes which maintain the proliferation and survival of tumor cells when the oncogenic signal is blocked (Lue et al., 2017). Reprogramming of energy metabolism, known as a hallmark of cancers, has been lately verified to take part in the initiation, progression, and drug resistance in lung cancer (Hensley et al., 2016; Chen et al., 2019). There is a distinct difference in carbohydrate, amino acid, and lipid metabolism between tumor cells and normal cells (Yu et al., 2019). Taking carbohydrate metabolism as an example, normal tissue cells decompose glucose into pyruvate by glycolysis, and in addition to glucose decomposition, oxidative phosphorylation in mitochondria generates vast energy. On the other hand, in cancer cells, the glucose is catabolized into lactate with an insufficient generation of energy; thus, cancer cells consume much more glucose than normal cells (Faubert et al., 2017). As demonstrated by Xue et al., cellular pyruvate metabolism of LUAD changed, including the reduction of expression of mitochondrial pyruvate carrier 1 (MPC1) compared with adjacent normal tissues. Xue et al. (2021) also revealed that higher MPC1 expression was related to a favorable prognosis. In addition to carbohydrate metabolism, lipid metabolism is also a potential hallmark for cancers. Lipogenesis, lipid uptake, and lipid storage are highly upregulated in malignant tumors to meet the augmented demands of membrane biogenesis and promote cancer cell proliferation and survival, especially under conditions of insufficient nutrition and oxygen (Menendez and Lupu, 2007; Nath et al., 2015; Qiu et al., 2015; Röhrig and Schulze, 2016; Geng and Guo, 2017). Presently, fatty acid metabolism, involved in many biological activities including signaling molecule synthesis, cell membrane formation, and energy storage in carcinogenesis, has been widely researched (Currie et al., 2013). For example, Ding et al. (2021) exhibited that the signature of fatty acid metabolism can predict the prognosis of colorectal cancer and was associated with chemoresistance and TME characteristics. In another study conducted by Svensson et al. (2016), ND-646, an allosteric inhibitor of the acetyl-CoA carboxylase (ACC) enzymes ACC1 and ACC2 which suppress ACC subunit dimerization, prevented the synthesis of fatty acid *in vitro* and *in vivo*. Thus, ND-646 significantly inhibited lung cancer growth in the KRAS p53 and KRAS Lkb1 mouse models of NSCLC, indicating the therapeutic potential of the ACC inhibitor in malignant tumors (Svensson et al., 2016). Based on previous studies, it is clear that an analysis of the metabolic pathway of lung cancer can help us comprehend the molecular mechanism of lung cancer and develop novel personalized therapeutic regimens (Sayin et al., 2019). However, the characterization of the genes related to fatty acid metabolism in LUAD has not been systematically investigated.

To detect the underlying genomic mechanism of fatty acid metabolism of LUAD, we used the genomic information on the clinicopathological features of 1,087 LUAD patients from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) database to reveal the pattern of fatty acid metabolism and establish a fatty acid-related risk score (FARS) model. The FARS model proved to be an independent prognostic factor in the survival outcome of LUAD patients. Moreover, the FARS can recognize patients who are suitable for anti-PD1/PDL1 antibody immunotherapy, indicating that fatty acid metabolism is highly related to individual characterizations of the TME. All these findings give a new insight into lipid metabolic mechanism and potential therapeutic targets.

MATERIAL AND METHODS

Data Acquisition and Processing

The clinicopathological features and fragments per kilobase per million mapped reads (FPKM) of 573 LUAD samples were extracted from UCSC Xena (http://xena.ucsc.edu/; accessed October 8 2021). The exclusion criteria were as follows: patients (a) with no survival information or survival time less than 30 days; (b) with no age, sex, or the American Joint Committee on Cancer Tumor Node Metastasis (AJCC TNM) stage information; and (c) has received neoadjuvant therapy. As a result, 468 LUAD samples with complete clinicopathological characteristics, including age, sex, AJCC TNM stage, and overall survival (OS) data, were put into the analysis as the training cohort.

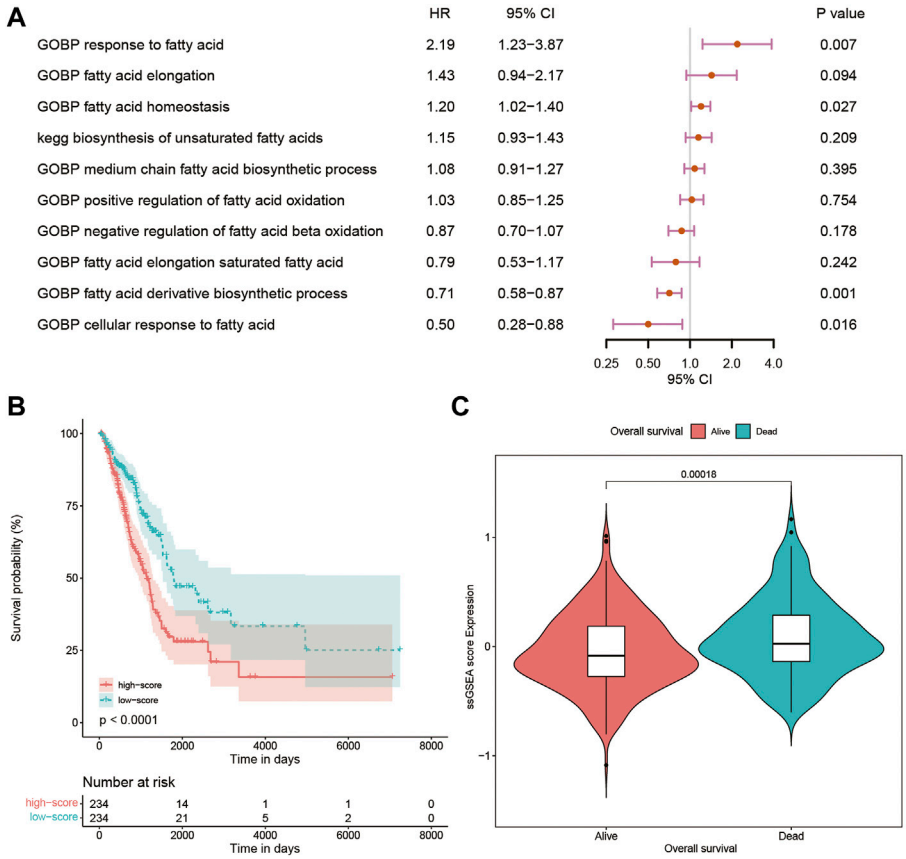


FIGURE 1 | (A) Multivariate Cox analysis of 10 pathways and biological processes; **(B)** Kaplan–Meier curves of patients with high and low ssGSEA scores; **(C)** violin plot comparing ssGSEA scores of survival and dead patients. GOBP, gene ontology biological process; ssGSEA, single-sample Gene Set Enrichment Analysis.

The microarray dataset GSE72094 was downloaded from GEO datasets (https://www.ncbi.nlm.nih.gov/gds/) and was used as the first validation cohort. This dataset was produced by using a Rosetta/Merck Human RSTA Custom Affymetrix 2.0 microarray and contained 393 samples of lung adenocarcinoma. In addition, 226 lung adenocarcinoma samples from GSE31210 [(HG-U133_Plus_2) Affymetrix Human Genome U133 Plus 2.0 Array] were used as the second independent validation cohort. The microarray and RNA-seq data in this study were normalized and $\log_2$ transformed.

Fatty Acid-Related Pathways and Biological Processes

A total of 68 fatty acid-related pathways or biological processes were extracted from hallmark gene sets (H collection), curated gene sets (C2 collection), and ontology gene sets (C5 collection) in Molecular Signatures Database (MSigDB, http://software.broadinstitute.org/gsea/index.jsp). All these pathways, biological processes, and their corresponding genes were fused into a file in CSV format (**Supplementary Material S1**). The related infiltration and activity levels of those pathways or biological processes in LUAD samples were quantified using the single-sample Gene Set Enrichment Analysis (ssGSEA) in the R package GSVA and normalized by using the Z-score method. The univariate and multivariate Cox analyses were used to assess the significance of these pathways or biological processes, and a two-side p-value <0.05 was considered to be statistically significant.

Candidate Gene Selection and Signature Construction in the Training Cohort

Using the R package WGCNA, weighted gene co-expression network analysis (WGCNA) was performed. The analysis was performed on the top 5,000 genes with the highest standard deviation by the standard protocol. A gene correlation matrix with an optimal soft thresholding of power was used to derive the adjacency matrix. Modules were obtained with the following criteria: module size ≥30 and height for merging modules ≥0.2. The relationship between the modules and fatty acid-related pathways or biological processes was calculated based on ssGSEA scores, and modules correlated with fatty acid-related pathways or biological processes were extracted based on p-value < 0.05 and $|r| > 0.3$.

Thereafter, the genes related to prognosis in those modules were selected by the univariate Cox analysis. Then, the least absolute shrinkage and selection operator (LASSO) regression model was applied to further detect the most robust prognostic markers. A fatty acid-related risk score (FARS) of each sample was established using the formula

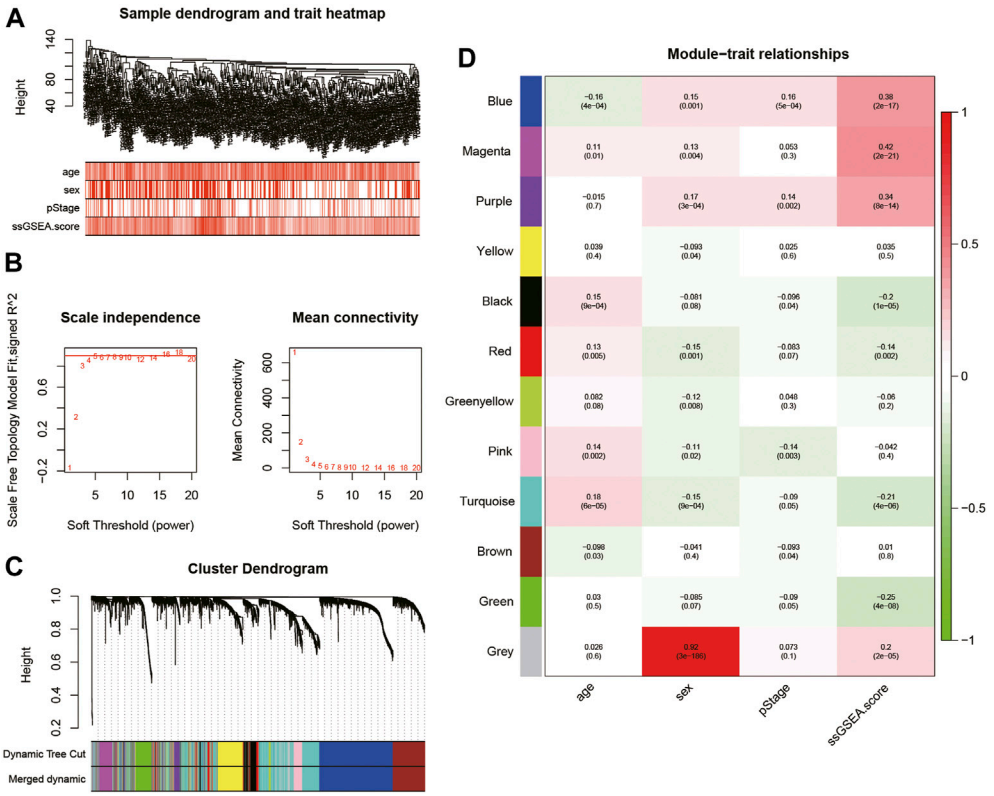


FIGURE 2 | (A) Cluster was based on the expression data of the training cohort, which contained 468 lung adenocarcinoma samples. The top 5,000 genes with the highest standard deviation were used for the analysis by WGCNA. The color intensity was proportional to age, sex, pStage, and ssGSEA score; **(B)** detection of the optimal soft thresholding power; **(C)** cluster dendrogram of genes in the training cohort. Every branch in the figure denotes one gene and each color below denotes one co-expression module; **(D)** heatmap of the relationship between module eigengenes and the ssGSEA score. The blue, magenta, and purple modules were most positively correlated with the ssGSEA score. WGCNA, Weighted Gene Co-expression Network Analysis; ssGSEA, single-sample Gene Set Enrichment Analysis.

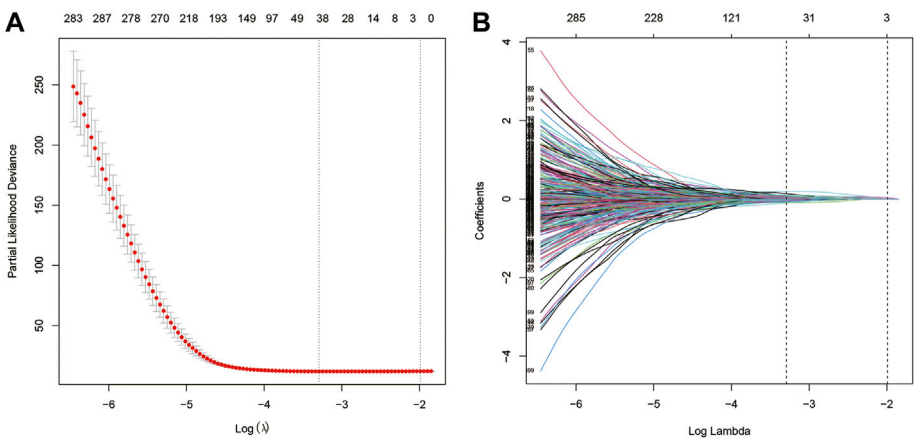


FIGURE 3 | (A) Screening of optimal parameters (lambda) at which the vertical lines were drawn; **(B)** Lasso coefficient profiles of the fatty acid metabolism-related genes with non-zero coefficients determined by the optimal lambda.

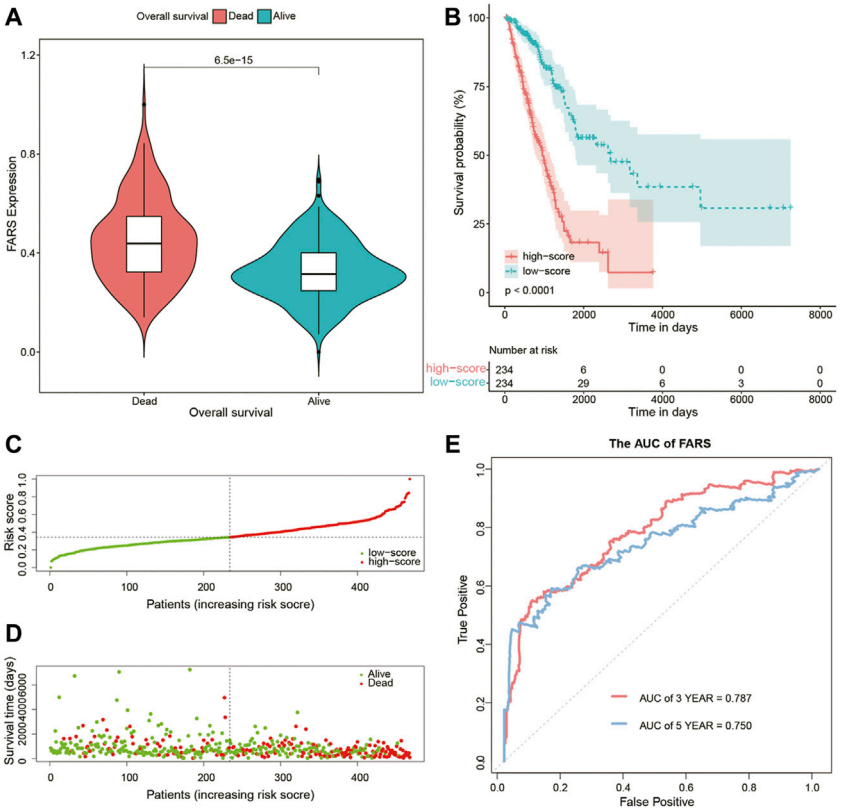


FIGURE 4 | (A) Violin plot comparing the FARS of survival and dead patients for the training cohort; **(B)** Kaplan–Meier curves of patients with high and low FARS for the training cohort; **(C)** distribution patterns of risk scores for the training cohort; **(D)** distribution patterns of survival status for the training cohort; **(E)** ROC curves of the FARS for the training cohort. FARS, fatty acid-related risk score; ROC, receiver operating characteristic.

$$FARS = \sum_{i}^{1} Coefficient(mRNAi) \times Expression(mRNAi).$$

Predictive Power of the Fatty Acid-Related Risk Score in Training and Validation Cohorts

A violin plot was drawn to compare the FARS of the dead and alive patients in the training cohort. Then, the FARS was divided into high-score and low-score groups using the median value as the cut-off value. For different risk score groups in the training cohort, the Kaplan–Meier curve, risk factor curve, and survival status scatter plot were plotted. To authenticate the predictive power of the FARS in the training cohort, the area under the curve (AUC) according to 3- and 5-year OS was calculated and plotted, respectively, using the survivalROC package. In addition, the predictive performance and applicability of the FARS were further verified in the two validation cohorts.

Subgroup Analysis of the Fatty Acid-Related Risk Score in Different Clinicopathological Features

To determine the robustness of the FARS for predicting the survival outcome in different subgroups of clinicopathological features containing age, sex, and pathological stage (pStage), the Kaplan–Meier curve was plotted to evaluate the discriminative capacity of the FARS based on the median value.

Immune Profile

To explore the relationship between immune status and the FARS, the expression of 22 immune cells was calculated in high and low FARS groups using the CIBERSORT package. Furthermore, the expression levels of the immune checkpoint containing PD-1, PD-L1, and CTLA4 of high and low FARS groups were compared.

Establishment of a Prognostic Nomogram

All samples of the training cohort from TCGA database were used to construct the nomogram. The FARS and clinicopathological features were included in the univariate and multivariate Cox analyses with p-value <0.05 as the screening criterion. The variables selected were included in the nomogram using the "RMS" package; the predictive performance of the nomogram was assessed by the time-dependent receiver operating characteristic curves (tROC) and calibration curves.

Statistical Analysis

R software (www.r-project.org) was used in statistical analyses. Manuals for the R packages used in the present study could be

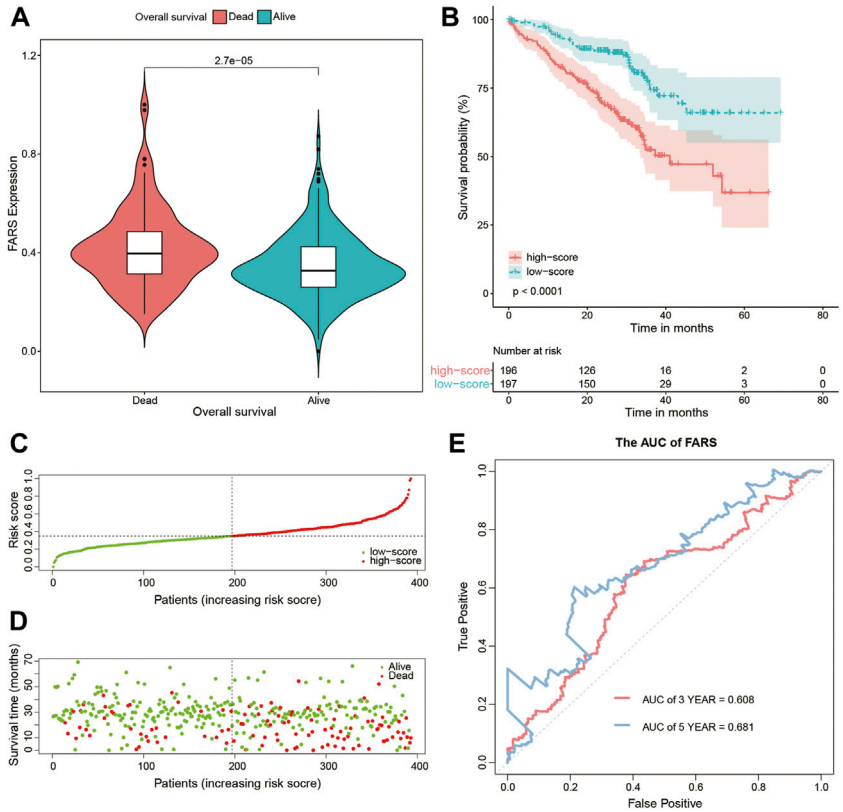


FIGURE 5 | (A) Violin plot comparing the FARS of survival and dead patients for the first validation cohort (GSE72094); **(B)** Kaplan–Meier curve of patients with high and low FARS for the first validation cohort (GSE72094); **(C)** distribution patterns of risk scores for the first validation cohort (GSE72094); **(D)** distribution patterns of the survival status for the first validation cohort (GSE72094); **(E)** ROC curves of the FARS for the first validation cohort (GSE72094). FARS, fatty acid-related risk score; ROC, receiver operating characteristic.

downloaded from the Internet (https://cran.r-project.org/web/packages/available_packages_by_name.html).

Clinicopathological characteristics of the training and validation cohorts were compared using the Kruskal–Wallis test and chi-square test as appropriate.

RESULTS

Data Processing

Expression data of FPKM and clinicopathological features of 468 LUAD cancer samples from TCGA were selected and downloaded according to the screening criteria, and these samples were analyzed as the training cohort. A total of 393 cancer samples in GSE72094 were the first validation cohort and the corresponding expression data and clinicopathological features were downloaded using the GEOquery package. Similarly, the information of 226 samples from GSE31210 was obtained as the second validation set. The clinicopathological characteristics of these three cohorts are presented in **Supplementary Material S2**. A total of 170 (36.3%), 111 (28.2%), and 35 (15.5%) patients died in the training, first validation, and second validation cohorts, respectively. The median (interquartile range) follow-up times for these three cohorts were 658 (435-1118), 825 (541-1012), and 1744 (1246-2050) days, respectively.

Identification of Prognosis-Related Pathways and Biological Processes

Using the CSV file of fatty acid-related pathways and biological processes and the method of ssGSEA, separate enrichment scores for each sample from LUAD were calculated. A total of 10 pathways and biological processes were selected for the univariate Cox analysis in terms of p-value < 0.05. Finally, only four biological processes, namely, gene ontology biological process (GOBP) response to fatty acid, GOBP fatty acid homeostasis, GOBP fatty acid derivative biosynthetic process, and GOBP cellular response to fatty acid proved to be independent risk factors in the multivariate Cox analysis (**Figure 1A**). The Kaplan–Meier curve and violin plot demonstrated similar results that the ssGSEA score was related to prognosis (**Figures 1B,C**).

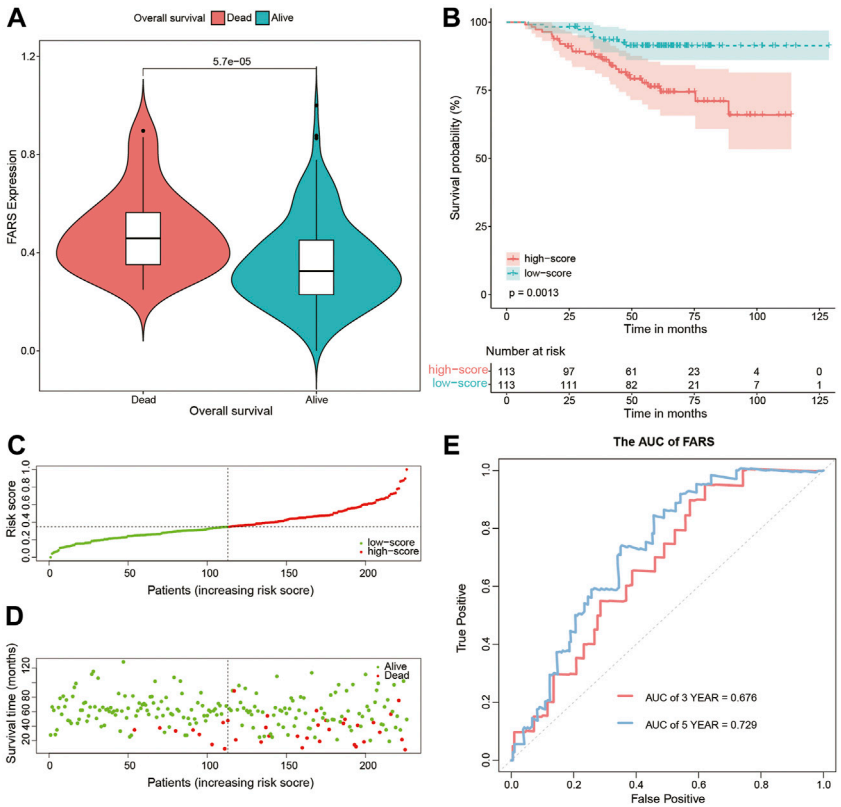


FIGURE 6 | (A) Violin plot comparing the FARS of survival and dead patients for the second validation cohort (GSE31210); **(B)** Kaplan–Meier curves of patients with high and low FARS for the second validation cohort (GSE31210); **(C)** distribution patterns of risk scores for the second validation cohort (GSE31210); **(D)** distribution patterns of the survival status for the second validation cohort (GSE31210); **(E)** ROC curves of the FARS for the second validation cohort (GSE31210). FARS, fatty acid-related risk score; ROC, receiver operating characteristic.

Establishment of the Fatty Acid-Related Risk Score

WGCNA was performed on the top 5,000 genes with the highest standard deviation in TCGA cohort to explore the genes correlated with the fatty acid-associated pathways, and the sample dendrogram was exhibited (**Figure 2A**). The quality of all samples was very good. The optimal soft thresholding power of 4 was selected to obtain the adjacency matrix (**Figure 2B**). A total of 12 modules were constructed under the minimum module size of 30 and there was no module merged with the minimum height for merging modules of 0.2 (**Figure 2C**). In **Figure 2D**, the blue, magenta and purple modules met the criteria of p-value <0.05 and $|r| > 0.3$.

Univariate Cox analysis was performed for each 1,567 genes in the three modules, and those 653 genes with p-value <0.05 in the univariate analysis were included in the LASSO regression analysis (**Supplementary Material S3**) to construct the FARS. The ten-fold cross-validation was used to determine the optimal penalty parameter (λ) of the model (**Figure 3**). A total of 38 genes (LDHA, TM4SF1, HPCAL1, P4HA1, TP53I3, HGSNAT, MYO6, SQLE, IVD, KLHDC8B, GNPNAT1, PAQR4, ENPP5, JAG1, MCTP2, PLEKHA6, MAOB, ANKRD29, ELOVL6, ABAT, ZNF738, BEX5, LETM2, WASF1, INPP5J, DKK1, SLC4A5, CDC25C, FAIM2, BAIAP2L2, GPR37, TM4SF4, TCN1, GALNT13, CNTNAP2, IGFBP1, IGF2BP1, and SALL1) were included in the LASSO model. The coefficient of each gene is given in **Supplementary Material S4**.

Predictive Power of the Fatty Acid-Related Risk Score in Training and Validation Cohorts

As shown in **Figure 4A**, the dead patients had a significantly higher FARS than the alive patients during the follow-up period. A total of 468 samples of the training cohort were divided into the high FARS group and low FARS group by the median value of the FARS, and the high FARS group exhibited worse OS than the low FARS group, with p-value <0.01 (**Figure 4B**). Moreover, the plot distribution patterns of risk scores and survival status showed good results (**Figures 4C,D**). As the FARS increased, the OS time decreased and mortality increased. Calculated by the "survivalROC" package, the 3- and 5-year AUC (AUC = 0.787 and 0.750) of the FARS is presented in **Figure 4E**.

These results indicated that the high FARS showed a worse prognosis, and FARS had a good performance in predicting the prognosis of LUAD patients. The same result was also demonstrated in the two validation cohorts. The FARS of the two validation cohorts was calculated according to the LASSO

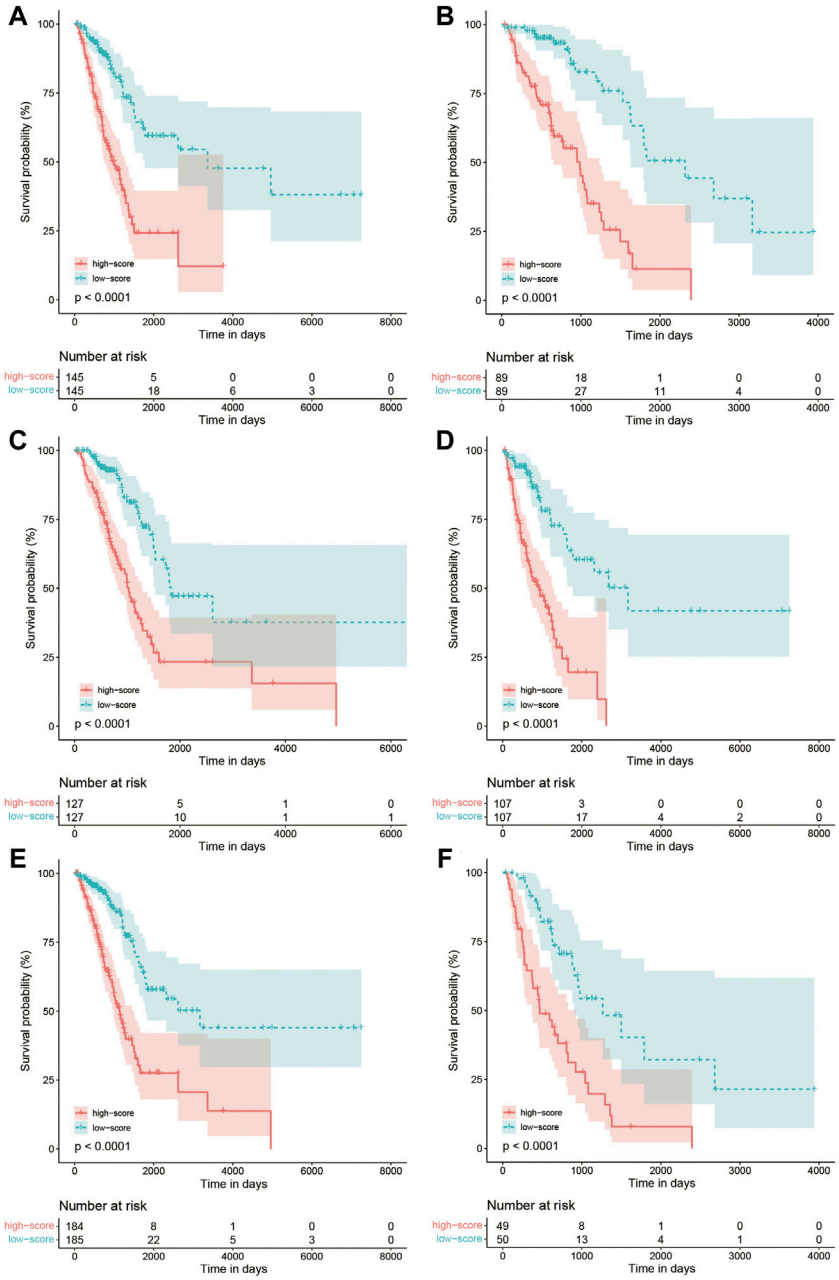


FIGURE 7 | (A) Kaplan–Meier curves of patients with age<70 in the training cohort comparing high and low FARS; **(B)** Kaplan–Meier curves of patients with age ≥ 70 in the training cohort comparing high and low FARS; **(C)** Kaplan–Meier curves of female patients in the training cohort comparing high and low FARS; **(D)** Kaplan–Meier curves of male patients in the training cohort comparing high and low FARS; **(E)** Kaplan–Meier curves of I–II pStage patients in the training cohort comparing high and low FARS; **(F)** Kaplan–Meier curves of III–IV pStage patients in the training cohort comparing high and low FARS. FARS, fatty acid-related risk score; pStage, pathological stage.

formula. The dead patients had significantly higher FARS than the alive patients in the two validation cohorts (p-value <0.01, **Figure 5A**, **Figure 6A**). The high FARS patients showed a worse OS compared with the low FARS patients (p-value <0.01, **Figures 5B**, **6B**). The trend of the survival time and mortality in the validation cohorts was similar to those in the training cohort (**Figures 5C**,D, **6C**,D). We also found that the 3- and 5-year AUC (AUC of validation 1 = 0.608, 0.681; AUC of validation 2 = 0.676, 0.729) of the FARS was similar to those of the training group (**Figures 5E**, **6E**).

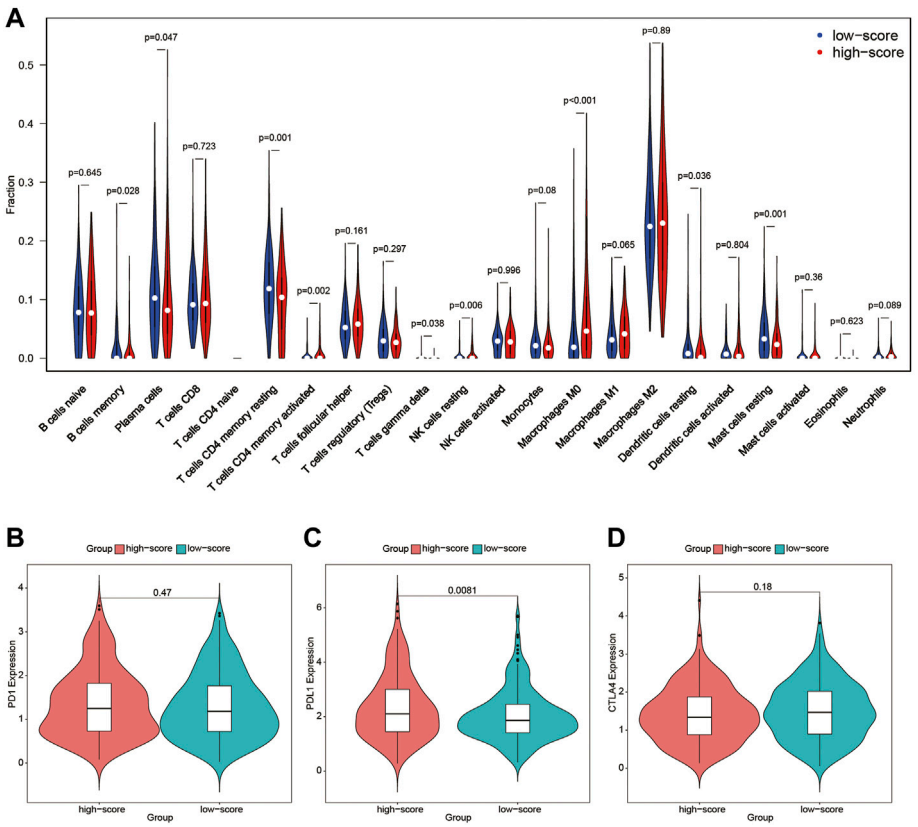


FIGURE 8 | (A) Immune profile of high and low FARS; **(B)** expression of programmed cell death protein 1 (PD1) of high and low FARS; **(C)** expression of programmed death-ligand 1 (PDL1) of high and low FARS; **(D)** expression of cytotoxic T-lymphocyte-associated protein 4 (CTLA4) of high and low FARS. FARS, fatty acid-related risk score.

Subgroup Analysis of the Fatty Acid-Related Risk Score in Different Clinicopathological Features

The Kaplan–Meier curve comparing the survival outcome of high and low FARS in subgroups of clinicopathological features including age, sex, and pStage is shown in **Figures 7A–F**, where the high FARS patients exhibited worse OS than the low FARS patients in each subgroup.

Immune Profile

To further detect the correlation between the immune profile and FARS, the degree of immune cell infiltration of the high and low FARS groups was compared (**Figure 8A**). B cells memory, plasma cells, T cells CD4 memory resting, T cells gamma delta, dendritic cells resting, and mast cells resting were significantly more prevalent in the low FARS patients. Whereas, T cells CD4 memory activated, NK cells resting, and macrophages M0 were significantly more prevalent in the high FARS patients.

The expression of immune checkpoint inhibitors targeting immune checkpoint proteins including programmed cell death protein 1 (PD1), programmed death-ligand 1 (PDL1), and cytotoxic T-lymphocyte-associated protein 4 (CTLA4) was also calculated in high and low FARS patients (**Figures 8B–D**). The high FARS group exhibited higher expression of PDL1 (**Figure 8C**), suggesting that patients with a high FARS may be more susceptible to immune checkpoint inhibitors targeting PD1/PDL1.

Establishment of a Prognostic Nomogram

In the end, a nomogram was established to predict the OS more conveniently. The FARS and clinicopathological features of all TCGA samples were included in the univariate and multivariate Cox analyses, and FARS and pStage proved to be independent risk factors of LUAD (p-value <0.01). A nomogram based on the FARS and pStage was plotted using the "rms" package (**Figures 9A,B**). The tAUC of FARS, clinicopathological characteristics, and nomogram is shown in **Figure 10A**. The figure showed that the prediction ability of the FARS (AUC = 0.787, 0.796, 0.787, 0.806, and 0.750) was higher than that of other clinicopathological features and the predictive performance of the nomogram (AUC = 0.789, 0.807, 0.798, 0.809, and 0.753) was higher than that of other features. Furthermore, the calibration curves of the nomogram predicting 1-, 3-, and 5-year OS were plotted, and the predicted OS probability was very close to the actual OS probability in each calibration curve (**Figures 10B–D**).

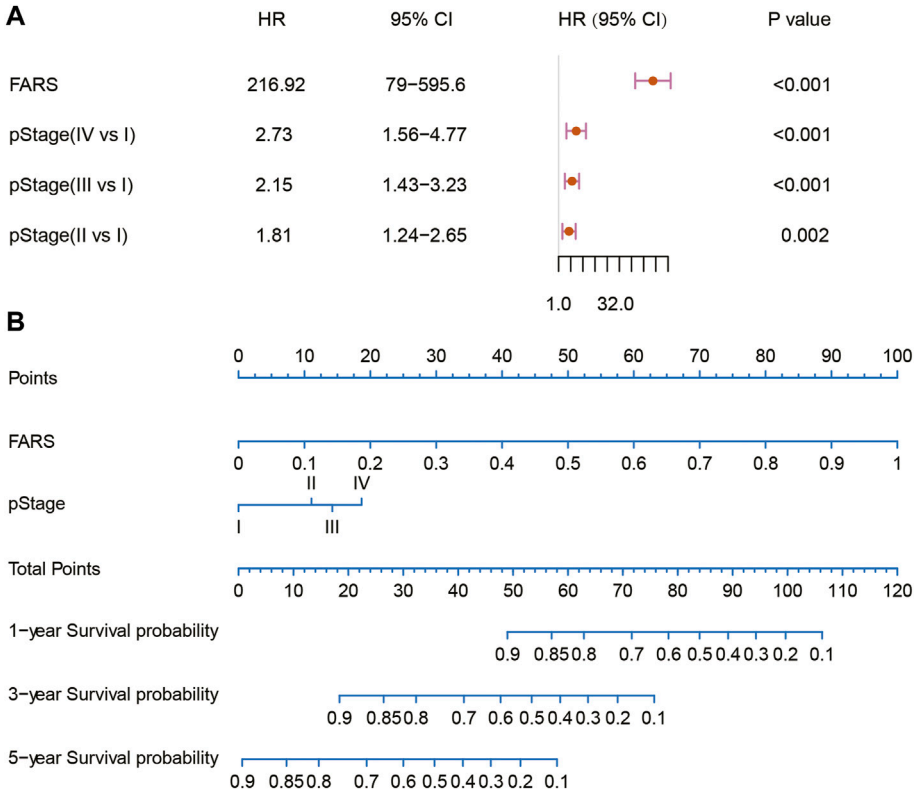


FIGURE 9 | (A) Multivariate Cox analysis of the variables selected for the nomogram; **(B)** nomogram predicting 1-, 3-, and 5-year OS for LUAD patients. OS, overall survival; LUAD, lung adenocarcinoma.

DISCUSSION

Metabolic reprogramming, as a defining characteristic of malignant tumors, plays an important role in cancer development and opens up novel opportunities for cancer therapies (Kapadia et al., 2018). Lipid metabolism has been demonstrated as a pivotal regulator of malignant tumor immunology (Chen et al., 2021a). Most malignant tumor cells exhibit characteristic modifications in *de novo* lipid biosynthesis, lipogenic characteristics, and lipid metabolism (Wymann and Schneiter, 2008; Santos and Schulze, 2012). Upregulated lipid metabolism is one of the physiological features of human cancers and supports signal molecule synthesis and transduction and energy generation (Zaugg et al., 2011; Louie et al., 2013). Molecules regulating lipid metabolism can be potential therapeutic targets. For example, Montal et al. (2015) revealed that phosphoenolpyruvate carboxykinase (PEPCK) promotes lipid synthesis in cancer cells, helping to coordinate a pivotal feature of cancer metabolism. In addition, Svensson et al. (2016)'s study proved that the ACC inhibitor, ND-646, could prevent the biosynthesis of fatty acids in a mouse model of NSCLC. Blockage of fatty acid oxidation leads to the death of lung and ovarian cancer cells (Sullivan et al., 2014). Inhibition of fatty acid oxidation can enhance cancer therapies by modulating the immunosuppressive functions of myeloid-derived suppressor cells (Hossain et al., 2015). Previous studies focused on the influence of a single regulator of pathways related to fatty acid metabolism; however, the comprehensive roles of genes related to fatty acid metabolism have not been presented.

Until now, this is the first research to investigate the relationship between LUAD and fatty acid metabolism-related genes. Based on TCGA and GEO dataset and using the methods of ssGSEA, WGCNA, univariable Cox regression model, and LASSO Cox regression model, a FARS model with 38 fatty acid metabolism-related genes was built and proved to be an independent predictive factor for the survival outcome of LUAD patients. The high-risk score patients showed worse OS than low-risk score patients both in TCGA cohort and two GEO cohorts. Subgroup analyses of different clinicopathological features confirmed the stable prediction of the FARS model. These results indicated the FARS could discriminate patients with poor prognoses. Furthermore, compared to age, gender, and pStage, the FARS model had a larger AUC. As a matter of fact, the AUC of the FARS in predicting survival probability within 5 years was as high as close to 0.8, marking a relatively accurate prediction. Based on the FARS and pStage, a nomogram was established and the calibration plots exhibited good accuracy. It was interesting to find that the FARS model and the nomogram combining the pStage and FARS had a similar AUC, indicating that the pStage brought little improvement to the model. As shown in the nomogram, the FARS contributes much more than the pStage in the points, which means that the FARS maybe more

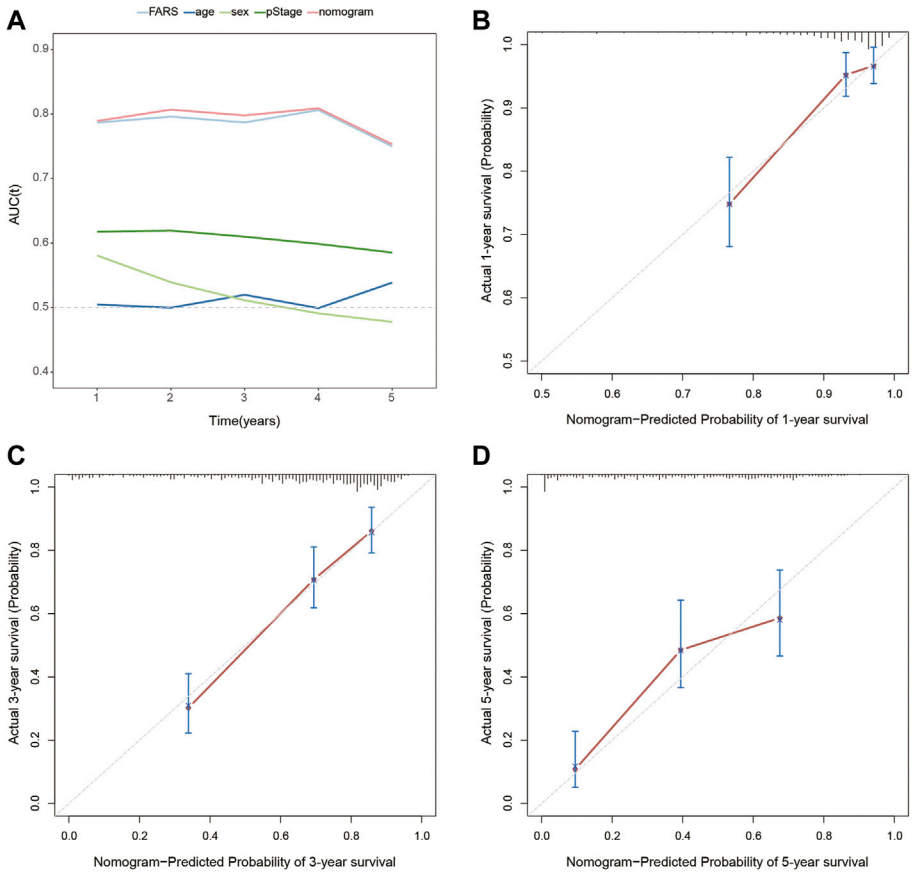


FIGURE 10 | (A) Time-dependent AUC for the FARS, age, sex, pStage, and nomogram; **(B)** calibration plot for 1-year OS prediction of the nomogram; **(C)** calibration plot for 3-year OS prediction of the nomogram; **(D)** calibration plot for 5-year OS prediction of the nomogram. AUC, area under the curve; FARS, fatty acid-related risk score; pStage, pathological stage; OS, overall survival.

important than the pStage in predicting the survival outcome of LUAD patients.

Of all these 38 genes included in the FARS, 30 have been reported to be related to LUAD/NSCLC in the PubMed database. However, no research studies were found to concentrate on the relationship between LUAD with the other eight genes: HGSNAT, ENPP5, MCTP2, PLEKHA6, ANKRD29, ZNF738, SLC4A5, and CNTNAP2. ENPP5, MCTP2, PLEKHA6, ANKRD29, SLC4A5, and CNTNAP2 were shown to play a role in malignancy other than NSCLC (Bralten et al., 2010; Langevin et al., 2012; Smith et al., 2012; Parris et al., 2014; Yang and Loh, 2019; Gopalakrishnan et al., 2020; Song et al., 2020; Sun et al., 2020; Chen et al., 2021b). No previous research studies were published concerning the relationship between cancer and HGSNAT or ZNF738. Overall, the underlying mechanisms of these genes included in the FARS relating to fatty acid metabolism and LUAD prognosis are worth further investigation.

