

Bioinformatics Analysis Reveals Esophageal Squamous Cell Carcinoma-Associated Lysine-Related Genes as Poor Prognostic Factors and Potential Therapeutic Targets

Liping Huang¹, Binxiang Cai¹, Yanyan Lin¹, Zhijian Hu², Dahua Fu^{1,*}

¹School of Basic Medicine, Zhangzhou Health Vocational College, 363005 Zhangzhou, Fujian, China

²School of Public Health, Fujian Medical University, 350108 Fuzhou, Fujian, China

*Correspondence: zzhc306@163.com (Dahua Fu)

Submitted: 27 February 2024 Revised: 21 March 2024 Accepted: 16 April 2024 Published: 30 August 2024

Background: Esophageal squamous cell carcinoma (ESCC) represents a prevalent malignancy within the digestive tract, often diagnosed at advanced stages, leading to treatment challenges and poor prognosis. Despite its essential roles in biosynthesis, antioxidation, and energy metabolism, the involvement of lysine in cancer pathogenesis remains poorly understood. This study aimed to elucidate the significance of lysine-associated genes in ESCC, potentially enhancing treatment outcomes.

Methods: mRNA expression data for ESCC were retrieved from The Cancer Genome Atlas (TCGA) for limma differential expression analysis. Lysine-related genes were extracted from the GeneCards database and overlapped with differentially expressed genes. Lysine levels were quantified using single-sample Gene Set Enrichment Analysis (ssGSEA), followed by metabolic pathway analysis comparing high and low lysine levels. The random survival forest (RSF) algorithm identified lysine-related genetic signatures associated with poor prognosis of ESCC. Further analysis encompassed drug susceptibility profiling and assessment of immune cell infiltration. Subsequently, single-cell analysis was conducted using data from the Gene Expression Omnibus (GEO) database (GSE196756) to validate key lysine-related genes.

Results: Ubiquitination Factor E4A (*UBE4A*) and Apolipoprotein C1 (*APOC1*) were identified as key lysine-associated genes in ESCC. Functional enrichment analysis revealed that the lysine-related genes remarkably enriched lysine-regulated pathways, including histone modification, histone lysine methylation, and lysine degradation pathways. Moreover, these lysine-related differentially expressed genes (DEGs) served as independent prognostic factors, and a nomogram incorporating these genes and clinical factors accurately predicted 1- and 2-year survival rates in ESCC patients. Drug databases highlighted *AICAR*, *BI.D1870*, *MS.275*, and lapatinib as potential therapeutic agents. Immunomicroenvironment infiltration analysis indicated that *UBE4A* was actively connected with T cells CD4 memory resting and passively connected with B cells memory. *APOC1* was actively linked to macrophages M2, T cells CD4 memory activated, T cells CD8, and was passively linked to the activated Dendritic cells, T cells CD4 memory resting and Eosinophils. We further performed single-cell analysis, and 21 subtypes were obtained by t-Distributed Stochastic Neighbor Embedding (tSNE) algorithm. The total clusters were labeled to tissue stem cells, neurons, epithelial cells, B cells, neutrophils, monocytes, T cells, smooth muscle cells, NK cells, and endothelial cells.

Conclusions: Our findings underscore the pivotal role of lysine-related gene regulation in ESCC. The identified lysine-related genes hold promise as prognostic indicators for ESCC patients, shedding light on potential mechanisms underlying poor prognosis and facilitating the search for targeted therapeutics to enhance the success rate of ESCC treatment.

Keywords: ESCC; TME; lysine; gene signature; therapeutic targets

Introduction

Esophageal cancer (EC) ranks as the eighth most prevalent malignancy globally and stands as the sixth leading cause of cancer-related mortality. Esophageal squamous cell carcinoma (ESCC) accounts for approximately 90% of EC cases worldwide [1]. The primary strategies for treating ESCC include radical surgery, chemotherapy, and radiotherapy, with early-stage patients exhibiting prolonged survival post-surgery. However, due to the subtle

onset of symptoms and the lack of specific clinical manifestations, 80% of patients are diagnosed at advanced stages, precluding surgical interventions and resulting in a dismal 5-year survival rate of merely 20% [2]. Currently, ESCC lacks molecular markers characterized by high specificity and good efficacy, and conventional targeted therapies utilized in other cancers have shown limited benefits in ESCC cases. Targeting epigenetic and post-translational abnormalities implicated in ESCC pathogenesis holds promise as an alternative approach to traditional treatments. Hence,

it is paramount to address ESCC by exploring its molecular mechanisms and identifying effective therapeutic interventions from a novel perspective. The metabolic profile of tumor cells markedly diverges from that of normal cells. Recent study underscores metabolic reprogramming as a pivotal hallmark of tumorigenesis [3]. Tumor cells adeptly rewire metabolic pathways to sustain viability and proliferation under stressors, epitomized by the renowned “Warburg effect”, wherein cancer cells preferentially metabolize glucose into lactic acid under aerobic conditions. In glucose-deprived conditions, normal cells utilize ketone bodies derived from fatty acid oxidation and amino acid catabolism as alternative energy sources. Lysine, an essential amino acid crucial for protein synthesis, also plays a pivotal role in fueling the rapid proliferation of cancer cells. Post-translational modifications such as acetylation, ubiquitination, methylation, succinylation, and malonylation on lysine residues modulate protein stability, activity, and sub-cellular localization, intricately participating in tumor cell metabolic reprogramming.

Findings have revealed that acetylation at Lys-305 facilitates the interaction between pyruvate kinase M2 (*PKM2*) and heat shock cognate protein 70 (HSC70), leading to *PKM2* degradation via chaperone-mediated autophagy, ultimately promoting tumor proliferation [4]. Moreover, the degree of acetylation modification at the Lys-245 of glyceraldehyde 3-phosphate dehydrogenase (GAPDH) is influenced by glucose concentration. Elevated glucose levels enhance GAPDH Lys-254 acetylation, enhancing GAPDH enzyme activity and fostering lung cancer proliferation [5]. In addition, amino acids can directly impact epigenetic regulation through substrate availability. Lysine is broken down into intermediate acetyl coenzyme A (acetyl-CoA), which can be used by histone acetyltransferases (HATs). Lysine also provides acetyl-CoA to the acetyltransferase recombinant E1A-binding protein P300 (EP300). EP300-mediated acetylation, notably at the K1097 site of Raptor, activates the mechanistic target of rapamycin complex 1 (mTORC1), regulating tumor cell growth, proliferation, and apoptosis, thereby impacting the tumor onset and progression.

The relationship between abnormal amino acid metabolism (AAM) and cancer development, along with its underlying mechanisms, has emerged as a new area of research. However, investigations into the association between lysine and tumor development remain sparse. Recent studies have underscored the pivotal role of amino acids as metabolic substrates within tumor cells and the tumor microenvironment (TME). The balance of amino acid metabolism significantly impacts tumor cell proliferation rates and the immune function of tumor-associated macrophages (TAMs). Tumor cells compete with TAMs for amino acids to fulfill their nutritional requirements, disrupting the amino acid equilibrium in TME. AAM regulates

T cell activation and function, while *in vivo* tumor progression can be constrained by dietary amino acid interventions [6–8].

Huairui Yuan’s investigation [9] revealed lysine metabolic reprogramming in glioblastoma stem cells, leading to the generation of crotonyl CoA, elevating cellular crotonylation levels, especially the crotonylation modification of histone H4. Histone crotonylation modulates H3K27ac and H3K9me3, interferon signaling, and CD8⁺ T cell infiltration, ultimately promoting tumor growth [9]. Given the negative impact of metabolic imbalance on tumor immunity, inhibiting relevant amino acid metabolism to rectify amino acid imbalances in TME represents a promising adjunct antitumor immunotherapy.

Accordingly, this research investigated the correlation between lysine-related genes and esophageal squamous cell carcinoma, unveiling a novel target for chemotherapy and ESCC immunotherapy from the perspective of amino acid metabolism.

Materials and Methods

Data Collection and Identification of Cancer-Associated Lysine-Associated Genes in ESCC

We retrieved the original mRNA expression data for esophageal squamous cell carcinoma data (ESCC) from publicly available repositories. A total of 99 samples were collected for analysis. Additionally, we retrieved the dataset GSE196756 from the NCBI Gene Expression Omnibus (GEO) public database for single-cell analysis (<https://www.ncbi.nlm.nih.gov/>). The “Limma” package, an R software (version 4.2.2, <https://mirrors.tuna.tsinghua.edu.cn/CRAN/>), for differential expression analysis was utilized to identify genes that exhibited significant differential expression between normal and tumor samples. Differential expression analysis was conducted with the criteria of $p\text{-value} < 0.05$ and $\log_2$ fold change (FC) > 0.585 [$\log_2$ FC]. Subsequently, volcano plots and heatmaps were generated to visualize the differentially expressed genes (DEGs). Following the identification of DEGs in the The Cancer Genome Atlas (TCGA) dataset, we extracted lysine-related genes from the GeneCards database and intersected them with the DEGs to identify lysine-associated genes.

Functional Annotation and Enrichment Analysis

Functional annotation of genes was performed using the “ClusterProfiler” package in R to comprehensively explore the functional correlation of the intersecting targets. Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) databases were employed to evaluate the relevant functional categories. Enrichment analysis was considered significant for KEGG and GO pathways with p and q values below 0.05.

Random Survival Forest Algorithm

Random survival forest (RSF) is a machine-learning algorithm that analyzes survival data. In this study, the RSF algorithm was implemented using the randomForestSRC package (version 4.2.2, <https://www.randomforestsrc.org/> <https://ishwaran.org/>) to screen for feature genes and rank the significance of prognostic-associated genes. The algorithm was run with 1000 iterations in Monte Carlo simulation (nrep = 1000). Genes with relative importance >0.1 were considered final feature genes.

