The Oncologist, 2025, **30**, oyaf191
https://doi.org/10.1093/oncolo/oyaf191
Advance access publication 24 June 2025
Original Article

Integrative multiomic and immune profiling of lung adenocarcinoma: molecular landscapes, gene expression, and treatment response insights

Yahui Tian[1,‡], Honghong Dong[1,‡], Yujie Guo[1,2], Shaowei Xin[1,3], Suxin Jiang[1], Zitong Wan[1,4], Huaiyu Wang[1,*], Yong Han[1,*]

[1]Department of Thoracic Surgery, Air Force Medical Center, Fourth Military Medical University, Beijing 100142, China
[2]Graduate School of China Medical University, Shenyang 110122, China
[3]Department of Thoracic Surgery, 962 Hospital of the joint Logistics Support Force, Harbin 150080, China
[4]College of Life Sciences, Northwestern University, Xi'an 710069, China

[*]Corresponding authors. Yong Han, Department of Thoracic Surgery, Air Force Medical Center, Fourth Military Medical University, No. 28 Fucheng Road, Haidian District, Beijing 100142, China. E-mail: hanyong_td@163.com; Huaiyu Wang, Department of Thoracic Surgery, Air Force Medical Center, Fourth Military Medical University, No. 28 Fucheng Road, Haidian District, Beijing 100142, China. E-mail: alexpepsi@126.com.
[‡]Y. Tian and H. Dong contributed equally to this work.

Abstract

Background: Lung adenocarcinoma (LUAD) is a major cause of cancer death. Traditional histopathological classification overlooks molecular heterogeneity, limiting personalized treatment. This study used multiomic data to define LUAD subtypes, assess prognostic significance, and analyze immune features, aiming to improve targeted therapy and clinical outcomes.

Methods: This study used Consensus Clustering and Gap Statistics to analyze LUAD multiomic data, including mRNA, lncRNA, miRNA, DNA methylation, and mutations. Clustering was validated by silhouette plots and heatmaps. Molecular characterization involved regulon activity, immune and metabolic profiling. Functional assays (qPCR, WB, CCK-8, flow cytometry) assessed neuronal differentiation factor (NDNF)'s role in LUAD.

Results: Two molecular LUAD subtypes showed distinct clustering and survival outcomes. One subtype had worse prognosis and unique immune features, including checkpoint expression and microenvironment differences. Gene signatures and metabolism varied by subtype. Neuronal differentiation factor was downregulated in tumors; its overexpression suppressed LUAD cell viability and promoted apoptosis, suggesting tumor-suppressive function.

Conclusions: This study identifies 2 LUAD subtypes with distinct molecular and immune features linked to prognosis and therapy response. Neuronal differentiation factor downregulation and its tumor-suppressive effects highlight its therapeutic potential. These findings support improved LUAD stratification and personalized treatment strategies.

Key words: lung adenocarcinoma; multiomics; molecular subtypes; immune profiling; personalized medicine; differential gene expression.

Implications for Practice

This study reveals 2 lung adenocarcinoma (LUAD) molecular subtypes with distinct prognoses and immune features, offering a basis for more precise treatment strategies. The identification of neuronal differentiation factor as a potential tumor suppressor suggests its value as both a biomarker and therapeutic target. These findings support improved LUAD stratification and personalized therapy.

Introduction

Lung adenocarcinoma (LUAD) represents a prevalent form of nonsmall cell lung cancer, accounting for a significant proportion of all lung cancer cases.[1] Originating from the glandular cells in the lungs, responsible for the secretion of mucus and other fluids, LUAD is characterized by its complex molecular structure and diverse clinical manifestations.[2] This complexity poses challenges in its diagnosis and treatment. Despite advances in radiotherapy, chemotherapy, and surgical interventions, LUAD's intricate nature demands a thorough understanding to devise effective treatment strategies, underscoring the need for comprehensive research and analysis in this domain. The molecular landscape of LUAD is notably complex, marked by a wide array of genetic mutations, variations in gene expression, and diverse epigenetic alterations.[3] This heterogeneity not only complicates the understanding of the disease's pathogenesis but also poses significant challenges in the development of effective, targeted treatment strategies. Traditional one-size-fits-all treatment approaches are often ineffective due to the unique molecular characteristics of

Received: 16 October 2024; Accepted: 18 June 2025.

each patient's tumor. Consequently, there is a pressing need for a comprehensive analysis of LUAD's molecular landscape. Such an analysis would involve integrating multiomic data—genomics, transcriptomics, proteomics, and epigenomics—to gain a holistic view of the disease.[4] This integrative approach is crucial for identifying specific molecular subtypes of LUAD, understanding the underlying mechanisms driving each subtype, and uncovering potential therapeutic targets.

Recent advancements in multiomic technologies have ushered in a new era in our understanding of diseases like LUAD.[5] Multiomics refers to the collective analysis and integration of datasets from various biological layers, including genomics, transcriptomics, proteomics, and epigenomics.[6] Each omics layer reveals distinct aspects of cancer biology: genomics identifies mutations, transcriptomics captures gene expression, proteomics shows protein activity, and epigenomics reveals regulatory changes. Integrating these data offers a comprehensive view of LUAD, linking genetic, transcriptional, protein, and epigenetic alterations. This approach uncovers molecular subtypes, improves prognostic accuracy, and identifies new therapeutic targets.

This study aims to integrate multiomic data to identify molecular subtypes of LUAD, using the combined insights of genomics, transcriptomics, proteomics, and epigenomics to clarify disease heterogeneity and improve prognostic and therapeutic precision.

Methods

Data collection and preprocessing

The dataset included 11 427 LUAD samples collected from The Cancer Genome Atlas (TCGA) (https://tcga-data.nci.nih.gov/tcga/), GEO datasets (GSE31210, GSE50081, GSE64865, GSE68465), and single-cell cohorts. This dataset encompassed multiomic data, including messenger RNA (mRNA), long non-coding RNA (lncRNA), microRNA (miRNA), DNA CpG methylation, and gene mutation profiles (Supplementary dataset).

Missing data imputation was applied to mRNA expression, lncRNA expression, miRNA expression, and DNA methylation data, specifically addressing missing values caused by low gene expression levels and incomplete sequencing coverage. The *k*-Nearest Neighbors (*k*-NN) algorithm was used to impute these missing values while preserving the intrinsic data structure. Probes and samples with excessive missing values beyond a predefined threshold were excluded to ensure data quality.

Normalization techniques were employed to adjust for batch effects and other potential confounders in the data. Specifically, the ComBat algorithm was applied to correct for batch effects, ensuring that variations due to different batches were minimized while preserving the biological signal [7]. Additionally, log transformation was used to stabilize variance and reduce skewness in the distribution of expression levels. To further standardize the data, *Z*-score standardization was applied to each dataset independently, transforming the data to have a mean of 0 and a standard deviation of 1 [8]. This combination of the ComBat algorithm and *Z*-score standardization facilitates direct comparisons across different types of molecular data and between samples, by mitigating the influence of both batch effects and disparate scale and distribution.

Clustering analysis

The clustering process was guided by 2 principal metrics: the Cluster Prediction Index (CPI) [9] and Gap Statistics. The CPI was employed to assess the internal coherence and separation between potential clusters, providing a quantifiable measure of cluster quality based on within-cluster variance relative to between-cluster variance [9]. The Gap Statistics, on the other hand, compared the log-transformed within-cluster sums of squares across different *k*-values with their expected values under a null reference distribution of the data [10].

