scientific reports

Check for updates

OPEN

Bioinformatic analysis of glycolysis and lactate metabolism genes in head and neck squamous cell carcinoma

Huanyu Jiang[1,2], Lijuan Zhou[2], Haidong Zhang[1,2] & Zhenkun Yu[1,2]✉

Head and neck squamous cell carcinoma (HNSCC) is a heterogeneous cancer with significant global incidence. This study investigates glycolysis- and lactate metabolism-related genes (GALMRGs) in HNSCC, focusing on their impact on prognosis, the tumor immune microenvironment, and their potential as therapeutic biomarkers. Analysis of data from the Cancer Genome Atlas and Gene Expression Omnibus identified 16 GALMRGs that were differentially expressed in HNSCC compared to normal tissues. Functional analysis revealed the involvement of lactate and pyruvate metabolism and HIF-1 signaling pathways. Weighted gene co-expression network analysis identified two module genes, *CDKN3* and *SLC2A1*. Five key genes (*CAV1*, *CDKN3*, *LDHA*, *MB*, and *PER2*) were identified through univariate, multivariate, and LASSO regression analyses and used to construct a prognostic model. This model demonstrated strong predictive accuracy for overall survival, stratifying patients into high- and low-risk groups. Immune cell infiltration analysis showed a negative correlation between resting and activated mast cells, and low-risk patients had higher tumor mutational burden, suggesting a better response to immunotherapy. Consensus clustering classified HNSCC into two distinct molecular subtypes with differing expression of the key genes. This GALMRG-based prognostic model is a promising biomarker for predicting HNSCC outcomes and immunotherapy responses, providing valuable insights for personalized treatment strategies.

Keywords Head and neck squamous cell carcinoma, Glycolysis, Lactate metabolism, Tumor immune microenvironment, Prognostic risk model

Head and neck squamous cell carcinoma (HNSCC) is a prevalent malignancy of the nasal cavity, oral cavity, larynx, and pharynx. By 2030, the incidence of HNSCC is projected to increase by 30%, resulting in approximately 1.08 million new cases annually[1,2]. Despite advancements in surgery, radiotherapy, and chemotherapy, many patients continue to experience poor outcomes because of tumor heterogeneity and drug resistance. PD-1/PD-L1 immune checkpoint inhibitors (ICIs) show promise, but their efficacy varies among patients due to the complex interplay between tumor metabolic reprogramming and the tumor immune microenvironment (TME)[3,4].

Metabolic reprogramming, particularly in glycolysis and lactate metabolism, is vital for tumor growth and survival. Glycolysis supplies the energy and biosynthetic precursors required for rapid cell proliferation[5]. In HNSCC, glycolytic activity is elevated by key enzymes, such as pyruvate kinase M2, which is significantly upregulated and promotes tumor cell survival[6]. Even when oxygen is available, tumor cells predominantly use glycolysis instead of oxidative phosphorylation, a process termed the "Warburg effect," leading to elevated lactate production and TME acidification[7]. This acidic environment not only promotes tumor proliferation and metastasis but also inhibits tumor-infiltrating immune cells, facilitating immune evasion[8]. Although glycolysis and lactate metabolism are inherently interconnected, most HNSCC studies have examined these pathways separately, which may have failed to capture their combined influence on the TME and immune treatment response.

In this study, we aimed to create a comprehensive framework that integrates the roles of glycolysis and lactate metabolism in the TME, thereby addressing the limitations of previous studies. Through bioinformatics analyses, we analyzed the expression patterns and functional implications of genes involved in these pathways to determine their collective impact on tumor progression, immune evasion, and patient prognosis. This integrated

[1]School of Medicine, Southeast University, 87 Dingjiaqiao, Hunan Road, Nanjing 210009, Jiangsu, China. [2]Department of Otolaryngology Head and Neck Surgery, The Affiliated Benq Hospital of Nanjing Medical University, Nanjing 210019, Jiangsu, China. ✉email: yuzhenkun65@hotmail.com

perspective is expected to reveal novel therapeutic targets, identify potential drug target combinations that address both glycolysis and lactate metabolism, and lay the groundwork for personalized treatment strategies.

Results

Technical roadmap

The technical roadmap of this study is presented in Fig. 1.

Merging HNSCC datasets

Batch effects in the GSE107591, GSE6631, and GSE9638 datasets were removed using the sva R package, resulting in a combined Gene Expression Omnibus (GEO) dataset. Boxplots (Supplementary Fig. S1a, b) were used to visualize expression differences before and after correction, while principal component analysis (PCA) plots (Supplementary Fig. S1c, d) showed a better distribution of low-dimensional features. Both visualizations indicate the successful elimination of batch effects.

Differential expression, somatic mutation (SM), and copy number variation (CNV) analyses of glycolysis and lactate metabolism-related genes (GALMRGs) in HNSCC

Using the R package limma for differential expression analysis of The Cancer Genome Atlas (TCGA) dataset revealed 4,288 differentially expressed genes (DEGs) ($|\log FC| > 0.5$, adj.$P < 0.05$), comprising 2,972 upregulated and 1,316 downregulated genes. Figure 2a shows a volcano plot of the DEGs. The 16 glycolysis and lactate metabolism-related differentially expressed genes (GALMRDEGs) were *ABCC1*, *ADRB2*, *AKR1B1*, *BSG*, *CAV1*, *CDKN3*, *FOXM1*, *GPT*, *HIF1A*, *IL1B*, *LDHA*, *LDHB*, *MB*, *PC*, *PER2*, and *SLC2A1* (Fig. 2b). The differential expression of the genes was visualized using a heatmap (Fig. 2c), and chromosomal locations were mapped in a Circos plot (Fig. 2d), revealing that these genes were located on chromosomes 7, 11, 12, and 14.

We analyzed the SMs in the 16 GALMRDEGs using the R package maftools, which revealed nine main types of mutations, with missense mutations being the most prevalent (Fig. 3a). Single nucleotide polymorphisms were the most common, particularly the C > T transition, while *PC* showed the highest mutation rate, at 2%. (Fig. 3b). We analyzed CNVs using GISTIC 2.0 (Broad Institute, Cambridge, MA, USA), which identified CNVs