Immune Cell Infiltration Analysis

The CIBERSORT method is widely used for assessing immune cell subtypes within the microenvironment. This approach utilizes support vector regression to deconvolute the matrix of immune cell subtype expressions, which includes 547 biomarkers distinguishing 22 human immune cell phenotypes, such as B cells, myeloid cell subsets, plasma cells, and T cells. In this study, the CIBERSORT algorithm was employed to analyze patient data, predicting the relative percentage of 22 categories of immune infiltration cells. Additionally, Spearman's correlation analysis was conducted to explore the relationship between gene expression and immune cell populations.

Drug Sensitivity Analysis

The Genomics of Drug Sensitivity in Cancer (GDSC) database (<https://www.cancerrxgene.org/>), housing extensive information on drug susceptibility, served as the primary source for drug sensitivity analysis. The “pRRophetic” package (version 4.2.2, “<http://bioconductor.org/biocLite.R>”) in R was used to predict the chemotherapy susceptibility of all tumor samples. Estimations of Half maximal inhibitory concentration (IC₅₀) values for specific chemotherapy drugs were obtained through a regression approach, with the precision of regression and prediction assessed via 10-fold cross-validation using the GDSC training set. Default parameter values, including “combat” for batch effect elimination and mean repeated gene expression, were selected.

GSEA Path Enrichment Analysis

Gene Set Enrichment Analysis (GSEA) was employed to further elucidate differences between high- and low-expression signaling pathways. Background gene sets sourced from version 7.0 of the Molecular Signatures Database (MsigDB) were utilized for pathway annotation. Differential pathway expression analysis was conducted between subtypes, with significantly enriched gene sets (regulated $p < 0.05$) sequenced based on consistency scores. GSEA analysis facilitates the examination of strong associations between biological significance and disease categorization.

Gene Set Variation Analysis (GSVA)

Gene Set Variation Analysis (GSVA) is an unsupervised, nonparametric approach for evaluating transcriptome gene set enrichment. By scoring gene sets of interest, GSVA transforms gene-level data into pathway-level information, allowing for the assessment of biological function within samples. Gene sets sourced from the Molecular Signatures Database (<http://www.broadinstitute.org/gsea/msigdb/index.jsp>) were employed, and the GSVA algorithm was utilized to generate integrated scores for each gene set, enabling the evaluation of potential biological functional variations across samples.

Nomogram Model Construction

A nomogram was constructed based on regression analysis to illustrate the relationship between variables within the predictive model, integrating gene expression and clinical symptoms on a unified plane. A multi-factor regression model was established to assign scores to each influencing factor based on their contribution to the outcome variable, determined by the size of regression coefficients. These individual scores were then aggregated to yield total scores, facilitating the prediction of the outcome.

Transcriptional Regulation Analysis of Key Genes

The R package “RcisTarget” (version 4.2.2, <https://resources.aertslab.org/cistarget/>) was employed to predict transcription factors associated with key genes. Calculations within RcisTarget are based on motif analysis, with the normalized enrichment score (NES) for motifs dependent on the total quantity of motifs. Additionally, annotated files were inferred from gene sequences and motif similarity beyond those provided by the source data. The first step involves assessing the overexpression of each motif within the gene set by calculating the area under the curve (AUC) for each motif-motif pair derived from the recovery curve of the gene set sequence. Subsequently, the NES was computed for each motif based on the AUC of all motifs.

miRNA Network Construction

MicroRNAs (miRNAs) are small, non-coding RNAs that regulate gene expression by facilitating mRNA degradation or inhibiting mRNA translation. In this study, we conducted a comprehensive analysis to identify miRNAs associated with key genes regulating gene degradation or transcription of deleterious genes. Key gene-associated miRNAs were obtained from the miRcode database (<http://www.mircode.org/index.php>), and the gene miRNA network was visualized using Cytoscape software (version 4.2.2, <https://cytoscape.org/>)

Single-Cell Analysis

Gene expressions were processed using the “Seurat” package for single-cell analysis, and low-expression genes

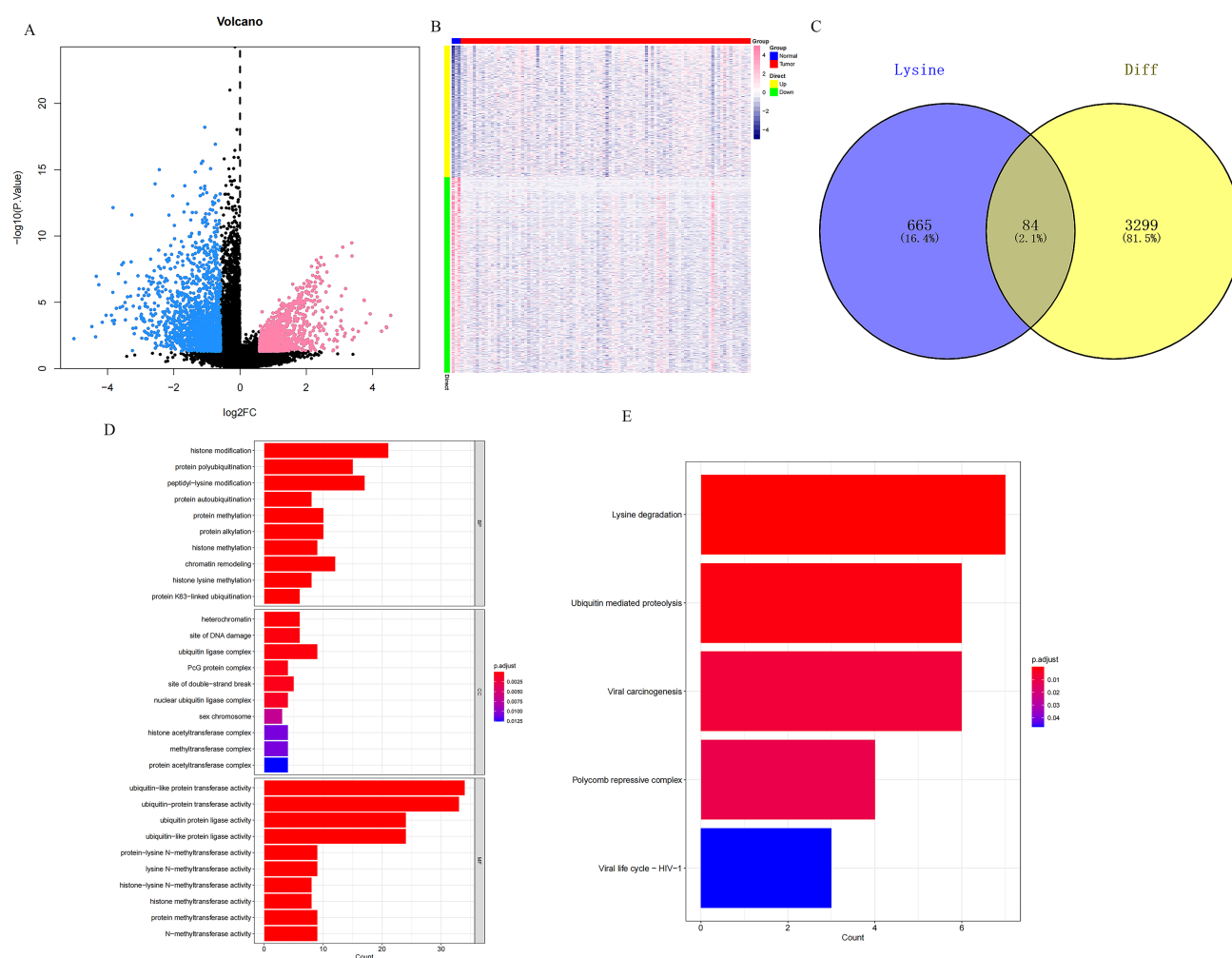


Fig. 1. Identification and functional enrichment analysis of differentially expressed genes in esophageal squamous cell carcinoma. (A) Volcano plots of differentially expressed genes (DEGs) in TCGA datasets, displaying the number of upregulated and downregulated genes, and the total number of DEGs detected in each dataset. (B) Heat map of DEGs in The Cancer Genome Atlas (TCGA). (C) Overlapping DEGs. (D) Most enriched Gene Ontology (GO) terms for all DEGs. (E) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways for all DEGs.

(defined as nFeature RNA >50 & percent.mt <5) were filtered. Data were standardized, normalized, and subjected to principal component analysis (PCA). The optimal number of principal components was determined using the Elbow-Plot method. Additionally, t-Distributed Stochastic Neighbor Embedding (tSNE) analysis was conducted to investigate the positional correlation between clusters. Cluster labeling was performed using the “cellDex” package to identify cells of importance in disease occurrence.

Statistical Analysis

All statistical analyses were conducted using R language (version 4.2.2). All statistical tests were two-tailed, and a p -value < 0.05 was considered statistically significant.

