



OPEN ACCESS

EDITED BY

Xu Chen,
The First Affiliated Hospital of Soochow
University, China

REVIEWED BY

Qi-Tong Xu,
Tongji University, China
Yanpeng Zhang,
First Affiliated Hospital of Xi'an Jiaotong
University, China

*CORRESPONDENCE

Daqiang Sun
✉ sdqmdsr@163.com
Xin Li
✉ 653206889@qq.com
Kai Wang
✉ nickwang1996@163.com

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

RECEIVED 05 January 2024

ACCEPTED 13 March 2024

PUBLISHED 26 March 2024

CITATION

Zhang H, Zhang P, Lin X, Tan L, Wang Y, Jia X,
Wang K, Li X and Sun D (2024) Integrative
single-cell analysis of LUAD: elucidating
immune cell dynamics and prognostic
modeling based on exhausted CD8+ T cells.
Front. Immunol. 15:1366096.
doi: 10.3389/fimmu.2024.1366096

COPYRIGHT

© 2024 Zhang, Zhang, Lin, Tan, Wang, Jia,
Wang, Li and Sun. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).
The use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Integrative single-cell analysis of LUAD: elucidating immune cell dynamics and prognostic modeling based on exhausted CD8+ T cells

Han Zhang^{1†}, Pengpeng Zhang^{2†}, Xuefeng Lin^{3†}, Lin Tan^{4†},
Yuhang Wang¹, Xiaoteng Jia¹, Kai Wang^{5*}, Xin Li^{5*}
and Daqiang Sun^{5*}

¹Clinical School of Thoracic, Tianjin Medical University, Tianjin, China, ²Department of Lung Cancer Surgery, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China, ³Tianjin Medical College, Tianjin, China, ⁴Qingdao Hospital, University of Health and Rehabilitation Sciences, Qingdao Municipal Hospital, Qingdao, China, ⁵Tianjin Chest Hospital, Tianjin University, Tianjin, China

Background: The tumor microenvironment (TME) plays a pivotal role in the progression and metastasis of lung adenocarcinoma (LUAD). However, the detailed characteristics of LUAD and its associated microenvironment are yet to be extensively explored. This study aims to delineate a comprehensive profile of the immune cells within the LUAD microenvironment, including CD8+ T cells, CD4+ T cells, and myeloid cells. Subsequently, based on marker genes of exhausted CD8+ T cells, we aim to establish a prognostic model for LUAD.

Method: Utilizing the Seurat and Scanpy packages, we successfully constructed an immune microenvironment atlas for LUAD. The Monocle3 and PAGA algorithms were employed for pseudotime analysis, pySCENIC for transcription factor analysis, and CellChat for analyzing intercellular communication. Following this, a prognostic model for LUAD was developed, based on the marker genes of exhausted CD8+ T cells, enabling effective risk stratification in LUAD patients. Our study included a thorough analysis to identify differences in TME, mutation landscape, and enrichment across varying risk groups. Moreover, by integrating risk scores with clinical features, we developed a new nomogram. The expression of model genes was validated via RT-PCR, and a series of cellular experiments were conducted, elucidating the potential oncogenic mechanisms of GALNT2.

Results: Our study developed a single-cell atlas for LUAD from scRNA-seq data of 19 patients, examining crucial immune cells in LUAD's microenvironment. We underscored pDCs' role in antigen processing and established a Cox regression model based on CD8_Tex-LAYN genes for risk assessment. Additionally, we contrasted prognosis and tumor environments across risk groups, constructed a new nomogram integrating clinical features, validated the expression of model genes via RT-PCR, and confirmed GALNT2's function in LUAD through cellular experiments, thereby enhancing our understanding and approach to LUAD treatment.

Conclusion: The creation of a LUAD single-cell atlas in our study offered new insights into its tumor microenvironment and immune cell interactions, highlighting the importance of key genes associated with exhausted CD8+ T cells. These discoveries have enabled the development of an effective prognostic model for LUAD and identified GALNT2 as a potential therapeutic target, significantly contributing to the improvement of LUAD diagnosis and treatment strategies.

KEYWORDS

LUAD, myeloid, CD4+ T cells, CD8+ T cells, pDCs, GALNT2

1 Introduction

In the past decade, LUAD has emerged as the predominant subtype of lung cancer, accounting for approximately 40% of all lung cancer cases (1, 2). The incidence of this form of lung cancer is not only increasing among chronic smokers but also shows an upward trend in former and even non-smokers. Despite significant advancements in treatment modalities such as the introduction of PD-L1-targeted immunotherapies (3, 4) and targeted therapies for specific genetic mutations like KRAS and EGFR (5, 6), LUAD continues to pose a major public health challenge globally. The five-year survival rate for this malignancy varies based on region, treatment accessibility, and individual differences, generally remaining low, typically ranging from 5% to 20% (7).

TME refers to the environment surrounding tumor cells, encompassing nearby blood vessels, immune cells, fibroblasts, the extracellular matrix, and secreted molecules (8). It plays a critical role in tumor development, metastasis, and response to treatment. In LUAD, characteristics of the TME such as immune cell infiltration, local inflammatory responses, and interactions between tumor cells and their surrounding environment are key factors affecting disease progression and treatment efficacy (9). Increasing research emphasizes the significance of the TME in predicting cancer prognosis and responses to immunotherapy (10, 11). Understanding these aspects not only aids in deepening our comprehension of the biological characteristics of LUAD but also may facilitate the development of new treatment strategies, such as the use of immunomodulatory drugs or therapies targeting specific features of the TME.

Single-cell RNA sequencing (scRNA-seq) represents a significant advancement over conventional bulk RNA-sequencing (RNA-seq) by offering detailed transcriptomic analysis at the individual cell level (12). This advanced methodology has been transformative in tumor microenvironment research, yielding high-resolution insights across various cancers. Studies in malignancies such as lymphoma, melanoma, and liver cancer have utilized scRNA-seq to uncover complex details regarding immune cell heterogeneity, composition, and regulatory mechanisms within

their respective tumor microenvironments (13–15). Despite these advancements, the exploration of the LUAD tumor microenvironment using scRNA-seq is still in its nascent stage, underscoring an urgent need for more in-depth studies in this area to enhance our understanding.

In this study, our objective was to construct a comprehensive single-cell atlas of LUAD by performing scRNA-seq on 29 samples collected from 19 LUAD patients (16). Utilizing this atlas, we meticulously analyzed the canonical immune cell types within the LUAD microenvironment, with a focus on myeloid cells, CD4+ T cells, and CD8+ T cells. We identified plasmacytoid dendritic cells (pDCs) as key regulators in antigen processing and presentation via the MHC-II signaling pathway. By integrating scRNA-seq data with bulk RNA-seq data from LUAD patients, we successfully identified a set of risk genes based on CD8_Tex-LAYN (exhausted CD8+ T cells) marker genes. Subsequently, we developed a robust and reliable Cox regression model that accurately assesses the risk levels of LUAD patients. Based on risk stratification, we systematically evaluated differences in prognosis, tumor microenvironment, and mutation landscape between high-risk and low-risk groups. Furthermore, combining clinical characteristics and risk scores, we constructed a novel nomogram. Lastly, we validated the expression of model genes via RT-PCR and determined GALNT2 as a potential therapeutic target for LUAD through a series of cellular experiments. These findings not only enhance our understanding of the LUAD microenvironment but also offer promising prospects for improving diagnostics and prognosis in LUAD clinical settings.

2 Methods

2.1 Data collection of LUAD single-cell datasets

The original scRNA-seq data used in this study were obtained from 19 treatment-naïve LUAD patients (16), including primary lung tumors (tLung, n = 11), distant normal lung tissues (nLung, n = 11),

and lymph node metastases (mLN, $n = 7$). These data were acquired from the Gene Expression Omnibus (GEO) database under the accession number GSE131907. The training data for our prognostic model, including the gene expression matrix (FPKM format), clinical specifics, and mutation details for LUAD patients, were directly downloaded from The Cancer Genome Atlas (TCGA) via the following portal: <https://portal.gdc.cancer.gov/repository>. Additionally, the external validation cohorts, comprising gene expression and clinical data (GSE31210, GSE37745, GSE50081, GSE68465, GSE3141), were acquired from the GEO database to assess the model's predictive accuracy. To make the data from TCGA and GEO more compatible, we converted the format from FPKM to TPM and then used the R package 'SVA' to correct for batch effects between the datasets.

2.2 Evaluating data from single-cell RNA sequencing

The gene-cell matrix for each sample was individually imported to Scanpy (version 1.9.1) for downstream analysis. Cells with the unusual number of UMIs ($\geq 8,000$), number of detected genes (≤ 500 or $\geq 4,000$) or mitochondrial gene percent ($\geq 10\%$) were excluded. The UMI count for the genes in each cell was normalized by the "LogNormalize" method as the following formula:

The gene-cell matrix for each sample was imported into Scanpy (version 1.9.1) (17) for downstream analysis. To ensure data quality, cells with unusual characteristics were excluded from the analysis. These characteristics included an abnormal number of UMIs ($\geq 8,000$), a low number of detected genes (≤ 500) or a high number of detected genes ($\geq 4,000$), as well as a high percentage of mitochondrial genes ($\geq 10\%$). Subsequently, the UMI count for each gene in every cell was normalized using the "LogNormalize" method as the following formula:

$$\text{Gene A expression level} = \log\left(1 + \frac{UMI_A}{UMI_{Total}} \times 10^5\right)$$

To alleviate the impact of batch effects during the clustering process, we utilized Harmony to integrate all the samples. Specifically, we identified 2,000 highly variable genes in each sample using a variance stabilizing transformation. Next, we determined anchors between the individual datasets and computed correction vectors to create an integrated expression matrix. This integrated matrix was then utilized for subsequent cell clustering, enhancing the robustness and reliability of the analysis by effectively addressing batch effects. The integrated expression matrix was utilized to calculate the principal components (PCs). From these PCs, a subset of significant ones was selected, and the cells were categorized into sub-clusters using the Louvain algorithm. To visualize the clustering results, we employed the Uniform Manifold Approximation and Projection (UMAP) technique. The cell type annotations for each cluster were determined based on the expression patterns of known marker genes using CellTypist (<https://github.com/Teichlab/celltypist>). The

expression of each gene in a given cluster was compared to the rest of the cells using the Wilcoxon rank sum test. Genes meeting the following criteria were considered as significantly upregulated: Firstly, they exhibited a $\log_2(\text{foldchange})$ value ≥ 1 or ≤ -1 , indicating substantial overexpression in the target cluster. Secondly, these genes were expressed in more than 25% of the cells within the target cluster. Finally, the adjusted p-value for each gene was required to be less than 0.05, demonstrating statistical significance in the differential expression analysis. Monocle3 R package (<https://cole-trapnell-lab.github.io/monocle3/>) was used for pseudotime trajectory analysis (18).

2.3 Cell-cell communication analysis

We employed the CellChat toolkit (<https://github.com/sqjin/CellChat>) within the R programming environment (19). This analysis aimed to elucidate the differential interactions and signaling pathways among various cell types, including DC1, DC2, pDC, Monocytes, Migratory DCs, and Mast cells. CellChat, a cutting-edge tool, enables the quantitative inference of intercellular communication networks from scRNA-seq data. It leverages a comprehensive database of human ligand-receptor interactions and advanced pattern recognition techniques to predict the primary signaling mechanisms among cells, thereby illuminating the coordination of cellular functions. To ensure the analysis's relevance and accuracy, only ligand-receptor pairs exhibiting a P-value less than 0.05 were considered, allowing for a focused evaluation of the intricate relationships between diverse cell types.

2.4 Simultaneous gene regulatory network analysis

To assess the transcriptional distinctions among cell clusters (CD8_Tn-LEF1, CD8_Tem-GZMK, CD8_Trm-KLRB1, and CD8_Tex-LAYN), based on transcription factors and their target genes, we conducted pySCENIC (version 0.10.0) analysis on all single cells. This approach facilitated the identification of regulons preferentially expressed within these clusters, employing the Limma package for calculation. Our analysis focused solely on regulons that exhibited significant upregulation or downregulation in at least one of the clusters, considering only those with an adjusted p-value of less than 0.05 for further exploration.

2.5 Construction and validation of prognostic signature

The prognostic relevance of CD8_Tex-LAYN and CD8_Tn-LEF1 as cellular biomarkers for overall survival in patients with lung adenocarcinoma (LUAD) from the TCGA database was initially evaluated through univariate Cox regression, pinpointing genes of prognostic significance at a threshold of $p < 0.01$. Subsequently, we refined our prognostic gene assessment using the "glmnet" R package

(20) to implement a LASSO Cox proportional hazards model. Further refinement was achieved by integrating this model with a multivariate Cox regression analysis, thereby constructing a risk model that incorporates gene expression levels weighted by their associated risk coefficients. This composite approach yielded six principal genes warranting additional prognostic scrutiny. To differentiate between high and low-risk patient cohorts, a median value cutoff was applied. The survival outcomes for these cohorts within the TCGA-LUAD dataset and across five independent GEO datasets (GSE3141, GSE50081, GSE68465, GSE37745, and GSE31210) were analyzed using Kaplan-Meier survival curves, facilitated by the “survival” and “survminer” R packages. The prognostic potency of the derived risk score was corroborated by generating ROC curves and computing the AUC with the “survivalROC” R package (21), thereby quantifying the risk model’s predictive precision.

2.6 Nomogram development

In developing the nomogram, we first executed both univariate and multivariate Cox regression analyses on the risk scores and clinical characteristics of patients with lung adenocarcinoma (LUAD) from the TCGA database. The objective was to pinpoint variables that held significant prognostic weight for LUAD. Following this, using the ‘rms’ R package (22), we constructed the final nomogram. To evaluate its effectiveness, we employed calibration curves and decision curve analysis (DCA).