The genes with the largest hazard ratio with a value >1 in the FARS were LDHA, HPCAL1, and IGF2BP1. MYO6, BEX5, and ABAT exhibited the smallest hazard ratio with a value <1. LDHA is a critical enzyme which can catalyze the mutual transformation of lactic acid and pyruvic acid in glycolysis (Massari et al., 2016), and pyruvic acid can be converted into acetyl-CoA which is the substrate for fat synthesis. As reported in previous studies, malignant tumors have a higher level of lipogenesis (Menendez and Lupu, 2007; Röhrig and Schulze, 2016). In the present study, higher expression of LDHA indicated worse OS, which suggests that higher LDHA in LUAD might help enhance fat synthesis by promoting acetyl-CoA generation in glycolysis, thus providing more energy for tumor progression. HPCAL1 was the gene with the second-largest hazard ratio in our study, and interestingly, a recent study demonstrated that HPCAL1 could directly bind to LDHA and enhance its activation, thus influencing fatty acid synthesis and promoting NSCLC growth (Wang et al., 2022). IGF2BP1 was reported to be associated with lipid accumulation in macrophages or serve as a biomarker for NSCLC; however, the relationship between IGF2BP1 and fatty acid metabolism in LUAD had not been shown (Kato et al., 2007; Liu et al., 2022). As for MYO6, its relationship with fatty acid metabolism or LUAD was also obscure. The expression of BEX5 was significantly decreased in several LUAD cell lines compared with normal lung epithelial cells

in vitro and also downregulated in LUAD tissues compared with adjacent normal tissues (Zhang et al., 2019), while the association between BEX5 and fatty acid metabolism in LUAD was never studied. A previous study revealed that ABAT was related to the survival of LUAD patients, but no research disclosed whether ABAT influenced LUAD survival *via* fatty acid metabolism (Wang et al., 2021).

As immunotherapy has drastically improved the survival outcome of many cancer patients as compared to chemotherapy or radiotherapy, more studies concerning the relationship between fatty acid metabolism and immunotherapy were carried out recently (Bleve et al., 2021). For instance, in a Lewis lung carcinoma model, inhibiting CTP1, a rate-limiting enzyme in the fatty acid oxidation cycle, could significantly enhance adoptive cell transfer therapy and reduce tumor progression (Hossain et al., 2015). Another study revealed that lipofermata could reduce the uptake of fatty acid by polymorphonuclear myeloid-derived suppressor cells and suppress tumor progression when combined with anti-CTLA4 or anti-PD1 antibodies in lung carcinoma models (Veglia et al., 2019). Several other studies also proved that targeting fatty acid metabolism could inhibit the immunosuppressive function of myeloid-derived suppressor cells and reduce cancer cell growth (Li et al., 2016; Kim et al., 2017; Prima et al., 2017; Goyal et al., 2018). An improved understanding of the relationship between fatty acid metabolism and the immune profile of the TMEs can help find other therapeutic targets and extend the clinical benefit of immunotherapy to more patients. In this study, the difference in the immune profile of LUAD was compared between the high-risk and low-risk score patients. Although there was no difference in the expression level of PD1 and CTLA4, the high-risk score cohort showed higher PDL1 expression, indicating that the high-risk score patients can potentially benefit from anti-PD1/PDL1 antibodies.

There are some limitations to be noted in the study for the reference of future studies. First, selection bias potentially exists for the inevitable retrospective nature, and the genomic and clinical data are extracted from the public database. Second, the sample size of this study is small, and further studies are warranted to validate the results.

CONCLUSION

In summary, we identified 38 fatty acid metabolism-related gene-based FARS which could accurately predict the survival outcome of LUAD patients. Patients with higher FARS can potentially benefit from anti-PD1/PDL1 immunotherapy. In addition, the mechanisms of the genes in the FARS affecting prognosis are worthy of further research to develop new gene-targeted drugs.

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

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

Conception/design: SW, XL, and ML. Collection and/or assembly of data: AC and WZ. Data analysis and interpretation: SW, AC, and DF. Manuscript writing: SW, JW, QL, and XS. Final approval of manuscript: All authors. Funding support: XL. All authors contributed to the manuscript and approved the submitted version.

FUNDING

This study was funded by Program of Shanghai Academic Research Leader (No. 21XD1402800) and Shanghai "Rising Stars of Medical Talent" Youth Development Program: Outstanding Youth Medical Talents.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.905508/full#supplementary-material

REFERENCES

Bleve, A., Durante, B., Sica, A., Consonni, F. M., Wei, L., Shi, D., et al. (2021). Lipid Metabolism and Cancer Immunotherapy: Immunosuppressive Myeloid Cells at the Crossroad. *Int. J. Mol. Sci.* 14, 5845–5869. doi:10.3390/ijms21165845

Bralten, L. B. C., Gravendeel, A. M., Kloosterhof, N. K., Sacchetti, A., Vrijenhoek, T., Veltman, J. A., et al. (2010). The CASPR2 Cell Adhesion Molecule Functions as a Tumor Suppressor Gene in Glioma. *Oncogene* 29 (46), 6138–6148. doi:10.1038/onc.2010.342

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

Chen, P.-H., Cai, L., Huffman, K., Yang, C., Kim, J., Faubert, B., et al. (2019). Metabolic Diversity in Human Non-small Cell Lung Cancer Cells. *Mol. Cell.* 76 (5), 838–851. e5. doi:10.1016/j.molcel.2019.08.028

Chen, P., Wu, S., Yu, J., Tang, X., Dai, C., Qi, H., et al. (2021). mRNA Network: Solution for Tracking Chemotherapy Insensitivity in Small-Cell Lung Cancer. *J. Healthc. Eng.* 2021, 1–11. doi:10.1155/2021/2105176

Chen, Y.-J., Liao, W.-X., Huang, S.-Z., Yu, Y.-F., Wen, J.-Y., Chen, J., et al. (2021). Prognostic and Immunological Role of CD36: A Pan-Cancer Analysis. *J. Cancer* 12 (16), 4762–4773. doi:10.7150/jca.50502

Chen, Z., Fillmore, C. M., Hammerman, P. S., Kim, C. F., and Wong, K.-K. (2014). Non-small-cell Lung Cancers: a Heterogeneous Set of Diseases. *Nat. Rev. Cancer* 14 (8), 535–546. doi:10.1038/nrc3775

Currie, E., Schulze, A., Zechner, R., Walther, T. C., and Farese, R. V., Jr. (2013). Cellular Fatty Acid Metabolism and Cancer. *Cell. Metab.* 18 (2), 153–161. doi:10.1016/j.cmet.2013.05.017

Ding, C., Shan, Z., Li, M., Chen, H., Li, X., and Jin, Z. (2021). Characterization of the Fatty Acid Metabolism in Colorectal Cancer to Guide Clinical Therapy. *Mol. Ther. - Oncolytics* 20, 532–544. doi:10.1016/j.omto.2021.02.010

Faubert, B., Li, K. Y., Cai, L., Hensley, C. T., Kim, J., Zacharias, L. G., et al. (2017). Lactate Metabolism in Human Lung Tumors. *Cell.* 171 (2), 358–371. doi:10.1016/j.cell.2017.09.019

Geng, F., and Guo, D. (2017). Lipid Droplets, Potential Biomarker and Metabolic Target in Glioblastoma. *Imrj* 3 (5), 10. doi:10.18103/imr.v3i5.443

Gopalakrishnan, K., Aushev, V. N., Manservisi, F., Falcioni, L., Panzacchi, S., Belpoggi, F., et al. (2020). Gene Expression Profiles for Low-Dose Exposure to Diethyl Phthalate in Rodents and Humans: a Translational Study with Implications for Breast Carcinogenesis. *Sci. Rep.* 10 (1), 7067. doi:10.1038/s41598-020-63904-w

Goyal, G., Wong, K., Nirschl, C. J., Souders, N., Neuberg, D., Anandasabapathy, N., et al. (2018). PPARγ Contributes to Immunity Induced by Cancer Cell Vaccines that Secrete GM-CSF. *Cancer Immunol. Res.* 6 (6), 723–732. doi:10.1158/2326-6066.CIR-17-0612

Hensley, C. T., Faubert, B., Yuan, Q., Lev-Cohain, N., Jin, E., Kim, J., et al. (2016). Metabolic Heterogeneity in Human Lung Tumors. *Cell.* 164 (4), 681–694. doi:10.1016/j.cell.2015.12.034

Hossain, F., Al-Khami, A. A., Wyczechowska, D., Hernandez, C., Zheng, L., Reiss, K., et al. (2015). Inhibition of Fatty Acid Oxidation Modulates Immunosuppressive Functions of Myeloid-Derived Suppressor Cells and Enhances Cancer Therapies. *Cancer Immunol. Res.* 3 (11), 1236–1247. doi:10.1158/2326-6066.CIR-15-0036

Hu, J., Yu, H., Sun, L., Yan, Y., Zhang, L., Jiang, G., et al. (2021). Identification of an Individualized Metabolism Prognostic Signature and Related Therapy Regimens in Early Stage Lung Adenocarcinoma. *Front. Oncol.* 11, 650853. doi:10.3389/fonc.2021.650853

Kapadia, B., Nanaji, N. M., Bhalla, K., Bhandary, B., Lapidus, R., Beheshti, A., et al. (2018). Fatty Acid Synthase Induced S6Kinase Facilitates USP11-eIF4B Complex Formation for Sustained Oncogenic Translation in DLBCL. *Nat. Commun.* 9 (1), 829. doi:10.1038/s41467-018-03028-y

Kato, T., Hayama, S., Yamabuki, T., Ishikawa, N., Miyamoto, M., Ito, T., et al. (2007). Increased Expression of Insulin-like Growth Factor-II Messenger RNA-Binding Protein 1 Is Associated with Tumor Progression in Patients with Lung Cancer. *Clin. Cancer Res.* 13 (2 Pt 1), 434–442. doi:10.1158/1078-0432.CCR-06-1297

Kim, S. H., Li, M., Trousil, S., Zhang, Y., Pasca di Magliano, M., Swanson, K. D., et al. (2017). Phenformin Inhibits Myeloid-Derived Suppressor Cells and Enhances the Anti-tumor Activity of PD-1 Blockade in Melanoma. *J. Investigative Dermatology* 137 (8), 1740–1748. doi:10.1016/j.jid.2017.03.033

Langevin, S. M., Koestler, D. C., Christensen, B. C., Butler, R. A., Wiencke, J. K., Nelson, H. H., et al. (2012). Peripheral Blood DNA Methylation Profiles Are Indicative of Head and Neck Squamous Cell Carcinoma: an Epigenome-wide Association Study. *Epigenetics* 7 (3), 291–299. doi:10.4161/epi.7.3.19134

Li, B., Cui, Y., Diehn, M., and Li, R. (2017). Development and Validation of an Individualized Immune Prognostic Signature in Early-Stage Nonsquamous Non-small Cell Lung Cancer. *JAMA Oncol.* 3 (11), 1529–1537. doi:10.1001/jamaoncol.2017.1609

Li, Y., Fang, M., Zhang, J., Wang, J., Song, Y., Shi, J., et al. (2016). Hydrogel Dual Delivered Celecoxib and Anti-PD-1 Synergistically Improve Antitumor Immunity. *Oncoimmunology* 5 (2), e1074374. doi:10.1080/2162402X.2015.1074374

Liu, M., Tao, G., Cao, Y., Hu, Y., and Zhang, Z. (2022). Silencing of IGF2BP1 Restrains Ox-LDL-Induced Lipid Accumulation and Inflammation by Reducing RUNX1 Expression and Promoting Autophagy in Macrophages. *J. Biochem. Mol. Toxicol.* 36 (4), e22994. doi:10.1002/jbt.22994

Louie, S. M., Roberts, L. S., Mulvihill, M. M., Luo, K., and Nomura, D. K. (2013). Cancer Cells Incorporate and Remodel Exogenous Palmitate into Structural and Oncogenic Signaling Lipids. *Biochimica Biophysica Acta (BBA) - Mol. Cell. Biol. Lipids* 1831 (10), 1566–1572. doi:10.1016/j.bbalip.2013.07.008

Lue, H.-w., Podolak, J., Kolahi, K., Cheng, L., Rao, S., Garg, D., et al. (2017). Metabolic Reprogramming Ensures Cancer Cell Survival Despite Oncogenic Signaling Blockade. *Genes Dev.* 31 (20), 2067–2084. doi:10.1101/gad.305292.117

Massari, F., Ciccarese, C., Santoni, M., Iacovelli, R., Mazzucchelli, R., Piva, F., et al. (2016). Metabolic Phenotype of Bladder Cancer. *Cancer Treat. Rev.* 45, 46–57. doi:10.1016/j.ctrv.2016.03.005

Menendez, J. A., and Lupu, R. (2007). Fatty Acid Synthase and the Lipogenic Phenotype in Cancer Pathogenesis. *Nat. Rev. Cancer* 7 (10), 763–777. doi:10.1038/nrc2222

Montal, E. D., Dewi, R., Bhalla, K., Ou, L., Hwang, B. J., Ropell, A. E., et al. (2015). PEPCK Coordinates the Regulation of Central Carbon Metabolism to Promote Cancer Cell Growth. *Mol. Cell.* 60 (4), 571–583. doi:10.1016/j.molcel.2015.09.025201510.1016/j.molcel.2015.09.025

Nath, A., Li, I., Roberts, L. R., and Chan, C. (2015). Elevated Free Fatty Acid Uptake via CD36 Promotes Epithelial-Mesenchymal Transition in Hepatocellular Carcinoma. *Sci. Rep.* 5, 14752. doi:10.1038/srep14752

Parris, T. Z., Aziz, L., Kovács, A., Hajizadeh, S., Nemes, S., Semaan, M., et al. (2014). Clinical Relevance of Breast Cancer-Related Genes as Potential Biomarkers for Oral Squamous Cell Carcinoma. *BMC Cancer* 14, 324. doi:10.1186/1471-2407-14-324

Patz, E. F., Jr., Pinsky, P., Gatsonis, C., Sicks, J. D., Kramer, B. S., Tammemägi, M. C., et al. (2014). Overdiagnosis in Low-Dose Computed Tomography Screening for Lung Cancer. *JAMA Intern Med.* 174 (2), 269–274. doi:10.1001/jamainternmed.2013.12738

Prima, V., Kaliberova, L. N., Kaliberov, S., Curiel, D. T., and Kusmartsev, S. (2017). COX2/mPGES1/PGE 2 Pathway Regulates PD-L1 Expression in Tumor-Associated Macrophages and Myeloid-Derived Suppressor Cells. *Proc. Natl. Acad. Sci. U.S.A.* 114 (5), 1117–1122. doi:10.1073/pnas.1612920114

Qiu, B., Ackerman, D., Sanchez, D. J., Li, B., Ochocki, J. D., Grazioli, A., et al. (2015). HIF2α-Dependent Lipid Storage Promotes Endoplasmic Reticulum Homeostasis in Clear-Cell Renal Cell Carcinoma. *Cancer Discov.* 5 (6), 652–667. doi:10.1158/2159-8290.CD-14-1507

Röhrig, F., and Schulze, A. (2016). The Multifaceted Roles of Fatty Acid Synthesis in Cancer. *Nat. Rev. Cancer* 16 (11), 732–749. doi:10.1038/nrc.2016.89

Santos, C. R., and Schulze, A. (2012). Lipid Metabolism in Cancer. *FEBS J.* 279 (15), 2610–2623. doi:10.1111/j.1742-4658.2012.08644.x

Sayin, V. I., LeBoeuf, S. E., and Papagiannakopoulos, T. (2019). Targeting Metabolic Bottlenecks in Lung Cancer. *Trends Cancer* 5 (8), 457–459. doi:10.1016/j.trecan.2019.06.001

Siegel, R. L., Miller, K. D., and Jemal, A. (2016). Cancer Statistics, 2016. *CA A Cancer J. Clin.* 66 (1), 7–30. doi:10.3322/caac.21332

Smith, S. J., Tilly, H., Ward, J. H., Macarthur, D. C., Lowe, J., Coyle, B., et al. (2012). CD105 (Endoglin) Exerts Prognostic Effects via its Role in the Microvascular Niche of Paediatric High Grade Glioma. *Acta Neuropathol.* 124 (1), 99–110. doi:10.1007/s00401-012-0952-1

Song, H., Sun, J., Kong, W., Ji, Y., Xu, D., and Wang, J. (2020). Construction of a circRNA-Related ceRNA Prognostic Regulatory Network in Breast Cancer. *Ott* Vol. 13, 8347–8358. doi:10.2147/OTT.S266507

Sullivan, E. J., Kurtoglu, M., Brenneman, R., Liu, H., and Lampidis, T. J. (2014). Targeting Cisplatin-Resistant Human Tumor Cells with Metabolic Inhibitors. *Cancer Chemother. Pharmacol.* 73 (2), 417–427. doi:10.1007/s00280-013-2366-8

Sun, G., Li, Z., He, Z., Wang, W., Wang, S., Zhang, X., et al. (2020). Circular RNA MCTP2 Inhibits Cisplatin Resistance in Gastric Cancer by miR-99a-5p-Mediated Induction of MTMR3 Expression. *J. Exp. Clin. Cancer Res.* 39 (1), 246. doi:10.1186/s13046-020-01758-w

Svensson, R. U., Parker, S. J., Eichner, L. J., Kolar, M. J., Wallace, M., Brun, S. N., et al. (2016). Inhibition of Acetyl-CoA Carboxylase Suppresses Fatty Acid

Synthesis and Tumor Growth of Non-small-cell Lung Cancer in Preclinical Models. *Nat. Med.* 22 (10), 1108–1119. doi:10.1038/nm.4181

Veglia, F., Tyurin, V. A., Blasi, M., De Leo, A., Kossenkov, A. V., Donthireddy, L., et al. (2019). Fatty Acid Transport Protein 2 Reprograms Neutrophils in Cancer. *Nature* 569 (7754), 73–78. doi:10.1038/s41586-019-1118-2

Wang, Z., Pei, H., Liang, H., Zhang, Q., Wei, L., Shi, D., et al. (2021). Construction and Analysis of a circRNA-Mediated ceRNA Network in Lung Adenocarcinoma. *Onco Targets Ther.* 14, 3659–3669. doi:10.2147/OTT.S305030

Wang, X., Xie, X., Zhang, Y., Ma, F., Pang, M., Laster, K. V., et al. (2022). Hippocalcin-like 1 Is a Key Regulator of LDHA Activation that Promotes the Growth of Non-small Cell Lung Carcinoma. *Cell. Oncol.* 45 (1), 179–191. doi:10.1007/s13402-022-00661-0

Wymann, M. P., and Schneiter, R. (2008). Lipid Signalling in Disease. *Nat. Rev. Mol. Cell. Biol.* 9 (2), 162–176. doi:10.1038/nrm2335

Xue, C., Li, G., Bao, Z., Zhou, Z., and Li, L. (2021). Mitochondrial Pyruvate Carrier 1: a Novel Prognostic Biomarker that Predicts Favourable Patient Survival in Cancer. *Cancer Cell. Int.* 21 (1), 288. doi:10.1186/s12935-021-01996-8

Yang, O. C. Y., and Loh, S.-H. (2019). Acidic Stress Triggers Sodium-Coupled Bicarbonate Transport and Promotes Survival in A375 Human Melanoma Cells. *Sci. Rep.* 9 (1), 6858. doi:10.1038/s41598-019-43262-y

Yu, T., Wang, Y., Fan, Y., Fang, N., Wang, T., Xu, T., et al. (2019). CircRNAs in Cancer Metabolism: a Review. *J. Hematol. Oncol.* 12 (1), 90. doi:10.1186/s13045-019-0776-8

Zaugg, K., Yao, Y., Reilly, P. T., Kannan, K., Kiarash, R., Mason, J., et al. (2011). Carnitine Palmitoyltransferase 1C Promotes Cell Survival and Tumor Growth under Conditions of Metabolic Stress. *Genes Dev.* 25 (10), 1041–1051. doi:10.1101/gad.1987211

Zhang, Z. H., Luan, Z. Y., Han, F., Chen, H. Q., Liu, W. B., Liu, J. Y., et al. (2019). Diagnostic and Prognostic Value of the BEX Family in Lung Adenocarcinoma. *Oncol. Lett.* 18 (5), 5523–5533. doi:10.3892/ol.2019.10905

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 26 August 2022
DOI 10.3389/fgene.2022.968601

OPEN ACCESS

EDITED BY
Jinhui Liu,
Nanjing Medical University, China

REVIEWED BY
Chao Ma,
First Affiliated Hospital of Zhengzhou University, China
Ti-wei Miao,
Sichuan University, China

*CORRESPONDENCE
Peng Lin,
linpeng@sysucc.org.cn
Lanjun Zhang,
zhanglj@sysucc.org.cn

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 14 June 2022
ACCEPTED 04 August 2022
PUBLISHED 26 August 2022

CITATION
Zhang R, Zhang X, Yang H, Lin Y, Wen Y, Zhao D, Chen L, Lin P and Zhang L (2022), Ferroptosis-related lncRNAs signature to predict the survival and immune evasion for lung squamous cell carcinoma.
Front. Genet. 13:968601.
doi: 10.3389/fgene.2022.968601

Ferroptosis-related lncRNAs signature to predict the survival and immune evasion for lung squamous cell carcinoma

Rusi Zhang[1,2], Xuewen Zhang[1,3], Han Yang[1,2], Yongbin Lin[1,2], Yingsheng Wen[1,2], Dechang Zhao[1,2], Lianjuan Chen[1,2], Peng Lin[1,2]* and Lanjun Zhang[1,2]*

[1]State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China, [2]Department of Thoracic Surgery, Sun Yat-sen University Cancer Center, Guangzhou, China, [3]Department of Anesthesiology, Sun Yat-sen University Cancer Center, Guangzhou, China

Introduction: the investigation on the interactions between ferroptosis and lncRNAs for lung squamous cell carcinoma (LUSC) has been scare, and its impact on tumor immune microenvironment remained unknown. We aim to not only identify a ferroptosis-related lncRNAs signature for LUSC prognosis, but also evaluate its correlation to tumor immune evasion.

Methods: RNA sequencing data and survival information were obtained from The Cancer Genome Atlas database. A ferroptosis-related lncRNAs signature (FerRLSig) was developed and validated by univariate Cox regression, Least Absolute Shrinkage and Selection Operator regression and multivariate Cox regression. The tumor immune microenvironment and immune evasion were subsequently evaluated based on the FerRLSig stratification.

Results: the FerRLSig consisted of 10 ferroptosis-related lncRNAs and significantly associated with overall survival with satisfactory area under curve (HR = 2.240, 95% CI: 1.845–2.720, $p < 0.001$, 5-years AUC: 0.756). Based on the FerRLSig stratification, the high-risk group demonstrated not only significantly higher immune infiltration, but also more profound T cell dysfunction and immune evasion, which might ultimately lead to the resistance to current immune checkpoint inhibitors.

Conclusion: a robust prognostic FerRLSig for LUSC has been developed and validated, demonstrating a close association not only with tumor immune cell infiltration, but also with T cell dysfunction and immune evasion. Further investigation is warranted to better improve the survival of LUSC patients based on the FerRLSig stratification.

KEYWORDS

lung cancer, ferroptosis, lncRNA, tumor immune microenvironment, T cell dysfunction, immune evasion

Introduction

Lung cancer has been the top one cause of cancer death with one of the highest incidence rates second only to breast cancer worldwide (Sung et al., 2021). Lung squamous cell carcinoma (LUSC) accounts for approximately 20% of all lung cancer cases and constitutes the bulk of non-small cell lung cancer with lung adenocarcinoma (Barta et al., 2019). There has been numerous effective targeted therapies for lung adenocarcinoma, which has significantly prolonged the survival of lung adenocarcinoma patients with certain mutations (Sordella et al., 2004; Solomon et al., 2014; Ramalingam et al., 2020; Shaw et al., 2020; Wu et al., 2020). Moreover, a large number of researches have utilized the transcriptome data of lung adenocarcinoma to build various prognostic and predictive tools, to derive useful risk stratification, and to provide valuable insights on the development of sensitive drugs (Guo et al., 2021; Lu et al., 2021; Zheng et al., 2021). However, compared to lung adenocarcinoma, LUSC lacks effective targeted therapy and is generally less well-defined in terms of gene expression profile.

On the other hand, tumor immune evasion has been identified as one of the hallmarks of cancer and closely related to the tumor immune microenvironment (Hanahan and Weinberg, 2011; Gajewski et al., 2013). And fortunately, immune checkpoint inhibitors tackling tumor immune evasion have made remarkable breakthroughs in LUSC and significantly improved the LUSC patients' survival (Brahmer et al., 2015; Paz-Ares et al., 2018). But still some LUSC patients were resistant to the current treatments including immunotherapy, leading to intractable progression or relapse, and ultimately cancer death. Multiple studies have identified increased $CD8^+$ T cell infiltration as a favorable prognostic factor (Lee and Ruppin, 2019; Morad et al., 2021). However, T cell dysfunction has also been recognized as an important mechanism of immunotherapy resistance (Thommen and Schumacher, 2018). Therefore, further evaluation of the LUSC gene expression pattern's impact on tumor immune microenvironment and immune evasion are still needed for LUSC patients.

Ferroptosis is an iron-dependent, oxidatively regulated cell death activated by extrinsic blockade of the cystine/glutamate transporter or intrinsic blockade of intracellular antioxidants. Recent studies have demonstrated that ferroptosis plays a significant part in tumorigenesis and various treatment sensitivity, thus might be a useful tool in cancer prognosis and patient stratification (Dixon et al., 2012; Chen et al., 2021). Moreover, long non-coding RNA (lncRNA), with more than 200 nucleotide and without functional protein translation, has been found to be closely related to tumorigenesis and tumor progression *via* ferroptosis in recent studies (Mao et al., 2018; Wang et al., 2019; Zhang et al., 2020; Statello et al., 2021). Therefore, interactions between ferroptosis and lncRNAs are likely to be critical to overcome cancer progression. However, investigation on the ferroptosis-related lncRNAs signature on LUSC has been scare and its impact on tumor immune microenvironment remained unknown, thus warranting further investigation.

In this study, we utilized The Cancer Genome Atlas database on lung squamous cell carcinoma (TCGA-LUSC) (Cancer Genome Atlas Research Network, 2012), and we aimed to develop a ferroptosis-related lncRNAs signature for LUSC prognosis. Furthermore, explorations on tumor immune microenvironment, tumor immune evasion and T cell dysfunction were also performed based on the FerRLSig stratification.

Materials and methods

Data acquisition

RNA sequencing data of TCGA-LUSC patients and all available clinical data were downloaded from the Genomic Data Commons portal (https://portal.gdc.cancer.gov/) (Grossman et al., 2016). Overall survival (OS) was calculated from the time of lung cancer diagnosis to the time of death or the last follow-up. Patients with incomplete survival information or with OS less than 30 days were excluded.

This is a retrospective study based on publicly available TCGA database. The Ethics Committee of our hospital has confirmed that no additional ethical approval or informed consent is required.

A list of 108 validated ferroptosis genes was obtained from FerrDb database (http://www.zhounan.org/ferrdb) (Zhou and FerrDb, 2020). LncRNAs was identified in TCGA-LUSC RNA sequencing data through GENCODE annotation (https://www.gencodegenes.org/) (Frankish et al., 2021). The correlations between 108 ferroptosis genes and the lncRNAs expression in the entire set of TCGA-LUSC were analyzed with Pearson's correlation, and 4,259 ferroptosis-related lncRNAs were identified by the selection criterion of $|\text{Pearson R}| > 0.3$ and $p < 0.001$ with any one of the 108 ferroptosis genes.

Development and validation of the ferroptosis-related LncRNAs signature

The TCGA-LUSC patients were randomized into a training set and a testing set by the "caret" R package at the ratio of 7:3 (Kuhn, 2008). The FerRLSig was established with the training set while the validation was performed with the testing set and the entire set. Univariate Cox regression, Least Absolute Shrinkage and Selection Operator (LASSO) regression and multivariate Cox regression were applied in order to establish the final FerRLSig while avoiding overfitting with cross-validation (Friedman et al., 2010). The correlation between FerRLSig lncRNAs and

ferroptosis genes was further visualized in both heatmap and Sankey diagram. The risk score of every patient was calculated as $\sum_{i=1}^{n} Coef(i) \times Expr(i)$, with Coef(i) and Expr(i) representing the regression coefficient and expression level for each FerRLSig lncRNA respectively. The entire set of TCGA-LUSC patients were stratified into low-risk and high-risk groups by the median of the calculated risk scores.

Further evaluation of the FerRLSig and establishment of the prognostic nomogram

The prognosis effect of the FerRLSig was further evaluated with Kaplan–Meier survival plot, subgroup analysis by universal clinical characteristics, correlation with universal clinical variables, univariate and multivariate Cox regression, and the area under the receiver operating characteristic curve (AUC). In addition, the discriminative ability of the FerRLSig was further evaluated through the comparison to the Principal component analysis (PCA) and the t-distributed stochastic neighbor embedding (t-SNE).

A nomogram incorporating the FerRLSig, age, gender and TNM stage, was developed to visualize the prognostic model and facilitate its clinical application for the OS of LUSC patients. The prognostic power of the nomogram was evaluated with AUC curves at 1, 3, and 5 years' OS, calibration curves were also plotted and visually assessed. Furthermore, decision curve analysis (DCA) was also conducted to demonstrate the actual net benefit gain of the nomogram (Vickers and Elkin, 2006).

Tumor immune microenvironment exploration and drug sensitivity screening based on the FerRLSig stratification

To analyze and compare the tumor immune microenvironment based on the FerRLSig stratification, xCell and ESTIMATE were both applied to inferred the immune and stromal cell infiltration (Yoshihara et al., 2013; Aran et al., 2017). In addition, to further characterize the potential underlying molecular pathways, gene set enrichment analysis (GSEA) was performed to identify the significantly enriched pathways in low-risk and high-risk groups respectively. The hallmark gene sets and C5 gene sets from the Gene Ontology (GO) were downloaded from the Molecular Signatures Database as the reference files (Subramanian et al., 2005). A nominal p value < 0.05 and a false discovery rate (FDR) q value < 0.25 were set as the statistically significant thresholds for GSEA GO analysis. Moreover, the expression level of several immune checkpoints and immune inhibitory factors, including CCL2, CD274 (PD-L1), CTLA4, CXCR4, IL6, LAG3, PDCD1 (PD-1), and TGFB1 were compared between low-risk and high-risk group. To further evaluate the predictive application for immune checkpoint inhibitors of the FerRLSig, TIDE score were compared between low-risk and high-risk groups (Fu et al., 2020). Furthermore, drug sensitivity screening was also performed with the 198 compounds available from the Genomics of Drug Sensitivity in Cancer (GDSC) database (Yang et al., 2013). And the half-maximal inhibitory concentration (IC50) of the available compounds on low-risk and high-risk groups was extracted with the Oncopredict R package (Maeser et al., 2021).

Statistical analysis

Discrete variables were described as counts and percentages, their differences between groups were statistically evaluated with Pearson chi-square test or Fisher's exact test (any expected values less than 5). On the other hand, continuous variables were described as median, mean and/or interquartile range, and their differences between groups were compared with Mann–Whitney–Wilcoxon test.

All statistical analyses and visualizations were performed with R version 4.1.0 (http://www.R-project.org) and corresponding packages. The Kaplan-Meier method was utilized in survival analysis and survival curves were compared with log-rank test. Two-sided $p < 0.05$ was considered statistically significant.

Results

Development and validation of the FerRLSig

In total, 493 patients with RNA sequencing data were downloaded from the TCGA-LUSC database, and 473 patients with necessary survival information were included and randomly assigned into training set and testing set by the ratio of 7:3 as demonstrated in the study workflow (Figure 1). The mean age of the patients in the entire set was 67 years old, 74.2% were male, 71.0% were Caucasian and 67.7% were early stage (I-II). No statistically significant difference was found between the training set and the testing set (Table 1). In the entire set, 4,259 lncRNAs were significantly correlated with any one of the 108 ferroptosis genes (|Pearson R|>0.3 and $p < 0.001$, Supplementary Figure S1A), among which 43 were significantly associated with OS by univariate Cox regression in the training set ($p < 0.01$, Figure 2A). Within the training set, 21 ferroptosis-related lncRNAs were further identified as significant prognostic factors *via* LASSO regression with the λ set at lambda. min

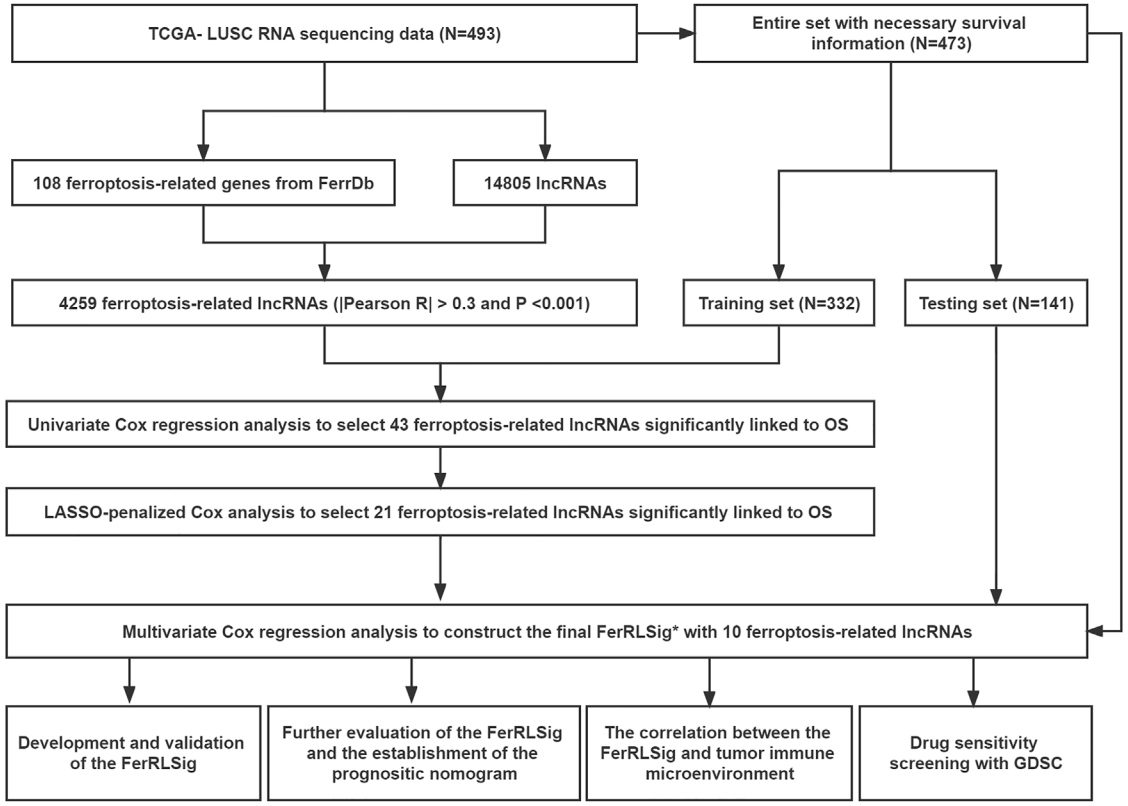


FIGURE 1
Flow chart of this study; *FerRLSig: the ferroptosis-related lncRNAs signature.

TABLE 1 The baseline characteristics of LUSC patients in TCGA database.

	Subgroups	Total, n = 473	Training set, n = 332	Testing set, n = 141	p Value
Age, mean(SD)		67 (9)	67 (8)	67 (9)	0.807
Gender, n (%)	Female	122 (25.8)	83 (25.0)	39 (27.7)	0.624
	Male	351 (74.2)	249 (75.0)	102 (72.3)	
Race, n (%)	Caucasian	336 (71.0)	228 (68.7)	108 (76.6)	0.104
	Other ethnicities	137 (29.0)	104 (31.3)	33 (23.4)	
Stage, n (%)	I-II	320 (67.7)	217 (65.4)	103 (73.0)	0.260
	III	74 (15.6)	56 (16.9)	18 (12.8)	
	IV	79 (16.7)	59 (17.8)	20 (14.2)	
T, n (%)	T1	108 (22.8)	71 (21.4)	37 (26.2)	0.587
	T2	276 (58.4)	200 (60.2)	76 (53.9)	
	T3	68 (14.4)	46 (13.9)	22 (15.6)	
	T4	21 (4.4)	15 (4.5)	6 (4.3)	
N, n (%)	N0	300 (63.4)	204 (61.4)	96 (68.1)	0.320*
	N1	125 (26.4)	89 (26.8)	36 (25.5)	
	N2	39 (8.2)	31 (9.3)	8 (5.7)	
	N3	9 (1.9)	8 (2.4)	1 (0.7)	
M, n (%)	M0	393 (83.1)	272 (81.9)	121 (85.8)	0.369
	M1	80 (16.9)	60 (18.1)	20 (14.2)	

*p value of fisher's exact test for at least one expected count less than 5.

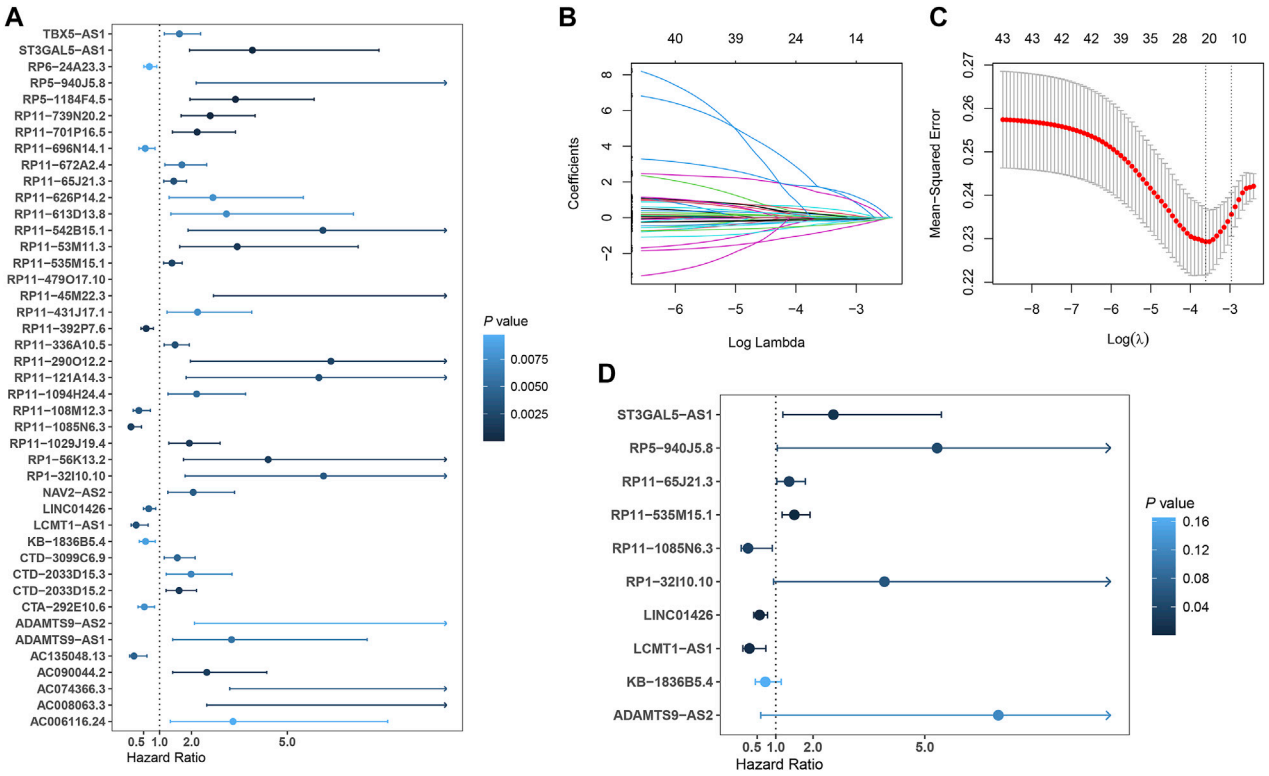


FIGURE 2
Development of prognostic ferroptosis-related lncRNAs signature for LUSC patients; **(A)** Forest plot of the 43 selected lncRNAs significantly associated with overall survival by univariate Cox regression analysis; **(B)** The coefficient profile of 21 OS-related lncRNAs chosen by LASSO regression; **(C)** The mean-squared error curve with different tuning parameters (logλ) and perpendicular dotted lines were drawn at the logλ corresponding to the minimum mean squared error (lambda.min) and the most regularized model within one standard error of the minimum mean squared error (lambda.1-se); **(D)** Forest plot of the 10 final selected lncRNAs by multivariate Cox regression analysis.

that gave the minimum mean squared error (Figures 2B,C). The final FerRLSig was established with multivariate Cox regression, consisting of 10 lncRNAs (Figure 2D). The correlation between FerRLSig lncRNAs and ferroptosis genes were visualized in both heatmap and Sankey diagram for the entire set (Supplementary Figures S1B,C). The training set was divided into high-risk group and low-risk group by the median of the FerRLSig. Compared to low-risk group, high-risk group had evidently more death event and shorter overall survival time (Figure 3A). Within the FerRLSig, RP11-65J21.3, ST3GAL5-AS1, ADAMTS9-AS2, RP5-940J5.8, RP11-535M15.1, and RP1-32I10.10 were over-expressed in the high-risk group with positive coefficients and classified as risk promoters. On the other hand, RP11-1085N6.3, KB-1836B5.4, LCMT1-AS1, and LINC01426 were over-expressed in the low-risk group with negative coefficients and classified as risk inhibitors (Figure 3A; Supplementary Figure S1C). And the final FerRLSig Formula equal to 0.305*(RP11-65J21.3) + 0.934*(ST3GAL5-AS1) + (−1.366)*(RP11-1085N6.3) + 1.942*(ADAMTS9-AS2) + 1.674*(RP5-940J5.8) + (−0.331)*(KB-1836B5.4) + (−1.229)*(LCMT1-AS1) + 0.404*(RP11-535M15.1) + (−0.572)*(LINC01426) + 1.365*(RP1-32I10.10). The high-risk group was significantly associated with worse OS compared to low-risk group in the training set (HR = 3.345, $p < 0.001$, Figure 3C). For validation, the same model was applied to the testing set and the entire set, all FerRLSig lncRNAs demonstrated similar expression profiles in the high-risk and the low-risk groups. In addition, the high-risk group in both testing and entire set was also significantly associated with worse OS compared to low-risk group (testing set HR = 2.290, $p < 0.001$, Figures 3B,D; entire set HR = 2.606, $p < 0.001$, Supplementary Figure S2).

Evaluation of the FerRLSig and the establishment of prognostic nomogram with the FerRLSig

To compare the whole expression profile and the FerRLSig, both PCA and t-SNE were applied to the RNA sequencing of the entire set and annotated with the FerRLSig risk

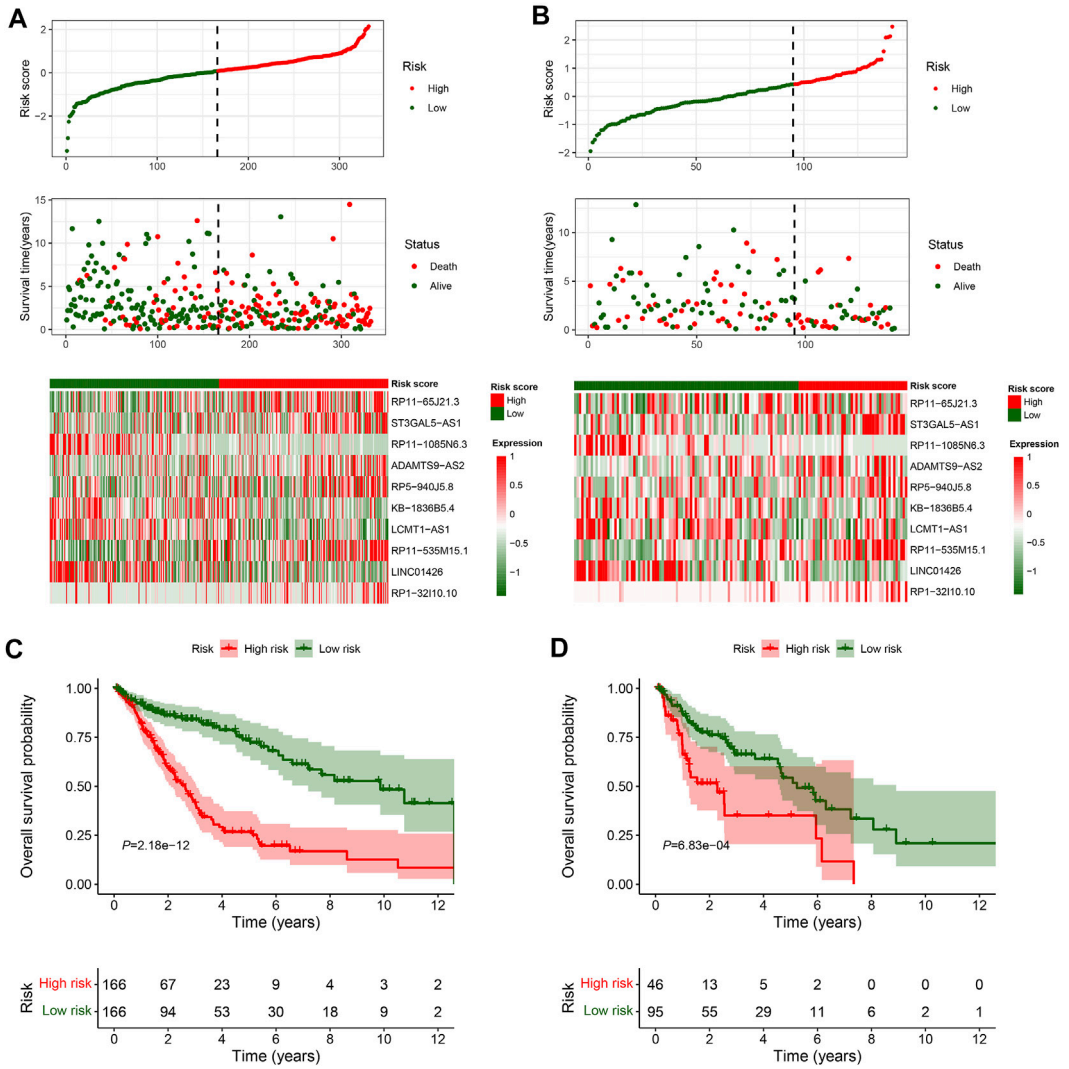


FIGURE 3
Overall survival analysis and validation of the ferroptosis-related lncRNAs signature; **(A–B)** Distribution of risk score, OS time, OS status and heatmap of the 10 prognostic ferroptosis-related lncRNAs signature in the TCGA-LUSC training set **(A)** and TCGA-LUSC testing set **(B)**. **(C–D)** Kaplan-Meier survival curves of the OS of the patients in the high- and low-risk groups for the TCGA-LUSC training set **(C)** and TCGA-LUSC testing set **(D)**.

stratification. The high-risk group and low-risk group demonstrated distinctly different distribution in both PCA and t-SNE, indicating that the FerRLSig risk stratification recapitulated the major variability of the TCGA-LUSC RNA sequencing (Figures 4A,B). The correlations of the FerRLSig risk score with clinical characteristics including age, gender, and TNM stage were explored and no statistically significant correlation was found (Supplementary Figure S3). Univariate and multivariate Cox OS analysis were further performed with FerRLSig and other universal clinical characteristics. Both TNM stage and FerRLSig (univariate: HR = 2.281, $p < 0.001$; multivariate: HR = 2.240, $p < 0.001$) demonstrated significant prognostic effect in both univariate and multivariate Cox regression (Figures 4C,D). In addition, time-dependent AUC of the FerRLSig and other clinical characteristics were plotted, and the FerRLSig demonstrated consistently higher AUC compared to other clinical characteristics, including TNM stage (5-years AUC: FerRLSig 0.756, TNM stage 0.607, Figures 4E,F).