Results

Identification of Cancer-Associated Lysine-Associated Genes in Esophageal Squamous Cell Carcinoma

To identify lysine-associated differentially expressed genes (DEGs) in esophageal squamous cell carcinoma (ESCC), we first obtained the dataset related to ESCC from the TCGA database, comprising expression profiles from 99 patients, including 3 normal samples and 96 tumor samples. The “Limma” package was utilized to compute the DEGs between the normal and tumor groups, with screening conditions set at $p < 0.05$ and $|\log_2FC| > 0.585$. In total, 3383 differential genes were identified, comprising 1361 upregulated and 2022 downregulated genes (Fig. 1A,B). Subsequently, lysine-related genes were extracted from the GeneCards database and intersected with the identified dif-

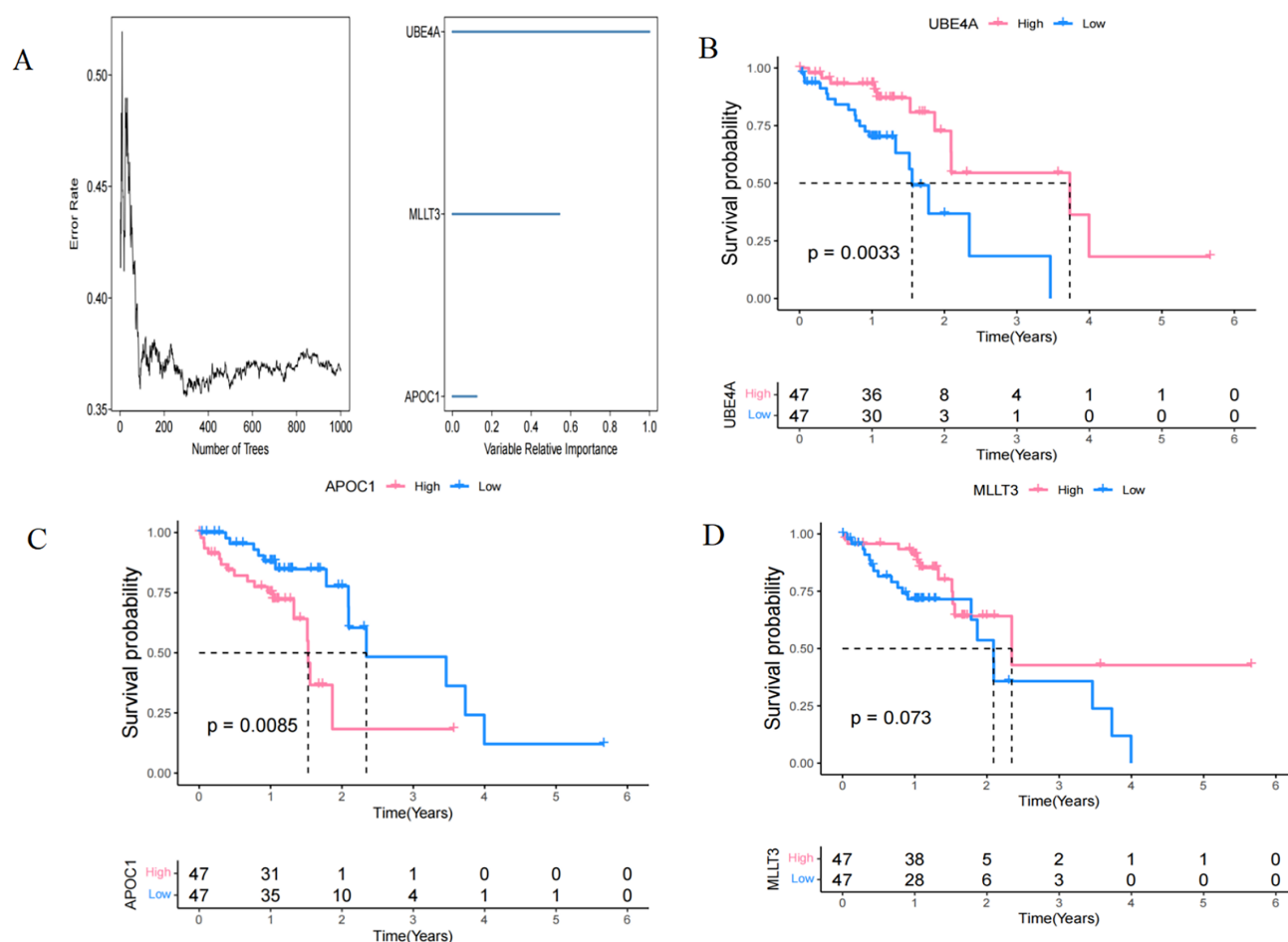


Fig. 2. Random survival forest analysis. (A) Random survival forest analysis identified three genes with relative importance >0.1 as final markers. (B–D) Kaplan-Meier survival analysis of the three marker genes in esophageal squamous cell carcinoma. TCGA data, including three normal esophageal tissues and 96 primary esophageal squamous cell carcinoma tissues, were used for analysis. *UBE4A*, Ubiquitination Factor E4A; *APOC1*, Apolipoprotein C1; *MLLT3*, MLLT3 Super Elongation Complex Subunit.

ferential genes, resulting in a total of 84 intersection genes (Fig. 1C). Further pathway analysis was conducted on the intersection genes.

GO enrichment analysis revealed that the intersection genes predominantly enriched in pathways related to histone modification, protein polyubiquitination, and ubiquitin-like protein transferase activity, among others (Fig. 1D). KEGG analysis exhibited that intersection genes were predominantly enriched in lysine degradation, ubiquitin mediated proteolysis, polycomb repressive complex and other pathways (Fig. 1E).

Screening of Important Prognostic Genes by Random Survival Forest Algorithm

To comprehensively identify key genes impacting ESCC, we conducted random survival forest (RSF) analysis on the set of 84 intersecting genes. Genes with relative importance >0.1 were considered final markers. Fig. 2A displays the importance ranking of three selected genes.

Subsequently, survival analysis was performed for these three genes. It was observed that the survival difference between the high and low expression groups of Ubiquitination Factor E4A (*UBE4A*) and Apolipoprotein C1 (*APOC1*), based on the median values as the cut-off point, was statistically significant (Fig. 2B,C), while no significant difference was observed for MLLT3 Super Elongation Complex Subunit (*MLLT3*) (Fig. 2D). Hence, *UBE4A* and *APOC1* were identified as key genes for further investigation in our study.

Amino Acid Metabolic Pathway Analysis

Utilizing the 84 lysine-associated genes as the background set, we used the single-sample Gene Set Enrichment Analysis (ssGSEA) method to quantify the lysine score of each sample. Subsequently, samples were stratified into high and low lysine groups based on the median score value. We quantified lysine levels by ssGSEA and further analyzed the discrepancies in metabolic pathways between high and low lysine levels.



Fig. 3. Lysine metabolic pathway analysis. (A) Lysine levels were quantified by single-sample Gene Set Enrichment Analysis (ss-GSEA) to analyze differences in metabolic pathways between high and low lysine levels. (B) Correlation between the two differential genes and amino acid metabolism (* $p < 0.05$, ** $p < 0.01$).

Our findings reveal that patients with elevated lysine metabolism may exhibit enhanced activity in amino acid metabolism pathways (Fig. 3A). Furthermore, a detailed analysis of the correlation between two DEGs and AAM demonstrated a significant association of *UBE4A* and *APOC1* with lysine degradation, remethylation, cysteine and methionine metabolism, as well as valine, leucine, and isoleucine biosynthesis pathways (Fig. 3B). These results suggest that these two key genes may influence the onset and progression of esophageal squamous cell carcinoma by modulating lysine metabolism.

GSEA and GSA

To further investigate the potential molecular mechanisms underlying the influence of key genes on ESCC development, we conducted a GSEA pathway analysis. Our findings revealed that *UBE4A*-enriched pathways included the chemokine signaling pathway, Forkhead box O (FoxO) signaling pathway, and NF-kappa B signaling pathway (Fig. 4A). On the other hand, pathways enriched by *APOC1* included phagosome, protein export, and ribosome (Fig. 4B). Additionally, GSA results demonstrated that *UBE4A* exhibited elevated expression primarily in the MTORC1 SIGNALING and phosphatidylinositol 3 kinase-protein kinase B-mammalian target of rapamycin (PI3K-AKT MTOR) signaling pathways (Fig. 4C). Conversely, el-