2.7 Mutation analysis

For the analysis of mutation data and clinical details, we utilized the ‘maftools’ R package. The function `read.maf` was applied to import this information into a maf file format. We then employed `plotmafSummary` to examine the mutation profile of patients with LUAD in the TCGA dataset. For visualizing the mutation characteristics in both high and low-risk groups, the `oncplot` function was used to create heatmaps that integrated clinical data with mutation information. To investigate co-mutation patterns among key genes and the top 10 most frequently altered genes in TCGA-LUAD, the `somaticInteractions` function was deployed. Tumor Mutational Burden (TMB) refers to the total number of non-synonymous mutations within the genome of tumor cells. We downloaded the maf files for LUAD from the TCGA database, then assessed the TMB levels for each patient. Subsequently, we applied a logarithmic transformation to reduce data skewness and minimize the impact of outliers on the analysis results. Finally, we compared the differences in TMB between the high-risk and low-risk groups, and calculated the correlation between the risk score and TMB.

2.8 Enrichment analysis

Our study commenced with the implementation of the ‘GSVA’ algorithm (23), aiming to identify Hallmarker pathways that were predominantly enriched in the high-risk group compared to the

low-risk group. Following this, we employed the ‘GSEA’ algorithm (24) to distinguish and analyze pivotal Kyoto Encyclopedia of Genes and Genomes Pathways (KEGG) (25) in both high and low-risk groups. The ssGSEA algorithm (26) played a crucial role in evaluating the correlation between enrichment scores, tumor immunity cycle, and tumor-related pathways. It also proved instrumental in measuring variances in immune cell types and immune functions across the high and low-risk categories.

2.9 Assessment of the tumor microenvironment

We procured data regarding immune cell infiltration from seven different databases via the Timer2.0 platform, accessible at (<http://timer.comp-genomics.org/>) (27). Subsequently, we conducted an analysis to compare the levels of immune cell infiltration between high and low-risk groups. Utilizing the ‘estimate’ R package (28), we calculated the stromal, immune, and TumorPurity scores, along with the ESTIMATE scores for each specimen in the TCGA-LUAD dataset.

2.10 Immunotherapy response evaluation

Differences in the expression of immune checkpoint genes and Major Histocompatibility Complex (MHC) genes were compared between the high and low-risk groups. Correlations between central genes, risk scores, and these immune-related genes were also calculated. Variations in Immunophenoscore (IPS) between the high and low-risk groups were assessed. Additionally, the potential for immune escape in these groups was evaluated using the tumor immune dysfunction and exclusion (TIDE) algorithm (29).

2.11 Cultivation of cell lines

In our laboratory setting, normal human lung epithelial cells (BEAS-2B) and LUAD cell lines (A549, H1299, H1975, H1650) were procured from the Cell Resource Center at the Shanghai Institute for Biological Sciences. These cells were cultured in RPMI-1640 medium, produced by Gibco BRL, USA, supplemented with 10% fetal bovine serum (FBS) obtained from Cell-Box, Hong Kong, and 1% penicillin-streptomycin solution, supplied by Biosharp, China. Cultivation of these cells was carried out in a controlled environment, maintained at 37°C with 5% CO₂ and 95% humidity.

2.12 RNA extraction and reverse transcription PCR analysis

Total RNA was isolated from the cell lines using the TRIzol reagent (15596018, Thermo Fisher Scientific), adhering to the instructions provided by the manufacturer. The PrimeScriptTM RT kit (R232-01, Vazyme) was then utilized for synthesizing cDNA.

This was followed by quantitative RT-PCR analysis, conducted using the SYBR Green Master Mix (Q111-02, Vazyme), with GAPDH mRNA serving as the normalization control. The relative gene expression levels were determined employing the $2^{-\Delta\Delta Ct}$ method. Primers used in this study were sourced from Nanjing Sunbio Technology Co., Ltd. (Nanjing, China), detailed in [Supplementary Table 1](#).

2.13 Migration and invasion analysis via transwell assays

Migration and invasion capacities were tested using 24-well transwell inserts, with A549 and H1299 cells seeded at 1×10^5 cells in the upper chamber. In the invasion assays, we prepared the chambers by pre-coating them with matrigel from BD Biosciences, USA. For migration assays, however, the chambers were left uncoated. Following the migration or invasion process, cells located on the lower side of the membrane were fixed and subsequently stained using crystal violet sourced from Solarbio, China.

2.14 Assessment of cellular proliferation using the CCK-8 method

In this procedure, cells were plated in 96-well plates at a concentration of 3×10^3 cells per well. Following cell seeding, 10 mL of CCK-8 solution (A311-01, Vazyme) was added to each well. The plates were then incubated in the dark at 37°C for 2 hours. The proliferation of cells was determined by measuring absorbance at 450 nm with a spectrophotometer (A33978, Thermo Fisher Scientific) at time points of 0, 24, 48, 72, and 96 hours.

2.15 Assessment of colony formation

In this part of the study, cells were seeded at a density of 1×10^3 cells per well in 6-well plates, followed by a growth period of 14 days. After this incubation, cells were washed with PBS and fixed with 4% paraformaldehyde for 15 minutes. The colonies were then stained with Crystal Violet, provided by Solarbio, China.

2.16 Statistical methods

Bioinformatics data were analyzed using R (version 4.3.1), while Graphpad and ImageJ were employed for experimental data. Intergroup differences were evaluated using the T-test or one-way ANOVA for normal distributions, and Wilcoxon or Kruskal-Wallis tests for non-normal data. Survival was analyzed via Kaplan-Meier curves and Log-rank tests. Spearman's correlation assessed relationships between datasets. Significance was set at $p < 0.05$, with $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ indicating varying significance levels.

3 Results

3.1 Construction of a scRNA-seq Atlas for LUAD

To gain valuable insights into the cellular composition of the tumor microenvironment in LUAD, we performed scRNA-seq analysis on 19 treatment-naïve LUAD patients. This included collecting samples from primary lung tumors (tLung, $n = 11$), distant normal lung tissues (nLung, $n = 11$), and lymph node metastases (mLN, $n = 7$). We carefully addressed batch effects and annotated canonical cell markers (detailed in the Methods section), enabling us to classify a total of 107,751 cells into seven major cell lineages: epithelial cells, stromal cells (including fibroblasts and endothelial cells), and immune cells (T cells, B cells, myeloid cells, and NK cells) ([Figure 1A](#)). Notably, the cellular composition demonstrated consistency across different patients ([Figure 1B](#)), further validating our analysis. As anticipated, these major cell lineages exhibited the expression of specific canonical marker genes ([Figure 1C](#)) and formed distinct clusters ([Figure 1D](#)), affirming the accuracy of our lineage classification. In summary, our integration of scRNA-seq data from 19 LUAD patients successfully led to the construction of a comprehensive single-cell atlas of LUAD. This atlas serves as a solid foundation for our subsequent analyses, which aim to delve into the cellular composition and function of the LUAD tumor microenvironment.

3.2 Analysis of macrophage profiles in LUAD

In addition to T lymphocytes, the significance of myeloid cells as crucial components of tumor-infiltrating cells and regulators of tumor inflammation and angiogenesis has been emphasized in numerous studies ([30, 31](#)). To gain insights into the subsets of myeloid cells in LUAD, we categorized a total of 33,561 myeloid cells into 31 sub-clusters using the Louvain algorithm (resolution = 1.0) ([Figure 2A](#)). These sub-clusters were further assigned to nine major subsets based on canonical cell markers ([Figure 2B](#)). The major myeloid subsets identified include macrophages (alveolar M ϕ , monocyte-derived M ϕ , and interstitial M ϕ perivascular), monocytes, mast cells, and dendritic cells (DC1, DC2, pDC, and migratory DCs). The expression of gene signatures specific to each subset confirmed the accuracy of our cell annotation ([Figure 2C](#)). These findings contribute to the characterization of the diverse myeloid cell populations within the LUAD tumor microenvironment, providing a foundation for further understanding their roles in tumor progression and immunomodulation.

Subsequently, we examined the relative proportion and cell numbers of these myeloid cell subsets across different tissues. Interestingly, macrophages exhibited significant variations both in proportion ([Figure 2D](#)) and cell numbers ([Figure 2E](#)) compared to other cell lineages. To ensure the reliability and validity of our analysis, we specifically focused on macrophages as a distinct population within the total myeloid cell population, in order to explore their unique properties and functions. By re-clustering the

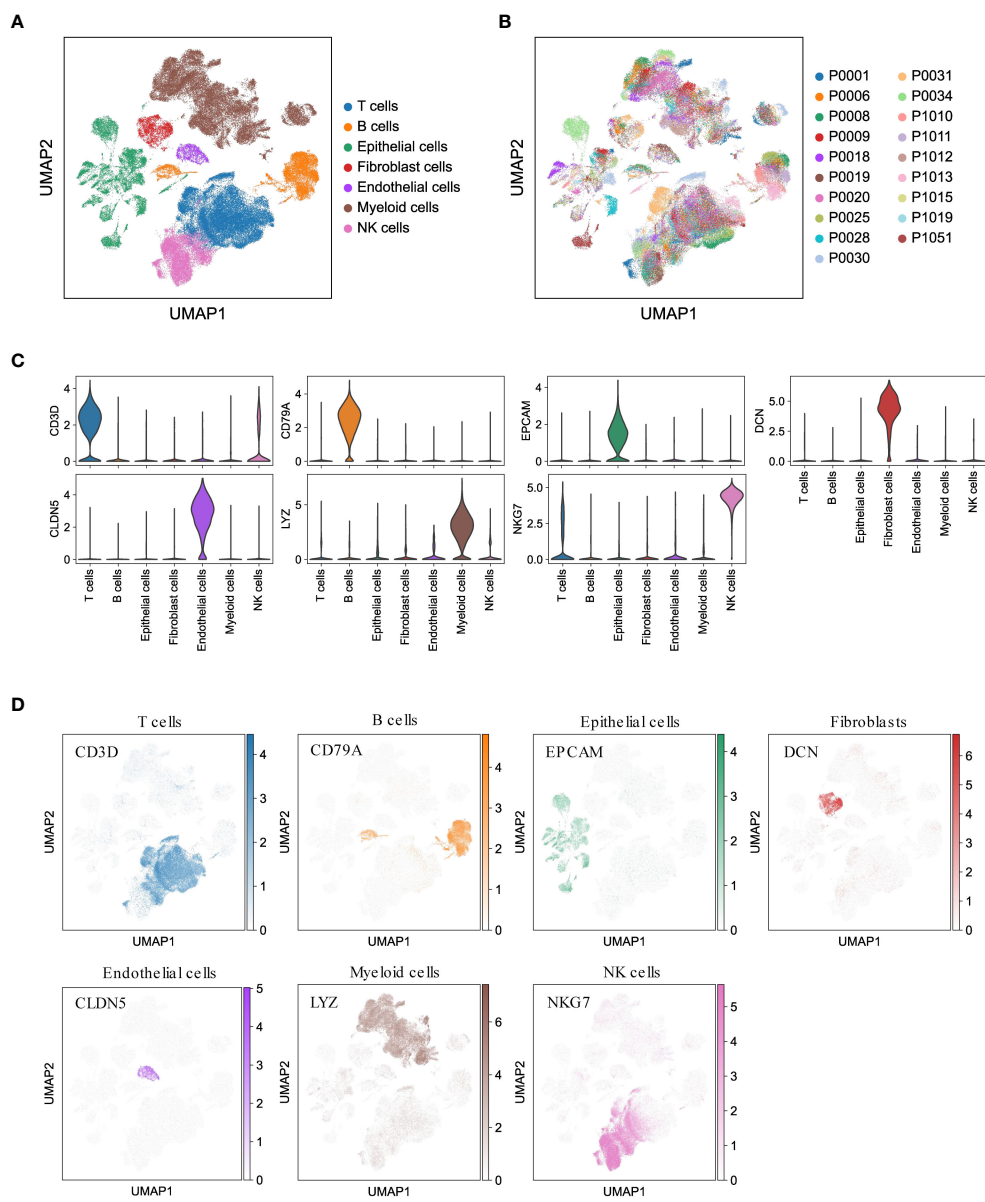


FIGURE 1
 Comprehensive LUAD single-cell atlas. (A, B) UMAP plots of 107,751 cells from 19 patients, colored by major cell lineages and patients. (C) Violin plots for expression of canonical marker genes in each major cell lineage. (D) UMAP plots of seven major cell lineages, colored by corresponding canonical marker genes.

22,959 macrophages using UMAP (Figure 2F) analyses, we identified three distinct macrophage subsets: alveolar M ϕ , monocyte-derived M ϕ , and interstitial M ϕ perivascular. Each of these subsets exhibited unique gene signatures (Figure 2G). Furthermore, KEGG pathway analysis revealed that these tumor-associated macrophages (TAMs) in LUAD primarily participate in immune regulation, cell proliferation, and antigen processing and presentation (Figure 2H). These findings are consistent with the well-established role of macrophages in shaping the tumor microenvironment and engaging in tumor immunity (32, 33). By investigating the distinct subsets and functions of macrophages within the LUAD tumor microenvironment, we gain a better understanding of their contributions to tumor progression and

immune responses. This knowledge holds potential for targeted therapeutic strategies aimed at modulating macrophage-mediated immune regulation in LUAD.