The optimal number of clusters was determined using silhouette plots and Gap Statistics [11]. Silhouette analysis involves calculating a score for each sample in a cluster, which measures how similar that sample is to samples in its own cluster compared to samples in other clusters. This metric ranges from −1 to 1, where a high value indicates that the sample is well matched to its own cluster and poorly matched to neighboring clusters. The silhouette plot provides a visual representation of how each cluster stands apart from others, which aided in determining the most coherent and distinct clustering solution. The Gap Statistic further corroborated these findings by providing a statistical basis for choosing the number of clusters that best captured the inherent grouping in the data.

To robustly cluster the multiomic data, 10 advanced clustering methods were utilized, each chosen for its ability to handle the complexity and dimensionality of the data effectively [12]. Molecular subtypes of LUAD were identified using robust consensus clustering approaches. These methods integrated multiomic data to uncover distinct and reproducible subtypes validated across external cohorts.

Subtype identification and validation

Consensus clustering was employed to integrate the results from multiple clustering algorithms, enhancing the robustness and reproducibility of the identified subtypes [13]. This method uses a consensus clustering matrix to aggregate the clustering outcomes from different methods, where each entry in the matrix represents the proportion of algorithms that assign the same pairs of samples to the same cluster. The consensus matrix thus provides a visual and quantitative measure of agreement among the various methods, highlighting consistent groupings. The final subtypes were determined based on the stability of clusters across algorithms, as evidenced by high consensus values and low variability, ensuring that the chosen clusters represent genuine underlying biological patterns rather than artifacts of any single clustering technique.

Survival analysis was conducted to assess the prognostic differences among the identified molecular subtypes. Kaplan–Meier survival curves were plotted to visualize the survival probabilities over time for each subtype, providing an intuitive representation of survival distribution [14]. Additionally, the Cox proportional hazards model was applied to quantify the impact of subtype classification on survival outcomes [15]. This model adjusts for various covariates and computes hazard ratios, offering insights into the relative risk of mortality associated with each subtype. The significance of differences in survival among subtypes was evaluated using log-rank tests, establishing a statistically grounded basis for prognostic assessments.

Validation of the identified LUAD subtypes was performed using external cohorts, namely META-LUAD and

IMvigor-LUAD, to confirm generalizability and robustness. Two principal methods were utilized: Nearest Template Prediction (NTP) and Prediction Analysis for Microarrays (PAM). NTP involves comparing the molecular profile of new samples against predefined template profiles representing each subtype to assign the most similar template to each sample [16]. PAM, on the other hand, uses a class prediction strategy based on gene expression data to classify new samples into one of the established subtypes [17]. These methods were crucial in demonstrating that the subtypes not only hold in the original TCGA cohort but also retain their prognostic relevance and distinct molecular signatures in diverse and independent patient populations. Statistical measures such as classification accuracy, sensitivity, and specificity were calculated to evaluate the performance of subtype assignments in these validation cohorts.

Immune microenvironment analysis

To evaluate the immune cell infiltration within the tumor microenvironment (TME), several computational tools were utilized, each offering unique insights into the composition and function of immune cells in the context of LUAD. CIBERSORT was employed to deconvolute the cell composition of complex tissues based on gene expression data, providing estimates of the relative proportions of immune cell types within the tumor [18]. MCPcounter estimated the abundance of various immune and stromal cell populations using a set of marker genes characteristic of each cell type, thus offering quantification of these cells in the tumor samples [19]. TIMER utilized a similar approach but with an emphasis on the influence of tumor purity, thereby adjusting the infiltration scores to reflect the true immune contexture more accurately [20]. Lastly, quantISEQ was applied to specifically quantify tumor-infiltrating lymphocytes, using a machine learning model trained on digital cytometry data [15].

The expression of immune checkpoint genes, which play critical roles in regulating immune responses, was analyzed to understand their differential regulation across the identified molecular subtypes. Using RNA-Seq data, expression levels of key checkpoint genes such as PD-1, CTLA-4, LAG-3, and others were quantified. Statistical tests, including ANOVA and post-hoc Tukey tests, were conducted to detect significant differences in expression among the subtypes. This analysis was complemented by correlation studies that linked expression patterns with clinical data such as treatment response and survival outcomes. Further, the data were visualized using box plots and heatmaps, facilitating an intuitive understanding of the variation in checkpoint gene expression across subtypes. This analysis not only highlighted the potential mechanisms underlying immune evasion but also suggested targeted therapeutic strategies that could be employed to enhance the efficacy of immunotherapy based on the specific immune profiles of different LUAD subtypes.

Biomarker expression and drug response prediction

To analyze the expression levels of potential biomarkers and their correlation with patient risk groups, statistical methodologies were rigorously applied. Initially, biomarker candidates were identified based on their known or predicted involvement in LUAD pathogenesis or therapy response. Expression data for these biomarkers were then extracted from the multiomic datasets. For comparative analysis, patients were stratified into high-risk and low-risk groups based on a composite risk score derived from clinical data and molecular profiles. The expression levels of biomarkers across these groups were compared using the Mann–Whitney *U*-test for nonparametric data, ensuring that differences in median expression levels were not influenced by outliers or non-normal distributions. Additionally, for biomarkers with parametric expression profiles, *t*-tests were utilized.

The prediction of patient-specific drug responses was based on integrating molecular characteristics with pharmacogenomic data. This integration was facilitated by the use of predictive models that incorporate both genetic and epigenetic features relevant to drug metabolism and efficacy. Specifically, machine learning models such as Random Forests and Support Vector Machines were trained on datasets comprising genomic profiles and historical treatment outcomes [21]. These models were used to predict the efficacy of various chemotherapeutic agents and targeted therapies, providing personalized treatment recommendations. Additionally, the drug response prediction was enhanced by the use of pharmacogenomic databases like the Genomics of Drug Sensitivity in Cancer, which provides drug sensitivity profiles for various cancer cell lines [22].

Single-cell RNA sequencing and data analysis

LUAD and adjacent normal tissue samples were collected from patients undergoing surgery. Fresh tissue samples were immediately dissociated into single cells using enzymatic digestion, following established protocols. Single-cell suspensions were filtered to remove cell debris and then subjected to viability assessment using Trypan Blue staining. Cells with >85% viability were used for subsequent analysis.

Single-cell RNA sequencing libraries were prepared using the 10x Genomics Chromium platform according to the manufacturer's instructions [23]. Individual cells were captured in microdroplets, and barcoded mRNA was reverse-transcribed to generate cDNA libraries. Sequencing was performed on an Illumina platform, generating paired-end reads with an average depth of 50 000 reads per cell.

The raw sequencing data were processed using the Cell Ranger pipeline (v.3.1, 10x Genomics) to align reads to the human reference genome (GRCh38) and to generate the gene expression matrix. Low-quality cells with fewer than 200 detected genes, or cells with a mitochondrial gene content >10%, were excluded from the analysis. Genes expressed in fewer than 3 cells were also removed.

After normalization and log transformation of the expression data, highly variable genes were selected for downstream analysis. Principal component analysis (PCA) was used for dimensionality reduction, and the top principal components were selected for clustering. Cells were clustered using the Seurat R package (v.3.1) based on the Louvain algorithm. Uniform Manifold Approximation and Projection (UMAP) was employed to visualize the clustering results.

Clusters were annotated by comparing differentially expressed genes (DEGs) in each cluster with known cell type markers from published literature. Identified cell types included T cells, B cells, NK cells, epithelial cells (Epi), macrophages (Mac), monocytes (Mon), fibroblasts (Fib), dendritic cells (MDC), mast cells (Mast), plasma cells (Plasma), endothelial cells (Endo), and plasmacytoid dendritic cells (PDC).

