Original Article

Integrated multiomics machine learning and mediated Mendelian randomization investigate the molecular subtypes and prognosis lung squamous cell carcinoma

Zhanghao Huang[1,2,3#], Jing Li[1,2,3#], You Lang Zhou[3,4], Jiahai Shi[2,3]

[1]Graduate School, Medical School of Nantong University, Nantong University, Nantong, China; [2]Department of Thoracic Surgery, Affiliated Hospital of Nantong University, Nantong, China; [3]Nantong Key Laboratory of Translational Medicine in Cardiothoracic Diseases, and Research Institution of Translational Medicine in Cardiothoracic Diseases, Affiliated Hospital of Nantong University, Nantong, China; [4]Research Center of Clinical Medicine, Affiliated Hospital of Nantong University, Nantong, China

Contributions: (I) Conception and design: Z Huang, J Shi; (II) Administrative support: YL Zhou, J Shi; (III) Provision of study materials or patients: Z Huang, J Li; (IV) Collection and assembly of data: Z Huang, J Li; (V) Data analysis and interpretation: Z Huang, YL Zhou, J Shi; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Jiahai Shi, MD. Department of Thoracic Surgery, Affiliated Hospital of Nantong University, Nantong, China; Nantong Key Laboratory of Translational Medicine in Cardiothoracic Diseases, and Research Institution of Translational Medicine in Cardiothoracic Diseases, Affiliated Hospital of Nantong University, No. 20 Xisi Road, Nantong 226001, China. Email: sjh@ntu.edu.cn; You Lang Zhou, PhD. Research Center of Clinical Medicine, Affiliated Hospital of Nantong University, 20 Xisi Road, Nantong 226001, China; Nantong Key Laboratory of Translational Medicine in Cardiothoracic Diseases, and Research Institution of Translational Medicine in Cardiothoracic Diseases, Affiliated Hospital of Nantong University, Nantong, China. Email: zhouyoulang@ntu.edu.cn.

Background: Lung squamous cell carcinoma (LUSC) lacks specific early diagnostic markers. Given the critical role of 5'-Nucleotidase Ecto (NT5E) in immune evasion and therapy resistance of cancer cells and the involvement of Dual Specificity Phosphatase 4 (DUSP4) in tumor cell proliferation through inhibition of the ERK signaling pathway, incorporating NT5E and DUSP4 into the consensus machine learning signature (CMLS) system in this study holds significant potential for investigating the early diagnosis and immune microenvironment of LUSC. The objective of this study was to explore the prognostic targets of LUSC.

Methods: Employing integrated algorithms enhances the ability to identify molecular subtypes and key features from multiple perspectives. A combination of 10 clustering algorithms and multi-omics data from LUSC patients, merged with 10 machine learning algorithms, was used to analyze and identify high-resolution molecular subsets and develop a CMLS. Mediated Mendelian randomization (MR) was utilized to explore mediations between immune cells and metabolites associated with CMLS.

Results: Cluster 1 demonstrated elevated infiltration of immune and stromal components, indicating an immunosuppressive microenvironment predominantly driven by tumor-associated macrophages or other inhibitory cells. In contrast, Cluster 2 displayed a metabolism-driven phenotype associated with improved prognosis. Mediated MR provided further insights into the causal relationships among CMLS, macrophages, and metabolites in LUSC. Validation of the RAS-RAF-MEK-ERK signaling pathway in conjunction with CMLS reinforced the immune characteristics of CMLS.

Conclusions: The integration of CMLS with multi-omics offers a robust framework for predicting prognosis, elucidating the causal interactions between the immune microenvironment and metabolic reprogramming in LUSC, and identifying patient subgroups likely to benefit from immunotherapy.

Keywords: Multiomics; machine learning; mediated Mendelian randomization (mediated MR); immune cells; metabolic reprogramming

Submitted Sep 27, 2024. Accepted for publication Feb 19, 2025. Published online Mar 18, 2025.
doi: 10.21037/tlcr-24-891
View this article at: https://dx.doi.org/10.21037/tlcr-24-891

Introduction

Non-small cell lung cancer (NSCLC) represents the majority of lung tumors, with lung adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC) as the predominant subtypes. LUAD and LUSC exhibit distinct transcriptome profiles and cell regulation networks (1). LUSC is characterized by dense infiltration of inflammatory monocytes (IMs) and low survival rates. Immune checkpoint inhibitors (ICIs) have shown promising responses in a subset of LUSC patients (2). Early surgical intervention can yield favorable outcomes, particularly when tumors are localized and non-disseminated. In advanced stages, treatment options are limited to radiotherapy (RT), chemotherapy, and pharmacotherapy, which primarily aim to delay disease progression and enhance survival (3). Advances in immunotherapy have significantly improved cancer treatment outcomes; however, due to the low mutation burden in LUSC, driver-gene-targeted therapies remain largely ineffective, benefiting only about 20% of patients (4). Radiation therapy remains a cornerstone treatment for LUSC, but its efficacy is often hindered by radiation resistance (5). Notably, RT can induce immunogenic tumor cell death, eliciting a systemic anti-tumor immune response (6). Identifying prognostic markers for LUSC and exploring their roles in the immune microenvironment could further enhance treatment strategies.

DUSP4 interacts with pathways associated with lymphocytes, including T cell and cytokine signaling, while NT5E is predominantly involved in tumor-associated macrophage pathways, promoting an M2 immunosuppressive microenvironment. This interplay between immunity and metabolism highlights their potential as prognostic markers in LUSC. Loss of DUSP4 is associated with increased resistance to radiation and targeted therapies, whereas NT5E, along with epidermal growth factor receptor (EGFR) and KRAS mutations, accelerates tumor development and progression. The combined analysis of DUSP4 and NT5E provides a comprehensive understanding of tumor immune escape, sensitivity to chemoradiotherapy, and metabolic reprogramming in LUSC.

This study analyzed the prognostic genes DUSP4 and NT5E in relation to LUSC by integrating multi-omics data. Among these, NT5E has been identified as an independent prognostic factor for LUSC. Multiple studies have established a strong association between NT5E and immunotherapy, RT, and EGFR mutations across various cancers, underscoring the reliability of this prognostic signature. The protein kinase structure of ERK1/2 regulates cellular activities through substrate phosphorylation. Extracellular signals such as cytokines, hormones, and cellular stress are transmitted into cells via transmembrane receptors and the RAF-MEK-ERK signaling pathway. DUSP4 can deactivate ERK1/2, thereby suppressing

Highlight box

Key findings

- Multi-omics consensus clustering successfully identified the molecular subtypes of lung squamous cell carcinoma (LUSC). By integrating machine learning algorithms, it facilitated the construction of consensus machine learning signature (CMLS), elucidated the role of CMLS in macrophage-mediated immune responses, and established a causal relationship between metabolites and LUSC.

What is known and what is new?

- Previous studies have found that NT5E plays a crucial role in immune evasion and drug resistance in cancer cells, whereas DUSP4 is involved in tumor cell proliferation by inhibiting the ERK signaling pathway.
- In the present research, the integration of CMLS with multi-omics offers a robust framework for predicting prognosis, elucidating the causal interactions between the immune microenvironment and metabolic reprogramming in LUSC, and identifying patient subgroups likely to benefit from immunotherapy.

What is the implication, and what should change now?

- Modulating tumor cell metabolic pathways have the potential to enhance immunotherapy efficacy and improve anti-tumor immune responses.

the progression of LUSC. The glycine cleavage system, an essential contributor to serine/glycine metabolism, fuels single-carbon metabolism, which is critical for cell proliferation (7). Evidence indicates that monocytes play a significant role in the pathogenesis of LUSC, with their interactions partially explaining the relationship between LUSC and the tumor immune microenvironment.

Limiting exogenous serine and glycine availability reduces one-carbon metabolism in tumor cells, thereby inhibiting their growth and enhancing T cell-mediated cytotoxicity during immunotherapy. This approach increases tumor immunogenicity, facilitating recognition and clearance by the immune system. Metabolic interventions amplify immune cell responses to immunogenic cancers while simultaneously increasing cancer cell immunogenicity, broadening the therapeutic applications of immunotherapy (8). Tumor cells undergo metabolic reprogramming to adapt to their evolving microenvironment during growth and metastasis. This reprogramming allows evasion of immune surveillance, and the interplay between altered metabolism and immune responses provides opportunities to identify novel therapeutic targets. Modulating tumor cell metabolic pathways has the potential to enhance immunotherapy efficacy and improve anti-tumor immune responses. To address the interplay between LUSC, macrophages, and tumor metabolism while considering reverse causality, this study utilized the mediation Mendelian method to investigate the causal relationships among macrophages, metabolic reprogramming, and LUSC. We present this article in accordance with the STROBE-MR reporting checklist (available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-24-891/rc).

Methods

Multiomics data collation

Multiomics data were obtained from the The Cancer Genome Atlas (TCGA) LUSC cohort (https://portal.gdc.cancer.gov), encompassing transcriptome expression [messenger RNA (mRNA), long non-coding RNA (lncRNA), and microRNA (miRNA)], DNA methylation, somatic mutations, and clinical data. Validation and immunotherapy cohorts were derived from Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo), including datasets such as GEO Series 30219 (GSE30219), GSE50081, GSE73403 (merge cohort), GSE78220, GSE91061, and GSE135222. Additionally, the immunotherapy cohort IMvigor210 was accessed from http://research-pub.gene.com/IMvigor210CoreBiologies. To ensure data comparability across cohorts, Z-score normalization was applied to all datasets prior to analysis. Background correction, $\log_2$ transformation, and quantile normalization of data were performed using the limma package, and the boxplot function was used to visualize and assess the uniformity of sample expression abundance distributions.

Data integration and batch effect processing

Z-score unified standardization

Data from various platforms (e.g., high-throughput sequencing and microarray) were normalized using Z-scores, ensuring expression values from different sources were within the same numerical range. This approach minimized the effects of platform detection sensitivity, sequencing depth, and background impurities on subsequent analyses.

Data preprocessing

GEO microarray data underwent background correction, $\log_2$ transformation, and quantile normalization using the limma package. RNA-sequencing (RNA-seq) data from TCGA were $\log_2$-transformed and aligned to GEO data using the same Z-score standardization criteria. Uniformity of data distribution within each cohort was evaluated using boxplots and principal component analysis (PCA), with outlier samples excluded.

Multi-cohort balancing

For integration of GEO datasets (e.g., GSE30219, GSE50081) with TCGA data, preprocessing was conducted within each sub-cohort, followed by a secondary Z-score normalization across the combined datasets. This step aimed to minimize variability introduced by experimental platforms, batch effects, and other confounding factors. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Multiomics consensus analysis

The MOVICS package was employed for multi-omics integration and visualization in cancer classification. The "getElite" function of MOVICS was used to screen for genetic characteristics (9). Statistical metrics, including the Rand Index and adjusted mutual information (AMI), were applied to evaluate the similarity of cluster labels

generated by various tools. Visualization methods, such as heatmaps and scatter plots, were utilized to compare the clustering structure of samples post-classification. When the clustering of key samples identified by MOVICS aligned with results from other methods, the findings were considered consistent with "standard practice" in terms of overall trends.