3.3 Analysis of monocytes, mast cells, and dendritic cells in LUAD

To gain further insights into the properties and functions of myeloid cells other than macrophages, we employed UMAP and tSNE methods to re-cluster the remaining 10,602 myeloid cells. This analysis resulted in the identification of seven subsets, including monocytes, mast cells, and four types of dendritic cells (DC1, DC2,

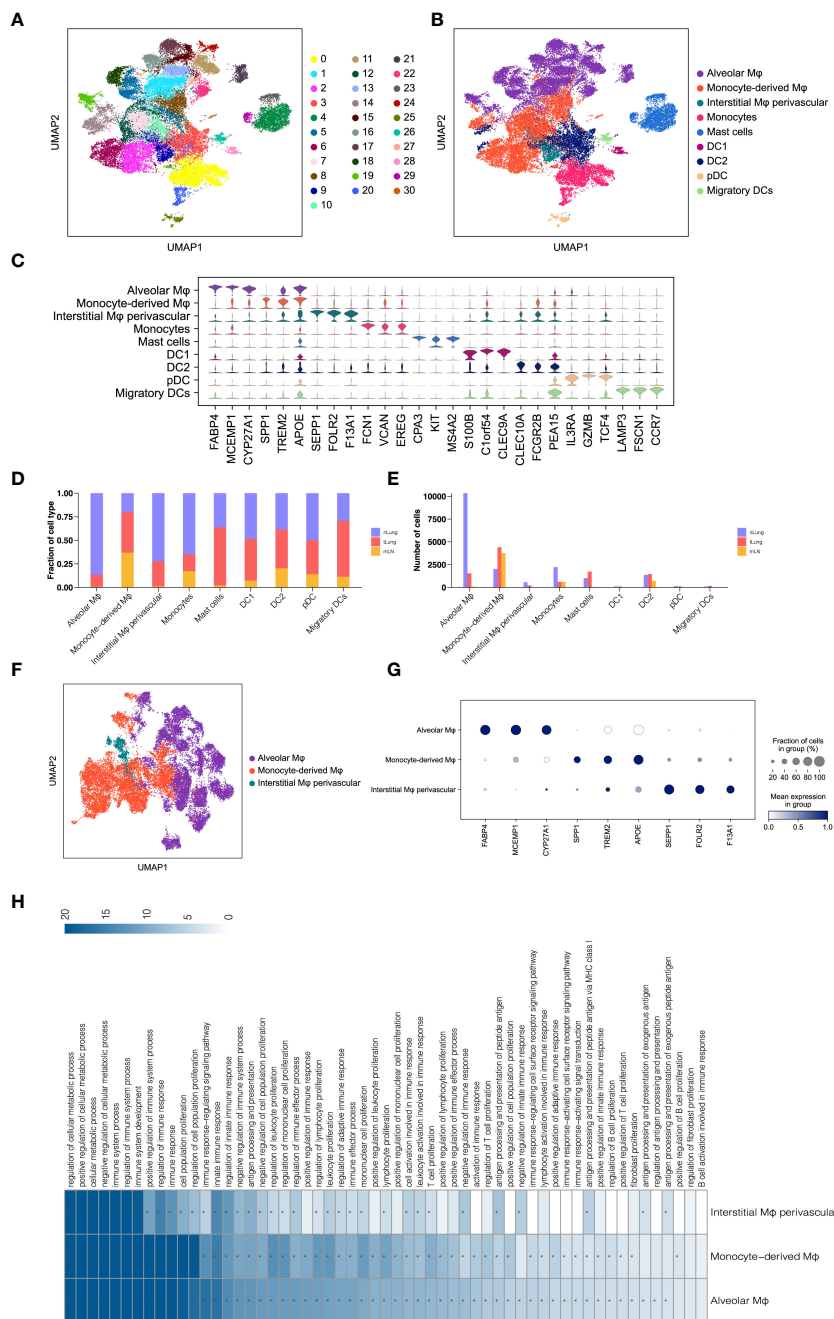


FIGURE 2 Analysis of Macrophages in LUAD. **(A, B)** UMAP of 33,561 myeloid cells, indicating clusters and cell types. **(C)** Violin plots for top threeDEGs in nine myeloid subsets. **(D, E)** Comparison of proportions and cell numbers in nine myeloid subsets across nLung, tLung, and mLN. **(F)** UMAP of 22,959 macrophages, categorized by cell types. **(G)** Dot plot of marker gene expression in three macrophage subsets. **(H)** KEGG pathway analysis results for three macrophage subsets.

pDC, and migratory DCs) (Figure 3A and Supplementary Figure 1A). As expected, each subset displayed specific enrichment of corresponding cell marker genes (Figures 3B, C, Supplementary Figures 1B, C), providing strong evidence for the accuracy of our cell type annotation. Next, we employed CellChat analysis (19) to explore the intercellular communication network among these myeloid cell subsets. Intriguingly, we observed potential intercellular communication in the MHC-II signaling pathway network among monocytes, DC1, DC2, pDC, and

migratory DCs (Figure 3D). Specifically, DC1, DC2, and pDC exhibited close connections within the MHC-II signaling pathway (Figure 3E). This finding suggests possible interactions and shared functionalities among these subsets, supported by their similar gene expression patterns (Supplementary Figure 1D) and distribution within LUAD tissues (Figures 2D, E). These results shed light on the intricate interplay and communication networks existing among myeloid cell subsets in the LUAD tumor microenvironment. Understanding these interactions provides valuable insights into

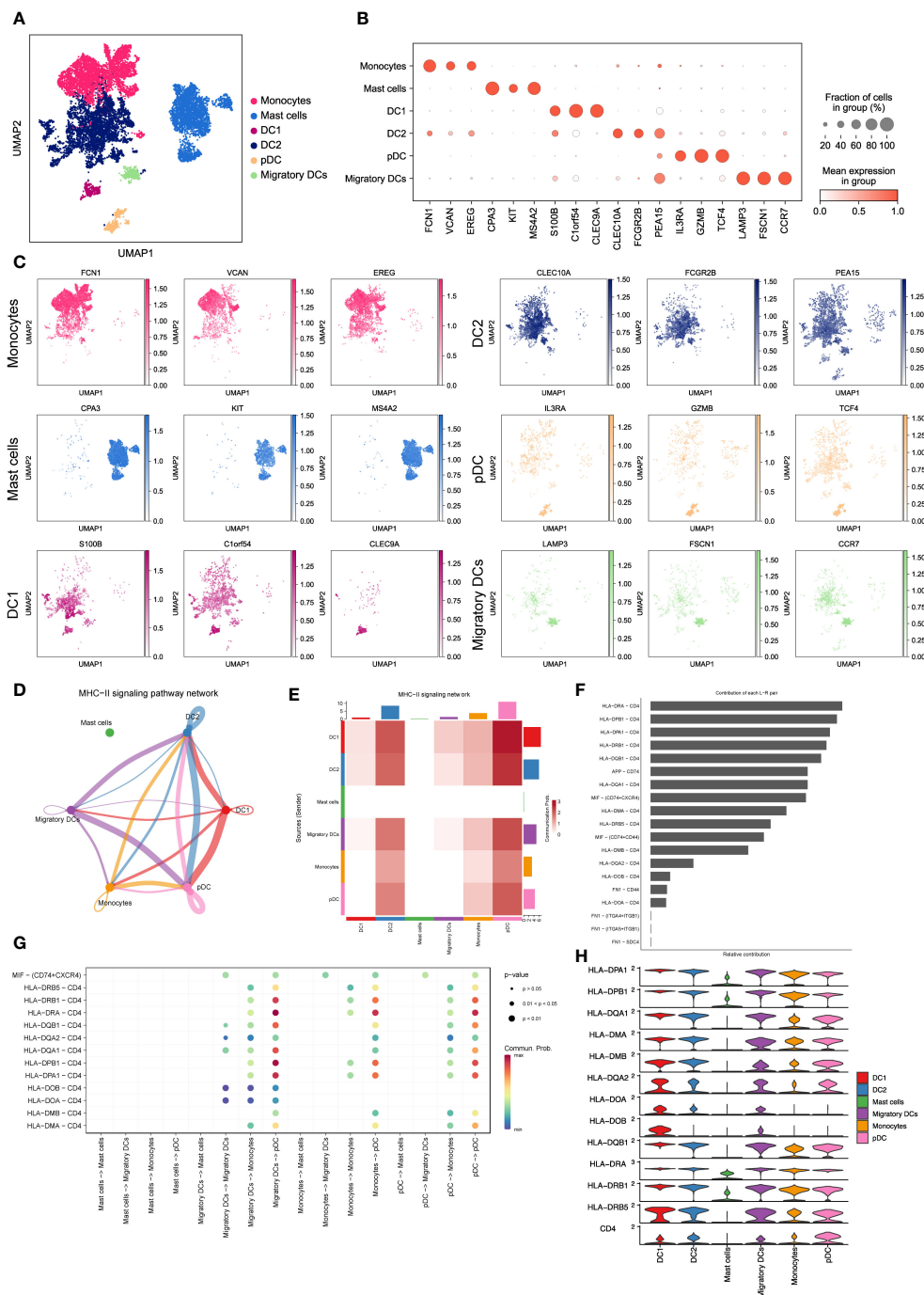


FIGURE 3 Characterizing monocytes, mast cells and dendritic cells in LUAD. **(A)** UMAP plot of 10,602 myeloid cells, including monocytes, mast cells and dendritic cells. **(B)** Dot plot for expression of top3 DEGs in six myeloid cell subsets. **(C)** UMAP plots of six myeloid cell subsets, colored by corresponding marker genes. **(D)** The inferred MHC-II signaling network by Cell-Chat. The edge width represents the communication probability. **(E)** Heatmap of the communication probability among six myeloid cell subsets in MHC-II signaling pathway. **(F)** Relative contributions of 19 ligand-receptor pairs to the overall communication network of MHC-II signaling pathway. **(G)** Significant ligand-receptor pairs that contribute to MHC-II signaling among the mast cell, migratory DC, monocyte and pDC. **(H)** Violin plot for expression of canonical MHC-II signaling genes in six myeloid cell subsets.

the regulatory mechanisms and potential therapeutic targets associated with myeloid cells in LUAD.

To validate the significance of activated MHC-II signaling in LUAD, we examined the contributions of 19 curated ligand-receptor pairs to the MHC-II signaling pathway network. As

depicted in **Figure 3F**, the HLA-DRA-CD4 pair emerged as a prominent contributor to this communication network, underscoring the pivotal role of antigen presentation in activated MHC-II signaling in LUAD. Moreover, our analysis predicted pDCs to be the predominant cell type in the MHC-II signaling

pathway network (Figure 3G). Notably, pDCs exhibited high expression of MHC-II signaling marker genes (Figure 3H), aligning with their crucial function in tumor antigen presentation during tumor immunity (19, 34). Furthermore, we have discovered that these myeloid cells engage in intercellular communication via the macrophage migration inhibitory factor (MIF) signaling pathway (Supplementary Figures 1E, F). This observation suggests a critical involvement of hypoxia in regulating the LUAD microenvironment and influencing metastatic processes (35). These findings collectively demonstrate that dendritic cells, particularly pDCs, heavily rely on antigen processing and presentation in the MHC-II signaling pathway to regulate LUAD progression. This highlights the significance of antigen presentation and the potential involvement of pDCs in shaping the tumor microenvironment and immune responses in LUAD.

3.4 Profiling diverse CD4+ T cell states in LUAD

To investigate the functions of CD4+ T cells in LUAD, we categorized a total of 15,711 CD4+ T cells into five major subsets based on marker gene expression (Figure 4A). Specifically, the CD4_Th17_SLC2A3 subset displayed an enrichment of genes related to T helper 17 (Th17) cells, such as SLC2A3, SFTPC, and C1QA. The CD4_IGKC subset exhibited high expression of immunoglobulin-related genes, including IGKC, IHA1, and IGLC2. The CD4_Treg_CTLA4 subset demonstrated elevated expression of inhibitory receptors like CTLA4, FOXP3, and LAYN, representing CD4+ regulatory T cells (Tregs). The CD4_Naive_CCR7 subset displayed higher expression of genes associated with naïve CD4+ cells, such as CCR7, SELL, and LEAF1. Lastly, the CD4_Pro_MKI67 subset showed an enrichment of proliferation markers, including MKI67, TUBA1B, and IDH2, indicating the presence of proliferating cells (Figure 4B). As expected, these subsets exhibited distinct gene expression patterns (Figure 4G), which reflect their diverse functions and regulatory roles within LUAD.

Subsequently, we conducted pairwise correlation analysis of gene expression levels among the five CD4+ T cell subsets. Interestingly, we observed a relatively higher correlation between the CD4_Pro_MKI67 and CD4_Treg_CTLA4 subsets (Figure 4C), which is consistent with their known immunosuppressive roles in non-small cell lung cancer (36, 37). Furthermore, pseudotime analysis of these CD4+ T cell subsets revealed a clear transition trend from a naïve state to an exhausted state (Figures 4D, E), indicating an ongoing process of immune exhaustion during the progression of LUAD (16). Importantly, we found that ten upregulated genes (Figure 5A) in the previously mentioned CD8_Tex-LAYN subset, particularly MKI67, TK1, MAD2L1, and TYMS, were also highly expressed in the CD4_Pro_MKI67 subset (Figure 4F). This implies a close collaboration between exhausted CD4+ and CD8+ T cells in promoting tumor immunosuppression (38). Taken together, these findings indicate that CD4+ T cells can differentiate into distinct states with specific gene expression patterns, and they work in concert with CD8+ T cells to regulate tumor progression in LUAD.

