

ORIGINAL ARTICLE OPEN ACCESS

Integrating Single-Cell Transcriptomics and Machine Learning to Define an ac4C Gene Signature in Lung Adenocarcinoma

Yuan Wang¹ | Wei Su^{2,3} | Guangyao Zhou²  | Yijie Wang² | Chunnuan Wu¹ | Pengpeng Zhang² | Lianmin Zhang² 

¹Key Laboratory of Cancer Prevention and Therapy, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Tianjin's Clinical Research Center for Cancer, Tianjin, China | ²Department of Lung Cancer, Tianjin Lung Cancer Center, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China | ³Department of Endoscopy, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Tianjin, China

Correspondence: Lianmin Zhang (zhanglianmin1031@163.com) | Pengpeng Zhang (zpp19940120@tmu.edu.cn) | Chunnuan Wu (chunnuan.wu@tmu.edu.cn)

Received: 5 June 2025 | **Revised:** 11 July 2025 | **Accepted:** 16 July 2025

Funding: This work was supported by the Tianjin Natural Science Foundation (21JCYBJC01020).

Keywords: lung cancer | machine learning | N4-acetylcytidine | prognostic signature | RNA modification

ABSTRACT

Introduction: Lung adenocarcinoma, the most common subtype of non-small cell lung cancer, faces challenges such as drug resistance and tumor heterogeneity. N4-acetylcytidine (ac4C) is an important RNA modification involved in cancer progression, but its role in lung adenocarcinoma remains unclear.

Methods: This study analyzed transcriptomic and single-cell RNA sequencing data from public databases to investigate the expression and clinical significance of ac4C-related genes in lung adenocarcinoma. Ten machine learning algorithms were applied to develop and validate an ac4C-related gene signature (ARGSig) for prognosis prediction across multiple independent cohorts.

Results: Cells with high ac4C activity showed increased intercellular communication and activation of tumor-associated pathways. The ARGSig model effectively stratified patients by survival outcomes and predicted sensitivity to immune checkpoint inhibitors and chemotherapy agents.

Conclusion: ac4C modification and its related genes play a critical role in lung adenocarcinoma development. The ARGSig model provides a promising molecular tool for prognosis evaluation and personalized treatment guidance in lung adenocarcinoma patients.

Abbreviations: ac4C, N4-acetylcytidine; ARGSig, ac4C-related gene signature; AUC, area under the dose–response curve; CCLE, Cancer Cell Line Encyclopedia; C-index, concordance index; CNV, copy number variation; DEGs, differentially expressed genes; EMT, epithelial–mesenchymal transition; FDR, false discovery rate; GEO, Gene Expression Omnibus; GSVA, gene set variation analysis; ICIs, immune checkpoint inhibitors; IPS, immunophenoscores; LUAD, lung adenocarcinoma; ML, machine learning; MSigDB, Molecular Signatures Database; NSCLC, non-small cell lung cancer; OS, overall survival; PCA, principal component analysis; ROC, receiver operating characteristic; RSF, random survival forest; SCLC, small cell lung cancer; scRNA-seq, single-cell RNA sequencing; ssGSEA, single-sample gene set enrichment analysis; survivalSVM, survival support vector machine; TCGA, The Cancer Genome Atlas; TMB, tumor mutation burden; TPM, transcripts per million; UMAP, Uniform Manifold Approximation and Projection; WHO, World Health Organization.

Yuan Wang, Wei Su, and Guangyao Zhou have contributed equally to this work and share first authorship.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Thoracic Cancer* published by John Wiley & Sons Australia, Ltd.

1 | Introduction

Lung cancer is the leading cause of cancer-related mortality in both sexes and represents the most common malignant tumor within the respiratory system [1]. According to histological classification, the World Health Organization (WHO) categorizes lung malignancies into non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) [2]. NSCLC accounts for approximately 85% of cases and is the predominant form of lung cancer [3]. Lung adenocarcinoma, the most frequent subtype of NSCLC, constitutes about 40% of NSCLC cases [4, 5]. Surgical intervention remains the most effective treatment for early-stage lung adenocarcinoma patients (stages I and II) [6]. However, the majority of patients are diagnosed at advanced stages, rendering complete surgical resection challenging. Over the past decade, significant therapeutic advancements have been achieved through the development of molecular targeted therapies and immunotherapy. Immune checkpoint inhibitors (ICIs) have notably improved clinical outcomes in lung adenocarcinoma and represent a promising option for neoadjuvant immunotherapy in early resectable disease [7]. Nevertheless, studies indicate that the 5-year survival rate remains low at 4%–10%, benefiting only a minority of patients [8–10]. Molecular targeted therapy and immunotherapy continue to face substantial obstacles, primarily drug resistance and increased toxicity. Combination treatments, in particular, often induce severe adverse effects that frequently necessitate treatment discontinuation. Furthermore, the high heterogeneity of lung adenocarcinoma contributes to diverse mechanisms of resistance among patients, likely representing a major factor in therapeutic failure [11]. Therefore, there is an urgent need to elucidate the underlying molecular mechanisms and develop robust molecular classification models to accurately assess prognosis and guide personalized therapeutic strategies for lung adenocarcinoma (LUAD) patients.

RNA modification constitutes a fundamental component of epigenetic regulation [12]. It not only participates in gene expression control and cellular fate determination but also exerts a profound influence on tumor initiation, progression, and recurrence [13]. Increasing evidence indicates that dysregulation of RNA modifications can contribute to oncogenesis, presenting potential targets for cancer therapy [14]. N4-acetylcytidine (ac4C) is a conserved nucleoside modification present on tRNA and rRNA, representing a chemical alteration of RNA [15]. As an mRNA modification, ac4C plays a critical role in maintaining mRNA stability and enhancing translation. Specifically, under physiological conditions, ac4C is catalyzed enzymatically by the addition of an acetyl group to the N4 position of cytidine, affecting RNA stability, structure, and translational efficiency. Beyond normal physiology, ac4C modification exerts regulatory functions in various cancers and has been implicated in processes including cancer cell proliferation, metastasis, metabolism, and chemoresistance. In esophageal cancer, ac4C is significantly upregulated and associated with poor prognosis [16]. Studies in gastric cancer have shown that inhibiting ac4C reduces the hypoxia tolerance of cancer cells, thereby suppressing tumor progression in vivo [17]. NAT10 is currently identified as the sole enzyme responsible for catalyzing this modification. In prostate cancer, the acetyltransferase NAT10 promotes ac4C synthesis, facilitating epithelial–mesenchymal transition (EMT) and enhancing cellular migration [18]. Given its differential

expression and pivotal role in cancer progression, ac4C represents a promising target for tumor diagnosis and therapy. However, the relationship between ac4C modification and lung adenocarcinoma remains insufficiently explored.

This study systematically evaluates the role of ac4C modification in lung adenocarcinoma using single-cell RNA sequencing (scRNA-seq) to dissect the cellular composition of LUAD patients and its association with ac4C modification activity. Based on differentially expressed genes (DEGs) within cellular subpopulations, we developed a predictive model employing 10 machine learning algorithms, achieving superior accuracy compared to previous models and demonstrating potential clinical applicability. Functionally, we further assessed the model's capability to predict responses to immunotherapy and chemotherapy. Collectively, this research offers novel insights into the application of ac4C modification in lung adenocarcinoma and provides a foundation for personalized treatment and improved patient prognosis.

2 | Methods

2.1 | Data Sources

This study obtained transcriptomic and copy number variation (CNV) data for lung adenocarcinoma (LUAD) from The Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov>) to serve as the training set for model development. For model validation, six transcriptomic datasets were retrieved from the Gene Expression Omnibus (GEO) database: GSE13213 ($n = 119$) [19], GSE26939 ($n = 115$) [20], GSE29016 ($n = 39$) [21], GSE30219 ($n = 86$) [22], GSE31210 ($n = 227$) [23], and GSE42127 ($n = 134$) [24]. Additionally, two single-cell RNA sequencing datasets were sourced from GEO (GSE189357) [25] and the Genome Sequence Archive (HRA001130) [26].

To ensure data consistency and comparability, all gene expression profiles were converted to transcripts per million (TPM) format. The “combat” function from the R package “sva” was applied to mitigate potential batch effects [27]. Moreover, all TCGA and GEO datasets were log-transformed to standardize data formatting. Principal component analysis (PCA) was conducted to evaluate batch effects among datasets.