To further demonstrate the model applicability in different population, subgroup OS analysis was performed in different age, gender, ethnicity and stage groups. The high-risk group was consistently associated with worse OS in all subgroups, which not only validated the model's wide applicability in

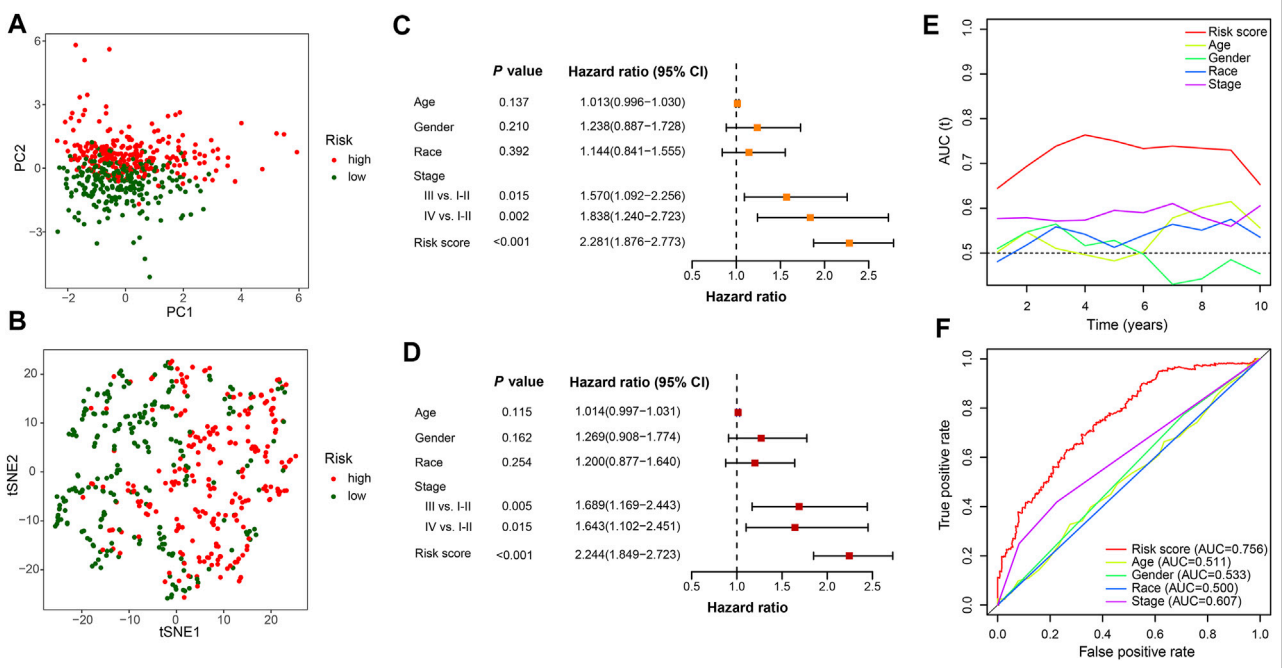


FIGURE 4
Evaluation of the ferroptosis-related lncRNAs signature (FerRLSig) in the entire set of TCGA-LUSC; **(A)** Principal component analysis of TCGA-LUSC RNA sequencing annotated with the FerRLSig stratification; **(B)** t-distributed stochastic neighbor embedding (t-SNE) analysis annotated with the FerRLSig stratification; **(C)** Univariate Cox regression overall survival analysis of the FerRLSig score and universal clinical characteristics; **(D)** Multivariate Cox regression overall survival analysis of the FerRLSig score and universal clinical characteristics; **(E)** Time-dependent area under curve plot of the risk score and clinical characteristics. **(F)** Receiver operating characteristic (ROC) curves of the universal clinical characteristics and risk score of the 5-years overall survival.

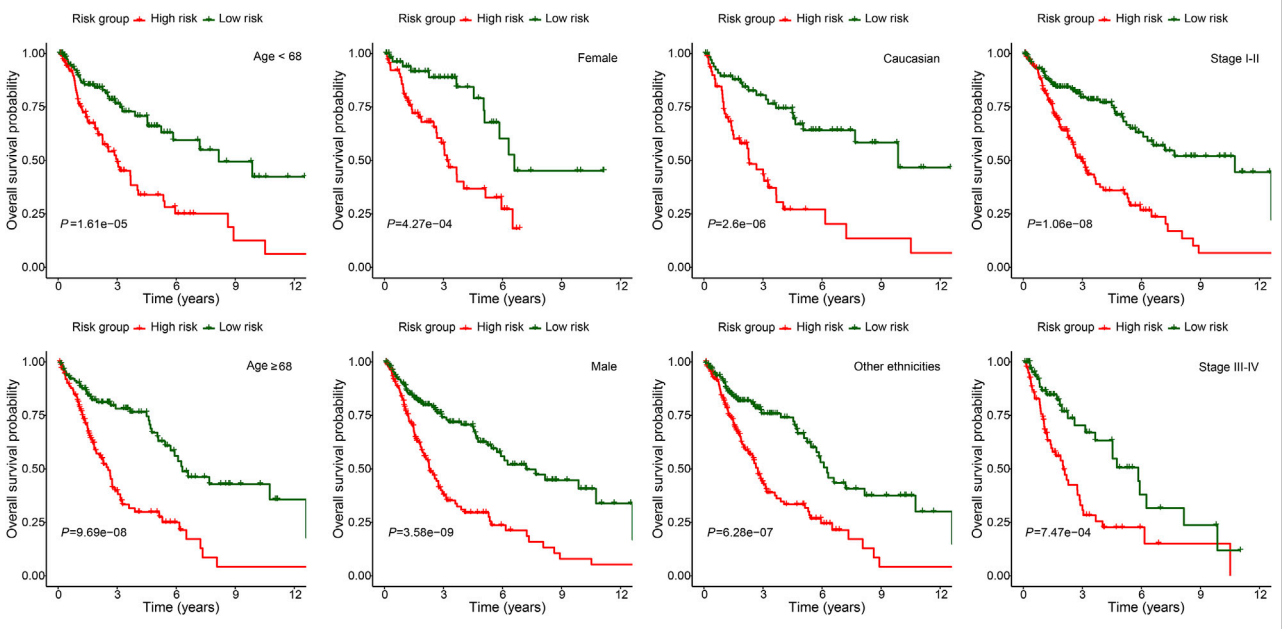


FIGURE 5
Subgroup overall survival analysis of the FerRLSig stratification by different age, gender, ethnicities and TNM stage subgroups in the TCGA-LUSC entire set.

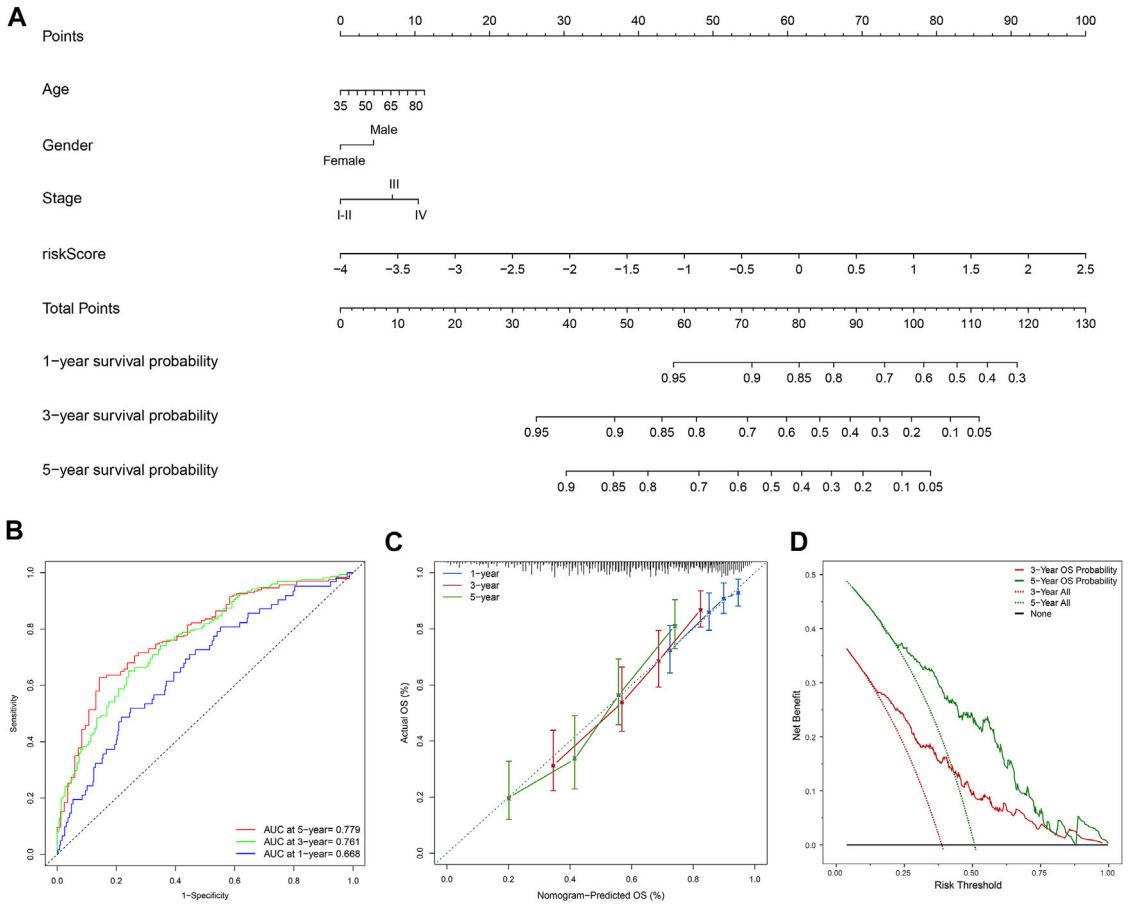


FIGURE 6
Development and evaluation of the prognostic nomogram; **(A)** A clinical prognostic nomogram was developed to predict the 1-, three- and 5-years overall survival (OS) probability. A vertical line between each variable and point scale can be drawn to determine the points for each variable, then all the points are summed up as the total points, and the predicted overall survival rate of the 1-, three- and 5-years were calculated by drawing a vertical line from the total points scale to the 1-, three- and 5-years survival scales; **(B)** Area under curve plot of the nomogram for the 1-, three- and 5-years OS; **(C)** Calibration curves of the nomogram for 1-, three- and 5-years overall survival: nomogram-predicted overall survival is plotted on the *x*-axis, actual overall survival is plotted on the *y*-axis, a plot along the 45-degree line indicates a satisfactory model in which the predicted probabilities are identical to the actual outcomes; **(D)** Decision curve analysis demonstrating the clinical benefit gain of the nomogram for the three- and 5-years OS: the *y*-axis measures the net benefit, which is calculated by summing the benefit (true positives) and subtracting the harms (false positives). The solid line indicates the prognostic model, and the two other lines indicate the "intervention for all" (dotted line) and "intervention for none" (black line). A model is considered of clinical value if it has a higher net benefit than other models at any given threshold.

different population, but also demonstrated its independent prognostic role (Figure 5).

Moreover, the FerRLSig and other universal clinical characteristics were incorporated into a prognostic nomogram to better predict the one, three and 5 years' OS probabilities (Figure 6A). Overall, the nomogram demonstrated satisfactory AUC on 1, 3, and 5 years' OS and calibration (AUC: 1-year 0.668, 3-years 0.761, 5-years 0.779, Figures 6B,C). In addition, the nomogram also demonstrated net benefit gain in decision curve analysis compared to both "intervention to none" and "intervention to all" in both 3 and 5 years' OS prediction (Figure 6D).

The correlation between the FerRLSig and tumor immune microenvironment

To explore the correlation between the FerRLSig and the tumor immune microenvironment, immune cell infiltration was inferred and compared between the high-risk and low-risk group with xCell analysis. The two groups exhibited apparently different tumor immune microenvironment. The high-risk group demonstrated significantly higher dendritic cells, B cells, class-switched memory B cells, $CD8^+$ T cells, and multiple myeloid cells infiltration while the low-risk group had significantly higher pro B cells, Th1 cells and Th2 cells infiltration (Figures 7A,B). Moreover, both immune score and stromal score were significantly higher while tumor purity was

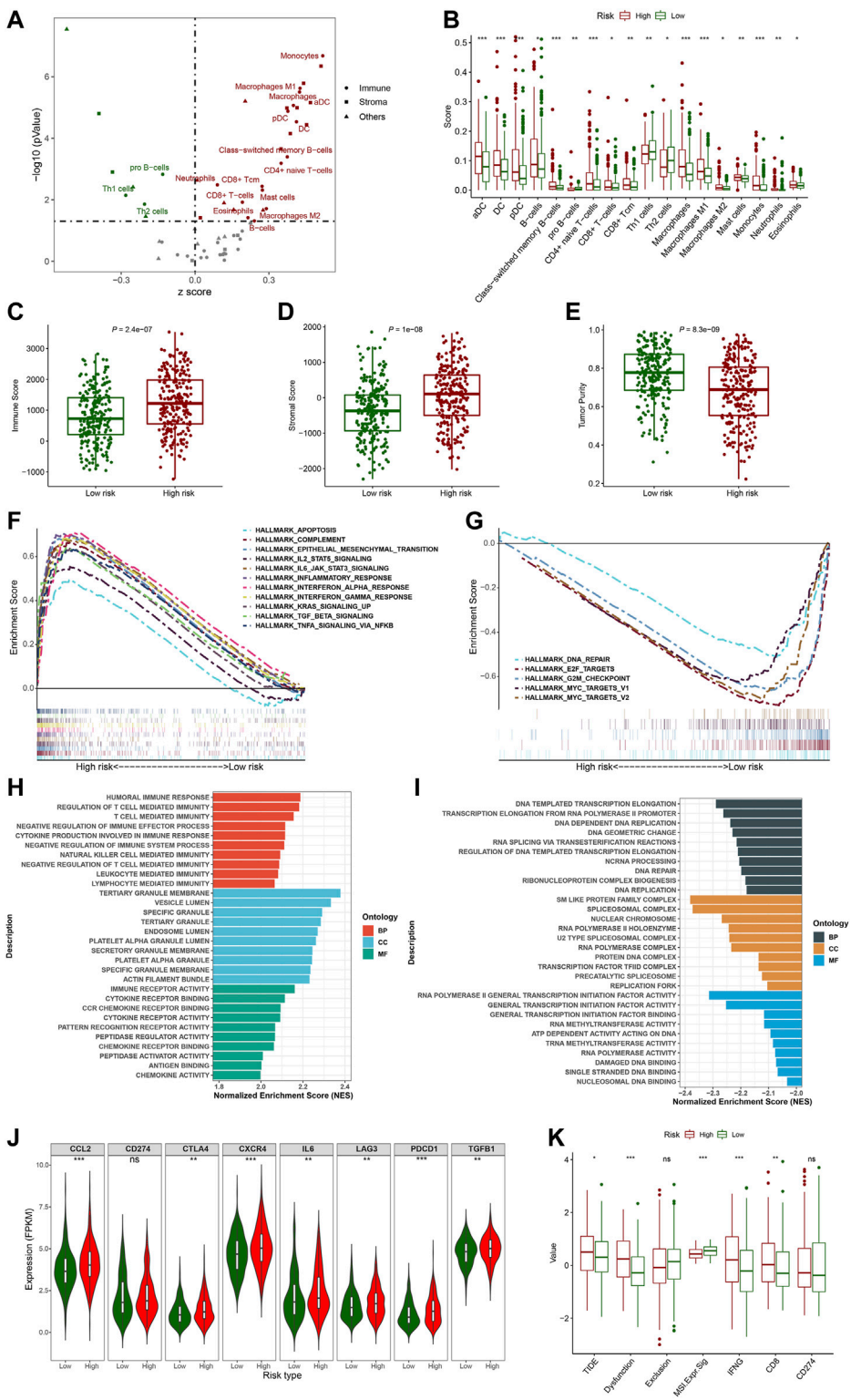


FIGURE 7
Evaluation of the tumor immune microenvironment based on the ferroptosis-related lncRNAs signature stratification; **(A)** The volcano plot depicted the different cellular landscapes of the tumor based on the ferroptosis-related lncRNAs signature (FerRLSig) stratification by the xCell analysis, the infiltration of green dot cells were significantly higher while red dot cells were significantly lower in low-risk group compared to high-risk group. **(B)** The boxplots depicted the significantly different infiltration of immune cell based on the FerRLSig stratification. **(C–E)** The boxplots compared the immune scores **(C)**, stromal scores **(D)** and tumor purity **(E)** based on the FerRLSig stratification by the ESTIMATE analysis. **(F–G)** GSEA
(*Continued*)

FIGURE 7
results demonstrated the differential gene set enrichment in Hallmark with high-risk **(F)** and low-risk **(G)** group. **(H–I)** GSEA results demonstrated the differential gene set enrichment in C5 of biological process (BP), cellular component (CC), and molecular function (MF) in high-risk **(H)** and low-risk **(I)** groups based on the FerRLSig stratification. **(J)** Comparison of the immune checkpoints and immune inhibitory factors, including CCL2, CD274, CTLA4, CXCR4, IL6, LAG3, and PDCD1 and TGFB1based on the FerRLSig stratification in violin plots and boxplots. **(K)** Comparison of TIDE score based on the FerRLSig stratification in boxplots. *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

significantly lower in high-risk group, confirming higher immune cell infiltration in high-risk group (Figures 7C–E). Furthermore, hallmark gene set enrichment analysis was performed to further characterize the different molecular pathways activated by different risk groups. Multiple immune-related hallmarks were significantly enriched in high-risk group, including complement, IL2-STAT5 signaling, IL6-JAK-STAT3 signaling, inflammatory response, interferon-alpha response, interferon-gamma response, TGF-BETA signaling and TNFA signaling *via* NFKB were enriched in high-risk group. On the other hand, low-risk group demonstrated multiple proliferation-related hallmarks and DNA damage repair hallmarks (Figures 7F, G). In addition, Gene Ontology pathway enrichment analysis also corroborated the above finding that high-risk group demonstrated multiple significantly enriched immune-related biological process and molecular function, while low-risk group was correlated with multiple proliferation and transcription ones(Figures 7H, I).

Apparently, the FerRLSig was strongly correlated with tumor immune microenvironment and high-risk group demonstrated increased immune activities compared to low-risk group. We continued to compare the expression level of immune checkpoints and immune inhibitory factors between the two groups to evaluate the potential predictive application of the FerRLSig on the current immune checkpoint blockade therapy. We found that the high-risk group was significantly associated with higher expression level of immune checkpoints and immune inhibitory factors, including CCL2, CTLA4, CXCR4, IL6, LAG3, PDCD1, and TGFB1, indicating tumor immune evasion (Figure 7J). We utilized the TIDE to further evaluate the tumor immune microenvironment of different FerRLSig groups. And surprisingly, although high-risk group consistently demonstrated higher effector T cell signatures including IFNG (interferon gamma) and $CD8^+$ T cell infiltration, the T cell dysfunction signature were significantly higher while microsatellite instability score was significantly lower in high-risk group. This ultimately led to significantly higher overall TIDE score in high-risk group, indicating resistance to immune checkpoint inhibitors (Figure 7K).

Drug sensitivity screening based on the FerRLSig

Besides immune checkpoint inhibitors, we also aimed to evaluate the predictive application of the FerRLSig on other drugs available from GDSC database. Therefore, drug sensitivity screening with GDSC was performed based on the FerRLSig. We identified 119 compounds from GDSC to have statistically significant different half-maximal inhibitory concentration (IC50) based on the FerRLSig stratification (Supplementary Table S1). Notably, low-risk group was significantly more sensitive to platinum and taxane compared to high-risk group (Figures 8A–C), which are the backbones for LUSC chemotherapy, therefore might account for the superior OS of the low-risk group. On the other hand, high-risk group was seemingly more intractable with fewer clinically available systemic therapies compared to low-risk group. Three representative drugs with significantly lower IC50 in high-risk group compared to low-risk group were identified, targeting WNT signaling, MAPK signaling and PI3K signaling pathway (Figures 8D,E).

Discussion

Previous studies have successfully developed prognostic ferroptosis-related lncRNAs signatures in lung adenocarcinoma (Guo et al., 2021; Lu et al., 2021; Zheng et al., 2021). However, the investigation of the ferroptosis-related lncRNAs signature in lung squamous cell carcinoma (LUSC) has been scare and has not evaluated the signature's impact on tumor immune microenvironment yet. In this study, we developed and validated a ferroptosis-related lncRNAs signature (FerRLSig) for the prognosis stratification of lung squamous cell carcinoma (LUSC). High-risk group had significantly worse OS compared to low-risk group (HR = 2.240, 95%CI: 1.845–2.720, $p < 0.001$), which was further corroborated in different age, gender, ethnicities and TNM stages subgroups, indicating the wide applicability and independent prognostic effect of the FerRLSig. And notably, compared to TNM stage, the FerRLSig demonstrated consistently improved AUCs (5-years AUC: FerRLSig 0.756, TNM stage 0.607, Figures 4E,F) on OS. Thus through this study, we have developed and validated a robust prognostic ferroptosis-related lncRNAs signature for LUSC.

A previous retrospective study utilizing TCGA database identified 29 ferroptosis-related lncRNAs with univariate Cox regression and constructed a prognostic ferroptosis-related lncRNAs. The 1-, 2-, and 3-years area under curve (AUC) of their signature were 0.658, 0.693, and 0.687 respectively (Yao et al., 2022). In our study, the TCGA-LUSC patients were

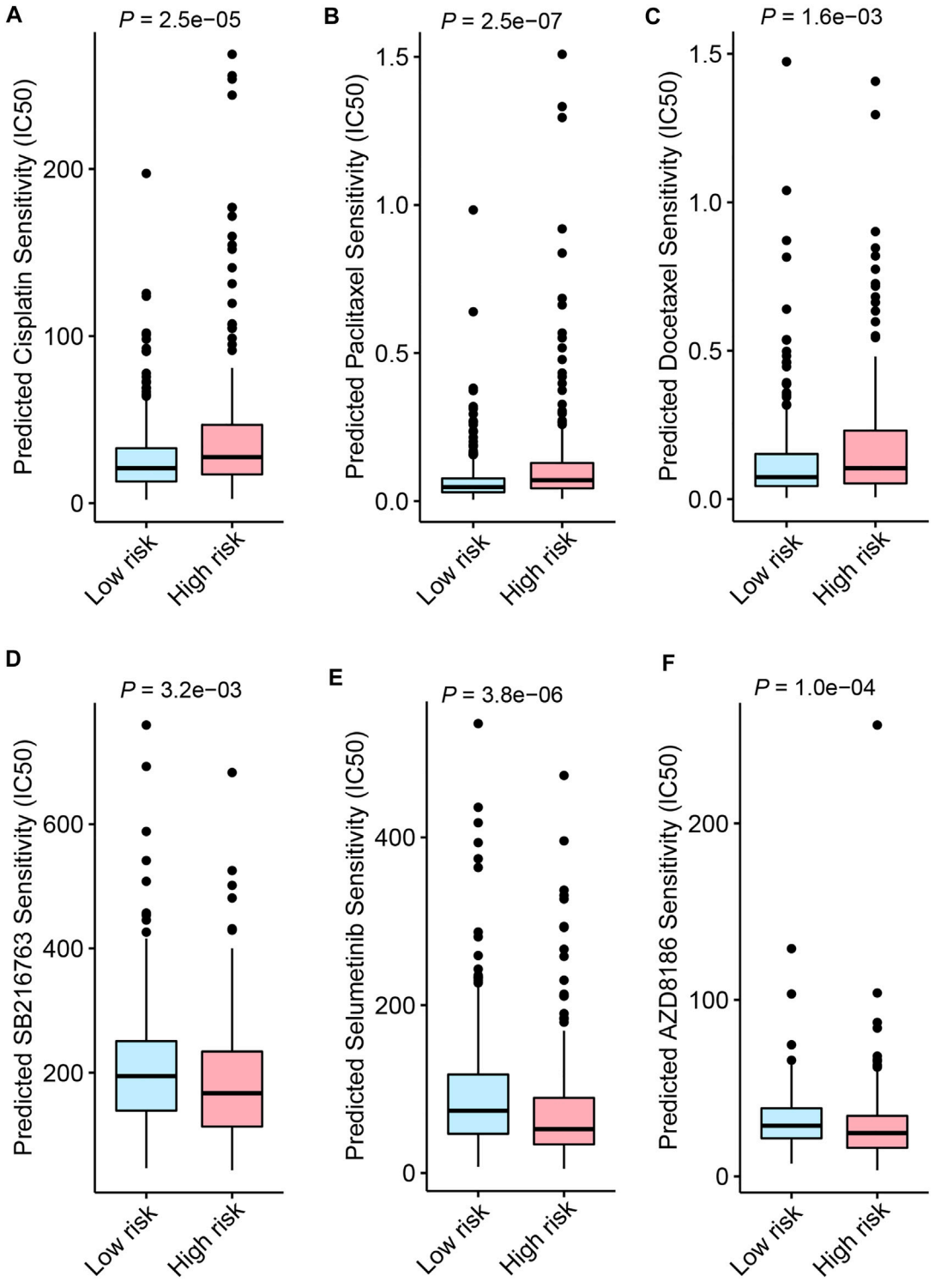


FIGURE 8
Drug sensitivity screening based on the ferroptosis-related lncRNAs signature (FerRLSig) stratification with the Genomics of Drug Sensitivity in Cancer (GDSC) database; **(A)** Boxplot of the cisplatin half-maximal inhibitory concentration (IC50) based on the FerRLSig stratification; **(B)** Boxplot of the paclitaxel IC50 based on the FerRLSig stratification; **(C)** Boxplot of the docetaxel IC50 based on the FerRLSig stratification; **(D)** Boxplot of the SB216763 (targeting WNT signaling pathway) IC50 based on the FerRLSig stratification; **(E)** Boxplot of the Selumetinib (targeting MAPK signaling pathway) IC50 based on the FerRLSig stratification; **(F)** Boxplot of the AZD8186 (targeting PI3K signaling pathway) IC50 based on the FerRLSig stratification.

randomized into a training set and a testing set by the "caret" R package at the ratio of 7:3. The FerRLSig was established with the training set while the validation was performed with the testing set and the entire set. Univariate Cox regression, Least Absolute Shrinkage and Selection Operator (LASSO) regression and multivariate Cox regression were applied in order to establish the final FerRLSig while avoiding overfitting with cross-validation. A more concise ferroptosis-related lncRNAs signature comprising 10 ferroptosis-related lncRNAs with an AUC of 0.756 for 5-years OS was established and validated. Compared to the previous study, we applied more stringent statistical methods to identify ferroptosis-related lncRNAs to avoid overfitting. Moreover, we underwent additional internal validation, which was absent in the previous study. Therefore, we believed our FerRLSig to be more statistically stringent with better prognostic effect compared to the previous study.

More importantly, we have also evaluated the correlation between the FerRLSig and the tumor immune microenvironment. Generally, high-risk group demonstrated significantly higher immune cell infiltration in the xCell analysis, notably by dendritic cells, $CD8^+$ T cells, M1 macrophages and M2 macrophages. On the other hand, low-risk group demonstrated significantly higher pro B cells, Th1 cells and Th2 cells infiltration. In addition, the ESTIMATE and GSEA analysis also corroborated previous results that high-risk group had significantly higher immune and tumor scores with multiple immune-related gene sets enrichment compared to low-risk group, including complement, IL2-STAT5 signaling, IL6-JAK-STAT3 signaling, interferon-alpha response, interferon-gamma response, TGF-BETA signaling and TNFA signaling *via* NFKB. $CD8^+$ cytotoxic T cells and its secreted interferon gamma are central to the tumor immune elimination, and increased immune cell infiltration indicates prominent immune response, but these do not necessarily lead to better tumor control or survival (Gajewski et al., 2013; Mojic et al., 2017; Morad et al., 2021). In addition, multiple immune inhibitory factors seen in high-risk group, including M2 macrophages and IL6, might render the infiltrating $CD8^+$ T cell dysfunction and leading to immune evasion (Xue et al., 2014; Kumari et al., 2016). To further evaluate the mechanism of the immune evasion based on the FerRLSig stratification, several immune checkpoints and immune inhibitory factors including CCL2, CD274 (PD-L1), CTLA4, CXCR4, IL6, LAG3, PDCD1 (PD-1), and TGFB1 were compared between low-risk and high-risk group, and all were significantly higher in high-risk group except PD-L1, strongly suggesting the immune evasion and inhibitory microenvironment in high-risk group.

Immune checkpoint inhibitors have made remarkable breakthroughs in LUSC and significantly improved the LUSC patients' survival (Brahmer et al., 2015; Paz-Ares et al., 2018). Considering the immune evasion and inhibitory microenvironment in high risk group based on the FerRLSig stratification, TIDE score was utilized to estimate the immune checkpoint inhibitor sensitivity. $CD8^+$ T cell infiltration and interferon gamma signature were significantly higher in high-risk group while no statistically significant difference was found on CD274 (PD-L1) signature, which corroborated previous results. And high-risk group was significantly less sensitive to immune checkpoint inhibitor with significantly higher TIDE score and dysfunction score compared to low-risk group. Besides the inherent limitations of the TIDE analysis, one possible explanation would likely be that the T cell dysfunction with multiple alternative immune checkpoints including CTLA-4 and LAG3 within the tumor microenvironment is beyond the salvage of the single-target immune checkpoint inhibitor (Thommen and Schumacher, 2018). Therefore, the trials of combination immunotherapy targeting multiple immune checkpoints and further innovation are needed for the future improvement of LUSC patients. Drug sensitivity screening was also performed based on the FerRLSig with drugs available from GDSC database. Low-risk group was significantly more sensitive to platinum and taxane compared to high-risk group, which might partially account for its better OS. On the other hand, high-risk group were significantly more sensitive to three representative drugs that targeting WNT signaling, MAPK signaling and PI3K signaling pathways. And intriguingly, all three pathways are known to cancer immune evasion (Sumimoto et al., 2006; Dituri et al., 2011; Martin-Orozco et al., 2019) and these kinds of drugs might be combined with immunotherapy to further improve the survival for LUSC patients. However, further investigation is needed to verify these possibilities.

Several limitations are worth mentioning in this study. Firstly, this is a retrospective study from a single database, thus external validation and further prospective study are required. Secondly, several important clinical variables including the extent of resection, resection margin, comorbidities are currently unavailable, which warrant further investigation in future studies. Thirdly, the distance to actual clinical application remains long as whole transcriptome RNA sequencing for lncRNAs identification has not been easily accessible in clinical practice yet. The last but not the least, both *in vitro* and *in vivo* experiments are required to further explore the molecular mechanism underlying the ferroptosis-related lncRNAs signature.

In conclusion, a robust prognostic FerRLSig for LUSC has been developed and validated, demonstrating a close association not only with tumor immune cell infiltration, but also with T cell dysfunction and immune evasion. Further investigation and innovation are required to validate the results from our study and better improve the survival of LUSC patients based on the FerRLSig stratification.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://portal.gdc.cancer.gov/.

Author contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by RZ, XZ, HY, YL, YW, DZ, and LC. The first draft of the manuscript was written by RZ and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

This study was funded by the National Key Research and Development Program of China (Grant number 2021YFC2500905).

Acknowledgments

The results shown here are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

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.2022.968601/full#supplementary-material

References

Aran, D., Hu, Z., and Butte, A. J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 18, 220. doi:10.1186/s13059-017-1349-1

Barta, J. A., Powell, C. A., and Wisnivesky, J. P. (2019). Global epidemiology of lung cancer. *Ann. Glob. Health* 85, 8. doi:10.5334/aogh.2419

Brahmer, J., Reckamp, K. L., Baas, P., Crinò, L., Eberhardt, W. E. E., Poddubskaya, E., et al. (2015). Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N. Engl. J. Med.* 373, 123–135. doi:10.1056/NEJMoa1504627

Cancer Genome Atlas Research Network (2012). Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 489, 519–525. doi:10.1038/nature11404

Chen, X., Kang, R., Kroemer, G., and Tang, D. (2021). Broadening horizons: the role of ferroptosis in cancer. *Nat. Rev. Clin. Oncol.* 18, 280–296. doi:10.1038/s41571-020-00462-0

Dituri, F., Mazzocca, A., Giannelli, G., and Antonaci, S. (2011). PI3K functions in cancer progression, anticancer immunity and immune evasion by tumors. *Clin. Dev. Immunol.* 2011, 947858. doi:10.1155/2011/947858

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

Frankish, A., Diekhans, M., Jungreis, I., Lagarde, J., Loveland, J. E., Mudge, J. M., et al. (2021). GENCODE 2021. *Nucleic Acids Res.* 49, D916–D923. doi:10.1093/nar/gkaa1087

Friedman, J. H., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33, 1–22. doi:10.18637/jss.v033.i01

Fu, J., Li, K., Zhang, W., Wan, C., Zhang, J., Jiang, P., et al. (2020). Large-scale public data reuse to model immunotherapy response and resistance. *Genome Med.* 12, 21. doi:10.1186/s13073-020-0721-z

Gajewski, T. F., Schreiber, H., and Fu, Y-X. (2013). Innate and adaptive immune cells in the tumor microenvironment. *Nat. Immunol.* 14, 1014–1022. doi:10.1038/ni.2703

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

Guo, Y., Qu, Z., Li, D., Bai, F., Xing, J., Ding, Q., et al. (2021). Identification of a prognostic ferroptosis-related lncRNA signature in the tumor microenvironment of lung adenocarcinoma. *Cell Death Discov.* 7, 190. doi:10.1038/s41420-021-00576-z

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell* 144, 646–674. doi:10.1016/j.cell.2011.02.013

Kuhn, M. (2008). Building predictive models in R using the caret package. *J. Stat. Softw.* 28, 1–26. doi:10.18637/jss.v028.i05

Kumari, N., Dwarakanath, B. S., Das, A., and Bhatt, A. N. (2016). Role of interleukin-6 in cancer progression and therapeutic resistance. *Tumour Biol.* 37, 11553–11572. doi:10.1007/s13277-016-5098-7

Lee, J. S., and Ruppin, E. (2019). Multiomics prediction of response rates to therapies to inhibit programmed cell death 1 and programmed cell death 1 ligand 1. *JAMA Oncol.* 5, 1614–1618. doi:10.1001/jamaoncol.2019.2311

Lu, L., Liu, L-P., Zhao, Q-Q., Gui, R., and Zhao, Q-Y. (2021). Identification of a ferroptosis-related LncRNA signature as a novel prognosis model for lung adenocarcinoma. *Front. Oncol.* 11, 675545. doi:10.3389/fonc.2021.675545

Maeser, D., Gruener, R. F., and Huang, R. S. (2021). oncoPredict: an R package for predicting *in vivo* or cancer patient drug response and biomarkers from cell line screening data. *Brief. Bioinform.* 22, bbab260. doi:10.1093/bib/bbab260

Mao, C., Wang, X., Liu, Y., Wang, M., Yan, B., Jiang, Y., et al. (2018). A G3BP1-interacting lncRNA promotes ferroptosis and apoptosis in cancer via nuclear sequestration of p53. *Cancer Res.* 78, 3484–3496. doi:10.1158/0008-5472.CAN-17-3454

Martin-Orozco, E., Sanchez-Fernandez, A., Ortiz-Parra, I., and Ayala-San Nicolas, M. (2019). WNT signaling in tumors: The way to evade drugs and immunity. *Front. Immunol.* 10, 2854. doi:10.3389/fimmu.2019.02854

Mojic, M., Takeda, K., and Hayakawa, Y. (2017). The dark side of IFN-γ: Its role in promoting cancer immunoevasion. *Int. J. Mol. Sci.* 19, E89. doi:10.3390/ijms19010089

Morad, G., Helmink, B. A., Sharma, P., and Wargo, J. A. (2021). Hallmarks of response, resistance, and toxicity to immune checkpoint blockade. *Cell* 184, 5309–5337. doi:10.1016/j.cell.2021.09.020

Paz-Ares, L., Luft, A., Vicente, D., Tafreshi, A., Gümüş, M., Mazières, J., et al. (2018). Pembrolizumab plus chemotherapy for squamous non–small-cell lung cancer. *N. Engl. J. Med.* 379, 2040–2051. doi:10.1056/NEJMoa1810865

Ramalingam, S. S., Vansteenkiste, J., Planchard, D., Cho, B. C., Gray, J. E., Ohe, Y., et al. (2020). Overall survival with osimertinib in untreated, EGFR-mutated advanced NSCLC. *N. Engl. J. Med.* 382, 41–50. doi:10.1056/NEJMoa1913662

Shaw, A. T., Bauer, T. M., de Marinis, F., Felip, E., Goto, Y., Liu, G., et al. (2020). First-line lorlatinib or crizotinib in advanced ALK-positive lung cancer. *N. Engl. J. Med.* 383, 2018–2029. doi:10.1056/NEJMoa2027187

Solomon, B. J., Mok, T., Kim, D-W., Wu, Y-L., Nakagawa, K., Mekhail, T., et al. (2014). First-line crizotinib versus chemotherapy in ALK-positive lung cancer. *N. Engl. J. Med.* 371, 2167–2177. doi:10.1056/NEJMoa1408440

Sordella, R., Bell, D. W., Haber, D. A., and Settleman, J. (2004). Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways. *Science* 305, 1163–1167. doi:10.1126/science.1101637

Statello, L., Guo, C-J., Chen, L-L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* 22, 96–118. doi:10.1038/s41580-020-00315-9

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102, 15545–15550. doi:10.1073/pnas.0506580102

Sumimoto, H., Imabayashi, F., Iwata, T., and Kawakami, Y. (2006). The BRAF-MAPK signaling pathway is essential for cancer-immune evasion in human melanoma cells. *J. Exp. Med.* 203, 1651–1656. doi:10.1084/jem.20051848

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

Thommen, D. S., and Schumacher, T. N. T. (2018). T cell dysfunction in cancer. *Cancer Cell* 33, 547–562. doi:10.1016/j.ccell.2018.03.012

Vickers, A. J., and Elkin, E. B. (2006). Decision curve analysis: a novel method for evaluating prediction models. *Med. Decis. Mak.* 26, 565–574. doi:10.1177/0272989X06295361

Wang, M., Mao, C., Ouyang, L., Liu, Y., Lai, W., Liu, N., et al. (2019). Long noncoding RNA LINC00336 inhibits ferroptosis in lung cancer by functioning as a competing endogenous RNA. *Cell Death Differ.* 26, 2329–2343. doi:10.1038/s41418-019-0304-y

Wu, Y-L., Tsuboi, M., He, J., John, T., Grohe, C., Majem, M., et al. (2020). Osimertinib in resected EGFR-mutated non-small-cell lung cancer. *N. Engl. J. Med.* 383, 1711–1723. doi:10.1056/NEJMoa2027071

Xue, J., Schmidt, S. V., Sander, J., Draffehn, A., Krebs, W., Quester, I., et al. (2014). Transcriptome-based Network analysis reveals a spectrum model of human macrophage activation. *Immunity* 40, 274–288. doi:10.1016/j.immuni.2014.01.006

Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., et al. (2013). Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* 41, D955–D961. doi:10.1093/nar/gks1111

Yao, N., Zuo, L., Yan, X., Qian, J., Sun, J., Xu, H., et al. (2022). Systematic analysis of ferroptosis-related long non-coding RNA predicting prognosis in patients with lung squamous cell carcinoma. *Transl. Lung Cancer Res.* 11, 632–646. doi:10.21037/tlcr-22-224

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612

Zhang, X., Wang, L., Li, H., Zhang, L., Zheng, X., and Cheng, W. (2020). Crosstalk between noncoding RNAs and ferroptosis: New dawn for overcoming cancer progression. *Cell Death Dis.* 11, 580. doi:10.1038/s41419-020-02772-8

Zheng, Z., Zhang, Q., Wu, W., Xue, Y., Liu, S., Chen, Q., et al. (2021). Identification and validation of a ferroptosis-related long non-coding RNA signature for predicting the outcome of lung adenocarcinoma. *Front. Genet.* 12, 690509. doi:10.3389/fgene.2021.690509

Zhou, N., and FerrDb, Bao J. (2020). FerrDb: a manually curated resource for regulators and markers of ferroptosis and ferroptosis-disease associations. *Database.* 2020, baaa021. doi:10.1093/database/baaa021

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 29 August 2022
doi: 10.3389/fgene.2022.905353

Identification of a 5-lncRNA-Based Signature for Immune Characteristics and Prognosis of Lung Squamous Cell Carcinoma and Verification of the Function of lncRNA SPATA41

Sheng Huan[1,2†], *Miao Chen*[1,3†], *Sumin Sun*[4], *Yanling Zhong*[2], *Yu Chen*[1,5], *Yihao Ji*[1,6] *and Guoping Yin*[1,2*]

[1]Nanjing Hospital Affiliated to Nanjing University of Chinese Medicine, Nanjing, China, [2]Department of Anesthesiology, Nanjing Second Hospital, Nanjing, China, [3]Department of Surgery, Nanjing Second Hospital, Nanjing, China, [4]Nanjing University of Chinese Medicine, Nanjing, China, [5]Nanjing Medical University, Nanjing, China, [6]Department of Critical Medicine, Nanjing Second Hospital, Nanjing, China

OPEN ACCESS

Edited by:
Wei Li,
Lanzhou University Second Hospital, China

Reviewed by:
Yunpeng Cao,
Chinese Academy of Sciences (CAS), China
Tao Yang,
Guizhou University of Traditional Chinese Medicine, China

***Correspondence:**
Guoping Yin
yinguoping0304@hotmail.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *27 March 2022*
Accepted: *15 June 2022*
Published: *29 August 2022*

Citation:
Huan S, Chen M, Sun S, Zhong Y, Chen Y, Ji Y and Yin G (2022) Identification of a 5-lncRNA-Based Signature for Immune Characteristics and Prognosis of Lung Squamous Cell Carcinoma and Verification of the Function of lncRNA SPATA41. Front. Genet. 13:905353. doi: 10.3389/fgene.2022.905353

Lung squamous cell carcinoma (LUSC) is one of the most lethal cancers worldwide. Traditional tumor-node-metastasis (TNM) staging system has many insufficiencies in predicting immune characteristics, overall survival (OS), and prognosis of LUSC. LncRNA is currently found involved in tumor development and effectively predicts tumor prognosis. We screened potential tumor-related lncRNAs for immune characteristics and constructed a nomogram combining lncRNA and traditional clinical indicators for prognosis prediction. We obtained the large-scale gene expression profiles of samples from 492 LUSC patients in The Cancer Genome Atlas database. SPATA41, AL034550.2, AP003721.2, AC106786.1, and AC078889.1 were finally screened to construct a 5-lncRNA-based signature. The risk score of the signature divided patients into subgroups of high-risk and low-risk with significant differences in OS. Their area under the curve (AUC) reached more than 0.70 in 1, 3, and 5 years. In addition, compared with the high-risk subgroup, the low-risk subgroup exhibited a remarkably favorable prognosis and TME score, along with a higher immune infiltration score and lower TIDE score. The signature also significantly related to chemotherapy response, especially in cisplatin, vinorelbine, and paclitaxel. Importantly, the nomogram we constructed had good reliability with the assessment of the calibration chart and consistency index (c-index). GO and KEGG enrichment analysis indicated that co-expression mRNAs of the 5 lncRNAs were mainly focused on RNA splicing, DNA replication, and protein serine/threonine kinase activity. Functional assays demonstrated that SPATA41, one of the five OS-related lncRNAs, regulated invasion, migration, proliferation, and programmed death *in vitro*. In summary, our 5-lncRNA-based signature has a good performance in predicting immune characteristics and prognosis of LUSC patients.

Keywords: lncRNA, lung squamous cell carcinoma, nomogram, prognosis, SPATA41

BACKGROUND

Lung cancer is the most common chest tumor in the world, with a high incidence and mortality. More than 2.2 million people in the world have been newly diagnosed with lung cancer in 2021, accounting for 11.1% of all newly diagnosed cancer (Bu et al., 2020; Xu et al., 2020). Lung squamous cell carcinoma (LUSC), deriving from chronic stimulation and injury of bronchial columnar epithelial cells, takes up a large percentage of lung cancer (Sánchez Danés and Blanpain, 2018). Although LUSC grows slowly and metastases late, it is less sensitive to radiotherapy than undifferentiated lung carcinoma. In addition, LUSC lacks effective chemotherapeutic drugs compared with lung adenocarcinoma (LUAD) and has a poor prognosis. Presently, the best treatment for LUSC is still surgical resection, which has been demonstrated to have a good 5-year overall survival (OS) rate (Bozinovski et al., 2016; Socinski et al., 2018). Therefore, studies related to LUSC immune checkpoint blockade (ICB) therapy were urgent to perform for addressing the gap in this research field.

Tumor-node-metastasis (TNM) staging was often used to evaluate tumor development and prognosis in the past. However, many clinicians found it was not accurate enough in diagnosing LUSC due to complex disease pathology and high heterogeneity between patients (Lian et al., 2020). Recent research showed that lncRNA could regulate gene expression by influencing mRNA transcription, binding nucleic acids, and participating in posttranscriptional modification, such as DNA methylation and acetylation (Quinn et al., 2014; Engreitz et al., 2016; Storti et al., 2020). LncRNAs had been found to have great importance in the diagnosis, treatment, and prognosis of LUSC.

Tumor microenvironment (TME) plays a critical role in neoplasia and tumor development. Immune cells and immune stromal cells are the core components of TME, regulating tumor differentiation, proliferation, and metastasis. For example, a continuous abundance of specific T cell subtypes in tumors contributed to a better prognosis of patients (Jiang et al., 2018). Macrophage polarization was of great importance in subverting adaptive immunity and inducing tumor metastasis (Mantovani et al., 2002). Identifying the useful lncRNAs that influence the immune cells and immune stromal cells contributes to deciphering the carcinogenic mechanism of lncRNA.