3.5 Profiling diverse CD8+ T cell states in LUAD

CD8+ T cells are widely recognized as crucial contributors to anti-tumor immunity across different cancer types (13, 39, 40) and have been closely linked to immunotherapy response (14, 41, 42). In order to enhance our understanding of the CD8+ T cell composition in LUAD, we utilized the Louvain algorithm in Scanpy to classify a total of 15,092 CD8+ T cells into 16 sub-clusters (Figure 5A, left panel). Subsequently, based on the expression of marker genes, we further categorized these cells into four major subsets (Figure 5A, right panel). In Figures 5B, C, we displayed the expression of marker genes for each cell type. Importantly, the proportions (Figure 5D) and cell numbers (Figure 5E) of these four major CD8+ T cell subsets exhibited variations across different tissues, aligning with the complex and dynamic composition of T cells in LUAD progression and metastasis. To gain a deeper understanding of the characteristics of distinct CD8+ T cell subsets in LUAD, we employed the partition-based graph abstraction (PAGA) algorithm to construct a cell fate map and infer transitional trajectories of these subsets. As illustrated in Figures 5F, G, each CD8+ T cell subset displayed a unique transitional trajectory to the other subsets. Notably, the CD8_Tem-GZMK subset was relatively closer to the CD8_Tex-LAYN subset compared to the CD8_Tn-LEF1 and CD8_Trm-KLRB1 subsets (Figures 5F, G), aligning with recent studies that highlight the transition from effector memory T cells to exhausted T cells during tumor progression. Interestingly, the CD8_Tn-LEF1 subset, representing the group of naïve T cells, was predicted to be located at the initial position (Figures 5H), suggesting a shared origin among these CD8+ T cell subsets in different states. Furthermore, we utilized the Single-Cell rEgulatory Network Inference and Clustering (SCENIC) (43) approach to analyze the activities of transcription factors (TFs) in the different CD8+ T cell subsets. Consistent with marker gene expression, these subsets also demonstrated distinct patterns of TF activity (Figures 5I, J), emphasizing the critical role of TFs in T cell differentiation. Overall, these findings highlight the differentiation of CD8+ T cells in LUAD patients, showcasing distinct patterns of gene expression and TF activity. This diversity leads to a range of cell states and functions during tumor progression and metastasis, underscoring the complex dynamics of CD8+ T cell subsets in the context of LUAD (44, 45).

3.6 Development of a risk assessment model using CD8_Tex-LAYN signatures in LUAD

Considering the significant enrichment of CD8_Tex-LAYN (exhausted CD8+ T cells) in tLung and mLN (Figures 2D, E), our next objective was to explore the potential of utilizing signatures within exhausted CD8+ T cells to accurately assess the risk of LUAD patients and develop appropriate treatment plans. To identify relevant signatures for risk assessment, we conducted differential expression gene (DEG)

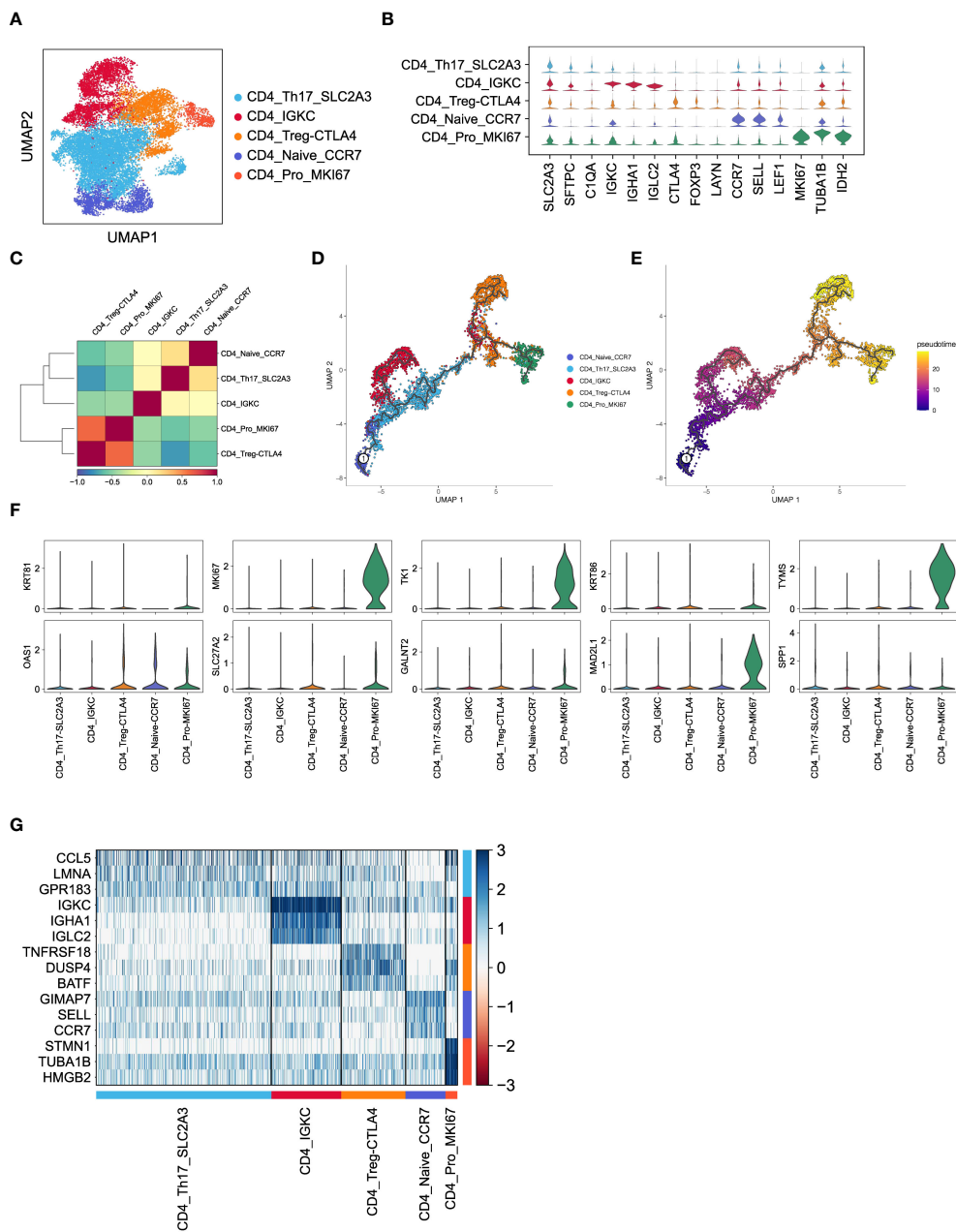
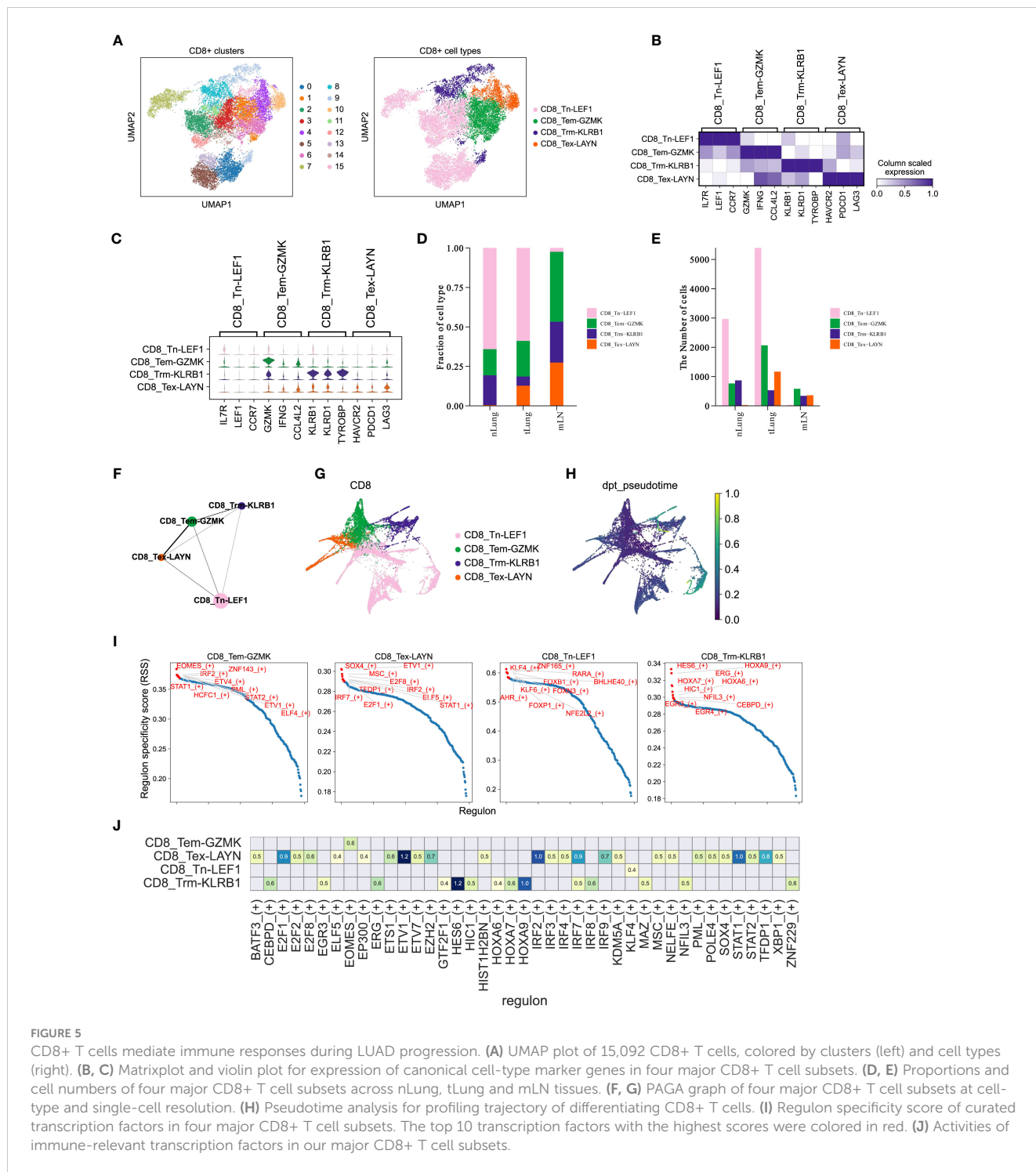


FIGURE 4
 Inferred CD4+ T cell states in LUAD patients. **(A)** UMAP plot of 15,711 CD4+ T cells, colored by cell types. **(B)** Violin plot for expression of canonical cell-type marker genes in five major CD4+ T cell subsets. **(C)** Correlations of gene expression among five major CD4+ T cell subsets. **(D, E)** Pseudotime analysis for profiling trajectory of differentiating CD4+ T cells, colored by cell types and pseudotime. **(F)** Violin plots for expression of ten CD8_Tex-LAYN signatures in five major CD4+ T cell subsets. **(G)** Heatmap of top3 DEGs in five major CD4+ T cell subsets.

analysis and identified 487 genes specifically upregulated in CD8_Tex-LAYN (Figure 6A), which are likely associated with the progression of CD8+ T cell exhaustion. To streamline the selection of signatures, we employed Lasso Cox regression analysis, as previously reported (46), resulting in the identification of 12 candidate genes (Figure 6B, left) with lambda = 0.0155, indicating the lowest partial likelihood deviance (Figure 3B, right). Subsequently, using multivariate Cox regression with a stepwise regression method, we further narrowed down the selection to six final risk signatures: polypeptide N-acetylgalactosaminyltransferase 2

(GALNT2), methylenetetrahydrofolate dehydrogenase, cyclohydrolase, and formyltetrahydrofolate synthetase 1 (MTHFD1), family with sequence similarity 207 member A (FAM207A), keratin 81 (KRT81), ORMDL sphingolipid biosynthesis regulator 3 (ORMDL3), and IKAROS family zinc finger 3 (IKZF3). These six risk signatures were incorporated into a formula for classifying LUAD patients, as follows: RiskScore = 0.454 * GALNT2 + 0.339 * MTHFD1 + 0.288 * FAM207A + 0.071 * KRT81 - 0.276 * ORMDL3 - 0.323 * IKZF3 (Figure 6C).

To assess the accuracy of our risk assessment model, we applied the RiskScore calculation to a cohort of LUAD bulk RNA-seq samples



obtained from TCGA database. After Z-mean normalization, the samples were divided into two distinct groups: high-risk (n = 234) and low-risk (n = 266). Importantly, the high-risk group exhibited significantly poorer survival outcomes (p < 0.0001) (Figure 6D, left). Moreover, the model demonstrated promising predictive performance, with AUC values of 0.67, 0.70, and 0.67 for 1-year, 3-year, and 5-year survival, respectively (Figure 6D, right). To further validate the robustness of our model, we subjected it to testing using five additional GEO datasets (Figures 6E-I). The consistent results across these datasets further confirm the reliability and effectiveness of our

approach. Overall, by identifying signatures in CD8_Tex-LAYN cells, we have successfully established an accurate and reliable model for the risk assessment of LUAD patients, highlighting the strong association between exhausted CD8+ T cells and poor prognosis in LUAD.