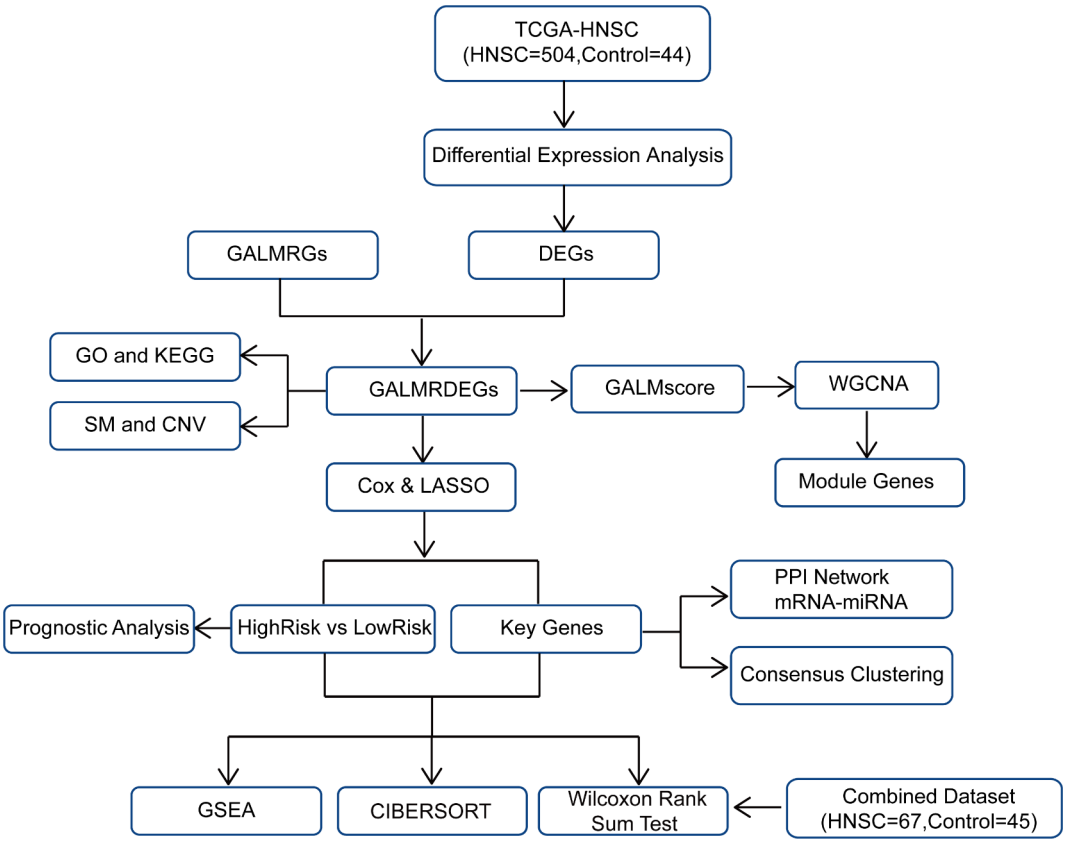

Fig. 1. Flow chart for the comprehensive analysis of GALMRGs. TCGA: The Cancer Genome Atlas, HNSC: Head and neck squamous cell carcinoma, DEGs: Differentially expressed genes, GALMRGs: Glycolysis and lactate metabolism-related genes, GALMRDEGs: Glycolysis and lactate metabolism-related differentially expressed genes, CNV: Copy number variation, SM: Somatic mutation, GALM.Score: Glycolysis and lactate metabolism score, WGCNA: Weighted correlation network analysis, LASSO: Least absolute shrinkage and selection operator, PPI network: Protein-protein interaction network, GO: Gene Ontology, KEGG: Kyoto Encyclopedia of Genes and Genomes, GSEA: Gene set enrichment analysis.

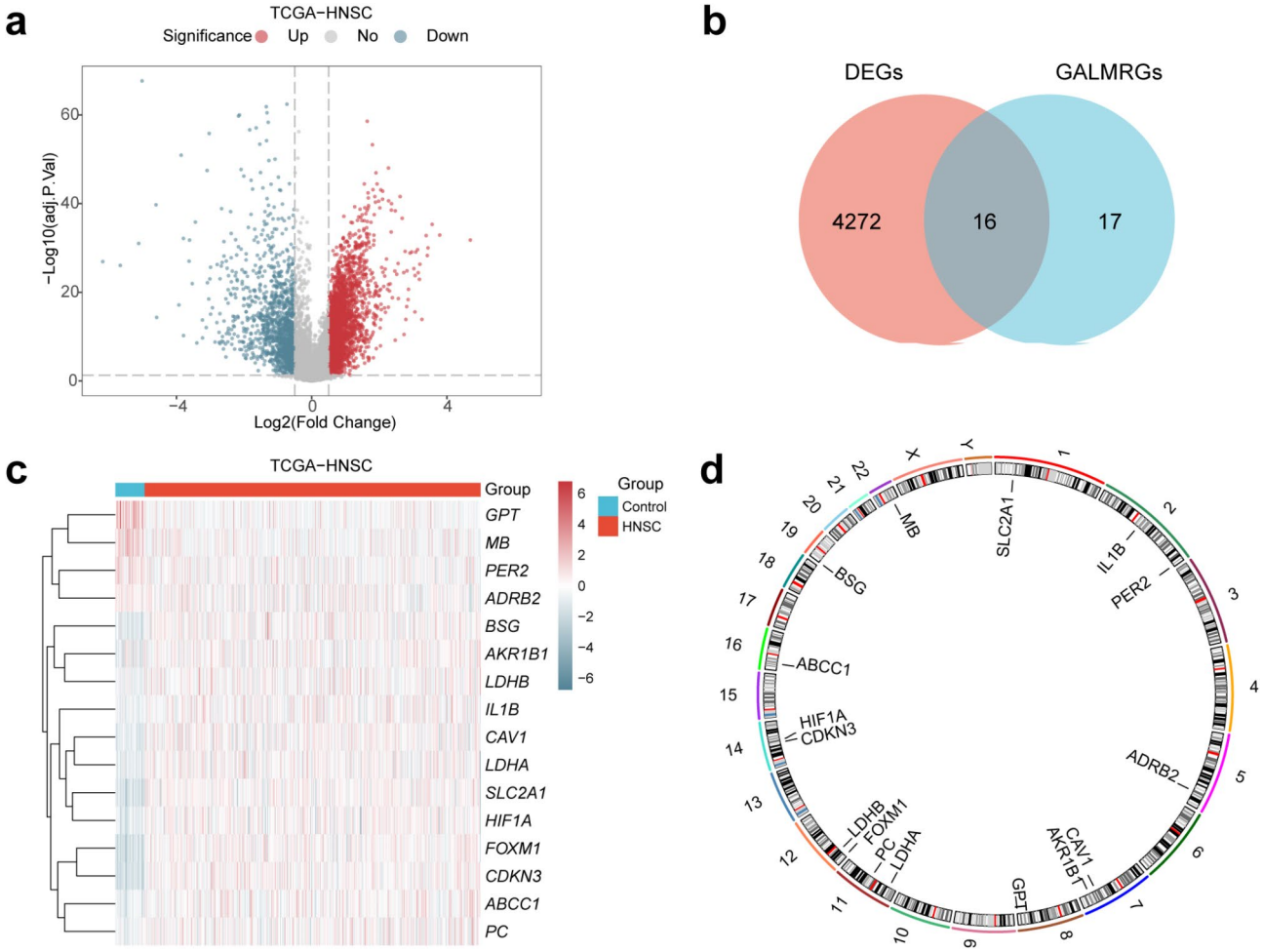

Fig. 2. Differential expression analysis of GALMRGs in TCGA-HNSC. (**a**) Volcano plot of the DEG analysis between the HNSCC and control groups. (**b**) DEG and GALMRG Venn diagram in TCGA-HNSC. (**c**) Heatmap of GALMRDEGs in TCGA-HNSC. (**d**) Chromosomal mapping of GALMRDEGs. TCGA: The Cancer Genome Atlas, HNSC: Head and neck squamous cell carcinomas, DEGs: Differentially expressed genes, GALMRGs: Glycolysis and lactate metabolism-related genes, GALMRDEGs: Glycolysis and lactate metabolism-related differentially expressed genes.

in 15 of the 16 GALMRDEGs (Fig. 3c, d): *ABCC1*, *ADRB2*, *AKR1B1*, *BSG*, *CAV1*, *CDKN3*, *FOXM1*, *HIF1A*, *IL1B*, *LDHA*, *LDHB*, *MB*, *PC*, *PER2*, and *SLC2A1*.

Enrichment analysis of gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) pathways

GO and KEGG pathway analyses revealed associations between the 16 GALMRDEGs and HNSCC (Table S1). The 16 GALMRDEGs in HNSCC were mainly involved in biological processes (BP), such as lactate, hexose, monosaccharide, and pyruvate metabolic processes and associated with cellular components (CC), including membrane rafts, microdomains, sarcolemma, acrosomal membrane, and basolateral plasma membrane. The identified molecular functions (MF) included oxidoreductase activities involving CH-OH groups with NAD/NADP, xenobiotic and vitamin transmembrane transporter activities, and potassium channel regulation. These genes were associated with KEGG pathways such as central carbon metabolism in cancer, HIF-1 signaling, pyruvate metabolism, glucagon signaling, and antifolate resistance. Bubble plots were used to visualize the enrichment results from GO and KEGG analyses (Supplementary Fig. S2a), and network diagrams for BP, CC, MF, and pathways are presented in Supplementary Fig. S2b–e.