2.2 | Identification of Prognosis-Related ac4C Genes

Differentially expressed ac4C-related genes between normal and tumor tissues were identified using the limma package, with criteria set as false discovery rate (FDR) < 0.05 and $|\log_2$ fold change (FC)| > 1. Univariate Cox regression analysis was subsequently performed to assess the prognostic significance of these genes in LUAD patients. Based on 10-fold cross-validation, a comprehensive model selection was conducted using 10 machine learning algorithms—stepwise Cox regression, Lasso, Ridge, partial least squares regression for Cox models (plsRcox), CoxBoost, random survival forest (RSF), generalized boosted regression model (GBM), elastic net (Enet), supervised principal components (SuperPC), and

survival support vector machine (survival-SVM)—resulting in 101 model combinations. The optimal ac4C-related gene signature (ARGSig) was selected according to the highest concordance index (C-index). Our approach was adapted from Liu et al., who utilized similar algorithms but employed leave-one-out cross-validation; here, 10-fold cross-validation was adopted [28]. Model accuracy was evaluated by receiver operating characteristic (ROC) curves and PCA. Gene set variation analysis (GSVA) was employed to explore underlying biological mechanisms. Immune cycle and immunotherapy response-related pathways were assessed according to established protocols, with pathway gene sets obtained from the Molecular Signatures Database (MSigDB).

2.3 | Integrated Analysis of Genomic Variations and Immune Cell Infiltration

GISTIC 2.0 was utilized to identify recurrent genomic amplifications and deletions. Tumor mutation burden (TMB) was calculated using the R package “maftools.” Tumor immune gene expression and immune cell infiltration data were acquired from the TIDE website (<http://tide.dfci.harvard.edu/>), The Cancer Immunome Atlas (<https://tcia.at/home>), TIMER2.0 database, and the IOBR R package. Immune cell abundance and pathway activity were quantified using the single-sample gene set enrichment analysis (ssGSEA) algorithm.

2.4 | Single-Cell Gene Expression Data Preprocessing and Integration

Preprocessing was performed using the Seurat package (version 4.2.0). Genes expressed in at least 10 cells were retained. Cells expressing fewer than 200 or more than 5000 genes, or exhibiting mitochondrial gene expression exceeding 10%, were excluded. Sample integration was conducted via the Harmony package. Highly variable genes were selected for principal component analysis (PCA), with the top 30 components retained. Uniform Manifold Approximation and Projection (UMAP) was applied for dimensionality reduction and visualization. DEGs within cell subpopulations were identified using the “FindAllMarkers” function, and cell types were annotated based on known marker genes.

2.5 | Cell–Cell Communication Analysis

CellChat was employed to integrate gene expression data and analyze differences in intercellular communication. The default CellChatDB ligand–receptor database was utilized. Overexpressed ligands or receptors were detected to infer cell-specific interactions and identify enhanced ligand–receptor pairs.

2.6 | Drug Sensitivity Analysis in Cancer Cell Lines and LUAD Patients

Human cancer cell line expression data were sourced from the Broad Institute’s Cancer Cell Line Encyclopedia (CCLE). CERES

scores, reflecting gene dependency for cell survival and proliferation, were obtained from the DepMap portal, with lower scores indicating higher gene essentiality. Drug sensitivity data were derived from the CTRP and PRISM databases and quantified by the area under the dose–response curve (AUC), where lower AUC values represent greater sensitivity. Missing values were imputed using the k-nearest neighbors method; compounds and hematopoietic or lymphoid lineage cell lines with over 20% missing data were excluded. The pRRophetic package was applied, integrating drug sensitivity and genomic datasets to predict chemotherapy response in LUAD patients based on ridge regression models.

2.7 | Statistical Analysis

All data processing and statistical analyses were conducted in R software (version 4.2.0). Overall survival (OS) differences among subgroups were assessed using Kaplan–Meier survival curves and log-rank tests. Comparisons of continuous variables between two groups were performed using Wilcoxon rank-sum tests or Student’s *t* tests, whereas categorical variables were compared via χ^2 tests or Fisher’s exact tests. Multiple testing correction employed the FDR method. Pearson correlation coefficients were calculated to evaluate variable associations. All statistical tests were two-sided, with *p* values less than 0.05 considered statistically significant.

3 | Results

3.1 | Cell Type Identification and Marker Gene Screening

We obtained a publicly available single-cell RNA sequencing dataset from the GEO database to analyze cellular populations from various non-small cell lung cancer patients. After dimensionality reduction, 21 distinct cellular subpopulations were identified (Figure 1A). Based on specific gene expression patterns within these clusters, cell types were annotated as epithelial, Macro_FABP4, Macro_CCL17, fibroblast, T cells, endothelial, B/plasma, proliferating, mast cells, CD8_NK_like, Macro_IFI27, and Neutrophils_CSF3R (Figure 2B,C). Additionally, 11 lung adenocarcinoma samples were obtained from HRA005794, with the percentage distribution of cell types per sample depicted in Figure 2E. Guided by an extensive literature review, genes most closely associated with ac4C modification were selected for further analysis. Using the AUCell algorithm, cells were scored and subsequently divided into high (ac4C-high) and low (ac4C-low) scoring groups based on the median score. The results indicated that Epithelial and Macro_FABP4 cells exhibited higher ac4C-related gene activity, whereas T cells showed lower scores, suggesting differential ac4C gene expression activity among these populations (Figure 2A,B).

3.2 | Transcriptomic Similarity and Differences Between ac4C-High and ac4C-Low Cells

Correlation analyses and clustering generated heatmaps revealing that in ac4C-low cells, Macro_FABP4 and Macro_CCL17,