3.7 Nomogram development

We assessed the differences in clinical characteristics between patients stratified by different risk levels, finding that those in the

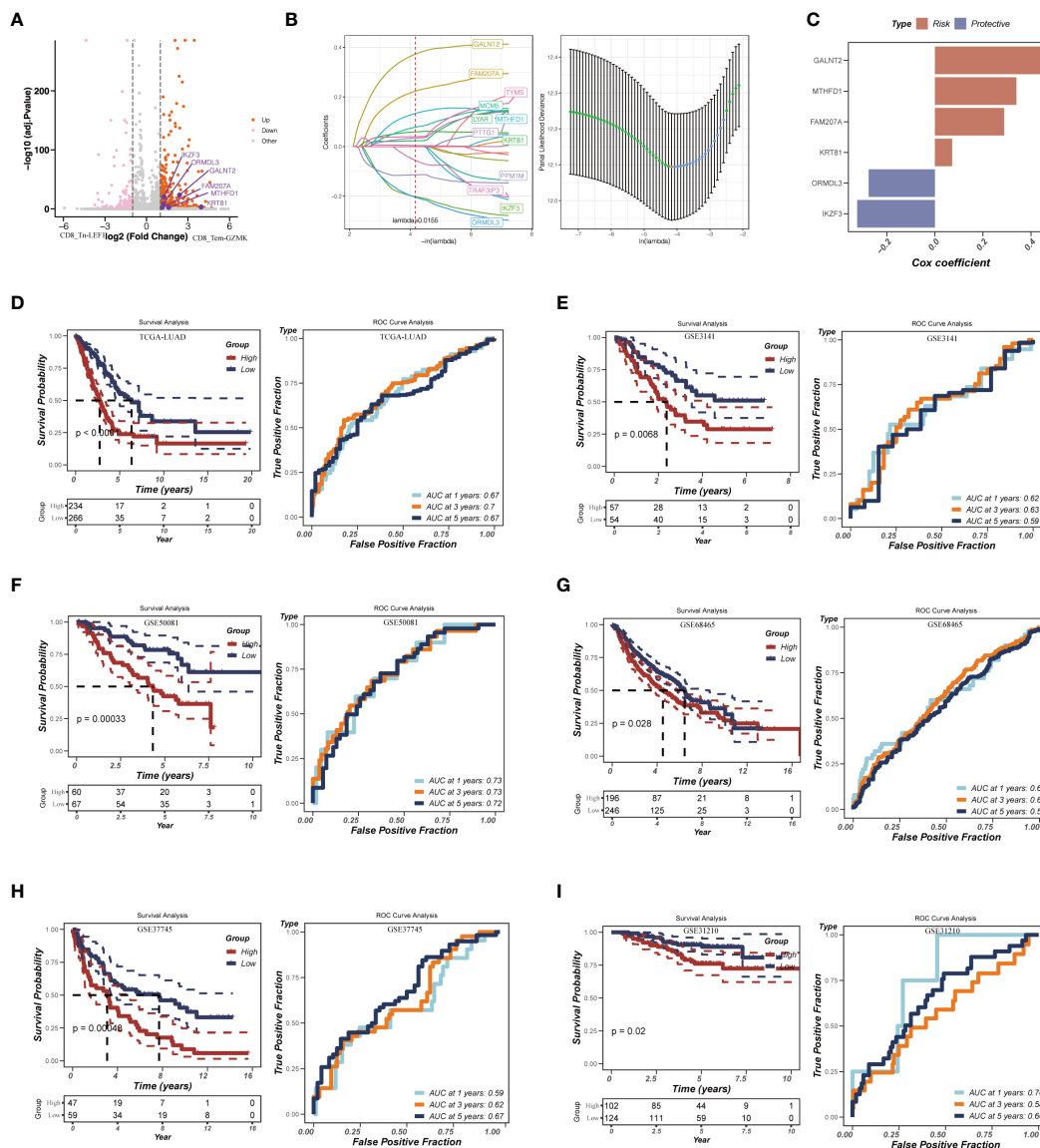


FIGURE 6 Construction of the risk assessment model with CD8_Tex-LAYN signatures. (A) Volcano plot for expression of differentially expressed genes between CD8_Tn-LEF1 and CD8_Tex-LAYN cells. The six risk genes used to construct model were highlighted. (B) Plots of each independent variable (left) and coefficient distributions for parameter selection (right). (C) The multivariate Cox coefficients for six risk genes. (D–I) Kaplan–Meier curves (left in each panel) and ROC curves (right in each panel) for TCGA and GEO LUAD cohorts.

low-risk group were younger, with a higher proportion of females, and typically presented with earlier pathological stages (Figure 7A). Both univariate and multivariate Cox analyses indicated that pathological stage and risk score are independent risk factors for LUAD prognosis (Figures 7B, C). Subsequently, we constructed a nomogram for predicting the prognosis of LUAD (Figure 7D). Calibration and decision curves demonstrated the robustness and accuracy of the nomogram (Figures 7E, F). The ROC curve highlighted the superior predictive performance of the nomogram over other features (Figure 7G).

3.8 Mutation analysis

Through an extensive mutation analysis of the TCGA-LUAD dataset, we detailed the mutational profile of lung adenocarcinoma, identifying the 20 most commonly mutated genes and correlating them with clinical data across both high and low-risk patient groups (Figure 8A). As shown in Figure 8B, in TCGA-LUAD patients, the most common Variant Classification was Missense Mutation, the prevalent Variant Type was SNP, and the most frequent SNV Class was C>A. Subsequent co-mutation analysis did not reveal significant

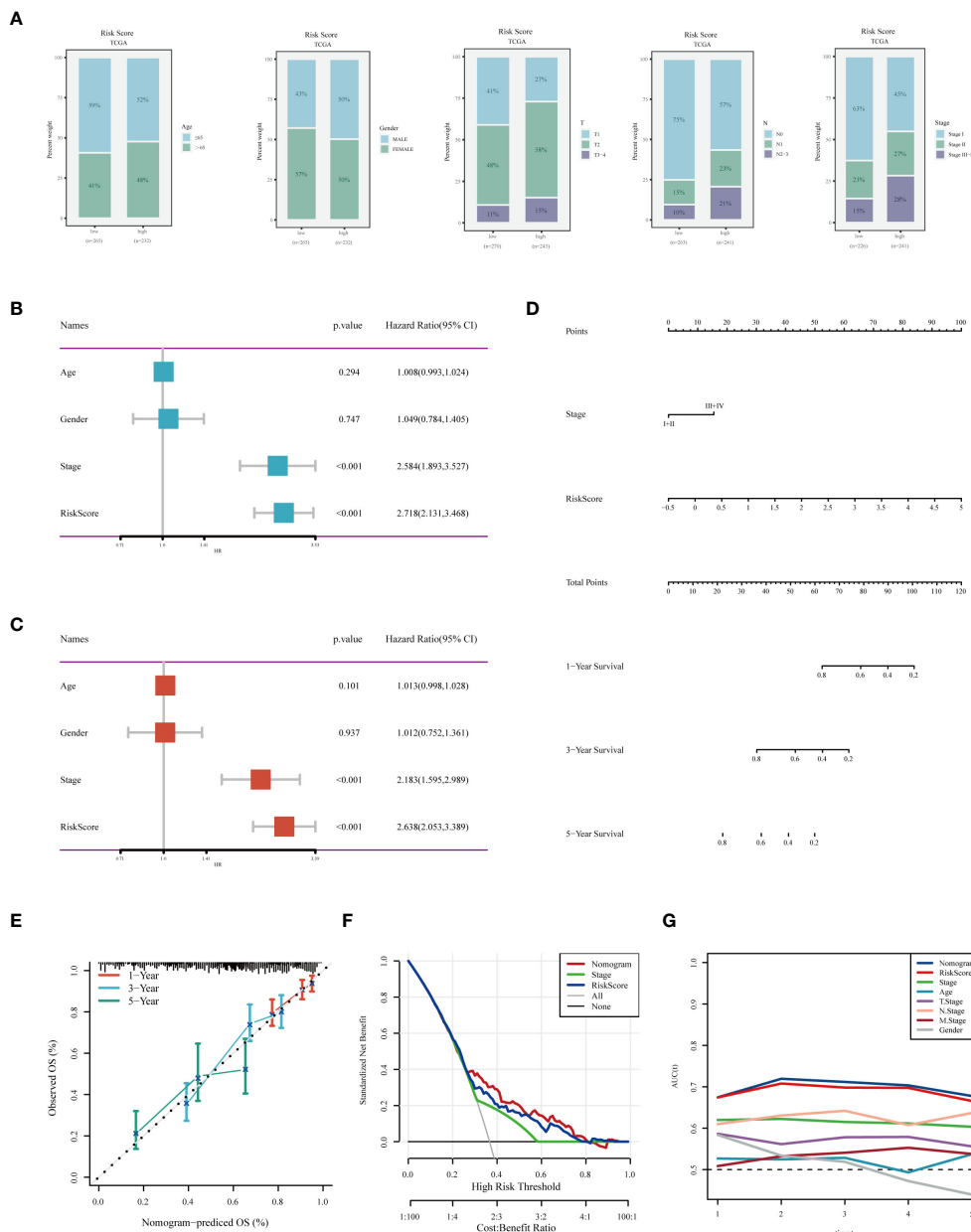


FIGURE 7 Analysis of Clinical Characteristics and Risk Evaluation in the TCGA-LUAD Dataset. **(A)** Analysis of the distribution of different clinical attributes across various risk groups. **(B)** Univariate Cox regression results indicating the relationship of clinical features with survival outcomes. **(C)** Evaluation of the independent prognostic significance of clinical factors using multivariate Cox regression. **(D)** A nomogram was developed by integrating clinical characteristics with the risk score. **(E)** Calibration plot for checking the accuracy of the prognostic model. **(F)** Decision curve analysis to determine the clinical utility of the prognostic model. **(G)** ROC curve analysis for assessing the predictive accuracy of the model over different time points.

co-mutations between model genes and the most commonly mutated genes (Figure 8C). The high-risk group exhibited a higher TMB (Figure 8D), and there was a significant positive correlation between risk score and TMB (Figure 8E). Theoretically, tumors with a higher TMB generate a greater number of neoantigens, which may be recognized as foreign substances by the patient’s immune system, thereby triggering an attack on the tumor. Survival analysis of patients with high and low TMB demonstrated better prognosis in

the high TMB group (Figure 8F). When combining TMB with risk score for survival analysis, patients in the low-TMB and high-risk category showed the poorest prognosis, while those in the high-TMB and low-risk category had a better prognosis (Figure 8G). In summary, these findings highlight the intricate relationship between genetic mutations, TMB, and patient prognosis in LUAD, underscoring the importance of comprehensive genomic profiling in risk assessment and treatment stratification.

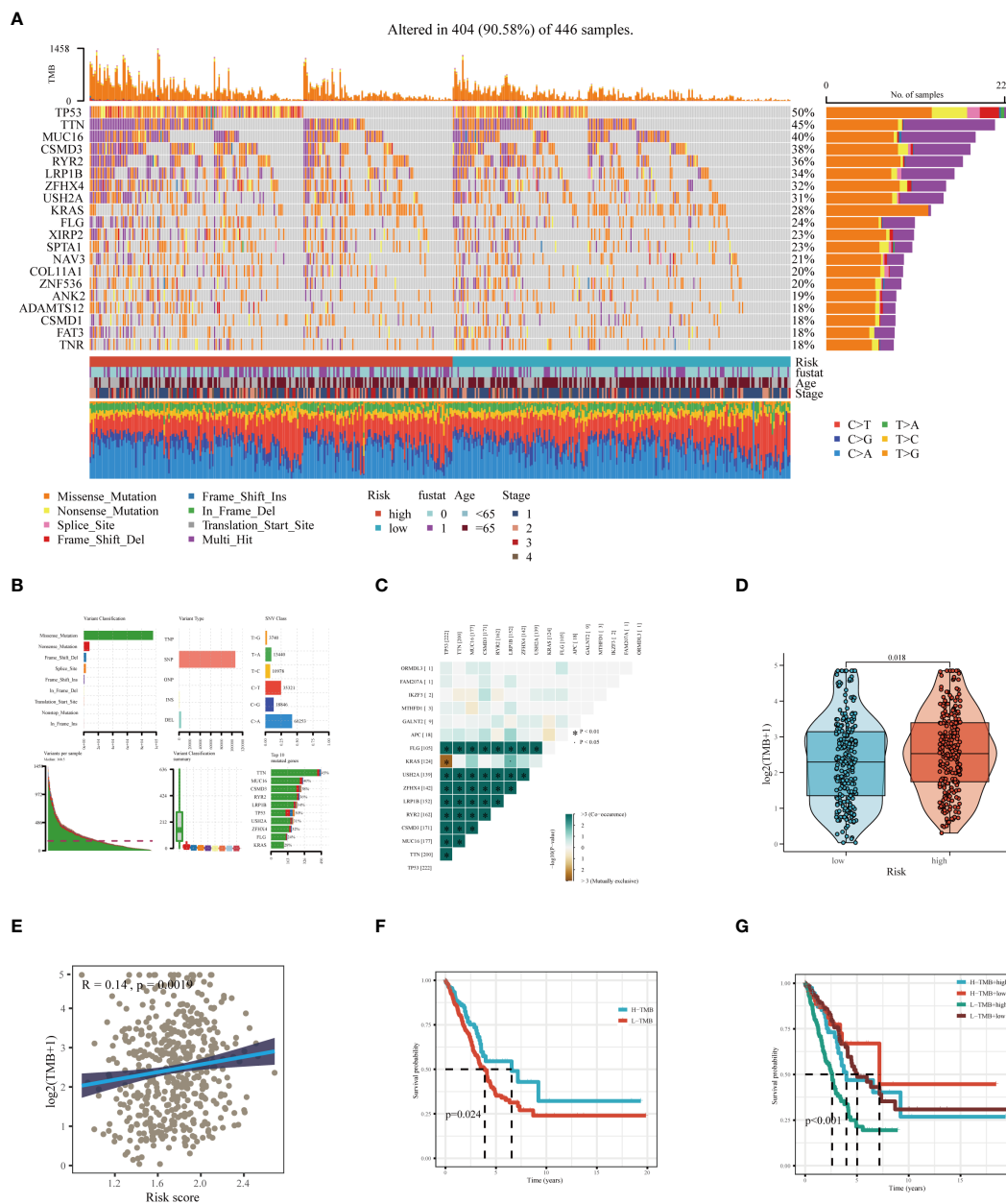


FIGURE 8 Mutation landscape analysis in LUAD. **(A)** Analysis of the mutation spectrum in high and low risk categories, highlighting the top 20 genes with the highest mutation rates in the TCGA-LUAD dataset. **(B)** Examination of mutation details in patients from the TCGA-LUAD study. **(C)** Interaction patterns between model genes and those frequently mutated. **(D)** Comparative assessment of Tumor Mutation Burden (TMB) across varying risk groups. **(E)** Investigation of the relationship between the risk score and TMB. **(F)** Kaplan-Meier curves depicting survival outcomes for groups with high and low TMB. **(G)** Combined multivariate Kaplan-Meier survival analysis using both TMB and risk score.