Statistical methods and software tools

The analyses described in this study were supported by a comprehensive suite of software tools and packages, primarily within the R and Python ecosystems. Key software used included: R (version 4.1.2): Utilized for statistical analyses and visualization, with packages such as survival for survival analyses, limma for differential expression studies, and ConsensusClusterPlus for consensus clustering. Python (version 3.8): Employed for machine learning models and data preprocessing, with libraries including scikit-learn for modeling and pandas for data manipulation. CIBERSORT: a standalone tool for characterizing cell composition of complex tissues from their gene expression profiles. TIMER: used within a web-based platform for immune infiltration analysis. quantISEQ: implemented via a Python script to quantify tumor-infiltrating lymphocytes based on transcriptomic data.

A variety of statistical tests were applied to ensure the validity of the findings across different types of data analyses: Kaplan–Meier Analysis and Cox Proportional Hazards Model: these methods were used to evaluate survival differences, with log-rank tests for comparing Kaplan–Meier curves to determine statistical significance. ANOVA and *t*-tests: Employed for comparing gene expression across multiple groups, with post-hoc Tukey tests to adjust for multiple comparisons when significant ANOVA results were found. Mann–Whitney *U* and Wilcoxon Signed-Rank Tests: these nonparametric tests were used where data did not assume normal distribution, suitable for comparing two independent or paired groups, respectively. Benjamini–Hochberg Procedure: this false discovery rate controlling procedure was used to adjust *P*-values in the context of multiple hypothesis testing, particularly in the gene expression and biomarker analyses, ensuring that the likelihood of type I errors was minimized.

Cell culture

Normal lung cell samples and LUAD cell lines, A549 cells, were cultured in Dulbecco's Modified Eagle Medium supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin [24]. The cells were maintained at 37°C in a humidified atmosphere containing 5% CO_2. The medium was replaced every 2–3 days, and cells were passaged at approximately 80%–90% confluence using trypsin-EDTA solution for detachment.

qPCR analysis

Total RNA was extracted from the cell samples using an RNA extraction kit. The RNA was then reverse transcribed into cDNA using a reverse transcription kit. Quantitative PCR (qPCR) was performed using specific primers for NDNF and a SYBR Green master mix. The relative expression levels of NDNF were calculated using the 2^-ΔΔCt method, with GAPDH as the internal control.

Western blot analysis

Proteins were extracted from the cell samples using a lysis buffer. The extracted proteins were separated by SDS-PAGE and transferred onto a PVDF membrane. The membrane was then probed with anti-NDNF primary antibodies, followed by incubation with HRP-conjugated secondary antibodies. Protein bands were visualized using an enhanced chemiluminescence detection system, and the relative protein levels were quantified using image analysis software.

Adenoviral vector construction

Adenoviral vectors carrying the full-length NDNF cDNA were generated through standard cloning techniques. The NDNF cDNA was inserted into the adenoviral vector, driven by a strong promoter to ensure robust expression. A549 cells were subsequently transduced with these adenoviral vectors using the Lipofectamine 3000 reagent, facilitating efficient gene delivery.

CCK8 Assay

A549 cells, both NDNF overexpressing and control, were seeded into 96-well plates at a density of 5000 cells per well. After allowing the cells to adhere and grow for 24 hours, the Cell Counting Kit-8 (CCK8) reagent was added to each well according to the manufacturer's instructions [25]. The plates were incubated for 2 hours at 37°C. The absorbance of each well was measured at 450 nm using a microplate reader. Cell viability was assessed by comparing the absorbance values of NDNF overexpressing cells to those of control cells.

Results

Prognostic relevance of S-MMR score in LUAD and its association with clinical characteristics

We applied the S-MMR score (somatic mismatch repair score) to stratify LUAD patients into high- and low-risk groups across TCGA, GSE68465, GSE50081, and GSE31210 datasets (Figure 1A–D). Kaplan–Meier survival curves consistently showed that patients with high-S-MMR scores had significantly poorer overall survival compared to those with low scores. This trend was statistically significant across all datasets, as indicated by log-rank *P*-values. A combined survival analysis across all cohorts (Figure S1A) further confirmed the prognostic value of the S-MMR score, with the high-S-MMR group exhibiting markedly worse survival outcomes (log-rank $P < .0001$). Boxplots (Figure S1B) demonstrated significant correlations between risk scores and clinical characteristics: males, patients with higher T, N, and M stages, advanced overall stage, and older age (>65) had higher-risk scores, linking these features with worse prognosis. Donut plots (Figure S1C) illustrated the distribution of clinical characteristics between high- and low-S-MMR groups. Chi-squared tests revealed significant associations between the S-MMR score and gender ($P = .0154$), T stage ($P = .0197$), N stage ($P < .0001$), M stage ($P = .0111$), and age ($P = .0088$), reinforcing the link between high S-MMR scores and more aggressive clinical features. These findings establish the S-MMR score as a strong prognostic indicator in LUAD, demonstrating its association with poorer survival and aggressive disease characteristics. This score could serve as a valuable tool for patient stratification and personalized treatment strategies.

Immune landscape and immunotherapy response in LUAD

Figure 2A compares immune cell infiltration between high-risk (HR) (red) and low-risk (LR) (green) LUAD patient groups using CIBERSORT, MCPcounter, and XCell. Boxplots show that CD4+ memory T cells, fibroblasts, neutrophils, and dendritic cells are significantly enriched in the LR group, suggesting a more immune-activated tumor microenvironment. In contrast, myeloid-derived suppressor cells, known for promoting immune suppression, are elevated in HR