Phenotype score construction, weighted gene co-expression network analysis (WGCNA), and expression correlation analysis

The GALM.Score for all TCGA samples was determined using the single-sample gene set enrichment analysis (ssGSEA) algorithm based on the expression levels of 16 GALMRDEGs. A comparison plot (Fig. 4a) revealed significant differences ($P < 0.01$) in GALM.Score between the HNSCC and control groups. WGCNA was conducted on the top 70% of genes, ranked by absolute median deviation, to identify co-expression modules in

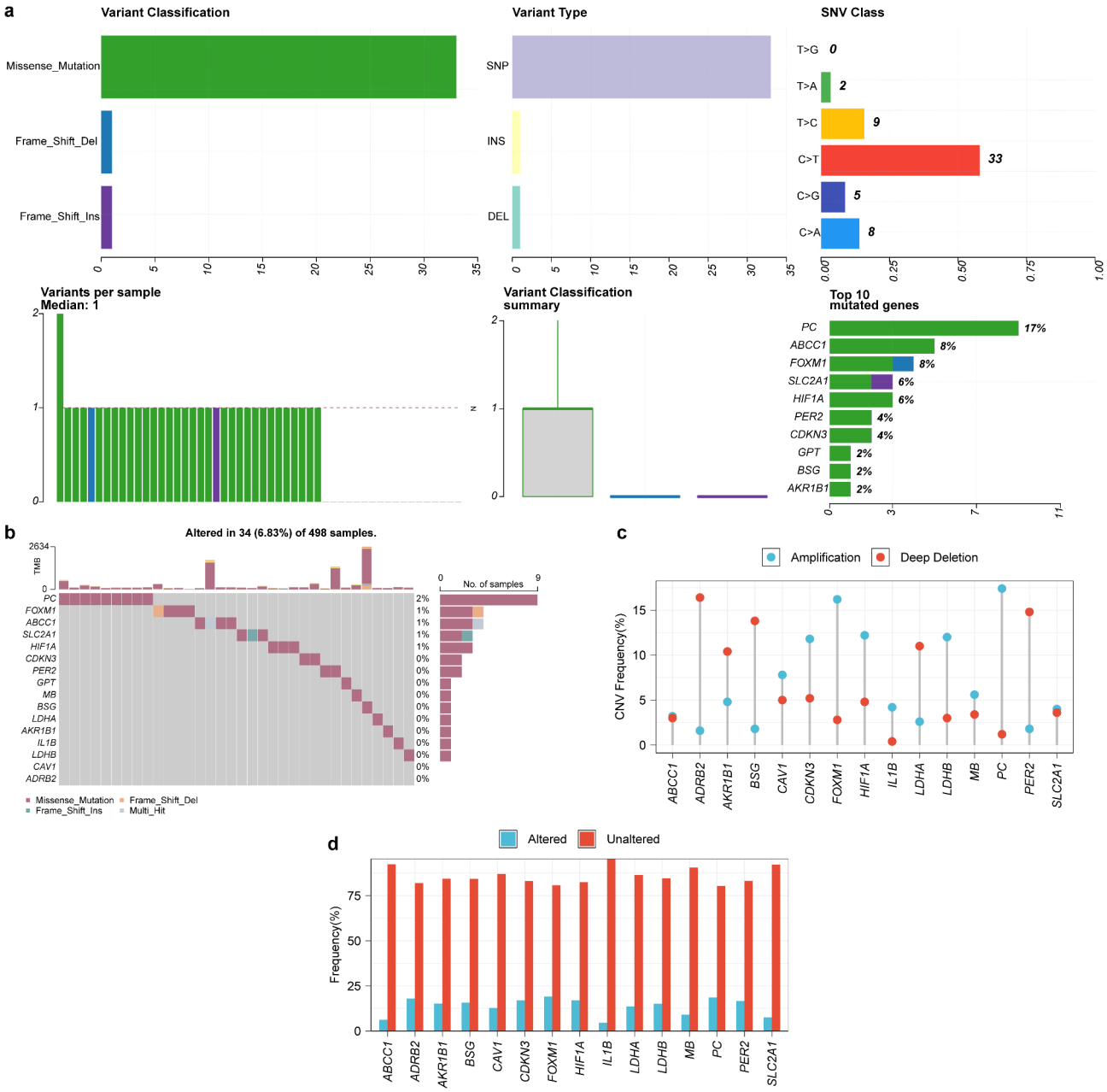

Fig. 3. SM and CNV analysis of GALMRDEGs in TCGA-HNSC. (**a**) Overview of mutation characteristics in the HNSCC. (**b**) OncoPrint showing the distribution of somatic mutations within the 16 GALMRDEGs across individual HNSCC samples. (**c**) Dot plot showing the frequency of CNV for each GALMRDEG. (**d**) Frequency of samples with SM for each GALMRDEG. SM: Somatic mutation, CNV: Copy number variation, GALMRDEGs: Glycolysis and lactate metabolism-related differentially expressed genes, TCGA: The Cancer Genome Atlas, HNSC: Head and neck squamous cell carcinomas, SNV: Single nucleotide variant, SNP: Single nucleotide polymorphism, INS: insertion, DEL: deletion.

TCGA-HNSC group. The scale-free topology fit index was evaluated across different soft thresholds, identifying an optimal threshold of 8 with a fit index of 0.89 (Fig. 4b). This resulted in a co-expression network clustering the top 70% of genes into 15 modules at a cutoff value of 0.25, which were visualized along with their relationships (Fig. 4c, d). The correlation between module gene expression patterns and the GALM.Score was assessed for all 15 modules (Fig. 4e). Genes from the green and purple modules, with $|r| > 0.3$ were selected for further analysis. A Venn diagram (Fig. 4f, g) revealed that two genes, *CDKN3* and *SLC2A1*, were shared between the 16 GALMRDEGs and the selected module genes.

The differential expression analysis of *CDKN3* and *SLC2A1* between the HNSCC and control groups in TCGA-HNSC dataset showed significant differences ($P < 0.001$), as shown in the boxplots (Fig. 5a). Receiver operating characteristic curves (ROC) indicated that *CDKN3* effectively distinguished patients with HNSCC