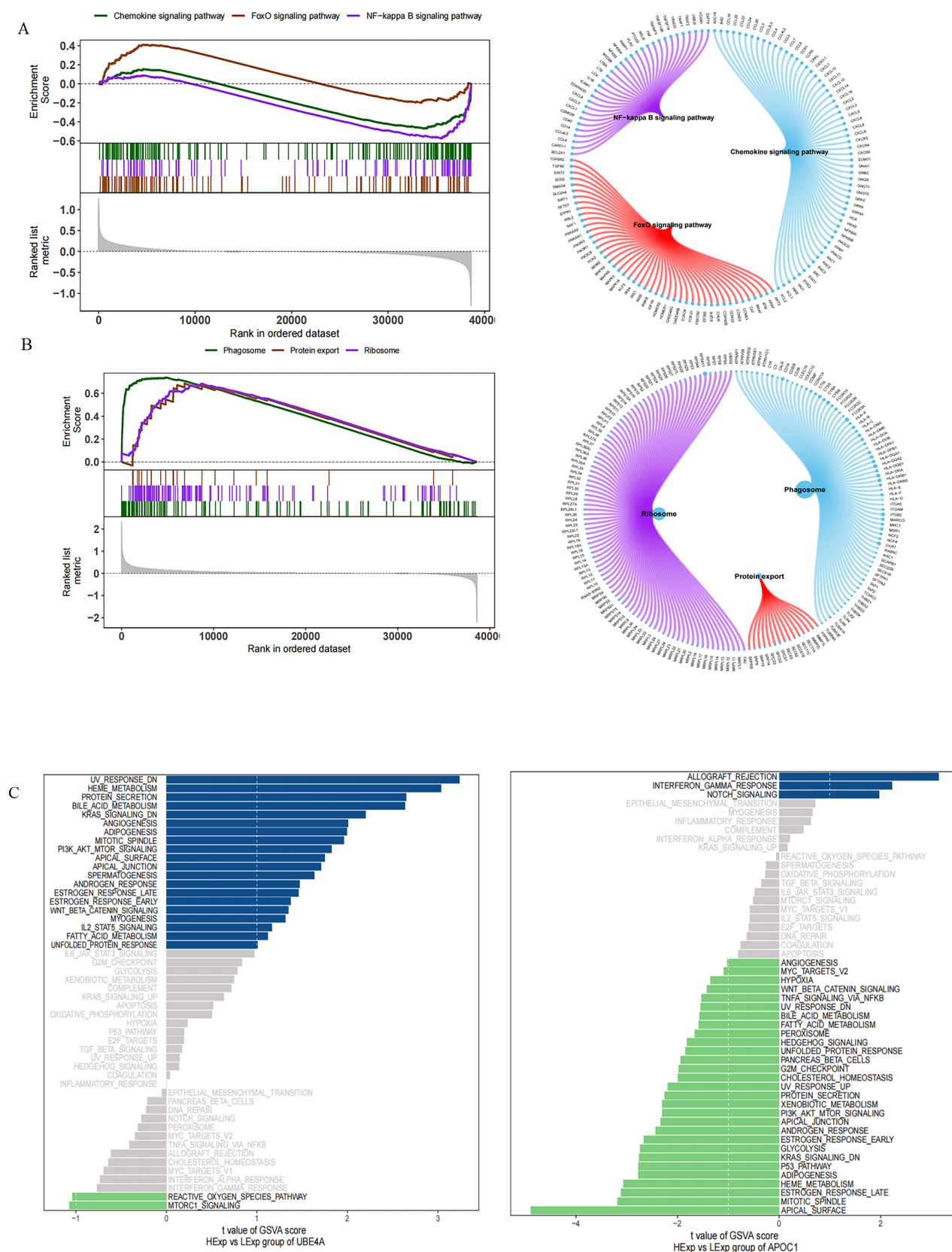


Fig. 4. Gene Set Enrichment Analysis (GSEA) pathway enrichment analysis and Gene Set Variation Analysis (GSVA). (A) GSEA pathway enrichment analysis revealed the main enrichment pathway of the Ubiquitination Factor E4A (*UBE4A*) gene. (B) GSEA pathway enrichment analysis revealed the main enrichment pathway of the Apolipoprotein C1 (*APOC1*) gene. (C) GSVA analyzed the signaling pathways for the high expression of two key genes (*UBE4A* and *APOC1*).

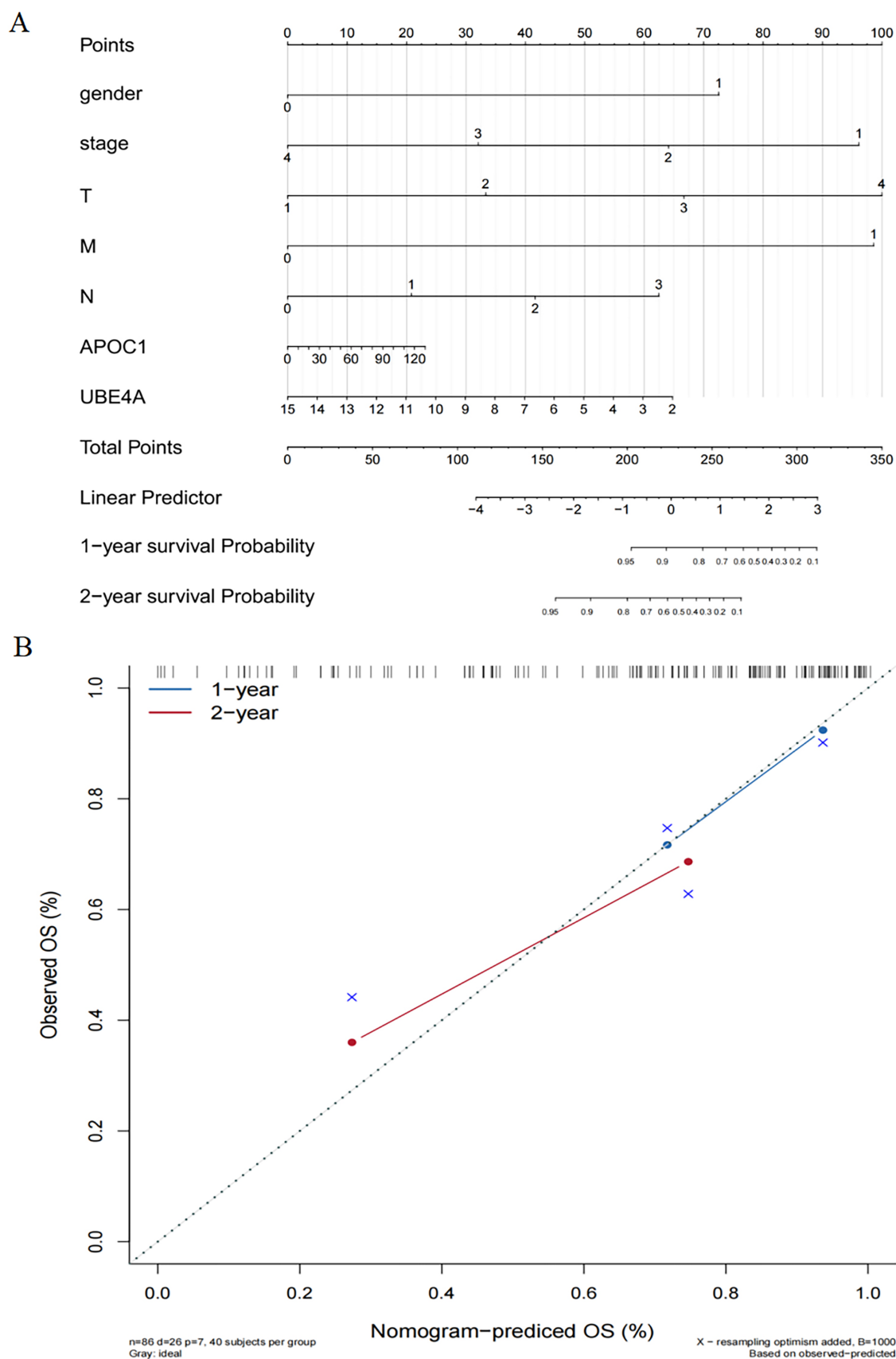


Fig. 5. Nomogram prognostic model construction. (A) Predicted 1- and 2-year survival rates of esophageal squamous cell carcinoma patients based on the prognostic nomogram constructed using lysine-related genes and clinicopathological parameters. (B) Calibration curves showing the concordance between predicted and actual 1- and 2-year survival rates of patients. OS, overall survival.

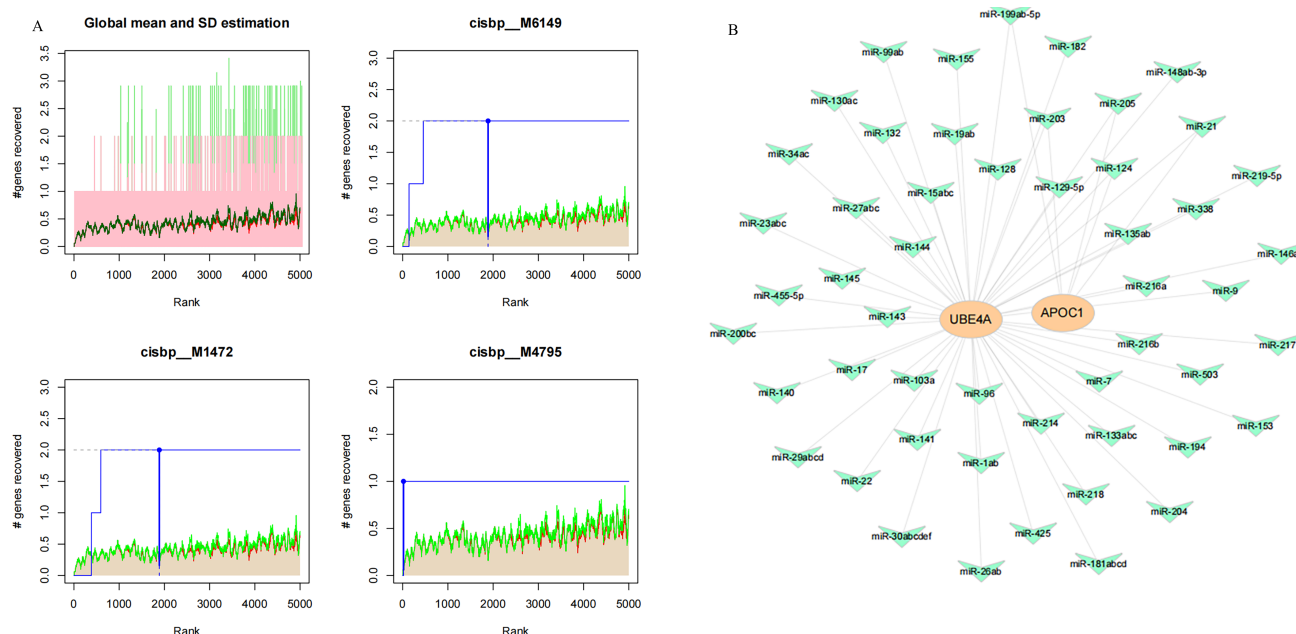


Fig. 6. Transcriptional regulation analysis of two lysine-related genes and construction of microRNA (miRNA) networks. (A) Transcription factor prediction. (B) miRNA network construction.

evated expression of *APOC1* was enriched in the Interferon-Gamma Response and P53 pathways (Fig. 4C). These findings suggest that these two key genes may exert their influence on ESCC development through the modulation of these pathways.

Nomogram Prognostic Model Construction

A nomogram was constructed using the two lysine-related genes, clinical stage, Tumor Node Metastasis (TNM) stage, and gender, to accurately estimate the 1- and 2-year survival probabilities of esophageal squamous carcinoma patients. Regression analysis was used to visualize the contribution of disparate clinical indices of ESCC and the expression distributions of *UBE4A* and *APOC1* in scoring, depicted in a column graph (Fig. 5A). Synchronously, prediction analysis was conducted for one year and two years (Fig. 5B). It was shown that the predicted overall survival (OS) was consistent with the observed OS. Additionally, the nomogram model demonstrated excellent predictive performance.