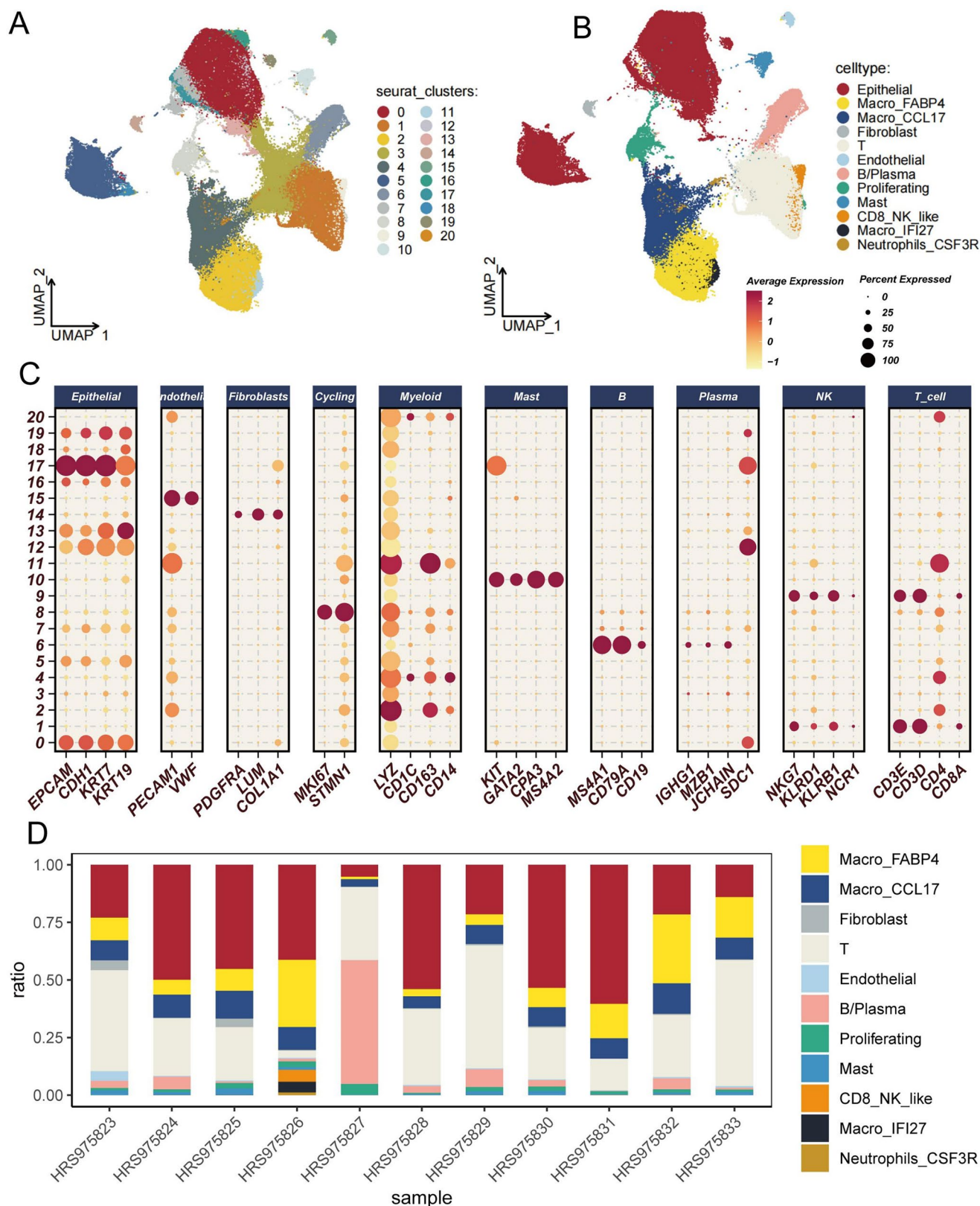


FIGURE 1 | Cell type annotation based on distinct cell markers in lung cancer samples. (A) UMAP visualization of 21 clusters. (B) UMAP visualization of 12 cell types, including Epithelial, Macro_FABP4, Macro_CCL17, Fibroblast, T, Endothelial, B/Plasma, Proliferating, Mast, CD8_NK_like, Macro_IFI27, Neutrophils-CSF3R. (C) The dot plot illustrates the average and percent expression levels of marker genes across various cell subtypes. (D) The line chart displays the distribution of each cell type throughout the samples.

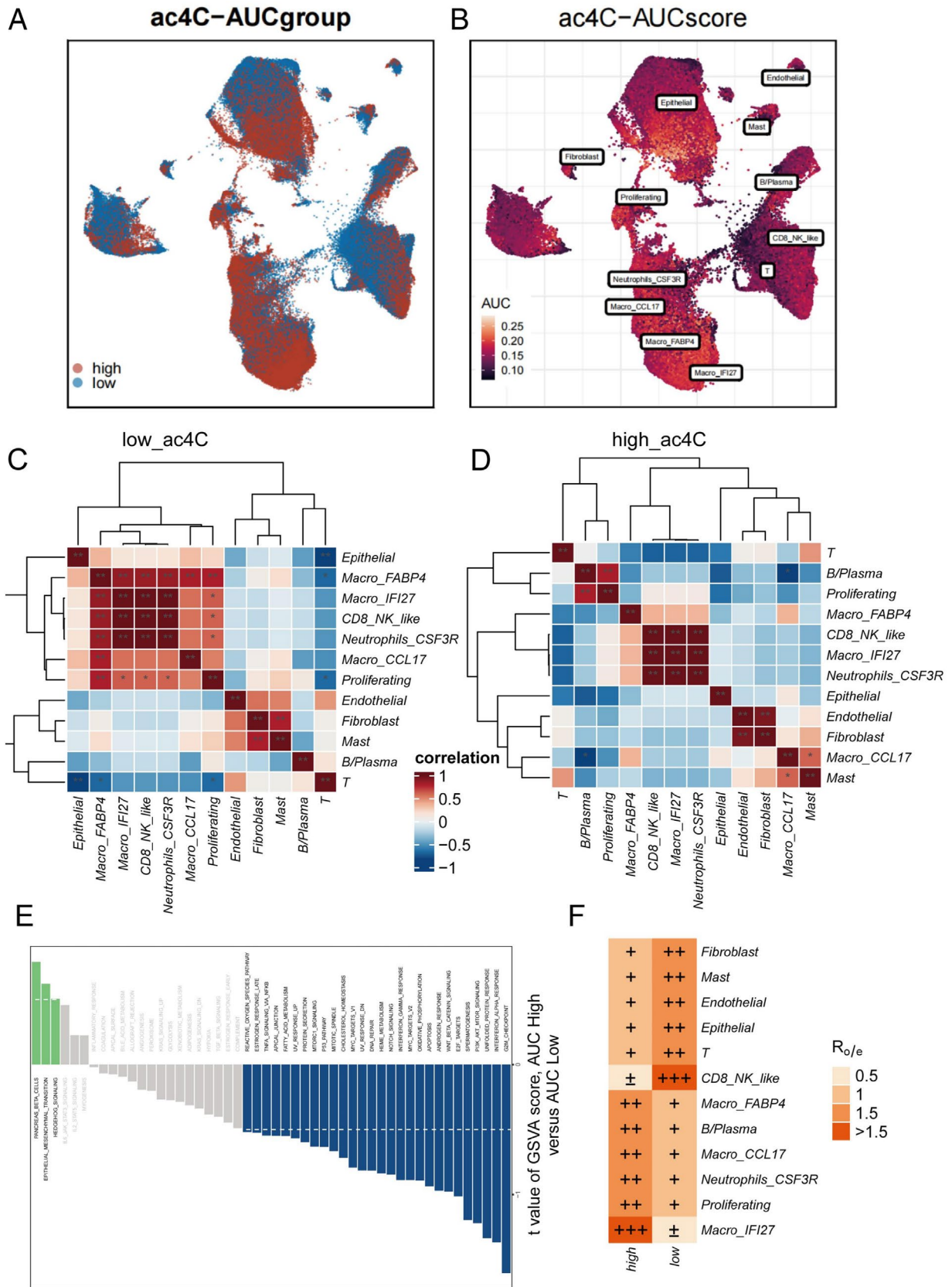


FIGURE 2 | Similarities and differences between ac4C-high and ac4C-low cells. (A) The distribution of high-score cells and low-score cells. (B) Scores of ac4C-related genes in clusters, by AUCcell. (C) The correlation of cell clusters in ac4C-low cells. (D) The correlation of cell clusters in ac4C-high cells. (E) The difference of GSVA scores. (F) The enrichment of ac4C-high and ac4C-low cells in different cell clusters, by single-cell tissue preference analysis.

CD8_NK_like, Neutrophils_CSF3R, Macro_IFI27, and Proliferating cells demonstrated strong transcriptomic correlations ($p < 0.01$). Proliferating cells showed moderate correlation with Macro_IFI27, CD8_NK_like, Neutrophils_CSF3R, and Mast cells ($p < 0.05$). Conversely, in ac4C-high cells, significant correlations were observed between B/Plasma and Proliferating cells, CD8_NK_like with Macro_IFI27 and Neutrophils_CSF3R, and endothelial with fibroblast populations ($p < 0.01$). Moderate correlations were found between NK/T cells and T regulatory cells ($p < 0.05$) (Figure 2C,D). To explore differences between ac4C-high and ac4C-low cells, canonical gene sets from the Broad Institute were applied. Compared to ac4C-low cells, ac4C-high cells showed significant upregulation of tumor-associated pathways, including metabolic processes and epithelial-mesenchymal transition, with three pathways upregulated, 28 downregulated, and 19 unchanged (Figure 2E).

Single-cell tissue preference analysis revealed significant enrichment of Macro_IFI27, Macro_FABP4, B/Plasma, Macro_CCL17, and Proliferating cells in the ac4C-high group, particularly Macro_IFI27, while CD8_NK_like cells were diminished. In contrast, CD8_NK_like, T cells, epithelial, endothelial, mast cells, and fibroblasts were enriched in the ac4C-low group, where Macro_IFI27 was reduced (Figure 2E).