Although previous studies have predicted and identified some molecular biomarkers in LUSC patients, most of them may have limited research meaning due to small sample sizes, differences in platforms, or a lack of combining diverse variables (Friedlaender et al., 2019; Li and Guo, 2020; Ma et al., 2020; Ren et al., 2020). For these reasons, we conducted this study to integrate relevant data and identify credible prognostic biomarkers for clinical guidance.

MATERIALS AND METHODS

Workflow and Dataset Processing

The flowchart was shown in **Figure 1**. We downloaded lncRNA expression data, mRNA expression data, and corresponding clinical data of LUSC from The Cancer Genome Atlas (TCGA, HYPERLINK "http://cancergenome.nih.gov/") (Kourou et al., 2015; Gibson et al., 2016; Rokavec et al., 2017).

Screening of Differentially Expressed lncRNAs and mRNAs

Differentially expressed lncRNAs (DELs) and differentially expressed mRNAs (DEMs) between LUSC samples and control samples were analyzed with the "DESeq2" package in R software. Then, we used "ggplot2" and "pheatmap" packages to draw volcanic and thermal maps of them (Robinson et al., 2010).

Identification of lncRNA Signature and Calculation of Risk score

We constructed primary dataset (n = 484) and entire dataset (n = 239) to conduct data processing. Specifically, we first used univariate Cox proportional hazard regression (CPHR) to screen the lncRNAs significantly associated with the OS of LUSC patients, making them candidates (p = 0.01) (Tibshirani, 1997). Then we continued with multivariate CPHR analysis (stepwise model) to find the most suitable OS-related lncRNAs from the aforementioned candidate lncRNAs. Akaike information criterion (AIC) was used in order to avoid overfitting. We used "pheatmap" and "survminer" packages to draw thermal maps and survival probability curves.

Sankey Diagram

We drew a Sankey diagram based on life-status, gender, TNM staging, and 5-lncRNA risk score with the "ggalluvial" package (Grazziotin et al., 2022).

Nomogram and ROC Curves for Predicting OS

We drew ROC curves for predicting OS in the primary dataset and the entire dataset with the "survivalROC" package (Zhou et al., 2020).

Generation of ImmuneScore, StromalScore, and ESTIMATEScore

The ESTIMATE algorithm was used to estimate the ratio of the immune-stromal components in TME for each sample with the "ggpubr" and "estimate" packages, exhibited in the form of three kinds of scores: Immune Score, StromalScore, and ESTIMATEScore, which positively correlated with the ratio of immune, stromal, and the sum of both, respectively (Yoshihara et al., 2013).

Nomogram Construction and Reliability Assessment

We used the "RMS" package to draw the nomogram of the primary dataset and the entire dataset (Zhang and Kattan, 2017). The calibration chart and c-index were conducted to evaluate their value in calibrating and identifying the OS of LUSC.

GSVA and Chemotherapeutic Response Prediction

We performed gene set variation analysis (GSVA) to evaluate the related immune cell and molecular pathway variation

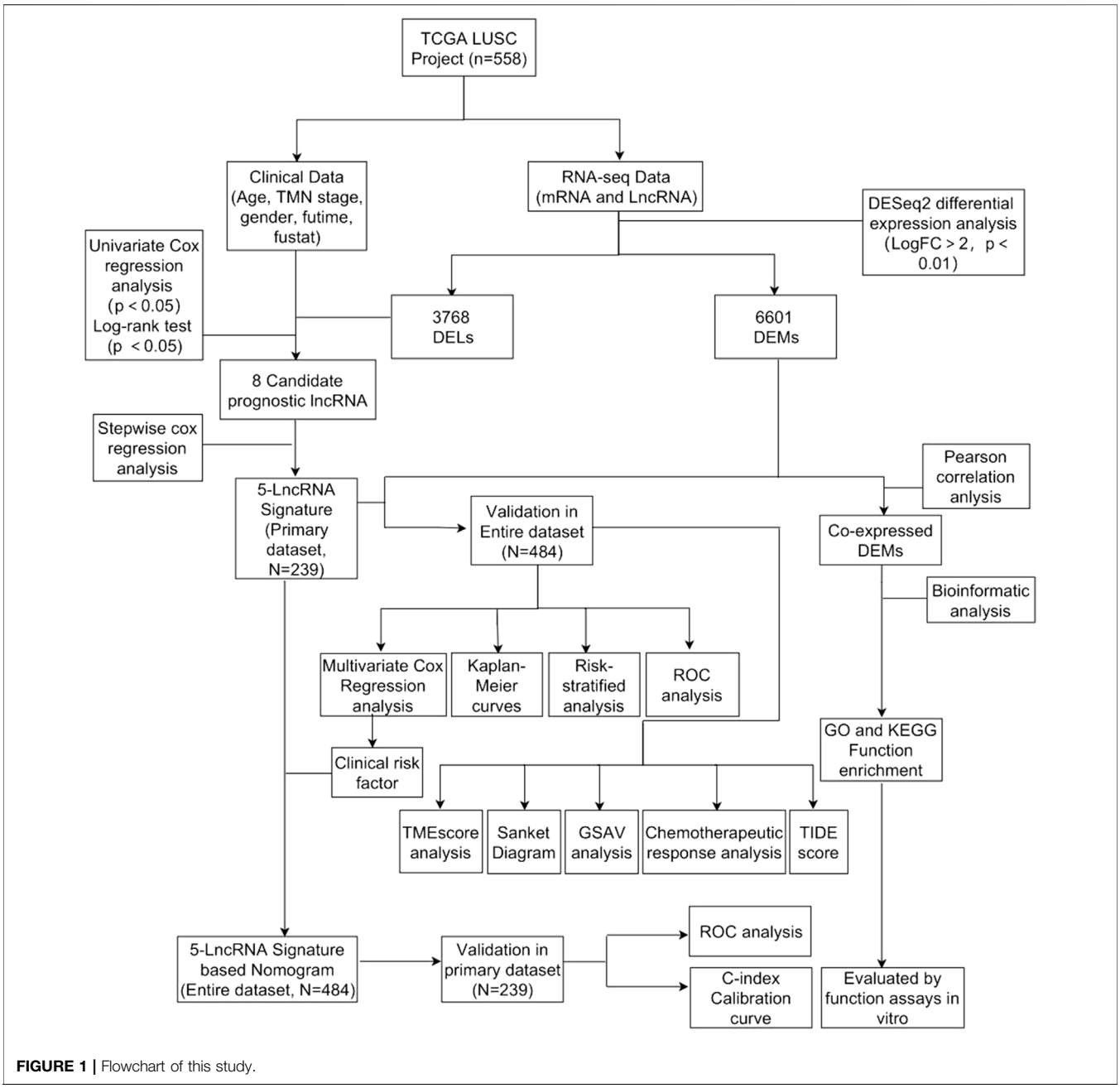


FIGURE 1 | Flowchart of this study.

between the high-risk and low-risk groups with the "GSVA" package (Subramanian et al., 2005). Then, we conducted chemotherapeutic response prediction for LUSC based on our expression profiles with the "pRRophetic" and "ggExtra" packages. Five commonly used drugs were selected, namely, docetaxel, vinorelbine, cisplatin, paclitaxel, and gemcitabine (Reinhold et al., 2012). Moreover, tumor immune dysfunction and exclusion (TIDE) was also applied to predict the clinical response to immunotherapy of LUSC patients with the "ggpubr" package (Jiang et al., 2018).

Enrichment Analysis of Co-related mRNA of the 5-lncRNA Signature

We used the "limma" package to select the mRNA co-expressed with these 5-lncRNA signatures (Pearson

correlation coefficient >0.25). Then we used the "enrichplot" package to perform Gene ontology (GO) functional annotation and Kyoto Encyclopedia of Gene and Genomes (KEGG) functional annotation to verify the biological process and potential pathway (Yu et al., 2012). We also used the GSVA enrichment analysis to investigate the underlying biological activities between the high-risk subgroups and low-risk subgroups (Subramanian et al., 2005).

Box Plots and Differential Gene Comparison

The microarray data of normal tissue and tumor tissue were downloaded from the GEO database (GSE73402) (http://www.ncbi.nih1.gov/geo). The raw data were downloaded as MINiML files. Box plots are drawn with the package "boxplot" (Zhou et al., 2020). Differential gene comparison was performed with Kruskal–Wallis rank sum test.

Cell Culture and siRNA Transfection

BEAS-2B and H520 cells were cultured in RPMI-1640 medium (GIBCO, Shanghai, China), and SK-MES-1 cells were cultured in MEM medium (GIBCO, Shanghai, China). All media were supplemented with 10% fetal bovine serum (FBS) (GIBCO, Shanghai, China) and 1% penicillin and streptomycin (Solarbio, Beijing, China). The plasmids of SPATA41 siRNA were designed by Kaiji Biological Company (Nanjing, China) and transfected into H520 and SK-MES-1 cells with Lipofectamine 8000 according to the manufacturer's instructions.

RNA Extraction and Quantitative Real-Time PCR

The cells were lysed by Trizol. Total RNA was collected with isopropanol from Aladdin (Shanghai, China) and RNA content was determined by an enzyme labeling instrument. Total RNA was reverse-transcribed into cDNA with Hifair® II 1st Strand cDNA Synthesis Kit from Yeasen (Shanghai, China). The cDNA was amplified with Hieff Q real-time PCR SYBR Green main mixture from Yeasen (Shanghai, China) and detected through Applied Biosystems 7500 from Thermo Fisher (Massachusetts, USA) according to the manufacturer's instructions. The expression of the target gene was normalized to GADPH expression.

Transwell Assay

The intervened cells (1×10^5 cells/well) were seeded into the upper chamber of the transwell culture plate while 500 ul medium with 20% FBS was filled with the bottom chamber. After incubation for 24 h at 37 °C in 5% CO_2, the cells were fixed in cooled methanol for 30 min and were stained with hematoxylin for 1 min, and washed with 1 X PBS. The representative images were obtained under an inverted light microscope.

Wound-Healing Assay

The intervened cells (3×10^5 cells/well) were seeded into a 6-well microplate until forming a fusion monolayer. We used the 200 μL sterile pipettes to make an artificial and uniform wound and carefully cleaned the unattached cells with 1 x PBS. Then the cells were incubated in a serum-free medium at 37 °C in 5% CO_2. The representative images of 0 and 24 h were obtained under an inverted light microscope.

Western Blotting

The total protein concentration was determined using the BCA protein analysis kit (Pierce, #23225). The same amount of proteins was added and separated by SDS-PAGE. Then it was transferred to PVDF membrane (microporous, IPVH00010) and blocked with 5% BSA for 1 h. The PVDF membranes were incubated with the primary antibody overnight and the secondary antibody for 2 h. Finally, the PVDF membranes were incubated with ECL (Pierce, #32109) and the strip strength was quantified by ImageJ software.

Colony Formation Assay

The cells (1×10^5 cells) were seeded into the culture chamber and the culture medium was covered. After incubation for 24 h at 37°C in 5% CO_2, the cells were transfected with siRNA and then cultured for 7 days. Cell colonies suspended gradually formed in the medium on the matrix gel. The images were obtained with a common camera.

Cell Counting Kit-8 Assay

The cells (5×10^4 cells/well) were seeded into 96-well microplates and incubated for 24 h at 37°C in 5% CO_2. After the intervention, cells were treated with 10 μL CCK8 reagents and incubated at 37°C for 4 h. The optical density was determined at 450 nm in a plate reader from Thermo Fisher (Massachusetts, USA).

Annexin V-FITC/PI Double Staining Assay

The intervened cells were lysed, collected, and dissolved in buffer at the density of 1.0×10^5 cells/mL. 100 μL sample solution was added with 5 μL FITC-conjugated Annexin V reagent and 5 μL Propidium iodide (PI) reagent and incubated for 15 min in a dark greenhouse. Percentages of cells within each cell death compartment (Q1, Q3, Q3, and Q4) were determined by flow cytometry. The results were analyzed by FlowJo software.

Statistical Analysis

Our data were analyzed with the deviation of mean and standard. Results related to mapping were analyzed with GraphPad Prism 7.0. Statistical significance was determined using an unpaired Student's t test for comparisons between two groups, while one-way ANOVA followed Tukey's post hoc test was used for comparisons among more than two groups. The difference was considered significant when the p value was less than 0.05. Each experiment was repeated at least three times.

RESULTS

Construction and Assessment of a 5-lncRNA Signature in Predicting Prognosis of LUSC Patients

In total, 510 LUSC tissue samples and 48 normal tissue samples were eventually included. After analyzing their lncRNA and

TABLE 1 | 5 lncRNAs significantly associated with the OS of 492 LUSC patients.

Gene Name	Coefficient	Type	Down/up-Regulated	HR	95%CI	P Value
SPATA41	0.535264763	Risky	Up	1.707900371	1.239-2.354	0.001080441
AC106786.1	0.337590814	Risky	Up	1.401566884	1.067-1.841	0.015266432
AL034550.2	−0.857359527	Protect	Down	0.424280907	0.231-0.778	0.005590659
AP003721.2	−0.617768427	Protect	Down	0.53914624	0.353-0.822	0.004127988
AC078889.1	−0.692738229	Protect	Down	0.500204517	0.334-0.749	0.000781794

mRNA expression profiles, we found that there were 6601 DEMs and 3768 DELs. Their volcanic and thermal maps were shown as **Supplementary Figures S1, S2**. Then, we combined clinical data with lncRNA expression and received data of 505 LUSC patients. After excluding 1 patient without survival time and 11 patients with insufficient survival data, relevant data of 492 patients were retained. 8 lncRNA were found to have a significant correlation with OS in LUSC patients ($p < 0.05$) (**Supplementary Table S1**). Kaplan–Meier curve was performed in accordance with the result of the univariate CPHR analysis (**Supplementary Figure S3**).

In order to further find the most suitable OS-related lncRNA, we continued to conduct multivariate CPHR. The result showed that five lncRNAs (SPATA41, AL034550.2, AP003721.2, AC106786.1, and AC078889.1) had the lowest AIC value and highest likelihood ratios (**Table 1**). The formula was shown as follows: Risk Score = (0.535 × Expression SPATA41) + (0.338 × Expression AC106786.1) + (−0.857 × Expression AL034550.2) + (−0.618 × Expression AP003721.2) + (−0.692 × Expression AC078889.1).

8 patients without TNM staging or age data were excluded. Within the remaining 484 LUSC patients, 239 patients were randomly classified into the subgroup of "primary dataset" and all 484 patients was assigned to the subgroup of "entire dataset". The characteristics of these 484 patients in the entire dataset and 239 patients in the primary dataset are shown in **Table 2**. The five lncRNA expression profile, OS status, and risk score distribution of the primary dataset and the entire dataset were presented in **Figures 2A–F**. Patients were equally divided into a subgroup of high-risk and low-risk in accordance with median risk score. Kaplan–Meier curve also showed that patients in a subgroup of high-risk had a worse prognosis than that of low-risk in both the entire data set (p = 5.935e-11) and the primary dataset (p = 6.5790e-09) (**Figures 2G, H**).

TABLE 2 | Baseline clinical characteristics of LUSC patients involved in this study.

Characteristic	Primary Dataset	Entire Dataset	*p* value
	n = 239	*n* = 484	
Age (years)	—	—	0.917534
≥65	157 (65.69%)	317(65.50%)	—
<65	82 (34.31%)	167 (34.50%)	—
Gender	—	—	0.642574
Female	59 (24.69%)	127 (26.24%)	—
Male	180 (75.31%)	357 (73.76%)	—
TNM stage	—	—	0.987752
I	115 (48.12%)	238 (49.17%)	—
II	82 (34.31%)	157 (32.44%)	—
III	39 (16.32%)	82 (16.94%)	—
IV	3 (1.26%)	7 (1.45%)	—
Tumor stage	—	—	0.719912
T0-T2	197 (82.43%)	393 (81.20%)	—
T3-T4	42 (17.57%)	91 (18.80%)	—
Lymph node metastasis	—	—	0.944372
Nx	2 (0.84%)	5 (1.03%)	—
No	154 (64.44%)	309 (63.84%)	—
Yes	83 (34.73%)	170 (35.12%)	—
Distant metastasis	—	—	0.830401
Mx	37 (15.48%)	77 (15.91%)	—
No	199 (83.26%)	400 (82.64%)	—
Yes	3 (1.26%)	7 (1.45%)	—

We constructed a receiver operating characteristic (ROC) curve based on the primary dataset and the entire dataset. As shown in **Figures 2I, J**, the area under the curve (AUC) of the 5-lncRNA signature reached 0.714 at 1 year, 0.789 at 3 years, and 0.810 at 5 years in the primary dataset while 0.665 at 1 year, 0.703 at 3 years, and 0.709 at 5 years in the entire dataset. The result of univariate and multivariate CPHR analysis in the entire dataset showed that the HR of the high-risk subgroup versus low-risk subgroup was 1.142 ($p < 0.001$, 95% CI = 1.082-1.205) and 1.152 ($p < 0.001$, 95% CI = 1.094-1.213), indicating that 5-lncRNA signature was independent of traditional clinical risk factors in predicting the prognosis of LUSC patients (**Table 3**). The univariate and multivariate CPHR analysis in the primary dataset showed consistent results (**Supplementary Table S2**). Risk stratification analysis was also performed on the entire dataset in consideration of the big sample of LUSC patients. We conducted a hierarchical analysis based on gender (female or male), age (≥65 or<65), TNM staging (I, II, III, or IV), T staging (T1, T2, T3, and T4), N staging (N0, N1, and N2), and M staging (M0 or M1). Each group was further assigned into subgroups of high-risk group and low-risk group in accordance with the risk score. As shown in the Kaplan–Meier curve, except for the condition of TNM III, TNM IV, T1, T4, N2, and M1, patients in the other conditions had a worse prognosis with higher risk score (**Supplementary Figure S4**).

Assessment of the TME Scores, Immune Characteristics, and Drug Sensitivity Related to the 5-lncRNA-Based Signature.

For determining the relationship between the TME Scores with traditional TNM staging and 5-lncRNA-based signature, we analyzed the corresponding clinicopathological information. As shown in **Figure 3A**, there is no significant difference in TME Scores among TNM staging, T staging, M staging, and N staging. In contrast, the low-risk subgroup showed higher StromalScore,

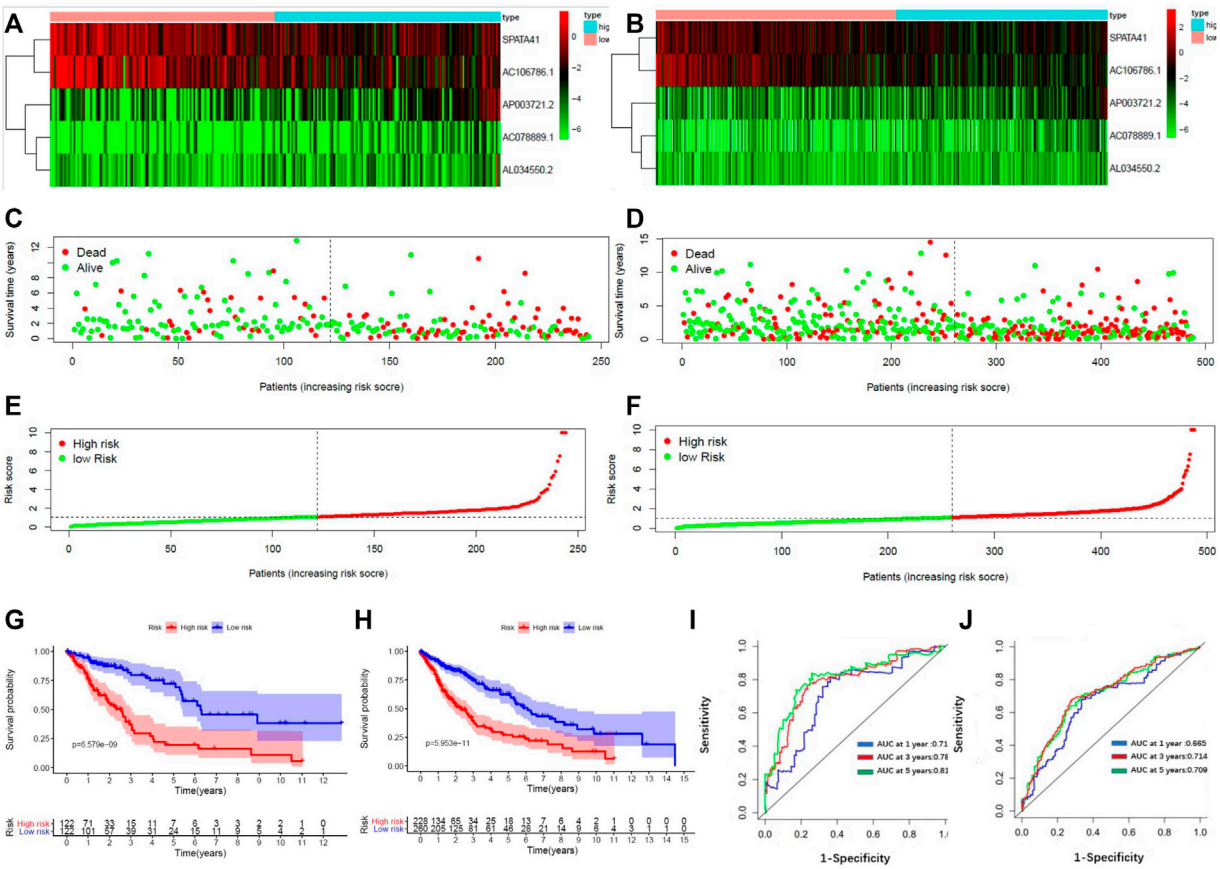


FIGURE 2 | Construction of a 5-lncRNA signature in predicting prognosis of LUSC patients in the primary dataset and entire dataset. **(A)** The heatmap of 5-lncRNA signature in LUSC patients of the primary dataset. **(B)** The OS status of 5-lncRNA signature in LUSC patients of the primary dataset. **(C)** The risks score distribution of 5-lncRNA signature in LUSC patients of primary dataset. **(D)** The heatmap of 5-lncRNA signature in LUSC patients of the entire dataset. **(E)** The OS status of 5-lncRNA signature in LUSC patients of the entire dataset. **(F)** The risks score distribution of 5-lncRNA signature in LUSC patients of the entire dataset. **(G)** Kaplan–Meier curves are based on the 5-lncRNA signature of LUSC patients in the primary dataset. **(H)** Kaplan–Meier curves are based on the 5-lncRNA signature of LUSC patients in the entire dataset. **(I)** Time-dependent ROC curve based on 5-lncRNA signature of LUSC patients in the primary dataset. **(J)** Time-dependent ROC curve based on 5-lncRNA signature of LUSC patients in the entire dataset.

TABLE 3 | Univariate and multivariate Cox proportional hazards regression analysis of 5-lncRNA signature and clinical risk factors in the entire dataset.

Characteristic	Univariate analysis		Multivariate analysis	
	HR (95%CI)	*p*-Value	HR (95%CI)	*p*-Value
Age (≥65 *vs*. <65)	1.415 (0.996-2.010)	0.053	1.547 (1.079-2.219)	0.018
Gender (male *vs*. female)	1.256 (0.858-1.838)	0.241	1.271 (0.865-1.867)	0.222
TNM stage (III-IV *vs*. I-II)	1.248 (1.035-1.506)	0.020	1.128 (0.821-1.550)	0.458
Tumor stage (T3-T4 *vs*. T0-T2)	1.268 (1.033-1.557)	0.023	1.268 (0.932-1.600)	0.148
Lymph node metastasis (yes *vs*. no)	1.432 (0.895-2.291)	0.134	1.344 (0.727-2.486)	0.345
Distant metastasis (yes *vs*. no)	1.837 (0.583-5.784)	0.299	1.161 (0.307-4.387)	0.826
Risk score (high *vs*. low)	1.142 (1.082-1.205)	< 0.001	1.152 (1.094-1.213)	< 0.001

ImmuneScore, and ESTIMATEScore in comparison with the high-risk subgroup divided by 5-lncRNA-based signature (**Figure 3B**). However, there seemed to be no significant difference in 5-lncRNA based risk score among patients with different TNM staging, T staging, M staging, and N staging (**Figure 3C**). Sankey diagram was used to visualize the relationship between life status, gender, TNM staging, and risk score. The results showed that male patients had a more terrible life status and patients in the high-risk subgroup tended to have worse TNM staging (**Figure 3D**).

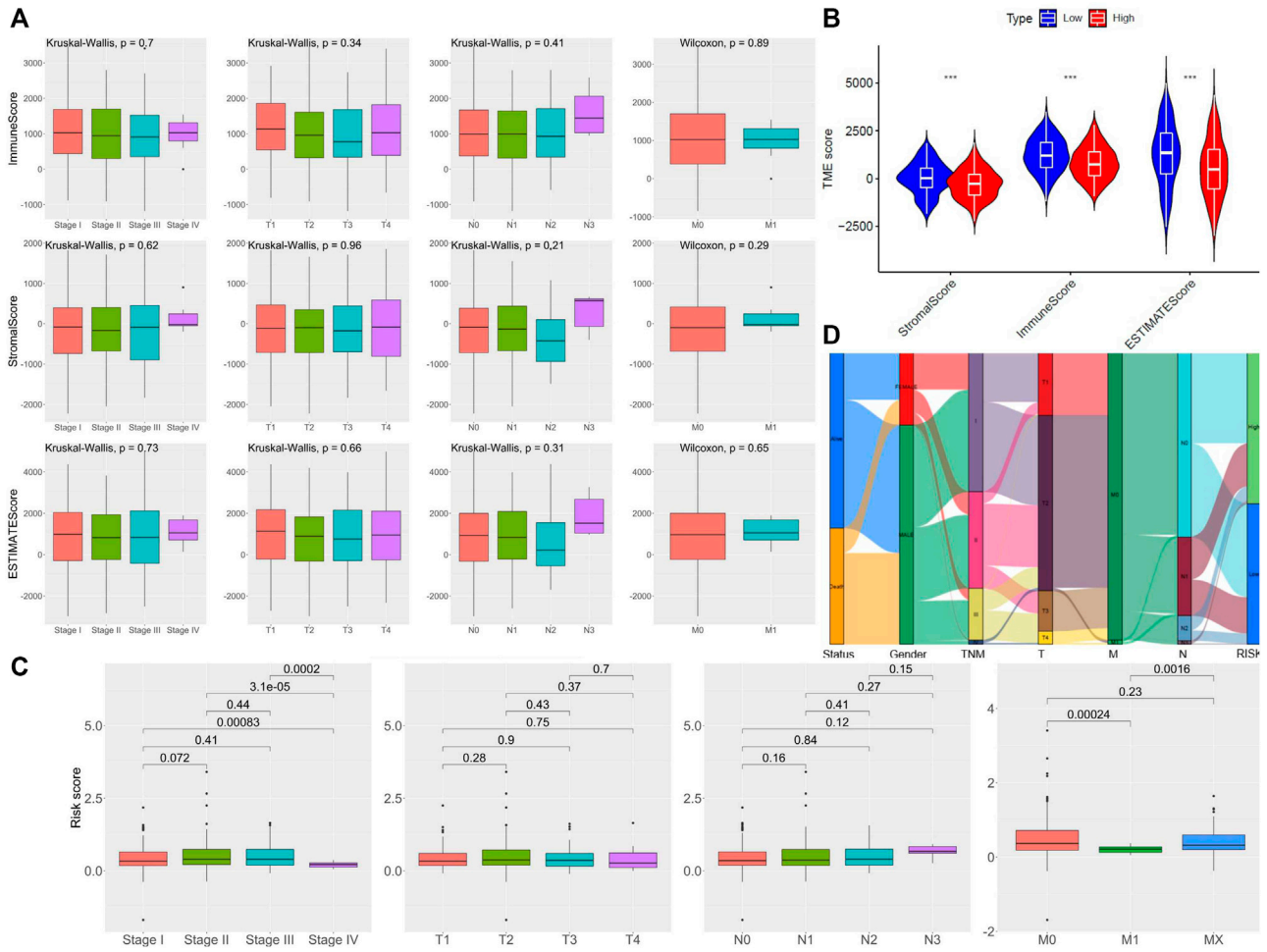


FIGURE 3 | Assessment of the TME Scores among TNM staging and 5-lncRNA-based signature. **(A)** Comparison of ImmuneScore and StromalScore in different clinicopathological staging characteristics. **(B)** Comparison of risk score of LUSC patients of different TNM staging, T staging, N staging, and M staging. **(C)** Comparison of 5 lncRNA-based riskscore between LUSC patients of different TNM staging. **(D)** Alluvial diagram of subgroups based on 5-lncRNA signature with different life status, gender, and TNM staging.

Due to the close relationship between tumor prognosis and immunity, we further performed GSVA immune infiltration analysis with the enrichment scores of 16 types of immune cells and 13 types of immune-related pathways between the high-risk subgroup and low-risk subgroup. As the results show, the high-risk group expressed a low level of immune cells in the mass. Except for the aDCs and mast cells, significant differences appeared in the other 11 immune cells, especially in tumor-infiltrating lymphocytes (TILs), T helper (Th) cells (Th1 and Th2 cells), pDCs, natural killer (NK) cells, neutrophils, and $CD8^+$ T cells (**Figure 4A**). In terms of the immune pathway, all of them had lower activation in the high-risk group than in the low-risk group, and only the pathway of APC co-inhibition did not show a significant difference between groups (**Figure 4B**). Then, we conformed to the tumor immune dysfunction and exclusion (TIDE) analysis to assess the possibility of immune escape. The result showed high-risk subgroup had a higher TIDE score than that of the low-risk subgroup, indicating that immunotherapy may be less effective in high-risk patients (**Figure 4C**).

Since the GSVA and TIDE analysis revealed that low-risk patients are more likely to have an active immune system and had a better prognosis after immunotherapy, we would like to assess the difference in chemotherapy response between the two subgroups. The five widely used drugs in clinical therapy for LUSC were included (docetaxel, vinorelbine, cisplatin, paclitaxel, and gemcitabine) (**Figures 4D–H**). We found that the estimated IC50 of cisplatin, paclitaxel, and vinorelbine chemotherapy was significantly higher in the low-risk group while that of docetaxel and gemcitabine chemotherapy had no significant difference, showing that

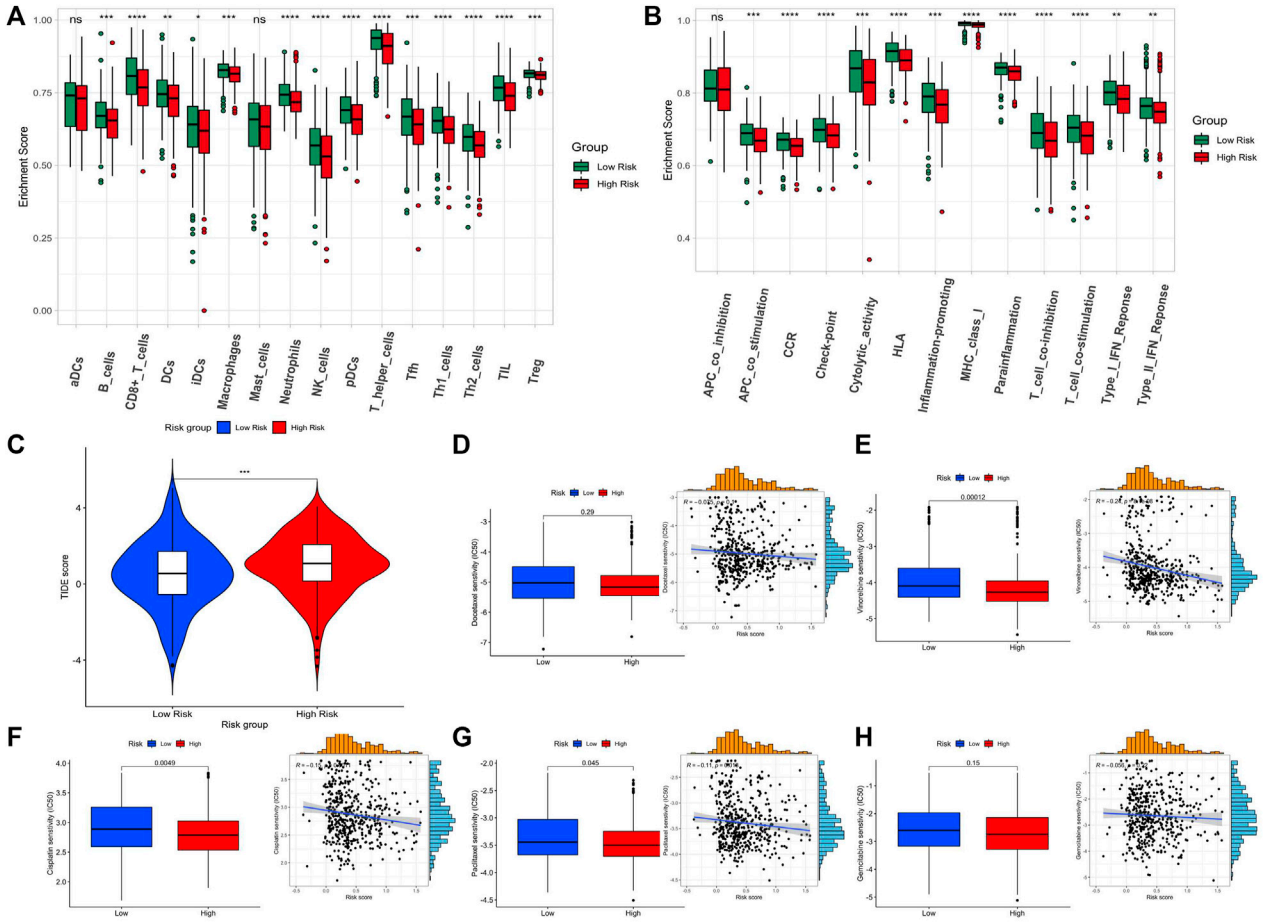


FIGURE 4 | Assessment of the immune level and drug sensitivity related to the 5-lncRNA-based signature. **(A)** enrichment scores of 16 types of immune cells between the two risk subgroups. **(B)** enrichment scores of 13 immune-related pathways between the two risk subgroups. **(C)** Tide score predicting immunotherapy response between the two risk subgroups. **(D–H)** The estimated half-maximal inhibitory concentration (IC50) of docetaxel, vinorelbine, cisplatin, paclitaxel, and gemcitabine for response between the two risk subgroups.

LUSC patients in the high-risk group may be more resistant to chemotherapy of cisplatin, paclitaxel, and vinorelbine.

Development of the Nomogram Combining the 5-lncRNA Signature and Clinical Indicators

We drew a nomogram of the primary dataset and the entire dataset, which consisted of 5-lncRNA-based signatures and three clinical indicators (gender, age, and TNM staging) (**Figure 5A**). Then, we used a calibration chart to evaluate the recognition and calibration ability of the nomogram in the entire dataset (**Figures 5B–D**). An internal validation using bootstrap with 1000 resamplings showed that our nomogram was effective and reliable: the c-index of the primary dataset was 0.678 (95% CI = 0.613-0.743) and the c-index of the entire dataset was 0.613 (95% CI = 0.568-0.658) (**Supplementary Figure S5**). In addition, the AUC of the nomogram reached 0.791 at 1 year, 0.817 at 3 years, and 0.804 at 5 years in the primary dataset while 0.680 at 1 year, 0.721 at 3 years, and 0.729 at 5 years in the entire dataset, which was superior to the predictive performance of 5-lncRNA signature (**Figures 5E, F**). Importantly, our nomogram has a better performance (AUC = 0.828) than that of age (AUC = 0.527), gender (AUC = 0.558), TNM staging (AUC = 0.602), T staging (AUC = 0.622), N staging (AUC = 0.561), and M staging (AUC = 0.514) (**Figures 5G,H**).

Finally, we analyzed the co-expression mRNA of OS-related lncRNAs through GO and KEGG. 1025 DEM levels were positively correlated with the 5-lncRNA signatures (Pearson correlation coefficient >0.25). Results of GO enrichment showed that these co-expressed DEMs involved 121 biological processes, 26 cellular components, and 14 molecular functions. The mRNA functions are mainly focused on DNA replication, nuclear speck, and protein serine/threonine kinase activity (**Figure 5I**). Results of KEGG enrichment showed that these significantly differential co-expressed mRNAs were primarily involved in the pathway of axon guidance, Human T-cell leukemia virus 1 infection, and spliceosome (**Figure 5J**). GSVA enrichment analysis showed a

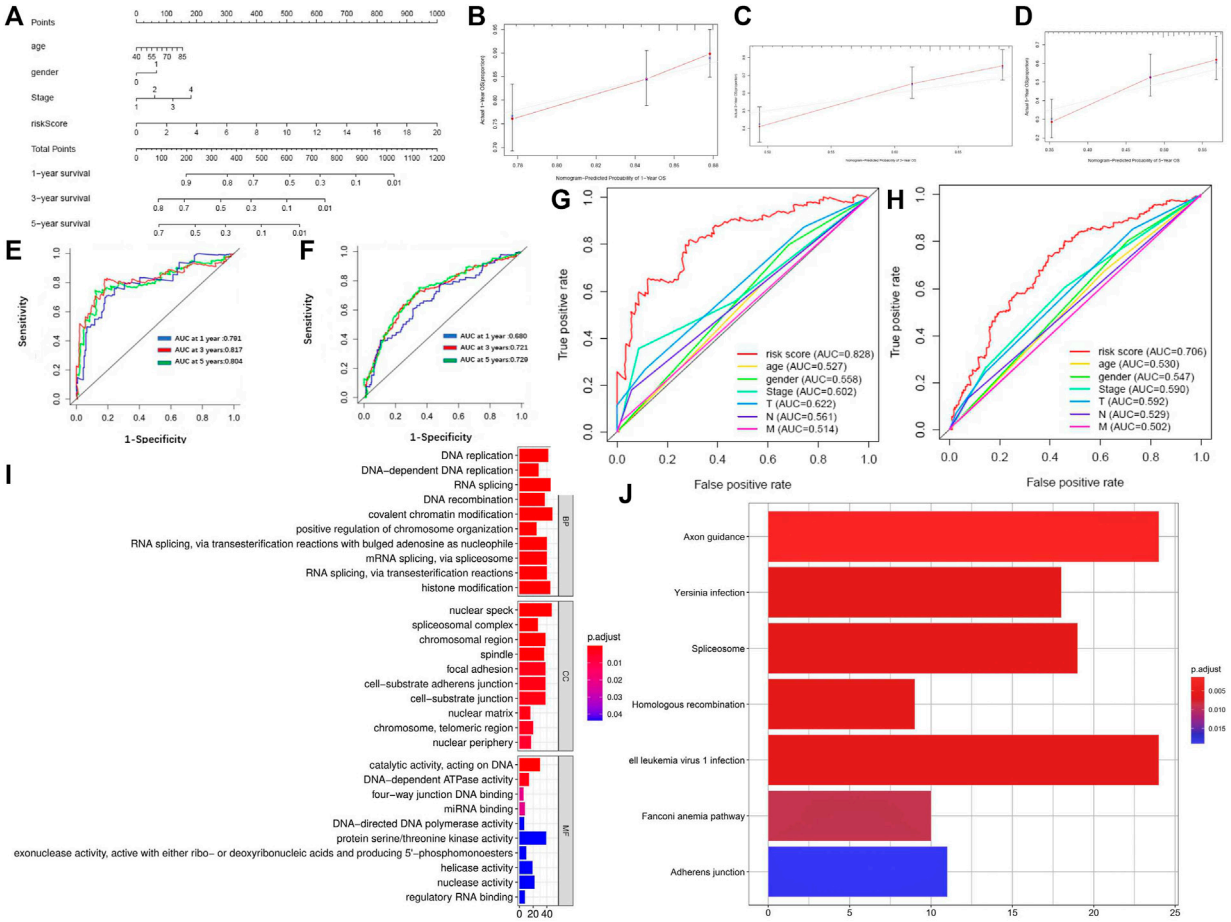


FIGURE 5 | Identification of reliability and accuracy of the nomogram binding the 5-lncRNA signature and clinical indicators. **(A)** Nomogram based on 5-lncRNA in predicting prognosis of LUSC patients. **(B–D)** Calibration chart valuating the survival probability of nomogram at one, three, and five years. **(E,F)** Time-dependent ROC curves of the nomogram in the primary and entire dataset. **(G,H)** Comparison of prognostic ability between nomogram and gender, age, and TNM stage in the primary and entire dataset. **(I)** GO enrichment of mRNAs co-expressed with 5-lncRNA signature. **(J)** KEGG enrichment of mRNAs co-expressed with 5-lncRNA signature.

similar result that the differential genes between high-risk and low-risk subgroups were markedly enriched in the pathway of "SPLICEOSOME", "CELL_ADHESION_MOLECULES_CAMS" and "DNA_REPLICATION" (**Supplementary Figure S6**).

SPATA41 Regulated Alternative Splicing, Apoptosis and Autophagy of LUSC Cells *In Vitro*

We evaluated whether these OS-related lncRNAs influenced the development of LUSC. The number of DEMs co-expressed with the 5-lncRNA signature was examined, and we selected SPATA41 for further functional analysis (**Supplementary Table S3**). Then, two LUSC cell lines (H520 and SK-MES-1) were used to further explore the role of SPATA41 in LUSC. Quantitative real-time PCR was used to compare the expression of SPATA41 between normal lung epithelial cells BEAS-2B and that of H520 and SK-MES-1 cells. The result showed that SPATA41 expression of H520 and SK-MES-1 was higher (**Figure 6A**). We analyzed the differential expression of SPATA41 with LUSC related GEO database (GSE73402). Consistent with our previous result, SPATA41 expression increased significantly in tumor and adjacent tumor tissue in comparison with normal lung tissue (**Figures 6B, C**). We also treated H520 and SK-MES-1 cells with cisplatin, which is a special therapeutic drug for lung cancer. PCR results showed that SPATA41 was also significantly decreased by cisplatin (**Figure 6D**).

We then transfected SPATA41 siRNA into H520 and SK-MES-1 cells, respectively. PCR results revealed that SPATA41 expression was significantly down-regulated in H520 and SK-MES-1 cells after transfection (**Figure 6E**). The results of CCK8 assays also showed that SPATA41 knockdown may impair cell viability (**Figure 6F**). Notably, transwell assay, wound-healing assays, and colony formation assay demonstrated that the knockdown of SPATA41 dramatically attenuated the invasive, migratory, and proliferation abilities of H520 and SK-MES-1 (**Figures 6G–I**). We cannot get the transwell assay result of SK-MES-1 due to its low invasion ability.

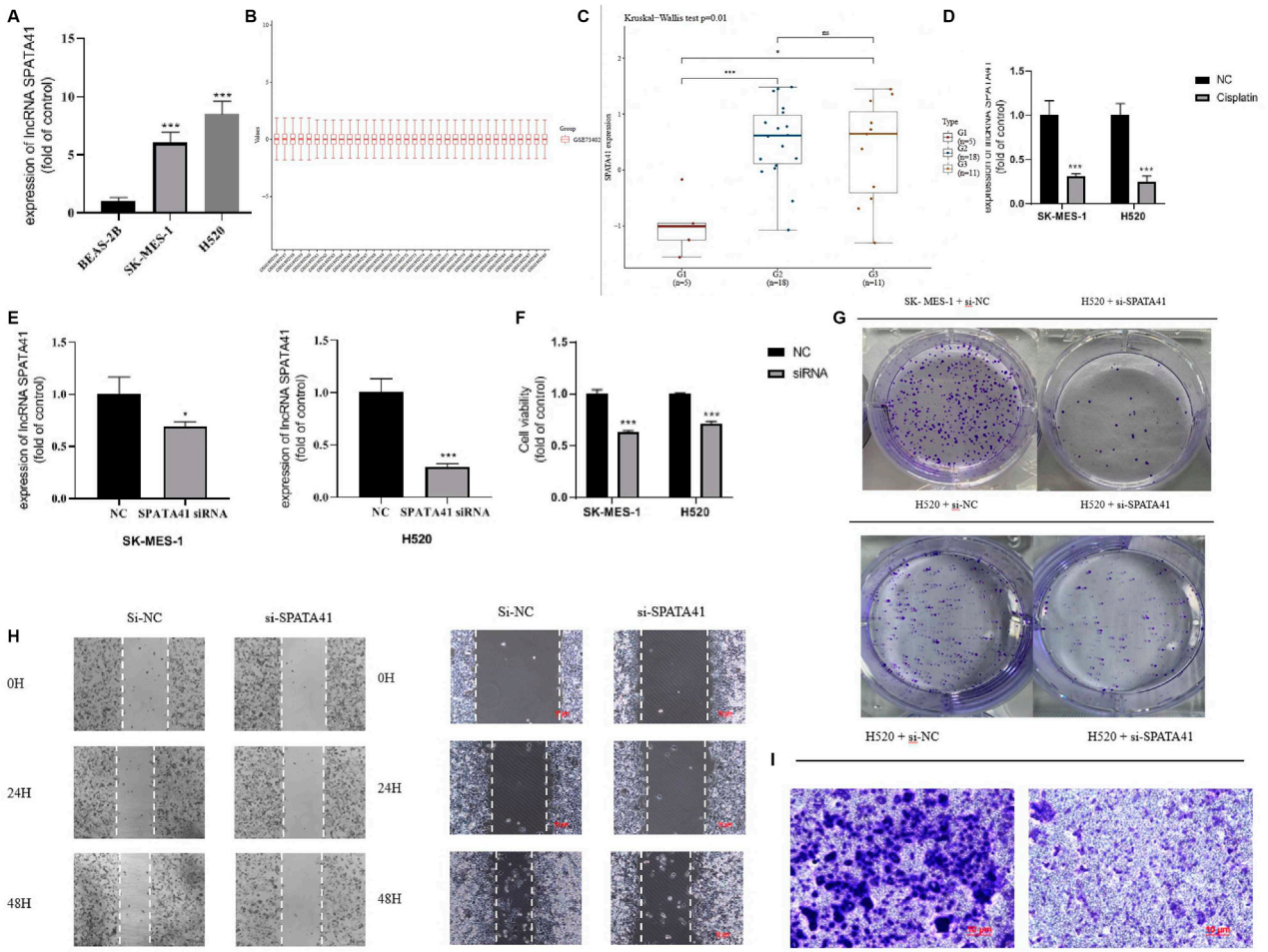


FIGURE 6 | Assessment of the function of SPATA41 in cellular survival, invasion, migration, and colony. **(A)** Quantitative real-time PCR analysis of SPATA41 expression in BEAS-2B, SK-MES-1, and H520 cells. **(B)** Boxplot of the tumor, adjacent tumor, and normal lung tissue. **(C)** Comparison of SPATA41 expression in tumor, adjacent tumor, and normal lung tissue. **(D)** Quantitative real-time PCR analysis of SPATA41 expression between NC group and siRNA group in SK-MES-1 and H520 cells. **(E)** Quantitative real-time PCR analysis of SPATA41 expression between NC group and siRNA group in SK-MES-1 and H520 cells. **(F)** Results of CCK8 assays between NC group and siRNA group in SK-MES-1, and H520 cells. **(G)** Results of colony formation assays between NC group and siRNA group in H520 cells and SK-MES-1 cells. **(H)** Results of wound healing assays between NC group and siRNA group in H520 cells and SK-MES-1 cells. **(I)** Results of transwell assays between NC group and siRNA group in H520 cells. Data are expressed as mean ± SD (n = 3); $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Variable splicing and programmed cell death are critical processes during tumor development. To further investigate the involvement of SPATA41 in the molecular pathological course of LUSC, we predicted the possible downstream proteins of SPATA41 by CATRAPID software (**Supplementary Figure S7**). Finally, four potential genes (SRSF1, SRSF9, FUS, and SFPQ) were screened, which were all related to alternative splicing. PCR and WB experiments also confirmed that SPATA41 knockdown significantly affected these four splicing-associated genes (**Figures 7A, B**). The sequences of genes are listed in **Supplementary Table S4**. In addition, the expression of autophagy protein (p62, Beclin-1, and LC3B) was also influenced (**Figure 7C**). Results of Annexin V-FITC/PI double staining experiments also showed that knockdown of SPATA41 could cause significant apoptosis (**Figure 7D**). These results indicated that SPATA41 regulated the expression of splicing-associated genes in tumor cells and further influenced cellular survival.