Transcriptional Regulation Analysis of the Two Lysine-Related Genes and Construction of miRNA Networks

The two lysine-related genes were utilized as the gene sets for transcriptional regulation analysis. It was observed that these genes were regulated by various mechanisms, including multiple transcription factors, which were enriched through cumulative curves, motif-TF annotation, and identification of key genes. Notably, the motif with the highest normalization score (NES: 9.56) was cisbp M6149. As

depicted in Fig. 6A, we presented the total enriched motifs and their relevant transcription factors. Additionally, reverse prediction of important genes was conducted using the miRcode database, identifying 51 miRNAs and a total of 56 mRNA-miRNA relation pairs. They were visualized using Cytoscape (Fig. 6B).

The Correlation between Two Lysine-Related Genes and Disease Genes in Esophageal Squamous Cell Carcinoma

We utilized the GeneCards database to retrieve genes associated with esophageal squamous carcinoma. Gene expression levels with high relevance scores were analyzed, and differences in expression of disease genes between groups were examined. Our analysis revealed significant differences in the expression levels of *RHBDF2*, *TGFBR2*, *PVT1*, *CASC9*, and *LZTS1* between the two groups. Furthermore, we observed a strong association between the expression levels of the two important genes and the expression levels of some ESCC-related genes. Primarily, *UBE4A* exhibited a significant positive correlation with SMAD Family Member 4 (*SMAD4*) (Pearson $r = 0.364$), while *APOC1* displayed a significant negative correlation with Epidermal Growth Factor Receptor (*EGFR*) (Pearson $r = -0.332$) (Fig. 7).

Sensitivity Analysis of Chemotherapy Drugs

Early-stage ESCC often responds well to treatment with surgery combined with chemotherapy. Leveraging drug susceptibility information from the Genomics of Drug Sensitivity in Cancer GDSC database, we employed the

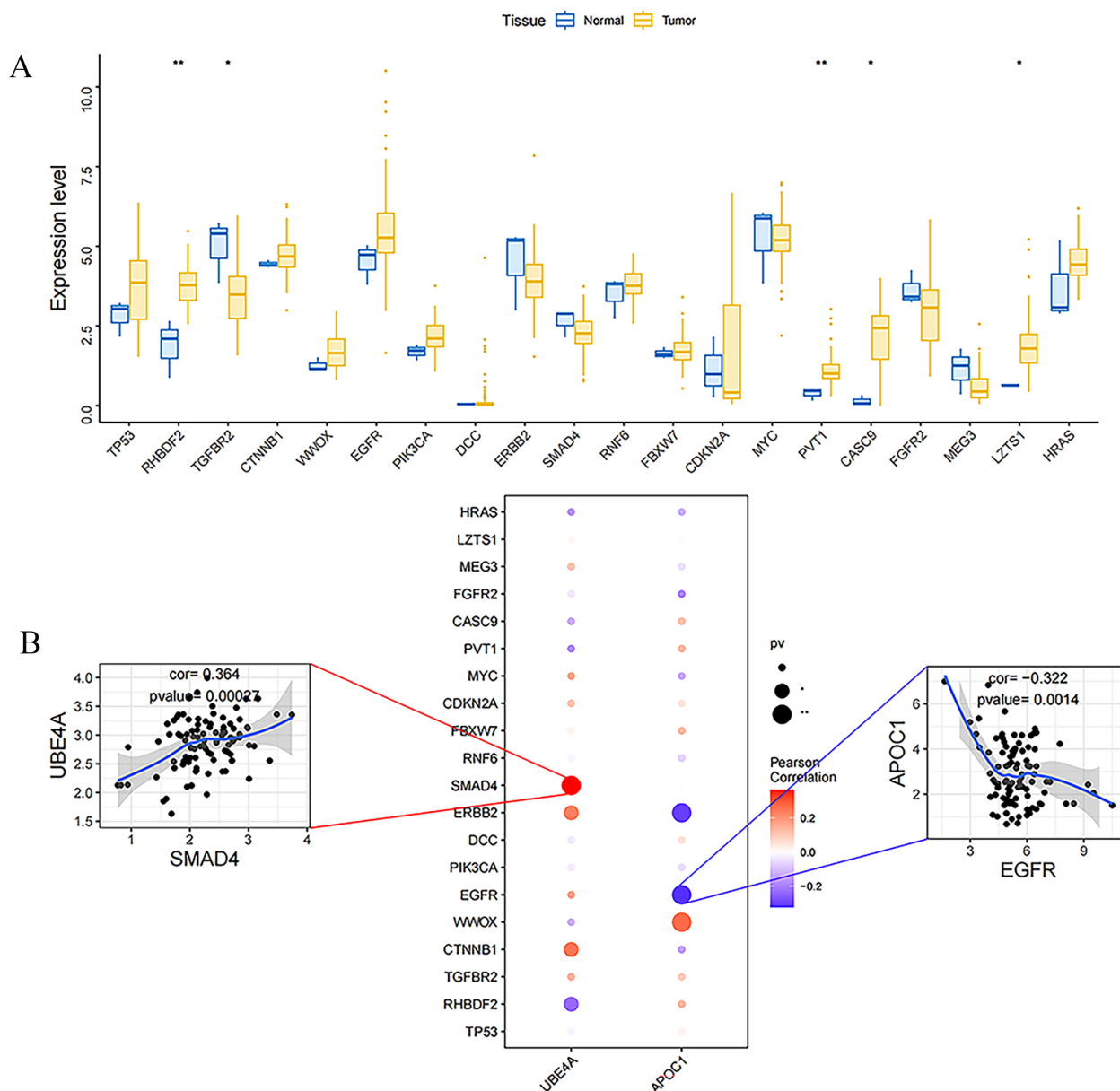


Fig. 7. Correlation between two lysine-related genes and disease genes in esophageal squamous cell carcinoma. (A) Utilizing the GeneCards database to obtain esophageal squamous carcinoma-related disease genes, expressions of *RHBDF2*, *TGFB2*, *PVT1*, *CASC9*, and *LZTS1* were significantly different in the two groups. (B) The expression levels of two key genes were significantly positively correlated: *UBE4A* and *SMAD4*. *APOC1* and *EGFR* were significantly negatively correlated ($*p < 0.05$, $**p < 0.01$). *SMAD4*, SMAD Family Member 4; *EGFR*, Epidermal Growth Factor Receptor.

“pRRophetic” in R to predict the chemotherapy susceptibility of each tumor specimen. The analysis investigated the susceptibility of important genes and chemotherapy drugs in depth.

Our findings revealed a significant association between gene expression levels and the sensitivity of specific chemotherapy drugs. Specifically, *UBE4A* was significantly linked to the sensitivity of *AICAR*, *BI.D1870*, and *GPP.082996*, while *APOC1* was closely associated with the sensitivity of *MS.275*, *lapatinib*, and *c-Jun N-terminal kinases Inhibitor VIII (JNk. inhibitor VIII)* (Fig. 8A,B).

These findings suggest the potential for developing novel chemotherapy regimens for ESCC patients who are resistant to chemotherapy.

The Correlation between Two Lysine-Related Genes and Immune Cells

The tumor microenvironment consists predominantly of tumor-associated fibroblasts, extracellular matrix components, immune cells, inflammatory factors, various growth factors, and unique physicochemical features, as well as cancer cells themselves. This microenvironment

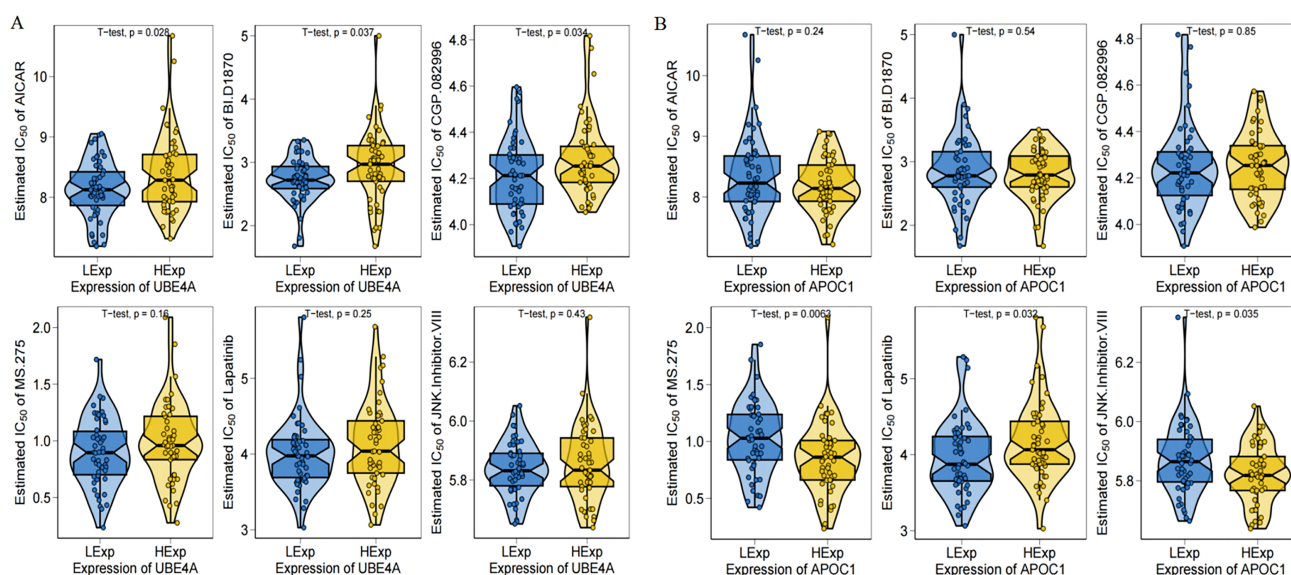


Fig. 8. Sensitivity analysis of chemotherapy drugs. (A) Sensitive chemotherapeutic agents of *UBE4A*. (B) Sensitive chemotherapeutic agents of *APOC1*. IC_{50} , Half maximal inhibitory concentration; Jnk, c-Jun N-terminal kinases.

significantly impacts tumor survival outcomes and clinical therapy sensitivity. In this study, we analyzed the correlation between lysine-related genes and immune cell infiltration in ESCC dataset.