3.3 | Differences in Cell–Cell Communication Between Groups

Cell–cell communication analysis demonstrated that ac4C-high cells exhibited significantly increased numbers and intensities of intercellular interactions compared to ac4C-low cells (Figure 3A). To identify key signaling pathways driving these differences, information flow between ac4C-high and ac4C-low cells was compared. Pathways including PTPRM, PARs, CD6, ALCAM, CD22, and MIF were active in both groups without significant differences. However, signaling via GRN, CLEC, PECAM1, EGF, VISFATIN, and ICAM pathways was markedly upregulated in ac4C-high cells (Figure 3B). All 12 cellular subtypes in the ac4C-high group exhibited enhanced communication, indicative of more active intercellular interactions. Among these, Fibroblasts showed the highest outbound interactions, whereas Macro_IFI27 cells displayed the highest inbound communication. Fibroblasts engaged robustly with the other 11 cell types, predominantly epithelial cells, endothelial cells, and macrophages, with these interactions being more pronounced in ac4C-high cells relative to ac4C-low cells (Figure 3C,D). Previous research has elucidated the role of ac4C in modulating the function of macrophages within the tumor microenvironment. In esophageal squamous cell carcinoma, NAT10 stabilizes the expression of fatty acid synthase (FASN) by mediating ac4C modification, thereby promoting the M2 polarization of macrophages [29]. In nasopharyngeal carcinoma, the inflammatory macrophage apoptosis inhibitor lncRNA (SIMALR) is highly expressed in tumor tissues. NAT10 enhances the stability of SIMALR through ac4C modification, which in turn facilitates the proliferation and metastasis of nasopharyngeal carcinoma cells [30]. Regarding NSCLC, the stability of GAS5 is regulated by ac4C modification. Elevated expression of GAS5 promotes the recruitment of macrophages and T cells, thereby inhibiting the progression of NSCLC [31]. In the future, we need to conduct

more systematic and comprehensive studies on the role of ac4C in other cell subpopulations.

3.4 | Elucidation of ac4C-Associated Biomarkers

To deepen understanding of ac4C's role in cancer, we identified a panel of ac4C-related biomarkers with specific focus on LUAD. Differential expression analysis integrating normal lung tissue data from GTEx and LUAD samples from TCGA revealed differentially expressed ac4C-associated genes (Figure 4A). Batch effects were mitigated across seven LUAD transcriptomic cohorts to enhance comparability (Figure 4B). Univariate Cox regression identified ac4C genes significantly associated with prognosis (Figure 4C). CNV analysis indicated frequent genomic alterations in ac4C-related genes, with S100A10 and S100A6 exhibiting the most extensive CNV amplifications (Figure 4D).

3.5 | Model Development, Validation, and Clinical Applicability for Immunotherapy and Chemotherapy

Utilizing the expression profiles of prognosis-related ac4C genes, an ARGSig was developed via an ensemble of machine learning algorithms. The TCGA dataset served as the training cohort, while six GEO datasets were employed for validation. The average C-index across the six validation cohorts guided model selection, ultimately identifying the combination of RSF and survival support vector machine (survivalSVM) as optimal (Figure 5A). ARGSig scores stratified patient prognosis effectively across all seven cohorts (Figure 5B), with high-ARGSig patients exhibiting poorer survival outcomes compared to low-ARGSig counterparts. ROC curves confirmed the robust prognostic predictive capability of ARGSig, consistently validated across datasets with 1-, 3-, and 5-year AUC values typically exceeding 0.65 (Figure 6A). Principal component analysis based on ARGSig gene expression further demonstrated clear separation between high- and low-risk LUAD patients, indicating distinct clustering patterns (Figure 6B). Moreover, immunophenoscores (IPS) derived from the TCIA database suggested that patients in the low-ARGSig group are more likely to benefit from PD-1, CTLA-4, or combination checkpoint inhibitor therapies (Figure 7A). Chemotherapy sensitivity predictions indicated enhanced responsiveness of the low-ARGSig group to agents such as 5-Fluorouracil, AZD6738, and AZD7762 (Figure 7B). Collectively, these findings substantiate ARGSig as a reliable prognostic tool for LUAD and highlight its potential utility in guiding personalized immunotherapeutic and chemotherapeutic interventions.

4 | Discussion

Significant advancements have been made in the field of epigenetics concerning RNA acetylation, particularly NAT10-mediated ac4C modification [32, 33]. Beyond its involvement in various physiological processes, ac4C contributes to cancer progression by modulating mRNA stability, thereby influencing cellular proliferation, metastasis, lipid metabolism, and drug resistance [15, 34–36]. In this study, scRNA-seq revealed

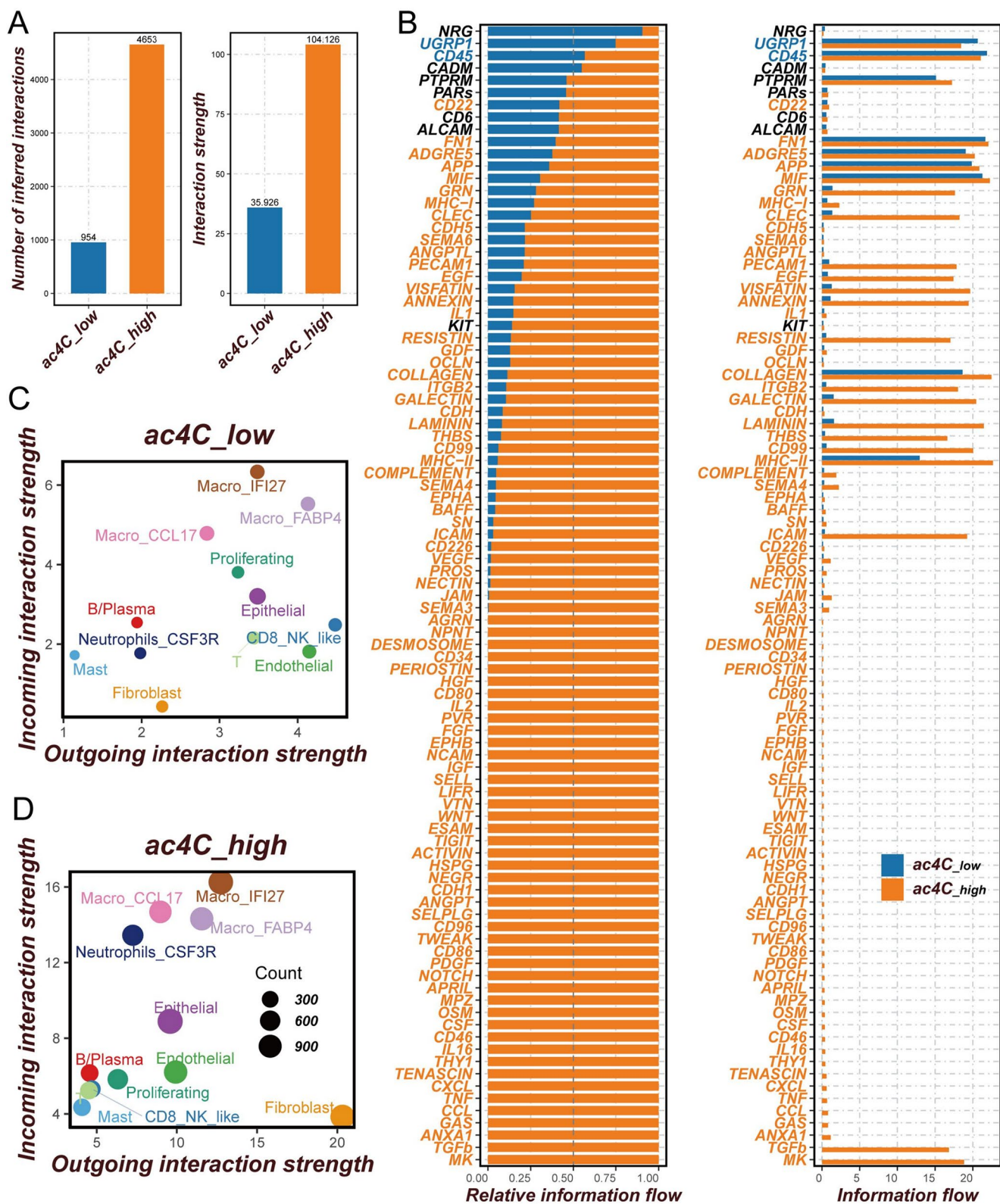


FIGURE 3 | Cell-cell communication analysis between ac4C-high and ac4C-low cells. (A) Number of inferred interactions and interaction strength in ac4C-high and ac4C-low cells. (B) Information flow in ac4C-high and ac4C-low cells. (C) Interaction strength of 12 cell clusters in ac4C-high cells. (D) Interaction strength of 12 cell clusters in ac4C-low cells.