The IOBR package facilitated a comprehensive analysis of signature estimation, tumor microenvironment (TME) deconvolution, and signature construction using multi-omics data. It enabled batch analysis of these signatures and their associations with clinical phenotypes, lncRNA spectra, genomic features, and signatures derived from single-cell RNA-seq (scRNA-seq) data in diverse cancer environments (10).

The GSE method was applied as a key endpoint in bioinformatics analysis, while gene set variation analysis (GSVA) served as the starting point for constructing pathway-centered biological models (11). The single-sample gene set enrichment analysis (ssGSEA) algorithm was used to quantify the enrichment of specific molecular features in samples. This approach aimed to identify and explore the distinct molecular characteristics and biological functions of these clusters (CSs).

Construction of multiple machine learning prognostic signature

Ten machine learning algorithms were employed to construct the consensus machine learning signature (CMLS), including Cox Boost, Stepwise Cox, least absolute shrinkage and selection operator (LASSO), Ridge, Elastic Network (Enet), survival support vector machine (survival SVM), generalized boosted regression models (GBMs), supervised principal component analysis (SuperPC), Partial Least Cox (plsRcox), and random survival forest (RSF). The optimal penalty value in the CoxBoost model was determined using the "optimal CoxBoost penalty" function. A 10-fold cross-validation approach was applied to refine the model by determining the optimal number of boosting steps, and the "CoxBoost" function was used for model fitting. Stepwise Cox analysis was conducted using the survival package, with Akaike information criteria (AIC) applied to assess the statistical complexity of the signature. The LASSO, Ridge, and Enet models were implemented using the glmnet package, with regularization parameter lambda determined through 10-fold cross-validation, and the trade-off parameter alpha set at intervals between 0 and 1 (interval =0.1).

Molecular characterization of immunotherapy

CMLS was compared against six previously published LUSC-related signatures, which focused on TME cell types, immunotherapy response, immunosuppression, and immune rejection. Drug sensitivity data from the Cancer Therapeutics Response Portal (CTRP), comprising 481 compounds tested on 835 cancer cell lines (CCLs), and Parameter-elevation Regressions on Independent Slopes Model (PRISM), which includes sensitivity data for 1,448 compounds tested on 482 CCLs, were integrated to identify associations between genetic, lineage, and cellular characteristics with small molecule sensitivity (12). Immunotherapy responses were evaluated using the Tumor ImmunoPhenotype (TIP) algorithm, subclass mapping, and the tumor immune dysfunction and exclusion (TIDE) algorithm, analyzing response and delayed survival rates. Differential drug sensitivity analysis was performed to identify drugs specific to CMLS in LUSC. Binding target predictions for identified compounds were further explored through molecular docking.

Mendelian randomization (MR) analysis

Unidentified pooled data were acquired from publicly available genome-wide association studies (GWAS), all of which received prior ethical approval. Summary statistical data from GWAS relevant to the traits analyzed in this study include immune cells (GCST0001391-GCST0002121) and metabolites (GCST90199621-GCST90204063). The original GWAS for immune traits involved data from 3,757 European individuals from non-overlapping cohorts, while metabolite data comprised 1,091 metabolites and 309 metabolite ratios derived from 8,299 participants in the Canadian Longitudinal Study on Aging (CLSA) cohort. Data on LUSC (NCase =365, NControl =218,427, finn-b-C3-NCLC_SQUAM) were used to investigate the relationship between the immune microenvironment and metabolic reprogramming.

Single nucleotide polymorphisms (SNPs) exhibiting linkage disequilibrium were excluded, and weak instrumental variables were filtered using an *F*-test, applying a threshold of F-value >10. The inverse variance weighted (IVW) method was applied under the assumption that all SNPs function as valid instrumental variables and remain independent. The MR Egger method, which does not constrain the regression line to pass through the origin, was employed to provide a broader assessment of heterogeneity.

Criteria for SNP selection included $P<5\times10^{-8}$, kb=10,000, and $r^2=0.001$ (13). SNPs significantly associated with immune cells and metabolites were selected under the thresholds $P<1\times10^{-5}$, kb=10,000, and $r^2=0.001$. Forward MR analysis ensured positive results (P<0.05) for both exposure and outcome factors, while reverse MR analysis confirmed the absence of positive associations (P>0.05) between the factors. To increase the number of tools available for each cytokine, SNPs with a threshold P value of 5×10^{-6} were selected. A network MR analysis was conducted to explore the potential mediating role of immune cells in the relationship between metabolites and LUSC. By employing multiple methods, such as IVW and MR Egger, to analyze the same dataset, bias resulting from gene pleiotropy was minimized when results from these methods demonstrated consistent direction and significance.

Three estimates were performed for network MR analysis: (I) the total effect of immune cells on LUSC; (II) the direct effect of immune cells on metabolites (denoted as α); and (III) the direct effect of metabolites on LUSC (denoted as β). The mediation effect was calculated using the formula: indirect effect =α×β. The proportion of mediating effects was determined as the total causal effect of immune cells on LUSC divided by the mediating effect (14,15). The MR-Egger and weighted median methods were implemented using the TwoSampleMR package within R Studio. Statistical significance was determined when the P value from the IVW method was less than 0.05, indicating robust causal inferences.

Statistical analysis

The cutoff value of the CMLS score was established using the "surv-cutpoint" function from the survminer package in R. This method determines the optimal cutoff point by maximizing the log-rank statistic for survival analysis, enabling the division of patients into high and low CMLS groups. For comparison between groups, unpaired Student's *t*-test was employed for normally distributed variables, while the Wilcoxon rank-sum test was applied to non-normally distributed variables.

Results

Prognostic consensus molecular subtypes for 5 LUSC omics data

Molecular subtypes were identified using clustering prediction indices, gap statistical analysis, and contour scores derived from multi-component ensemble clustering algorithms. When the clustering result was set to two groups, significant differences were observed in the unique molecular expression patterns of the transcriptome (*Figure 1A-1C*). Clinical data analysis further demonstrated that survival outcomes differed significantly between the clusters, with CS2 exhibiting superior survival rates (*Figure 1D*). Validation in the merge cohort produced consistent findings (*Figure 1E*), confirming the robustness and reliability of the multi-cluster classification approach.

Analysis of prognostic features of LUSC patients

The ssGSEA analysis revealed distinct pathway enrichment patterns between molecular subtypes. CS2 demonstrated enrichment in EGFR ligands, cell cycle, and DNA replication pathways, while CS1 was enriched in pathways related to myofibroblasts, immune differentiation, interferon response, and EMT differentiation (*Figure 2A*) (16). Differential regulatory patterns among CSs are associated with regulatory activity profiles linked to carcinogenic chromatin remodeling, with epigenetically driven transcriptional networks serving as key differentiators of molecular subtypes. Variations in the transcriptional origins of CSs between potential regulators involved in cancer chromatin remodeling and 23 transcription factors (TFs) in LUSC underscore the significance of clustering (*Figure 2B*). The extent of immune infiltration further highlights the impact of distinct CSs on the progression and development of LUSC within the immune microenvironment. CS1 exhibited higher immune, matrix, and MeTIL scores, whereas CS2 displayed lower levels of immune infiltration (*Figure 2C*). CS2 may exhibit a more favorable tumor microenvironment, characterized by higher infiltration of beneficial immune cells (e.g., cytotoxic T cells) and lower levels of immunosuppressive cells, compared to CS1. The presence of specific cytokines or chemokines that enhance anti-tumor immunity appears to be more pronounced in CS2. Univariate Cox regression was applied to identify genes significantly associated with OS, leading to the development of CMLS using 99 algorithms. Among these, the algorithm combining LASSO and RSF demonstrated the highest average C-index and was selected to construct the final signature (*Figure 2D*). The consistency of CSs was assessed using the NTP and partition around medoids (PAM) algorithms (*Figure 2E-2G*). To further validate the predictive performance of CMLS, its efficacy was compared

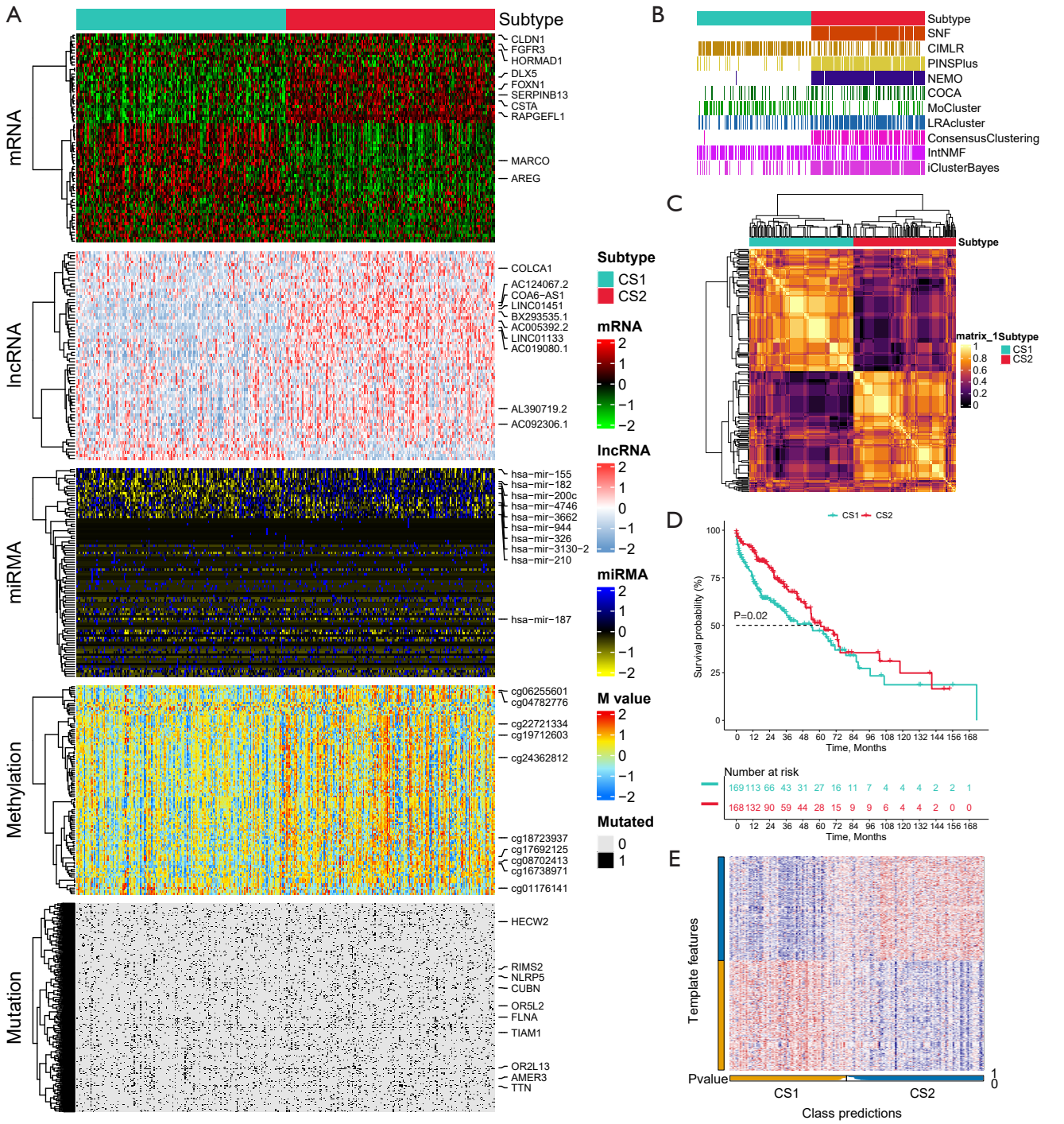


Figure 1 LUSC's multi-omics integration consensus subtype. (A) Comprehensive heatmaps of consensus set subtypes, including mRNA, lncRNA, miRNA, DNA CpG methylation sites, and mutated genes. (B) Ten multi-omics clustering methods were used to cluster LUSC patients. (C) A consistent cluster matrix for two prognostic subtypes in the TCGA cohort. (D) Survival analysis of two subtypes. Cluster 2 survives better. (E) A consistent cluster matrix for two prognostic subtypes in the merge cohort. CS1, cluster 1; CS2, cluster 2; lncRNA, long non-coding RNA; LUSC, lung squamous cell carcinomas; miRNA, microRNA; mRNA, messenger RNA; TCGA, The Cancer Genome Atlas.