3.9 Enrichment analysis

Our study comprehensively analyzed the relationship between risk scores, tumor-related pathways, and the tumor immune cycle. Findings revealed that risk scores were significantly positively correlated with most tumor-related pathways, yet exhibited a negative correlation with the majority of the tumor immune cycle stages (Supplementary Figure 2A). Further Gene Set Variation Analysis (GSVA) enrichment analysis indicated that high-risk groups were primarily associated with pathways like

MYC_TARGETS_V2, GLYCOLYSIS, and MYC_TARGETS_V1 (Supplementary Figure 2B). GSEA conducted separately for high and low-risk groups showed significant enrichment of the high-risk group in pathways such as KEGG_CELL_CYCLE and DNA_REPLICATION, whereas the low-risk group was mainly enriched in KEGG_ALLOGRAFT_REJECTION and KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_ICA pathways (Supplementary Figures 2C, D). In summary, these insights reveal a close link between risk scores and tumor biological characteristics, highlighting molecular pathway activity

differences in patients with varying risk levels and providing an important perspective for understanding tumor development mechanisms.

3.10 Assessment of the tumor microenvironment

Utilizing seven algorithms - TIMER, CIBERSORT, CIBERSORT-ABS, QUANTISEQ, MCPOUNTER, XCELL, and EPIC, we thoroughly evaluated the differences in immune cell infiltration levels between high and low-risk groups (Supplementary Figure 3A). Subsequent radar charts highlighted the disparities in immune-related functions and immune cell infiltration (Supplementary Figures 3B, C). Overall, the low-risk group exhibited more active immune-related functions and higher levels of immune infiltration. Additionally, using the 'estimate' R package, we assessed the differences in tumor purity between the high and low-risk groups. The results indicated that the high-risk group had higher tumor purity, and a significant positive correlation was observed between risk scores and tumor purity (Supplementary Figure 3D).

3.11 Immunotherapy response evaluation

In light of the substantial advancements in immunotherapy for LUAD, we aimed to assess the differential responsiveness to immunotherapy between high and low-risk groups through a series of analyses. Initially, patients in the low-risk group exhibited higher expression levels of immune checkpoint-related genes (Figure 9A). There was a tendency for the risk score to be inversely correlated with these genes, interestingly, IKZF3 showed a significant positive correlation with them (Figure 9B). Similarly, the expression levels of MHC-related genes were significantly higher in the low-risk group compared to the high-risk group (Figure 9C), with an inverse correlation observed between risk score and MHC-related genes (Figure 9D). The detailed list of the genes mentioned can be found in Supplementary Table 2. Based on the TIDE database, patients in the high-risk group were found to be more prone to immune escape (Figure 9E). Subsequent analysis of Immunophenoscore (IPS) differences revealed that patients in the low-risk group, especially those positive for CTLA4, are more likely to benefit from immunotherapy (Figure 9F). Overall, our multi-faceted evaluation of immunotherapy sensitivity across different groups demonstrated that the low-risk group not only expresses higher levels of immune checkpoint and MHC-related genes but is also more likely to benefit from immunotherapy, particularly in CTLA4 positive patients.

3.12 Experimental validation *in vitro*

Given the reliance of our analysis on publicly available databases, we undertook a series of fundamental experiments to validate the bioinformatics findings. Initially, we collected tumor

and adjacent normal tissue samples from eight LUAD patients who underwent surgical resection at Tianjin Chest Hospital. RT-PCR was employed to validate the expression of model genes (Figures 10A-F). Our results indicated significant overexpression of GALNT2, MTHFD1, FAM207A, KRT81, and IKZF3 in tumor tissues, consistent with our bioinformatics analysis. No expression difference was noted for ORMDL3 between tumor and adjacent tissues.

With the highest risk coefficient in the risk model (0.454), GALNT2 was further investigated for its role in LUAD using various cellular assays. Expression levels of GALNT2 were first examined across normal pulmonary epithelial (BEAS-2B) and 4 LUAD cell lines (A549, H1299, H1975, H1299) (Figure 11A), revealing a significant upregulation in LUAD cells, particularly in A549 and H1299. Post-transfection RT-PCR confirmed the knockdown efficiency (Figures 11B, C). Transwell assays demonstrated that GALNT2 knockdown notably inhibited the migratory and proliferative abilities of LUAD cells (Figure 11D). Colony formation assays showed a reduction in colony numbers post GALNT2 knockdown (Figure 11E). Finally, CCK8 assays confirmed that knocking down GALNT2 significantly suppressed the proliferation of LUAD cell lines (Figures 11F-G). In summary, our experimental validation reinforces the bioinformatics predictions by confirming the overexpression of several risk genes, including GALNT2, in LUAD tissues compared to adjacent non-tumor tissues. The significant inhibitory effects on migration, proliferation, and colony formation in LUAD cell lines upon GALNT2 knockdown highlight its potential role as a therapeutic target.

4 Discussion

This study aimed to construct a comprehensive single-cell atlas of LUAD through scRNA-seq of 29 samples from 19 treatment-naïve LUAD patients. The atlas detailed cellular compositions and functionalities within the LUAD microenvironment, particularly focusing on epithelial cells, stromal cells (including fibroblasts and endothelial cells), and a variety of immune cells (comprising T cells, B cells, myeloid cells, and NK cells). By identifying seven major cell lineages and delving into the immune cells, we elucidated their collaborative roles in tumor progression and immune response regulation within TME. Furthermore, leveraging marker genes associated with exhausted CD8+ T cells, we established a robust prognostic model for stratifying LUAD patient risk. This model facilitated a systematic assessment of differences in prognosis, TME, mutation landscape, and immune therapy responses across varied risk groups. Our findings not only deepen the understanding of tumor immunology in LUAD but also provide new directions for future therapeutic strategies and improvement of patient prognosis.

In this study, we elucidated the integral roles of DC1, DC2, and pDC within the MHC-II signaling pathway in the context of the LUAD microenvironment. It was observed that MHC-II, predominantly expressed on antigen-presenting cells, is also present in cancer cells, potentially correlating with improved immunotherapy outcomes (47). Evidence from recent studies

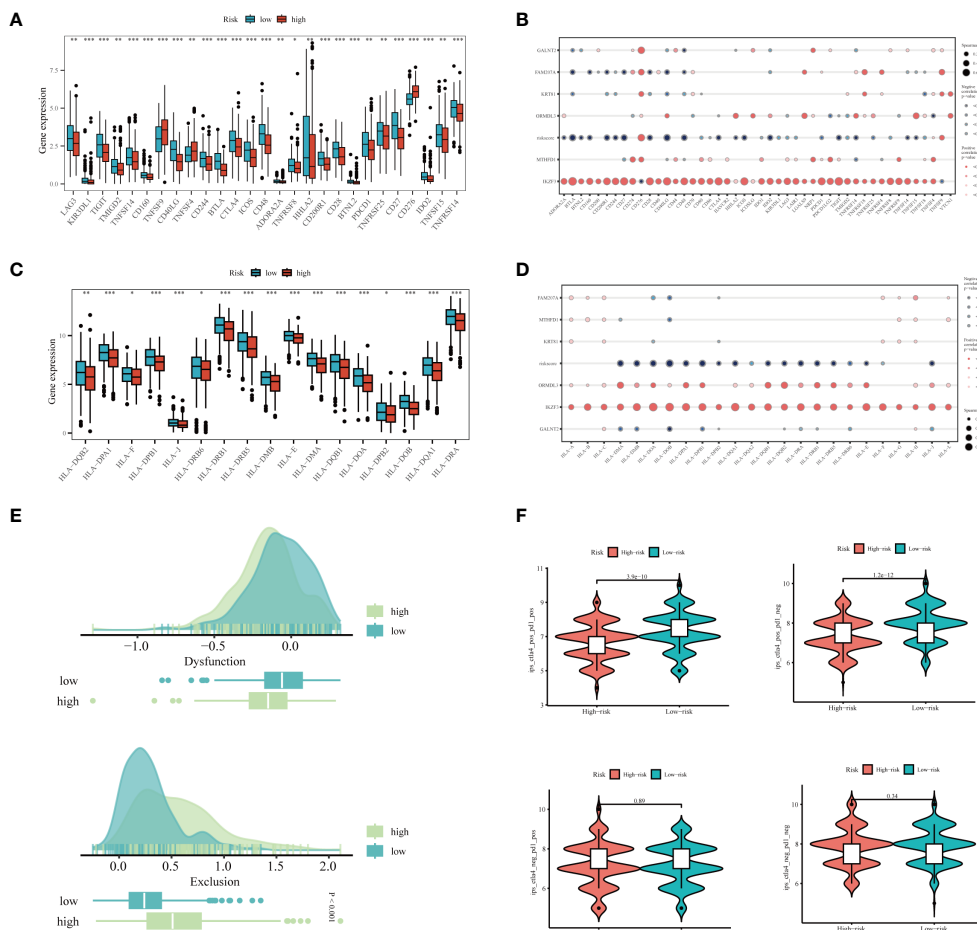


FIGURE 9 Prediction of Immunotherapy Efficacy. **(A)** Differences in immune checkpoint gene expression between high and low-risk groups. **(B)** Correlations between immune checkpoint gene expression, risk scores, and Hub genes. **(C)** Expression differences of major histocompatibility complex (MHC) genes between risk groups. **(D)** Correlations of MHC-related genes and immune checkpoint genes with risk scores and Hub genes. **(E)** Differences in TIDE scores across the high and low-risk categories. **(F)** Comparative Immune Predictive Scores (IPS) between high and low-risk groups. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

suggests a positive correlation between MHC-II expression in tumor cells and the effectiveness of immunotherapeutic interventions, underscoring its pivotal role in tumor immunology (48). Tumor cells expressing MHC-II potentially secrete immunostimulatory exosomes, engage in direct interactions with CD4+ T cells to influence their polarization and activation, or secrete antigens that are endocytosed and presented by professional antigen-presenting cells (pAPCs). These mechanisms collectively contribute to modulating the immune landscape within the tumor microenvironment. Adjusting MHC-II expression in melanoma may enhance responsiveness to immunotherapies (49). Furthermore, our investigation revealed dynamic interactions within the Macrophage Migration Inhibitory Factor (MIF) signaling pathway among myeloid cell subtypes, with a notable emphasis on the critical function of pDCs. The MIF pathway is essential in orchestrating the immune responses within the tumor milieu, especially in modulating cellular interactions and immune cell activity. Overexpression of MIF in various cancer types is associated with increased tumor aggressiveness and adverse prognostic outcomes (50). This study also highlights the role of

MIF in cell survival via Akt pathway activation and the involvement of CSN5/JAB1 in the regulation of autocrine MIF activity (51).

Our scRNA-seq study revealed a dynamic shift in CD8+ T cells from a naïve to an exhausted phenotype as LUAD progresses, mirroring recent discoveries in other cancers such as breast and colorectal (40, 52). Based on the marker genes of exhausted CD8+ T cells, we constructed a risk model composed of 6 genes, namely *GALNT2*, *MTHFD1*, *FAM207A*, *KRT81*, *ORMDL3*, *IKZF3*. The function of *GALNT2* and a series of basic experimental validations will be discussed in the following sections. Next, we discuss the potential roles of the remaining five model genes in the progression of LUAD. *MTHFD1* encodes a protein that plays a critical role in the folate metabolic pathway, which is essential for DNA synthesis, repair, and methylation (53). Alterations in folate metabolism can contribute to carcinogenesis by affecting DNA methylation patterns and thus gene expression (54). In the context of LUAD, *MTHFD1* could influence tumor progression through effects on DNA synthesis and methylation, potentially affecting cell proliferation and survival (55). Given its classification as a risk signature gene, *FAM207A* may be involved