The analysis revealed the distribution of immune cells in each patient and illustrated the relationship between different immune cell types (Fig. 9A,B). Notably, T cells follicular helper, macrophages M0, and M1 were markedly elevated in ESCC patients compared to the normal group (Fig. 9C). Further examination of the relationship between immune cells and lysine-related genes unveiled significant associations. Specifically, *UBE4A* was significantly positively associated with T cells CD4 memory resting and negatively correlated with B cells memory (Fig. 9D). On the other hand, *APOC1* was proactively associated with macrophages M2, T cells CD4 memory activated, and T cells CD8, while significantly passively connected with activated dendritic cells, T cells CD4 memory resting, and Eosinophils (Fig. 9E).

Moreover, we investigated the correlation between various immune factors and key genes using data from the TISIDB (<http://cis.hku.hk/TISIDB/>), including receptors, immunosuppressors, chemokines, and immunostimulators. These analyses suggest the strong association between lysine-related genes and immune cell infiltrating levels, highlighting their pivotal role in the immune microenvironment (Fig. 10).

Analysis of the Expression Patterns of Hub Genes and Specific Cell Subtypes in Single Cells

The single-cell dataset GSE196756 was analyzed using the “Seurat package” (<https://github.com/satijalab/seurat/issues>). The tSNE algorithm was employed to cluster the cells, identifying 21 subtypes (Fig. 11A). Subsequently,

each cluster was annotated using the “SingleR” package in R (version 4.2.2, <https://bioconductor.org/packages/release/bioc/html/SingleR.html>). All clusters were annotated to represent various cell types, including T cells, NK cells, B cells, neutrophils, monocyte, smooth muscle cells, epithelial cells, endothelial cells, tissue stem cells, and neurons (Fig. 11B). Furthermore, we analyzed the expression of *UBE4A* and *APOC1* across different cell types, including T cells, NK cells, B cells, neutrophils, monocyte, smooth muscle cells, epithelial cells, endothelial cells, tissue stem cells, and neurons (Fig. 11C,D). This analysis provided insights into the expression patterns of these hub genes across specific cell subtypes, shedding light on their potential roles in cellular processes within the tumor microenvironment.

Discussion

The energy metabolism of tumor cells regulates cell proliferation, progression, invasion and metastasis, significantly influencing tumor onset and progression. Targeting tumor metabolism has emerged as a promising strategy for combating tumor onset and progression. Although research on tumor metabolism dates back to the mid-20th century, progress in developing targeted therapies remains slow, with only a few metabolically-based antitumor drugs currently in clinical trials. Therefore, it is of utmost importance to investigate the underlying mechanisms of tumor metabolism and develop corresponding targeted therapies for tumor treatment.

Amino acids are the fundamental building blocks for proteins, energy sources, and metabolic substrates for cellular activities. Numerous studies have demonstrated that amino acid metabolism is closely associated with cancer onset, development, and drug resistance [10,11]. Currently,

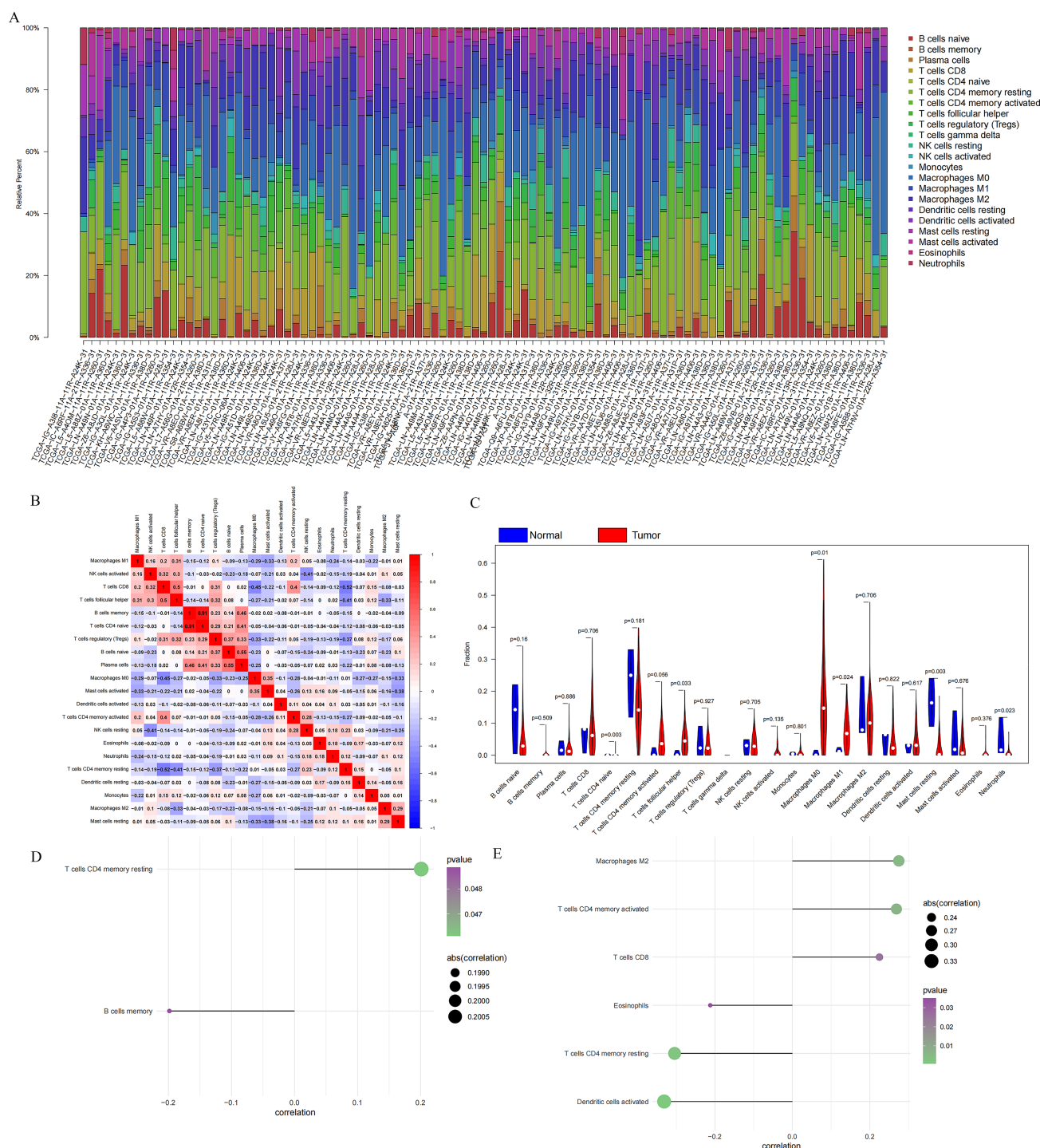
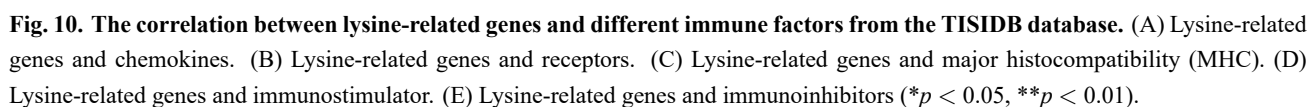


Fig. 9. The correlation between two lysine-related genes and immune cells. (A,B) Correlation between immune cell content and proportion of patients with esophageal squamous cell carcinoma from the TCGA database. (C) Comparison of infiltration ratio of immune cells from the TCGA database for esophageal squamous cell carcinoma tissue and normal tissue. (D,E) Correlation between lysine-related genes and immune cells.

research on metabolic abnormalities in tumor cells primarily focuses on two aspects: abnormal glucose metabolism and abnormal amino acid metabolism. Glucose metabolism in tumors involves converting a significant amount of pyruvate, produced by glucose glycolysis, into lactic acid via the lactate dehydrogenase complex, a phenomenon known

as the Warburg effect [12–16]. However, the rapid proliferation of tumor cells necessitates a substantial supply of carbon scaffolding for synthesizing macromolecules such as cell membranes and nucleic acids. The intermediate metabolites generated by the Warburg effect alone are insufficient to fulfill these requirements.



works concurrent with changes in their glucose metabolic pathways. Tumor cells significantly depend on glutamate,