heterogeneous cellular clusters within lung adenocarcinoma, including Epithelial, Macro_FABP4, T cells, B/Plasma, and CD8_NK_like populations, underscoring the complexity of the tumor microenvironment. Notably, ac4C modification levels

were markedly elevated in Macro_IFI27 cells compared to other cell types. Cell-cell communication analyses demonstrated that cells in the high-ac4C group exhibited stronger intercellular interactions, suggesting that ac4C-associated genes may

facilitate enhanced cellular communication. Further information flow analysis identified increased expression of PECAM1, EGF, VISFATIN, ANNEXIN, and LAMININ in the high-ac4C group, molecules known to play critical roles in tumor initiation, progression, and metastasis [37–41]. These factors predominantly regulate tumor cell proliferation, migration, immune

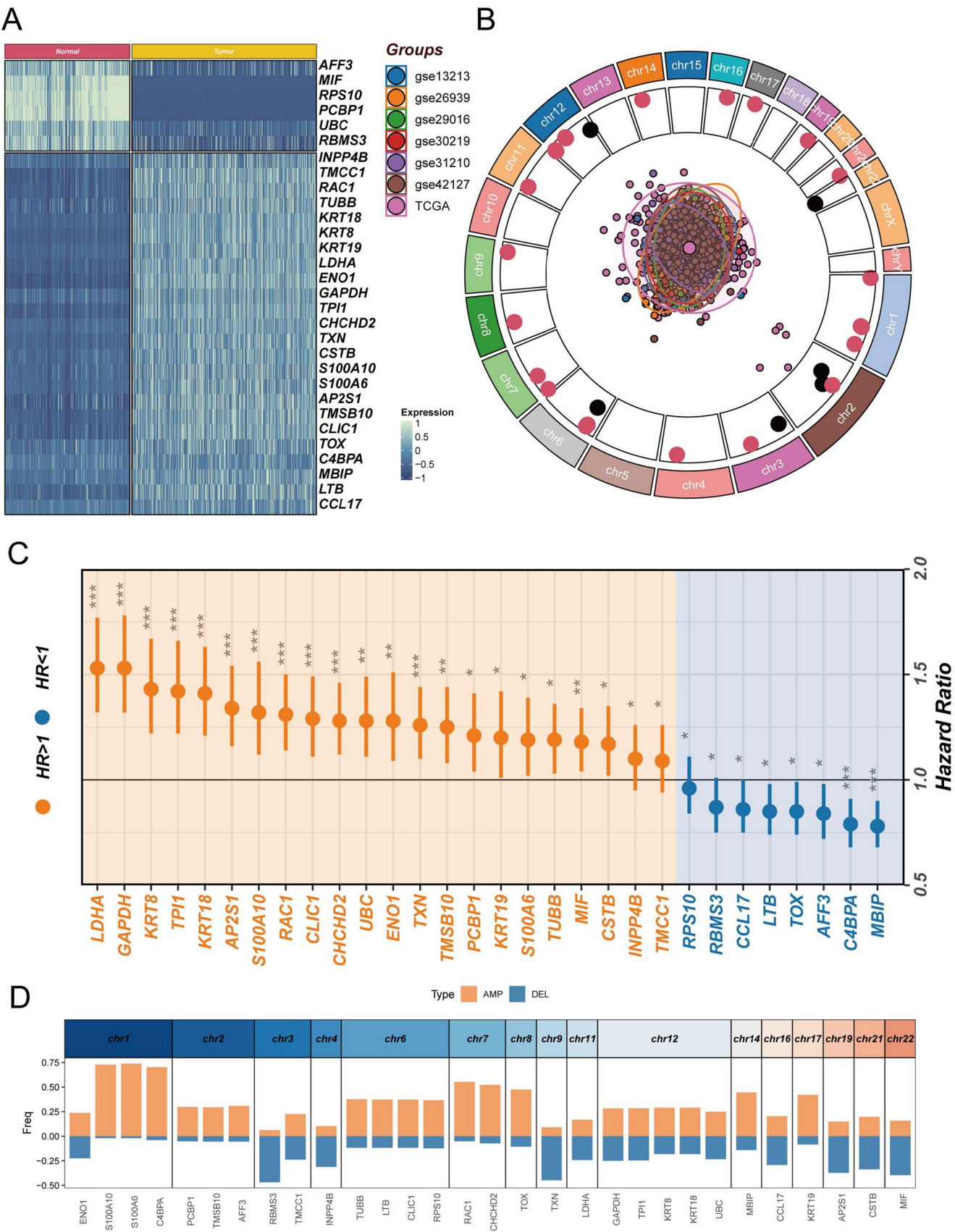
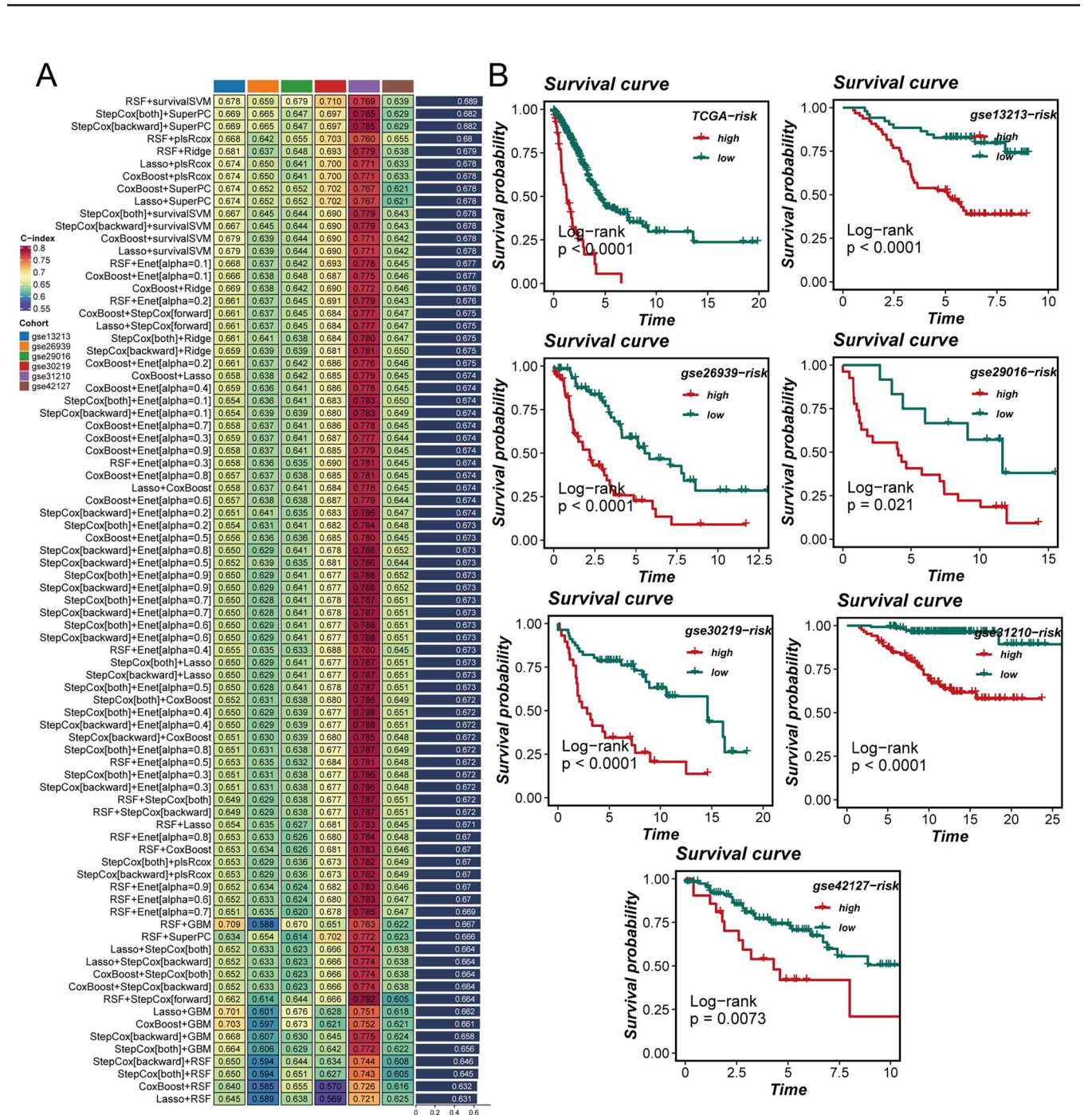


FIGURE 4 | Legend on next page.