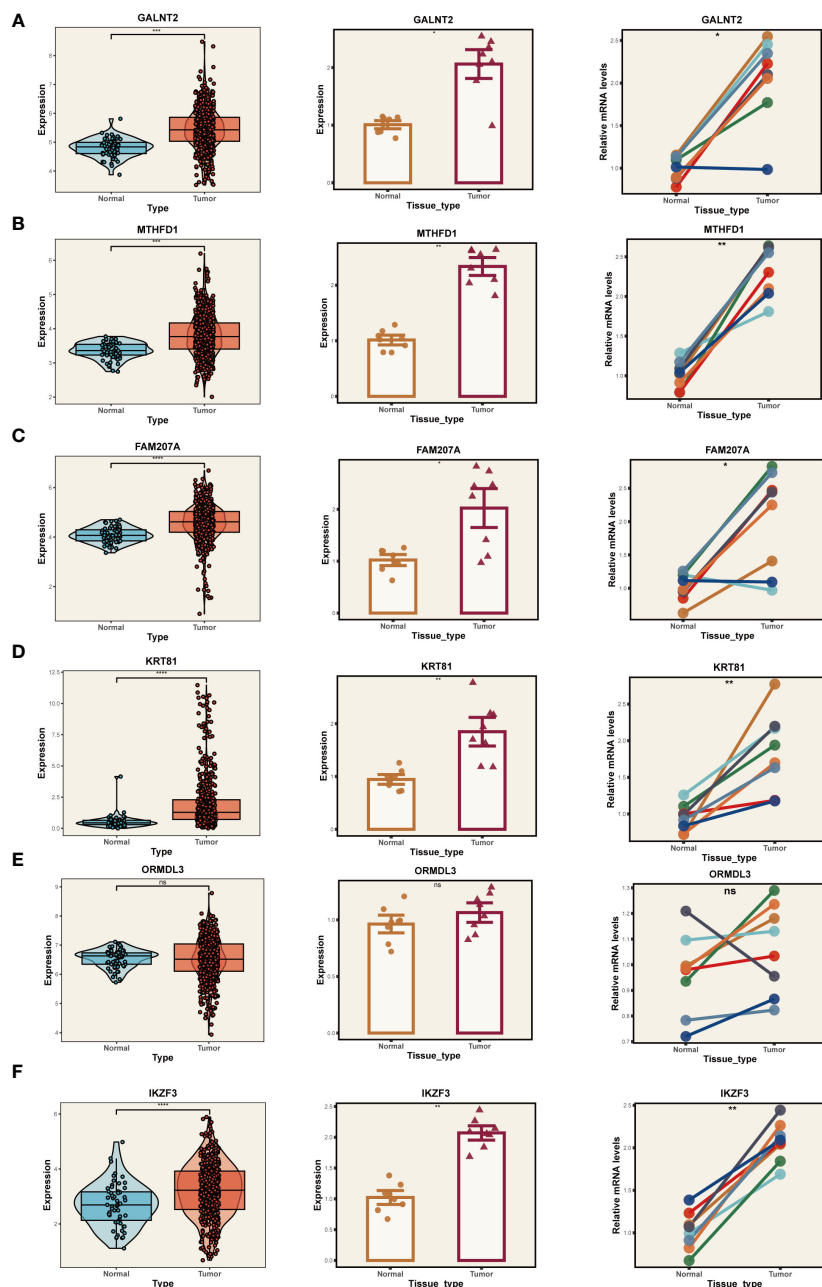


FIGURE 10
 PCR Validation of Gene Expression. Expression of genes GALNT2 (A), MTHFD1 (B), FAM207A (C), KRT81 (D), ORMDL3 (E), and IKZF3 (F) in tumor and normal tissue samples. Left panels: Distribution of expression in the TCGA database. Middle panels: Expression in tumor vs. normal tissue from Tianjin Chest Hospital. Right panels: Paired comparison between individual tumor and adjacent normal tissues. *P < 0.05, **P < 0.01, ***P < 0.001.

in processes that promote tumor growth or metastasis, possibly through affecting cell adhesion, migration, or communication within the TME (56). *KRT81*, as a member of the keratin family, is involved in the structural integrity of epithelial cells (57). Keratins are often implicated in cancer through their roles in epithelial cell stability, migration, and invasion (58). *KRT81* could contribute to LUAD pathogenesis by affecting tumor cell mechanical properties, facilitating invasion and metastasis (58). *ORMDL3* is involved in sphingolipid metabolism, which has been linked to various cellular processes important in cancer, including cell growth, apoptosis, and

response to therapy (59). *ORMDL3*'s role in modulating sphingolipid metabolism could influence the survival and proliferation of LUAD cells, as well as their sensitivity to chemotherapeutic agents (60). *IKZF3*, as a member of the IKAROS family of zinc-finger transcription factors, plays a role in lymphocyte differentiation and function (61). Its involvement in LUAD could be related to immune evasion mechanisms, where altered expression of *IKZF3* affects the immune microenvironment's ability to recognize and eliminate tumor cells. Leveraging these signatures, a Cox regression model was

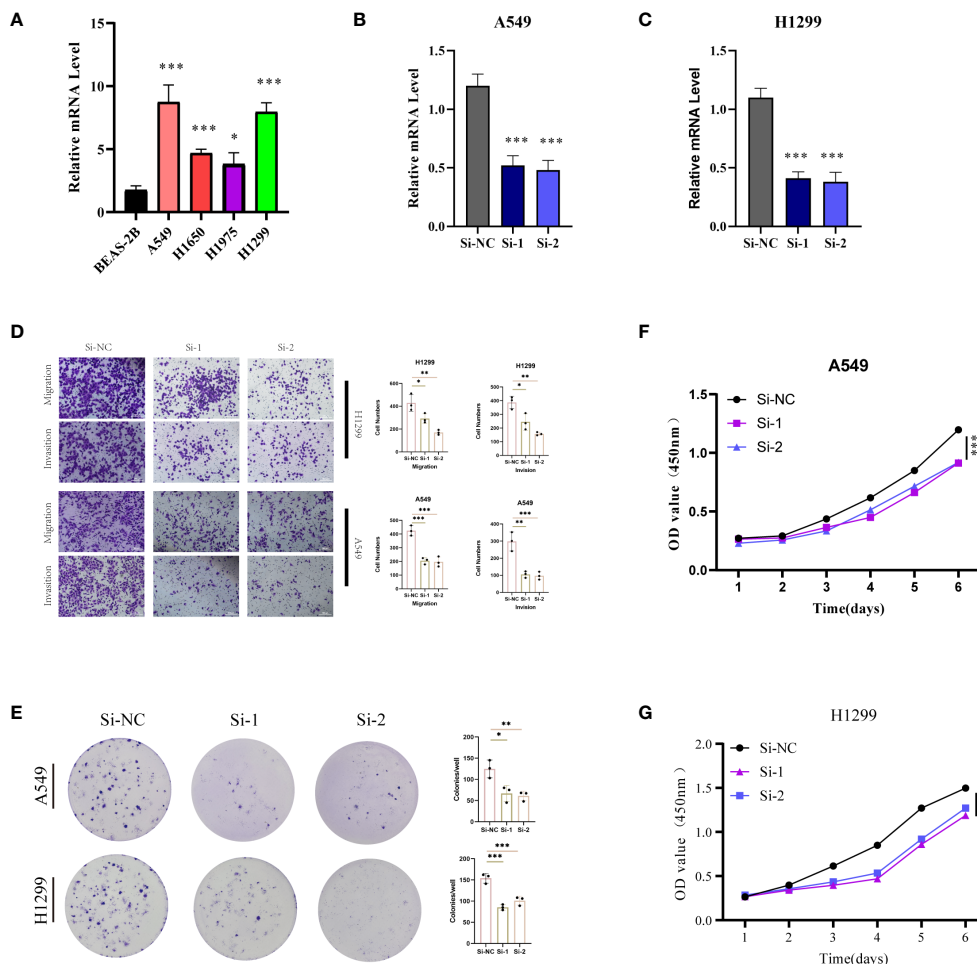


FIGURE 11
 Experimental Validation *In Vitro*. (A) RT-PCR validation of GALNT2 expression across five cell lines. Post-transfection (48 hours), RT-PCR was utilized to confirm GALNT2 RNA expression levels in (B) A549 and (C) H1299 cell lines. (D) Transwell assays were conducted to evaluate the impact of GALNT2 knockdown on the migratory and invasive capabilities of A549 and H1299 cells. (E) Colony formation assays were performed to assess the proliferation potential of A549 and H1299 cells post GALNT2 knockdown. CCK-8 assays detected the effects of GALNT2 knockdown on cell proliferation abilities in (F) A549 and (G) H1299 cell lines. *P < 0.05, **P < 0.01, ***P < 0.001.

formulated, demonstrating substantial predictive accuracy and consistency, corroborated by data from the TCGA and six additional GEO cohorts. Our findings emphasize that patients categorized in the high-risk group, based on these gene signatures, showed significantly worse outcomes, underscoring the critical role of these genes in the advancement of LUAD. In advancing our research, we have integrated risk scores with clinical features to construct a novel nomogram. The diagnostic efficacy of this nomogram was superior compared to other clinical characteristics, as evidenced by its ROC curve analysis. Mutation analysis revealed a higher mutation frequency in the high-risk group. Furthermore, patients with high TMB exhibited markedly better survival outcomes compared to those with low TMB. Kaplan-Meier analysis for various factors indicated that patients in the high-TMB and low-risk category had a more favorable prognosis. This comprehensive approach, combining genetic risk factors with clinical parameters, offers a more refined and predictive model for patient outcomes in LUAD, potentially guiding more personalized therapeutic strategies.

Through enrichment analysis, we established a link between risk scores and tumor-related pathways in LUAD. This revealed a positive correlation with tumor-promoting pathways and a negative one with tumor immune cycle stages, highlighting the complex interplay between genetic risk and tumor environment. Distinct molecular pathways were associated with high-risk (cell cycle and DNA replication) and low-risk (immune-related) groups, suggesting different tumor progression mechanisms. Additionally, higher tumor homogeneity in high-risk groups, potentially influencing invasiveness, was noted, with low-risk groups showing greater potential for immunotherapy response, particularly in CTLA4 positive cases. These findings are crucial for understanding LUAD and developing personalized, immunotherapy-focused treatment strategies.

Our study then shifted its focus to GALNT2, which emerged as the gene with the highest risk score. Historically identified as a member of the glycosyltransferase family, GALNT2 was found to modulate adipogenesis and insulin signaling in adipocytes, impacting metabolic processes associated with obesity and

diabetes (62). Its importance also extends to tumor development and progression, emphasizing its role in both metabolic and oncological pathways (63). Elevated GALNT2 levels are linked to unfavorable outcomes in patients with glioblastoma multiforme (GBM). In a functional context, inhibiting GALNT2 disrupts the growth, self-renewal, and aggressive behavior of glioma stem-like cells, primarily by downregulating CD44 expression (64). GALNT2 has been shown to contribute to the enhanced aggressiveness of colorectal cancer cells, doing so in part by modulating the AXL pathway (65). In this research, RT-PCR analysis revealed a marked upregulation of GALNT2 in LUAD. Subsequent cellular assays linked GALNT2 with enhanced proliferation and migration in LUAD cells, indicating its potential as a viable target for LUAD treatment strategies.

Our focus was primarily on primary lung tumors and lymph node metastasis in LUAD, without delving into the diverse clinical stages of the disease. More research is needed to explore LUAD's progression at various stages. Although we've pinpointed critical genes and pathways related to LUAD, further external validation with functional experiments is required for confirmation. Despite these limitations, our study significantly aids in understanding LUAD's microenvironment and provides a solid risk assessment model. Identifying risk genes linked to LUAD could lead to novel treatments. Targeting these genes and pathways may enhance treatment outcomes for LUAD patients, and we anticipate that our findings will be valuable for personalized treatment decisions, thereby improving LUAD's clinical management. For hereditary tumors or high-risk populations, based on our predictive model, we can assess the risk of developing LUAD, thereby enabling early screening and monitoring of high-risk groups. Additionally, we found that the low-risk group is more likely to benefit from immunotherapy, which can help determine the best combination and sequence of treatment drugs to maximize efficacy and prolong survival. At the same time, by avoiding ineffective treatment plans, personalized treatment plans can help reduce the waste of medical resources, thereby reducing treatment costs and improving the efficiency and quality of medical services.

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 Ethics Committee of Tianjin Chest Hospital. 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

HZ: Writing – original draft. PZ: Writing – original draft. XL: Writing – original draft. LT: Writing – original draft. YW: Writing – original draft. XJ: Writing – original draft. KW: Writing – review & editing. XL: Writing – review & editing. DS: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by Tianjin Key Medical Discipline (Specialty) Construction Project (TJYXZDXK-018A).

Acknowledgments

The authors of this article would like to thank the online databases such as TCGA and GEO for providing the data. We also appreciate the sincere and constructive comments from the editors and peer reviewers. Our article's preprint has been published on Research Square, titled: "Characterizing Tumor Microenvironment and Risk Genes in Lung Adenocarcinoma through scRNA-seq and Bulk RNA-seq Data" (<https://doi.org/10.21203/rs.3.rs-3135938/v1>).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2024.1366096/full#supplementary-material>