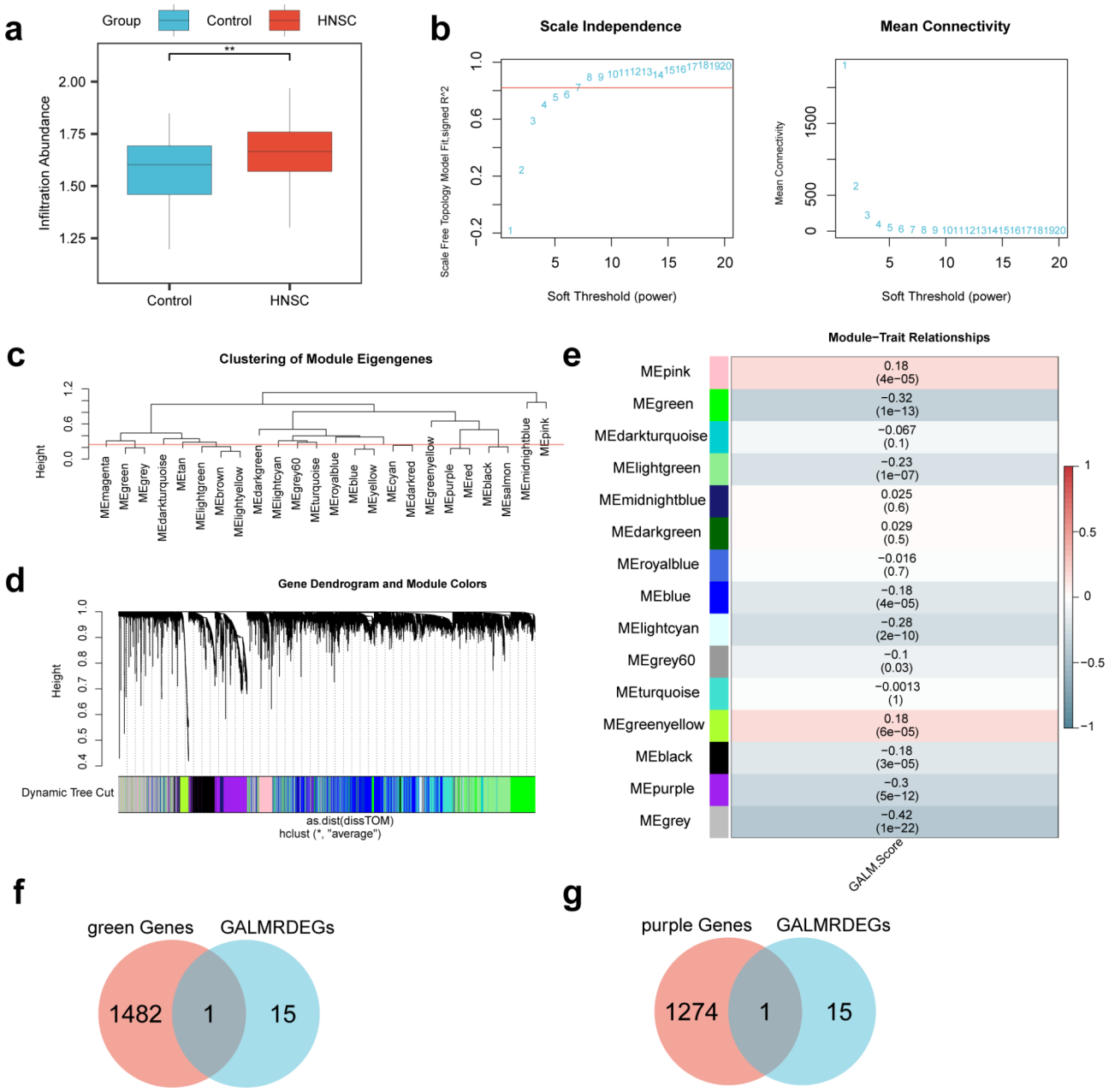

Fig. 4. WGCNA and module-phenotype correlations for TCGA-HNSCC. (**a**) Group comparison plot of the GALM. Scores between the HNSCC and control groups in TCGA-HNSC are presented. (**b**) Scale-free network display of the best soft threshold from WGCNA; the left panel shows the best soft threshold, and the right panel shows the network connectivity under different soft threshold conditions. (**c**) Module clustering results of genes with the top 70% absolute median difference. (**d**) Cluster results for genes with the top 70% absolute median difference. The upper part is divided into hierarchical clustering dendrograms, and the lower part is divided into gene modules. (**e**) The results of the correlation analysis between the top 70% gene cluster modules with absolute median difference and GALM.Score are presented. (**f–g**) GALMRDEGs and green (**f**) and purple (**g**) Venn diagrams display module genes. ns: $P \geq 0.05$, not statistically significant, *: $P < 0.05$, statistically significant, **: $P < 0.01$, highly statistically significant, ***: $P < 0.001$, extremely highly statistically significant. TCGA: The Cancer Genome Atlas, HNSCC: Head and neck squamous cell carcinoma, WGCNA: Weighted gene co-expression network analysis, GALMRDEGs: Glycolysis and lactate metabolism-related differentially expressed genes, GALM.Score: Glycolysis and lactate metabolism score.

from controls with high accuracy (area under the curve [AUC] > 0.9), whereas *SLC2A1* indicated moderate accuracy (0.7 < AUC < 0.9) (Fig. 5b). Correlation analysis of the expression matrix for both module genes revealed a weak positive correlation ($r = 0.104$, $P < 0.05$) between *CDKN3* and *SLC2A1*, as shown in scatter plots (Fig. 5c).

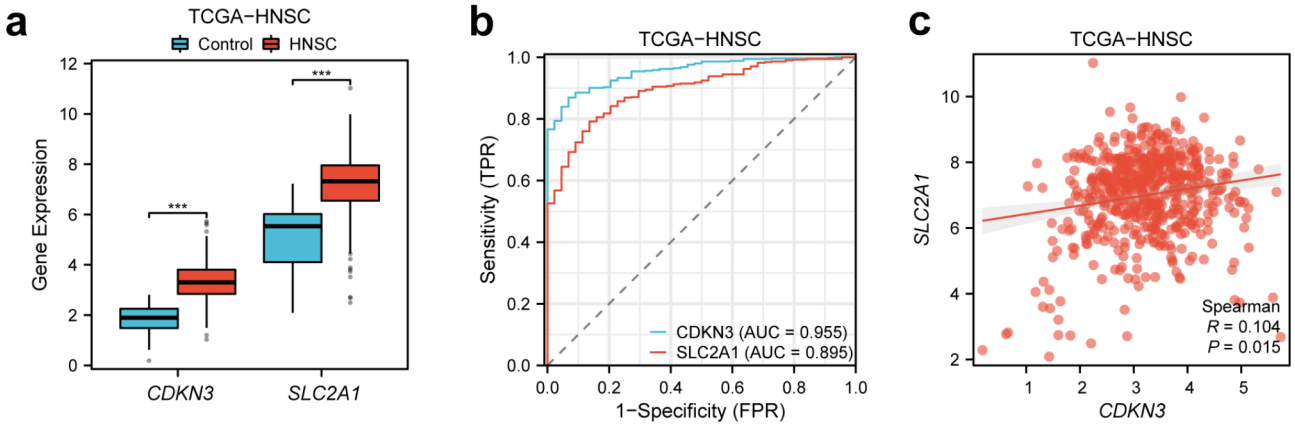

Fig. 5. Analysis of differences in the expression of module genes. (**a**) Module genes in TCGA-HNSC dataset are displayed in grouping comparison boxplots. (**b**) ROC curve for module genes in TCGA-HNSC dataset. (**c**) Scatter plot of the association between *CDKN3* and *SLC2A1*. ns: $P \geq 0.05$, not statistically significant, *: $P < 0.05$, statistically significant, **: $P < 0.01$, highly statistically significant, ***: $P < 0.001$, extremely highly statistically significant. TCGA: The Cancer Genome Atlas, HNSC: Head and neck squamous cell carcinoma, ROC: Receiver operating characteristic.

Development of a prognostic risk model for HNSCC

We developed a prognostic risk model for HNSCC by performing univariate Cox regression analysis on the 16 GALMRDEGs, which identified five key genes: *CAV1*, *CDKN3*, *LDHA*, *MB*, and *PER2* (Fig. 6a). Least absolute shrinkage and selection operator (LASSO) regression confirmed the prognostic significance of these genes, with consistent results in the LASSO coefficient and variable trajectory plots (Fig. 6b, c). Multivariate Cox regression analysis further assessed the association between the RiskScore and clinical prognosis (Fig. 6d). The RiskScore was calculated as follows:

$$\begin{aligned}\mathrm{Risk}Score \;=\; & CDKN3 \times (0.149052) + LDHA \times (0.179586) + MB \\ & \times (0.059684) + CAV1 \times (0.014609) + PER2 \times (-0.219785)\end{aligned}$$

The samples were divided into high- and low-risk categories using the median RiskScore as the threshold. Kaplan–Meier analysis revealed a significant difference in overall survival (OS) between the groups in TCGA dataset ($P < 0.05$) (Fig. 6e). A risk factor plot (Fig. 6f) illustrates the HNSCC grouping, displaying survival times and outcomes, along with a heatmap of the expression levels of the five key genes.

Univariate and multivariate Cox regression analyses based on the riskscore

Univariate and multivariate Cox regression analyses identified RiskScore and four clinical variables (sex, age, pathologic stage, and histological grade) as significant predictors ($P < 0.1$), as shown in the forest plots (Fig. 7a, b). A nomogram was developed to evaluate the prognostic model, with histological grade as the most informative predictor and sex as the least useful (Fig. 7c). Calibration analysis at 1-, 3-, and 5-year intervals (Fig. 7d–f) demonstrated the best performance at 1 year, followed by 3 and 5 years.

Analysis of differential expression and gene correlation

Boxplots were used to compare the expression levels of the five key genes between the low- and high-risk groups in TCGA dataset (Supplementary Fig. S3a). Highly significant differences in *MB* expression were observed ($P < 0.01$), whereas extremely highly significant differences were observed in *CAV1*, *CDKN3*, *LDHA*, and *PER2* ($P < 0.001$). We further analyzed the differential expression of these key genes in HNSCC using the combined GEO dataset. The RiskScore was derived from gene expression levels and multivariate Cox regression coefficients in the HNSCC prognostic model, with samples categorized into low- and high-risk groups based on the median. The boxplots showed extremely highly significant expression differences for *CAV1*, *CDKN3*, *LDHA*, and *PER2* ($P < 0.001$) between the groups (Supplementary Fig. S3b). Correlation chord diagrams were used to visualize the five key genes in TCGA-HNSC and combined GEO datasets (Supplementary Fig. S3c, d). The findings revealed a notable positive correlation between *CAV1* and *LDHA* across both datasets.

Differential expression analysis and GSEA in TCGA-HNSC dataset

TCGA samples were divided into high- and low-risk groups based on the median RiskScore for analysis. Fifty DEGs were identified, with 21 upregulated and 29 downregulated, based on |logFC| > 0.5 and adj.$P < 0.05$. Supplementary Fig. S4a shows a volcano plot of the DEGs, and Supplementary Fig. S4b presents a heatmap of the top 20 DEGs by |logFC|. GSEA was performed to assess the impact of gene expression on HNSCC development, comparing high- and low-risk groups based on logFC values. The results (ridge plot, Supplementary Fig. S4c; Table S2) revealed significant enrichment in pathways such as the Buffa Hypoxia Metagene (Supplementary Fig.