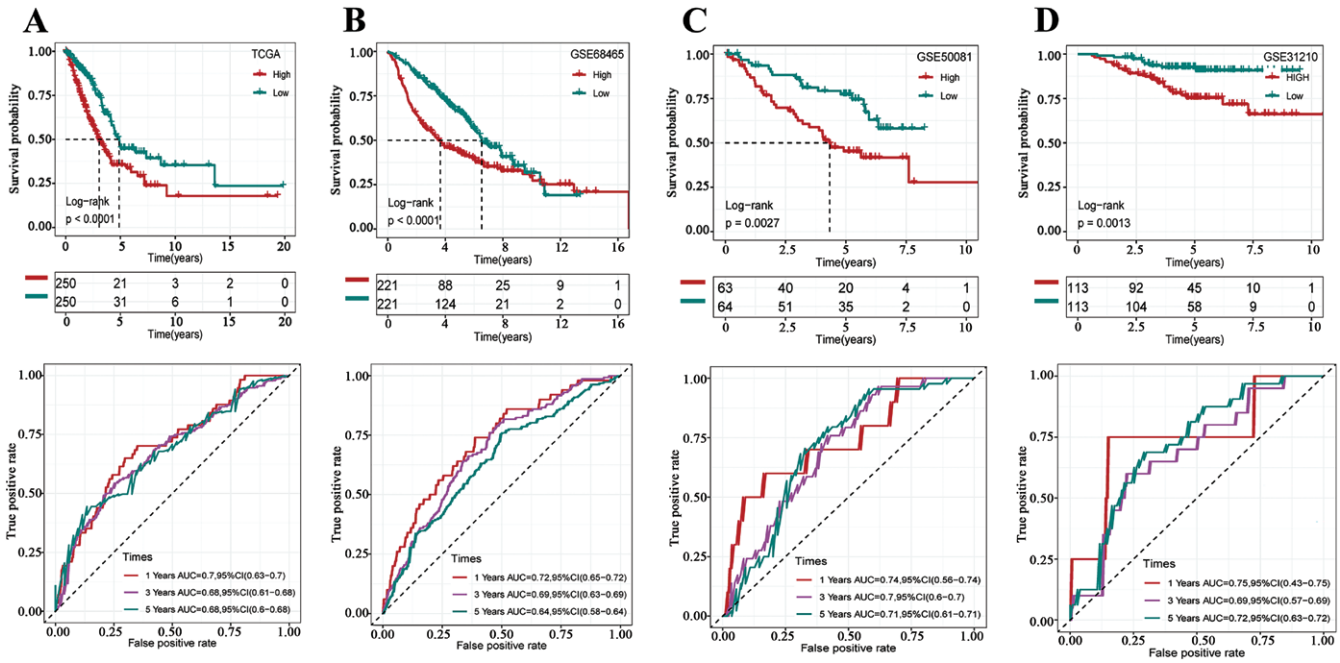

Figure 1. Survival analysis of S-MMR score in LUAD: multicohort insights and clinical correlations. (A) Kaplan–Meier survival curves and ROC curves of 3 S-MMR score in the training set 1 (TCGA-LUAD). (B) Kaplan–Meier survival curves and ROC curves of 3 S-MMR score in the training set 2 (GSE68465). (C) Kaplan–Meier survival curves and ROC curves of 3 S-MMR score in the training set 3 (GSE50081). (D) Kaplan–Meier survival curves and ROC curves of 3 S-MMR score in the training set 4 (GSE31210). **P* < .05; ***P* < .01; ****P* < .001.

patients, contributing to poorer outcomes. Figure 2B highlights differential expression of immune recruitment and activation signatures. The LR group shows higher recruitment signatures for CD4+ T cells, supporting a more active antitumor response. Conversely, macrophage-related signatures are enriched in HR patients, indicating a more immunosuppressive tumor microenvironment (TME). Figure 2C demonstrates that HR patients exhibit increased expression of immune checkpoint molecules, including PD-L1 (CD274) and LAG3, which dampen immune responses and facilitate tumor immune evasion. Figure 2D presents the distribution of risk scores among patients with different responses to immune checkpoint inhibitors. Kruskal–Wallis analysis (P = 7.6e-05) shows that lower-risk patients are more likely to achieve Partial or Complete Response, while higher-risk patients tend to have Progressive Disease. Figure 2E utilizes Tumor Immune Dysfunction and Exclusion (TIDE) analysis to predict ICI response, revealing that 75.6% of HR patients are likely nonresponders, whereas 48.4% of LR patients are predicted responders (P = .000). Figure S1D further breaks down the TIDE score, showing significantly higher immune dysfunction and exclusion in HR patients (P < .0001), suggesting impaired antitumor immunity. These findings underscore significant differences in immune infiltration, immune checkpoint expression, and predicted ICI response between risk groups, highlighting the potential for stratifying LUAD patients for immunotherapy.

Differential gene expression and its clinical implications in LUAD risk stratification

Figure 3A presents Spearman rank correlation analysis between the 3S-MMR score and gene expression in the TCGA-LUAD cohort. The volcano plot highlights significant correlations, with genes such as PNP, SMS, VDAC2, and SLCO1B1 showing positive correlations with HR scores, suggesting their elevated expression in HR patients. Scatter plots demonstrate a strong positive correlation between the 3S-MMR score and tumor mutation burden as well as tumor neoantigen burden (r > 0.5, P = 0), indicating that HR patients have higher mutation and neoantigen loads, which may influence immune evasion and treatment response. Figure 3B identifies genes negatively correlated with the 3S-MMR score, including TOP2A, KIF18A, and SLC25A13 (R < −0.3). Their lower expression in HR patients suggests potential tumor-suppressive roles. Figure 3C evaluates drug sensitivity, showing that HR patients have significantly higher half-maximal inhibitory concentration (IC50) values for Paclitaxel, Gemcitabine, and Cisplatin, indicating reduced sensitivity to these chemotherapy agents. Figure S1E further explores drug resistance through area under the curve (AUC) comparisons. High-risk patients exhibit significantly higher AUC values for multiple chemotherapeutic and targeted agents, including BI-2536, KX2-391, Leptomycin B, and SB-743921, suggesting greater drug resistance. Additional drugs, such as Docetaxel, Cabazitaxel, Vincristine, and SNS-314, also show higher AUC values in HR patients, reinforcing the association between the 3S-MMR score and poor treatment response. These findings highlight the link between the 3S-MMR score, gene expression, mutation burden, and chemotherapy resistance, emphasizing its potential role in guiding treatment strategies for LUAD patients.

Comprehensive immune and molecular landscape analysis in LUAD

Figure 4A presents a radar chart comparing molecular subtypes of LUAD between HR and LR groups. High-risk patients show stronger associations with aggressive features such as epithelial-mesenchymal transition and EGFR mutations,