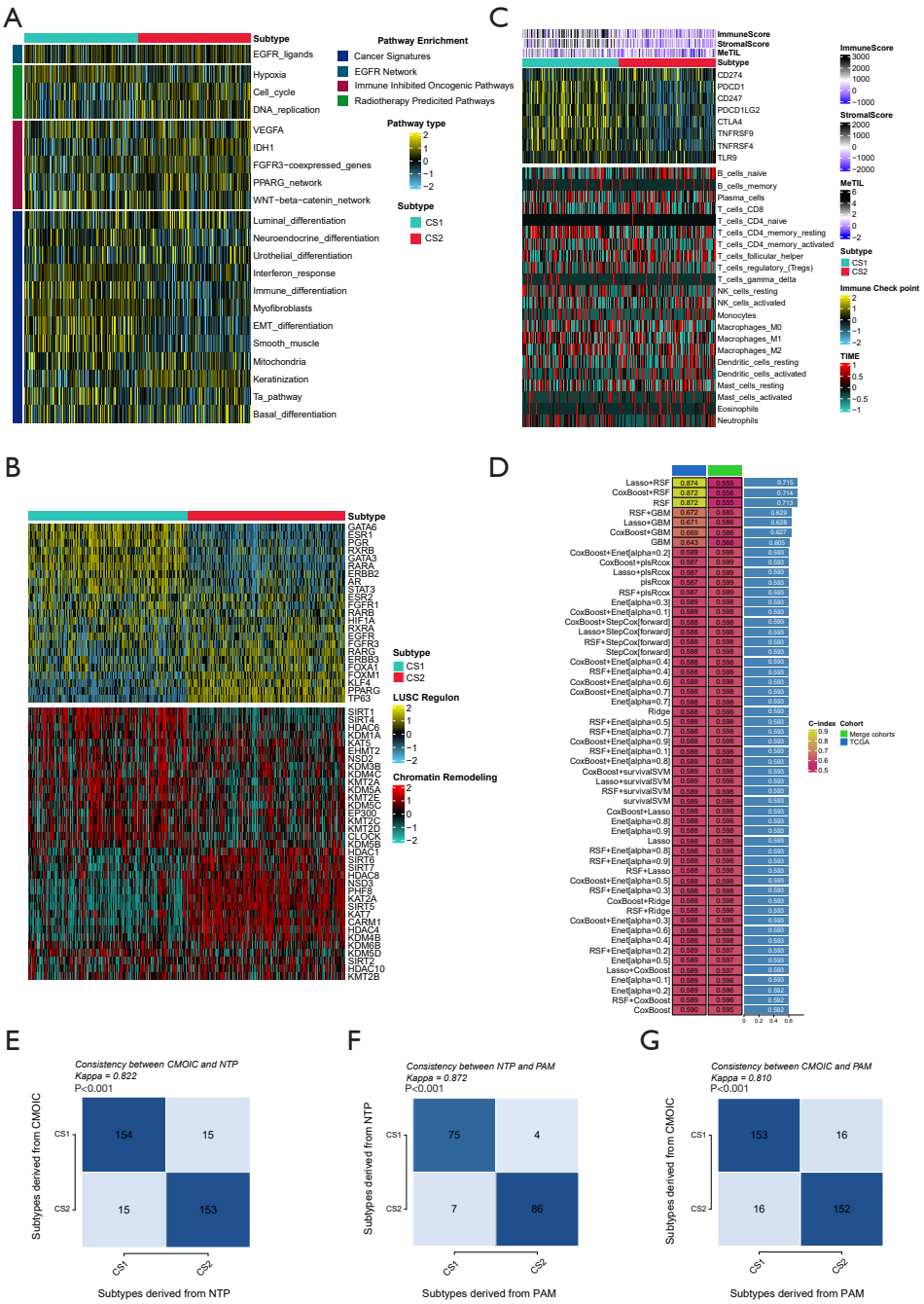


Figure 2 Machine learning. (A) The two clusters are enriched in multiple tumor pathways. (B) Association of two clusters with regulatory activity profiles of 23 TFs (top) and potential regulators associated with chromatin remodeling (bottom). (C) Immune profiles in the TCGA-LUSC cohort. Containing immune scores, stromal scores and MeTIL of lymphocytes, expression of immune checkpoint genes, and 22 TME-associated immune cells. (D) A combination of 99 machine learning algorithms calculates the C-index of each signature and sorts it by the average C-index of the validation set. (E) Consistency between CSs and NTP in the TCGA-LUSC queue. (F) Consistency between PAMs and NTP. (G) Consistency between PAMs and CSs. CS, cluster; CS1, cluster 1; CS2, cluster 2; EGFR, epidermal growth factor receptor; EMT, epithelial-mesenchymal transition; GBM, generalized boosted regression model; LUSC, lung squamous cell carcinoma; NTP, Nearest Template Prediction; PAM, partition around medoids; PPARG, peroxisome proliferator-activated receptor gamma; RSF, random survival forest; TCGA, The Cancer Genome Atlas; TF, transcription factor; TIME, tumor immune microenvironment; TME, tumor micro-environment; VEGFA, vascular endothelial growth factor A.

with that of six previously published prognostic signatures for LUSC (*Figure 3A-3D*) (17-22). The signature comprises two genes, DUSP4 and NT5E, both of which exhibit high expression levels in tumor tissues of pancarcinoma (Figure S1A,S1B). The signature genes also demonstrated variability in microsatellite instability and TMB across multiple tumors (Figure S1C-S1F). Prognostic analysis indicated that the high-CMLS group was associated with poorer outcomes, whereas the low-CMLS group exhibited a more favorable prognosis (*Figure 3E,3F*). Among the two genes, NT5E was identified as an independent prognostic factor for LUSC, in contrast to DUSP4 (*Figure 3G,3H*; *Table 1*). GSEA hallmark analysis revealed that DUSP4 in LUSC was predominantly enriched in KRAS signaling down and the P53 pathway, while NT5E was mainly associated with E2F targets and the G2M checkpoint (*Figure 3I,3J*) (23).

Multiomics analysis of signature genes

The ceRNA network analysis identified miRNAs and lncRNAs associated with the signature genes DUSP4 and NT5E (*Figure 4A*). The methylation site cg27130302 showed a negative correlation with DUSP4, where higher cg27130302 expression was associated with improved survival (*Figure 4B*). Among miRNAs, miR-374a-5p, miR-378a-3p, and miR-93-5p were negatively correlated with NT5E, with higher miRNA expression linked to better survival. Similarly, miR-149-5p exhibited a negative correlation with DUSP4, and higher expression of miR-149-5p was also associated with better survival outcomes (*Figure 4C*).

Further analysis revealed correlations between lncRNAs and miRNAs. AC092756.1 was negatively correlated with miR-93-5p; TENM3-AS1 was negatively correlated with miR-374a-5p; and MAGI2-AS3 was negatively correlated with miR-378a-3p. These lncRNAs were positively correlated with NT5E, and lower expression levels of these lncRNAs were associated with better survival. In contrast, AC037198.1 and MIR29B2CHG were negatively correlated with miR-149-5p and positively correlated with DUSP4. For these lncRNAs, lower expression levels were associated with improved survival outcomes (*Figure 4D*) (24).

Immune properties of CMLS

Analysis of CMLS differences in the TME revealed significant variations in cell types, particularly macrophages, B cells, and NK cells (*Figure 5A,5B*). Using the CIBERSORT algorithm, NT5E, a core gene of CMLS, was predominantly associated with macrophage-related cell types, while DUSP4 was primarily linked to T cell-related populations (Figure S2). The high-CMLS group showed significant enrichment in the TAM pathway (*Figure 5C,5D*). Conversely, the low-CMLS group exhibited higher scores in the mismatch repair pathway, suggesting enhanced immune activity (*Figure 5E,5F*). In terms of immunosuppression, the high-CMLS group demonstrated elevated scores in the fibroblast pathway, indicating a more suppressive TME (*Figure 5G,5H*).

Survival analysis highlighted the prognostic relevance of CMLS, particularly in conjunction with TMB, TNB and M1 macrophage expression. The low-CMLS, low-TMB, and low-TNB groups displayed improved survival outcomes. Moreover, within the low-CMLS group, high M1 macrophage expression correlated with a better prognosis, further supporting the observed immune infiltration patterns (*Figure 6A-6C*).

To evaluate the impact of CMLS in mucosal immunotherapy, restricted mean survival (RMS) at 6 and 12 months was compared to assess long-term survival following 3 months of immunotherapy. Patients with low-CMLS scores had superior outcomes, reflecting the delayed clinical benefits of immunotherapy (*Figure 6D,6E*). Additionally, CMLS significantly differentiated partial response (PR) from complete response (CR), stable disease (SD), and progressive disease (PD), further validating its clinical utility (*Figure 6F*) (25).

Mediated screening of immune cells and metabolites

Using the two-sample Mendelian randomization (TSMR) technique, the relationships between immune cells, metabolites, and LUSC were analyzed (*Figure 7A*). Significant immune cells and metabolites were identified for mediation effect analysis, with heterogeneity evaluated using the I^2 test statistic presented in tables available at https://cdn.amegroups.cn/static/public/tlcr-24-891-1.xlsx and https://cdn.amegroups.cn/static/public/tlcr-24-891-2.xlsx. Prognostic model results indicated a favorable prognosis for LUSC patients with low CMLS scores and high M1 macrophage expression. Mediation Mendelian analysis further established a causal relationship between these immune cells and LUSC.