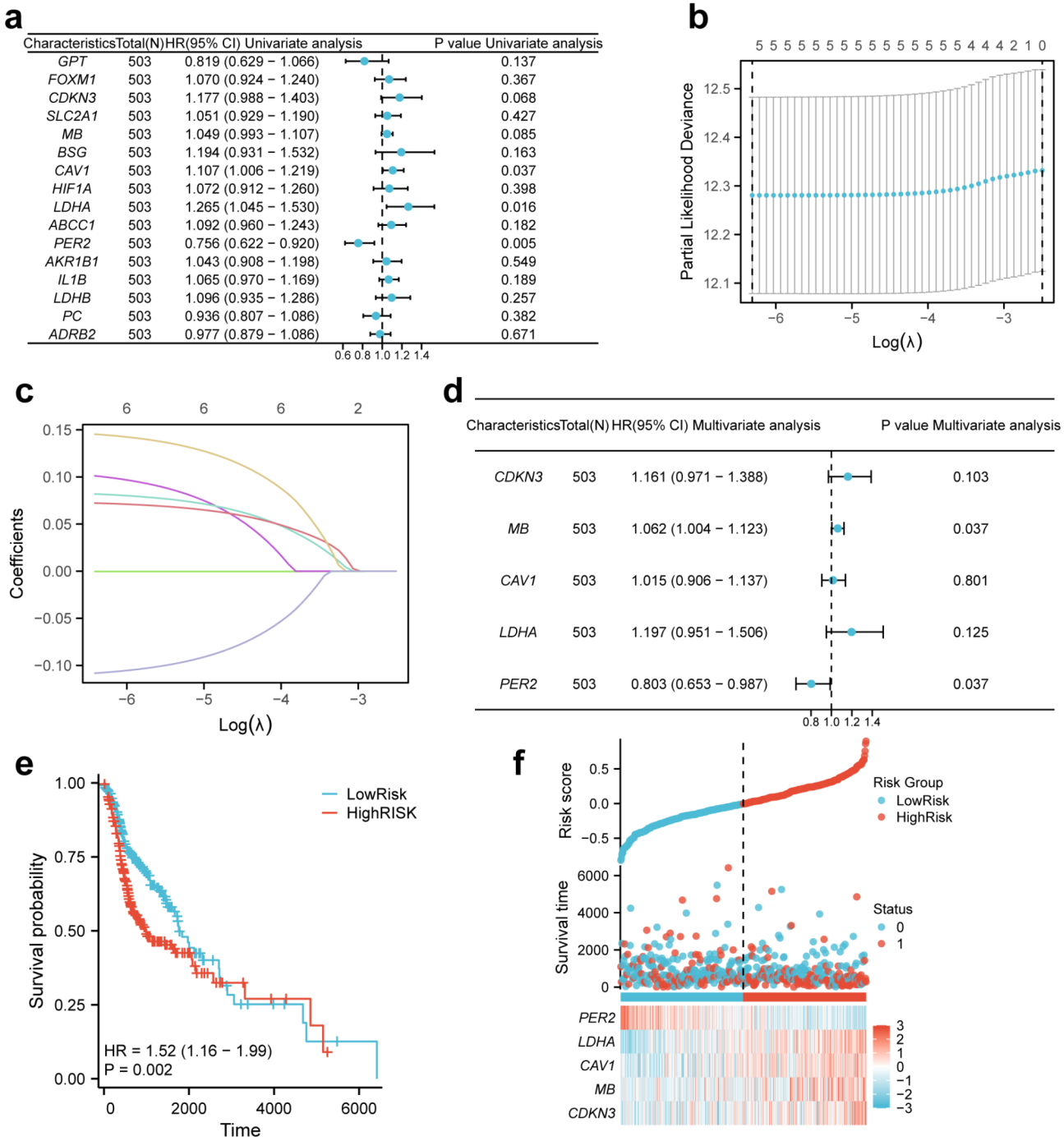

Fig. 6. Prognostic model construction and survival analysis. (**a**) Forest plot of 16 GALMRDEGs in the univariate Cox regression model. (**b**) LASSO regression partial likelihood deviance plot, selecting the optimal λ for the model. (**c**) Variable trajectory plot from LASSO regression. (**d**) Forest plot of five key genes in the multivariate Cox regression model. (**e**) Prognostic Kaplan–Meier survival curves between the RiskScore high and low groups and OS in HNSCC. (**f**) Risk factor plot of a prognostic risk model for HNSCC. GALMRDEGs: Glycolysis and lactate metabolism-related differentially expressed genes, LASSO: Least absolute shrinkage and selection operator, HNSCC: Head and neck squamous cell carcinoma, OS: Overall survival.

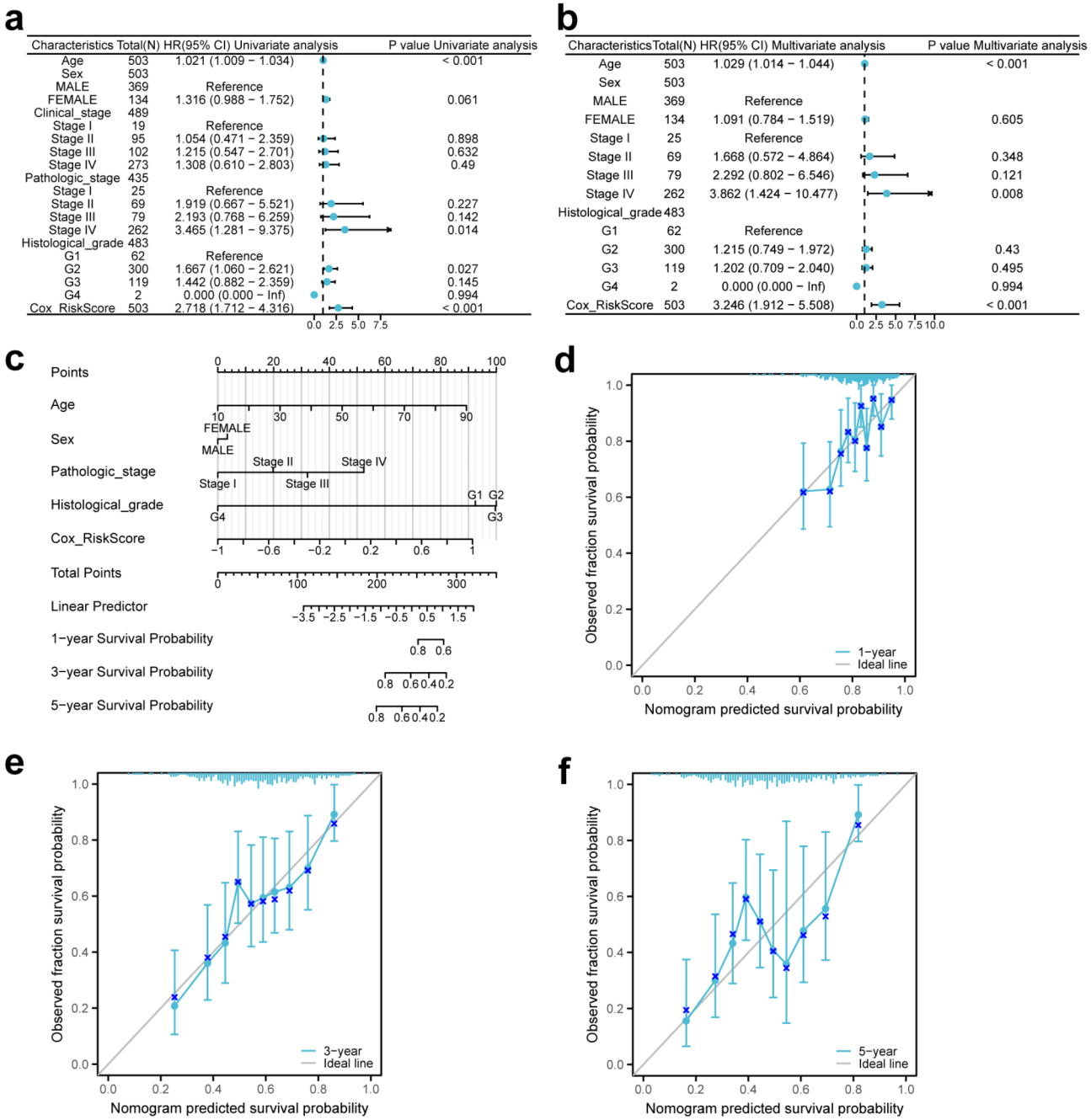

Fig. 7. Validation of the prognostic model based on Cox regression analysis. (**a–b**) Forest plot of the RiskScore and clinical information in the univariate (**a**) and multivariate Cox regression models (**b**). (**c**) Nomogram of the RiskScore and clinical information in the univariate and multivariate Cox regression models. (**d–f**) Calibration curves for 1- (**d**), 3- (**e**), and 5-year (**f**) survival.