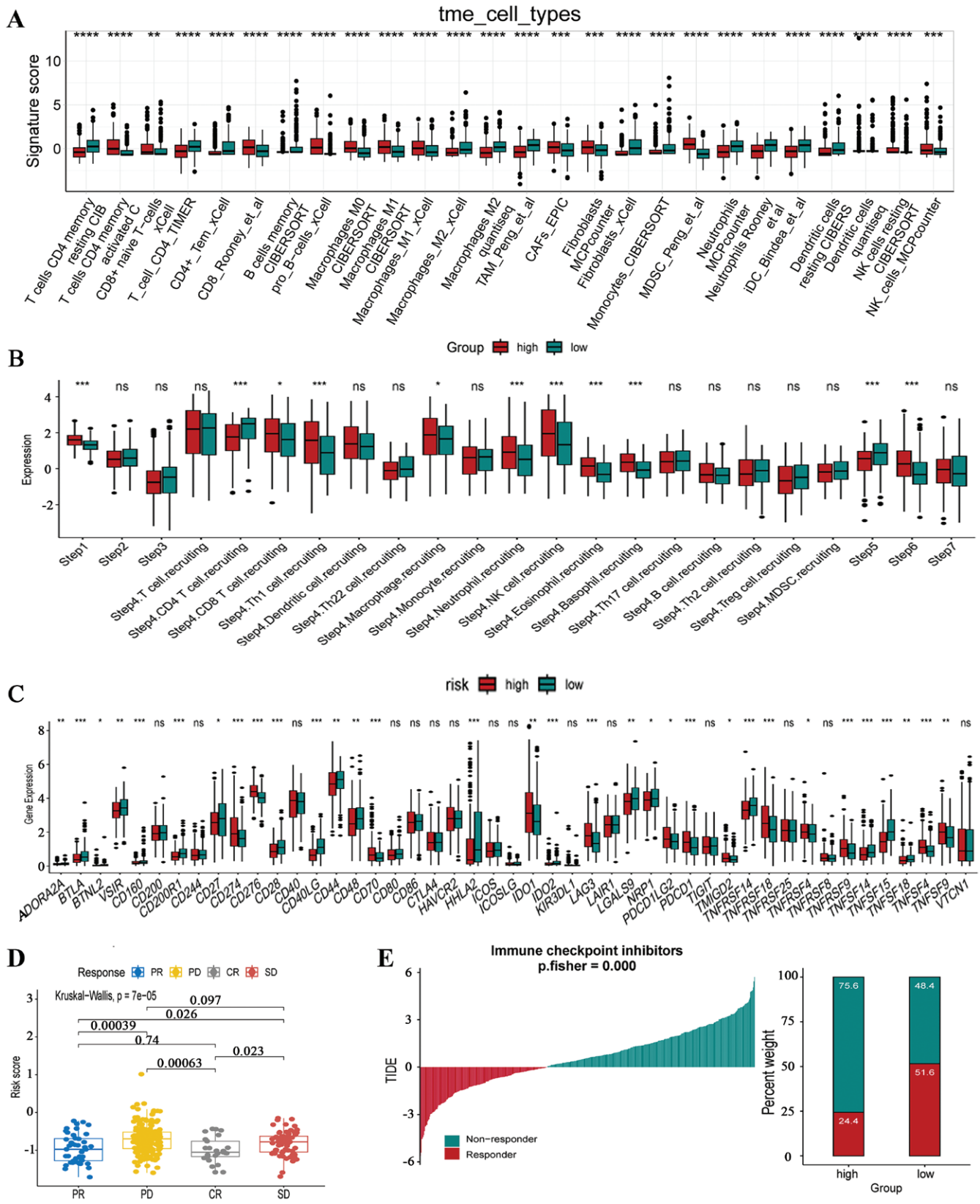

Figure 2. Integrated immune landscape profiling in cancer patient cohorts. (A) Boxplot illustrating the relative abundance of infiltrating immune cell types in patients belonging to high- and low-risk groups. (B and C) Boxplot of relative expression levels at 27 immune checkpoints profiles between the high- and low-risk patients. (D) Risk score distribution for treatment responses (PR, PD, CR, SD) displayed in color-coded scatter and box plots. (E) Bar Graph of Immune Checkpoint Inhibitor Response.

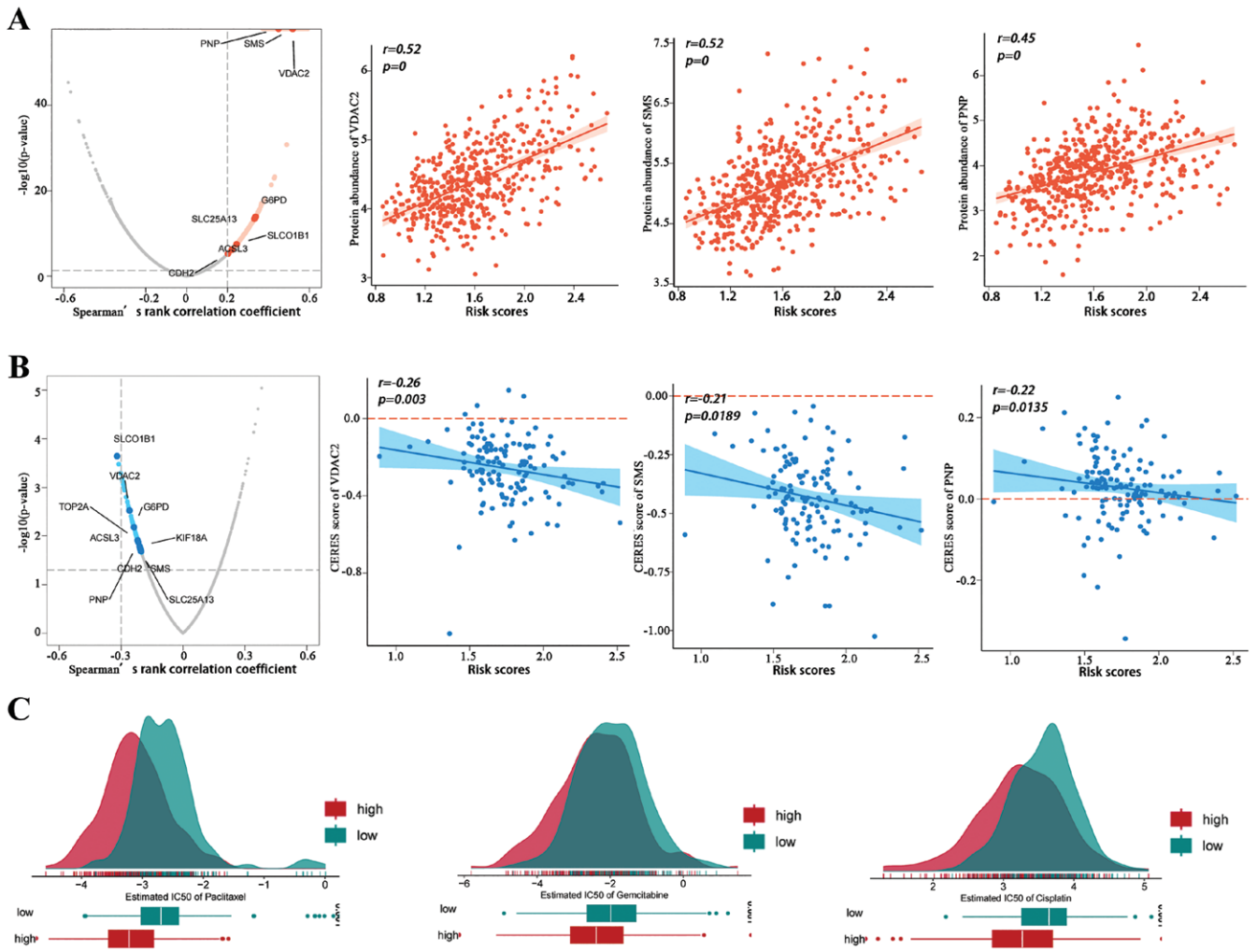

Figure 3. Comprehensive analysis comparing HR and LR patient groups in LUAD. (A) Volcano plot and scatter plots of the correlation coefficients derived from Spearman's rank correlation analysis between 3 S-MMR score and druggable mRNA expression in the TCGA-LUAD cohort. Red dots indicate the significant positive correlations ($P < .05$, and Spearman's $r > 0.2$). (B) Volcano plot and scatter plots of the correlations and significance between 3 S-MMR score and CERES score of drug targets. Green dots indicate the significant negative correlations ($P < .05$, and Spearman's $r < -0.2$). (C) The comparison of IC50 values between high- and low 3 S-MMR score groups of Paclitaxel, Gemcitabine, Cisplatin. *$P < .05$; **$P < .01$; ***$P < .001$.

while LR patients exhibit a greater association with microsatellite instability (MSI), suggesting distinct biological behaviors between risk groups. Figure 4B displays a heatmap of pathway enrichment, revealing that HR patients have significant activation of oncogenic pathways, including NOTCH, TGF-beta, and WNT signaling, along with DNA repair pathways such as Mismatch Repair and Nucleotide Excision Repair, consistent with more aggressive tumor biology. In contrast, LR patients show enrichment in metabolic pathways (eg, Glycolysis/Gluconeogenesis, Oxidative Phosphorylation) and immune-related pathways (eg, Cytokine-cytokine receptor interaction), suggesting a more favorable immune response and prognosis. Figure 4C compares immune-related gene expression between risk groups. High-risk patients display elevated levels of inhibitory immune checkpoints (eg, PD-L2, CTLA4), indicating an immunosuppressive environment, while human leukocyte antigen genes related to antigen presentation are downregulated, suggesting impaired immune function. This pattern indicates that HR patients may exhibit reduced immune responses, potentially affecting their response to immunotherapy. These findings highlight key molecular and immune differences between LUAD risk groups, providing insights into tumor biology and potential therapeutic strategies.