FIGURE 4 | Identification and analysis of prognostic ac4C-related genes. (A) Identification of differentially expressed ac4C-related genes in tumor compared with normal lung tissue. (B) Chromosomal locations of prognostic ac4C-related genes, with red dots representing higher expression in tumors relative to normal lung tissue (depicted by log2FC values; the larger the value, the more peripheral the position), and black dots representing the opposite. The central pie chart shows all LUAD datasets used for modeling and principal component analysis plot after batch effect removal in seven LUAD datasets. (C) Univariate Cox analysis to determine potential prognostic ac4C-related genes. (D) Chromosomal locations and copy number variations of prognostic ac4C-related genes.



evasion, angiogenesis, and metabolic reprogramming. Machine learning (ML) has emerged as a powerful tool for extracting meaningful insights from large-scale datasets. Employing an

ensemble of 101 ML algorithm combinations, we developed the ARGSig, which demonstrated superior performance in predicting OS compared to previously reported prognostic signatures.

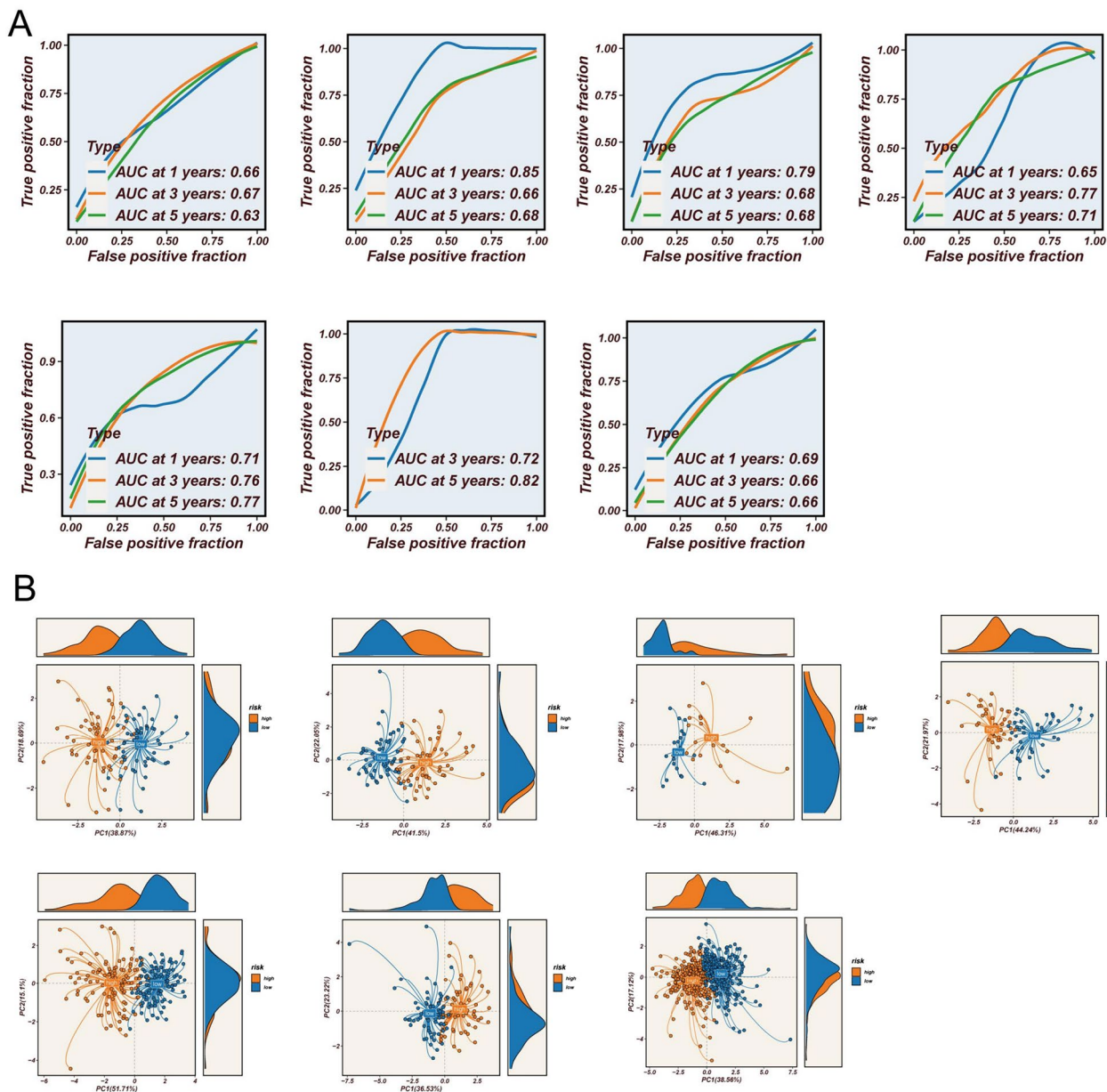


FIGURE 6 | Prediction value. (A) ROC curve for 1-year, 3-year, and 5-year survival predictions. (B) PCA of high-risk samples and low-risk samples across seven datasets.

Validation across six independent cohorts further confirmed ARGSig's robust predictive capability, highlighting its potential utility in clinical practice.

Immunotherapy has fundamentally transformed the therapeutic landscape of lung adenocarcinoma, particularly through ICIs such as PD-1 and PD-L1 blockers that disrupt tumor immune evasion mechanisms [10, 42, 43]. However, not all patients derive benefit, emphasizing the critical need for precise biomarkers to identify likely responders [44–46]. Our analysis revealed significant differences in immune status between the two ac4C-based risk groups. The low-risk group exhibited higher levels of immune cell infiltration, suggesting enhanced responsiveness to immunotherapy. Additionally, predictions of chemotherapeutic response indicated greater sensitivity of the low-ARGSig group to agents including 5-Fluorouracil,

AZD6738, and AZD7762, underscoring ARGSig's potential as a biomarker for both immunotherapeutic and chemotherapeutic efficacy. These findings support ARGSig as a valuable tool for stratifying lung adenocarcinoma patients most likely to benefit from these treatments.

The ARGSig developed herein exhibits strong prognostic relevance, correlating significantly with patient outcomes and thereby underscoring its promise as a clinical biomarker. Nonetheless, several limitations warrant consideration. First, this study focused exclusively on lung adenocarcinoma to reflect ac4C modification activity at the single-cell level; other NSCLC subtypes, such as squamous cell carcinoma, were not addressed. Future studies should investigate these subtypes as suitable datasets become available. Second, the ARGSig gene list may be incomplete, which could introduce bias.