References

- Cao M, Li H, Sun D, Chen W. Cancer burden of major cancers in China: A need for sustainable actions. *Cancer Commun (Lond)*. (2020) 40:205–10. doi: 10.1002/cac2.12025
- Ferlay J, Colombet M, Soerjomataram I, Dyba T, Randi G, Bettio M, et al. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. *Eur J Cancer*. (2018) 103:356–87. doi: 10.1016/j.ejca.2018.07.005
- Fehrenbacher L, Spira A, Ballinger M, Kowanzet M, Vansteenkiste J, Mazieres J, et al. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. *Lancet*. (2016) 387:1837–46. doi: 10.1016/S0140-6736(16)00587-0
- Garon EB, Rizvi NA, Hui R, Leigh N, Balmanoukian AS, Eder JP, et al. Pembrolizumab for the treatment of non-small-cell lung cancer. *N Engl J Med*. (2015) 372:2018–28. doi: 10.1056/NEJMoa1501824
- Blumenschein GR Jr., Smit EF, Planchard D, Kim DW, Cadranel J, Pas De T, et al. A randomized phase II study of the MEK1/MEK2 inhibitor trametinib (GSK1120212) compared with docetaxel in KRAS-mutant advanced non-small-cell lung cancer (NSCLC)dagger. *Ann Oncol*. (2015) 26:894–901. doi: 10.1093/annonc/mdv072
- Janne PA, Shaw AT, Pereira JR, Jeannin G, Vansteenkiste J, Barrios C, et al. Selumetinib plus docetaxel for KRAS-mutant advanced non-small-cell lung cancer: a randomised, multicentre, placebo-controlled, phase 2 study. *Lancet Oncol*. (2013) 14:38–47. doi: 10.1016/S1470-2045(12)70489-8
- Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Curran WJ Jr, Wu YL, et al. Lung cancer: current therapies and new targeted treatments. *Lancet*. (2017) 389:299–311. doi: 10.1016/S0140-6736(16)30958-8
- Zhou Y, Cheng L, Liu L, Li X. NK cells are never alone: crosstalk and communication in tumour microenvironments. *Mol Cancer*. (2023) 22:34. doi: 10.1186/s12943-023-01737-7
- Jin MZ, Jin WL. The updated landscape of tumor microenvironment and drug repurposing. *Signal Transduct Target Ther*. (2020) 5:166. doi: 10.1038/s41392-020-00280-x
- Wculek SK, Malanchi I. Neutrophils support lung colonization of metastasis-initiating breast cancer cells. *Nature*. (2015) 528:413–7. doi: 10.1038/nature16140
- Hanna RN, Cekic C, Sag D, Tacke R, Thomas GD, Nowyhed H, et al. Patrolling monocytes control tumor metastasis to the lung. *Science*. (2015) 350:985–90. doi: 10.1126/science.aac9407
- Roider T, Seufert J, Uvarovskii A, Frauhammer F, Bordsa M, Abedpour N, et al. Dissecting intratumour heterogeneity of nodal B-cell lymphomas at the transcriptional, genetic and drug-response levels. *Nat Cell Biol*. (2020) 22:896–906. doi: 10.1038/s41556-020-0532-x
- Zheng C, Zheng L, Yoo JK, Guo H, Zhang Y, Guo X, et al. Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. *Cell*. (2017) 169:1342–56.e16. doi: 10.1016/j.cell.2017.05.035
- Sade-Feldman M, Yizhak K, Bjorgaard SL, Ray JP, Boer CG, Jenkins RW, et al. Defining T cell states associated with response to checkpoint immunotherapy in melanoma. *Cell*. (2018) 175:998–1013.e20. doi: 10.1016/j.cell.2018.10.038
- Tirosh I, Izar B, Prakadan SM, Wadsworth MH 2nd, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*. (2016) 352:189–96. doi: 10.1126/science.aad0501
- Kim N, Kim HK, Lee K, Hong Y, Cho JH, Choi JW, et al. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat Commun*. (2020) 11:2285. doi: 10.1038/s41467-020-16164-1
- Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol*. (2018) 19:15. doi: 10.1186/s13059-017-1382-0
- Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods*. (2017) 14:979–82. doi: 10.1038/nmeth.4402
- Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun*. (2021) 12:1088. doi: 10.1038/s41467-021-21246-9
- Chi H, Zhao S, Yang J, Gao X, Peng G, Zhang J, et al. T-cell exhaustion signatures characterize the immune landscape and predict HCC prognosis via integrating single-cell RNA-seq and bulk RNA-sequencing. *Front Immunol*. (2023) 14:1137025. doi: 10.3389/fimmu.2023.1137025
- Huang R, Chen Z, Li W, Fan C, Liu J. Immune system-associated genes increase Malignant progression and can be used to predict clinical outcome in patients with hepatocellular carcinoma. *Int J Oncol*. (2020) 56:1199–211. doi: 10.3892/ijo.2020.4998
- Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol*. (2008) 26:1364–70. doi: 10.1200/JCO.2007.12.9791
- Manoranjan B, Venugopal C, Bakhshinyan D, Adile AA, Richards L, Kameda-Smith MM, et al. Wnt activation as a therapeutic strategy in medulloblastoma. *Nat Commun*. (2020) 11:4323. doi: 10.1038/s41467-020-17953-4
- Grover A, Sanjuan-Pla A, Thongjuea S, Carrelha J, Giustacchini A, Gambardella A, et al. Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells. *Nat Commun*. (2016) 7:11075. doi: 10.1038/ncomms11075
- Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. (2000) 28:27–30. doi: 10.1093/nar/28.1.27
- Zhuang W, Sun H, Zhang S, Zhou Y, Weng W, Wu B, et al. An immunogenomic signature for molecular classification in hepatocellular carcinoma. *Mol Ther Nucleic Acids*. (2021) 25:105–15. doi: 10.1016/j.omtn.2021.06.024
- Zhang S, Iyer S, Ran H, Dolgalev I, Gu S, Wei W, et al. Genetically defined, syngeneic organoid platform for developing combination therapies for ovarian cancer. *Cancer Discovery*. (2021) 11:362–83. doi: 10.1158/2159-8290.CD-20-0455
- Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. (2013) 4:2612. doi: 10.1038/ncomms3612
- Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med*. (2018) 24:1550–58. doi: 10.1038/s41591-018-0136-1
- Engblom C, Pfirschke C, Pittet MJ. The role of myeloid cells in cancer therapies. *Nat Rev Cancer*. (2016) 16:447–62. doi: 10.1038/nrc.2016.54
- Schmid MC, Varner JA. Myeloid cells in the tumor microenvironment: modulation of tumor angiogenesis and tumor inflammation. *J Oncol*. (2010) 2010:201026. doi: 10.1155/2010/201026
- DeNardo DG, Ruffell B. Macrophages as regulators of tumour immunity and immunotherapy. *Nat Rev Immunol*. (2019) 19:369–82. doi: 10.1038/s41577-019-0127-6
- Cheng S, Li Z, Gao R, Xing B, Gao Y, Yang Y, et al. A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell*. (2021) 184:792–809.e23. doi: 10.1016/j.cell.2021.01.010
- Jahrsdorfer B, Vollmer A, Blackwell SE, Maier J, Sontheimer K, Beyer T, et al. Granzyme B produced by human plasmacytoid dendritic cells suppresses T-cell expansion. *Blood*. (2010) 115:1156–65. doi: 10.1182/blood-2009-07-235382
- Demoulin S, Herfs M, Delvenne P, Hubert P. Tumor microenvironment converts plasmacytoid dendritic cells into immunosuppressive/tolerogenic cells: insight into the molecular mechanisms. *J Leukoc Biol*. (2013) 93:343–52. doi: 10.1189/jlb.0812397
- Mitchell KG, Parra ER, Nelson DB, Wistuba Zhang J II, Fujimoto J, et al. Tumor cellular proliferation is associated with enhanced immune checkpoint expression in stage I non-small cell lung cancer. *J Thorac Cardiovasc Surg*. (2019) 158:911–19.e6. doi: 10.1016/j.jtcvs.2019.04.084
- Zhang Q, He Y, Luo N, Patel SJ, Han Y, Gao R, et al. Landscape and dynamics of single immune cells in hepatocellular carcinoma. *Cell*. (2019) 179:829–45.e20. doi: 10.1016/j.cell.2019.10.003
- Zheng B, Wang D, Qiu X, Luo G, Wu T, Yang S, et al. Trajectory and functional analysis of PD-1(high) CD4(+)/CD8(+) T cells in hepatocellular carcinoma by single-cell cytometry and transcriptome sequencing. *Adv Sci (Weinh)*. (2020) 7:2000224. doi: 10.1002/advs.202000224
- Guo X, Zhang Y, Zheng L, Zheng C, Song J, Zhang Q, et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat Med*. (2018) 24:978–85. doi: 10.1038/s41591-018-0045-3
- Zhang L, Yu X, Zheng L, Zhang Y, Li Y, Fang Q, et al. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature*. (2018) 564:268–72. doi: 10.1038/s41586-018-0694-x
- Fairfax BP, Taylor CA, Watson RA, Nassiri I, Danielli S, Fang H, et al. Peripheral CD8(+) T cell characteristics associated with durable responses to immune checkpoint blockade in patients with metastatic melanoma. *Nat Med*. (2020) 26:193–99. doi: 10.1038/s41591-019-0734-6
- Wu TD, Madireddi S, Almeida PE, Bancheau R, Chen YJ, Chitre AS, et al. Peripheral T cell expansion predicts tumour infiltration and clinical response. *Nature*. (2020) 579:274–78. doi: 10.1038/s41586-020-2056-8
- Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res*. (2021) 49:D545–D51. doi: 10.1093/nar/gkaa970
- Wang Y, Li X, Peng S, Hu H, Wang Y, Shao M, et al. Single-cell analysis reveals spatial heterogeneity of immune cells in lung adenocarcinoma. *Front Cell Dev Biol*. (2021) 9:638374. doi: 10.3389/fcell.2021.638374
- He D, Wang D, Lu P, Yang N, Xue Z, Zhu X, et al. Single-cell RNA sequencing reveals heterogeneous tumor and immune cell populations in early-stage lung adenocarcinomas harboring EGFR mutations. *Oncogene*. (2021) 40:355–68. doi: 10.1038/s41388-020-01528-0
- Yu L, Shen N, Shi Y, Shi X, Fu X, Li S, et al. Characterization of cancer-related fibroblasts (CAF) in hepatocellular carcinoma and construction of CAF-based risk signature based on single-cell RNA-seq and bulk RNA-seq data. *Front Immunol*. (2022) 13:1009789. doi: 10.3389/fimmu.2022.1009789
- Rodig SJ, Gusenleitner D, Jackson DG, Gjini E, Giobbie-Hurder A, Jin C, et al. MHC proteins confer differential sensitivity to CTLA-4 and PD-1 blockade in

- untreated metastatic melanoma. *Sci Transl Med.* (2018) 10. doi: 10.1126/scitranslmed.aar3342
48. Zeng Z, Gu SS, Ouardaoui N, Tymms C, Yang L, Wong CJ, et al. Hippo signaling pathway regulates cancer cell-intrinsic MHC-II expression. *Cancer Immunol Res.* (2022) 10:1559–69. doi: 10.1158/2326-6066.CIR-22-0227
49. Axelrod ML, Cook RS, Johnson DB, Balko JM. Biological consequences of MHC-II expression by tumor cells in cancer. *Clin Cancer Res.* (2019) 25:2392–402. doi: 10.1158/1078-0432.CCR-18-3200
50. Yang S, He P, Wang J, Schetter A, Tang W, Funamizu N, et al. A novel MIF signaling pathway drives the Malignant character of pancreatic cancer by targeting NR3C2. *Cancer Res.* (2016) 76:3838–50. doi: 10.1158/0008-5472.CAN-15-2841
51. Lue H, Thiele M, Franz J, Dahl E, Speckgens S, Leng L, et al. Macrophage migration inhibitory factor (MIF) promotes cell survival by activation of the Akt pathway and role for CSN5/JAB1 in the control of autocrine MIF activity. *Oncogene.* (2007) 26:5046–59. doi: 10.1038/sj.onc.1210318
52. Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S, et al. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell.* (2018) 174:1293–308.e36. doi: 10.1016/j.cell.2018.05.060
53. Piskounova E, Agathocleous M, Murphy MM, Hu Z, Huddleston SE, Zhao Z, et al. Oxidative stress inhibits distant metastasis by human melanoma cells. *Nature.* (2015) 527:186–91. doi: 10.1038/nature15726
54. Green AC, Marttila P, Kiweler N, Chalkiadaki C, Wiita E, Cookson V, et al. Formate overflow drives toxic folate trapping in MTHFD1 inhibited cancer cells. *Nat Metab.* (2023) 5:642–59. doi: 10.1038/s42255-023-00771-5
55. Yao S, Peng L, Elakad O, Kuffer S, Hinterthaler M, Danner BC, et al. One carbon metabolism in human lung cancer. *Transl Lung Cancer Res.* (2021) 10:2523–38. doi: 10.21037/tlcr-20-1039
56. Shao H, Ge M, Zhang J, Zhao T, Zhang S. Osteoclasts differential-related prognostic biomarker for osteosarcoma based on single cell, bulk cell and gene expression datasets. *BMC Cancer.* (2022) 22:288. doi: 10.1186/s12885-022-09380-z
57. Lee SY, Choi JE, Jeon HS, Hong MJ, Choi YY, Kang HG, et al. A genetic variation in microRNA target site of KRT81 gene is associated with survival in early-stage non-small-cell lung cancer. *Ann Oncol.* (2015) 26:1142–48. doi: 10.1093/annonc/mdv100
58. Nanashima N, Horie K, Yamada T, Shimizu T, Tsuchida S. Hair keratin KRT81 is expressed in normal and breast cancer cells and contributes to their invasiveness. *Oncol Rep.* (2017) 37:2964–70. doi: 10.3892/or.2017.5564
59. Zhang Y, Willis-Owen SAG, Spiegel S, Lloyd CM, Moffatt MF, Cookson WOCM, et al. The ORMDL3 asthma gene regulates ICAM1 and has multiple effects on cellular inflammation. *Am J Respir Crit Care Med.* (2019) 199:478–88. doi: 10.1164/rccm.201803-0438OC
60. Siow D, Sunkara M, Dunn TM, Morris AJ, Wattenberg B. ORMDL/serine palmitoyltransferase stoichiometry determines effects of ORMDL3 expression on sphingolipid biosynthesis. *J Lipid Res.* (2015) 56:898–908. doi: 10.1194/jlr.M057539
61. Landau DA, Tausch E, Taylor-Weiner AN, Stewart C, Reiter JG, Bahlo J, et al. Mutations driving CLL and their evolution in progression and relapse. *Nature.* (2015) 526:525–30. doi: 10.1038/nature15395
62. Marucci A, Antonucci A, Bonis De C, Mangiacotti D, Scarale MG, Trischitta V, et al. GALNT2 as a novel modulator of adipogenesis and adipocyte insulin signaling. *Int J Obes (Lond).* (2019) 43:2448–57. doi: 10.1038/s41366-019-0367-3
63. Hu Q, Tian T, Leng Y, Tang Y, Chen S, Lv Y, et al. The O-glycosylating enzyme GALNT2 acts as an oncogenic driver in non-small cell lung cancer. *Cell Mol Biol Lett.* (2022) 27:71. doi: 10.1186/s11658-022-00378-w
64. Liu Y, Chen P, Xu L, Wang B, Zhang S, Wang X, et al. GALNT2 sustains glioma stem cells by promoting CD44 expression. *Aging (Albany NY).* (2023) 15:2208–20. doi: 10.18632/aging.204609
65. Liao YY, Chuang YT, Lin HY, Lin NY, Hsu TW, Hsieh SC, et al. GALNT2 promotes invasiveness of colorectal cancer cells partly through AXL. *Mol Oncol.* (2023) 17:119–33. doi: 10.1002/1878-0261.13347