S4d), Winter Hypoxia Up (Supplementary Fig. S4e), Foroutan Prodrank TGF-β EMT Up (Supplementary Fig. S4f), and Bosco Epithelial Differentiation Module (Supplementary Fig. S4g).

Immune-cell infiltration, tumor immune dysfunction and exclusion (TIDE), microsatellite instability (MSI), and tumor mutation burden (TMB)

CIBERSORT analysis of TCGA dataset revealed the abundance of 22 immune cell types in HNSCC, highlighting significant differences in immune cell infiltration between the groups (Fig. 8a). Significant differences ($P < 0.05$) in the abundance of seven immune cell types were observed, including plasma cells, T follicular helper (Tfh) cells, regulatory T cells (Tregs), resting natural killer cells, M1 macrophages, and both resting mast cells (rMCs) and activated mast cells. Correlation heatmaps (Fig. 8b, c) depicting the infiltration levels of these immune cells revealed strong correlations between the low- and high-risk groups. rMCs and activated mast cells showed

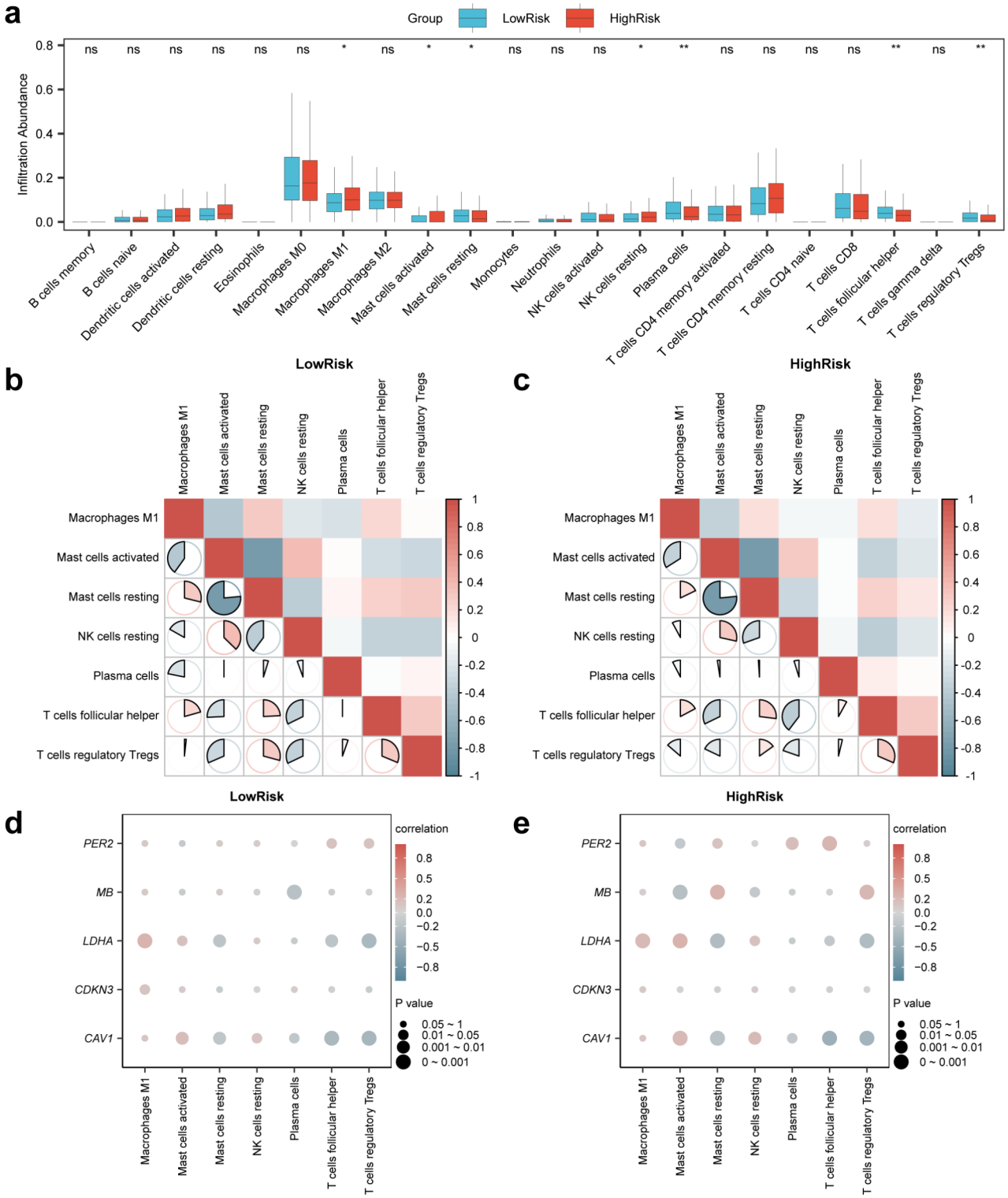

Fig. 8. Risk group immune cell infiltration analysis using the CIBERSORT algorithm. (**a**) Immune cells in HNSCC of the low-risk group and the group comparison chart in the high-risk group. (**b–c**) Correlation heatmap of immune cells in the low-risk (**b**) and high-risk groups (**c**). (**d–e**) Bubble plot of the correlation between infiltrated immune cell abundance and key genes in the low-risk (**d**) and high-risk (**e**) groups. ns stands for $P \geq 0.05$, not statistically significant, $^{*}P < 0.05$, statistically significant, and $^{**}P < 0.01$, highly statistically significant. Red and blue indicate positive and negative correlations, respectively. The color intensity reflects the strength of the correlation. HNSCC: Head and neck squamous cell carcinoma.

the strongest negative correlation in both the low-risk ($r = -0.765$, $P < 0.05$) and high-risk groups ($r = -0.766$, $P < 0.05$). The association between key genes and immune-cell infiltration was illustrated using bubble plots (Fig. 8d, e). *CAV1* exhibited the strongest negative correlation with Tregs ($r = -0.354$, $P < 0.05$) in the low-risk group and with Tfh cells ($r = -0.434$, $P < 0.05$) in the high-risk group.

The distribution of TIDE, MSI, and TMB was evaluated in the low-risk and high-risk groups. No statistically significant differences were observed in TIDE (Fig. 9a) and MSI (Fig. 9b) scores between the two groups ($P > 0.05$). The TMB score in the low-risk group was higher than that in the high-risk group, a difference that was statistically significant ($P < 0.05$) as shown in Fig. 9c.

Construction of HNSCC subtypes

Consensus clustering of TCGA-HNSC dataset revealed two distinct HNSCC subtypes based on five key gene expressions: subtype A (cluster 1, $n = 355$) and subtype B (cluster 2, $n = 149$) (Fig. 10a–c). The 3D t-SNE plot (Fig. 10d), heatmap (Fig. 10e) and group comparison chart (Fig. 10f) confirmed significant differences between the subtypes, with notable expression changes in *CAV1* and *MB* ($P < 0.001$).

Protein-protein interaction (PPI) and regulatory networks

A PPI network for the five key genes was constructed using the STRING database, revealing interactions among *CAV1*, *CDKN3*, *LDHA*, *MB*, and *PER2* (Supplementary Fig. S5a). An interaction network for these genes and 20 functionally similar genes was constructed using GeneMANIA (Supplementary Fig. S5b). The StarBase database was used to identify associated miRNAs, and an mRNA–miRNA regulatory network of four key genes and 39 miRNAs was visualized using Cytoscape (Supplementary Fig. S5c, Table S3).

Discussion

HNSCC is the sixth most common cancer globally, often diagnosed at an advanced stage, leading to poor prognosis[9]. Surgical treatment and radiotherapy can cause facial disfigurement and impaired swallowing, depending on tumor location. Immunotherapy is a promising new treatment option, but only a subset of patients benefit, likely due to immune evasion mechanisms within the TME[10]. This evasion is closely linked to metabolic reprogramming, particularly the upregulation of glycolysis and lactate metabolism, which supports tumor survival in hypoxic conditions. Understanding these processes is crucial for enhancing treatment precision.

To address this, we analyzed glycolysis and lactate metabolism in HNSCC and identified 16 GALMRDEGs related to metabolism, cell cycle regulation, inflammation, and tumor progression. Using a subset of these key genes, we developed a prognostic risk model that stratifies patients into high- and low-risk groups, which exhibit distinct immune cell infiltration patterns. Additionally, we identified two distinct molecular subtypes of HNSCC, providing insights into prognosis and personalized treatment.