DISCUSSION

Bioinformatics is a comprehensive subject of statistics, computer science, and biology (Friedlaender et al., 2019; Ren et al., 2020). It reveals the hidden biological mystery by collecting, counting, and analyzing numerous complex biological data. In the transcripts of the human genome, only 2% of the messenger RNAs encode proteins, while the rest 98% of the RNA molecules do not encode proteins, which are called noncoding RNAs (Rokavec et al., 2017).

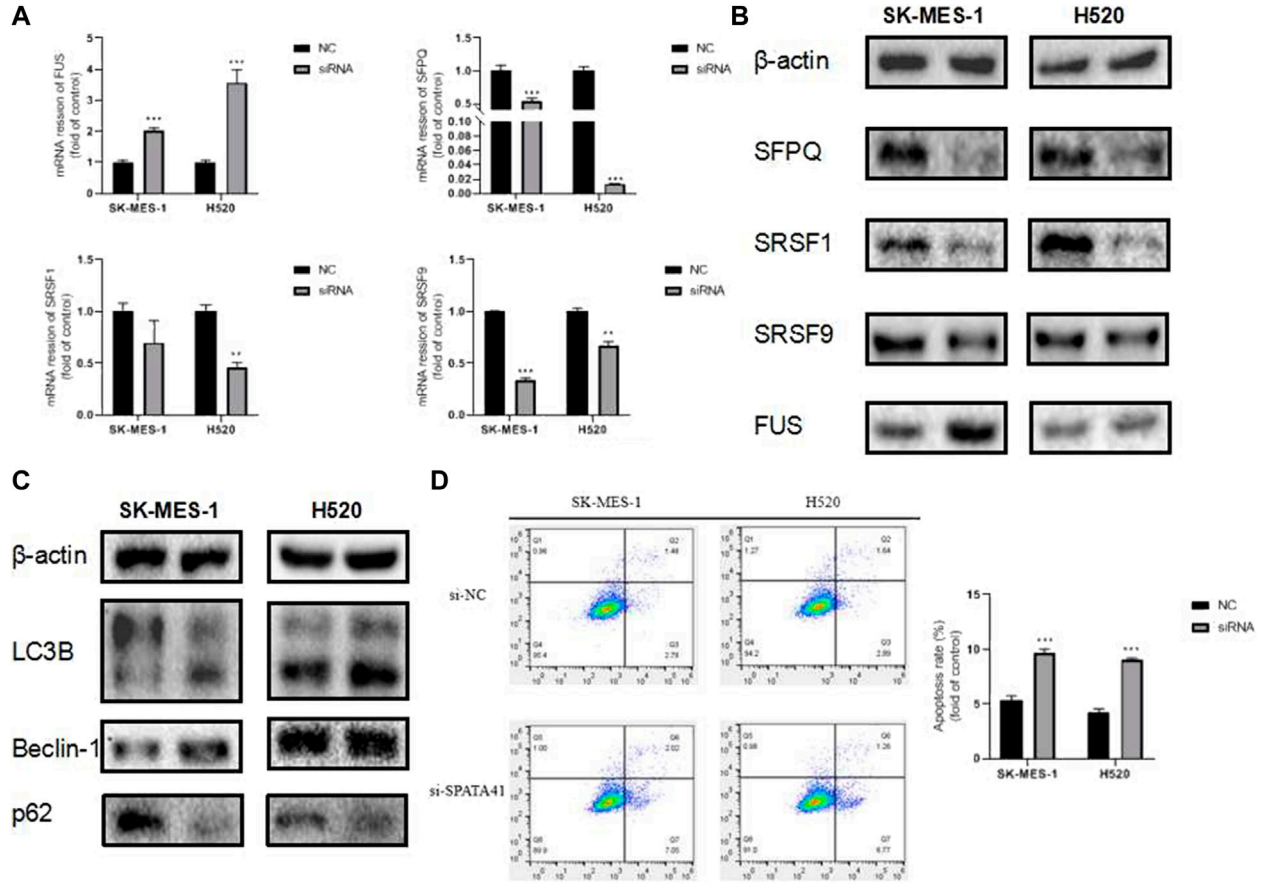


FIGURE 7 | Assessment of the influence of SPATA41 on splicing factor, autophagy, and apoptosis. **(A)** Quantitative real-time PCR analysis of FUS, SFPQ, SRSF1, and SRSF9 expression between NC group and siRNA group in H520 cells, and SK-MES-1 cells. **(B)** Western blotting of FUS, SFPQ, SRSF1, and SRSF9 expression between NC group and siRNA group in H520 cells and SK-MES-1 cells. **(C)** Western blotting of LC3B, Beclin-1, and p62 expression between NC group and siRNA group in H520 cells and SK-MES-1 cells. **(D)** Results of Annexin V-FITC/PI double staining assay between NC group and siRNA group in SK-MES-1 and H520 cells. Data are expressed as mean ± SD (n = 3); $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

In the past, lncRNA was considered as "noise" in gene transcription. However, it has been found that lncRNA is involved in the physiological functions of cells with the development of the biological genome, including chromatin modification, post transcriptional regulation, and nuclear transport (Kourou et al., 2015).

At present, the prediction of survival time of patients with LUSC mainly depends on the TNM staging system. However, patients with similar TNM staging sometimes show opposite responses to the same treatment, which is considered caused by the heterogeneity between tumor genes (Pan et al., 2020). Therefore, an effective prognosis method for LUSC is urgently required, especially in the era of individual treatment (White et al., 2017). As people pay increasing attention to personalized medicine, many genetic markers related to LUSC prognosis have been screened (Guo et al., 2016; Ibrahim et al., 2018). However, most of these studies only focused on the statistical ability of molecular marker screening without considering its clinical effectiveness. Some studies found that in addition to classical TNM staging, gender and age are also important in predicting LUSC prognosis (Gauthier et al., 2019; Chen et al., 2020). Now in this study, we combined these clinical indicators (age, gender, and TNM staging) with 5-lncRNA signature and constructed a nomogram to quantify the survival probability of LUSC patients. We found that the predictive performance of the prognostic nomogram was better than the 5-lncRNA signature and traditional TNM staging. The c-index and calibration curve also verified the reliability of the nomogram. At the same time, simplicity is also one of its advantages, which can guide clinicians to evaluate the disease progression and prognosis more conveniently and accurately (Ferrè et al., 2016; Schmitt and Chang, 2016; Peng et al., 2017; Kopp and Mendell, 2018). Our prognostic model aimed to determine the association between prognosis and basic characteristics and should be accurate and economical (Iasonos et al., 2008; Balachandran et al., 2015; König et al., 2017; Goh et al., 2020). The nomogram included multiple independent variables and was easy for clinicians to evaluate

the results and choose individual treatments for LUSC patients.

Although some famous lncRNAs have been widely reported, the specific mechanism of lncRNA still needs to be entirely explored (Li et al., 2014; Chen, 2016). The functional expression pattern of lncRNA is often related to its highly specific transcript abundance. In this study, we inferred the functions of main effective lncRNAs (SPATA41, AL034550.2, AP003721.2, AC106786.1, and AC078889.1) according to the functional evaluation of their co-expressed DEMs. Enrichment analysis of GO showed that the co-expressed DEMs were mainly enriched in DNA alternative splicing and nuclear speckles while KEGG enrichment indicated axon guidance, Human T-cell leukemia virus 1 infection, and spliceosome may be the downstream signaling pathway of the five related lncRNAs.

TME was of great importance in the initiation and development of tumorigenesis. Previous studies had identified that the immune microenvironment regulated tumorigenesis. Therefore, exploring TME remodeling has great development potential in tumor prediction and prognosis, and further fostering the transition of TME from tumor-friendly to tumor-suppressed. We found that the traditional TNM staging indicators could not distinguish the difference in TME score, while our lncRNA-based signature showed a great performance that Stromal score, Immune score, and ESTIMATE Score of low-risk subgroup were significantly higher than that of the high-risk subgroup. Our results showed that immune components in TME contributed to the prognosis of patients. Particularly, the proportion of immune and stromal components in TME was significantly correlated with the progression of LUSC, such as invasion and metastasis. These results exhibited the reliability and efficiency of our 5-lncRNA-based signature in immunity prediction.

Due to the immune-related pathways shown in functional analyses, we would like to further explore immune infiltration and immune escape between low-risk and high-risk groups. The result was consistent with our previous conclusion that the high-risk group had universally decreased levels of infiltrating immune cells, decreased activity of immune-related pathways, and higher TIDE score, indicating that low-risk patients may have opportunities for better prognosis when receiving immunotherapy. TIDE prediction scores were associated not only with poor efficacy of immune checkpoint inhibition therapy but also with poor survival of patients treated with anti-PD1 and anti-CTLA4 (Jiang et al., 2018). The lower the TIDE score was, the worse the effect of the ICB therapy may get. Importantly, we further performed an analysis of chemotherapy response to assess the drug sensitivity and resistance of our 5-lncRNA signature. The results showed that the estimated IC50 for cisplatin, paclitaxel, and vinorelbine was significantly higher in the low-risk group. High-risk patients may be more sensitive to chemotherapy of these drugs. A previous study reported that spliceosome-mediated RNA trans-splicing (SMaRT) could effectively overcome the obstacle that chemotherapy could not clear cancer cells with tumor specificity (Woess et al., 2022). Their study could explain that the different chemoresistance between groups may be related to alternative splicing.

In addition, further function assays were performed on lncRNA SPATA41, which is most associated with co-expressed DEMs among the 5 lncRNAs. Our results showed that knockdown of SPATA41 significantly affected the invasive, migratory, and proliferative abilities of LUSC cells and influenced the expression of splicing factors (SRSF1, SRSF9, FUS, and SFPQ). The apoptosis double staining experiment verified that SPATA41 knockdown could effectively induce apoptosis of SK-MES-1 and H520 cells. Moreover, increased expression of LC3 II/LC3 I and decreased expression of p62 indicated that autophagy may also be involved in the regulative process of SPATA41 in cancer cells. Thus, silencing SPATA41 in LUSC cells may prevent the development of tumors.

In conclusion, we identified the importance of lncRNA expression patterns in LUSC patients and confirmed our 5-lncRNA signature had a great advantage in assessing immune reaction, chemotherapy sensitivity, and the risk level of patients. The nomogram combining 5-lncRNA signature and clinical indicators provides an effective and reliable predictive model to help the individual treatment of LUSC patients.

LIMITATION

Our study had several limitations. First, we did not refer to Lasso Cox regression to screen the differential lncRNA. Second, we did not collect clinical samples from LUSC patients for comparison. Third, our study only involved two kinds of LUSC cell lines. Fourth, we did not validate the direct interaction between lncRNA and protein by RIP or pull-down assays.

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

SH contributed to the conception of the study; SH and MC performed the experiment; MC, YJ, and SS contributed significantly to analysis and manuscript preparation; SH, YZ, and MC wrote and revised the manuscript; All authors performed the data analyses and approved the final manuscript.

FUNDING

This work was supported in part by the Natural Science Foundation of Nanjing University of Chinese medicine (XZR2020072) and Graduate Research and Innovation Projects of Jiangsu Province (KYCX21_1637).

ACKNOWLEDGMENTS

We are very grateful for the help provided by the experiment center for science and technology at the Nanjing University of Chinese Medicine.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.905353/full#supplementary-material

REFERENCES

Balachandran, V. P., Gonen, M., Smith, J. J., and DeMatteo, R. P. (2015). Nomograms in Oncology: More Than Meets the Eye. *Lancet Oncol.* 16 (4), e173–e180. doi:10.1016/s1470-2045(14)71116-7

Bozinovski, S., Vlahos, R., Anthony, D., McQualter, J., Anderson, G., Irving, L., et al. (2016). COPD and Squamous Cell Lung Cancer: Aberrant Inflammation and Immunity Is the Common Link. *Br. J. Pharmacol.* 173 (4), 635–648. doi:10.1111/bph.13198

Bu, J., Zhang, P., Zhu, K., Yan, Y., Shi, B., Wang, J., et al. (2020). Constructing a Global Transcriptional Regulatory Landscape for Early Non-small Cell Lung Cancer to Identify Hub Genes and Key Pathways. *Aging* 12 (18), 17948–17957. doi:10.18632/aging.103475

Chen, C., Hou, J., Tanner, J. J., and Cheng, J. (2020). Bioinformatics Methods for Mass Spectrometry-Based Proteomics Data Analysis. *Ijms* 21 (8), 2873. doi:10.3390/ijms21082873

Chen, L.-L. (2016). Linking Long Noncoding RNA Localization and Function. *Trends Biochem. Sci.* 41 (9), 761–772. doi:10.1016/j.tibs.2016.07.003

Engreitz, J. M., Haines, J. E., Perez, E. M., Munson, G., Chen, J., Kane, M., et al. (2016). Local Regulation of Gene Expression by lncRNA Promoters, Transcription and Splicing. *Nature* 539 (7629), 452–455. doi:10.1038/nature20149

Ferrè, F., Colantoni, A., and Helmer-Citterich, M. (2016). Revealing Protein-lncRNA Interaction. *Brief. Bioinform* 17 (1), 106–116. doi:10.1093/bib/bbv031

Friedlaender, A., Banna, G., Malapelle, U., Pisapia, P., and Addeo, A. (2019). Next Generation Sequencing and Genetic Alterations in Squamous Cell Lung Carcinoma: Where Are We Today? *Front. Oncol.* 9, 166. doi:10.3389/fonc.2019.00166

Gauthier, J., Vincent, A. T., Charette, S. J., and Derome, N. (2019). A Brief History of Bioinformatics. *Brief. Bioinform* 20 (6), 1981–1996. doi:10.1093/bib/bby063

Gibson, W. J., Hoivik, E. A., Halle, M. K., Taylor-Weiner, A., Cherniack, A. D., Berg, A., et al. (2016). The Genomic Landscape and Evolution of Endometrial Carcinoma Progression and Abdominopelvic Metastasis. *Nat. Genet.* 48 (8), 848–855. doi:10.1038/ng.3602

Goh, Y. M., Antonowicz, S. S., Boshier, P., and Hanna, G. B. (2020). Metabolic Biomarkers of Squamous Cell Carcinoma of the Aerodigestive Tract: A Systematic Review and Quality Assessment. *Oxidative Med. Cell. Longev.* 2020, 1–13. doi:10.1155/2020/2930347

Grazziotin, L. R., Currie, G., Twilt, M., Ijzerman, M. J., Kip, M. M. A., Koffijberg, H., et al. (2022). Real-world Data Reveals the Complexity of Disease Modifying Anti-rheumatic Drug Treatment Patterns in Juvenile Idiopathic Arthritis: an Observational Study. *Pediatr. Rheumatol.* 20 (1), 25. doi:10.1186/s12969-022-00682-x

Guo, H., Ahmed, M., Zhang, F., Yao, C. Q., Li, S., Liang, Y., et al. (2016). Modulation of Long Noncoding RNAs by Risk SNPs Underlying Genetic Predispositions to Prostate Cancer. *Nat. Genet.* 48 (10), 1142–1150. doi:10.1038/ng.3637

Iasonos, A., Schrag, D., Raj, G. V., and Panageas, K. S. (2008). How to Build and Interpret a Nomogram for Cancer Prognosis. *Jco* 26 (8), 1364–1370. doi:10.1200/jco.2007.12.9791

Ibrahim, B., McMahon, D. P., Hufsky, F., Beer, M., Deng, L., Mercier, P. L., et al. (2018). A New Era of Virus Bioinformatics. *Virus Res.* 251, 86–90. doi:10.1016/j.virusres.2018.05.009

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T Cell Dysfunction and Exclusion Predict Cancer Immunotherapy Response. *Nat. Med.* 24 (10), 1550–1558. doi:10.1038/s41591-018-0136-1

König, I. R., Fuchs, O., Hansen, G., von Mutius, E., and Kopp, M. V. (2017). What Is Precision Medicine? *Eur. Respir. J.* 50 (4), 1700391. doi:10.1183/13993003.00391-2017

Kopp, F., and Mendell, J. T. (2018). Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* 172 (3), 393–407. doi:10.1016/j.cell.2018.01.011

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., and Fotiadis, D. I. (2015). Machine Learning Applications in Cancer Prognosis and Prediction. *Comput. Struct. Biotechnol. J.* 13, 8–17. doi:10.1016/j.csbj.2014.11.005

Li, G., and Guo, X. (2020). LncRNA STARD13-AS Blocks Lung Squamous Carcinoma Cells Growth and Movement by Targeting miR-1248/C3A. *Pulm. Pharmacol. Ther.* 64, 101949. doi:10.1016/j.pupt.2020.101949

Li, J.-H., Liu, S., Zheng, L.-L., Wu, J., Sun, W.-J., Wang, Z.-L., et al. (2014). Discovery of Proteinâ€"lncRNA Interactions by Integrating Large-Scale CLIP-Seq and RNA-Seq Datasets. *Front. Bioeng. Biotechnol.* 2, 88. doi:10.3389/fbioe.2014.00088

Lian, S., Liu, Z., Zhou, Y., Guo, J., Gong, K., and Wang, T. (2020). The Differential Expression Patterns and Co-expression Networks of Paralogs as an Indicator of the TNM Stages of Lung Adenocarcinoma and Squamous Cell Carcinoma. *Genomics* 112 (6), 4115–4124. doi:10.1016/j.ygeno.2020.07.019

Ma, J., Qi, G., and Li, L. (2020). LncRNA NNT-AS1 Promotes Lung Squamous Cell Carcinoma Progression by Regulating the miR-22/FOXM1 axis. *Cell Mol. Biol. Lett.* 25, 34. doi:10.1186/s11658-020-00227-8

Mantovani, A., Sozzani, S., Locati, M., Allavena, P., and Sica, A. (2002). Macrophage Polarization: Tumor-Associated Macrophages as a Paradigm for Polarized M2 Mononuclear Phagocytes. *Trends Immunol.* 23, 549–555. doi:10.1016/s1471-4906(02)02302-5

Pan, Y.-B., Zhu, Y., Zhang, Q.-W., Zhang, C.-H., Shao, A., and Zhang, J. (2020). Prognostic and Predictive Value of a Long Non-coding RNA Signature in Glioma: A lncRNA Expression Analysis. *Front. Oncol.* 10, 1057. doi:10.3389/fonc.2020.01057

Peng, W.-X., Koirala, P., and Mo, Y.-Y. (2017). LncRNA-mediated Regulation of Cell Signaling in Cancer. *Oncogene* 36 (41), 5661–5667. doi:10.1038/onc.2017.184

Quinn, J. J., Ilik, I. A., Qu, K., Georgiev, P., Chu, C., Akhtar, A., et al. (2014). Revealing Long Noncoding RNA Architecture and Functions Using Domain-specific Chromatin Isolation by RNA Purification. *Nat. Biotechnol.* 32 (9), 933–940. doi:10.1038/nbt.2943

Reinhold, W. C., Sunshine, M., Liu, H., Varma, S., Kohn, K. W., Morris, J., et al. (2012). CellMiner: AWeb-Based Suite of Genomic and Pharmacologic Tools to Explore Transcript and Drug Patterns in the NCI-60 Cell Line Set. *Cancer Res.* 72, 3499–3511. doi:10.1158/0008-5472.can-12-1370

Ren, P., Xing, L., Hong, X., Chang, L., and Zhang, H. (2020). LncRNA PITPNA-AS1 Boosts the Proliferation and Migration of Lung Squamous Cell Carcinoma Cells by Recruiting TAF15 to Stabilize HMGB3 mRNA. *Cancer Med.* 9 (20), 7706–7716. doi:10.1002/cam4.3268

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

Rokavec, M., Horst, D., and Hermeking, H. (2017). Cellular Model of Colon Cancer Progression Reveals Signatures of mRNAs, miRNA, lncRNAs, and Epigenetic Modifications Associated with Metastasis. *Cancer Res.* 77 (8), 1854–1867. doi:10.1158/0008-5472.Can-16-3236

Sánchez-Danés, A., and Blanpain, C. (2018). Deciphering the Cells of Origin of Squamous Cell Carcinomas. *Nat. Rev. Cancer* 18 (9), 549–561. doi:10.1038/s41568-018-0024-5

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

Socinski, M. A., Obasaju, C., Gandara, D., Hirsch, F. R., Bonomi, P., Bunn, P. A., Jr., et al. (2018). Current and Emergent Therapy Options for Advanced Squamous Cell Lung Cancer. *J. Thorac. Oncol.* 13 (2), 165–183. doi:10.1016/j.jtho.2017.11.111

Storti, C. B., de Oliveira, R. A., de Carvalho, M., Hasimoto, E. N., Cataneo, D. C., Cataneo, A. J. M., et al. (2020). Telomere-associated Genes and Telomeric lncRNAs Are Biomarker Candidates in Lung Squamous Cell Carcinoma (LUSC). *Exp. Mol. Pathology* 112, 104354. doi:10.1016/j.yexmp.2019.104354

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

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

White, N. M., Zhao, S. G., Zhang, J., Rozycki, E. B., Dang, H. X., McFadden, S. D., et al. (2017). Multi-institutional Analysis Shows that Low PCAT-14 Expression Associates with Poor Outcomes in Prostate Cancer. *Eur. Urol.* 71 (2), 257–266. doi:10.1016/j.eururo.2016.07.012

Woess, K., Sun, Y., Morio, H., Stierschneider, A., Kaufmann, A., Hainzl, S., et al. (2022). Evaluating a Targeted Cancer Therapy Approach Mediated by RNA Trans-splicing *In Vitro* and in a Xenograft Model for Epidermolysis Bullosa-Associated Skin Cancer. *Ijms* 23 (1), 575. doi:10.3390/ijms23010575

Xu, F., Zhang, H., Chen, J., Lin, L., and Chen, Y. (2020). Immune Signature of T Follicular Helper Cells Predicts Clinical Prognostic and Therapeutic Impact in Lung Squamous Cell Carcinoma. *Int. Immunopharmacol.* 81, 105932. doi:10.1016/j.intimp.2019.105932

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

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

Zhang, Z., and Kattan, M. W. (2017). Drawing Nomograms with R: Applications to Categorical Outcome and Survival Data. *Ann. Transl. Med.* 5, 211. doi:10.21037/atm.2017.04.01

Zhou, T., Cai, Z., Ma, N., Xie, W., Gao, C., Huang, M., et al. (2020). A Novel Ten-Gene Signature Predicting Prognosis in Hepatocellular Carcinoma. *Front. Cell Dev. Biol.* 8, 629. doi:10.3389/fcell.2020.00629

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 09 June 2023
DOI 10.3389/fgene.2023.1212465

OPEN ACCESS

EDITED BY
Xiaogang Wu,
University of Texas MD Anderson Cancer Center, United States

REVIEWED BY
Le Son Tran,
Medical Genetics Institute, Vietnam
Tian Tian,
Children's Hospital of Philadelphia, United States

*CORRESPONDENCE
Rongfeng Liu,
liurongfeng82@163.com

RECEIVED 26 April 2023
ACCEPTED 30 May 2023
PUBLISHED 09 June 2023

CITATION
Zhang X and Liu R (2023), Pyroptosis-related genes *GSDMB*, *GSDMC*, and *AIM2* polymorphisms are associated with risk of non-small cell lung cancer in a Chinese Han population.
Front. Genet. 14:1212465.
doi: 10.3389/fgene.2023.1212465

Pyroptosis-related genes *GSDMB*, *GSDMC*, and *AIM2* polymorphisms are associated with risk of non-small cell lung cancer in a Chinese Han population

Xia Zhang[1] and Rongfeng Liu[2]*

[1]Department of Respiratory Medicine, Shanxi Province Cancer Hospital/Shanxi Hospital Affiliated to Cancer Hospital, Chinese Academy of Medical Sciences/Cancer Hospital Affiliated to Shanxi Medical University, Taiyuan, Shanxi, China, [2]Department of Medical Oncology, The Fourth Hospital of Hebei Medical University, Shijiazhuang, Hebei, China

Background: Pyroptosis is essential for the remodeling of tumor immune microenvironment and suppression of tumor development. However, there is little information available about pyroptosis-related gene polymorphisms in non-small cell lung cancer (NSCLC).

Methods: Six SNPs in the *GSDMB*, *GSDMC*, and *AIM2* were genotyped in 650 NSCLC cases and 650 healthy controls using a MassARRAY platform.

Results: Minor alleles of rs8067378, rs2305480, and rs77681114 were associated with a lower risk of NSCLC ($p < 0.005$), whereas rs2290400 and rs1103577 were related to an increased risk ($p < 0.00001$). Moreover, rs8067378-AG/GG, rs2305480-GA/AA, and rs77681114-GA/AA genotypes were associated with a decrease in NSCLC risk ($p < 0.005$). In contrast, the TC/CC genotypes of rs2290400 and rs1103577 were associated with an elevated NSCLC risk ($p < 0.0001$). Based on the analysis of genetic models, minor alleles of rs8067378, rs2305480 and rs77681114 were related to reduced risk of NSCLC ($p < 0.05$); whereas rs2290400 and rs1103577 were related to increased risk ($p < 0.01$).

Conclusion: Our findings provided new insights into the roles of pyroptosis-related genes in NSCLC, as well as new factors to be considered for assessing the risk of developing this cancer.

KEYWORDS

non-small cell lung cancer (NSCLC), pyroptosis, gasdermin (GSDM), absent in melanoma 2 (AIM2), polymorphisms

Introduction

The incidence and mortality of lung cancer are stubbornly high in spite of the great effort put into the related field, with approximately 1.3 million deaths worldwide each year (Siegel et al., 2021). Nearly 85% of all lung cancer patients were diagnosed with non-small cell lung cancer (NSCLC) (Jonna and Subramaniam, 2019), including the following three pathological types (Wu et al., 2021): most of lung adenocarcinoma originates from the bronchial mucosal epithelium, squamous cell carcinoma mostly originates in the larger bronchi, and large cell carcinoma often occurs in the upper lobe of the lung (Rodriguez-Canales et al., 2016). Although there are many ways to treat lung cancer, including surgery, radiotherapy, targeted

drugs and chemotherapy, lung cancer is still a major challenge to human health and life around the world because of its high metastasis, high recurrence and low cure (Jones and Baldwin, 2018). According to statistics, approximately 40% and 60% of patients with stage I and II NSCLC still die from distant metastases within 5 years in patients undergoing tumor resection surgery (Torre et al., 2016). Therefore, early detection and prevention of NSCLC are crucial for improving the survival rate of the patients. Investigation of single-nucleotide polymorphisms (SNPs) in driver genes has proven to be a potential strategy to elaborate the hereditary susceptibility to NSCLC (Feng et al., 2020; Luo et al., 2021). Combination of the SNPs strategy and nowadays tumor related research hotspot might generate novel significative genotyping data and provide theoretical basis for early prevention of the disease.

Pyroptosis was first proposed to describe the process of programmed cell death caused by *Salmonella* infection of macrophages leading to their inflammatory death (D'Souza and Heitman, 2001). Pyroptosis can protect cells from infection by eliminating pathogen host cells and triggering an inflammatory response, with the symptoms of cell swelling, nuclear clotting, membranolysis, and the secretion of inflammatory cytokines and damage-related molecular patterns (Man and Kanneganti, 2016). Pyroptosis has been considered as a Caspase-1/11-induced programmed cell death, but the specific mechanism of Caspase-induced pyroptosis has been studied for a long time until the role of gasdermin (GSDM) family was revealed (Case et al., 2013; Li et al., 2020). It has been found several members in GSDM family, including GSDMA/B/C/D/E and DFNB59 (Broz et al., 2020). The GSDMs could be cleaved and activated by protease and then mediating pyroptosis, the GSDMB and GSDMC were processed by Caspase-3/6/7/granzyme A and Caspase-8 into their active form, respectively (Wang et al., 2017). GSDMB is high expressed in several types of cancers, and its expression level is related with poor prognosis of patients (Li et al., 2020). Overexpression of GSDMC also has relativity with a bad outcomes of lung adenocarcinoma patients (Wei et al., 2020). Moreover, inflammasome absent in melanoma 2 (AIM2) can recruit and activate Caspase-1, subsequently enhance the release of interleukin (IL)-1β and IL-18, and finally induce the pyroptosis (Sharma et al., 2019). The high expression of AIM2 has been found in NSCLC tissues, and functioned as an oncogene by influencing the formation of inflammasome (Zhang et al., 2019). However, little study focused on the SNPs in *GSDMB*, *GSDMC*, and *AIM2* among patients with NSCLC.

Considering the above research background, we finally focused on six SNPs in *GSDMB*, *GSDMC*, and *AIM2* based on previous studies. The rs8067378 (Li et al., 2022) and rs2305480 (Karunas et al., 2021) in *GSDMB* were found to be protective SNPs for cervical squamous intraepithelial lesion and asthma, respectively; while *GSDMB*-rs2290400 was correlated with asthma combined with allergic rhinitis (Karunas et al., 2021). Moreover, *GSDMC*-rs77681114 was related to reduced risk of lumbar disc herniation (Wu et al., 2020). In addition, *AIM2*-rs1103577 has been found protective role on risk of tuberculosis (Figueira et al., 2021), and rs2298803 in *AIM2* was investigated in patients with rectal cancer and have no correlation with adverse events of postoperative chemoradiotherapy (Chen et al., 2023). We distinguished the alleles and genotypes of these SNPs in our study cohort, and made a disease risk prediction using genetic model analysis.

Materials and methods

Participants

We enrolled 650 patients with histopathologically diagnosed NSCLC and 650 healthy controls for this case-control study. Each of the participants was recruited from Shanxi Province Cancer Hospital. There were no previous treatments for any of the cases, and all were newly diagnosed. Blood donors without a history of cancer, immune disorders, or serious diseases were used as controls. We obtained written informed consent from each subject, and the study was approved by the Ethics Department of Shanxi Province Cancer Hospital and was carried out in accordance with the World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects.

Genotyping

Five milliliters of whole blood was collected from each subject in tubes containing ethylenediaminetetraacetic acid. DNA was extracted using a QIAamp DNA Blood Midi Kit (QIAGEN, Germany). Spectrometry (DeNovix DS-11FX Ultramicro spectrophotometer, United States) was used to measure the DNA concentration. Primers were designed using Sequenom MassARRAY Assay Design 3.0 software. The primers used for this study is listed in Table 1. SNP genotyping was performed on a Mass ARRAY iPLEX platform (Sequenom, San Diego, CA, United States) according to the manufacturer's instructions. Assay design and mass spectrometric genotyping were performed as previously described (Gabriel et al., 2009).

Statistical analyses

The statistical analyses were carried out using SPSS package version 20.0 (SPSS, Chicago, IL, United States). The chi-square test was used to compare the gender and smoking status, and the Student t-test was used to compare the age between cancer patients and healthy subjects, respectively. Controls were checked for deviations from Hardy-Weinberg equilibrium (HWE) by measuring minor allele frequencies (MAFs). SNPstats (https://www.snpstats.net/start.htm) was used to evaluate the associations between SNPs and NSCLC risk, and the results are presented in odds ratios (ORs) and 95% confidence intervals (CIs) with adjustments for sex, age and smoking status. Statistical significance was established when $p < 0.05$.

Results

Table 2 shows the sex, age, and smoking status of the participants. Sex, age, and smoking status did not differ significantly between case and control groups ($p > 0.05$). Adenocarcinoma, squamous cell carcinoma, adenosquamous carcinoma, and large cell lung cancer account for 50.5%, 41.1%, 4.4%, and 4.0% of NSCLC cases, respectively.

The gene location information of candidate SNPs and their MAFs in cases and controls are listed in Table 3. Rs2305480 was

TABLE 1 PCR primers used for this study.

SNP	1st-PCR primer sequences	2nd-PCR primer sequences	UEP sequences
rs8067378	ACGTTGGATGCTGTGAGTGGAAAGCTTGAC	ACGTTGGATGACCTGGCAGTGATATAAACG	GATATAAACGTTTTTCCC
rs2305480	ACGTTGGATGCTAGGTATCTGAGGTCCTGA	ACGTTGGATGAAAAGGCTGCTTAGGAGAGG	AGGAGAGGCTTGTCTG
rs2290400	ACGTTGGATGGTTTTCCAGTCTCAGAAGCG	ACGTTGGATGTAAGGATCTCAGGGCCTTAC	CTCCCACTGACTCTT
rs77681114	ACGTTGGATGCCCATTATGGCTTCAAGGAG	ACGTTGGATGCCTAAAGAAACTTCAACAGG	TTCAACAGGATTCAAA
rs1103577	ACGTTGGATGTACTTCCACTACCTATCCCC	ACGTTGGATGGATGATTCCCGGCTTTCTG	CTTTCTGGCTTGAGC
rs2298803	ACGTTGGATGGTCCTCTGCTAGTTAAGCTC	ACGTTGGATGAGCTCCTCTATGGTGCTTAC	TGGTGCTTACCTCCTGA

TABLE 2 The demographic characteristics of the participants.

Characteristics	Case (n = 650)	Control (n = 650)	χ^2/t	p
Sex (%)				
Male	426 (65.5)	403 (62.0)	1.611	0.204
Female	224 (34.5)	247 (38.0)	0.934	0.350
Age				
mean ± SD	56.98 ± 10.17	56.45 ± 10.25	1.603	0.205
Smoking (%)				
Yes	423 (65.1)	400 (61.5)		
No	227 (34.9)	250 (38.5)		
Pathological types				
Adenocarcinoma	328 (50.5)			
Squamous cell carcinoma	267 (41.1)			
Adenosquamous carcinoma	29 (4.4)			
Large cell lung cancer	26 (4.0)			
Tumor staging				
I or II	236 (36.3)			
III or IV	414 (63.7)			

a missense variant and led to a changed amino acid Pro > Ser, rs77681114 was a synonymous variant (Asn > Asn), other SNPs were intron, downstream or non-coding transcript variant. All SNPs matched HWE (p > 0.05). Three beneficial SNPs, rs8067378, rs2305480 and rs7768114, were significantly associated with a significantly lower risk of NSCLC after comparing the MAFs between cases and controls (rs8067378: OR = 0.666, 95% CI: 0.548–0.810, p = 0.00004; rs2305480: OR = 0.663, 95% CI: 0.549–0.802, p = 0.00002; rs77681114: OR = 0.751, 95% CI: 0.617–0.913, p = 0.00401). As well, two additional SNPs, rs2290400 and rs1103577, were linked to increased NSCLC risk (rs2290400: OR = 1.540, 95% CI: 1.296–1.830, p < 0.00001; rs1103577: OR = 1.497, 95% CI: 1.263–1.774, p < 0.00001).

Table 4 shows the genotype frequencies of candidate SNPs. It was considered that the wild type genotype was the reference genotype. On the basis of the genotype frequencies of SNPs in cases and controls, the OR and 95% confidence interval were calculated for homozygous mutation genotypes and heterozygous mutation genotypes. There was a significant decrease in risk of NSCLC for the AG and GG genotypes of rs8067378 compared to the wild type AA (p = 0.0001). Similarly, the GA and AA genotypes of rs2305480 (p < 0.0001) and rs77681114 (p = 0.0025) were also determined to be protective genotypes for NSCLC. In contrast, the TC/CC genotypes of rs2290400 and rs1103577 were associated with different levels of elevated NSCLC risk (p < 0.0001).

Furthermore, we introduced three classical genetic models—dominant, recessive, and log-additive—so we could better evaluate SNPs' impact on NSCLC risk. Table 5 shows five SNPs that increase or decrease the risk of the disease. All three genetic models showed reduced risk of NSCLC for minor alleles of rs8067378, rs2305480, and rs77681114; while rs2290400 and rs1103577 showed increased risk of the disease (p < 0.01).

TABLE 3 The MAF and HWE of candidate SNPs between NSCLC cases and healthy controls.

SNP	Gene	Position	Allele	Region	MAF-case	MAF-control	HWE *p*	OR (95% CI)	*p*
rs8067378	GSDMB	chr17:39895095	A>G	Downstream Variant	0.17	0.23	0.91	0.666 (0.548–0.810)	0.00004*
rs2305480	GSDMB	chr17:39905943	G>A	Missense Variant	0.18	0.25	0.25	0.663 (0.549–0.802)	0.00002*
rs2290400	GSDMB	chr17:39909987	T>C	Intron Variant	0.33	0.24	0.67	1.540 (1.296–1.830)	<0.00001*
rs77681114	GSDMC	chr8:129750045	G>A	Synonymous Variant	0.17	0.22	0.1	0.751 (0.617–0.913)	0.00401*
rs1103577	AIM2	chr1:159130525	T>C	Intron Variant	0.34	0.26	0.081	1.497 (1.263–1.774)	<0.00001*
rs2298803	AIM2	chr1:159076640	T>C	Non Coding Transcript Variant	0.33	0.31	0.2	1.080 (0.917–1.274)	0.35622

SNP, single nucleotide polymorphism; MAF, minor allele frequency; HWE, Hardy–Weinberg equilibrium.
$p < 0.05$ indicates statistical significance.

TABLE 4 Genotype frequency distributions between NSCLC cases and healthy controls.

SNP	Genotype	Control	Case	OR (95% CI)	*p*
rs8067378	AA	385 (59.2%)	454 (69.8%)	1	0.0001*
	AG	232 (35.7%)	177 (27.2%)	0.63 (0.50–0.80)	
	GG	33 (5.1%)	19 (2.9%)	0.47 (0.26–0.85)	
rs2305480	GG	362 (55.7%)	433 (66.6%)	1	<0.0001*
	GA	254 (39.1%)	201 (30.9%)	0.64 (0.50–0.82)	
	AA	34 (5.2%)	16 (2.5%)	0.38 (0.21–0.70)	
rs2290400	TT	379 (58.3%)	292 (44.9%)	1	<0.0001*
	TC	232 (35.7%)	293 (45.1%)	1.64 (1.30–2.07)	
	CC	39 (6%)	65 (10%)	2.21 (1.44–3.39)	
rs77681114	GG	392 (60.3%)	435 (66.9%)	1	0.0025*
	GA	235 (36.1%)	207 (31.9%)	0.79 (0.62–1.00)	
	AA	23 (3.5%)	8 (1.2%)	0.30 (0.13–0.68)	
rs1103577	TT	351 (54%)	279 (42.9%)	1	<0.0001*
	TC	265 (40.8%)	306 (47.1%)	1.51 (1.20–1.90)	
	CC	34 (5.2%)	65 (10%)	2.49 (1.59–3.89)	
rs2298803	TT	298 (45.9%)	294 (45.2%)	1	0.220
	TC	295 (45.4%)	281 (43.2%)	0.95 (0.76–1.20)	
	CC	57 (8.8%)	75 (11.5%)	1.34 (0.91–1.96)	

SNP, single nucleotide polymorphism; OR, odds ratio; CI, confidence interval; $p < 0.05$ indicates statistical significance.

Finally, the participants were divided into four subgroups according to the age and smoking status (Table 6). The rs8067378, rs2305480, and rs2290400 remained significant in each subgroup ($p < 0.05$). However, rs77681114 had no protective influence on the NSCLC in smokers ($p > 0.05$), and rs1103577 was not linked to NSCLC risk in nonsmokers ($p > 0.05$). Finally, we also looked at the connection between SNPs and risk of disease in patients with adenocarcinoma and squamous cell carcinoma, respectively (Table 7). All of the SNPs remained significant except rs2305480 ($p < 0.05$). The rs2305480 had no protective role for risk of squamous cell carcinoma ($p > 0.05$).

Discussion

Pyroptosis is a new type of programmed cell death, which has been widely studied in various diseases in recent years, and the importance of this pathway to regulate tissue development and

TABLE 5 Association between SNPs and NSCLC risk in genetic models.

SNP	Model	Genotype	Control	Case	OR (95% CI)	*p*
rs8067378	Dominant	AA	385 (59.2%)	454 (69.8%)	1	<0.0001*
		AG-GG	265 (40.8%)	196 (30.1%)	0.61 (0.49–0.77)	
	Recessive	AA-AG	617 (94.9%)	631 (97.1%)	1	0.041*
		GG	33 (5.1%)	19 (2.9%)	0.55 (0.31–0.99)	
	Log-additive	—	—	—	0.65 (0.54–0.80)	<0.0001*
rs2305480	Dominant	GG	362 (55.7%)	433 (66.6%)	1	<0.0001*
		GA-AA	288 (44.3%)	217 (33.4%)	0.61 (0.48–0.77)	
	Recessive	GG-GA	616 (94.8%)	634 (97.5%)	1	0.0076*
		AA	34 (5.2%)	16 (2.5%)	0.45 (0.25–0.83)	
	Log-additive	—	—	—	0.63 (0.52–0.78)	<0.0001*
rs2290400	Dominant	TT	379 (58.3%)	292 (44.9%)	1	<0.0001*
		TC-CC	271 (41.7%)	358 (55.1%)	1.72 (1.38–2.15)	
	Recessive	TT-TC	611 (94%)	585 (90%)	1	0.0056*
		CC	39 (6%)	65 (10%)	1.78 (1.18–2.70)	
	Log-additive	—	—	—	1.56 (1.30–1.85)	<0.0001*
rs77681114	Dominant	GG	392 (60.3%)	435 (66.9%)	1	0.013*
		GA-AA	258 (39.7%)	215 (33.1%)	0.75 (0.59–0.94)	
	Recessive	GG-GA	627 (96.5%)	642 (98.8%)	1	0.0047*
		AA	23 (3.5%)	8 (1.2%)	0.33 (0.15–0.75)	
	Log-additive	—	—	—	0.72 (0.59–0.89)	0.0023*
rs1103577	Dominant	TT	351 (54%)	279 (42.9%)	1	<0.0001*
		TC-CC	299 (46%)	371 (57.1%)	1.62 (1.30–2.03)	
	Recessive	TT-TC	616 (94.8%)	585 (90%)	1	0.0009*
		CC	34 (5.2%)	65 (10%)	2.04 (1.32–3.14)	
	Log-additive	—	—	—	1.54 (1.29–1.85)	<0.0001*
rs2298803	Dominant	TT	298 (45.9%)	294 (45.2%)	1	0.88
		TC-CC	352 (54.1%)	356 (54.8%)	1.02 (0.82–1.27)	
	Recessive	TT-TC	593 (91.2%)	575 (88.5%)	1	0.088
		CC	57 (8.8%)	75 (11.5%)	1.37 (0.95–1.97)	
	Log-additive	—	—	—	1.08 (0.91–1.28)	0.37

SNP, single nucleotide polymorphism; OR, odds ratio; CI, confidence interval; $p < 0.05$ indicates statistical significance.

homeostasis has also received attention (Jia et al., 2023). Pyroptosis -related factors have a dual mechanism of promoting or inhibiting tumorigenesis, and can affect tumor progression by modulating malignant phenotypes such as cell morphology, proliferation, invasion, migration, and chemotherapy tolerance through multiple molecular signaling pathways, and may affect a patient's prognosis (Frank and Vince, 2019). In our study, we identified five SNPs associated with increased or reduced risk of NSCLC associated with the pyroptosis-related genes *GSDMB*, *GSDMC*, and *AIM2*, which may shed light on the relationship between pyroptosis and NSCLC pathogenesis, as well as provide theoretical foundations for detecting and preventing the disease early.

GSDMB, located at 17q21, encodes the GSDMB that participate in pyroptosis as a key molecule. Ding's group reported that GSDMB could be cut by Caspase-1 and released the N-terminus domain that induce cell pytoptosis (Ding et al., 2016), while Chen's team demonstrated that GSDMB could not form pores on cytomembrane, but promoted non-classical pyroptosis through an enhancement of caspase-4 activity (Chen et al., 2019). With the research development, Chao' lab argued that GSDMB could not

TABLE 6 Associations of candidate SNPs with NSCLC risk in four subgroups.