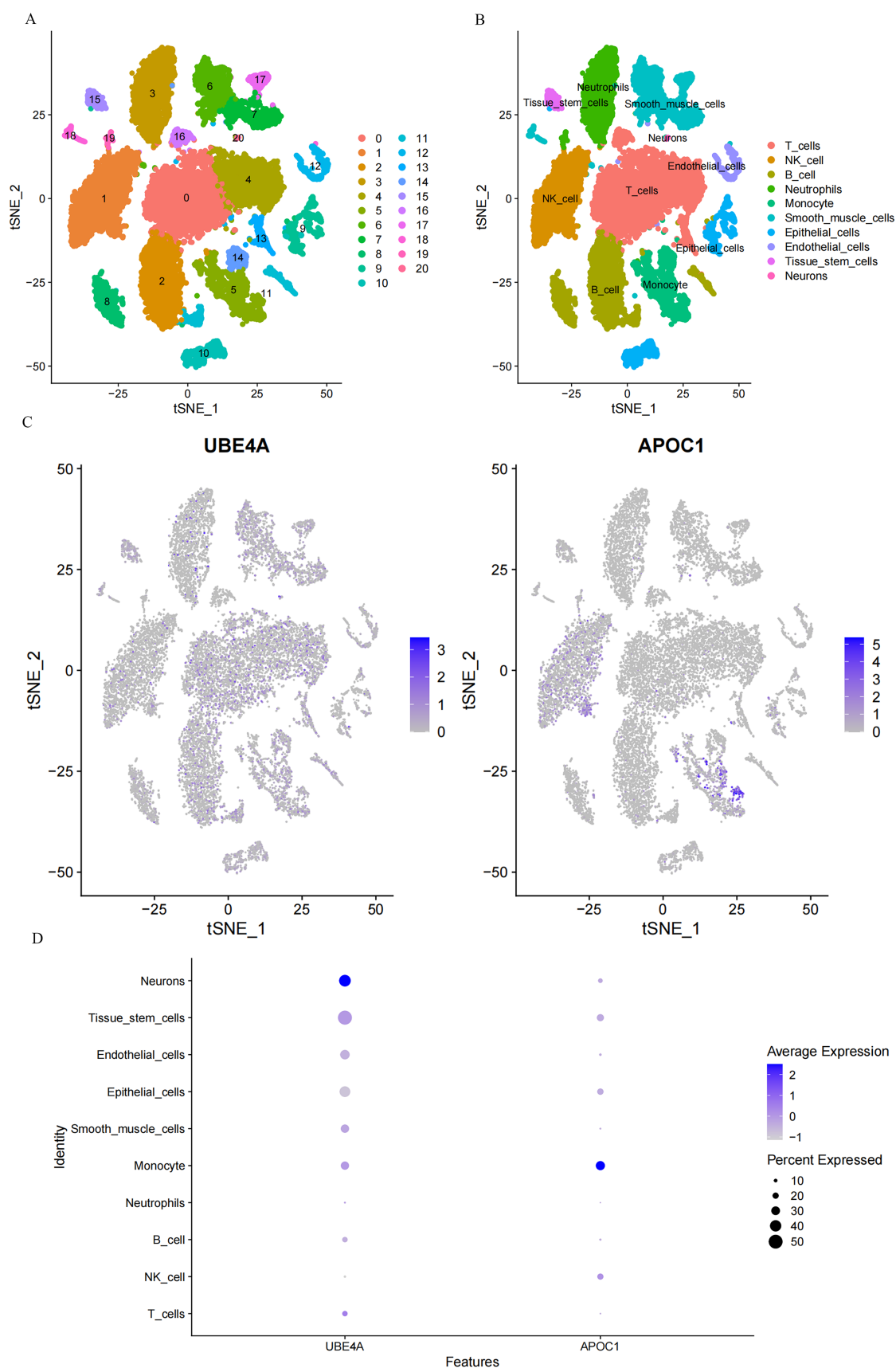


Fig. 11. Expression of lysine-related genes in scRNA-seq. (A) Twenty-one subtypes were obtained by t-Distributed Stochastic Neighbor Embedding (tSNE) and the annotation. (B–D) Expressions of *UBE4A* and *APOC1* genes in T cells, NK cells, B cells, neutrophils, monocytes, smooth muscle cells, epithelial cells, endothelial cells, and tissues.

emphasizing the importance of amino acid metabolism in the tumor onset and progression. This phenomenon, known as amino acid metabolic reprogramming, encompasses abnormal amino acid uptake rates, alterations in amino acid metabolic pathways, changes in key enzyme activities, and fluctuations in metabolite levels within cancer. Amino acids in tumor cells directly participate in protein synthesis and interact with other essential metabolic pathways crucial for tumorigenesis through catabolites.

A growing body of evidence indicates that amino acid metabolic reprogramming is implicated in tumor progression across various cancers, including hepatocellular carcinoma, ovarian cancer, breast cancer, and renal clear cell carcinoma [17–21]. Moreover, researchers have elucidated the crucial role of specific amino acids, such as arginine and tryptophan, in modulating immune responses, particularly in the activation and differentiation of T cells [22]. Study by Yuan H *et al.* [9] have shed light on the ability of glioblastoma stem cells to reprogram lysine metabolism, leading to the overproduction of crotonyl-coA. This metabolic alteration promotes overall crotonylation modification of cells, affecting interferon signaling, inhibiting CD8⁺ T cell infiltration, and ultimately suppressing tumor immune responses while fostering tumor growth.

Furthermore, the deficiency of specific amino acids in the tumor microenvironment (TME) or the production of amino acid metabolites also influences the functional maintenance of immune cells. This underscores the dual role of amino acids in providing energy for tumor cell proliferation and indirectly modulating cancer immunity by regulating the immune cell function.

In our research, comprehensive bioinformatic analyses were conducted to identify potential prognostic lysine-related genes with therapeutic implications for ESCC patients. Initially, a total of 3383 DEGs were identified from TCGA datasets, encompassing 1361 upregulated genes and 2022 downregulated genes. Subsequently, lysine-related genes were extracted from the GeneCards database and intersected with the DEGs, resulting in 84 interacting genes. Gene Ontology (GO) analysis revealed that these 84 lysine-related genes were significantly enriched in processes related to histone modification and protein polyubiquitination. Moreover, KEGG pathway analysis highlighted enrichment in pathways such as lysine degradation and ubiquitin-mediated proteolysis. These findings suggest potential connections between the identified DEGs and amino acid metabolism in ESCC, potentially influencing prognosis and treatment outcomes.

To identify lysine-related genes associated with poor prognosis in ESCC, a random survival forest analysis was conducted on the 84 lysine-related genes, leading to the identification of three key genes (*MLLT3*, *UBE4A*, and *APOC1*). Survival analysis demonstrated that *UBE4A* and *APOC1* were significantly correlated with the prognosis of ESCC. Further analysis involved quantifying lysine levels

using ssGSEA and assessing differences in metabolic pathways between samples with high and low lysine levels. It was observed that patients with high lysine metabolism exhibited enhanced activity in amino acid metabolism, lipid metabolism, and drug metabolism pathways.

Moreover, correlation analysis between AAM and the two identified differential genes (*UBE4A* and *APOC1*) revealed significant associations with lysine degradation, cysteine and methionine metabolism, and the biosynthesis of valine, leucine and isoleucine. These findings suggest potential roles for *UBE4A* and *APOC1* in lysine metabolism and drug metabolism processes in ESCC.

Amino acids play multifaceted roles in various processes via amino acid transporters, sensors, enzymes, and intermediates produced by their metabolism. They can also modulate the tumor microenvironment and intracellular signaling pathways, regulating the growth of tumor cells. Current studies have revealed that mTOR exerts its biological effects in tumor cells by interacting with other molecules, forming two complexes known as mTORC1 and mTORC2. Various amino acid transporters are crucial for promoting cell proliferation and activating mTORC1 in tumor cells. For instance, Linker for activation of T-cells family member 1 (LAT1) enhances cancer cell proliferation *in vitro* and *in vivo* by activating mTORC1. Conversely, the knockout of LAT1 disrupts tumor amino acids homeostasis and reduces mTORC1 activity, thereby inhibiting the progression of xenograft tumors in mice [23].

Another essential component, the cytosolic arginine sensor for mTORC1 subunit 2 Gene (GATOR2) complex, positively regulates the mTORC1 sensing amino acid pathway. Deficiency in leucine enhances the interaction between stress-inducing protein Sestrin2 and GATOR2, thereby impacting mTORC1 expression and affecting the onset and progression of colorectal cancer [24]. In this study, to further explore the molecular mechanisms underlying the influence of two lysine-related genes on ESCC development, GSEA revealed that *UBE4A* was enriched in pathways including the FoxO signaling pathway, chemokine signaling pathway, and NF-kappa B signaling pathway, while *APOC1* was associated with pathways such as phagosome, protein export, and ribosome. Gene Set Variation Analysis (GSVA) results indicated that *UBE4A* was highly expressed in mTORC1 signaling pathways and PI3K-AKT MTOR signaling pathways, whereas high *APOC1* expression was enriched in INTERFERON GAMMA RESPONSE and P53 signaling pathways. Therefore, these two genes may influence lysine metabolism through signaling pathways within tumor cells, impacting ESCC prognosis.

Certainly, one nomogram has emerged as a reliable clinical tool for assessing the survival of cancer patients. Currently, the clinical prognostic evaluation of ESCC patients primarily relies on their clinical stage. Therefore, a robust nomogram was developed, incorporating two lysine-

related genes, clinical stage, TNM stage, and gender, to improve the prognosis prediction for ESCC patients. Through calibration graphs, it was demonstrated that the predicted 1-year and 2-year survival rates based on the nomogram closely matched the actual rates, indicating the high predictive accuracy of the constructed nomogram model.

Drug resistance poses a significant challenge in tumor therapy, and resistance to chemotherapy or targeted therapy is a major cause of tumor recurrence and metastasis. Study has shown that *LAT2*, a neutral amino acid transporter independent of Na^+ , regulates glutamine-dependent mTOR activation, thereby promoting glycolysis and reducing gemcitabine sensitivity [25]. Additionally, P53 enhances arginine uptake by upregulating the expression of Solute Carrier Family 7 Member 3 (*SLC7A3*), thus promoting glutamine deprivation by cancer cells [26]. *SLC7A3* expression has also been associated with drug resistance in various cancers, with decreased expression observed in drug-resistant ovarian cancer cell lines [27]. In our signature gene analysis, we identified *UBE4A* and *APOC1*, targeting lysine-related genes, as having therapeutic potential for ESCC. A search of drug databases revealed that agents such as halofuginone, *AICAR*, lapatinib, and JNK inhibitor VIII act on *UBE4A* and *APOC1*.

AAM regulates the functions of various immune cell categories, upon which these cells depend for obtaining biological matrix and energy, and maintaining their effector functions by activating intermediate metabolites and key enzymes in metabolism [28]. Study has shown that bone marrow cells, as major components of the cancerous microenvironment, promote tumor growth and immune evasion by significantly increasing the level of inflammatory tumor-associated macrophages (TAM) by inhibiting glutamine metabolism, thereby reducing tumor growth and metastasis [29]. Asparagine, for instance, directly binds to lymphocyte-specific protein tyrosine kinase (LCK) to enhance T cell activation and LCK signaling, consequently enhancing CD8⁺ T cell antitumor response [30]. Additionally, immune cell proportions were analyzed in ESCC patients from the TCGA dataset and the correlation between immune cells. It was observed that T cells follicular helper, macrophages M0 and M1 were notably elevated in ESCC patients compared to the normal group. Further examination of the correlation between lysine-related genes and immune cells revealed that *UBE4A* was significantly positively associated with T cells CD4 memory resting and negatively associated with B cells memory, while *APOC1* was positively linked to macrophages M2, T cells CD4 memory activated, and T cells CD8, but exhibited significantly negative correlation with activated dendritic cells, T cells CD4 memory resting, and Eosinophils. Accordingly, B cells, T cells and macrophages are the key players in the onset and progression of ESCC.