Multiomic analysis reveals subtypes and prognostic implications in LUAD

To classify molecular subtypes of LUAD, we performed an integrative multiomic analysis incorporating mRNA, lncRNA, miRNA, DNA methylation, and mutation profiles. Cluster Prediction Index and Gap Statistics determined 2 optimal clusters, as supported by the elbow method (Figure S2A). Heatmap visualization revealed distinct molecular characteristics between the identified subtypes, including the differential expression of PCP4L1 (Purkinje cell protein 4-like 1) and LINC02489, suggesting potential variations in neuronal differentiation and regulatory pathways (Figure S2B). Notably, PCP4L1, a gene involved in neuronal differentiation and development, was upregulated in subtype C2 and downregulated in subtype C1. miRNA (eg, miR-133a-2) and DNA methylation patterns also varied between clusters, reinforcing molecular heterogeneity. The silhouette plot showed a clear separation

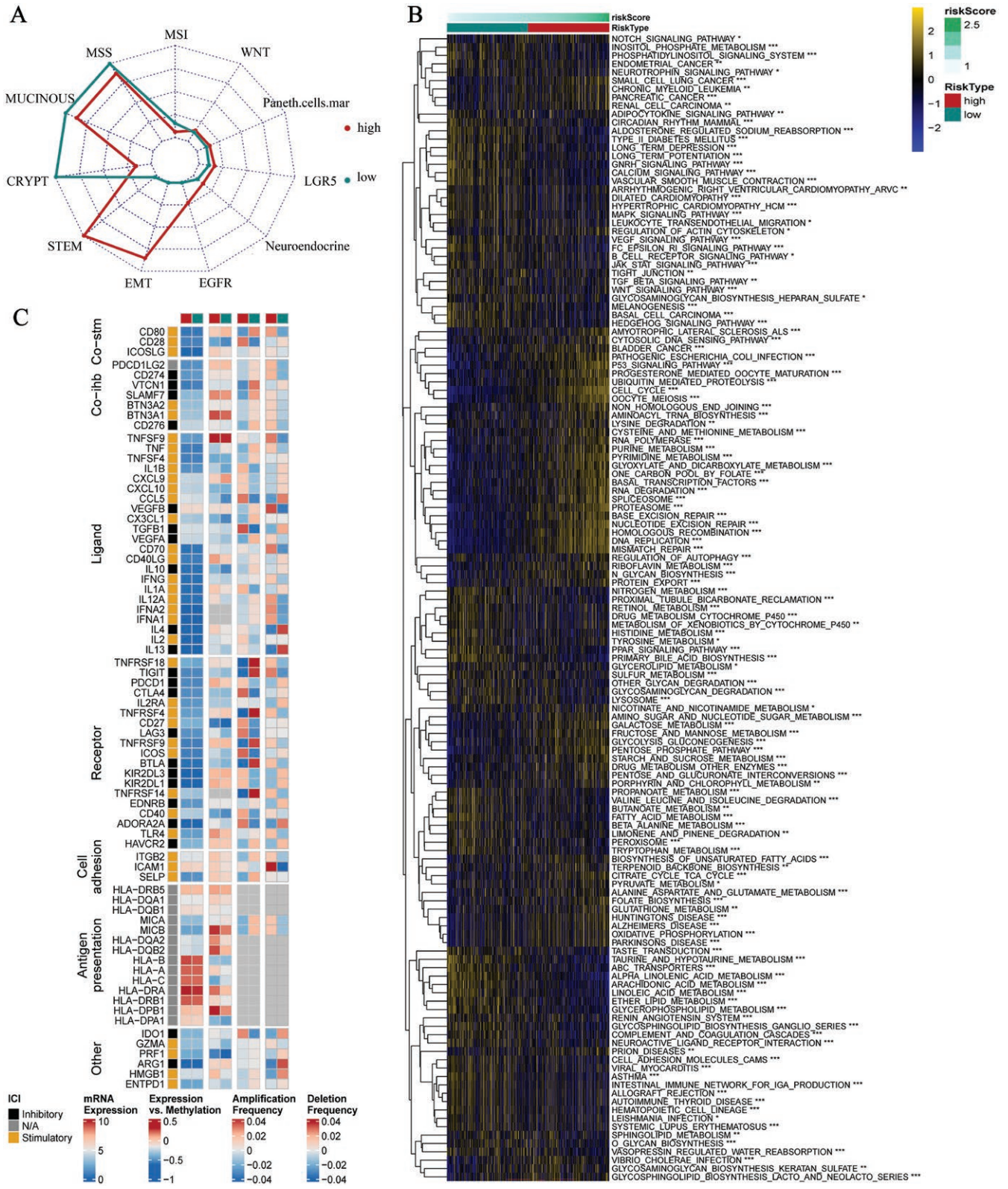

Figure 4. Comprehensive analysis of molecular subtypes and immune landscape in LUAD. (A) Radar chart comparing molecular characteristics between HR and LR LUAD patients. (B) Heatmap displaying the expression patterns of numerous genes across multiple patient samples, using hierarchical clustering to group similar profiles. (C) Heatmap of immune-related gene expression, revealing higher levels of inhibitory checkpoints in HR patients and lower HLA gene expression.

between C1 and C2 (average silhouette width = 0.6), indicating well-defined clusters (Figure S2C). Multiomics clustering of the Molecular Understanding of Cancer cohort using 10 advanced methods confirmed the robustness of these subtypes (Figure S2D). Figure S2E illustrates the consensus clustering matrix derived from the TCGA-LUAD dataset, demonstrating

the stability and robustness of the identified subtypes (C1 and C2). Survival analysis revealed a significant difference in overall survival ($P = .001$), with subtype C1 associated with worse outcomes, suggesting it may represent a more aggressive form requiring distinct treatment approaches (Figure S2F). These findings provide a basis for molecularly guided therapeutic strategies in LUAD.

Comprehensive molecular characterization and validation of LUAD CSs

To explore functional differences between the LUAD subtypes, we performed pathway enrichment analysis across multiple molecular layers. Figure S3A highlights differential activity in key pathways: CS1 demonstrates higher proliferative activity, as indicated by the upregulation of cell cycle regulators, hypoxia-related genes, and DNA replication pathways. CS2 shows elevated activity in pathways such as WNT/β-catenin signaling, PPARG network, and FGFR3-coexpressed genes, which are associated with cellular differentiation and metabolic regulation. Immune-related pathways, such as cytokine-cytokine receptor interaction and antigen processing, are modestly enriched in CS2, supporting its potential immune activity (Figure S3B). CS2 demonstrated higher immune cell infiltration, indicating a more immune-active tumor microenvironment (TME), whereas CS1 exhibited lower immune and stromal enrichment scores. These differences suggest CS2 may respond differently to immune checkpoint blockade therapies. To validate the robustness of these subtypes, we performed nearest template prediction (NTP) analysis in the META-LUAD cohort, confirming consistent classification of patients into CS1 and CS2 (Figure S3C). The strong agreement between CMOIC and NTP classifications (Kappa = 0.814, $P < .001$) further supported the reproducibility of our findings (Figure S3D). Survival analysis revealed a significant difference in prognosis, with CS1 associated with poorer overall survival ($P < .001$) (Figure S3E). To validate these findings, we applied the clustering results to the independent META-LUAD dataset using NTP and Prediction Analysis for Microarrays (PAM). Additional validation using CMOIC, NTP, and PAM demonstrated high classification consistency, with Kappa values of 0.828 and 0.909 (both $P < .001$) (Figure S3F and G), further confirming the robustness and reproducibility of the identified subtypes. Finally, enrichment analysis of biological processes (Figure S3H) identified distinct pathways, including small lung carcinoma-related processes, further differentiating CS1 and CS2.