SNP	Model	Genotype	≥50		<50		Smokers		Nonsmokers	
			OR (95% CI)	*p*	OR (95% CI)	*p*	OR (95% CI)	*p*	OR (95% CI)	*p*
rs8067378	Dominant	AA	1		1	0.051	1	0.001*	1	0.012*
		AG-GG	0.60 (0.46–0.79)	0.0002*	0.62 (0.38–1.00)		0.62 (0.47–0.83)		0.60 (0.40–0.90)	
	Recessive	AA-AG	1		/	/	1	0.8	/	/
		GG	0.95 (0.50–1.80)	0.87	/		0.92 (0.48–1.76)		/	
	Log-additive	/	0.69 (0.55–0.86)	0.0011*	0.54 (0.35–0.82)	0.0032*	0.71 (0.56–0.90)	0.0043*	0.55 (0.38–0.79)	0.001*
rs2305480	Dominant	GG	1	0.0005*	1	<0.0001*	1	0.0013*	1	0.0001*
		GA-AA	0.62 (0.47–0.81)		0.22 (0.11–0.46)		0.57 (0.40–0.80)		0.46 (0.31–0.68)	
	Recessive	GG-GA	1	0.013*	1	0.42	1	0.039*	1	0.083
		AA	0.45 (0.24–0.86)		0.50 (0.09–2.81)		0.47 (0.22–0.98)		0.41 (0.14–1.18)	
	Log-additive	/	0.64 (0.51–0.81)	0.0001*	0.28 (0.15–0.55)	<0.0001*	0.61 (0.46–0.81)	0.0006*	0.50 (0.35–0.70)	<0.0001*
rs2290400	Dominant	TT	1	0.0019*	1	<0.0001*	1	<0.0001*	1	0.007*
		TC-CC	1.49 (1.16–1.91)		4.01 (2.34–6.88)		1.89 (1.42–2.51)		1.67 (1.15–2.42)	
	Recessive	TT-TC	1	0.64	1	<0.0001*	1	0.14	1	0.0097*
		CC	1.12 (0.69–1.83)		5.76 (2.30–14.42)		1.49 (0.88–2.53)		2.36 (1.21–4.61)	
	Log-additive	/	1.31 (1.07–1.61)	0.0075*	3.20 (2.10–4.87)	<0.0001*	1.61 (1.28–2.03)	<0.0001*	1.58 (1.19–2.10)	0.0015*
rs77681114	Dominant	GG	1	0.29	1	0.0002*	1	0.27	1	0.0031*
		GA-AA	0.86 (0.66–1.13)		0.41 (0.25–0.66)		0.84 (0.62–1.14)		0.57 (0.40–0.83)	
	Recessive	GG-GA	1	0.0066*	/	/	1	0.11	1	0.017*
		AA	0.34 (0.15–0.78)		/	/	0.43 (0.14–1.27)		0.25 (0.07–0.89)	
	Log-additive	/	0.80 (0.63–1.02)	0.072	/	/	0.82 (0.62–1.08)	0.15	0.57 (0.41–0.80)	0.0008*
rs1103577	Dominant	TT	1	0.0007*	1	0.016*	1	<0.0001*	1	0.22
		TC-CC	1.55 (1.20–1.99)		1.81 (1.11–2.93)		2.40 (1.81–3.19)		0.79 (0.55–1.15)	
	Recessive	TT-TC	1	0.0011*	1	0.4	1	0.008*	1	0.054
		CC	2.22 (1.35–3.65)		1.46 (0.59–3.61)		2.13 (1.20–3.79)		1.88 (0.98–3.61)	
	Log-additive	/	1.52 (1.24–1.86)	0.0001*	1.58 (1.07–2.35)	0.021*	2.01 (1.59–2.54)	<0.0001*	0.99 (0.74–1.32)	0.94

SNP, single nucleotide polymorphism; OR, odds ratio; CI, confidence interval; *$p < 0.05$ indicates statistical significance.

10.3389/fgene.2023.1212465

TABLE 7 Association between Candidate SNPs and risk of Adenocarcinoma and Squamous cell carcinoma.

SNP	Model	Genotype	Adenocarcinoma		Squamous cell carcinoma	
			OR (95% CI)	*p*	OR (95% CI)	*p*
rs8067378	Dominant	AA	1	0.0007*	1	0.0001*
		AG-GG	0.61 (0.46–0.82)		0.53 (0.39–0.72)	
	Recessive	AA-AG	1	0.03*	1	0.57
		GG	0.43 (0.19–0.98)		0.82 (0.40–1.66)	
	Log-additive	—	0.63 (0.49–0.82)	0.0003*	0.62 (0.47–0.81)	0.0003*
rs2305480	Dominant	GG	1	<0.0001*	1	0.67
		GA-AA	0.54 (0.40–0.72)		0.93 (0.68–1.29)	
	Recessive	GG-GA	1	0.0079*	1	0.18
		AA	0.34 (0.14–0.83)		0.61 (0.28–1.30)	
	Log-additive	—	0.56 (0.43–0.72)	<0.0001*	0.89 (0.68–1.16)	0.4
rs2290400	Dominant	TT	1	0.0002*	1	0.024*
		TC-CC	1.67 (1.27–2.19)		1.40 (1.05–1.88)	
	Recessive	TT-TC	1	0.073	1	0.0023*
		CC	1.59 (0.96–2.62)		2.24 (1.34–3.73)	
	Log-additive	—	1.49 (1.21–1.84)	0.0002*	1.43 (1.14–1.79)	0.0021*
rs77681114	Dominant	GG	1	0.0003*	1	0.14
		GA-AA	0.59 (0.44–0.79)		1.25 (0.93–1.69)	
	Recessive	GG-GA	1	0.041	1	0.012*
		AA	0.39 (0.15–1.04)		0.21 (0.05–0.92)	
	Log-additive	—	0.61 (0.47–0.79)	0.0001*	1.09 (0.83–1.44)	0.52
rs1103577	Dominant	TT	1	0.0047*	1	<0.0001*
		TC-CC	1.49 (1.13–1.96)		1.91 (1.41–2.59)	
	Recessive	TT-TC	1	0.026*	1	0.012*
		CC	1.80 (1.08–3.00)		2.03 (1.18–3.50)	
	Log-additive	—	1.43 (1.15–1.79)	0.0013*	1.70 (1.34–2.16)	<0.0001*

SNP, single nucleotide polymorphism; OR, odds ratio; CI, confidence interval; *$p < 0.05$ indicates statistical significance.

be the substrate for human Caspase-1/4/5/11 due to lacking of the specific interdomain, but it could be cleaved by Caspase-3/6/7, which indicating that an apoptosis-pyroptosis cross-talk may be occurring (Chao et al., 2017). Also, lots of studies have shown that the genetic polymorphisms were linked to risk of autoimmune disease. Imraish et al. (2022) reported that *GSDMB*-rs7216389 has potential influence on IgE levels of patients with asthma in Jordanian population. Shamsi et al. (2023) reported that *GSDMB*-rs4795400, rs2305479, and rs12450091 were associated with risk of allergic rhinitis. As for cancer studies about *GSDMB* polymorphisms, Lutkowska et al. (2017) found that rs8067378 A>G variant may elevate the expression of GSDMB and increased the risk of the cervical squamous cell carcinomas in a Polish population, while Li et al. (2022) further reported that rs8067378 was a risk-reducing variant for cervical squamous intraepithelial lesion. We for the first time identified that a declined NSCLC risk was correlated with rs8067378 and rs2305480 in GSDMB, whereas an increased risk was associated with rs2290400. These results provided new evidence for the involvement of *GSDMB* in development and progression of NSCLC, while the molecular mechanism needed to be further explored. It is worth noting that rs2305480 is a missense variant and leads to Pro > Ser. Pro is a non-polar and hydrophobic amino acid, while Ser is a polar uncharged amino acid. We supposed that rs2305480 may has an effect on the progression of the disease through changing the conformation of GSDMB and its function in cell pyroptosis.

It has recently been revealed that GSDMC plays a role in cell pyroptosis as a member of the GSDM family. Hou's group found that Caspase-8 can cut GSDMC in hypoxic breast cancer cells, and its expression level was mediated by the PD-L1, following by TNF-α induced pyroptosis (Hou et al., 2020). Moreover, Miguchi et al.

established that an increase in GSDMC expression was linked to mutations in TGF-β receptor type II, and leading to a promotion of cell growth in colorectal cancer and xenograft tumor volum *in vivo* (Miguchi et al., 2016). A similar tumor-promoting role in lung adenocarcinoma was demonstrated by Wei's Lab, upregulation of GSDMC was linked to poor outcomes, making it be a promising target for the disease (Wei et al., 2020). Furthermore, Yan's group pointed that GSDMC functioned as an oncogene that promoting the cell proliferation and migration in pancreatic adenocarcinoma (Yan et al., 2022). In these studies, it was demonstrated that GSDMC played a crucial role in cancer development, especially in pyroptosis. However, little study focused on the genetic polymorphisms in *GSDMC*. Among Chinese Han, Wu et al. found that GSDMC-rs77681114 significantly decreased risk of lumbar disc herniation (Wu et al., 2020). We demonstrated that the variant rs77681114G>A is protective against NSCLC. However, smokers were not significantly affected by rs77681114 in a stratification analysis. We supposed that the protective role of this variant might be neutralized with cigarette smoking in NSCLC patients, while the hypothesis and detailed mechanisms should be verified and investigated in further studies.

AIM2 belongs to a family of inflammasomes and function as an intracellular DNA receptor that recognize double-stranded DNA released into the cytoplasm, activate downstream related effector proteins and induce cell pyroptosis (Wang et al., 2020). In addition to activating inflammasomes for immune function, studies have found that AIM2 also has the dual effects of promoting or inhibiting cancer development. Choubey et al. (2000) firstly reported the tumor suppressor role of AIM2, upregulation of AIM2 inhibited the cell proliferation and enhanced cell death in melanoma. Moreover, AIM2 was low expressed and played a tumor-inhibiting role in colon, liver, renal, breast and prostate cancers (Qin et al., 2022). In contrast, Farshchian et al. (2017) revealed a pro-tumorigenic role of AIM2, downregulation of AIM2 reduced the viability and invasion of cutaneous squamous cell carcinoma. Zhang et al. (2019) and Qi et al. (2020) also reported the oncogenic role of AIM2 in NSCLC through the inflammasome and modulation of mitochondrial dynamics, respectively. According to our understanding, few study focused on the AIM2 polymorphisms and cancer risk. We genotyped two SNPs, rs1103577, and rs2298803 in *AIM2*, and observed that rs1103577T>C was a risky variant for NSCLC. A subgroup analysis revealed that rs1103577T>C increased NSCLC risk in smokers but not in non-smokers, suggesting that rs1103577 may have interaction with smoking in the onset or development of NSCLC. The results provided an important detection site for early prevention of NSCLC.

In an association study, population stratification may lead to false positive or negative results. Thus, we stratified our analysis according to age and smoking status. Smokers and nonsmokers exhibited different results for GSMDC-rs77681114 and AIM2-rs1103577, suggesting that these two variants might interact with smoking in NSCLC progression. In addition, we also evaluated the association of candidate SNPs and different pathological type of NSCLC. The *GSDMB*-rs2305480 was a protective variant for adenocarcinoma, but not for squamous cell carcinoma, which could be explained by the different pathogenesis between the two pathological types of NSCLC.

Although the present study revealed the association between pyroptosis-related genes and NSCLC, there are also some potential limitations. Firstly, history of other lung diseases and family history of cancer might have associations with risk of the NSCLC; however, we have no related information to analyze, because the participants were collected in a very long time period, we did not design the factors from the very beginning. Secondly, the SNPs identified here could only represent the Chinese Han population, further validation study need to be done in other populations. Thirdly, our results needed to be further validated in functional studies.

In conclusion, we demonstrated that *GSDMB*-rs8067378, rs2305480, and *GSMDC*-rs77681114 were linked to a reduced NSCLC risk, while *GSDMB*-rs2290400 and *AIM2*-rs1103577 were related to an increased risk of the disease. Our findings provided new insights into the roles of pyroptosis-related genes in NSCLC, as well as new factors to be considered for assessing the risk of developing this cancer.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Department of Shanxi Province Cancer Hospital. The patients/participants provided their written informed consent to participate in this study.

Author contributions

XZ: investigation, formal analysis, writing- original draft preparation. RL: validation, supervision, writing- review and editing, funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Key Program of Medicine and Science Foundation of Hebei Province (20230874), the Program for Young Scholars of Medicine and Science Foundation of Hebei Province (20180576, 20120354).

Conflict of interest

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

Publisher's note

References

Broz, P., PelegríN, P., and Shao, F. (2020). The gasdermins, a protein family executing cell death and inflammation. *Nat. Rev. Immunol.* 20, 143–157. doi:10.1038/s41577-019-0228-2

Case, C. L., Kohler, L. J., Lima, J. B., Strowig, T., De Zoete, M. R., Flavell, R. A., et al. (2013). Caspase-11 stimulates rapid flagellin-independent pyroptosis in response to *Legionella pneumophila*. *Proc. Natl. Acad. Sci. U. S. A.* 110, 1851–1856. doi:10.1073/pnas.1211521110

Chao, K. L., Kulakova, L., and Herzberg, O. (2017). Gene polymorphism linked to increased asthma and IBD risk alters gasdermin-B structure, a sulfatide and phosphoinositide binding protein. *Proc. Natl. Acad. Sci. U. S. A.* 114, 1128–1137. doi:10.1073/pnas.1616783114

Chen, H. X., Ren, N. X., Yang, J., Chen, J. N., Lu, Q. X., Feng, Y. R., et al. (2023). Associations of genetic variations in pyroptosis related genes with acute adverse events in postoperative rectal cancer patients receiving concurrent chemoradiotherapy. *Zhonghua Zhong Liu Za Zhi* 45, 146–152. doi:10.3760/cma.j.cn112152-20220622-00447

Chen, Q., Shi, P., Wang, Y., Zou, D., Wu, X., Wang, D., et al. (2019). GSDMB promotes non-canonical pyroptosis by enhancing caspase-4 activity. *J. Mol. Cell Biol.* 11, 496–508. doi:10.1093/jmcb/mjy056

Choubey, D., Walter, S., Geng, Y., and Xin, H. (2000). Cytoplasmic localization of the interferon-inducible protein that is encoded by the AIM2 (absent in melanoma) gene from the 200-gene family. *FEBS Lett.* 474, 38–42. doi:10.1016/s0014-5793(00)01571-4

D'Souza, C. A., and Heitman, J. (2001). Dismantling the cryptococcus coat. *Trends Microbiol.* 9, 112–113. doi:10.1016/s0966-842x(00)01945-4

Ding, J., Wang, K., Liu, W., She, Y., Sun, Q., Shi, J., et al. (2016). Pore-forming activity and structural autoinhibition of the gasdermin family. *Nature* 535, 111–116. doi:10.1038/nature18590

Farshchian, M., Nissinen, L., SiljamäKI, E., Riihilä, P., Piipponen, M., Kivisaari, A., et al. (2017). Tumor cell-specific AIM2 regulates growth and invasion of cutaneous squamous cell carcinoma. *Oncotarget* 8, 45825–45836. doi:10.18632/oncotarget.17573

Feng, T., Feng, N., Zhu, T., Li, Q., Zhang, Q., Wang, Y., et al. (2020). A SNP-mediated lncRNA (LOC146880) and microRNA (miR-539-5p) interaction and its potential impact on the NSCLC risk. *J. Exp. Clin. Cancer Res.* 39, 157. doi:10.1186/s13046-020-01652-5

Figueira, M. B. A., De Lima, D. S., Boechat, A. L., Filho, M., Antunes, I. A., Matsuda, J. D. S., et al. (2021). Single-nucleotide variants in the AIM2 - absent in melanoma 2 gene (rs1103577) associated with protection for tuberculosis. *Front. Immunol.* 12, 604975. doi:10.3389/fimmu.2021.604975

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

Gabriel, S., Ziaugra, L., and Tabbaa, D. (2009). SNP genotyping using the Sequenom MassARRAY iPLEX platform. *Curr. Protoc. Hum. Genet.* 2, 2.12. doi:10.1002/0471142905.hg0212s60

Hou, J., Zhao, R., Xia, W., Chang, C. W., You, Y., Hsu, J. M., et al. (2020). PD-L1-mediated gasdermin C expression switches apoptosis to pyroptosis in cancer cells and facilitates tumour necrosis. *Nat. Cell Biol.* 22, 1264–1275. doi:10.1038/s41556-020-0575-z

Imraish, A., Abu-Thiab, T., Alhindi, T., and Zihlif, M. (2022). GSDM gene polymorphisms regulate the IgE level in asthmatic patients. *PLoS One* 17, 0274951. doi:10.1371/journal.pone.0274951

Jia, Y., Wang, X., Deng, Y., Li, S., Xu, X., Qin, Y., et al. (2023). Pyroptosis provides new strategies for the treatment of cancer. *J. Cancer* 14, 140–151. doi:10.7150/jca.77965

Jones, G. S., and Baldwin, D. R. (2018). Recent advances in the management of lung cancer. *Clin. Med. (Lond)* 18, s41–s46. doi:10.7861/clinmedicine.18-2-s41

Jonna, S., and Subramaniam, D. S. (2019). Molecular diagnostics and targeted therapies in non-small cell lung cancer (NSCLC): An update. *Discov. Med.* 27, 167–170.

Karunas, A. S., Fedorova, Y. Y., Gimalova, G. F., Etkina, E. I., and Khusnutdinova, E. K. (2021). Association of gasdermin B gene GSDMB polymorphisms with risk of allergic diseases. *Biochem. Genet.* 59, 1527–1543. doi:10.1007/s10528-021-10073-8

Li, L., Li, Y., and Bai, Y. (2020). Role of GSDMB in pyroptosis and cancer. *Cancer Manag. Res.* 12, 3033–3043. doi:10.2147/CMAR.S246948

Li, S., Li, X., Zhang, S., Feng, Y., Jia, T., Zhu, M., et al. (2022). Association between GSDMB gene polymorphism and cervical cancer in the northeast Chinese han population. *Front. Genet.* 13, 860727. doi:10.3389/fgene.2022.860727

Luo, J., Martucci, V. L., Quandt, Z., Groha, S., Murray, M. H., Lovly, C. M., et al. (2021). Immunotherapy-mediated thyroid dysfunction: Genetic risk and impact on outcomes with PD-1 blockade in non-small cell lung cancer. *Clin. Cancer Res.* 27, 5131–5140. doi:10.1158/1078-0432.CCR-21-0921

Lutkowska, A., Roszak, A., Lianeri, M., Sowińska, A., Sotiri, E., and Jagodziński, P. P. (2017). Analysis of rs8067378 polymorphism in the risk of uterine cervical cancer from a polish population and its impact on gasdermin B expression. *Mol. Diagn Ther.* 21, 199–207. doi:10.1007/s40291-017-0256-1

Man, S. M., and Kanneganti, T. D. (2016). Converging roles of caspases in inflammasome activation, cell death and innate immunity. *Nat. Rev. Immunol.* 16, 7–21. doi:10.1038/nri.2015.7

Miguchi, M., Hinoi, T., Shimomura, M., Adachi, T., Saito, Y., Niitsu, H., et al. (2016). Gasdermin C is upregulated by inactivation of transforming growth factor β receptor type II in the presence of mutated apc, promoting colorectal cancer proliferation. *PLoS One* 11, e0166422. doi:10.1371/journal.pone.0166422

Qi, M., Dai, D., Liu, J., Li, Z., Liang, P., Wang, Y., et al. (2020). AIM2 promotes the development of non-small cell lung cancer by modulating mitochondrial dynamics. *Oncogene* 39, 2707–2723. doi:10.1038/s41388-020-1176-9

Qin, Y., Pan, L., Qin, T., Ruan, H., Zhang, Y., Zhang, Y., et al. (2022). Pan-cancer analysis of AIM2 inflammasomes with potential implications for immunotherapy in human cancer: A bulk omics research and single cell sequencing validation. *Front. Immunol.* 13, 998266. doi:10.3389/fimmu.2022.998266

Rodriguez-Canales, J., Parra-Cuentas, E., and Wistuba, II (2016). Diagnosis and molecular classification of lung cancer. *Cancer Treat. Res.* 170, 25–46. doi:10.1007/978-3-319-40389-2_2

Shamsi, B. H., Chen, H., Yang, X., Liu, M., and Liu, Y. (2023). Association between polymorphisms of the GSDMB gene and allergic rhinitis risk in the Chinese population: A case-control study. *J. Asthma*, 1–10. doi:10.1080/02770903.2023.2185893

Sharma, B. R., Karki, R., and Kanneganti, T. D. (2019). Role of AIM2 inflammasome in inflammatory diseases, cancer and infection. *Eur. J. Immunol.* 49, 1998–2011. doi:10.1002/eji.201848070

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer statistics, 2021. *CA Cancer J. Clin.* 71, 7–33. doi:10.3322/caac.21654

Torre, L. A., Siegel, R. L., and Jemal, A. (2016). Lung cancer statistics. *Adv. Exp. Med. Biol.* 893, 1–19. doi:10.1007/978-3-319-24223-1_1

Wang, B., Bhattacharya, M., Roy, S., Tian, Y., and Yin, Q. (2020). Immunobiology and structural biology of AIM2 inflammasome. *Mol. Asp. Med.* 76, 100869. doi:10.1016/j.mam.2020.100869

Wang, Y., Gao, W., Shi, X., Ding, J., Liu, W., He, H., et al. (2017). Chemotherapy drugs induce pyroptosis through caspase-3 cleavage of a gasdermin. *Nature* 547, 99–103. doi:10.1038/nature22393

Wei, J., Xu, Z., Chen, X., Wang, X., Zeng, S., Qian, L., et al. (2020). Overexpression of GSDMC is a prognostic factor for predicting a poor outcome in lung adenocarcinoma. *Mol. Med. Rep.* 21, 360–370. doi:10.3892/mmr.2019.10837

Wu, F., Fan, J., He, Y., Xiong, A., Yu, J., Li, Y., et al. (2021). 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, J., Sun, Y., Xiong, Z., Liu, J., Li, H., Liu, Y., et al. (2020). Association of GSDMC polymorphisms with lumbar disc herniation among Chinese Han population. *Int. J. Immunogenet* 47, 546–553. doi:10.1111/iji.12488

Yan, C., Niu, Y., Li, F., Zhao, W., and Ma, L. (2022). System analysis based on the pyroptosis-related genes identifies GSDMC as a novel therapy target for pancreatic adenocarcinoma. *J. Transl. Med.* 20, 455. doi:10.1186/s12967-022-03632-z

Zhang, M., Jin, C., Yang, Y., Wang, K., Zhou, Y., Zhou, Y., et al. (2019). AIM2 promotes non-small-cell lung cancer cell growth through inflammasome-dependent pathway. *J. Cell Physiol.* 234, 20161–20173. doi:10.1002/jcp.28617

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 22 June 2023
DOI 10.3389/fgene.2023.1200262

OPEN ACCESS

EDITED BY
Mehdi Pirooznia,
Johnson & Johnson, United States

REVIEWED BY
Sacide Pehlivan,
Istanbul University, Türkiye
Guntulu Ak,
Eskişehir Osmangazi University, Türkiye

*CORRESPONDENCE
Aslı Tetik Vardarlı,
asli.tetik.vardarli@ege.edu.tr
Su Ozgur,
suozgur35@gmail.com

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

RECEIVED 04 April 2023
ACCEPTED 12 June 2023
PUBLISHED 22 June 2023

CITATION
Tetik Vardarlı A, Ozgur S, Goksel T, Korba K, Karakus HS, Asık A, Pelit L and Gunduz C (2023), Conversion of specific lncRNAs to biomarkers in exhaled breath condensate samples of patients with advanced stage non-small-cell lung cancer.
Front. Genet. 14:1200262.
doi: 10.3389/fgene.2023.1200262

Conversion of specific lncRNAs to biomarkers in exhaled breath condensate samples of patients with advanced stage non-small-cell lung cancer

Aslı Tetik Vardarlı [1,2*†], Su Ozgur [2,3*†], Tuncay Goksel [2,4], Korcan Korba [5,6], Hardar Soydaner Karakus [2,4], Aycan Asık [6], Levent Pelit [2,7] and Cumhur Gunduz [1,2]

[1]Department of Medical Biology, Faculty of Medicine, Ege University, Izmir, Türkiye, [2]EgeSAM-Ege University Translational Pulmonary Research Center, Izmir, Türkiye, [3]Regional Hub for Cancer Registration in Northern Africa, Central and Western Asia, WHO/IARC-GICR, Izmir, Türkiye, [4]Department of Pulmonary Medicine, Faculty of Medicine, Ege University, Izmir, Türkiye, [5]Department of Chemical Engineering, Faculty of Engineering, Ege University, Izmir, Türkiye, [6]Department of Medical Biology, Faculty of Medicine, Mugla Sıtkı Kocman University, Mugla, Türkiye, [7]Department of Chemistry, Faculty of Science, Ege University, Izmir, Türkiye

Objectives: Lung cancer (LC) is one of the most prevalent cancers with the highest fatality rate worldwide. Long noncoding RNAs (lncRNAs) are being considered potential new molecular targets for early diagnosis, follow-up, and individual treatment decisions in LC. Therefore, this study evaluated whether lncRNA expression levels obtained from exhaled breath condensate (EBC) samples play a role in the occurrence of metastasis in the diagnosis and follow-up of patients with advanced lung adenocarcinoma (LA).

Methods: A total of 40 patients with advanced primary LA and 20 healthy controls participated in the study. EBC samples were collected from patients (during diagnosis and follow-up) and healthy individuals for molecular analysis. Liquid biopsy samples were also randomly obtained from 10 patients with LA and 10 healthy people. The expression of lncRNA genes, such as MALAT1, HOTAIR, PVT1, NEAT1, ANRIL, and SPRY4-IT1 was analyzed using cfRNA extracted from all clinical samples.

Results: In the diagnosis and follow-up of patients with LA, lncRNA HOTAIR (5-fold), PVT1 (7.9-fold), and NEAT1 (12.8-fold), PVT1 (6.8-fold), MALAT1 (8.4-fold) expression levels were significantly higher than those in healthy controls, respectively. Additionally, the distinct lncRNA expression profiles identified in EBC samples imply that decreased ANRIL–NEAT1 and increased ANRIL gene expression levels can be used as biomarkers to predict the development of bone and lung metastases, respectively.

Conclusion: EBC is an innovative, easily reproducible approach for predicting the development of metastases, molecular diagnosis, and follow-up of LC. EBC has shown potential in elucidating the molecular structure of LC, monitoring changes, and discovering novel biomarkers.

KEYWORDS

lung cancer, cfRNA, lncRNA, ebc, biomarker

Introduction

Lung cancer (LC) is among the most common causes of mortality worldwide, accounting for 18.0% of all cancer deaths. The latest global cancer data indicated that the cancer burden increased to 19.3 million new cases and 10.0 million cancer deaths in 2020 and accounting for 2.21 million new cases and 1.80 million cancer deaths (Ferlay et al., 2021). LC is divided into two main classes: small-cell lung cancer (SCLC) and non-small-cell lung cancer (NSCLC). SCLC is the fastest growing and metastasizing kind of LC, accounting for 10%–15% of all cases, and NSCLC represents for 85%–90%. Adenocarcinoma, squamous cell carcinoma, and large-cell carcinoma are the most prevalent NSCLC subtypes (Hirsch et al., 2017). The adenocarcinoma subtype constitutes 40% of this group. Two-thirds of patients with LC are detected at an advanced stage, limiting the surgical therapeutic options (Shah and Masters, 2020). Immunotherapy and targeted medicines have been developed based on histology and molecular analysis in recent years, which have shown clinical benefits in patients with NSCLC (Shroff et al., 2018). Despite the development of new treatments, 5-year survival rates ranged from 15% to 20%, depending on the cancer stages (Lung Cancer Survival Rates). Drug resistance, metastasis development, and recurrence are defined as major factors that prevent the patients' survival. Therefore, the identification of molecular mechanisms of these factors has become the main goal nowadays (Shroff et al., 2018).

While cancer has traditionally been defined as a set of diseases caused by accumulating genetic mutations as major causes of neoplasia, this paradigm has now been expanded to include the disruption of epigenetic regulatory mechanisms commonly occurring in cancer (Gutschner and Diederichs, 2012). Similar to genetic changes, epigenetic events affect almost every stage of tumor development. Understanding the epigenetic changes associated with cancer onset, progression, and metastasis is critical to improving our ability to successfully diagnose, treat, and prevent.

Noncoding RNAs are involved in epigenetic regulation through transcriptional and posttranscriptional mechanisms (Ferreira and Esteller, 2018). Long noncoding RNAs (lncRNAs) play important cellular and physiological roles, such as chromatin dynamics, proliferation, differentiation, developmental regulation, and gene expression (Li et al., 2014). The specificity of lncRNA expression attests to its essential roles in regulating organismal function and repairing pathological processes (Smolarz et al., 2021). The lncRNA properties, such as tissue or cellular specificity and regulation of gene expression at transcriptional and posttranscriptional levels, suggest that lncRNAs may be important in the formation of malignant tumors. Additionally, some lncRNAs are regulated by oncogene products or cancer transformation suppressors, indicating that they indirectly fulfill tumorigenic functions (Huarte, 2015).

Current findings have implicated aberrantly expressed lncRNAs such as MALAT1, HOTAIR, PVT1, NEAT1, ANRIL, and SPRY4-IT1 in the development and progression of numerous malignancies, including LC. Additionally, these lncRNAs can be used as biomarkers for cancer diagnosis, prognosis, and follow-up (Gutschner and Diederichs, 2012; Hu et al., 2016; Sun et al., 2016; Wang et al., 2021; Zeng et al., 2021). Identification and characterization of lncRNAs have opened a new era in cancer management. lncRNAs can serve as cancer biomarkers for developing new diagnostic tools and the prediction of prognosis and as potential targets for new cancer treatment strategies. The expression of particular lncRNAs is a distinguishing feature of different tissues and even individual cells. Although only a few lncRNAs are typically found in the body, the majority of individuals have extremely high concentrations of them. The lncRNAs that are tissue- or cell-specific have drastically reduced levels of expression (Jiang et al., 2016).

Tumor tissue biopsies, which are considered the gold standard, are crucial for detecting molecular alterations that must be examined to make the correct diagnosis and treatment decisions throughout the disease progression. However, this type of biopsies has restrictions, as they are invasive and insufficient for revealing the real-time molecular structure and monitoring changes in disease. To repeat and monitor these molecular analyses, novel noninvasive analysis methods are required. In this context, liquid biopsy is the most promising technique. Liquid biopsy samples, the first noninvasive procedure, contain nucleic acids (cell-free DNA and RNA [cfDNA and cfRNA]) that are freely circulating outside the cell. Nowadays, cfDNA and cfRNA analyses are employed as a noninvasive tool for identifying multiple genetic alterations associated with cancer, which is, therefore, being recognized as an alternative material widely used for early cancer diagnosis and targeted treatment decisions (Sorber et al., 2017). Current advances in the analysis of lncRNA processes in liquid biopsies make these molecules important clinical targets for cancer treatment (Schmitt and Chang, 2016). The introduction of personalized treatment options and immunotherapy because of advancements in molecular oncology is the driving force behind efforts to improve molecular biomarker tests and liquid biopsy, specifically in detecting these biomarkers.

Exhaled breath condensate (EBC) samples are an additional noninvasive approach investigated in recent years to determine the real-time and repeatable molecular structure of LC. EBC is a sample matrix consisting of aerosolized droplets from the alveolar fluid that are further diluted in the distal and proximal respiratory tracts and collected by condensation during tidal breathing (Davis et al., 2012). Due to its compound composition, EBC can be used to diagnose LC. EBC samples can yield components, such as protein, DNA, RNA, and volatile organic chemicals. EBC biomarkers are found in the respiratory air because of direct breathing. They are also directly implicated in the metabolic activities of cancer cells and provide the opportunity for genomics, transcriptomics, epigenomics, proteomics, and metabolomics examinations (Campbell et al., 2008). With the advancements in technologies, genetic analysis studies can be conducted using EBC samples and liquid biopsies in LC (Tetik Vardarli et al., 2020). The concordance of DNA and RNA analysis results from the liquid biopsy and EBC samples demonstrates that EBC is a promising noninvasive approach for LC molecular diagnosis and follow-up. Therefore, this study determined whether the expression profiles of certain lncRNAs, which are candidates for new molecular targets in the diagnosis, follow-up, and individual treatment decisions of LC, could be detected in EBC samples. Moreover, the use of MALAT1, HOTAIR, PVT1, NEAT1, ANRIL, and SPRY4-IT1 expression levels detected in cfRNA from EBC samples was also investigated for the molecular diagnosis and follow-up of LC.

Materials and methods

Study population

This case-control study included 40 primary adenocarcinoma patients (stage IIIB or IV) with 1–3 synchronous metastases and without metastases, who were being followed up in the Chest Diseases Department of Ege University Medical School between 2021 and 2022. Adenocarcinoma was diagnosed in the Pathology Department of Ege University Medical School in Izmir, Türkiye. This study also included 20 age- and gender-matched healthy individuals. The control participants were recruited from healthy volunteers without a history of cancer. The difference between two independent means (two groups) approximation was used for sample size calculation. The effect size between the two groups (cancer patients and healthy individuals) was calculated to be 0.7146, and the *post hoc* power was achieved to be 0.82. Detailed demographic and clinical characteristics were recorded for patients and healthy controls. The Institutional Medical Investigation Ethics Committee accepted the study protocol (21-7T/27), and all participants provided written informed consent.

Clinic sample collection

EBC samples (Visit 0) were taken from 40 patients diagnosed with advanced stage (III and IVB) lung adenocarcinoma (LA) and 20 healthy individuals in the control group. Furthermore, peripheral blood samples were obtained at Visit 0 from 10 randomly selected patients with LA and one of every two individuals in the healthy control group to demonstrate the correlation of lncRNA expression levels in cfRNA samples obtained from EBC with liquid biopsy samples. Furthermore, EBC (Visit 1) samples were also collected from 26 patients diagnosed with LA during the third month of the follow-up.

EBC sample collection

EBC samples (2 mL) were collected from patients and placed in a DNA- and RNA-free 15-mL tube with a homemade glass EBC condenser, as described in our previous study (Tetik Vardarli et al., 2020).

Liquid biopsy collection

Peripheral blood was taken from patients included in this study in tubes with 4 mL ethylenediaminetetraacetic acid. Blood samples were centrifuged at 2000 × *g* for 10 min at 4°C within 2 h, and plasma samples were separated and transferred to a new centrifuge tube.

RT-PCR analysis

cfRNA was extracted from EBC and plasma specimens of patients and healthy individuals using Norgen Total RNA Purification Kit (Norgen Biotek Corp.) according to the manufacturer's instructions. cfRNA concentrations were measured using a Nanodrop 1000 instrument (Thermo Scientific). Then, cfRNAs were reverse transcribed into cDNA using the Transcriptor First Strand cDNA Synthesis Kit (Roche Applied Science) according to the manufacturer's protocols. The quantitative real-time RT-PCR method was used to determine MALAT1, HOTAIR, PVT1, NEAT1, ANRIL, and SPRY4-IT1 lncRNA expression levels using RT^2 SYBR Green qPCR Mastermix Kit (Roche) and gene-specific primers (Table 1) via LightCycler 480 Instrument II (Roche, Switzerland). GAPDH, a housekeeping gene, was used for normalization.

TABLE 1 lncRNA primer sequences.

ANRIL	F:5′-CCA TCA GAG GTA ACA GTA GAG AC-3′
	R:5′-GAG GCA GGA GAA TCG CTT G-3′
MALAT1	F:5′-CAG CAG TTC GTG GTG AAG ATA G-3′
	R:5′-GCC TCC TCC GTG TGG TTG-3′
PVT1	F:5′-ATA TGG ACT GTG ATG CGG AAG-3′
	R:5′-ATA ACC TGT GAT GAA CCA ATA AGC-3′
NEAT1	F:5′- TAC ACA GCG AGG CAC CAC-3′
	R: 5′-GTC AGC ACA GGA GCA GAG G-3′
SPRY4-IT1	F:5′-GAG GGG TTC TTA AAT AGG CAG C-3′
	R:5′-GAG GTT CTT TAA AAA CAG CCC A-3′
HOTAIR	F:5′- CTG ACT CGC CTG TGC TCT G-3′
	R:5′-CCG CCG TCT GTA ACT CTG G-3′

Statistical analysis

In lncRNA expression analysis, cycle threshold (CT) values obtained from the LC480 device were uploaded to the web-based online GeneGlobe Data Analysis Center (https://geneglobe.qiagen.com/us/analyze) software, normalized by the GAPDH gene expression, applied log2 transformation, and paired comparison of the groups. It was carried out in the same software with the $2^{-\Delta\Delta CT}$ method. After comparisons, those with significant lncRNA expression fold change ± two-fold and above and $p < 0.05$ with Student's *t*-test were listed in the web software. Additionally, comparative heat maps of the groups and correlation between EBC and liquid biopsy were performed with linear regression analysis in GraphPad Prism v.9.2. The importance of differentiated lncRNA expression analyses during the diagnosis using EBC and in reflecting the clinical outcomes of patients with LC was evaluated. The concordance among EBCs was interpreted using clinical data.

Results

Epidemiological and clinical findings of patients

The study included 20 healthy participants (10 F and 10 M) who had never been diagnosed with cancer and 40 patients with LA (11 F and 29 M). The average ages of patients diagnosed with LA and healthy control groups were 61.98 ± 9.59 (range: 40–79) and

TABLE 2 Demographic and clinical characteristics of patients with lung adenocarcinoma and the control group.

		Lung adenocarcinoma *n* = 40 (%)	Healthy control *n* = 20 (%)
Gender	Female	11 (27.5)	10 (50.0)
	Male	29 (72.5)	10 (50.0)
Age (year)	Female	61.1 ± 07.3	57.8 ± 7.3
	Male	62.3 ± 10.4	62.6 ± 5.7
Diagnosis	Stage IIIB	18 (45.0)	
	Stage IV	22 (55.0)	
Treatment	CT	8 (20.0)	
	CT and IT	10 (25.0)	
	CT and RT	10 (25.0)	
	CT, RT and IT	7 (17.5)	
	IT	4 (10.0)	
	RT and IT	1 (2.5)	
No Metastasis		18 (45.0)	
Metastasis		22 (55.0)	
Oligometastases	1	10 (45.0)	
	2	7 (31.8)	
	3	5 (27.7)	
Location of Oligometastases	Lung	14 (54.6)	
	Bone	8 (31.2)	
	Brain	5 (19.5)	
	Liver	3 (11.7)	
	Surrenal	3 (11.7)	
	Plevra	2 (7.8)	
	Soft tissue	2 (7.8)	
	Kidney	1 (3.9)	
	MSL	1 (3.9)	
Treatment duration (weeks)	Female	11.59 ± 11.91	
	Male	15.46 ± 11.87	

CT, chemotherapy; IT, immunotherapy; RT, radiotherapy; MSL, mediastinal subcarinal lymphadenopathy.

60.20 ± 6.85 (range: 42–72) years, respectively. The patient characteristics with clinical samples collected during the study are displayed in Table 2.

cfRNAs were successfully extracted from plasma (for liquid biopsy) and EBC samples which were taken for molecular analysis from patients (Visits 0 and 1) and healthy individuals (Visit 0).

lncRNA expression analyses

The qRT-PCR method was used to investigate the role of EBC materials in determining lncRNA expression levels in the diagnosis and follow-up of LA. MALAT1, HOTAIR, PVT1, NEAT1, ANRIL, and SPRY4-IT1 lncRNA gene expression levels were evaluated in cfRNA obtained from patients with LA and healthy individuals.

Fold changes of specific lncRNAs associated with proliferation and metastasis in NSCLC were determined by comparing patients diagnosed with LA and the control group at Visits 0 and 1. These findings revealed the expression levels of some lncRNAs that function as oncogenes in the LC process, namely, HOTAIR, PVT1, NEAT1, and MALAT1, were significantly increased in EBC samples of the patient group. The lncRNAs showing the supreme increases at Visit-0 and Visit-1 in LA patients were identified as HOTAIR (5 fold, $p = 0.004796$), PVT1 (7.9 fold, $p =$

TABLE 3 Fold changes of lncRNA expressions detected at Visits 0 and 1 compared with the control group in EBC samples.

lncRNA	EBC visit 0 Fold change	p	EBC visit 1 Fold change	p
HOTAIR	4.9588	0.004796	5.5008	0.085956
ANRIL	8.0700	0.101209	4.3349	0.306083
NEAT1	9.7075	0.154621	12.2778	0.009698
SPRY4-IT1	−4.3219	0.060601	−2.3959	0.397520
PVT1	7.8550	0.038149	6.7884	0.049306
MALAT1	7.6095	0.200276	8.4285	0.041893

TABLE 4 Fold changes of lncRNA expressions detected at Visit 1 compared with Visit 0 of EBC samples.

lncRNA	Fold change	p
HOTAIR	0.5460	**0.048808**
ANRIL	−3.6439	0.185442
NEAT1	2.5705	**0.017159**
SPRY4-IT1	1.9184	0.063224
PVT1	−1.0589	0.154379
MALAT1	0.8156	0.994734

The meaning of bold values emphasizes the *p*-values with significant.

0.038149) and NEAT1 (12.8 fold, $p = 0.009698$), PVT1 (6.8 fold, $p = 0.0493069$), MALAT (8.4fold, $p = 0.041893$), respectively (Table 3; Figure 1).

To determine the roles of these lncRNAs in the follow-up of LA, specific lncRNA expression levels detected in EBC samples obtained from 26 patients diagnosed with LC at Visit 1 were compared with expression levels detected at Visit 0; our results reveal that NEAT1 expression level increased two-fold from Visit 0 to Visit 1 ($p = 0.017159$ [Table 4; Figure 2]).

To evaluate the significance of lncRNAs in LA development, the expression patterns of lncRNAs in plasma samples from 10 patients with LA and 10 healthy individuals with qRT-PCR were also investigated. When we examined the lncRNA expression levels at Visit 0, only the MALAT1 expression level obtained from the plasma samples was nine-fold higher in the patient group compared with the control group ($p < 0.001$ [Table 5; Figure 3]). To assess whether the lncRNA expression levels in EBC and plasma samples from patients with LC are consistent with one another, the expression levels found in EBC and plasma specimens were collected during Visit 0. However, no concordance was observed in the lncRNA expression levels between the two groups ($p > 0.05$).

Specific lncRNAs are highly dysregulated in various cancers and induce tumor progression and metastasis. The expression levels of specific lncRNAs in EBC samples taken from patients with LA with and without metastasis were compared at Visit 0 to evaluate the roles of MALAT1, HOTAIR, PVT1, NEAT1, ANRIL, and SPRY4-IT1 expression in the development of metastasis. Our findings showed that PVT1 expression levels decreased by two-fold in patients with organ metastasis.

Comparison of lncRNA expression levels between patients with single or multiple organ metastases and those without metastasis revealed that HOTAIR expression level increased four-fold in patients with single organ metastasis and two-fold in those with multiple organ metastases, whereas PVT1 lncRNA expression levels decreased three-fold in those with multiple organ metastases. Changes

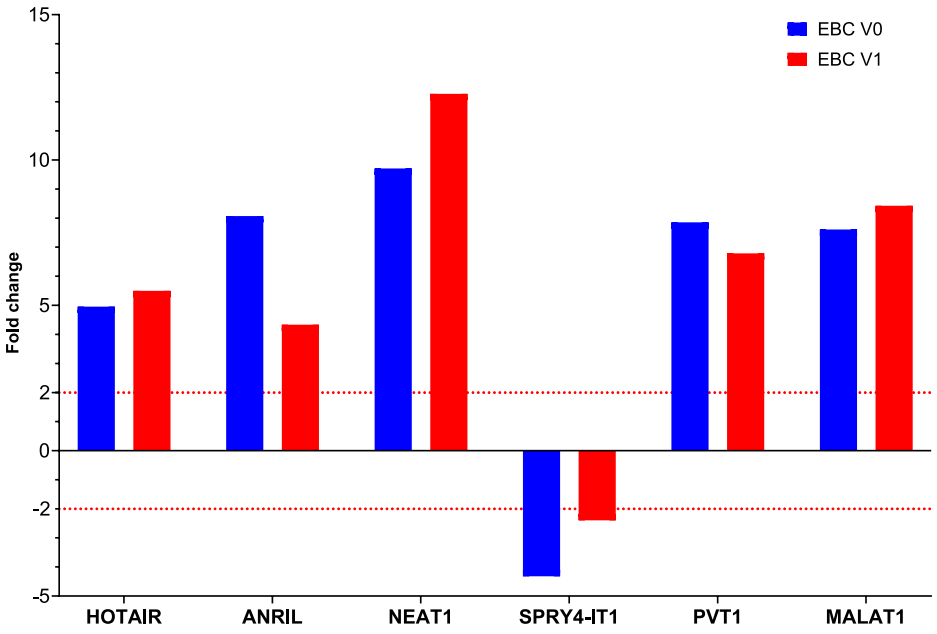


FIGURE 1
The fold changes in lncRNA expression levels were compared between EBC samples from patients at Visit 0 and Visit 1, and the control group.

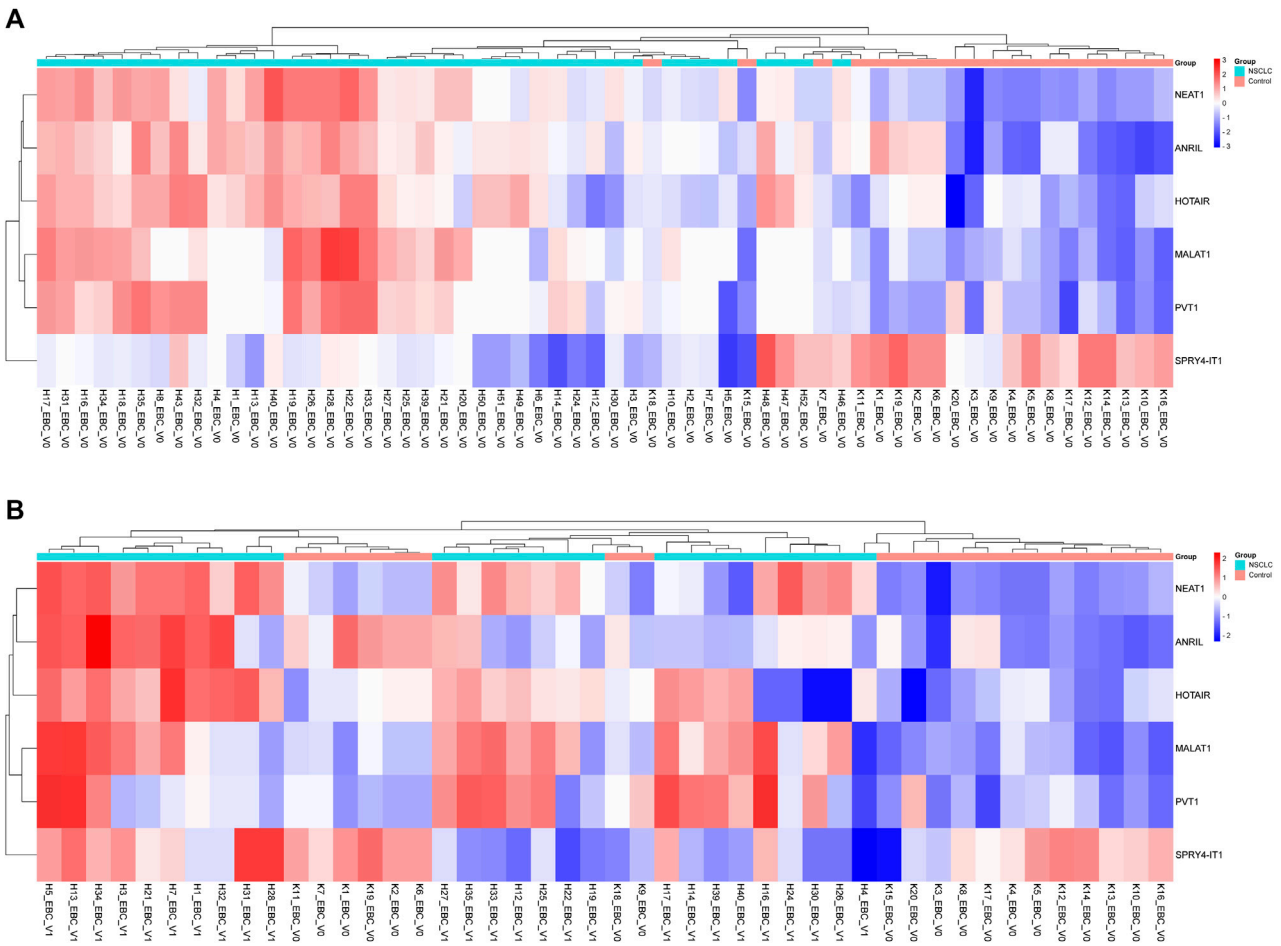


FIGURE 2
Displaying the expressions of lncRNAs in EBC samples of patients with NSCLC at Visits 0 and 1. **(A)** Illustration of a heatmap representing the 2-ΔΔCT changes in lncRNA expressions of Visit 0 EBC samples compared to the control group. **(B)** Displaying a heatmap representing the 2-ΔΔCT changes in lncRNA expressions of Visit 1 EBC samples compared to the control group.