CD80, a co-stimulatory molecule primarily expressed on antigen-presenting cells, plays a pivotal role in T cell activation and immune responses. Beyond immune

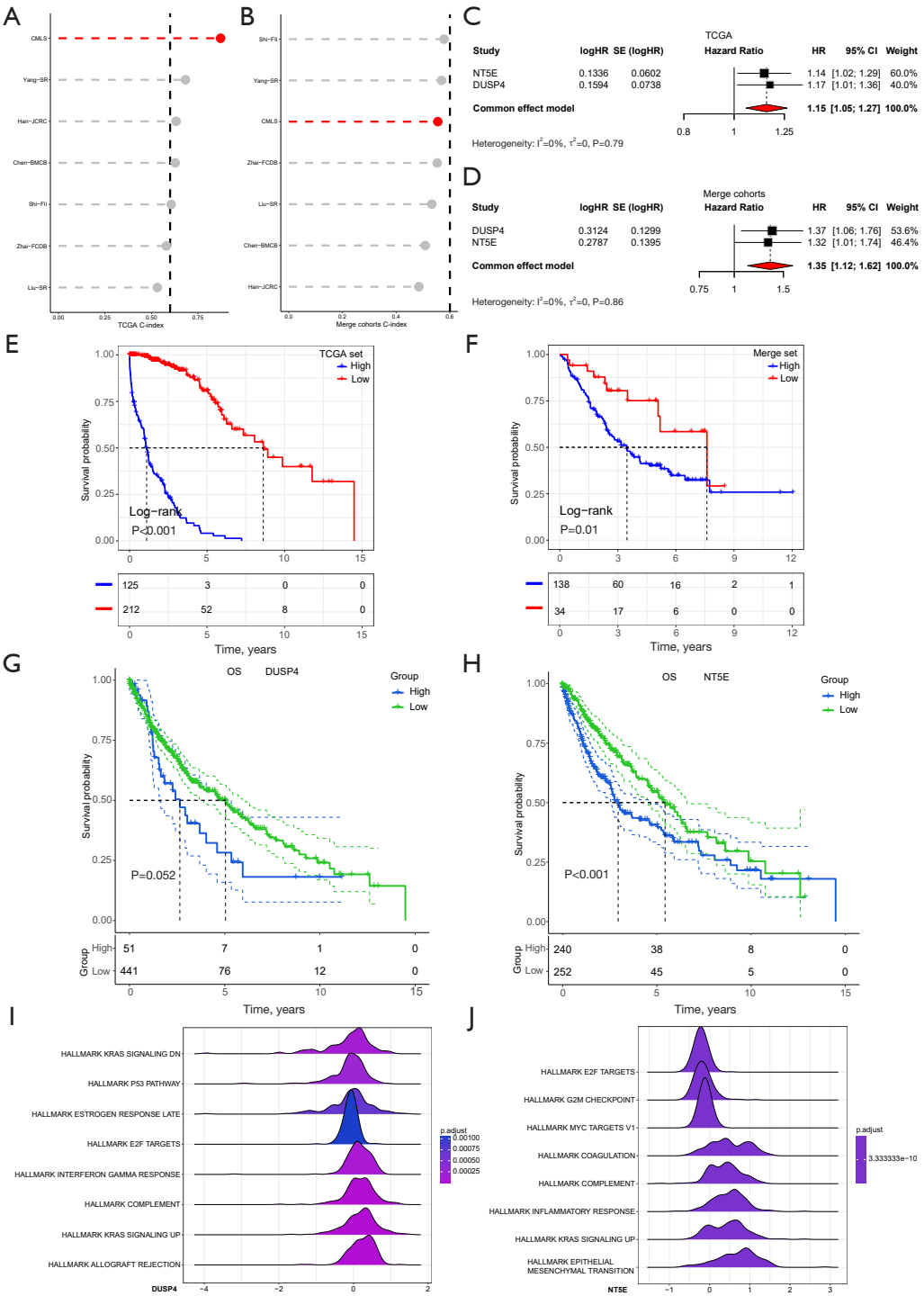
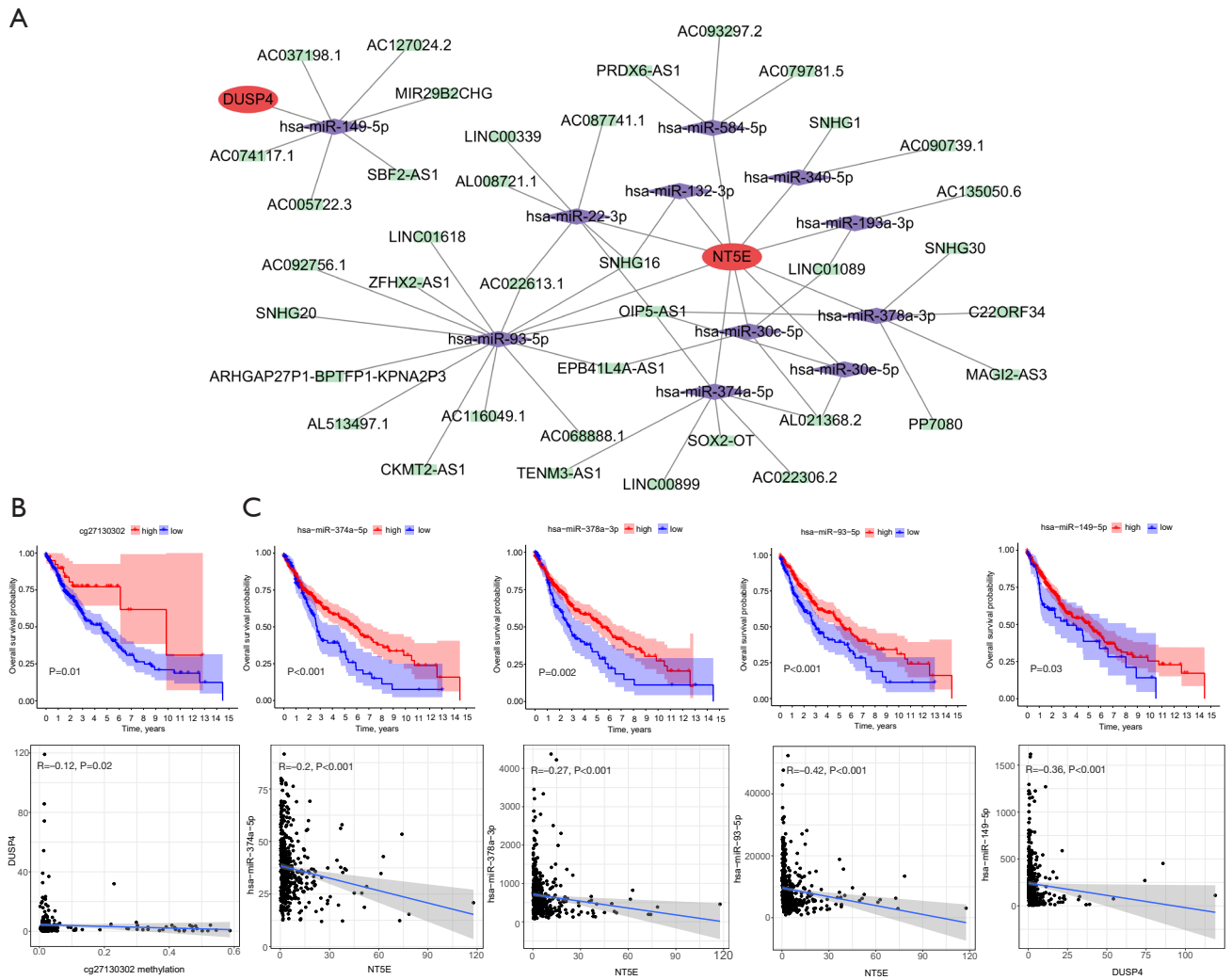


Figure 3 Construction of CMLS. (A) Comparison between CMLS and the other 6 published signatures in the TCGA cohort. (B) Comparison between CMLS and the other 6 published signatures in the merge cohort. (C) Risk coefficient of model gene in TCGA database. (D) Risk coefficient of model gene in the merge cohorts. (E) Survival analysis of CMLS in TCGA. (F) Survival analysis of CMLS in the merge cohorts. (G) Survival analysis of DUSP4 in TCGA. (H) Survival analysis of NT5E in TCGA. (I) Gene enrichment analysis of HALLMARK of DUSP4. (J) Gene enrichment analysis of HALLMARK of NT5E. CI, confidence interval; CMLS, consensus machine learning-driven signature; HR, hazard ratio; OS, overall survival; SE, standard error; TCGA, The Cancer Genome Atlas.

Table 1 Association of NT5E with lung cancer

Exposure	Outcome	Method	NSNP	Beta	SE	P value	Lo_CI	Up_CI	OR	OR_LCI95	OR_UCI95
NT5E	Lung cancer	MR Egger	13	0.051528307	0.09097767	0.58	−0.126787926	0.22984454	1.05287899	0.880920473	1.258404363
NT5E	Lung cancer	Weighted median	13	0.091586007	0.061200242	0.13	−0.028366467	0.211538482	1.095911029	0.972032083	1.235577512
NT5E	Lung cancer	Inverse variance weighted	13	0.106888652	0.052000348	0.04	0.004967969	0.208809334	1.112810338	1.00498033	1.232210036
NT5E	Lung cancer	Simple mode	13	0.114681242	0.102681121	0.29	−0.086573755	0.31593624	1.121515889	0.917067908	1.371542803
NT5E	Lung cancer	Weighted mode	13	0.080447467	0.062793404	0.22	−0.042627605	0.203522539	1.083771911	0.958268178	1.225712783

The relationship between NT5E and LUSC was analyzed by Mendelian randomization of two samples. CI, confidence interval; LCI, lower confidence interval; LO, lower bound; LUSC, lung squamous cell carcinoma; NSNP, number of single nucleotide polymorphism; OR, odds ratio; SE, standard error; UCI, upper confidence interval; UP, upper bound.