Integrative analysis of gene significance and predictive modeling in LUAD

To identify molecular drivers and prognostic factors in LUAD, we conducted an integrative analysis across multiple datasets. Principal Component Analysis visualized gene expression profiles, revealing distinct clustering of samples across GSE31210, GSE50081, GSE64865, and TCGA datasets, with some overlap but clear separation between molecular subtypes (Figure S4A). Subtype-specific clustering within individual datasets showed minimal outliers, reinforcing LUAD's molecular heterogeneity (Figure S4B). Cox proportional hazard models assessed gene expression impact on survival, categorizing genes as risk or protective factors (Figure S4C). Genes LYPD3, KET6A, and PLK1 were identified as potential LUAD drivers, frequently selected across classifier runs (Figure S4D). Stability analysis across 1,000 iterations confirmed PLK1, AURKB, CHEK1, and CDKN3 as consistently significant in LUAD pathogenesis and classification (Figure S4E and F). Hazard ratio analysis highlighted UBE2S, RRM2, PLK1, PBK, NEK2, and LYPD3 as associated with poorer survival, whereas SLC34A2, RNASE1, and NDNF correlated with better prognosis (Figure S4G). Predictive performance evaluation of survival models, including CoxBoost, Elastic Net (Enet), LASSO, piRsRox, Ridge, SuperPC, and XGBoost, showed that CoxBoost and Enet achieved the highest median C-index values, indicating superior predictive accuracy (Figure S4H). These models also had lower Integrated Brier Score values, confirming their robustness in survival prediction (Figure S4I). This analysis underscores the importance of gene signatures in LUAD prognosis and their potential for therapeutic targeting.

Differential expression of key genes in LUAD and normal tissues reveals potential biomarkers and therapeutic targets

Figure S5A presents boxplots comparing the expression levels of multiple key genes between normal lung tissues and LUAD tissues. Across the board, significant upregulation of these genes is observed in LUAD samples compared to normal tissues, with P-values < .001 for most genes. Genes like AURKB, CDK1, CDKN3, CDT1, and CHEK1 show particularly high overexpression in LUAD, indicating their potential role in tumor progression. Interestingly, the NDNF gene, a potential tumor suppressor, shows marked downregulation in LUAD tissues compared to normal samples. Figure S5B displays paired boxplots comparing the gene expression levels in matched normal and LUAD tissues from the same patients. The upward trend in gene expression from normal to LUAD tissues is evident across all analyzed genes, such as AURKB, CDK1, CDKN3, and CDT1, which further confirms their significant upregulation in cancerous tissue. The paired data visualization reinforces the robustness of these genes as potential biomarkers for distinguishing LUAD from normal lung tissue. Importantly, the downregulation of NDNF is consistent across matched samples, underscoring its relevance as a tumor suppressor and a potential therapeutic target in LUAD.

Single-cell analysis reveals differential expression of key genes across cell types in LUAD

Figure S6A presents a UMAP plot depicting the clustering of 32 distinct cell populations in LUAD based on gene expression profiles, highlighting the cellular heterogeneity of the tumor microenvironment. Figure S6B annotates the identified clusters, revealing major cell types, including T cells (T), B cells (B), Natural Killer cells (NK), epithelial cells (Epi), macrophages (Mac), monocytes (Mon), fibroblasts (Fib), dendritic cells (MDC), mast cells (Mast), plasma cells (Plasma), endothelial cells (Endo), and plasmacytoid dendritic cells. This classification underscores the diversity of immune and stromal cell populations within LUAD tumors. Figure S6C displays a dot plot illustrating the expression of key genes such as AURKB, CDK1, CDKN3, and FOSL1 across different cell types. EPCAM, KR77, and KRT19 are highly expressed in epithelial and proliferative immune cells, suggesting their roles in cell cycle regulation and tumor proliferation. Figure S6D presents violin plots showing expression levels of key genes, including AURKB, CDK1, CDKN3, CDT1, CHEK1, DLGAP5, EXO1, FOSL1, and KRT6A, across various cell

populations. These genes are predominantly expressed in epithelial cells (Epi), reinforcing their involvement in tumor growth and proliferation. This single-cell analysis highlights the heterogeneous expression of critical genes within different cell types in LUAD, providing insights into their potential roles in disease progression and tumor microenvironment interactions.

Role of NDNF in LUAD progression and its potential as a therapeutic target

We conducted a comprehensive analysis of 18 identified genes, discovering that 16 had been previously studied in LUAD, while NDNF and RNASE1 had not been reported in this context. Figure S7A–D demonstrates that NDNF expression is significantly reduced in LUAD samples compared to normal lung tissue. To confirm this result, we analyzed NDNF expression in normal lung cells and LUAD cells at both the gene and protein levels, revealing a marked downregulation of NDNF in LUAD cells.

To further explore the functional role of NDNF in LUAD, we constructed a LUAD cell model (A549) with overexpressed NDNF. qPCR results confirmed successful NDNF overexpression (Figure S8A–E). Using the CCK-8 assay, we observed a significant reduction in cell viability in NDNF overexpression cells, indicating that NDNF inhibits cell proliferation (Figure S8F). Additionally, flow cytometry and Western blot analyses demonstrated a substantial increase in apoptosis in A549 cells with NDNF overexpression (Figure S8G). Western blot results also showed elevated levels of cleaved caspase-3, a key apoptotic marker, in NDNF-overexpressing cells (Figure S8H and I).

These findings collectively indicate that NDNF plays a crucial role in LUAD progression. The downregulation of NDNF is associated with decreased cell viability and increased apoptosis, suggesting that NDNF may serve as a promising therapeutic target for the treatment of LUAD.

Discussion

This study identified 2 molecular LUAD subtypes through integrated analysis of mRNA, lncRNA, miRNA, DNA methylation, and mutations. Subtype C1 showed poorer survival and a proliferative, immune-cold profile, while C2 exhibited neuronal pathway activation and higher immune activity. Findings were validated across datasets and clustering methods. Neuronal differentiation factor emerged as a key gene, with tumor-suppressive effects in LUAD cells. These results reveal LUAD heterogeneity and suggest subtype-specific biomarkers and therapeutic targets.

Neuronal differentiation factor is a gene that has been implicated in various biological processes, particularly in neuronal differentiation and development.[26] In other disease studies, NDNF has been associated with functions related to kidney cancer,[26] but its role in LUAD has not been extensively explored until now. In this study, we discovered that NDNF expression is significantly reduced in LUAD samples compared to normal lung tissue. To further investigate its role, we created an A549 LUAD cell line model with overexpressed NDNF. This model demonstrated that NDNF overexpression inhibited cell viability and promoted apoptosis in LUAD cells. These findings indicate that NDNF plays a crucial role in the progression of LUAD, suggesting that NDNF could serve as a potential therapeutic target for LUAD treatment.

This study confirms the utility of multiomic analysis for identifying LUAD subtypes linked to immune profiles and prognosis, with validated robustness across datasets. However, clinical application is limited by sample quality, high costs, and data complexity. Standardized FFPE protocols, simplified pipelines, and key biomarker prioritization are needed to improve feasibility. Future research should validate these findings in trials and explore NDNF's therapeutic potential in LUAD.

Conclusion

This study utilized multiomic analysis to identify and validate 2 prognostic subtypes of LUAD, highlighting significant differences in patient survival and emphasizing the importance of subtype classification for personalized treatment strategies. The analysis of gene signatures, immune landscapes, and differential gene expression provided valuable insights into the molecular and immunological characteristics of LUAD. The role of NDNF in LUAD progression was elucidated, revealing its potential as a therapeutic target. Our findings underscore the need for further research to validate these results in clinical settings and explore the therapeutic implications, paving the way for improved prognostic assessments and personalized therapies in LUAD.