TABLE 5 Fold changes of lncRNA expression levels at Visit 0 compared with the control group in plasma samples.

lncRNA	Fold change	p
HOTAIR	1.5311	0.262673
ANRIL	0.5558	0.274300
NEAT1	1.9855	0.337700
SPRY4-IT1	2.6781	0.730624
PVT1	8.4650	0.315780
MALAT1	8.9890	0.000001

in these lncRNA expression levels were found to be clinically significant in predicting the development of metastasis (Figure 4).

Comparing the lncRNA expression levels at Visit 0 based on organ involvement in patients with various organ metastases to those without metastases revealed no variation in lncRNA expression levels among patients with brain metastasis. Interestingly, we found that ANRIL and NEAT1 gene expression levels were decreased four-fold in instances with bone involvement and elevated four-fold in those with lung involvement (Figure 5).

Discussion

The study reveals that cfRNA extracted from EBC samples and liquid biopsy samples can be used to detect alterations in lncRNA gene expression levels for the molecular diagnosis and follow-up of LC. The development of new treatments that target molecular pathways implicated in carcinogenesis is encouraging; however, these treatments should effectively eliminate cancer cells without damaging healthy cells. To prevent difficulties associated with invasive procedures, early cancer diagnosis must depend on noninvasive, inexpensive, rapid, and reproducible techniques (You and Jones, 2012; Crowley et al., 2013; Sorber et al., 2017).

Several lncRNAs, such as MALAT1, HOTAIR, PVT1, NEAT1, ANRIL, and SPRY4-IT1, have recently been identified, particularly in cell lines, liquid biopsy, and tissue biopsy samples and have been associated with cancer

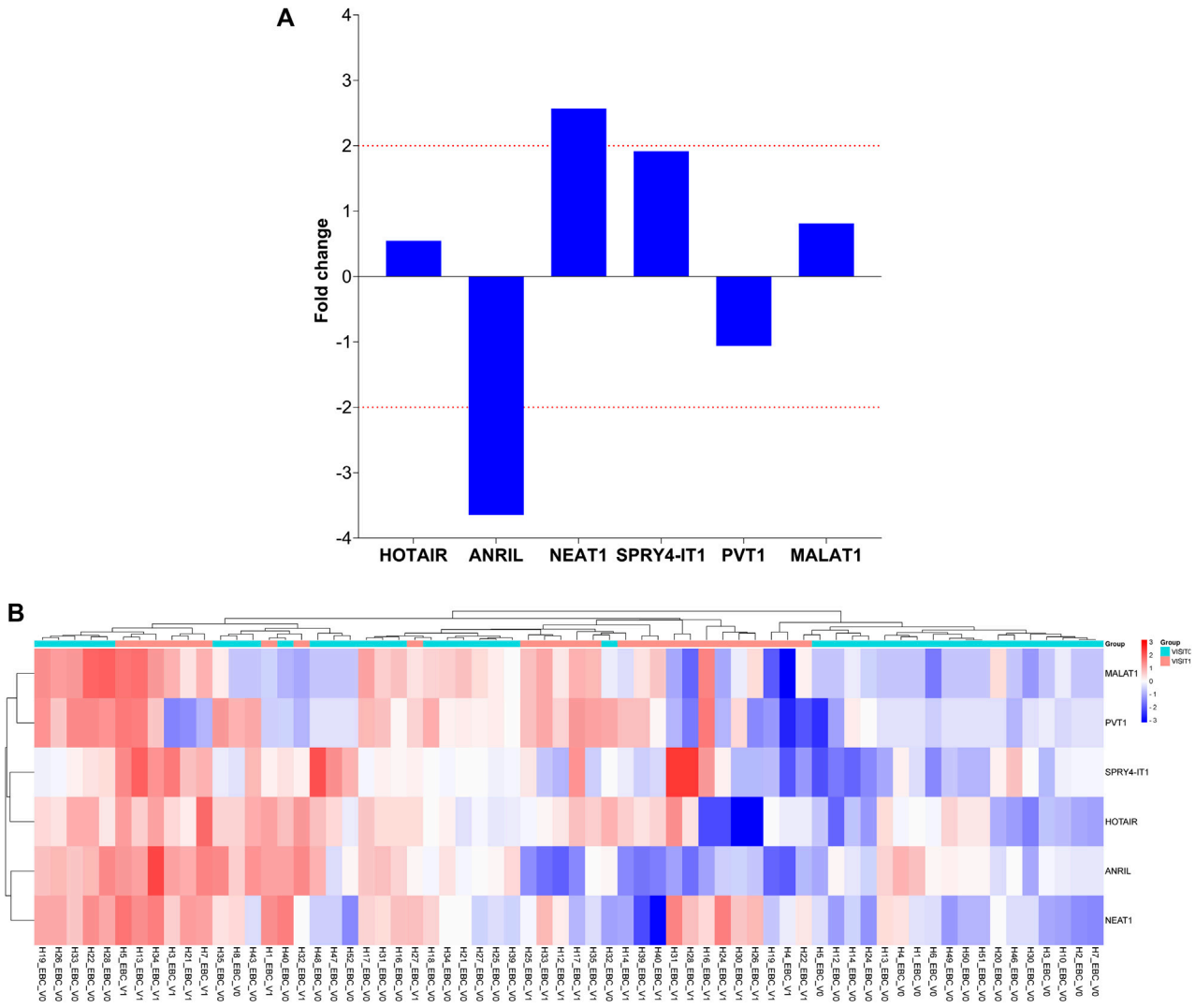


FIGURE 3
Illustration of the fold changes in lncRNA expression levels between Visit 1 and Visit 0 of EBC specimens. **(A)** Presenting a comparison of lncRNA expression levels at Visit 1 in EBC samples versus Visit 0. **(B)** Demonstration of a comparison of 2-ΔΔCT values for lncRNA expressions at Visit 1 in EBC samples compared to Visit 0.

pathogenesis. These findings suggest that lncRNAs provide new insights into the biology of this disease and could be used as a biomarker for early cancer detection. Consistent with other studies, we determined differentiated expression levels of lncRNAs in EBC samples from patients with LA. Additionally, HOTAIR and PVT1 (Visit 0) and NEAT1, PVT1, and MALAT1 (Visit 1) genes were overexpressed in EBC samples from patients with LA than those from healthy controls. The lncRNA gene expression profiles derived from the EBC sample can serve as both a diagnostic tool and a resource for patient follow-up. Increased HOTAIR gene expression levels have been identified as a diagnostic marker in patients with LA and lung squamous cell carcinoma in studies conducted with tumor tissue, plasma, and cell lines to determine the role of HOTAIR in NSCLC development (Wang et al., 2021; Yao et al., 2022). Yao et al. reported that HOTAIR expression is increased in LC tissues than in the control or surrounding normal tissues. Furthermore, they demonstrated an increased HOTAIR gene expression in both N and T stages. LncRNA HOTAIR has adequate diagnostic values for diagnosing LC (Yao et al., 2022). Similarly, HOTAIR gene expression was upregulated 5-fold and 5.5-fold in EBC samples from the patient group compared with the healthy control at Visits 0 and 1, respectively. HOTAIR seems to be an oncogene that plays a critical role in tumor progression. This study also shows that the HOTAIR gene expression profiles found in EBC samples are valuable for directing the molecular diagnosis of LC.

PVT1, an lncRNA, has been the focus of some preliminary studies in its potential function in NSCLC. PVT1 expression is higher in NSCLC tissue samples and cell lines, and PVT1 knockdown inhibits NSCLC cell proliferation, migration, and invasion (Wei et al., 2020; Wang et al., 2021). Additionally, Chen et al. revealed that A549 cells were made more sensitive to

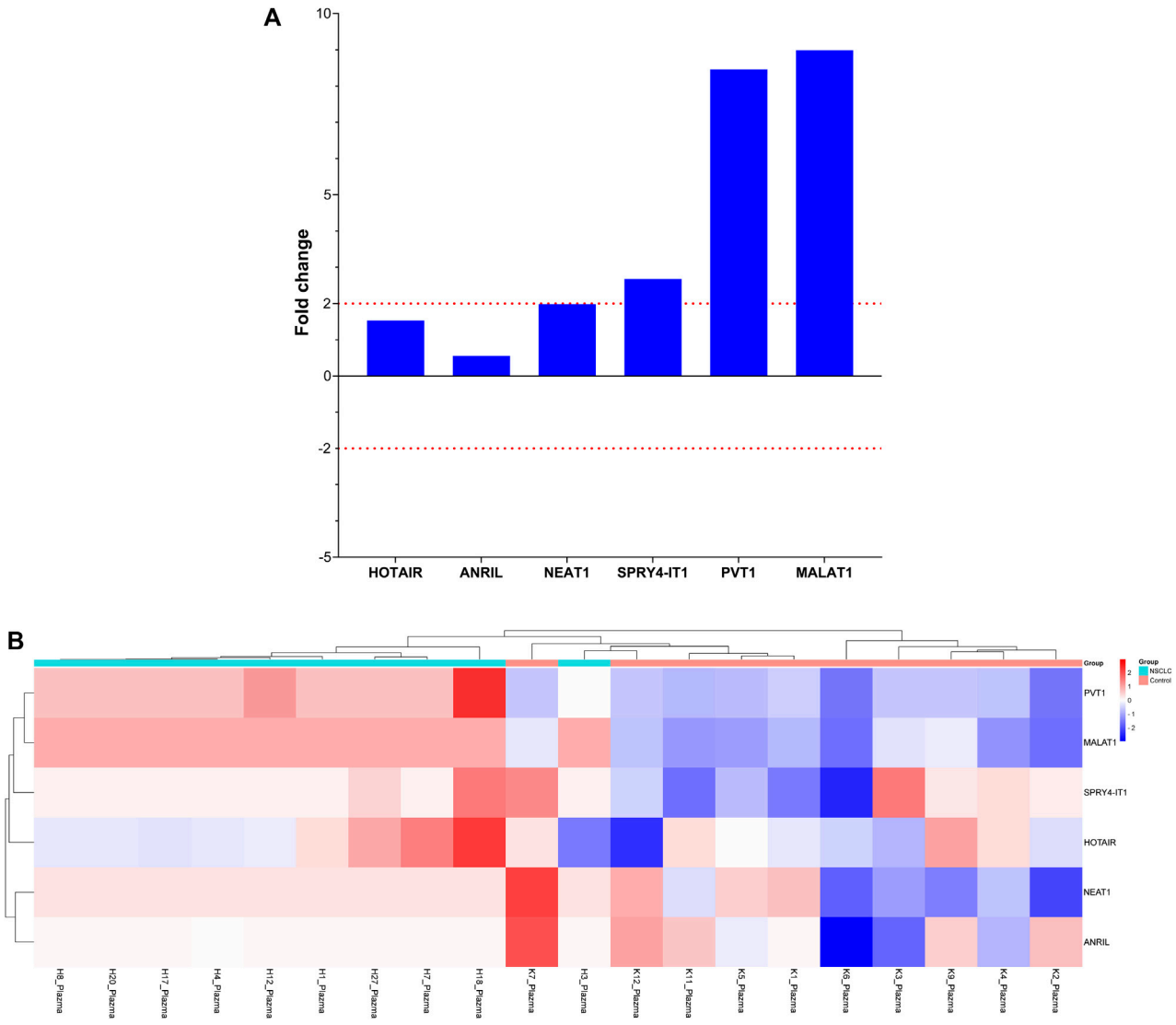


FIGURE 4
Depiction of the fold changes of specific lncRNAs in the plasma samples of patients with NSCLC at Visit 0. **(A)** Illustration of the fold changes in lncRNA expressions at Visit 0 of plasma cfRNA samples. **(B)** Displaying cases the 2-ΔΔCT changes in lncRNA expression of plasma cfRNA samples at Visit 0.

cisplatin after PVT1 was knocked down (Chen et al., 2019). Similarly, a higher expression pattern of the PVT1 gene in EBC samples of the patients was observed between both visits compared to healthy controls. We suggest that elevated PVT1 levels detected in EBC samples can be used as a diagnostic biomarker for NSCLC. These results lend support to the lncRNA analysis of EBC samples as a noninvasive method for detecting biomarkers in LA, as the lncRNA gene expression profile was shown to be similar to those of *in vivo* and *ex vivo* studies.

Several studies have profiled NEAT1 gene expression and its clinicopathological importance in NSCLC. Pan et al. reported that relative NEAT1 levels were considerably higher in NSCLC tissues than in nearby noncancerous lung tissues. Other studies also found a positive correlation between patient age and the presence of vascular invasion, lymph node metastases, and stage at diagnosis (Pan et al., 2015). Hu et al., (2016) reported that the levels of circulating NEAT1 were significantly higher in the plasma of patients with NSCLC. In a recent study that evaluated the effects of lncRNAs in plasma samples from 50 patients with NSCLC and 50 healthy individuals, researchers discovered 21 distinct lncRNAs and reported that SPRY4-IT1, ANRIL, and NEAT1 levels were significantly higher than those of other lncRNAs. It was emphasized that these three lncRNAs could serve as biomarkers for the early detection of NSCLC. Sun et al., (2016) also demonstrated that NEAT1 levels are elevated in patients with NSCLC, increasing cell proliferation and metastasis and inhibiting tumor cell death. Furthermore, they revealed that patients with elevated NEAT1 levels had lower overall survival than those with low NEAT1 gene expression, and that elevated NEAT1 levels accelerated NSCLC cell proliferation

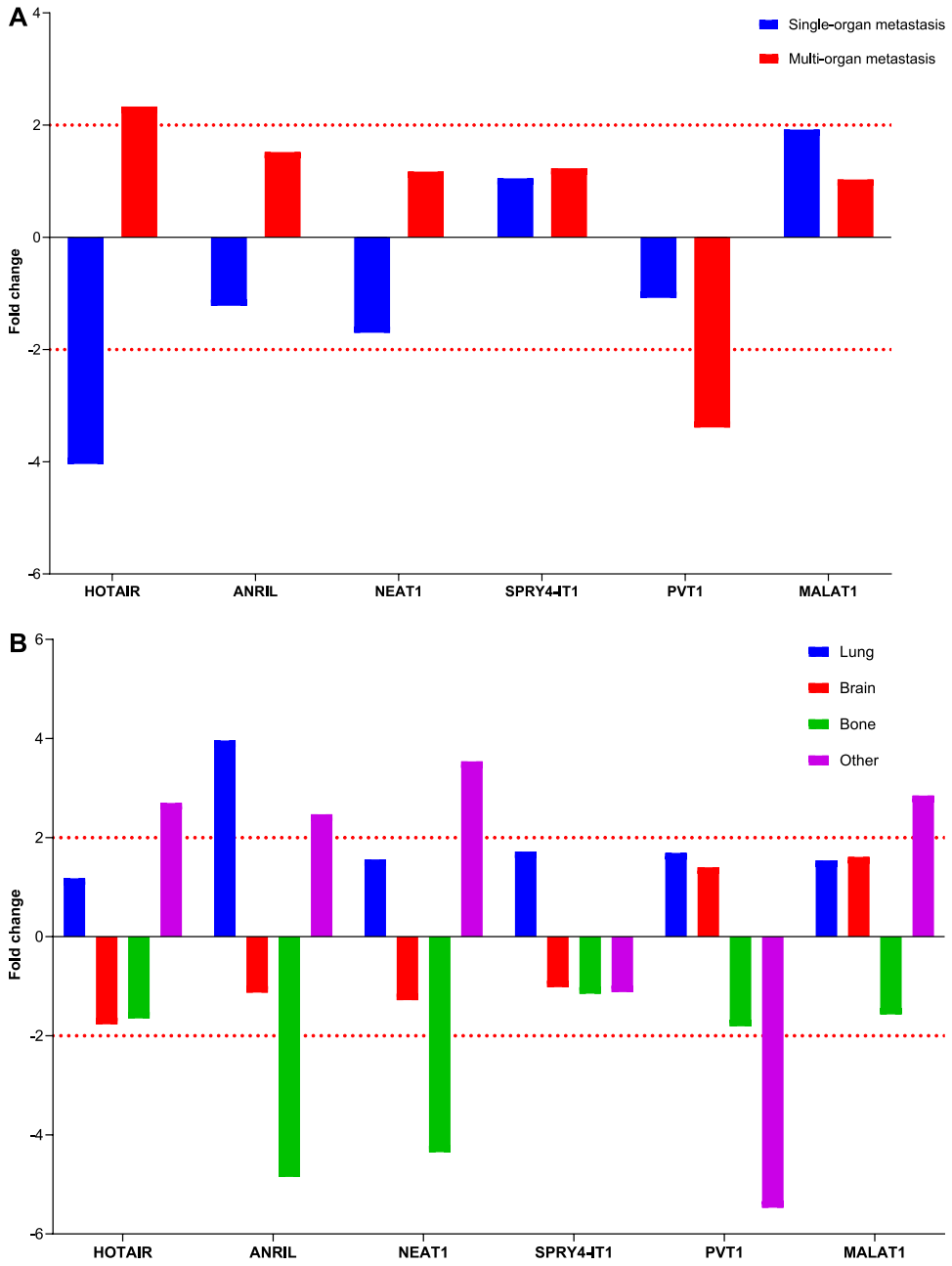


FIGURE 5
Illustration of the fold changes in lncRNA expressions in organ metastasis. **(A)** Depiction of the fold change of lncRNA expressions in patients with different organ metastases at Visit 0. **(B)** Displaying fold changes of lncRNA expression according to organ involvement.

and metastasis *in vitro* and *in vivo*. Both the diagnosis and follow-up of patients with LA showed elevated NEAT1 gene expression levels, which is consistent with the findings of earlier studies. These data suggest that NEAT1 can be one of the recently found regulators of the NSCLC progression. The lncRNA MALAT1 is a type of lncRNA greatly overexpressed in many tumors, including NSCLC. However, the mechanism of MALAT1 in NSCLC remains unclear. Several studies have shown that MALAT1 is significantly upregulated in human NSCLC cell lines and patients with LC. The findings demonstrated that MALAT1 overexpression is closely associated with NSCLC malignancy, vascular invasion, pathological differentiation, and recurrence and can be used as a biomarker to diagnose malignant tumors (Weber et al., 2013; Rong et al., 2020). To assess MALAT1's utility as a blood-based biomarker for NSCLC, Weber et al. compared the expression levels of MALAT1 in peripheral blood samples from patients with NSCLC and healthy individuals and discovered differences between the two groups (Weber et al., 2013). Rong et al., (2020) also suggested that MALAT1 levels in serum exosomes are higher in patients with

NSCLC and that exosome-derived MALAT1 also reflect biological changes in NSCLC cells. Similarly, Zhang et al., (2017) reported that MALAT1 expression was upregulated in serum exosomes of patients with NSCLC and that exosomal MALAT1 level was positively correlated with tumor stage and lymph node metastasis. To demonstrate a correlation between variations in the lncRNA expression levels in EBC and the related plasma samples, lncRNA expression analysis was performed on liquid biopsy samples from both patients with LA and the control group. We found that the MALAT1 lncRNA expression was altered in both groups, although no concordance was noted in the expression levels. Liquid biopsy provides the opportunity to detect and monitor cancer in various body fluids by detecting free circulating tumor cells, circulating tumor DNA and RNA fragments, and exosomes (Sorber et al., 2017). It is advantageous in reducing the harm of biopsy by noninvasive sampling and is of great importance for the early detection of cancer; however, a study reported that the low lncRNA expression level in the blood may complicate sensitive analysis (Weber et al., 2013). We hypothesize that the late entry and reduced tumor cell expression in the blood circulation can cause difficulty in recognizing a significant correlation between the results. We estimate that the molecular profile obtained from EBC samples in patients diagnosed with NSCLC reflects the molecular profile of tissue specimens more accurately. These results suggest that MALAT1 gene expression levels in EBC samples, in addition to the tumor tissue, peripheral blood, and serum exosomes, can be used as a tumor biomarker for the diagnosis and follow-up of NSCLC.

Hu et al., (2016) reported that SPRY4-IT1 gene is overexpressed in the plasma of patients with NSCLC compared with healthy volunteers, and the amplification of SPRY4-IT1 in the plasma of patients with NSCLC is closely related to the tumor size. Additionally, SPRY4-IT1 can be used as a biomarker for the early diagnosis of NSCLC tumor cells with a poor prognosis because of epithelial–mesenchymal transition activation by regulating the E-cadherin and vimentin expressions. However, Sun et al. suggested that the SPRY4-IT1 expression is significantly downregulated in NSCLC tissues, and specifically, decreased SPRY4-IT1 expression may be an important independent predictive factor for patients with NSCLC. They also reported that the upregulation of SPRY4-IT1 expression significantly inhibited cell proliferation, migration, invasion, and apoptosis, whereas reduced expression of SPRY4-IT1 promoted cell migration and invasion. They revealed that increased SPRY4-IT1 expression levels significantly reduced the number of metastatic nodules in the lungs *in vivo* (Sun et al., 2016). Similar to this study, we also found that SPRY4-IT1 lncRNA expression levels decreased four-fold and three-fold in EBC samples obtained from patients during the diagnosis (Visit 0) and treatment (Visit 1) compared with healthy individuals, respectively. These data support the hypothesis that SPRY4-IT1 downregulation may serve as a possible diagnostic biomarker for NSCLC and may play various functions not only in the diagnostic but also in the follow-up process.

To determine the clinical significance and potential role of ANRIL, Nie et al., (2015) investigated the expression levels of ANRIL in tumor and healthy tissues of 96 patients with NSCLC. They showed that ANRIL expression was increased in tumor tissues and the expression level was significantly correlated with TNM stages and tumor size. Moreover, using loss-of-function analysis in NSCLC cells, studies demonstrated that suppressed ANRIL expression can inhibit cell proliferation and induce cell apoptosis both *in vitro* and *in vivo*. These findings revealed that the ANRIL expression is associated with survival in patients with NSCLC and is critical to NSCLC development. This study shows novel perspectives on the function of lncRNA-directed carcinogenesis. A study performed with LC and cervical cancer cells reported that silencing ANRIL stops the cell cycle at the G1/G0 checkpoint and leads the cell to apoptosis. Our data show that in EBC samples collected from patients during Visit 0 and 1, ANRIL gene expression levels were eight- and four-fold higher, respectively, than that in the control group. We also determined that ANRIL expression levels in the patient group decreased by three-fold during treatment. These findings revealed that ANRIL may be a significant biomarker of adenocarcinoma progression. Low ANRIL expression levels may be associated with a favorable prognosis, given that the survival of our patients persisted throughout the study duration and the identification of decreasing ANRIL gene expression levels during the follow-up period.

Current studies indicated that lncRNAs are closely related to metastasis cancer progression and formation of various cancer types. However, the molecular basis for this observation in NSCLC advancement has not yet been thoroughly described. Zeng et al., (2021) used primary tumor and adjacent tissues from 30 patients who had not received any local or systemic therapy preoperatively and LC cell lines to investigate how PVT1 regulates tumorigenesis and NSCLC development. They found that PVT1 is upregulated in NSCLC tissues and lung cell lines and promotes NSCLC cell proliferation, migration, invasion, and metastasis. Weber et al., (2013) demonstrated that MALAT1 knockdown in NSCLC confirmed its antiproliferation, antimetastasis, and proapoptosis activities in the loss-of-function analysis in NSCLC tumor tissue and cell line models, in addition to its tumor-promoting effects in the xenograft mouse model. Our results indicate that PVT1 expression levels were two-fold lower at Visit 0 in patients with LA with diverse organ metastases. Moreover, HOTAIR gene expression is downregulated four-fold in single organ involvement and upregulated two-fold in multiorgan involvement, whereas PVT1 gene expression is downregulated three-fold. Moreover, ANRIL gene expression increased four-fold in patients with lung involvement, whereas ANRIL and NEAT1 gene expressions decreased four-fold in patients with bone metastases. Based on these results, we hypothesize that PVT1, HOTAIR, ANRIL, and NEAT1 are useful biomarkers for monitoring the development of metastasis in patients with LA. LncRNAs, which are candidates for new molecular targets in determining the early diagnosis, follow-up, and individual treatment decisions of patients with NSCLC, can also be detected by EBC within the scope of this study. In lncRNA expression analysis conducted on EBC samples derived from patients with NSCLC and healthy controls, the expression profiles of MALAT1, HOTAIR, PVT1, NEAT1, ANRIL, and

SPRY4-IT1 have been found, and an association between these expression profiles and molecular pathogenesis of LC has been established. In conclusion, molecular analysis of EBC samples from patients with LC will allow us to explore the mechanisms underlying the successful diagnosis, therapy, and prevention of cancer and identify novel therapeutic targets.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by The Ege University Medical Faculty Ethical Committee. The patients/participants provided their written informed consent to participate in this study.

Author contributions

AT and CG designed this study. TG, HK, LP, and KK collected data. AT and SO drafted the manuscript. AA and CG performed the lncRNA expression analysis. AT and SO conducted the literature search. All authors contributed to the article and approved the submitted version.

Funding

The research project received support from the Scientific and Technological Research Council of Turkey (TÜBİTAK-121S932). The article processing charges were funded by the Presidency of Strategy and Budget of the Republic of Turkey (2019K12-149080).

Acknowledgments

We are grateful to Ege University Planning and Monitoring Coordination of Organizational Development and Directorate of Library and Documentation for their support in editing and proofreading service of this study.

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.

References

Campbell, P. J., Stephens, P. J., Pleasance, E. D., O'Meara, S., Li, H., Santarius, T., et al. (2008). Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat. Genet.* 40 (6), 722–729. doi:10.1038/ng.128

Chen, L., Han, X., Hu, Z., and Chen, L. (2019). The PVT1/miR-216b/Beclin-1 regulates cisplatin sensitivity of NSCLC cells via modulating autophagy and apoptosis. *Cancer Chemother. Pharmacol.* 83 (5), 921–931. doi:10.1007/s00280-019-03808-3

Crowley, E., Di Nicolantonio, F., Loupakis, F., and Bardelli, A. (2013). Liquid biopsy: Monitoring cancer-genetics in the blood. *Nat. Rev. Clin. Oncol.* 10 (8), 472–484. doi:10.1038/nrclinonc.2013.110

Davis, M. D., Montpetit, A., and Hunt, J. (2012). Exhaled breath condensate: An overview. *Immunol. Allergy Clin. North Am.* 32 (3), 363–375. doi:10.1016/j.iac.2012.06.014

Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., et al. (2021). Cancer statistics for the year 2020: An overview. *Int. J. Cancer* 149 (4), 778–789. doi:10.1002/ijc.33588

Ferreira, H. J., and Esteller, M. (2018). Non-coding RNAs, epigenetics, and cancer: Tying it all together. *Cancer Metastasis Rev.* 37 (1), 55–73. doi:10.1007/s10555-017-9715-8

Gutschner, T., and Diederichs, S. (2012). The hallmarks of cancer: A long non-coding RNA point of view. *RNA Biol.* 9 (6), 703–719. doi:10.4161/rna.20481

Hirsch, F. R., Scagliotti, G. V., Mulshine, J. L., Kwon, R., Curran, W. J., Jr., Wu, Y. L., et al. (2017). Lung cancer: Current therapies and new targeted treatments. *Lancet* 389 (10066), 299–311. doi:10.1016/S0140-6736(16)30958-8

Hu, X., Bao, J., Wang, Z., Zhang, Z., Gu, P., Tao, F., et al. (2016). The plasma lncRNA acting as fingerprint in non-small-cell lung cancer. *Tumour Biol.* 37 (3), 3497–3504. doi:10.1007/s13277-015-4023-9

Huarte, M. (2015). The emerging role of lncRNAs in cancer. *Nat. Med.* 21 (11), 1253–1261. doi:10.1038/nm.3981

Jiang, C., Li, Y., Zhao, Z., Lu, J., Chen, H., Ding, N., et al. (2016). Identifying and functionally characterizing tissue-specific and ubiquitously expressed human lncRNAs. *Oncotarget* 7 (6), 7120–7133. doi:10.18632/oncotarget.6859

Li, M., Zhang, H., Zhao, X., Yan, L., Wang, C., Li, C., et al. (2014). SPRY4-mediated ERK1/2 signaling inhibition abolishes 17beta-estradiol-induced cell growth in endometrial adenocarcinoma cell. *Gynecol. Endocrinol.* 30 (8), 600–604. doi:10.3109/09513590.2014.912264

Lung Cancer Survival Rates *Lung cancer survival rates*. American Cancer Society. Available from: https://www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/survival-rates.html. (Accessed November 11, 2022)

Nie, F. Q., Sun, M., Yang, J. S., Xie, M., Xu, T. P., Xia, R., et al. (2015). Long noncoding RNA ANRIL promotes non-small cell lung cancer cell proliferation and inhibits apoptosis by silencing KLF2 and P21 expression. *Mol. Cancer Ther.* 14 (1), 268–277. doi:10.1158/1535-7163.MCT-14-0492

Pan, L. J., Zhong, T. F., Tang, R. X., Li, P., Dang, Y. W., Huang, S. N., et al. (2015). Upregulation and clinicopathological significance of long non-coding NEAT1 RNA in NSCLC tissues. *Asian Pac J. Cancer Prev.* 16 (7), 2851–2855. doi:10.7314/apjcp.2015.16.7.2851

Rong, F., Liu, L., Zou, C., Zeng, J., and Xu, Y. (2020). MALAT1 promotes cell tumorigenicity through regulating miR-515-5p/EEF2 Axis in non-small cell lung cancer. *Cancer Manag. Res.* 12, 7691–7701. doi:10.2147/CMAR.S242425

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

Shah, D. R., and Masters, G. A. (2020). Precision medicine in lung cancer treatment. *Surg. Oncol. Clin. N. Am.* 29 (1), 15–21. doi:10.1016/j.soc.2019.08.002

Shroff, G. S., de Groot, P. M., Papadimitrakopoulou, V. A., Truong, M. T., and Carter, B. W. (2018). Targeted therapy and immunotherapy in the treatment of non-small cell lung cancer. *Radiol. Clin. North Am.* 56 (3), 485–495. doi:10.1016/j.rcl.2018.01.012

Smolarz, B., Zadrozna-Nowak, A., and Romanowicz, H. (2021). The role of lncRNA in the development of tumors, including breast cancer. *Int. J. Mol. Sci.* 22 (16). doi:10.3390/ijms22168427

Sorber, L., Zwaenepoel, K., Deschoolmeester, V., Van Schil, P. E., Van Meerbeeck, J., Lardon, F., et al. (2017). Circulating cell-free nucleic acids and platelets as a liquid biopsy

in the provision of personalized therapy for lung cancer patients. *Lung Cancer* 107, 100–107. doi:10.1016/j.lungcan.2016.04.026

Sun, C., Li, S., Zhang, F., Xi, Y., Wang, L., Bi, Y., et al. (2016). Long non-coding RNA NEAT1 promotes non-small cell lung cancer progression through regulation of miR-377-3p-E2F3 pathway. *Oncotarget* 7 (32), 51784–51814. doi:10.18632/oncotarget.10108

Tetik Vardarli, A., Pelit, L., Aldag, C., Korba, K., Celebi, C., Dizdas, T. N., et al. (2020). Concordance in molecular genetic analysis of tumour tissue, plasma, and exhaled breath condensate samples from lung cancer patients. *J. Breath. Res.* 14 (3), 036001. doi:10.1088/1752-7163/ab739b

Wang, X., Cheng, Z., Dai, L., Jiang, T., Li, P., Jia, L., et al. (2021). LncRNA PVT1 facilitates proliferation, migration and invasion of NSCLC cells via miR-551b/FGFR1 Axis. *Onco Targets Ther.* 14, 3555–3565. doi:10.2147/OTT.S273794

Weber, D. G., Johnen, G., Casjens, S., Bryk, O., Pesch, B., Jockel, K. H., et al. (2013). Evaluation of long noncoding RNA MALAT1 as a candidate blood-based biomarker for the diagnosis of non-small cell lung cancer. *BMC Res. Notes* 6, 518. doi:10.1186/1756-0500-6-518

Wei, C. M., Zhao, X. F., Qiu, H. B., Ming, Z., Liu, K., and Yan, J. (2020). The long non-coding RNA PVT1/miR-145-5p/ITGB8 axis regulates cell proliferation, apoptosis, migration and invasion in non-small cell lung cancer cells. *Neoplasma* 67 (4), 802–812. doi:10.4149/neo_2020_190723N657

Yao, X., Wang, T., Sun, M. Y., Yuming, Y., Guixin, D., and Liu, J. (2022). Diagnostic value of lncRNA HOTAIR as a biomarker for detecting and staging of non-small cell lung cancer. *Biomarkers* 27 (6), 526–533. doi:10.1080/1354750X.2022.2085799

You, J. S., and Jones, P. A. (2012). Cancer genetics and epigenetics: Two sides of the same coin? *Cancer Cell.* 22 (1), 9–20. doi:10.1016/j.ccr.2012.06.008

Zeng, S. H. G., Xie, J. H., Zeng, Q. Y., Dai, S. H. H., Wang, Y., Wan, X. M., et al. (2021). lncRNA PVT1 promotes metastasis of non-small cell lung cancer through EZH2-mediated activation of hippo/NOTCH1 signaling pathways. *Cell. J.* 23 (1), 21–31. doi:10.22074/cellj.2021.7010

Zhang, R., Xia, Y., Wang, Z., Zheng, J., Chen, Y., Li, X., et al. (2017). Serum long non coding RNA MALAT-1 protected by exosomes is up-regulated and promotes cell proliferation and migration in non-small cell lung cancer. *Biochem. Biophys. Res. Commun.* 490 (2), 406–414. doi:10.1016/j.bbrc.2017.06.055

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 09 November 2023
DOI 10.3389/fonc.2023.1231094

OPEN ACCESS

EDITED BY
Mehdi Pirooznia,
Johnson & Johnson, United States

REVIEWED BY
Lin Wu,
Central South University, China
Ming Yi,
Zhejiang University, China
Mattia Falchetto Osti,
Sapienza University of Rome, Italy

*CORRESPONDENCE
Nong Yang
yangnong0217@163.com
Xiaorong Dong
xiaorongdong@hust.edu.cn

†These authors have contributed equally to this work

RECEIVED 30 May 2023
ACCEPTED 13 October 2023
PUBLISHED 09 November 2023

CITATION
Zhou R, Tong F, Zhang Y, Zhang R, Bin Y, Zhang S, Yang N and Dong X (2023) Genomic alterations associated with pseudoprogression and hyperprogressive disease during anti-PD1 treatment for advanced non-small-cell lung cancer. *Front. Oncol.* 13:1231094. doi: 10.3389/fonc.2023.1231094

Genomic alterations associated with pseudoprogression and hyperprogressive disease during anti-PD1 treatment for advanced non-small-cell lung cancer

Rui Zhou[1,2†], Fan Tong[1,2†], Yongchang Zhang[3†], Ruigang Zhang[1,2†], Yawen Bin[1,2], Sheng Zhang[1,2], Nong Yang[3*] and Xiaorong Dong[1,2*]

[1]Cancer Center, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [2]Institute of Radiation Oncology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [3]Department of Medical Oncology, Lung Cancer and Gastrointestinal Unit, Hunan Cancer Hospital/The Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha, China

Introduction: This study aimed to elucidate the relationship between dynamic genomic mutation alteration and pseudoprogression (PsPD)/hyperprogressive disease (HPD) in immunotherapy-treated advanced non-small-cell lung cancer (NSCLC), to provide clinical evidence for identifying and distinguishing between PsPD and HPD.

Method: Patients with advanced NSCLC who were treated with anti-PD1 were enrolled. Whole blood was collected at baseline and post image progression. Serum was separated and sequenced using 425-panel next-generation sequencing analysis (NGS).

Results: NGS revealed that not only single gene mutations were associated with PsPD/HPD before treatment, dynamic monitoring of the whole-blood genome mutation spectrum also varied greatly. Mutational burden, allele frequency%, and relative circulating tumor DNA abundance indicated that the fold change after image progression was much higher in the HPD group.

Discussion: The gene mutation profiles of PsPD and HPD not only differed before treatment, but higher genome mutation spectrum post image progression indicated true disease progression in patients with HPD. This suggests that dynamic whole-genome mutation profile monitoring as NGS can distinguish PsPD from HPD more effectively than single gene detection, providing a novel method for guiding clinical immune treatment.

KEYWORDS

next-generation sequencing analysis, pseudoprogression, hyperprogressive disease, anti-PD1 treatment, non-small cell lung cancer

1 Introduction

The widespread application of immune-checkpoint inhibitors (ICIs) in the treatment of non-small-cell lung cancer (NSCLC) has substantially prolonged patients' overall survival (1). However, they are not always effective for the entire population. Some studies have reported that programmed death-ligand 1 (PDL1), tumor mutational burden (TMB), mismatch repair, and $CD8^+T$ cells may be potential biomarkers for efficacy prediction, whereas Janus kinase 1 (JAK1), JAK2, and beta-2-microglobulin truncation may be predictors of primary resistance (2–4). In addition to the uncertain prediction of efficacy, the immunotherapy response also differs from that of classical radiotherapy and chemotherapy. Two distinct and atypical patterns of response to ICIs are pseudoprogression (PsPD) and hyperprogressive disease (HPD). PsPD is characterized by an increase in tumor size or the appearance of new lesions after ICI treatment, followed by tumor regression (5). HPD represents a novel pattern of progression, with an unexpected and rapid increase in both tumor volume and rate (6). The conventional imaging-based efficacy evaluation model may lead to tremendous bias, engendering confusion for clinicians. Hence, iRECIST has been redefined to differentiate response patterns between immunotherapy, chemotherapy, and radiotherapy (7). However, precise methods for distinguishing between PsPD and HPD are still lacking in clinical practice, necessitating the formulation of appropriate prediction strategies in such cases.

Next-generation sequencing (NGS), also termed high-throughput or massively parallel sequencing, is a technology that facilitates simultaneous and independent sequencing of thousands to billions of DNA fragments. This method not only detects the mutations in a single gene but also analyzes the total abundance and characteristics of all gene mutations (8). Continuous NGS of patient samples before and after treatment can provide information regarding the changes in gene profiles resulting from immunotherapy. Tissue detection is accurate, but continuous testing is encumbered by the invasiveness of the specimen acquisition procedure. Blood NGS testing can partially replace tissue sample testing, as it enables continuous testing owing to the ease of obtaining specimens (9, 10).

To date, no study has examined the relationship between dynamic genomic alterations in blood and the occurrence of PsPD and HPD in patients with NSCLC who undergo immunotherapy. To date, no study has examined the relationship between dynamic genomic alterations in blood and the occurrence of PsPD and HPD in patients with NSCLC who undergo immunotherapy. It provides a transformative and practical method for predicting the clinical response of NSCLC to immunotherapy and offers individualized scheme selection of immunotherapy based on dynamic NGS detection.

Abbreviations: AF, Allele frequency; BAM, Binary Alignment Map; CT, Computed tomography; DDR, DNA damage repair; HNSCC, Head and Neck Squamous Cell Carcinoma; HP, Hyper-progression; ICI, Immune-checkpoint inhibitors; NGS, Next-generation sequencing; NSCLC, Non-small-cell lung cancer; PD, Progressive disease; PR, Partial response; RECIST, Response Evaluation Criteria in Solid Tumors; SAM, Sequence Alignment/Map; TMB, Tumor mutational burden.

2 Materials and methods

2.1 Patients

This study enrolled 14 Chinese patients who were diagnosed with NSCLC based on histopathological examination at Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, between October 2018 and December 2021 and received anti-PD1 monoclonal antibody treatment. The inclusion criterion was a negative past history of cancer. An informed consent form was obtained from each participant. All patients underwent a follow-up. The clinical data were collected from the Union Hospital medical records database. The stage was classified according to the 8^{th} edition of TNM Staging of NSCLC by the International Association for the Study of Lung Cancer. The overall study protocol was approved by the Ethics Review Board of Wuhan Union Hospital, Huazhong University of Science and Technology, and the research was conducted in accordance with relevant ethical guidelines (2018-S271).

2.2 Study design

See Figure 1.

2.3 Blood sample collection

A 10 mL blood sample was drawn and stored in a Cell-Free DNA Storage Tube (Streck 218962) at 25°C. Blood was centrifuged at 1,600 ×*g* for 10 min at 25°C to obtained plasma. Plasma was centrifuged at 1,600 ×*g* for 10 min at 4°C, and the supernatant was placed in a new tube. Plasma was separated from blood

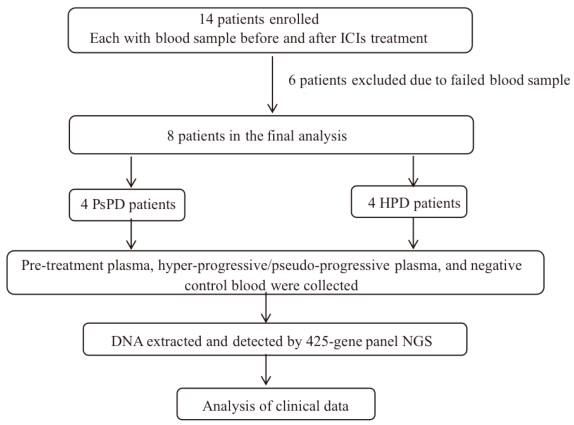


FIGURE 1
The design of this study.

(no apparent hemolysis) within 72 h after blood collection and stored at −80°C until DNA isolation.

2.4 Sample preparation, DNA isolation, and sequencing

Circulating free DNA was extracted using the QIAamp Circulating Nucleic Acid Kit (QIAGEN, Hilden, Germany). Genomic DNA obtained from buccal swabs was prepared using the DNeasy Blood & Tissue kit (QIAGEN) as a control for germline mutations. DNA was quantified using the dsDNA HS Assay Kit (Life Technologies, Eugene, Oregon), according to the manufacturer's recommendations. Sequencing libraries were prepared using the KAPA Hyper Prep Kit (KAPA Biosystems, Cape Town, South Africa), according to the manufacturer's instructions for different sample types. Customized xGen lockdown probes (Integrated DNA Technologies, Coralville, IA, USA) targeting 425 tumor-related genes were used for hybridization enrichment (425 genes, Supplementary Table 1).

NGS was performed, followed by CLIA-certified and CAP-accredited assay validation at a centralized clinical testing center (Nanjing Geneseq Technology, Inc., Nanjing, China). The libraries were sequenced on a HiSeq 4000 NGS platform (Illumina, San Diego, CA, USA), and the sequencing data were analyzed to detect genomic alterations. The mean coverage depth was ~100X for controls and ~3,000X for circulating free DNA samples. The resultant sequences were analyzed for base substitutions, small insertions and deletions, copy number alterations (focal amplifications and homozygous deletions), and gene fusions/rearrangements.

2.5 Analysis of DNA sequences

Sequencing data were processed as described previously (11, 12). Briefly, the data were first subjected to demultiplexing and FASTQ file quality control to remove low-quality data or N bases. Qualified reads were mapped to the reference human genome GRCh37/hg19 using the Burrows-Wheeler Aligner (13) and default parameters to create Sequence Alignment/Map (SAM) files (14). Picard was used to convert the SAM files to compressed Binary Alignment Map (BAM) files, which were then sorted according to the chromosomal coordinates. The Genome Analysis Toolkit (15) was used to locally realign the BAM files at intervals with insertion/deletion (indel) mismatches and recalibrate the base quality scores of the reads in the BAM files. VarScan2 (16) was employed to detect single-nucleotide variations (SNVs) and indel mutations. The resulting mutation lists were further filtered through an internally collected list (1,000 normal samples) of recurrent artefacts on the same sequencing platform. SNVs and indels were further filtered based on the following parameters: (1) minimum read depth=20, (2) minimum base quality=15, (3) minimum variant supporting reads=5, (4) variant supporting reads mapped to both strands, (5) strand bias no greater than 10%, (6) if present in >1% population frequency in the 1000G or ExAC database, and (7) through an internally collected list of recurrent sequencing errors using a normal pool of 100 samples. Copy number variations (CNVs) were analyzed with the CNVkit (17) Depth ratios above 2 and below 0.6 were considered as gains and losses in CNVs, respectively. Variants that are predicted to shift the translational reading frame should be described using either a short or a long form p.(Arg97fs) and p.(Arg97Profs*23), respectively. For 'fsTer#'/'fs*#', it is specified that '#' indicates at which codon number the new reading frame ends with a stop codon. The number of the stop in the new reading frame is calculated starting at the first amino acid that is changed by the frame shift, ending at the stop codon (*#) (18).

2.6 TMB calculation

TMB was defined as the total number of missense mutations. In addition, we profiled TMB of these samples by a targeted NGS panel (Geneseeq) to evaluate its correlation with whole-exome sequencing (WES) results. Panel TMB was counted by summing all base substitutions and indels in the coding region of targeted genes, including synonymous alterations to reduce sampling noise and excluding known driver mutations as they are over-represented in the panel, as previously described (19, 20).

2.7 Response assessment

Treatment efficacy was assessed by the treating physician and another independent physician and classified according to the Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1. Radiological evaluation of treatment efficacy by computed tomography (CT) was performed before treatment and on a schedule determined by each treating physician during treatment. HPD was defined as time-to-treatment failure <2 months, >50% increase in the tumor burden compared with that of pre-immunotherapy imaging, and >2-fold increase in the speed of progression. PsPD was defined as a partial response (PR) following RECIST-defined progressive disease (PD) during ICI treatment. The definition of PR in PsPD was assessed according to the changes observed from the time of PD and not from treatment initiation.