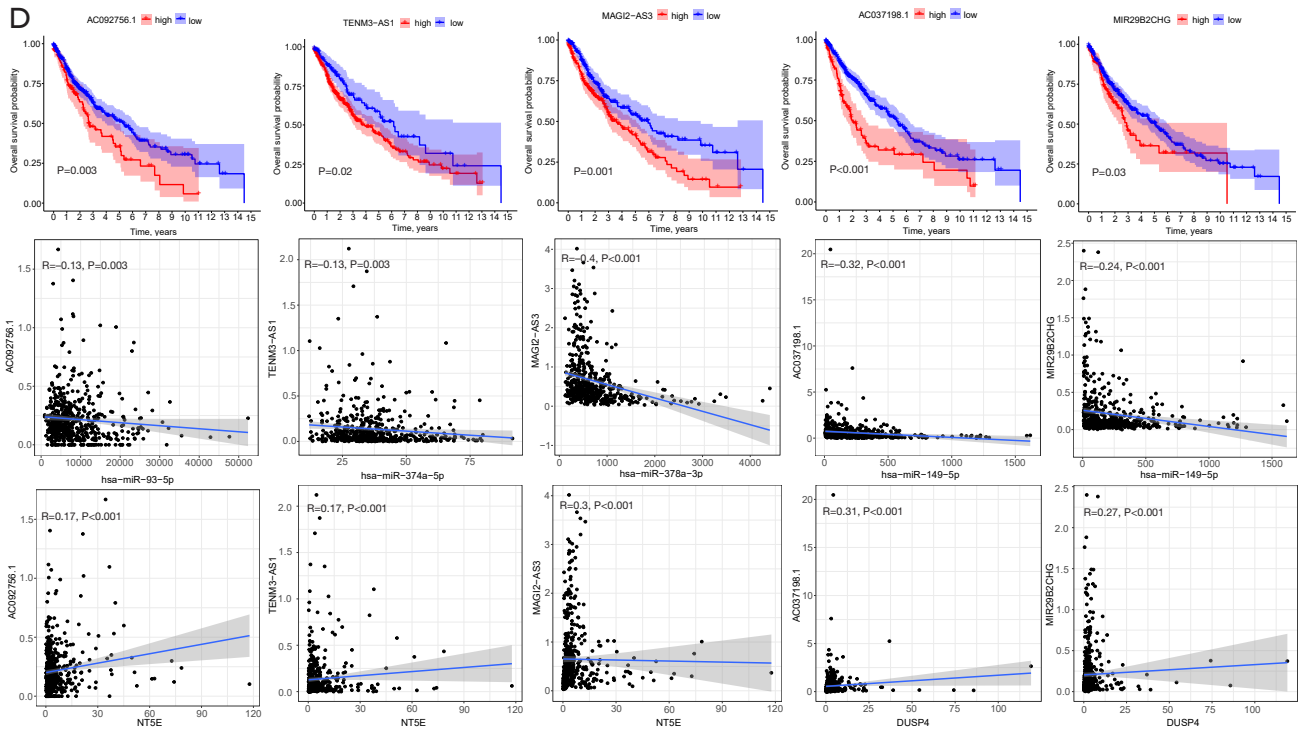


Figure 4 Multi-omics analysis of CMLS. (A) ceRNA network of genes in CMLS. (B) Survival and correlation analysis of methylation and DUSP4. (C) Survival and correlation analysis of miRNA and signature genes. (D) Survival and correlation analysis of lncRNA and signature genes. CMLS, consensus machine learning-driven signature; ceRNA, competing endogenous RNA; miRNA, microRNA; lncRNA, long non-coding RNA.

activation, CD80 has been implicated in promoting tumor cell growth by influencing tumor metabolism. Metabolic reprogramming in tumor cells facilitates their growth and adaptation to the microenvironment. Metabolites such as 3-hydroxycaproic acid and 3-aminoisobutyric acid, products of fatty acid and amino acid metabolism, are associated with tumor cell proliferation, survival, and transformation. CD80 on monocytes was found to modulate LUSC progression via metabolites like 3-hydroxyhexanoate, 3-aminoisobutyrate, and X-21353 (*Table 2*, tables available at https://cdn.amegroups.cn/static/public/tlcr-24-891-1.xlsx and https://cdn.amegroups.cn/static/public/tlcr-24-891-2.xlsx) (26,27).

These findings suggest that the CMLS signature impacts metabolic reprogramming through macrophage-mediated immune responses, influencing LUSC progression (*Figure 7B-7D*).

Drug therapy

Drugs closely associated with CMLS were identified, with seven being significantly linked to the Ras-Raf-MEK-ERK pathway, a critical component of the mitogen-activated protein kinase (MAPK) pathway. The Ras-Raf-MEK-ERK signaling cascade initiates with Ras transitioning from a GDP-bound to a GTP-bound state, activating Raf. Activated Raf phosphorylates serine residues on MEK1/2, which in turn phosphorylates threonine and tyrosine residues on ERK1/2. Activated ERK1/2 regulates nuclear TFs and cytoplasmic proteins, driving processes such as tumor proliferation, survival, and metabolic adaptation (*Figure 8A,8B*).

In highly activated states, the Ras-Raf-MEK-ERK pathway promotes tumor aggressiveness by modulating tumor immunogenicity. ERK activation can either enhance tumor-associated antigen expression or induce an immunosuppressive environment. Dysregulation of this pathway influences cytokine secretion and chemokine distribution in the tumor microenvironment, attracting immunosuppressive cells or reducing effector T cell infiltration and activation. Subtypes or CMLS phenotypes with minimal pathway dysregulation retain partial immunogenicity, potentially improving responsiveness to T

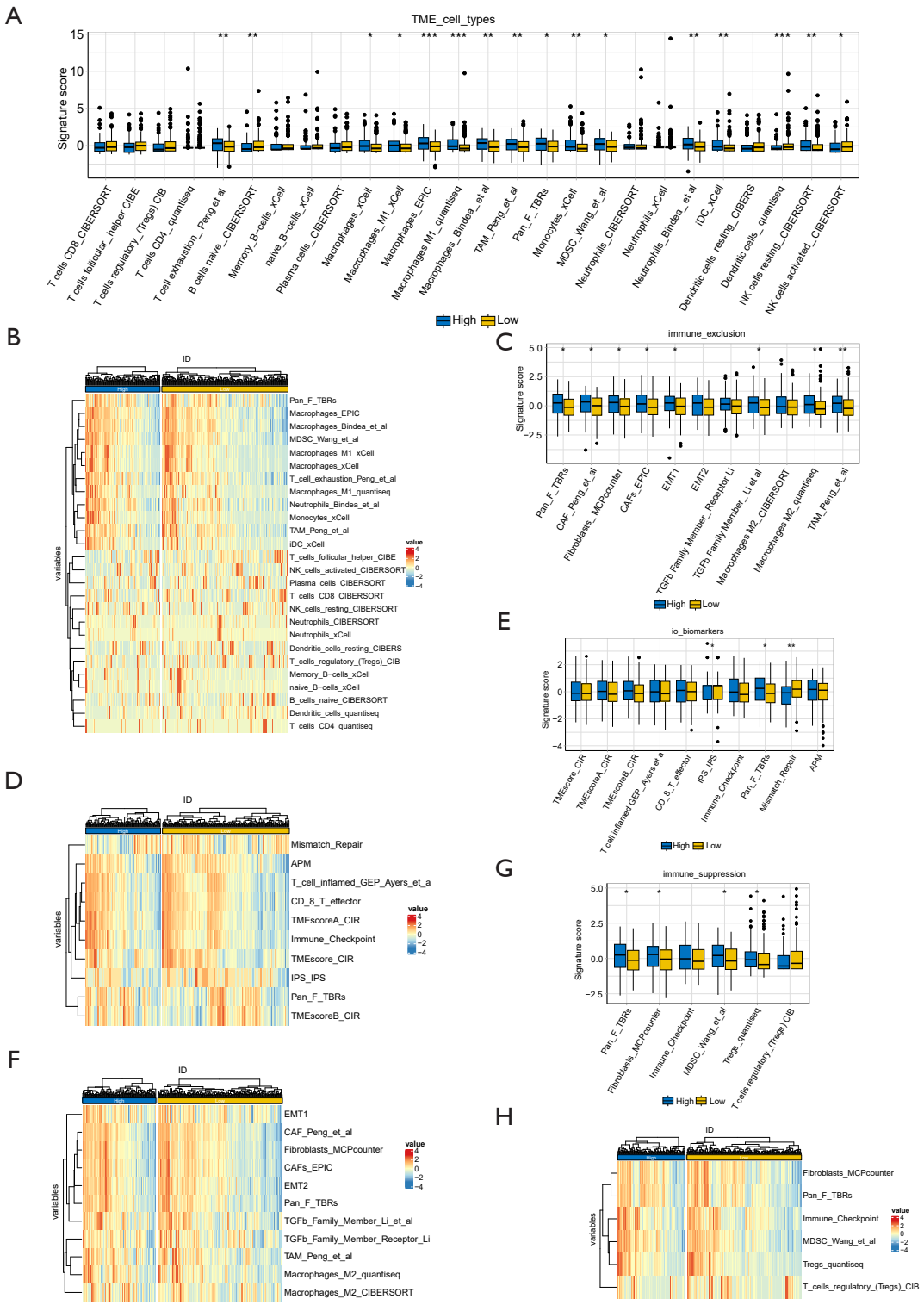


Figure 5 Immune properties of CMLS. (A,B) Distribution of TME immune cell type characteristics in CMLS patients. (C,D) Relationship between CMLS and immune exclusion. (E,F) Relationship between CMLS and biological targets. (G,H) Relationship between CMLS and immune suppression. *, P<0.05; **, P<0.01; ***, P<0.001. CMLS, consensus machine learning-driven signature; TME, tumor microenvironment.

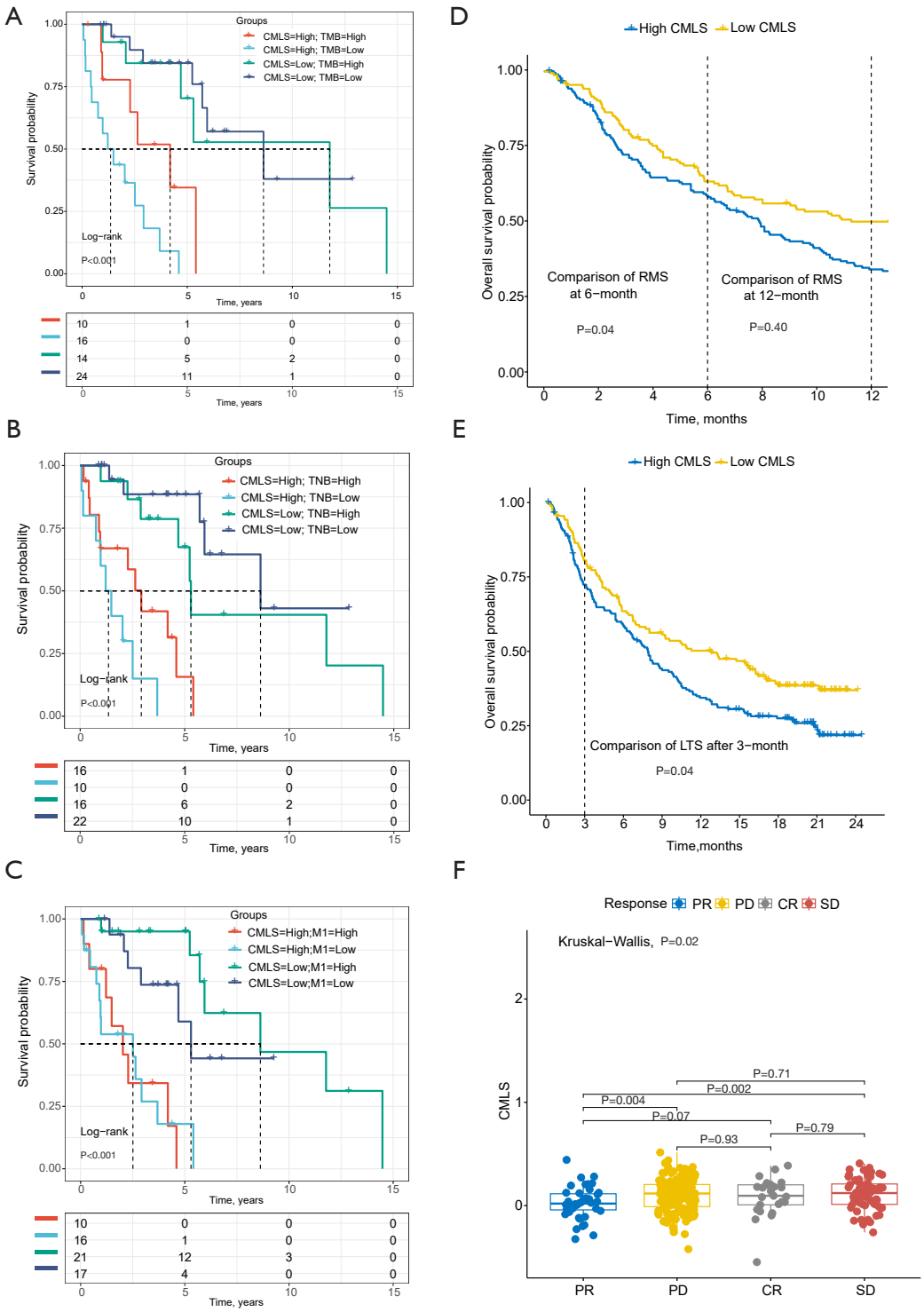
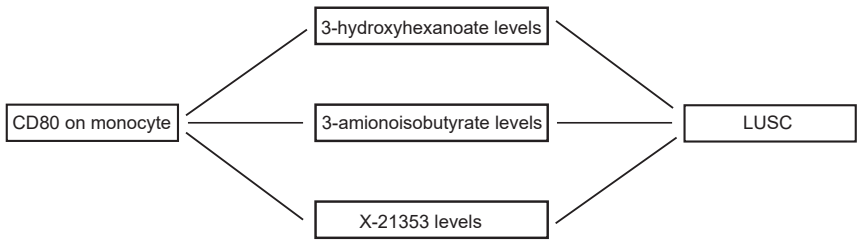