Author contributions

H.D. designed the study. Y.T. analyzed the data, participated in the data collection, and Y.G., S.X., S.J., and Z.W. prepared the manuscript. H.W. and Y.H. helped the analysis with constructive discussions. All authors critically revised the manuscript.

Funding

This work was funded by the National Natural Science Foundation of China (82373102 to H.Y.), the Natural Science Foundation of Shandong Province (ZR2022QH146 to T.Y.), the China Postdoctoral Science Foundation (2023MT44294 to T.Y.) and the National Natural Science Foundation of China (82403121 to T.Y.).

Conflicts of interest

None declared.

Data availability

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Supplementary material

Supplementary material is available at *The Oncologist* online.

References

1. Wang H, Chen F, Hu A, et al. Harmine loaded Au@MSNs@PEG@Asp6 nano-composites for treatment of spinal metastasis from lung adenocarcinoma by targeting ANXA9 in vivo experiment. *Transl Lung Cancer Res*. 2023;12:1062-1077. https://doi.org/10.21037/tlcr-23-191

2. Abdel-Maksoud MA, Hassan F, Mubarik U, et al. An in-silico approach leads to explore six genes as a molecular signatures of lung adenocarcinoma. *Am J Cancer Res*. 2023;13:727-757.
3. Ge X, Liu Z, Weng S, et al. Integrative pharmacogenomics revealed three subtypes with different immune landscapes and specific therapeutic responses in lung adenocarcinoma. *Comput Struct Biotechnol J*. 2022;20:3449-3460. https://doi.org/10.1016/j.csbj.2022.06.064
4. Gao AW, El Alam G, Lalou A, et al. Multi-omics analysis identifies essential regulators of mitochondrial stress response in two wild-type *C. elegans* strains. *iScience* 2022;25:103734. https://doi.org/10.1016/j.isci.2022.103734
5. Cao P, Li F, Xiao Y, et al. Identification and validation of 7-lncRNA signature of epigenetic disorders by comprehensive epigenetic analysis. *Dis Markers*. 2022;2022:5118444. https://doi.org/10.1155/2022/5118444
6. Avery CL, Howard AG, Ballou AF, et al. Strengthening causal inference in exposomics research: application of genetic data and methods. *Environ Health Perspect*. 2022;130:55001. https://doi.org/10.1289/EHP9098
7. Wu G, Ruben MD, Francey LJ, et al. A population-based gene expression signature of molecular clock phase from a single epidermal sample. *Genome Med*. 2020;12:73. https://doi.org/10.1186/s13073-020-00768-9
8. Koll FJ, Banek S, Kluth L, et al. Tumor-associated macrophages and Tregs influence and represent immune cell infiltration of muscle-invasive bladder cancer and predict prognosis. *J Transl Med*. 2023;21:124. https://doi.org/10.1186/s12967-023-03949-3
9. Ruan X, Ye Y, Cheng W, et al. Multi-omics integrative analysis of lung adenocarcinoma: an in silico profiling for precise medicine. *Front Med (Lausanne)* 2022;9:894338. https://doi.org/10.3389/fmed.2022.894338
10. Lv B, An Q, Zeng Q, et al. Single-cell RNA sequencing reveals regulatory mechanism for trophoblast cell-fate divergence in human peri-implantation conceptuses. *PLoS Biol*. 2019;17:e3000187. https://doi.org/10.1371/journal.pbio.3000187
11. Sharma S, Dincer C, Weidemüller P, Wright GJ, Petsalaki E. CENtools: an integrative platform to identify the contexts of essential genes. *Mol Syst Biol*. 2020;16:e9698. https://doi.org/10.15252/msb.20209698
12. Lu X, Meng J, Zhou Y, Jiang L, Yan F. MOVICS: an R package for multi-omics integration and visualization in cancer subtyping. *Bioinformatics*. 2021;36:5539-5541. https://doi.org/10.1093/bioinformatics/btaa1018
13. Lv J, Wang D, Li T. Autophagy-mediated expression clusters are involved in immunity regulation of coronary artery disease. *BMC Genom Data* 2022;23:24. https://doi.org/10.1186/s12863-022-01023-3
14. Parise CA, Caggiano V. The influence of socioeconomic status on racial/ethnic disparities among the ER/PR/HER2 breast cancer subtypes. *J Cancer Epidemiol*. 2015;2015:813456. https://doi.org/10.1155/2015/813456
15. Zhang Z, Huang L, Li J, Wang P. Bioinformatics analysis reveals immune prognostic markers for overall survival of colorectal cancer patients: a novel machine learning survival predictive system. *BMC Bioinf*. 2022;23:124. https://doi.org/10.1186/s12859-022-04657-3
16. Lee YA, Noon LA, Akat KM, et al. Autophagy is a gatekeeper of hepatic differentiation and carcinogenesis by controlling the degradation of Yap. *Nat Commun*. 2018;9:4962. https://doi.org/10.1038/s41467-018-07338-z
17. Galamb O, Spisák S, Sipos F, et al. Reversal of gene expression changes in the colorectal normal-adenoma pathway by NS398 selective COX2 inhibitor. *Br J Cancer*. 2010;102:765-773. https://doi.org/10.1038/sj.bjc.6605515
18. Lu J, Cao W, He Z, et al. A genomic instability-related long noncoding rna signature for predicting hepatocellular carcinoma prognosis. *J Oncol*. 2022;2022:3090523. https://doi.org/10.1155/2022/3090523
19. Yan Y, Cao X, Wang Z, et al. Pyroptosis-related patterns predict tumor immune landscape and immunotherapy response in bladder cancer. *Front Mol Biosci*. 2022;9:815290. https://doi.org/10.3389/fmolb.2022.815290
20. Zhou L, Ding L, Gong Y, et al. Identification of hub genes associated with the pathogenesis of diffuse large B-cell lymphoma subtype one characterized by host response via integrated bioinformatic analyses. *PeerJ*. 2020;8:e10269. https://doi.org/10.7717/peerj.10269
21. Zhao L, Wang W, Sedykh A, Zhu H. Experimental errors in QSAR modeling sets: what we can do and what we cannot do. *ACS Omega*. 2017;2:2805-2812. https://doi.org/10.1021/acsomega.7b00274
22. Bi G, Bian Y, Liang J, et al. Pan-cancer characterization of metabolism-related biomarkers identifies potential therapeutic targets. *J Transl Med*. 2021;19:219. https://doi.org/10.1186/s12967-021-02889-0
23. Ogawa Y, Corbo JC. Partitioning of gene expression among zebrafish photoreceptor subtypes. *Sci Rep*. 2021;11:17340. https://doi.org/10.1038/s41598-021-96837-z
24. Hossian A, et al. Nucleic acid delivery with α-tocopherol-polyethyleneimine-polyethylene glycol nanocarrier system. *Int J Nanomedicine*. 2020;15:6689-6703.
25. Pan J, Liu F, Xiao X, et al. METTL3 promotes colorectal carcinoma progression by regulating the m6A-CRB3-Hippo axis. *J Exp Clin Cancer Res*. 2022;41:19. https://doi.org/10.1186/s13046-021-02227-8
26. Xia L, Li S, Liu Y, et al. NDNF inhibits the migration and invasion of human renal cancer cells through epithelial-mesenchymal transition. *Oncol Lett*. 2019;17:2969-2975. https://doi.org/10.3892/ol.2019.9937