2.8 Statistical analysis

Figures were drawn using R-3.5.3 for Windows (32/64 bit), GraphPad Prism 8, and EXCEL. The difference for AF% of mutation gene between PsPD and HPD were analyzed by unpaired t test.

3 Results

3.1 Patient characteristics

Fourteen patients underwent eligibility assessment as described above, and eight patients were enrolled in this trial. The participants' baseline characteristics are listed in Supplementary Table 2.

All patients received anti-PD1 therapy, and their median age was 56 years (range, 27–72 years). Five of 14 (62.5%) received pembrolizumab (Keytruda®) treatment, while the others received sintilimab (Tyvyt®). The majority of patients were men (6/8, 75.0%) and had stage IV disease (7/8, 87.5%). Adenocarcinoma (6/8, 75.0%) was the predominant histopathological type, followed by squamous (1/14, 12.5%) or adenosquamous carcinoma (1/14, 12.5%). Four out of eight patients (50.0%) had a history of smoking, and 2/8 (25.0%) patients were non-smokers. Blood samples were collected from all patients at baseline and after image progression. NGS screening of plasma was performed for all patients (N=14), but only 8/14 were enrolled in the final analysis. Six patients were excluded owing to the poor quality of the blood sample. The median follow-up time was 4.0 months (range, 0.8–14.9 months), and the cut-off date was February 23, 2021. Till the cut-off date, 6/8 (75%) patients experienced disease progression, 4/8 (50%) had PsPD, and the others had HPD (Supplementary Table 3).

3.2 Mutational landscape

All enrolled patients underwent a 425-panel NGS at baseline and after image progression. We identified a list of frequently mutated genes in NSCLC on the basis of these data; the 20 most common genes are illustrated in Supplementary Figure 1. Upon segregation of the PsPD and HPD groups, the mutational landscape of patients with the two responses showed a differential pattern. As shown in Figure 2A, the most frequently mutated genes in the PsPD group were TP53 (8.8%), NOTCH2 (5.9%), SMARCA4 (5.9%), LRP1B (4.4%), and STAG2 (4.4%), while TP53 (10.8%), EGFR (6.0%), ARD2 (3.6%), ATM (3.6%), and PIK3CA (3.6%) were the top 5 genes in the HPD group (Supplementary Table 4).

We compared the differences in the most common mutant genes between PsPD and HPD at baseline and the results were consistent with those of previous findings (Figure 2B). TP53, SMARCA4, and KRAS were positively related to immunotherapy efficacy, while JAK2 was negatively related to immunotherapy efficacy. We also observed EGFR and DNMT3A mutations in patients with HPD but not in those with PsPD, akin to the results of a previous study. Subsequently, we explored the genomic alterations after PsPD/HPD (Figure 2C, D). The top mutant gene frequency did not change considerably in patients with PsPD (Table 1), whereas the top mutant gene frequency increased significantly in patients with HPD (Table 2). As TP53 increased from 75.0% to 100.0%, ATM and EGFR increased from 25% to 50%.

Herein, we investigated the single gene mutation pattern in PsPD and HPD before treatment and after image progression, as well as the variation tendency. It provided insights suggesting that alterations may reveal more information than a single mutation alone. Thus, the whole genomic spectrum was further analyzed.

3.3 Spectrum analysis of gene mutation

The aforementioned results provided a clue that the frequency of top mutational gene remained stable in patients with PsPD, while it increased significantly in patients with HPD. It also illustrated that the total number of mutated genes was elevated after HPD compared with that after PsPD. Based on the co-mutation gene data, we discovered that the frequency of co-mutated genes was higher in the PsPD group than in the HPD group (84.21% vs 38.33%) (Supplementary Figure 2A-C). The detailed gene spectrum is presented in Figure 3A, B, indicating that a higher co-mutation gene spectrum may indicate stable disease, while a lower co-mutation gene spectrum may suggest disease progression.

Several studies have found no association between blood-based TMB (bTMB) and the efficacy of immunotherapy. However, the association between change in the bTMB and the clinical outcomes of immunotherapy remains to be explored. Herein, we compared each patient's bTMB at baseline and after image progression. The data revealed that bTMB was elevated in 75% and 50% of patients with HPD and PsPD, respectively (Figure 3C). In the HPD group, the greatest elevation reached 538.1%, whereas the highest value in the PsPD group was merely 24.7%. Interestingly, the maximum allele frequency (AF) increased significantly in 75% of patients with HPD, whereas only a mild elevation was observed in 25% of patients with PsPD (Figure 3D). The same pattern was also observed in circulating tumor DNA (ctDNA) abundance (Figure 3E). These results indicated that spectrum analysis of gene mutation data possessed greater efficacy in distinguishing between PsPD and HPD than single mutant gene analysis.

3.4 Alterations in gene mutation abundance

Based on the NGS results, we discovered that the maximum AF for the top mutated genes decreased significantly in patients with PsPD (Figure 4A). On the contrary, it increased in patients with HPD (Figure 4B). The mutational abundance of the top three mutant genes, viz. TP53, KRAS, and SMARCA4, declined in patients with PsPD. These genes were also previously reported to be positive prognostic factors for immunotherapy. On the other hand, the mutational abundance of the top mutant genes such as EGFR, JAK2, and DNMT3A were elevated after progression in patients with HPD, which were previously reported to be negative prognostic factors of immunotherapy.

Thus, these data suggested that the genomic alteration after imaging progression can be a promising biomarker for distinguishing between PsPD and HPD in patients undergoing immunotherapy for NSCLC.

4 Discussion

ICIs confer tremendous survival benefits, especially in patients with NSCLC (1). However, they also pose great challenges for clinicians. In contrast to chemotherapy and radiotherapy, whose treatment response is evaluated using conventional radiographic methods, immunotherapy sometimes presents with unique treatment response patterns, such as PsPD and HPD (21). PsPD manifests with confusing imaging features, which may result in the

10.3389/fonc.2023.1231094

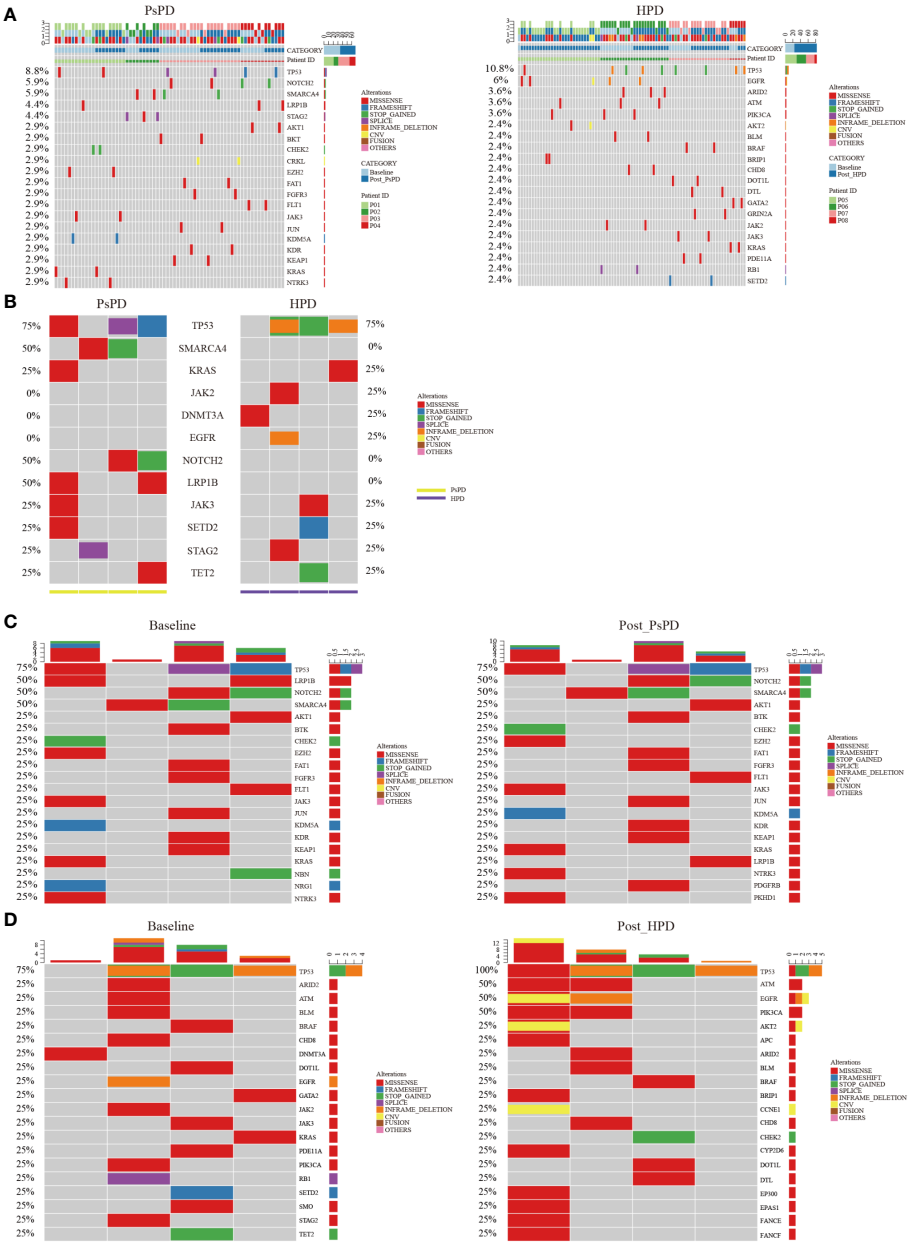


FIGURE 2
(A) Heat map of the top 20 mutant genes screened by plasma NGS in patients with PsPD/HPD (pooled data of baseline and after treatment). Left: PsPD, right: HPD. **(B)** Heat map of the 12 representative mutant genes at baseline in patients with PsPD/HPD. **(C)** Heat map of the top 20 mutant genes in patients with PsPD. Left: baseline, right: after PsPD. **(D)** Heat map of the top 20 mutant genes in patients with HPD. Left: baseline, right: after HPD. PsPD: pseudoprogression, HPD: hyperprogressive disease, NGS: next-generation sequencing.

loss of the potential benefits from ICIs. HPD, an extremely poor outcome of ICI therapy, often results in fatality, necessitating urgent attention. However, these atypical responses cannot be easily distinguished on conventional imaging, as the early stages of PsPD and HPD are both characterized by short-term enlargement of the tumor lesion (22). Some studies revealed that specific gene mutations may be associated with a corresponding unique response in patients (23); however, no single gene can accurately predict the efficacy of ICIs. NGS reveals more information about genomic mutation, whereas the evidence to link the whole genomic alteration and PsPD/HPD is lacking. Here, we have, for the first time, revealed the significant role of bTMB and shift in the mutational abundance before treatment and after imaging progression in the determination of PsPD/HPD in NSCLC. It highlights the clinical value of dynamic blood NGS monitoring to better understand the unique response patterns elicited by ICIs.

The incidence of HPD in NSCLC with ICIs treatment is reported to be 13.8%, highlighting a considerable variation in cases due to different causal factors (24). It is not triggered by a single factor, but by a series of events that occur simultaneously. T-regulatory cells lacking PD-1 signaling or tumor cells lacking PD-L1 have been shown to accelerate tumor development in HPD models

10.3389/fonc.2023.1231094

TABLE 1 Gene mutation frequency change in PsPD.

Gene Rank	Before treatment		After PsPD	
	Gene name	Frequency	Gene name	Frequency
1	TP53	75%	TP53	75%
2	LRP1B	50%	NOTCH2	50%
3	NOTCH2	50%	SMARCA4	50%
4	SMARCA4	50%	AKT1	25%
5	AKT1	25%	BTK	25%
6	BTK	25%	CHEK2	25%
7	CHEK2	25%	EZH2	25%
8	EZH2	25%	FAT1	25%
9	FAT1	25%	FGFR3	25%
10	FGFR3	25%	FLT1	25%
11	FLT1	25%	JAK3	25%
12	JAK3	25%	JUN	25%
13	JUN	25%	KDM5A	25%
14	KDM5A	25%	KDR	25%
15	KDR	25%	KEAP1	25%
16	KEAP1	25%	KRAS	25%
17	KRAS	25%	LRP1B	25%
18	NBN	25%	NTRK3	25%
19	NRG1	25%	PDGFRB	25%
20	NTRK3	25%	PKHD1	25%

(25, 26). Accumulating data demonstrate that high TGF-β is correlated with resistance to anti-PD-1/PD-L1 therapy, thus anti-TGF-β/PD-L1 bispecific antibodies such as YM101 and BiTP confer the resistance and exhibit enhanced antitumor activity in cancer treatment (27, 28). MDM2/MDM4 amplification may be associated with HPD (29, 30) as it promotes tumorigenesis directly or indirectly through the inhibition of p53. Pharmacological inhibitor results in an improvement in the antitumor immunity to anti-PD-1 treatment (31). EGFR is involved in immunotherapy-related resistance and HPD due to upregulation in the number of immunosuppressive receptors and induction of the secretion of cytokines (32–34). DNMT3A mutation is also related to poor outcomes with ICIs treatment in clinical research as well as EGFR and MDM2/4 mutations (35, 36). Loss-of-function mutations in JAK1/2 may play an important role in the lack of response to PD-1 inhibitors due to the reduced ability of immune T cells to recognize tumor cells. These mutations may result in the deficiency of T-cell infiltrates due to a deficit of chemokine production (37–39). PsPD is a rare phenomenon observed in <5% of cases of NSCLC (40). It is defined as the appearance of new lesions or tumor enlargement during therapy, followed by disease regression or stabilization at subsequent imaging (40). It indicates the true benefit of ICIs treatment, albeit with early pseudoprogression. Genetic studies investigating PsPD are even rarer, with most studies focusing on possible immunotherapy sensitivity. SMARCA4 mutation is reportedly associated with a favorable response to ICIs treatment in NSCLC. SMARCA4-mutant NSCLCs overlap genetically with frequent TP53 and KRAS mutations and a high TMB (41, 42). TP53 mutation has been proven to significantly increase the expression of immune checkpoints and activate the T-effector and interferon-γ signature, which contribute to the benefit conferred by ICIs. Specifically, the TP53/KRAS co-mutation subgroup manifested an exclusive increase in the expression of PD-L1 and mutational burden in the lung adenocarcinoma database, which also implies ICIs preference (43, 44). On the contrary, co-mutation of KRAS with STK11/LKB1 or KEAP1 indicated worse outcomes when ICIs are used for NSCLC therapy (44). Our study also found that EGFR, DNMT3A, and JAK2 were the most commonly mutated genes in the HPD group, while TP53, SMARCA4, and KRAS were the most commonly mutated genes in the PsPD group, consistent with previous research. Our study is rendered distinct by the fact that we used blood NGS to derive these results, which implies that blood gene testing could partly replace tissue detection.

However, some peculiarities were also observed during the course of this study. ATM belongs to the DNA damage repair pathway (DDR). Deficiency in the DDR activates innate immunity as well as tumor recognition of the adaptive immune system, leading to sensitivity to ICIs (45). However, our data showed that

TABLE 2 Gene mutation frequency change in HPD.

Gene Rank	Before treatment		After HPD	
	Gene name	Frequency	Gene name	Frequency
1	TP53	75%	TP53	100%
2	ARID2	25%	ATM	50%
3	ATM	25%	EGFR	50%
4	BLM	25%	PIK3CA	50%
5	BRAF	25%	AKT2	25%
6	CHD8	25%	APC	25%
7	DNMT3A	25%	ARID2	25%
8	DOT1L	25%	BLM	25%
9	EGFR	25%	BRAF	25%
10	GATA2	25%	BRIP1	25%
11	JAK2	25%	CCNE1	25%
12	JAK3	25%	CHD8	25%
13	KRAS	25%	CHEK2	25%
14	PDE11A	25%	CYP2D6	25%
15	PIK3CA	25%	COT1L	25%
16	RB1	25%	DTL	25%
17	SETD2	25%	EP300	25%
18	SMO	25%	EPAS1	25%
19	STAG2	25%	FANCE	25%
20	TET2	25%	FANCF	25%

ATM mutation is observed in HPD, contrary to the previous hypothesis. Similarly, we discovered PIK3CA mutation in HPD, but other studies have shown that it leads to an elevation in the level of public neoantigens, indicating sensitivity to ICI treatment (46). These differences may be attributed to the different genetic backgrounds of various neoplasms and also the limited sample size of the current study. Moreover, previous evidence about the role of ATM and PIK3CA in immunotherapy has been derived from basic research, which stands in stark contrast to clinical practice. More studies are needed on this topic in the future.

As regular lung cancer gene testing recommended by the guidelines does not fully cover the above-mentioned genes, NGS can comprehensively detect gene mutations in patients who may receive immunotherapy to predict the effect of ICIs. However, the prediction and identification of specific response patterns such as PsPD and HPD by a single gene is still inaccurate. A vast amount of data suggest that TMB could be another predictor of ICI efficacy (47). Studies have discovered that the tissue TMB (tTMB) was not correlated with PD-L1 expression, but both are associated with the clinical benefits from ICIs (48). The TMB can be evaluated using various techniques with different thresholds and can be determined using tissues and blood (49). Evidence on its predictive value is conflicting. In the CheckMate 227 study, tTMB was proven to be a prospective biomarker for PFS. However, other randomized controlled trials have failed to show a survival benefit upon stratifying patients by tTMB, and their findings do not currently support the prognostic or predictive value of tTMB in NSCLC (9, 50). On the contrary, bTMB is valuable in predicting the ICI response (10, 51, 52). To date, the prognostic and predictive value of tTMB or bTMB remains elusive, and evidence of their direct association with PsPD and HPD is lacking.

Our data showed that the bTMB before treatment partly represents the specific ICIs response, consistent with a previous study. Surprisingly, the absolute bTMB after imaging progression shows a vast difference between PsPD and HPD. More interestingly, comparison of the pre- and post-treatment values revealed that bTMB elevation was significantly higher in the HPD group than that in the PsPD group. Not only does it yield the bTMB, but also provides information about AF and ctDNA abundance. The fold change differences between PsPD and HPD were even more significant with respect to the changes in AF% and relative ctDNA abundance. In the HPD group, P08 progressed rapidly after treatment, and the interval between the two tests was only 3 weeks; therefore, the data may not represent the actual situation. If this patients' data are excluded, it can be seen that the ratios of bTMB, maximum AF%, and relative ctDNA abundance increased significantly in all patients with HPD after progression. Patients with PsPD showed the opposite result. These results may indicate

10.3389/fonc.2023.1231094

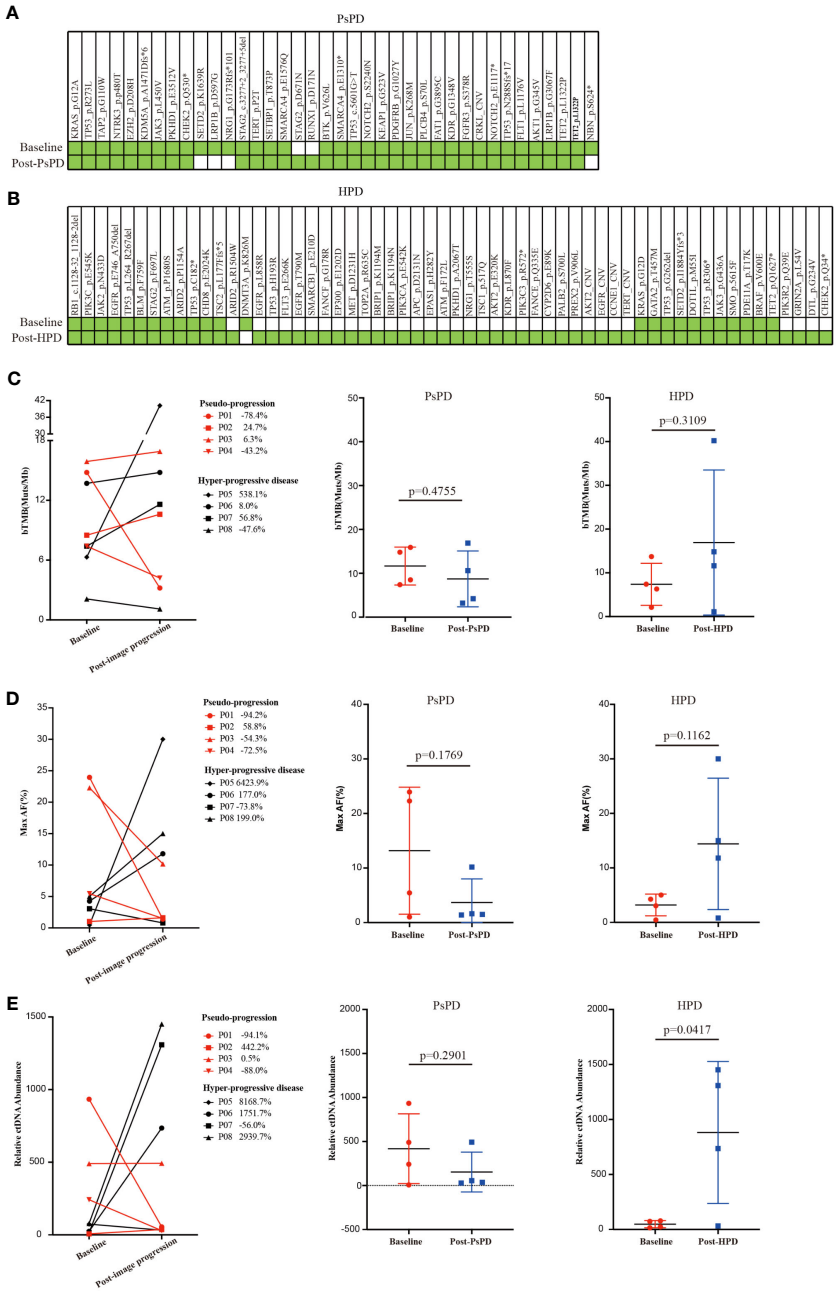


FIGURE 3
(A) Gene co-mutations in patients with PsPD at baseline and after imaging progression. **(B)** Gene co-mutations in patients with HPD at baseline and after imaging progression. **(C)** bTMB change from baseline to imaging progression in patients with PsPD/HPD left: connection diagram, middle: individual dotted diagram of bTMB at baseline, right: individual dotted diagram of bTMB after imaging progression. **(D)** Maximum AF% change from baseline to imaging progression in patients with PsPD/HPD left: connection diagram, middle: individual dots diagram of Max AF% at baseline, right: individual dotted diagram of maximum AF% after imaging progression. **(E)** Relative ctDNA abundance change from baseline to imaging progression in patients with PsPD/HPD. left: connection diagram, middle: individual dots diagram of relative ctDNA abundance at baseline, right: individual dots diagram of relative ctDNA abundance after imaging progression. Unpaired t test was used here. ct: circulating tumor, AF: allele frequency, bTMB: blood-based tumor mutational burden, PsPD: pseudoprogression, HPD: hyperprogressive disease, NGS: next-generation sequencing.

the association of the decrease in gene mutation abundance with predicting the decrease in the tumor burden; additionally, these findings underscore the potential value and superiority of dynamic blood genomic alteration in predicting the outcomes of ICIs treatment.

This study provides a novel method to distinguish HPD and PsPD in NSCLC patients receiving anti-PD1 treatment. To the best of our knowledge, this is the first study to illustrate a dynamic comparison of the whole genomic alteration by blood NGS to differentiate PsPD and HPD, thereby providing clinical evidence to evaluate the outcome of immunotherapy. The scarcity of eligible patients meeting our criteria posed a challenge in this study. However, given the clinical significance of our findings, we are committed to expanding our sample size for validation. We are

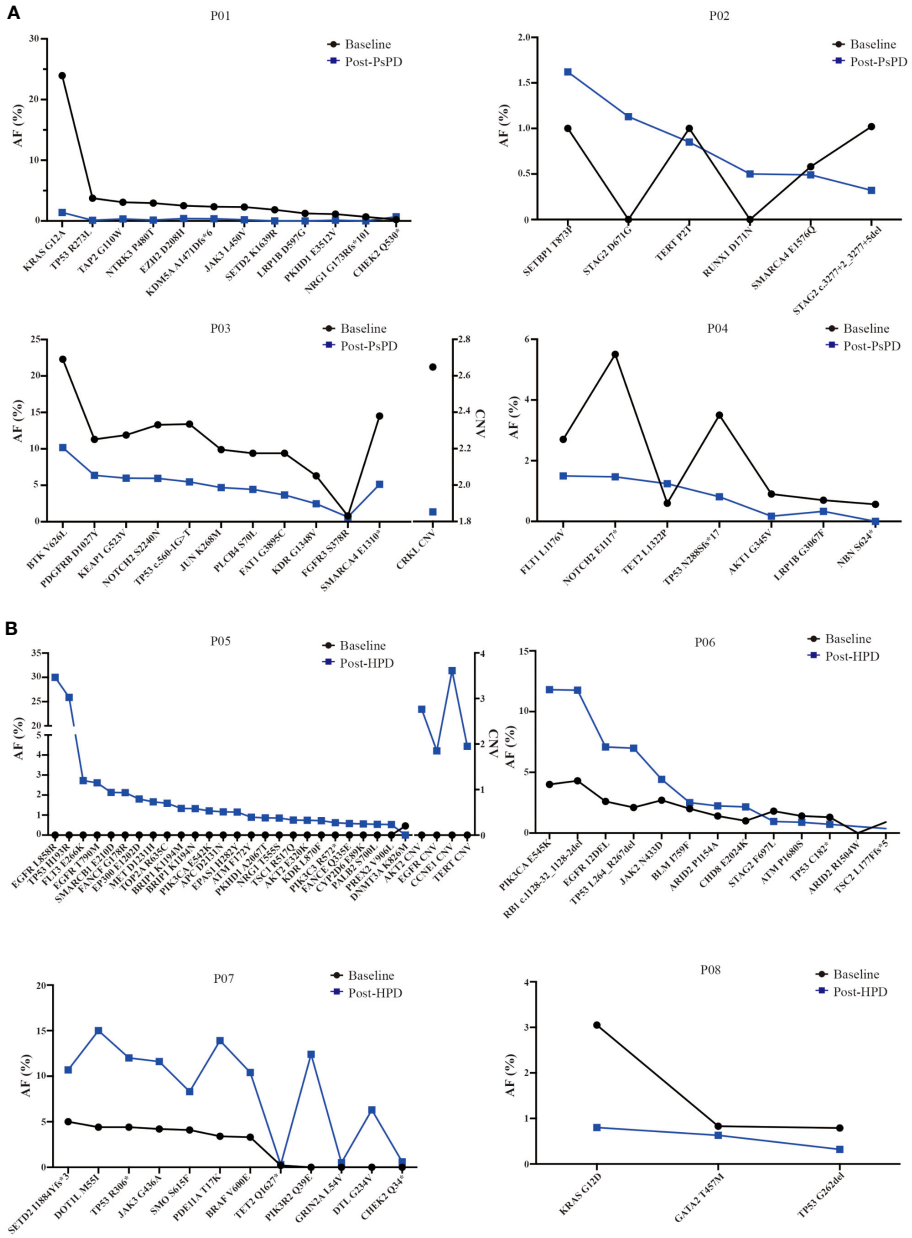


FIGURE 4
(A) AF% change in the total genomic mutation in patients with PsPD at baseline and after PsPD. The abbreviations represent different patients. **(B)** AF % change in the total genomic mutation in patients with HPD at baseline and after HPD. The abbreviations represent different patients. AF: allele frequency, CNV: copy number variation, PsPD: pseudoprogression, HPD: hyperprogressive disease.

currently employing ambulatory blood NGS detection on patients undergoing treatment at our center, and the preliminary results align with those presented here. We are also exploring cost-effective prediction methods, such as dynamic monitoring of ctDNA and CTC as other previous studies (53). Existing research has indeed highlighted disparities in gene mutation frequencies among different racial groups, particularly in targeted therapy for lung cancer (54). Immunotherapy has also shown the different efficacy across ethnicities (55). We acknowledge that racial differences in genetic testing evaluation warrant attention. It is important to note that our study focused on overall gene-level changes rather than specific gene mutations, which may not be clinically significant in this context. Nonetheless, we eagerly anticipate the accumulation of data from diverse regions and research centers to validate our findings, with the aim of benefits to lung cancer patients globally, transcending geographical boundaries, and not limited to the Chinese population. Genomic change may be associated with the adverse effects of ICIs, which is also a difficult problem. Unfortunately, the adverse effects were not included here. It is a practical issue that needs to be addressed in future research.

In conclusion, for patients with NSCLC receiving anti-PD1 treatment, the PsPD group showed a significant reduction in the bTMB, AF%, and relative ctDNA abundance in the whole genome as well as a decrease in the expression of all mutational genes, while

the opposite was observed in the HPD group. For the patients with image progression shortly after receiving anti-PD1 treatment despite better symptoms, a second blood NGS was preferred. These results broaden the scope of the dynamic genome-wide spectrum in differentiating PsPD from HPD and provide preliminary data to support the continuous blood-based NGS detection during ICI therapy for NSCLC.

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 humans were approved by the Ethics Review Board of Wuhan Union Hospital, Huazhong University of Science and Technology and the research was conducted in accordance with relevant ethical guidelines (2018-S271). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

XD: Conceptualization, Investigation, Writing – Review and Editing, Supervision, Funding acquisition. SZ: Methodology. YB: Software, Resources. FT: Validation, Data curation. RZ: Formal analysis, Writing – Original Draft, Writing – Review and Editing. NY: Investigation, Supervision. YZ: Visualization. RGZ: Project administration. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (No.82273323, No.81773233 to XD).

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.

The reviewer LW declared a shared affiliation with the authors YZ, NY to the handling editor at the time of review.

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1231094/full#supplementary-material

References

1. Reck M, Remon J, Hellmann MD. First-line immunotherapy for non-small-cell lung cancer. *J Clin Oncol Off J Am Soc Clin Oncol* (2022) 40:586–97. doi: 10.1200/JCO.21.01497

2. Moutafi M, Martinez-Morilla S, Divakar P, Vathiotis I, Gavrielatou N, Aung TN, et al. Discovery of biomarkers of resistance to immune checkpoint blockade in NSCLC using high-plex digital spatial profiling. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer* (2022) 17:991–1001. doi: 10.1016/j.jtho.2022.04.009

3. Patil NS, Nabet BY, MГjller S, Koeppen H, Zou W, Giltnane J, et al. Intratumoral plasma cells predict outcomes to PD-L1 blockade in non-small cell lung cancer. *Cancer Cell* (2022) 40:289–300.e4. doi: 10.1016/j.ccell.2022.02.002

4. Mamdani H, Matosevic S, Khalid AB, Durm G, Jalal SI. Immunotherapy in lung cancer: current landscape and future directions. *Front Immunol* (2022) 13:823618. doi: 10.3389/fimmu.2022.823618

5. Liao X, Liu M, Wang R, Zhang J. Potentials of non-invasive (18)F-FDG PET/CT in immunotherapy prediction for non-small cell lung cancer. *Front Genet* (2021) 12:810011. doi: 10.3389/fgene.2021.810011

6. Kas B, Talbot H, Ferrara R, Richard C, Lamarque JP, Pitre-Champagnat S, et al. Clarification of definitions of hyperprogressive disease during immunotherapy for non-small cell lung cancer. *JAMA Oncol* (2020) 6:1039–46. doi: 10.1001/jamaoncol.2020.1634

7. Seymour L, Bogaerts J, Perrone A, Ford R, Schwartz LH, Mandrekar S, et al. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. *Lancet Oncol* (2017) 18:e143–52. doi: 10.1016/S1470-2045(17)30074-8

8. Mosele F, Remon J, Mateo J, Westphalen CB, Barlesi F, Lolkema MP, et al. Recommendations for the use of next-generation sequencing (NGS) for patients with metastatic cancers: a report from the ESMO Precision Medicine Working Group. *Ann Oncol Off J Eur Soc Med Oncol* (2020) 31:1491–505. doi: 10.1016/j.annonc.2020.07.014

9. Powell SF, RodrГguez-Abreu D, Langer CJ, Tafreshi A, Paz-Ares L, Kopp HG, et al. Outcomes with pembrolizumab plus platinum-based chemotherapy for patients with NSCLC and stable brain metastases: pooled analysis of KEYNOTE-021, -189, and -407. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer* (2021) 16:1883–92. doi: 10.1016/j.jtho.2021.06.020

10. Rizvi NA, Cho BC, Reinmuth N, Lee KH, Luft A, Ahn MJ, et al. Durvalumab with or without tremelimumab vs standard chemotherapy in first-line treatment of metastatic non-small cell lung cancer: the MYSTIC phase 3 randomized clinical trial. *JAMA Oncol* (2020) 6:661–74. doi: 10.1001/jamaoncol.2020.0237

11. Xing L, Pan Y, Shi Y, Shu Y, Feng J, Li W, et al. Biomarkers of osimertinib response in patients with refractory, EGFR-T790M-positive non-small cell lung cancer and central nervous system metastases: the APOLLO study. *Clin Cancer Res* (2020) 26:6168–75. doi: 10.1158/1078-0432.CCR-20-2081

12. Wang X, Gao Y, Shan C, Lai M, He H, Bai B, et al. Association of circulating tumor DNA from the cerebrospinal fluid with high-risk CNS involvement in patients with diffuse large B-cell lymphoma. *Clin Trans Med* (2021) 11:e236. doi: 10.1002/ctm2.236

13. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv E-prints* (2013).

14. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. *Bioinf (Oxford England)* (2015) 31:2032–4. doi: 10.1093/bioinformatics/btv098

15. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* (2010) 20:1297–303. doi: 10.1101/gr.107524.110

16. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* (2012) 22:568–76. doi: 10.1101/gr.129684.111

17. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PloS Comput Biol* (2016) 12:e1004873. doi: 10.1371/journal.pcbi.1004873

18. den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, et al. HGVS recommendations for the description of sequence variants: 2016 update. *Hum Mutat.* (2016) (6):564–9. doi: 10.1002/humu.22981

19. Chalmers ZR, Connelly CF, Fabrizio D, Gay L, Ali SM, Ennis R, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med* (2017) 9:34. doi: 10.1186/s13073-017-0424-2

20. Fang W, Ma Y, Yin JC, Hong S, Zhou H, Wang A, et al. Comprehensive genomic profiling identifies novel genetic predictors of response to anti-PD-(L)1 therapies in non-small cell lung cancer. *Clin Cancer Res* (2019) 25:5015–26. doi: 10.1158/1078-0432.CCR-19-0585

21. Shen L, Fu H, Tao G, Liu X, Yuan Z, Ye X. Pre-immunotherapy contrast-enhanced CT texture-based classification: A useful approach to non-small cell lung cancer immunotherapy efficacy prediction. *Front Oncol* (2021) 11:591106. doi: 10.3389/fonc.2021.591106

22. Han J, Duan J, Bai H, Wang Y, Wan R, Wang X, et al. TCR repertoire diversity of peripheral PD-1(+)CD8(+) T cells predicts clinical outcomes after immunotherapy in patients with non-small cell lung cancer. *Cancer Immunol Res* (2020) 8:146–54. doi: 10.1158/2326-6066.CIR-19-0398

23. Ma Y, Wang Q, Dong Q, Zhan L, Zhang J. How to differentiate pseudoprogression from true progression in cancer patients treated with immunotherapy. *Am J Cancer Res* (2019) 9:1546–53.

24. Ferrara R, Mezquita L, Texier M, Lahmar J, Audigier-Valette C, Tessonnier L, et al. Hyperprogressive disease in patients with advanced non-small cell lung cancer treated with PD-1/PD-L1 inhibitors or with single-agent chemotherapy. *JAMA Oncol* (2018) 4:1543–52. doi: 10.1001/jamaoncol.2018.3676

25. Kamada T, Togashi Y, Tay C, Ha D, Sasaki A, Nakamura Y, et al. PD-1(+) regulatory T cells amplified by PD-1 blockade promote hyperprogression of cancer. *Proc Natl Acad Sci U.S.A.* (2019) 116:9999–10008. doi: 10.1073/pnas.1822001116

26. Ding P, Wen L, Tong F, Zhang R, Huang Y, Dong X. Mechanism underlying the immune checkpoint inhibitor-induced hyper-progressive state of cancer. *Cancer Drug Resistance (Alhambra Calif)* (2022) 5:147–64. doi: 10.20517/cdr.2021.104

27. Yi M, Niu M, Zhang J, Li S, Zhu S, Yan Y, et al. Combine and conquer: manganese synergizing anti-TGF-beta/PD-L1 bispecific antibody YM101 to overcome immunotherapy resistance in non-inflamed cancers. *J Hematol Oncol* (2021) 14:146. doi: 10.1186/s13045-021-01155-6

28. Yi M, Wu Y, Niu M, Zhu S, Zhang J, Yan Y, et al. Anti-TGF-beta/PD-L1 bispecific antibody promotes T cell infiltration and exhibits enhanced antitumor activity in triple-negative breast cancer. *J Immunother Cancer* (2022) 10.

29. Singavi AK, Menon S, Kilari D, Alqwasmi A, Ritch PS, Thomas JP, et al. 1140PD - Predictive biomarkers for hyper-progression (HP) in response to immune checkpoint inhibitors (ICI) – analysis of somatic alterations (SAs). *Ann Oncol* (2017) 28:v405.

30. Economopoulou P, Anastasiou M, Papaxoinis G, Spathas N, Spathis A, Oikonomopoulos N, et al. Patterns of response to immune checkpoint inhibitors in association with genomic and clinical features in patients with head and neck squamous cell carcinoma (HNSCC). *Cancers (Basel)* (2021) 13. doi: 10.3390/cancers13020286

31. Kato S, Ross JS, Gay L, Dayyani F, Roszik J, Subbiah V, et al. Analysis of MDM2 amplification: next-generation sequencing of patients with diverse Malignancies. *JCO Precis Oncol* (2018). doi: 10.1200/PO.17.00235

32. Akbay EA, Koyama S, Carretero J, Altabef A, Tchaicha JH, Christensen CL, et al. Activation of the PD-1 pathway contributes to immune escape in EGFR-driven lung tumors. *Cancer Discovery* (2013) 3:1355–63. doi: 10.1158/2159-8290.CD-13-0310

33. Wang X, Wang F, Zhong M, Yarden Y, Fu L. The biomarkers of hyperprogressive disease in PD-1/PD-L1 blockage therapy. *Mol Cancer* (2020) 19:81. doi: 10.1186/s12943-020-01200-x

34. Sun L, Zhang L, Yu J, Zhang Y, Pang X, Ma C, et al. Clinical efficacy and safety of anti-PD-1/PD-L1 inhibitors for the treatment of advanced or metastatic cancer: a systematic review and meta-analysis. *Sci Rep* (2020) 10:2083. doi: 10.1038/s41598-020-58674-4

35. Mao S, Zhang J, Guo Y, Zhang Z, Wu Y, Zhang W, et al. Hyperprogression after anti-programmed cell death ligand-1 therapy in a patient with recurrent metastatic urothelial bladder carcinoma following first-line cisplatin-based chemotherapy: a case report. *Drug Design Dev Ther* (2019) 13:291–300. doi: 10.2147/DDDT.S181122

36. Kato S, Goodman A, Walavalkar V, Barkauskas DA, Sharabi A, Kurzrock R. Hyperprogressors after immunotherapy: analysis of genomic alterations associated with accelerated growth rate. *Clin Cancer Res* (2017) 23:4242–50. doi: 10.1158/1078-0432.CCR-16-3133

37. Shin DS, Zaretsky JM, Escuin-Ordinas H, Garcia-Diaz A, Hu-Lieskovan S, Kalbasi A, et al. Primary resistance to PD-1 blockade mediated by JAK1/2 mutations. *Cancer Discovery* (2017) 7:188–201. doi: 10.1158/2159-8290.CD-16-1223

38. Zaretsky JM, Garcia-Diaz A, Shin DS, Escuin-Ordinas H, Hugo W, Hu-Lieskovan S, et al. Mutations associated with acquired resistance to PD-1 blockade in melanoma. *New Engl J Med* (2016) 375:819–29. doi: 10.1056/NEJMoa1604958

39. Saleh K, Khalife-Saleh N, Haddad EE, Kourie HR. Negative predictive biomarkers of checkpoint inhibitors in hyper-progressive tumors. *Biomarkers Med* (2017) 11:819–21. doi: 10.2217/bmm-2017-0168

40. Lopci E. Immunotherapy monitoring with immune checkpoint inhibitors based on [(18)F]FDG PET/CT in metastatic melanomas and lung cancer. *J Clin Med* (2021) 10.

41. Miao D, Margolis CA, Vokes NI, Liu D, Taylor-Weiner A, Wankowicz SM, et al. Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors. *Nat Genet* (2018) 50:1271–81. doi: 10.1038/s41588-018-0200-2

42. Nambirajan A, Jain D. Recent updates in thoracic SMARCA4-deficient undifferentiated tumor. *Semin Diagn Pathol* (2021) 38:83–9. doi: 10.1053/j.semdp.2021.06.001

43. Dong ZY, Zhong WZ, Zhang XC, Su J, Xie Z, Liu SY, et al. Potential predictive value of TP53 and KRAS mutation status for response to PD-1 blockade immunotherapy in lung adenocarcinoma. *Clin Cancer Res* (2017) 23:3012–24. doi: 10.1158/1078-0432.CCR-16-2554

44. Davis AP, Cooper WA, Boyer M, Lee JH, Pavlakis N, Kao SC. Efficacy of immunotherapy in KRAS-mutant non-small-cell lung cancer with comutations. *Immunotherapy* (2021) 13:941–52. doi: 10.2217/imt-2021-0090

45. Jiang M, Jia K, Wang L, Li W, Chen B, Liu Y, et al. Alterations of DNA damage response pathway: Biomarker and therapeutic strategy for cancer immunotherapy. *Acta Pharm Sinica B* (2021) 11:2983–94. doi: 10.1016/j.apsb.2021.01.003

46. Chandran SS, Ma J, Klatt MG, Dٌjndar F, Bandlamudi C, Razavi P, et al. Immunogenicity and therapeutic targeting of a public neoantigen derived from mutated PIK3CA. *Nat Med* (2022) 28:946–57. doi: 10.1038/s41591-022-01786-3

47. Duchemann B, Remon J, Naigeon M, Cassard L, Jouniaux JM, Boselli L, et al. Current and future biomarkers for outcomes with immunotherapy in non-small cell lung cancer. *Trans Lung Cancer Res* (2021) 10:2937–54. doi: 10.21037/tlcr-20-839

48. Rizvi H, Sanchez-Vega F, La K, Chatila W, Jonsson P, Halpenny D, et al. Molecular determinants of response to anti-programmed cell death (PD)-1 and anti-programmed death-ligand 1 (PD-L1) blockade in patients with non-small-cell lung cancer profiled with targeted next-generation sequencing. *J Clin Oncol Off J Am Soc Clin Oncol* (2018) 36:633–41. doi: 10.1200/JCO.2017.75.3384

49. Chen X, Fang L, Zhu Y, Bao Z, Wang Q, Liu R, et al. Blood tumor mutation burden can predict the clinical response to immune checkpoint inhibitors in advanced non-small cell lung cancer patients. *Cancer Immunol Immunother CII* (2021) 70:3513–24. doi: 10.1007/s00262-021-02943-2

50. Hellmann MD, Paz-Ares L, Bernabe Caro R, Zurawski B, Kim SW, Carcereny Costa E, et al. Nivolumab plus ipilimumab in advanced non-small-cell lung cancer. *New Engl J Med* (2019) 381:2020–31. doi: 10.1056/NEJMoa1910231

51. Socinski M, Velcheti V, Mekhail T, Chae YK, Leal TA, Dowell JE, et al. LBA83 - Final efficacy results from B-F1RST, a prospective phase II trial evaluating blood-based tumour mutational burden (bTMB) as a predictive biomarker for atezolizumab (atezo) in 1L non-small cell lung cancer (NSCLC). *Ann Oncol* (2019) 30:v919–20. doi: 10.1093/annonc/mdz394.081

52. Si H, Kuziora M, Quinn KJ, Helman E, Ye J, Liu F, et al. A blood-based assay for assessment of tumor mutational burden in first-line metastatic NSCLC treatment: results from the MYSTIC study. *Clin Cancer Res* (2021) 27:1631–40. doi: 10.1158/1078-0432.CCR-20-3771

53. Nikanjam M, Kato S, Kurzrock R. Liquid biopsy: current technology and clinical applications. *J Hematol Oncol* (2022) 15:131. doi: 10.1186/s13045-022-01351-y

54. James BA, Williams JL, Nemesure B. A systematic review of genetic ancestry as a risk factor for incidence of non-small cell lung cancer in the US. *Front Genet* (2023) 14:1141058. doi: 10.3389/fgene.2023.1141058

55. Qian J, Nie W, Lu J, Zhang L, Zhang Y, Zhang B, et al. Racial differences in characteristics and prognoses between Asian and white patients with nonsmall cell lung cancer receiving atezolizumab: An ancillary analysis of the POPLAR and OAK studies. *Int J Cancer* (2020) 146:3124–33. doi: 10.1002/ijc.32717

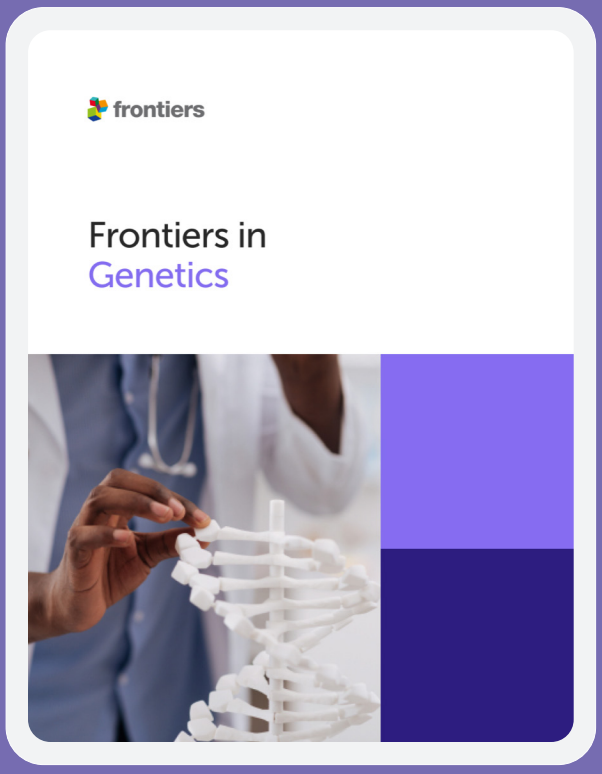
Frontiers in Genetics

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