Figure 6 Immunotherapy characteristics of CMLS. (A-C) Survival analysis of CMLS binding TMB, TNB and M1 macrophages. (D-E) The RMS time difference between 6 months and 1 year after treatment for CMLS. Differences in LTS after 3 months of CMLS treatment. (F) Distribution of CMLS in different immunotherapy response groups. CMLS, consensus machine learning-driven signature; CR, complete response; LTS, long-term survival; PD, progressive disease; PR, partial response; RMS, restricted mean survival; SD, stable disease; TMB, tumor mutation burden.

A



B

exposure	outcome	nsnp	method	P	OR(95% CI)
CD80 on monocyte	GCST90200012	21	MR Egger	0.052	1.066 (1.003 to 1.133)
		21	Weighted median	0.07	1.047 (0.995 to 1.102)
		21	Inverse variance weighted	**0.006**	1.056 (1.016 to 1.098)
		21	Simple mode	0.29	1.042 (0.968 to 1.121)
		21	Weighted mode	0.17	1.048 (0.983 to 1.117)
3−hydroxyhexanoate levels	LUSC	17	MR Egger	0.44	1.659 (0.478 to 5.756)
		17	Weighted median	**0.04**	2.062 (1.040 to 4.085)
		17	Inverse variance weighted	**0.03**	1.772 (1.073 to 2.927)
		17	Simple mode	0.12	2.486 (0.846 to 7.300)
		17	Weighted mode	0.08	2.448 (0.968 to 6.190)
CD80 on monocyte	LUSC	21	MR Egger	0.26	0.848 (0.643 to 1.119)
		21	Weighted median	0.19	0.840 (0.649 to 1.087)
		21	Inverse variance weighted	**0.04**	0.828 (0.690 to 0.993)
		21	Simple mode	0.34	0.814 (0.538 to 1.231)
		21	Weighted mode	0.35	0.861 (0.633 to 1.170)

C

exposure	outcome	nsnp	method	P	OR(95% CI)
CD80 on monocyte	GCST90200310	21	MR Egger	0.15	1.052 (0.984 to 1.125)
		21	Weighted median	0.10	1.050 (0.991 to 1.112)
		21	Inverse variance weighted	**0.04**	1.045 (1.003 to 1.090)
		21	Simple mode	0.07	1.100 (0.996 to 1.215)
		21	Weighted mode	0.07	1.076 (0.998 to 1.161)
3−aminoisobutyrate levels	LUSC	30	MR Egger	0.47	0.903 (0.688 to 1.185)
		30	Weighted median	0.30	0.874 (0.679 to 1.125)
		30	Inverse variance weighted	**0.03**	0.793 (0.647 to 0.971)
		30	Simple mode	0.23	0.640 (0.316 to 1.296)
		30	Weighted mode	0.23	0.862 (0.679 to 1.093)
CD80 on monocyte	LUSC	21	MR Egger	0.26	0.848 (0.643 to 1.119)
		21	Weighted median	0.20	0.840 (0.645 to 1.094)
		21	Inverse variance weighted	**0.04**	0.828 (0.690 to 0.993)
		21	Simple mode	0.33	0.814 (0.545 to 1.218)
		21	Weighted mode	0.33	0.861 (0.641 to 1.156)

D

exposure	outcome	nsnp	method	P	OR(95% CI)
CD80 on monocyte	GCST90200567	21	MR Egger	**0.03**	1.076 (1.013 to 1.143)
		21	Weighted median	0.34	1.028 (0.971 to 1.089)
		21	Inverse variance weighted	**0.03**	1.045 (1.005 to 1.086)
		21	Simple mode	0.71	1.020 (0.920 to 1.130)
		21	Weighted mode	0.55	1.023 (0.950 to 1.103)
X−21353 levels	LUSC	14	MR Egger	**0.046**	1.225 (1.024 to 1.464)
		14	Weighted median	**0.04**	1.219 (1.014 to 1.465)
		14	Inverse variance weighted	**0.01**	1.217 (1.043 to 1.421)
		14	Simple mode	0.70	1.149 (0.584 to 2.259)
		14	Weighted mode	**0.05**	1.219 (1.021 to 1.456)
CD80 on monocyte	LUSC	21	MR Egger	0.26	0.848 (0.643 to 1.119)
		21	Weighted median	0.19	0.840 (0.647 to 1.091)
		21	Inverse variance weighted	**0.04**	0.828 (0.690 to 0.993)
		21	Simple mode	0.32	0.814 (0.549 to 1.207)
		21	Weighted mode	0.36	0.861 (0.631 to 1.175)

Figure 7 Mediated Mendelian randomization analysis. (A) Model diagram of mediation Mendelian randomization diagram. (B-D) CD80 on monocyte affects LUSC through 3-aminoisobutyrate, 3-hydroxyhexanoate levels and X-21353 levels. CI, confidence interval; LUSC, lung squamous cell carcinoma; MR, Mendelian randomization; NSNP, number of single nucleotide polymorphism; OR, odds ratio.

Table 2 Mediating effect

Immune cell	Metabolite	Outcome	Mediated effect	Mediated proportion	P
CD80 on monocyte	3-hydroxyhexanoate levels	LUSC	0.0313 (0.009, 0.0535)	−16.5% (−4.77%, −28.3%)	0.006
CD80 on monocyte	3-aminoisobutyrate levels	LUSC	−0.0103 (−0.0202, −0.000493)	5.48% (10.7%, 0.261%)	0.04
CD80 on monocyte	X-21353 levels	LUSC	0.00859 (0.000767, 0.0164)	−4.55% (−0.406%, −8.69%)	0.03

Intermediate effect values of CD80 on monocyte with 3-aminoisobutyrate, 3-hydroxyhexanoate levels and X-21353 levels. Data are presented as $\alpha \times \beta \pm 1.96 \times$ SE. LUSC, lung squamous cell carcinoma; SE, standard error.

cell infiltration and immunotherapy.

Molecular docking analysis identified NT5E binding sites within the seven drugs, revealing frequent binding at serine residues (*Figure 8C-8I*). Serine plays a pivotal role in the Ras-Raf-MEK-ERK pathway, particularly in cell signal transduction and metabolic regulation. This highlights serine as a key amino acid for therapeutic targeting in this pathway (*Table 3*) (28).

Discussion

Comprehensive multi-omics analysis provides valuable insights into the genetic and epigenetic mechanisms underlying LUSC. By integrating ten clustering algorithms, two prognostic subtypes were identified, facilitating precise stratification and guiding therapeutic strategies. The immunoinfiltration analysis revealed a causal relationship between CMLS, the immune microenvironment, and metabolic reprogramming, offering a deeper understanding of LUSC development mechanisms and potential treatment directions. Drug screening through CMLS highlighted the RAS-RAF-MEK-ERK signaling pathway as a critical axis in LUSC. The activation of this pathway influences macrophage polarization and is associated with disruptions in lipid metabolism, further corroborating the interplay between CMLS, macrophages, and tumor metabolism.

The functional roles of CMLS model genes, DUSP4 and NT5E, underscore their potential as therapeutic targets. DUSP4 acts as a negative regulator of the MAPK/ERK pathway, maintaining signal balance by dephosphorylating ERK, JNK, and p38. Reduced DUSP4 expression leads to sustained ERK activation, which is linked to greater tumor aggressiveness. Overactivation of MAPK signaling can also impact immune cell recruitment and programmed death-ligand 1 (PD-L1) expression. Restoring or enhancing DUSP4 function could suppress ERK-driven tumorigenic signals, improve immune cell infiltration, and enhance tumor sensitivity to checkpoint inhibitors.

NT5E plays a pivotal role in the extracellular conversion of AMP to adenosine. Elevated adenosine levels contribute to an immunosuppressive microenvironment by inhibiting T cell activation and fostering regulatory T cell expansion. NT5E-targeted therapies aim to block this adenosine axis, reducing immunosuppression and enhancing the efficacy of checkpoint inhibitors. In preclinical tumor studies, combining CD73 blockade with programmed death-1 (PD-1)/PD-L1 inhibitors has shown encouraging results, highlighting the potential of targeting the adenosine axis for enhanced immunotherapy. DUSP4, widely recognized as a biomarker in various tumors, has demonstrated specific roles in modulating MAPKs. In esophageal cancer, DUSP4 dephosphorylates HSP90β at T214 and Y216, stabilizing JAK1 and JAK2, which subsequently activate downstream p-STAT3 translocation into the nucleus (29). Additionally, DUSP4 and TRAF4 were identified as key downstream targets of the ALDOA-Oct4 axis, and their knockout significantly reduced cellular stemness (30). Overactivation of ERK signaling has revealed DUSP4 and DUSP6 as dual-gene synthetic lethal targets in NRAS and BRAF mutant melanoma (31).