In tumor cells, amino acids directly contribute to protein synthesis and participate in tumor cell growth as inter-

mediate products of energy metabolism. Moreover, they regulate cancer function by engaging in numerous bioprocesses via intracellular amino acid transporters, sensors, enzymes, and intermediates, thereby modulating intracellular signaling pathways or the tumor microenvironment to regulate tumor cell growth. Notably, amino acid metabolism plays a critical role in activating or differentiating immune cells and can impact cancer immunity by regulating immune cells and factors. These observations underscore the potential of targeting amino acid metabolism as a therapeutic strategy for treating tumors and highlight the therapeutic potential of regulating AAM to enhance tumor immunity.

Therefore, future research on AAM in tumors will increasingly focus on the diverse signaling pathways of AAM. In this study, we used bioinformatics analysis to identify lysine metabolism-related genes in esophageal squamous cell carcinoma. Furthermore, we conducted pathway and drug sensitivity analyses to establish a theoretical foundation for uncovering the role of amino acid metabolism in esophageal squamous cell cancer cells and identifying potential drug targets.

Conclusions

Our findings underscore the pivotal role of lysine-related gene regulation in ESCC. The identified lysine-related genes hold promise as prognostic indicators for ESCC patients, shedding light on potential mechanisms underlying poor prognosis and facilitating the search for targeted therapeutics to enhance the success rate of ESCC treatment.

Availability of Data and Materials

Anonymized data for our analyses presented in the present report are available upon request from the corresponding author.

Author Contributions

DHF conceived of the study, participated in its design and reviewed the manuscript. LPH designed the study, performed the data analysis and drafted the manuscript. BXC and YYL drafted the part of manuscript. BXC, ZJH and YYL provided help and advice on the data analysis. All authors contributed to important editorial changes in the manuscript. All authors provided critical feedback and helped shape the research, analysis and manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

Not applicable.

Funding

This study was supported by grants from the National Natural Science Foundation of China (Nos.81402738); Research topic of Zhangzhou Traditional Chinese medicine (2023ZY02); Fujian Province young and middle-aged teacher education research project (JAT231262); Fujian Province young and middle-aged teacher education research project (JJ230086); Study on natural products and activities (kjcX-08).

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: a Cancer Journal for Clinicians*. 2021; 71: 209–249.
- [2] Lagergren J, Smyth E, Cunningham D, Lagergren P. Oesophageal cancer. *Lancet*. 2017; 390: 2383–2396.
- [3] Cantor JR, Sabatini DM. Cancer cell metabolism: one hallmark, many faces. *Cancer Discovery*. 2012; 2: 881–898.
- [4] Xiong Y, Lei QY, Zhao S, Guan KL. Regulation of glycolysis and gluconeogenesis by acetylation of PKM and PEPCK. *Cold Spring Harbor Symposia on Quantitative Biology*. 2011; 76: 285–289.
- [5] Li T, Liu M, Feng X, Wang Z, Das I, Xu Y, *et al.* Glyceraldehyde-3-phosphate dehydrogenase is activated by lysine 254 acetylation in response to glucose signal. *The Journal of Biological Chemistry*. 2014; 289: 3775–3785.
- [6] Wang W, Zou W. Amino Acids and Their Transporters in T Cell Immunity and Cancer Therapy. *Molecular Cell*. 2020; 80: 384–395.
- [7] Maddocks ODK, Athineos D, Cheung EC, Lee P, Zhang T, van den Broek NJF, *et al.* Modulating the therapeutic response of tumours to dietary serine and glycine starvation. *Nature*. 2017; 544: 372–376.
- [8] Gao X, Sanderson SM, Dai Z, Reid MA, Cooper DE, Lu M, *et al.* Dietary methionine influences therapy in mouse cancer models and alters human metabolism. *Nature*. 2019; 572: 397–401.
- [9] Yuan H, Wu X, Wu Q, Chatoff A, Megill E, Gao J, *et al.* Lysine catabolism reprograms tumour immunity through histone crotonylation. *Nature*. 2023; 617: 818–826.
- [10] Lukey MJ, Katt WP, Cerione RA. Targeting amino acid metabolism for cancer therapy. *Drug Discovery Today*. 2017; 22: 796–804.
- [11] Yoo HC, Han JM. Amino acid metabolism in cancer drug resistance. *Cells*. 2022; 11: 140.
- [12] Hosios AM, Vander Heiden MG. Acetate metabolism in cancer cells. *Cancer & Metabolism*. 2014; 2: 27.
- [13] Sadeghi RN, Karami-Tehrani F, Salami S. Targeting prostate cancer cell metabolism: impact of hexokinase and CPT-1 enzymes. *Tumour Biology*. 2015; 36: 2893–2905.
- [14] Gonzalez CD, Alvarez S, Ropolo A, Rosenzvit C, Bagnes MFG, Vaccaro MI. Autophagy, Warburg, and Warburg reverse effects in human cancer. *BioMed Research International*. 2014; 2014: 926729.
- [15] Dell'Antone P. Energy metabolism in cancer cells: how to explain the Warburg and Crabtree effects? *Medical Hypotheses*. 2012; 79: 388–392.
- [16] Bonuccelli G, Whitaker-Menezes D, Castello-Cros R, Pavlides S, Pestell RG, Fatatis A, *et al.* The reverse Warburg effect: glycolysis inhibitors prevent the tumor promoting effects of caveolin-1 deficient cancer associated fibroblasts. *Cell Cycle*. 2010; 9: 1960–1971.
- [17] Wang ZQ, Faddaoui A, Bachvarova M, Plante M, Gregoire J, Renaud MC, *et al.* BCAT1 expression associates with ovarian cancer progression: possible implications in altered disease metabolism. *Oncotarget*. 2015; 6: 31522–31543.
- [18] Erickson RE, Lim SL, McDonnell E, Shuen WH, Vadiveloo M, White PJ, *et al.* Loss of BCAA Catabolism during Carcinogenesis Enhances mTORC1 Activity and Promotes Tumor Development and Progression. *Cell Metabolism*. 2019; 29: 1151–1165.e6.
- [19] Qu YY, Zhao R, Zhang HL, Zhou Q, Xu FJ, Zhang X, *et al.* Inactivation of the AMPK-GATA3-ECHS1 Pathway Induces Fatty Acid Synthesis That Promotes Clear Cell Renal Cell Carcinoma Growth. *Cancer Research*. 2020; 80: 319–333.
- [20] Zhang L, Han J. Branched-chain amino acid transaminase 1 (BCAT1) promotes the growth of breast cancer cells through improving mTOR-mediated mitochondrial biogenesis and function. *Biochemical and Biophysical Research Communications*. 2017; 486: 224–231.
- [21] Martin SB, Reiche WS, Fifelski NA, Schultz AJ, Stanford SJ, Martin AA, *et al.* Leucine and branched-chain amino acid metabolism contribute to the growth of bone sarcomas by regulating AMPK and mTORC1 signaling. *The Biochemical Journal*. 2020; 477: 1579–1599.
- [22] Siska PJ, Rathmell JC. T cell metabolic fitness in antitumor immunity. *Trends in Immunology*. 2015; 36: 257–264.
- [23] Cormerais Y, Giuliano S, LeFloch R, Front B, Durivault J, Tambutti E, *et al.* Genetic Disruption of the Multifunctional CD98/LAT1 Complex Demonstrates the Key Role of Essential Amino Acid Transport in the Control of mTORC1 and Tumor Growth. *Cancer Research*. 2016; 76: 4481–4492.
- [24] Wolfson RL, Chantranupong L, Saxton RA, Shen K, Scaria SM, Cantor JR, *et al.* Sestrin2 is a leucine sensor for the mTORC1 pathway. *Science*. 2016; 351: 43–48.
- [25] Feng M, Xiong G, Cao Z, Yang G, Zheng S, Qiu J, *et al.* LAT2 regulates glutamine-dependent mTOR activation to promote glycolysis and chemoresistance in pancreatic cancer. *Journal of Experimental & Clinical Cancer Research*. 2018; 37: 274.
- [26] Lowman XH, Hanse EA, Yang Y, Ishak Gabra MB, Tran TQ, Li H, *et al.* p53 Promotes Cancer Cell Adaptation to Glutamine Deprivation by Upregulating Slc7a3 to Increase Arginine Uptake. *Cell Reports*. 2019; 26: 3051–3060.e4.
- [27] Januchowski R, Zawierucha P, Andrzejewska M, Ruciński M, Zabel M. Microarray-based detection and expression analysis of ABC and SLC transporters in drug-resistant ovarian cancer cell lines. *Biomedicine & Pharmacotherapy*. 2013; 67: 240–245.
- [28] Kelly B, Pearce EL. Amino Assets: How Amino Acids Support Immunity. *Cell Metabolism*. 2020; 32: 154–175.
- [29] Oh MH, Sun IH, Zhao L, Leone RD, Sun IM, Xu W, *et al.* Targeting glutamine metabolism enhances tumor-specific immunity by modulating suppressive myeloid cells. *The Journal of Clinical Investigation*. 2020; 130: 3865–3884.
- [30] Wu J, Li G, Li L, Li D, Dong Z, Jiang P. Asparagine enhances LCK signalling to potentiate CD8⁺ T-cell activation and anti-tumour responses. *Nature Cell Biology*. 2021; 23: 75–86.