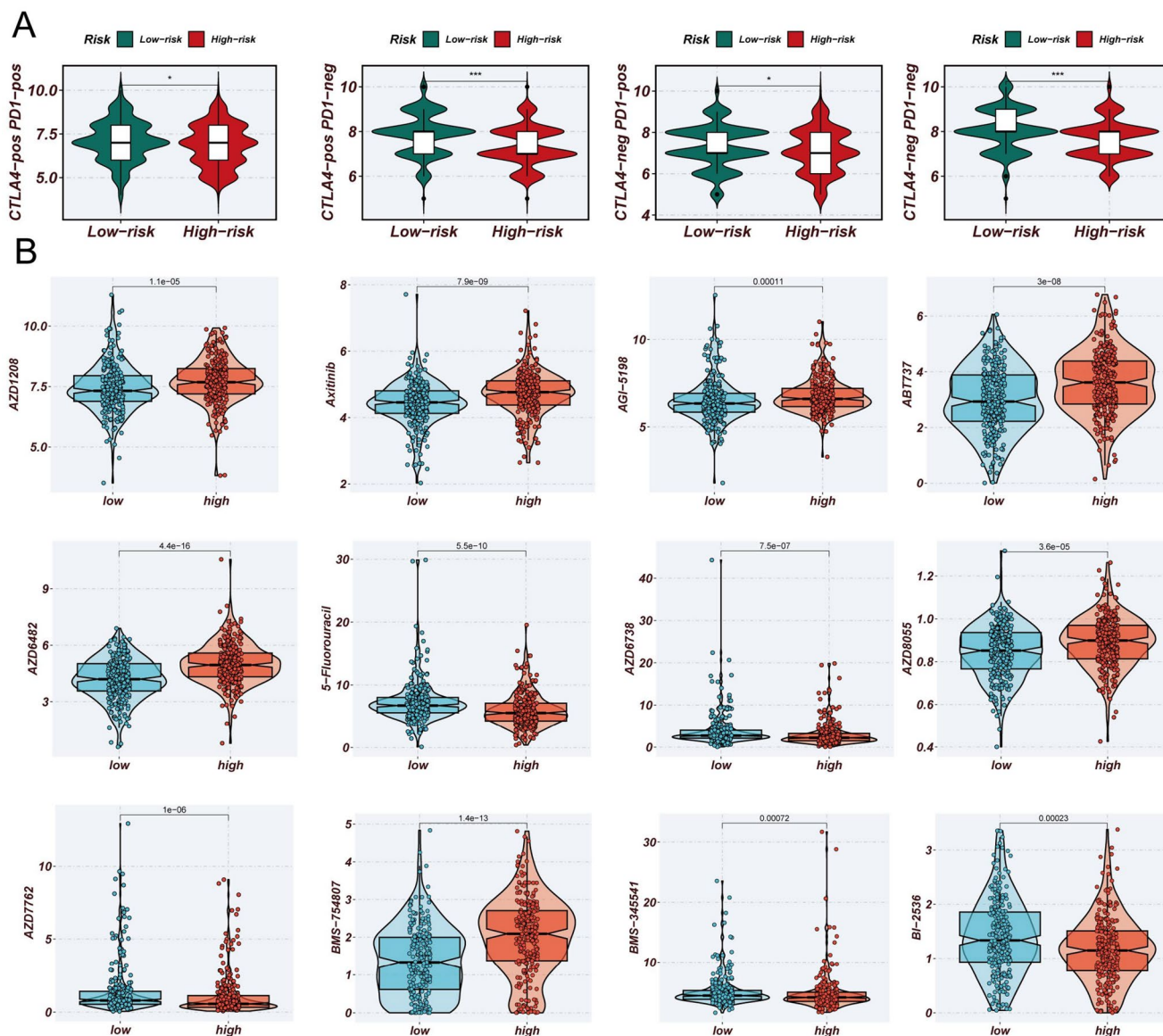


FIGURE 7 | Prediction value of chemotherapy and immune microenvironment. (A) Prediction of IPS for TCGA-LUAD patients using The Cancer Immunome Atlas, consistently indicating higher IPS and greater sensitivity to immunotherapy in the low-risk group. (B) Prediction of different chemotherapy for low-risk and high-risk group.

Lastly, the prognostic value of ARGSig requires validation in larger, multicenter cohorts. To further validate the clinical applicability of ARGSig, prospective studies and wet-lab experiments are warranted. Prospective cohorts could be designed to prospectively collect samples from lung adenocarcinoma patients and evaluate the predictive accuracy of ARGSig in real-time clinical settings. Additionally, wet-lab experiments, such as knockdown or overexpression studies of key genes within the ARGSig, could elucidate the underlying biological mechanisms and confirm the functional relevance of these genes in tumor progression and therapeutic response. Such validation efforts would provide robust evidence for the integration of ARGSig into clinical decision-making processes. Despite these limitations, this work provides critical insights into the role of ac4C in lung adenocarcinoma and presents a novel prognostic framework. The ARGSig model not only facilitates personalized therapeutic guidance but also

identifies new targets and directions for future LUAD treatment strategies.

5 | Conclusion

This study demonstrates that ac4C modification plays a vital role in lung adenocarcinoma progression. Using transcriptomic and single-cell data, we identified cellular differences in ac4C activity linked to tumor-related pathways and cell communication. The ARGSig developed through machine learning showed strong prognostic value and effectively predicted patient responses to immunotherapy and chemotherapy. These results suggest ARGSig as a useful biomarker for personalized treatment decisions. Future studies should validate this model in larger cohorts and explore its applicability to other lung cancer subtypes. Overall, this work provides new understanding of

ac4C's role and offers a tool to improve lung adenocarcinoma patient management.

Author Contributions

Yuan Wang, Wei Su, and Guangyao Zhou contributed equally as co-first authors, making substantial contributions to data collection, analysis, and manuscript drafting. Yijie Wang assisted in data processing and analysis. Chunnuan Wu supported the research work. Chunnuan Wu, Pengpeng Zhang, and Lianmin Zhang served as co-corresponding authors, providing overall research design, academic guidance, and critical manuscript revision, offering significant oversight to the research's academic quality and direction.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

This study used public data from online database. The data can be obtained freely from TCGA (<https://www.cancer.gov/ccg/research/genome-sequencing/tcga>) and GEO (<https://www.ncbi.nlm.nih.gov/geo/>).

References

1. K. Bajbouj, A. Al-Ali, R. K. Ramakrishnan, M. Saber-Ayad, and Q. Hamid, "Histone Modification in NSCLC: Molecular Mechanisms and Therapeutic Targets," *International Journal of Molecular Sciences* 22, no. 21 (2021): 11701.
2. K. Inamura, "Lung Cancer: Understanding Its Molecular Pathology and the 2015 WHO Classification," *Frontiers in Oncology* 7 (2017): 193.
3. M. G. Oser, M. J. Niederst, L. V. Sequist, and J. A. Engelman, "Transformation From Non-Small-Cell Lung Cancer to Small-Cell Lung Cancer: Molecular Drivers and Cells of Origin," *Lancet Oncology* 16, no. 4 (2015): e165–e172.
4. K. Funai, T. Yokose, G. Ishii, et al., "Clinicopathologic Characteristics of Peripheral Squamous Cell Carcinoma of the Lung," *American Journal of Surgical Pathology* 27, no. 7 (2003): 978–984.
5. J. Song, Y. Sun, H. Cao, et al., "A Novel Pyroptosis-Related lncRNA Signature for Prognostic Prediction in Patients With Lung Adenocarcinoma," *Bioengineered* 12, no. 1 (2021): 5932–5949.
6. F. R. Hirsch, G. V. Scagliotti, J. L. Mulshine, et al., "Lung Cancer: Current Therapies and New Targeted Treatments," *Lancet* 389, no. 10066 (2017): 299–311.
7. T. Cascone, J. Fradette, M. Pradhan, and D. L. Gibbons, "Tumor Immunology and Immunotherapy of Non-Small-Cell Lung Cancer," *Cold Spring Harbor Perspectives in Medicine* 12, no. 5 (2022): a037895.
8. S. Wang, R. Wang, D. Hu, C. Zhang, P. Cao, and J. Huang, "Machine Learning Reveals Diverse Cell Death Patterns in Lung Adenocarcinoma Prognosis and Therapy," *Npj Precision Oncology* 8, no. 1 (2024): 49.
9. T. Yang, Y. Xiong, Y. Zeng, et al., "Current Status of Immunotherapy for Non-Small Cell Lung Cancer," *Frontiers in Pharmacology* 13 (2022): 989461.
10. A. Lahiri, A. Maji, P. D. Potdar, et al., "Lung Cancer Immunotherapy: Progress, Pitfalls, and Promises," *Molecular Cancer* 22, no. 1 (2023): 40.
11. C. Wang, Q. Yu, T. Song, et al., "The Heterogeneous Immune Landscape Between Lung Adenocarcinoma and Squamous Carcinoma Revealed by Single-Cell RNA Sequencing," *Signal Transduction and Targeted Therapy* 7, no. 1 (2022): 289.
12. P. C. Teng, Y. Liang, A. A. Yarmishyn, et al., "RNA Modifications and Epigenetics in Modulation of Lung Cancer and Pulmonary Diseases," *International Journal of Molecular Sciences* 22, no. 19 (2021): 10592.
13. Q. Tang, L. Li, Y. Wang, et al., "RNA Modifications in Cancer," *British Journal of Cancer* 129, no. 2 (2023): 204–221.
14. I. Barbieri and T. Kouzarides, "Role of RNA Modifications in Cancer," *Nature Reviews. Cancer* 20, no. 6 (2020): 303–322.
15. L. Xie, X. Zhong, W. Cao, J. Liu, X. Zu, and L. Chen, "Mechanisms of NAT10 as ac4C Writer in Diseases," *Molecular Therapy* 32 (2023): 359–368.
16. W. Wei, S. Zhang, H. Han, et al., "NAT10-Mediated ac4C tRNA Modification Promotes EGFR mRNA Translation and Gefitinib Resistance in Cancer," *Cell Reports* 42, no. 7 (2023): 112810.
17. Q. Yang, X. Lei, J. He, et al., "N4-Acetylcytidine Drives Glycolysis Addiction in Gastric Cancer via NAT10/SEPT9/HIF-1alpha Positive Feedback Loop," *Advanced Science (Weinheim, Baden-Wurttemberg, Germany)* 10, no. 23 (2023): e2300898.
18. K. J. Li, Y. Hong, Y. Z. Yu, et al., "NAT10 Promotes Prostate Cancer Growth and Metastasis by Acetylating mRNAs of HMGA1 and KRT8," *Advanced Science (Weinheim, Baden-Wurttemberg, Germany)* 11, no. 32 (2024): e2310131.
19. S. Tomida, T. Takeuchi, Y. Shimada, et al., "Relapse-Related Molecular Signature in Lung Adenocarcinomas Identifies Patients With Dismal Prognosis," *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 27, no. 17 (2009): 2793–2799.
20. M. D. Wilkerson, X. Yin, V. Walter, et al., "Differential Pathogenesis of Lung Adenocarcinoma Subtypes Involving Sequence Mutations, Copy Number, Chromosomal Instability, and Methylation," *PLoS One* 7, no. 5 (2012): e36530.
21. J. Staaf, G. Jonsson, M. Jonsson, et al., "Relation Between Smoking History and Gene Expression Profiles in Lung Adenocarcinomas," *BMC Medical Genomics* 5 (2012): 22.
22. S. Rousseaux, A. Debernardi, B. Jacquiau, et al., "Ectopic Activation of Germline and Placental Genes Identifies Aggressive Metastasis-Prone Lung Cancers," *Science Translational Medicine* 5, no. 186 (2013): 186ra66.
23. H. Okayama, T. Kohno, Y. Ishii, et al., "Identification of Genes Up-regulated in ALK-Positive and EGFR/KRAS/ALK-Negative Lung Adenocarcinomas," *Cancer Research* 72, no. 1 (2012): 100–111.
24. H. Tang, G. Xiao, C. Behrens, et al., "A 12-Gene Set Predicts Survival Benefits From Adjuvant Chemotherapy in Non-Small Cell Lung Cancer Patients," *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 19, no. 6 (2013): 1577–1586.
25. J. Zhu, Y. Fan, Y. Xiong, et al., "Delineating the Dynamic Evolution From Preneoplasia to Invasive Lung Adenocarcinoma by Integrating Single-Cell RNA Sequencing and Spatial Transcriptomics," *Experimental & Molecular Medicine* 54, no. 11 (2022): 2060–2076.
26. Z. Wang, Z. Li, K. Zhou, et al., "Deciphering Cell Lineage Specification of Human Lung Adenocarcinoma With Single-Cell RNA Sequencing," *Nature Communications* 12, no. 1 (2021): 6500.
27. Y. Zhang, G. Parmigiani, and W. E. Johnson, "ComBat-Seq: Batch Effect Adjustment for RNA-Seq Count Data," *NAR Genomics and Bioinformatics* 2, no. 3 (2020): lqaa078.
28. Z. Liu, L. Liu, S. Weng, et al., "Machine Learning-Based Integration Develops an Immune-Derived lncRNA Signature for Improving Outcomes in Colorectal Cancer," *Nature Communications* 13, no. 1 (2022): 816.
29. C. Jin, J. Gao, J. Zhu, Y. Ao, B. Shi, and X. Li, "Exosomal NAT10 From Esophageal Squamous Cell Carcinoma Cells Modulates Macrophage