While DUSP4 is not an independent prognostic factor in LUSC, it plays a crucial role in the negative feedback regulation of the EGFR signaling pathway (32). Nimbolide, a plant-derived compound, upregulates DUSP4 expression, thereby inhibiting ERK1/2 activation. This leads to reduced MMP-3 and Snail expression, increased E-cadherin levels, and suppression of NSCLC invasion and migration (33). DUSP4 is an important gene associated with lenvatinib resistance in hepatocellular carcinoma (HCC), and loss of DUSP4 results in lenvatinib resistance through reactivation of ERK and MEK in the treatment of lenvatinib in HCC (34). DUSP4 inhibits type I interferon and inflammatory cytokine expression during infection response via TBK1 and ERK1/2 pathways (35). It also serves as a negative regulator of sorafenib-induced ferroptosis, acting as a biomarker and mediator of drug resistance mechanisms (36). DUSP4

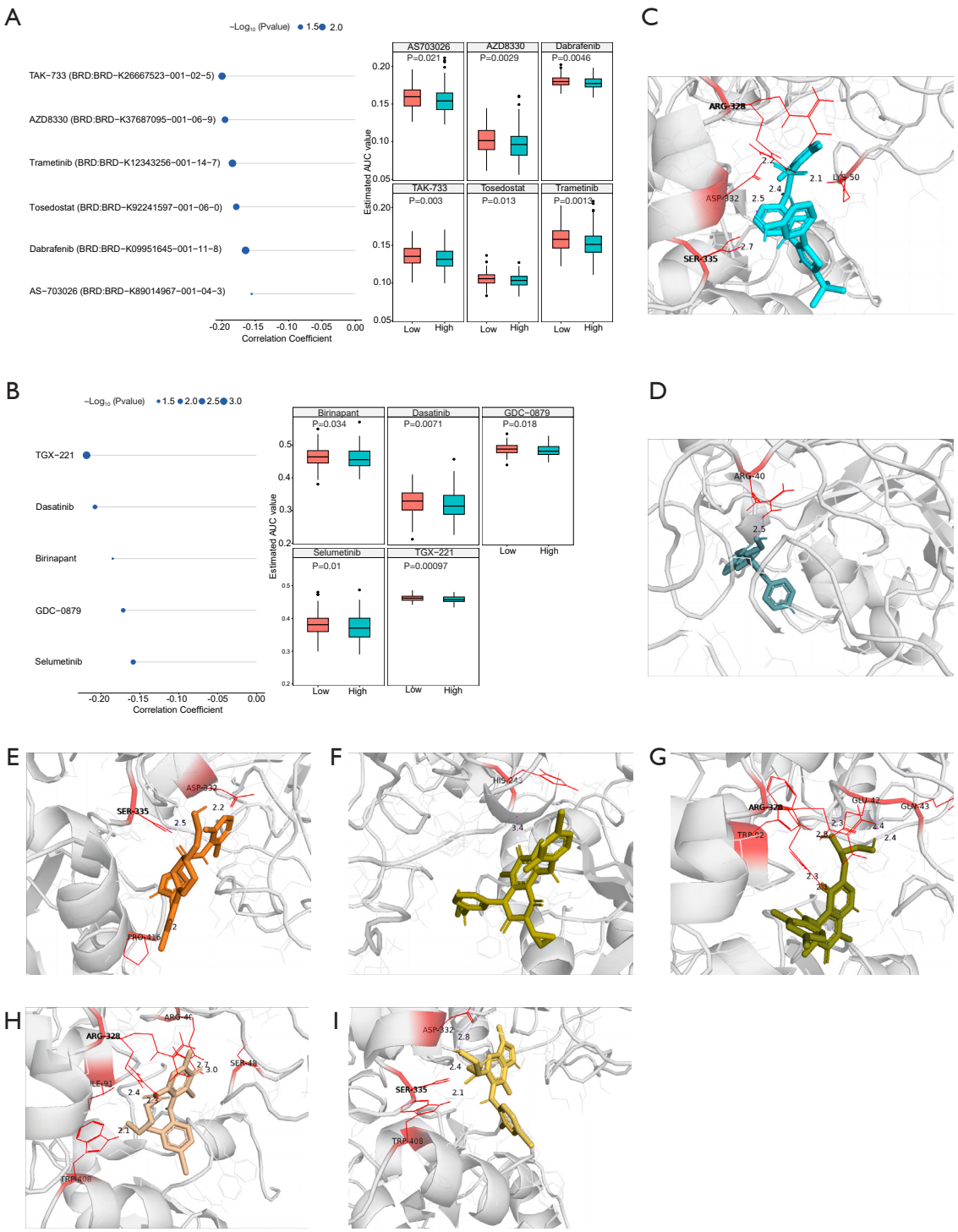


Figure 8 Drug properties. (A,B) Correlation and differential analysis of drug sensitivity of potential drugs screened from CTRP and PRISM datasets. (C-I) Seven drugs related to RS-RF-MEK-ERK pathway were screened for molecular docking with NT5E, and the docking site was more serine. AUC, area under the curve; CTRP, The Cancer Therapeutics Response Portal; PRISM, Parameter-elevation Regressions on Independent Slopes Model.

Table 3 Association of serine with LUSC

ID.exposure	ID.outcome	Outcome	Exposure	Method	NSNP	Bets	SE	P value	Lo_CI	Up_CI	OR	OR_LCI95	OR_UCI95
met-a-464	finn-b-C3_NSCLC_SQUAM_EXALLC	LUSC	Serine	MR Egger	3	−11.40249694	27.07514602	0.75	−64.46978315	41.66478927	1.11676E−05	1.0026E−28	1.24391E+18
met-a-464	finn-b-C3_NSCLC_SQUAM_EXALLC	LUSC	Serine	Weighted median	3	−6.60484146	2.664010015	0.01	−11.82630109	−1.383381829	0.001353798	7.30975E−06	0.250729194
met-a-464	finn-b-C3_NSCLC_SQUAM_EXALLC	LUSC	Serine	Inverse variance weighted	3	−6.344380767	2.409004794	0.008	−11.06603016	−1.622731371	0.00175659	1.56345E−05	0.197358902
met-a-464	finn-b-C3_NSCLC_SQUAM_EXALLC	LUSC	Serine	Simple mode	3	−7.172447544	3.167449762	0.15	−13.38064908	−0.96424601	0.000767442	1.54475E−06	0.381270566
met-a-464	finn-b-C3_NSCLC_SQUAM_EXALLC	LUSC	Serine	Weighted mode	3	−7.092906439	3.244590478	0.16	−13.45230378	−0.733509102	0.000830979	1.43793E−06	0.480220886
met-a-464	finn-b-C3_NSCLC_SQUAM	LUSC	Serine	MR Egger	3	−10.69119784	27.00983274	0.76	−63.63047001	42.24807433	2.27443E−05	2.3208E−28	2.22898E+18
met-a-464	finn-b-C3_NSCLC_SQUAM	LUSC	Serine	Weighted median	3	−6.590829689	2.716598463	0.02	−11.91536268	−1.266296701	0.0013729	6.68688E−06	0.281873553
met-a-464	finn-b-C3_NSCLC_SQUAM	LUSC	Serine	Inverse variance weighted	3	−6.372540746	2.403137151	0.008	−11.08268956	−1.662391929	0.001707815	1.53762E−05	0.189684725
met-a-464	finn-b-C3_NSCLC_SQUAM	LUSC	Serine	Simple mode	3	−7.128365793	3.278437706	0.16	−13.5541037	−0.70262789	0.000802029	1.29876E−06	0.495282046
met-a-464	finn-b-C3_NSCLC_SQUAM	LUSC	Serine	Weighted mode	3	−7.032398635	3.304649467	0.17	−13.50951159	−0.555285679	0.000882812	1.35798E−06	0.573908284

The relationship between serine and LUSC was analyzed by Mendelian randomization of two samples. CI, confidence interval; LCI, lower confidence interval; LO, lower bound; LUSC, lung squamous cell carcinoma; NSNP, number of single nucleotide polymorphism; OR, odds ratio; SE, standard error; UCI, upper confidence interval; UP, upper bound.

promotes megakaryocyte differentiation by inactivating p38 kinase, with PRMT1-mediated methylation facilitating DUSP4 degradation via E3 ligase HUWE1. This degradation is associated with transcriptional signatures linked to immune activation (37).

NT5E is a prognostic biomarker associated with cancer-associated fibroblasts (CAFs) and a unique tumor microenvironment across various cancers (38). In endothelial cells and CAFs, NT5E facilitates EMT and maintains CSC traits by upregulating SOX9 expression and stabilizing its protein. Targeting NT5E could eradicate CSCs and reverse lenvatinib resistance in HCC (39). As a critical enzyme in the purinergic signaling pathway, NT5E drives immunosuppressive adenosine production, a common resistance mechanism in cancer treatments (40). The expression of NT5E affects T cell receptor (TCR) diversity and T cell transcriptional profiles, suggesting their key role in T cell failure within tumors, and that AB680 (an NT5E inhibitor) improves the anticancer function of immunosuppressive cells such as Tregs and depleted T cells (41). NT5E has emerged as a potential novel biomarker in head and neck squamous cell carcinoma (HNSC), with its expression predominantly localized to CAFs (42). The adenosine pathway, where NT5E plays a pivotal role as a major producer of extracellular adenosine, represents an exciting therapeutic target in cancer immunotherapy. Elevated expression of NT5E has been linked to key oncogenic drivers of NSCLC, such as mutated EGFR and KRAS, further emphasizing its significance (43). Additionally, NT5E has shown promise as a therapeutic target for diclofenac, which promotes pancreatic cancer metastasis via a nucleotidase-independent mechanism (44). The NT5E/adenosine pathway contributes to tumor immune evasion, presenting a major obstacle to the effectiveness of ICIs and playing a critical role in therapeutic resistance (45). Combining the findings from multiple studies, DUSP4 and NT5E have been identified as critical players in LUSC, though their underlying mechanisms require further investigation. Tumor-associated metabolic alterations are increasingly recognized as emerging indicators of cancer progression, prompting the application of intermediary Mendelian analysis to better understand these dynamics. Metabolic changes within the tumor microenvironment drive cancer progression, with cancer cells exhibiting significant metabolic heterogeneity and plasticity (46). Studies have demonstrated that serine restriction can modulate sphingolipid diversity, effectively limiting tumor growth (47-49). Proteins such as SFXN1 and its homologues have been identified as crucial regulators of serine fate within cells. Many cancer cells depend on one-carbon units derived from serine metabolism for proliferation, with serine-related mitochondrial catabolic enzymes often upregulated in tumor cells (50). Although the predictive efficacy of CMLS has been discussed in multiple aspects, our study still has shortcomings. The variability in the size and sequencing platforms of the included cohorts represents a limitation and will be a focus of future research efforts. To strengthen the robustness and universal applicability of the Cancer Metabolism and Landscape Score, it is essential to validate CMLS across additional cohorts. Particular emphasis should be placed on testing in patients from diverse ethnic backgrounds, geographic locations, and clinical treatment regimens to ensure its broad applicability and relevance across varied clinical and demographic contexts.

Conclusions

Multi-omics consensus clustering successfully identified two molecular subtypes of LUSC and, through the integration of machine learning algorithms, facilitated the construction of the CMLS. The study elucidated the role of CMLS in macrophage-mediated immune responses and established a causal relationship between metabolites such as 3-hydroxyhexanoate, 3-aminoisobutyrate, and X-21353 levels and LUSC. The Ras-Raf-MEK-ERK pathway was used to validate this causal association, providing a mechanistic framework. These findings offer valuable insights for the early diagnosis and precision treatment of LUSC, paving the way for improved clinical outcomes.