Enrichment analysis of GO and KEGG pathways revealed that the 16 GALMRDEGs were primarily linked to metabolic pathways, including lactate and pyruvate metabolism. Their localization to membrane rafts and microdomains may facilitate HNSCC invasion and metastasis by enhancing cell signaling and transport. Molecular function analysis highlighted their involvement in oxidoreductase and transmembrane transporter activities, which are crucial for tumor survival in specific metabolic environments. The enrichment in pathways such as central carbon metabolism, HIF-1 signaling, and antifolate resistance indicates the role of these genes in tumor aggressiveness, adaptation to hypoxia, and drug resistance[13–15].

Our prognostic model for HNSCC was based on five key GALMRDEGs—*CAV1*, *CDKN3*, *LDHA*, *MB*, and *PER2*. This model effectively distinguished patient outcomes, showing a notable difference in OS across the high- and low-risk groups ($P < 0.002$). Compared to single-gene studies, our multi-gene model enhances

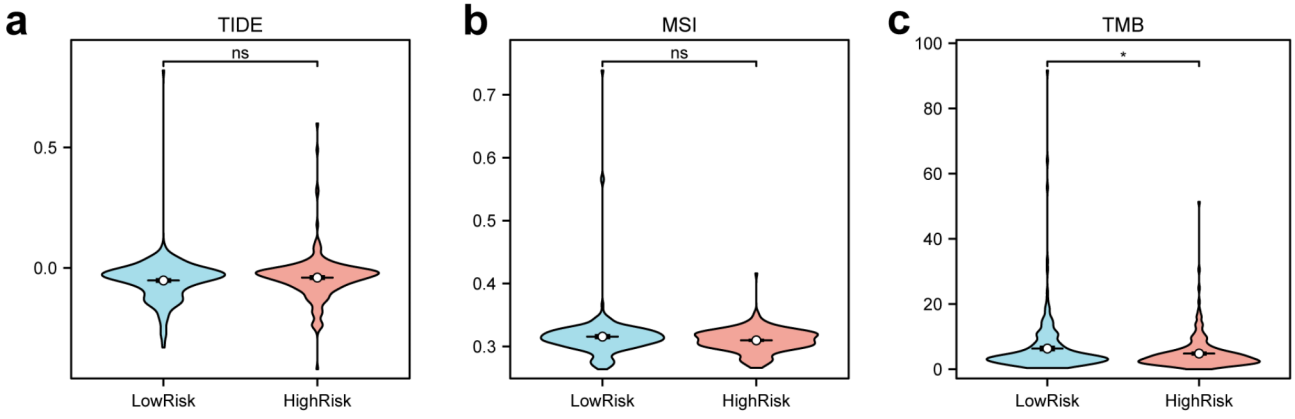

Fig. 9. TIDE, MSI, and TMB analyses. (**a–c**) Results of the TIDE immunotherapy score (**a**), MSI score (**b**), and TMB score (**c**) are displayed as a grouped comparison chart. ns: $P \geq 0.05$, not statistically significant, *: $P < 0.05$, statistically significant, **: $P < 0.01$, highly statistically significant, ***: $P < 0.001$, extremely highly statistically significant. TIDE: Tumor immune dysfunction and exclusion, MSI: Microsatellite instability, TMB: Tumor mutation burden.

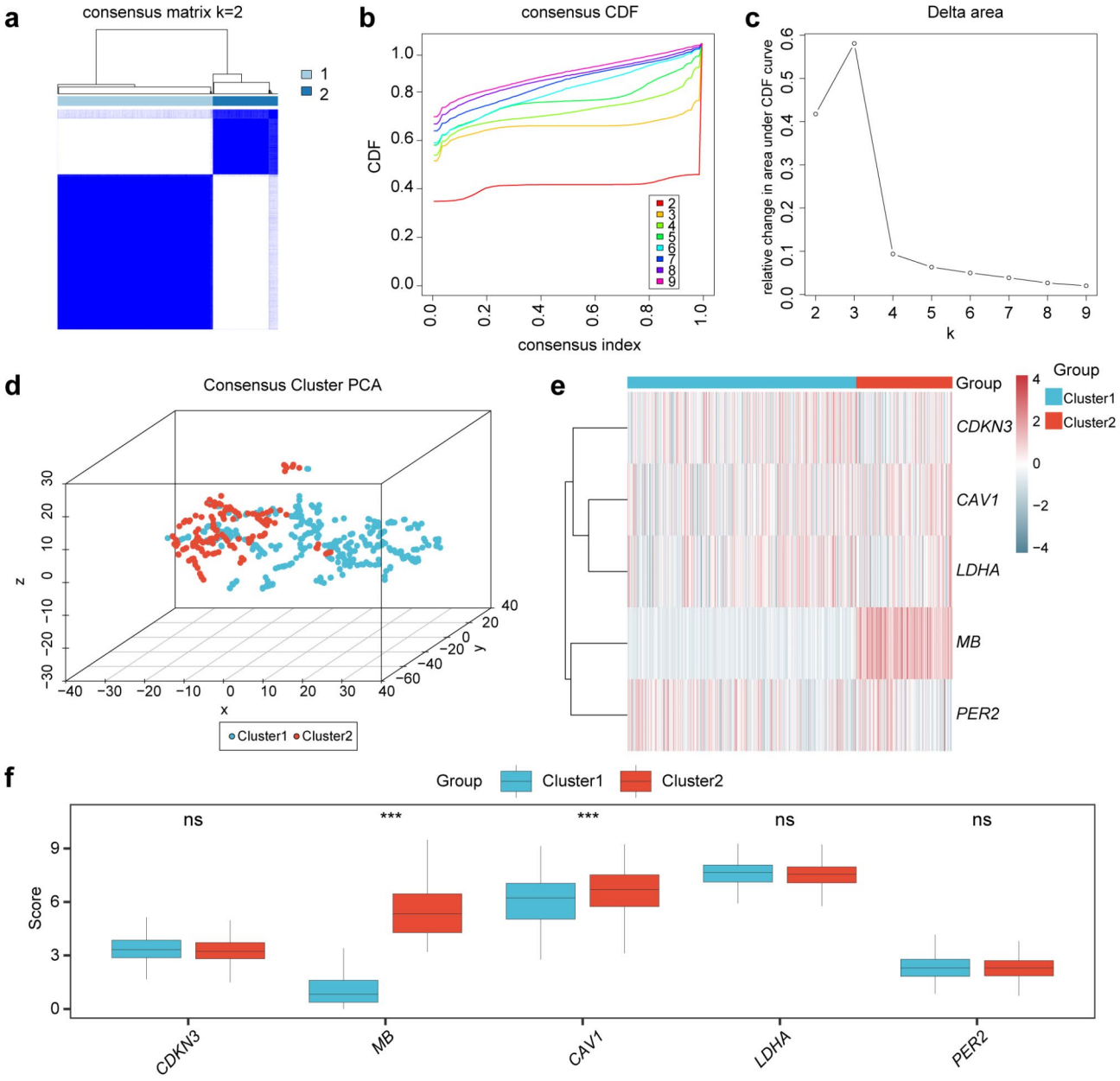

Fig. 10. Consensus clustering analysis for HNSCC. (**a**) Consensus clustering results of HNSCC from TCGA-HNSC dataset. (**b–c**) Consensus clustering analysis: cumulative distribution function (CDF) plot (**b**) and delta area plot (**c**). (**d**) Three-dimensional t-SNE clustering plot of two HNSCC subtypes. (**e**) Heatmap showing the expression levels of key genes in different HNSCC subtypes. (**f**) Group comparison chart of key gene expression between the two HNSCC subtypes. ns: $P \geq 0.05$, not statistically significant, *: $P < 0.05$, statistically significant, **: $P < 0.01$, highly statistically significant, ***: $P < 0.001$, extremely highly statistically significant. TCGA: The Cancer Genome Atlas, HNSCC: Head and neck squamous cell carcinoma, PCA: Principal component analysis.

predictive accuracy by capturing the combined influence of these genes on prognosis. Briefly, *CAV1*regulates energy metabolism and glycolysis via the AMPK and mTOR pathways[16]; LDHA, overexpressed in HNSCC, acts as an oncogene promoting HNSCC progression via glycolysis and EMT, and is correlated with advanced disease, metastasis, and poor prognosis[17,18]. *CDKN3*promotes tumor metabolism through lactate production[19]; and *PER2*inhibits tumor proliferation and glycolysis[20], with targeted modulation of *PER2*impacting PD-L1 expression and potentially augmenting HNSCC immunotherapy in combination with anti-PD-L1 treatment[21]. Although *MB*regulates the intracellular oxygen levels and oxidative stress, its role in HNSCC remains unclear[22]. Notably, the role of *CAV1*in HNSCC is complex. Some studies suggest that it promotes tumor progression[23], with high expression strongly associated with poor prognosis in HNSCC[24], while others propose a tumor-suppressive

function[25]. Our findings indicated that *CAV1* expression was reduced in the low-risk group and elevated in the high-risk group, offering novel insights into its prognostic significance.