Lipid Metabolism and Polarization Through ac4C Modification of FASN,” *Translational Oncology* 45 (2024): 101934.

30. S. Gong, H. Qiao, J. Y. Wang, et al., “Ac4C Modification of lncRNA SIMALR Promotes Nasopharyngeal Carcinoma Progression Through Activating eEF1A2 to Facilitate ITGB4/ITGA6 Translation,” *Oncogene* 43, no. 38 (2024): 2868–2884.

31. Z. Wang, J. Luo, H. Huang, et al., “NAT10-Mediated Upregulation of GAS5 Facilitates Immune Cell Infiltration in Non-Small Cell Lung Cancer via the MYBBP1A-p53/IRF1/Type I Interferon Signaling Axis,” *Cell Death Discovery* 10, no. 1 (2024): 240.

32. C. Yang, T. Wu, J. Zhang, et al., “Prognostic and Immunological Role of mRNA ac4C Regulator NAT10 in Pan-Cancer: New Territory for Cancer Research?,” *Frontiers in Oncology* 11 (2021): 630417.

33. Z. Xue, H. Xie, Y. Shan, et al., “NAT10 Inhibition Promotes ac4C-Dependent Ferroptosis to Counteract Sorafenib Resistance in Nasopharyngeal Carcinoma,” *Cancer Science* 115, no. 10 (2024): 3256–3272.

34. X. Chen, Y. Hao, Y. Liu, et al., “NAT10/ac4C/FOXP1 Promotes Malignant Progression and Facilitates Immunosuppression by Reprogramming Glycolytic Metabolism in Cervical Cancer,” *Advanced Science (Weinheim, Baden-Wurttemberg, Germany)* 10, no. 32 (2023): e2302705.

35. L. Liao, Y. He, S. J. Li, et al., “Lysine 2-Hydroxyisobutyrylation of NAT10 Promotes Cancer Metastasis in an ac4C-Dependent Manner,” *Cell Research* 33, no. 5 (2023): 355–371.

36. Z. Gu, L. Zou, X. Pan, et al., “The Role and Mechanism of NAT10-Mediated ac4C Modification in Tumor Development and Progression,” *MedComm* (2020) 5, no. 12 (2024): e70026.

37. Z. T. Liang, J. K. Li, J. Li, H. Tang, C. F. Guo, and H. Q. Zhang, “PECAM1 Plays a Role in the Pathogenesis and Treatment of Bone Metastases,” *Frontiers in Genetics* 14 (2023): 1151651.

38. T. Murata, H. Mizushima, I. Chinen, et al., “HB-EGF and PDGF Mediate Reciprocal Interactions of Carcinoma Cells With Cancer-Associated Fibroblasts to Support Progression of Uterine Cervical Cancers,” *Cancer Research* 71, no. 21 (2011): 6633–6642.

39. Y. Y. Wang, A. C. Hung, S. Lo, and S. F. Yuan, “Adipocytokines Visfatin and Resistin in Breast Cancer: Clinical Relevance, Biological Mechanisms, and Therapeutic Potential,” *Cancer Letters* 498 (2021): 229–239.

40. M. V. Christensen, C. K. Hogdall, K. M. Jochumsen, and E. V. S. Hogdall, “Annexin A2 and Cancer: A Systematic Review,” *International Journal of Oncology* 52, no. 1 (2018): 5–18.

41. J. Dai, P. J. Cimino, K. H. Gouin, 3rd, et al., “Astrocytic Laminin-211 Drives Disseminated Breast Tumor Cell Dormancy in Brain,” *Nature Cancer* 3, no. 1 (2022): 25–42.

42. R. S. Riley, C. H. June, R. Langer, and M. J. Mitchell, “Delivery Technologies for Cancer Immunotherapy,” *Nature Reviews. Drug Discovery* 18, no. 3 (2019): 175–196.

43. M. Feng, W. Jiang, B. Y. S. Kim, C. C. Zhang, Y. X. Fu, and I. L. Weissman, “Phagocytosis Checkpoints as New Targets for Cancer Immunotherapy,” *Nature Reviews. Cancer* 19, no. 10 (2019): 568–586.

44. Y. Wang, C. Li, Z. Wang, et al., “Comparison Between Immunotherapy Efficacy in Early Non-Small Cell Lung Cancer and Advanced Non-Small Cell Lung Cancer: A Systematic Review,” *BMC Medicine* 20, no. 1 (2022): 426.

45. C. Li, X. Zheng, P. Li, et al., “Heterogeneity of Tumor Immune Microenvironment and Real-World Analysis of Immunotherapy Efficacy in Lung Adenosquamous Carcinoma,” *Frontiers in Immunology* 13 (2022): 944812.

46. S. Li, W. Li, T. Ma, et al., “Assessing the Efficacy of Immunotherapy in Lung Squamous Carcinoma Using Artificial Intelligence Neural Network,” *Frontiers in Immunology* 13 (2022): 1024707.