Acknowledgments

None.

Footnote

Reporting Checklist: The authors have completed the STROBE-MR reporting checklist. Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-24-891/rc

Peer Review File: Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-24-891/prf

Funding: This work was supported by Postgraduate Research & Practice Program of Jiangsu Province (No. KYCX23_3427), National Natural Science Foundation

of China (No. 82370253), Jiangsu Provincial Research Hospital (No. YJXYY202204) and Innovation Team Project of Affiliated Hospital of Nantong University (No. XNBHCX31773).

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-24-891/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Relli V, Trerotola M, Guerra E, et al. Abandoning the Notion of Non-Small Cell Lung Cancer. Trends Mol Med 2019;25:585-94.
2. Porrello A, Leslie PL, Harrison EB, et al. Factor XIIIA-expressing inflammatory monocytes promote lung squamous cancer through fibrin cross-linking. Nat Commun 2018;9:1988.
3. Zhang P, Kang B, Xie G, et al. Genomic sequencing and editing revealed the GRM8 signaling pathway as potential therapeutic targets of squamous cell lung cancer. Cancer Lett 2019;442:53-67.
4. Yu T, Nie FQ, Zhang Q, et al. Effects of methionine deficiency on B7H3-DAP12-CAR-T cells in the treatment of lung squamous cell carcinoma. Cell Death Dis 2024;15:12.
5. Wen J, Zheng W, Zeng L, et al. LTF Induces Radioresistance by Promoting Autophagy and Forms an AMPK/SP2/NEAT1/miR-214-5p Feedback Loop in Lung Squamous Cell Carcinoma. Int J Biol Sci 2023;19:1509-27.
6. Chen Q, Chen J, Zhang Q, et al. Combining High-Z Sensitized Radiotherapy with CD73 Blockade to Boost Tumor Immunotherapy. ACS Nano 2023;17:12087-100.
7. Amelio I, Cutruzzolá F, Antonov A, et al. Serine and glycine metabolism in cancer. Trends Biochem Sci 2014;39:191-8.
8. Xia L, Oyang L, Lin J, et al. The cancer metabolic reprogramming and immune response. Mol Cancer 2021;20:28.
9. Lu X, Meng J, Zhou Y, et al. MOVICS: an R package for multi-omics integration and visualization in cancer subtyping. Bioinformatics 2021;36:5539-41.
10. Zeng D, Ye Z, Shen R, et al. IOBR: Multi-Omics Immuno-Oncology Biological Research to Decode Tumor Microenvironment and Signatures. Front Immunol 2021;12:687975.
11. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 2013;14:7.
12. Yang C, Huang X, Li Y, et al. Prognosis and personalized treatment prediction in TP53-mutant hepatocellular carcinoma: an in silico strategy towards precision oncology. Brief Bioinform 2021;22:bbaa164.
13. Zheng H, Shi YZ, Liang JT, et al. Modifiable factors for migraine prophylaxis: A mendelian randomization analysis. Front Pharmacol 2023;14:1010996.
14. Liu Y, Lai H, Zhang R, et al. Causal relationship between gastro-esophageal reflux disease and risk of lung cancer: insights from multivariable Mendelian randomization and mediation analysis. Int J Epidemiol 2023;52:1435-47.
15. Xu M, Zheng J, Hou T, et al. SGLT2 Inhibition, Choline Metabolites, and Cardiometabolic Diseases: A Mediation Mendelian Randomization Study. Diabetes Care 2022;45:2718-28.
16. Powers RK, Goodspeed A, Pielke-Lombardo H, et al. GSEA-InContext: identifying novel and common patterns in expression experiments. Bioinformatics 2018;34:i555-64.
17. Shi X, Dong A, Jia X, et al. Integrated analysis of single-cell and bulk RNA-sequencing identifies a signature based on T-cell marker genes to predict prognosis and therapeutic response in lung squamous cell carcinoma. Front Immunol 2022;13:992990.
18. Yang Q, Gong H, Liu J, et al. A 13-gene signature to predict the prognosis and immunotherapy responses of lung squamous cell carcinoma. Sci Rep 2022;12:13646.
19. Zhai WY, Duan FF, Chen S, et al. An Aging-Related Gene Signature-Based Model for Risk Stratification and

Prognosis Prediction in Lung Squamous Carcinoma. Front Cell Dev Biol 2022;10:770550.
20. Liu X, Zhao D, Shan Y, et al. Development and validation of a novel immune-related prognostic signature in lung squamous cell carcinoma patients. Sci Rep 2022;12:20737.
21. Chen L, Weng Y, Cui X, et al. Comprehensive analyses of a CD8(+) T cell infiltration related gene signature with regard to the prediction of prognosis and immunotherapy response in lung squamous cell carcinoma. BMC Bioinformatics 2023;24:238.
22. Han WJ, He P. A novel tumor microenvironment-related gene signature with immune features for prognosis of lung squamous cell carcinoma. J Cancer Res Clin Oncol 2023;149:13137-54.
23. Chu G, Ji X, Wang Y, et al. Integrated multiomics analysis and machine learning refine molecular subtypes and prognosis for muscle-invasive urothelial cancer. Mol Ther Nucleic Acids 2023;33:110-26.
24. Cao W, Zhou W, Li M, et al. A novel signature based on CeRNA and immune status predicts prognostic risk and drug sensitivity in gastric cancer patients. Front Immunol 2022;13:951135.
25. Tuninetti V, Virano E, Salutari V, et al. Real-life efficacy and safety of cemiplimab in advanced cervical cancer from a nominal use program in Italy: The MITO 44 study. Eur J Cancer 2024;203:114039.
26. Wang Q, Dai H, Hou T, et al. Dissecting Causal Relationships Between Gut Microbiota, Blood Metabolites, and Stroke: A Mendelian Randomization Study. J Stroke 2023;25:350-60.
27. Ji D, Chen WZ, Zhang L, et al. Gut microbiota, circulating cytokines and dementia: a Mendelian randomization study. J Neuroinflammation 2024;21:2.
28. Ham JM, Kim M, Kim T, et al. Structure-Based De Novo Design for the Discovery of Miniprotein Inhibitors Targeting Oncogenic Mutant BRAF. Int J Mol Sci 2024;25:5535.
29. Zhou L, Yao N, Yang L, et al. DUSP4 promotes esophageal squamous cell carcinoma progression by dephosphorylating HSP90β. Cell Rep 2023;42:112445.
30. Chang YC, Yang YF, Chiou J, et al. Nonenzymatic function of Aldolase A downregulates miR-145 to promote the Oct4/DUSP4/TRAF4 axis and the acquisition of lung cancer stemness. Cell Death Dis 2020;11:195.
31. Ito T, Young MJ, Li R, et al. Paralog knockout profiling identifies DUSP4 and DUSP6 as a digenic dependence in MAPK pathway-driven cancers. Nat Genet 2021;53:1664-72.
32. Chitale D, Gong Y, Taylor BS, et al. An integrated genomic analysis of lung cancer reveals loss of DUSP4 in EGFR-mutant tumors. Oncogene 2009;28:2773-83.
33. Lin H, Qiu S, Xie L, et al. Nimbolide suppresses non-small cell lung cancer cell invasion and migration via manipulation of DUSP4 expression and ERK1/2 signaling. Biomed Pharmacother 2017;92:340-6.
34. Huang S, Ma Z, Zhou Q, et al. Genome-Wide CRISPR/Cas9 Library Screening Identified that DUSP4 Deficiency Induces Lenvatinib Resistance in Hepatocellular Carcinoma. Int J Biol Sci 2022;18:4357-71.
35. Jiao H, James SJ, Png CW, et al. DUSP4 modulates RIG-I- and STING-mediated IRF3-type I IFN response. Cell Death Differ 2024;31:280-91.
36. Hao SH, Ma XD, Xu L, et al. Dual specific phosphatase 4 suppresses ferroptosis and enhances sorafenib resistance in hepatocellular carcinoma. Drug Resist Updat 2024;73:101052.
37. Su H, Jiang M, Senevirathne C, et al. Methylation of dual-specificity phosphatase 4 controls cell differentiation. Cell Rep 2021;36:109421.
38. Xue XM, Liu YY, Chen XM, et al. Pan-cancer analysis identifies NT5E as a novel prognostic biomarker on cancer-associated fibroblasts associated with unique tumor microenvironment. Front Pharmacol 2022;13:1064032.
39. Ma XL, Hu B, Tang WG, et al. CD73 sustained cancer-stem-cell traits by promoting SOX9 expression and stability in hepatocellular carcinoma. J Hematol Oncol 2020;13:11.
40. Bach N, Winzer R, Tolosa E, et al. The Clinical Significance of CD73 in Cancer. Int J Mol Sci 2023;24:11759.
41. Kim M, Min YK, Jang J, et al. Single-cell RNA sequencing reveals distinct cellular factors for response to immunotherapy targeting CD73 and PD-1 in colorectal cancer. J Immunother Cancer 2021;9:e002503.
42. Chen XM, Liu YY, Tao BY, et al. NT5E upregulation in head and neck squamous cell carcinoma: A novel biomarker on cancer-associated fibroblasts for predicting immunosuppressive tumor microenvironment. Front Immunol 2022;13:975847.
43. Kowash RR, Akbay EA. Tumor intrinsic and extrinsic functions of CD73 and the adenosine pathway in lung cancer. Front Immunol 2023;14:1130358.
44. Liu W, Yu X, Yuan Y, et al. CD73, a Promising Therapeutic Target of Diclofenac, Promotes Metastasis of Pancreatic Cancer through a Nucleotidase Independent Mechanism. Adv Sci (Weinh) 2023;10:e2206335.

45. Saigí M, Mesía-Carbonell O, Barbie DA, et al. Unraveling the Intricacies of CD73/Adenosine Signaling: The Pulmonary Immune and Stromal Microenvironment in Lung Cancer. Cancers (Basel) 2023;15:5706.
46. Schmidt DR, Patel R, Kirsch DG, et al. Metabolomics in cancer research and emerging applications in clinical oncology. CA Cancer J Clin 2021;71:333-58.
47. Muthusamy T, Cordes T, Handzlik MK, et al. Serine restriction alters sphingolipid diversity to constrain tumour growth. Nature 2020;586:790-5.
48. Ma S, Sandhoff R, Luo X, et al. Serine enrichment in tumors promotes regulatory T cell accumulation through sphinganine-mediated regulation of c-Fos. Sci Immunol 2024;9:eadg8817.
49. Qian L, Zhang F, Yin M, et al. Cancer metabolism and dietary interventions. Cancer Biol Med 2021;19:163-74.
50. Kory N, Wyant GA, Prakash G, et al. SFXN1 is a mitochondrial serine transporter required for one-carbon metabolism. Science 2018;362:eaat9528.

Cite this article as: Huang Z, Li J, Zhou YL, Shi J. Integrated multiomics machine learning and mediated Mendelian randomization investigate the molecular subtypes and prognosis lung squamous cell carcinoma. Transl Lung Cancer Res 2025;14(3):857-877. doi: 10.21037/tlcr-24-891