To elucidate the regulatory mechanisms of the five key GALMRDEGs, we constructed PPI and mRNA-miRNA regulatory networks. The PPI network revealed that these genes interact with 20 functionally related proteins, suggesting potential crosstalk between various signaling pathways. For instance, *PER2* interacts with both *CAV1* and *LDHA*, highlighting its potential role in coordinating metabolic reprogramming. Although *CDKN3* was identified in the PPI network, it was not included in the mRNA–miRNA network because no miRNAs met the stringent interaction criteria defined by the StarBase v3.0 database and our analysis parameters.

Immune-cell infiltration patterns differed between the low- and high-risk groups. The increased proportions of plasma cells and Tfh cells in the low-risk group were associated with improved disease-free survival and OS. An increase in Tfh cells is often correlated with higher plasma cell generation[33], and their activity is positively linked to patient survival[34], consistent with our findings. We observed an increased proportion of rMCs in the low-risk group. The role of rMCs in prognosis is complex. Some studies suggest that high rMC infiltration could benefit the immune response owing to weaker immunosuppression in the TME, while others indicate a poor prognosis in certain HNSCC subtypes, suggesting context-dependent roles[35]. Treg infiltration is typically correlated with poor prognosis in HNSCC[36], but our research suggests that elevated Treg numbers in the low-risk group may suppress pro-inflammatory cytokines, preventing excessive inflammation and indirectly inhibiting tumor progression. The low-risk group exhibited fewer M1 macrophages, likely due to predominant Tregs inhibiting M1 activation and diminishing pro-inflammatory responses. The balance between Treg-mediated immune suppression and M1 macrophage inflammation may mitigate excessive inflammation, leading to better outcomes in the low-risk group. The high-risk group showed elevated M1 macrophage infiltration, linked to increased glycolysis during metabolic reprogramming[37]. These findings highlight the importance of immune-cell composition in HNSCC and its influence on prognosis through modulating both pro-inflammatory and immunosuppressive responses.

A systematic evaluation of metabolic characteristics in patients with HNSCC and normal controls was performed by calculating the GALM.Score using the ssGSEA algorithm based on the 16 GALMRDEGs. The HNSCC group exhibited a significantly higher GALM.Score than the control group, suggesting increased glycolysis and lactate metabolism in the tumor cells of patients with HNSCC. This scoring method captures the overall activity of metabolic pathways in tumor samples rather than focusing on the expression of individual or a few metabolic markers, providing an innovative, comprehensive, and quantitative measure of metabolic activity, offering a more accurate depiction of metabolic reprogramming in tumors.

Distinct gene expression patterns in HNSCC samples correlated with key biological pathways, including those associated with hypoxia and metabolic reprogramming. The enrichment of pathways such as the Buffa Hypoxia Metagene, Winter Hypoxia Up, Foroutan Prodrank TGF-β, and the Bosco Epithelial Differentiation Module indicates that hypoxic conditions may enhance glycolysis and lactate metabolism. Hypoxia induces the expression of HIFs, upregulating glycolysis-related genes, such as *SLC2A1* and *LDHA*, thereby significantly enhancing glycolytic activity in tumor cells[38].

WGCNA identified two key module genes associated with GALMRDEGs: *CDKN3* and *SLC2A1*. *CDKN3*, a cell-cycle regulator that is often inactivated in cancer, can lead to uncontrolled cell proliferation and tumor progression[39]. *SLC2A1*, a key glucose transporter, is significantly expressed in numerous cancers and associated with tumor metabolic reprogramming[40]. The mRNA level of *SLC2A1*is significantly elevated in HNSCC patients and is associated with shorter OS and advanced tumor stages[41]. These genes were classified into cell proliferation and metabolic regulation modules, indicating that these processes may be co-regulated in HNSCC and synergistically drive tumor development. The significant differential expression of *CDKN3* and *SLC2A1* highlights their potential as diagnostic biomarkers for the early detection and risk assessment in HNSCC. Targeting excessive cell proliferation and abnormal energy metabolism may provide novel therapeutic strategies for clinical treatment.

We classified HNSCC into two subtypes based on the expression levels of five key genes: A (cluster 1) and B (cluster 2). Molecular heterogeneity, shown by distinct separation in the 3D t-SNE plot, highlighted differences in metabolic reprogramming and biological behavior between the subtypes. *CAV1* was highly expressed in subtype B, suggesting that it promotes tumor progression through enhanced energy metabolism and invasiveness. In contrast, elevated *MB* expression in subtype A indicated increased oxidative metabolism. These results suggest that *CAV1* and *MB* are biomarkers of HNSCC subtypes, highlighting the need for subtype-specific targeted therapies.

We evaluated TIDE immunotherapy scores, MSI, and TMB across different HNSCC risk groups. Although the low-risk group showed lower TIDE scores and higher MSI scores compared to the high-risk group, the differences were not statistically significant. Notably, the low-risk group exhibited significantly higher TMB scores, suggesting greater immunogenicity and a potentially better response to immunotherapy. Previous study has indicated that patients with high TMB are more likely to experience a robust tumor response to pembrolizumab monotherapy[42]. These findings underscore the complex relationship between genomic characteristics and immunotherapy outcomes, highlighting the need for further clinical investigation. Future studies incorporating advanced techniques, such as single-cell sequencing and spatial transcriptomics, could provide deeper insights into the specific roles of TMB and TIDE in shaping immune responses in HNSCC.

In this study, we identified high mutation and copy number amplification rates of *PC* in HNSCC, highlighting its essential role in tumor development and progression. *PC*encodes pyruvate carboxylase, which converts pyruvate to oxaloacetate for gluconeogenesis and the tricarboxylic acid (TCA) cycle, providing essential energy and metabolic intermediates[43]. Amplification of *PC*may enhance anaplerotic flux in the TCA cycle, enabling tumor cells to better adapt to hypoxic and nutrient-deprived conditions, which is vital for rapidly proliferating cancer cells[44]. Although research on *PC* in HNSCC is limited, its high mutation and amplification rates make it

a promising therapeutic target. Inhibiting *PC* activity has the potential to reduce the metabolic adaptability of HNSCC cells under stressful conditions, thereby suppressing tumor growth.

This study has several limitations. First, reliance on publicly available datasets restricts the inclusion of independent external validation cohorts, which are essential to confirm the robustness of the results. Although key genes in HNSCC were identified, their specific biological functions and mechanisms require further experimental validation. Second, the prognostic model and disease subtypes should to be evaluated for predictive performance and stability across diverse clinical settings. Future studies should integrate experimental validation with clinical data and validate the results in larger patient cohorts to enhance the precision of HNSCC diagnosis and treatment.

In conclusion, our comprehensive analysis of GALMRGs in HNSCC revealed their pivotal roles in prognosis, the TME, and disease subtypes. These results deepen our understanding of the molecular pathways involved in HNSCC and provide avenues for developing personalized treatments, ultimately enhancing patient outcomes.

Methods

Data acquisition

The TCGA-HNSC dataset from TCGA (https://www.cancer.gov/tcga) was obtained using the "TCGAbiolinks" R package (version 2.30.0)[45]for the training set, with associated clinical data retrieved from the UCSC Xena database[46]. Detailed information is provided in Table S4.

For validation, three HNSCC datasets were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) using the "GEOquery"[47]R package: GSE107591[48](GPL6244 platform) contains 24 HNSCC and 23 control samples; GSE6631[49](GPL8300 platform) contains 22 HNSCC and 22 control samples; and GSE9638[50] (GPL1291 platform) contains 21 HNSCC samples without controls. All samples were included in this study.

A total of 3,326 glycolysis-related genes (GRGs) and 54 lactate metabolism-related genes (LMRGs) were identified from the GeneCards[51]database by searching for protein-coding genes under the terms "Glycolysis" and "Lactate Metabolism," respectively. Additionally, 19 LMRGs were identified in PubMed[52]. Sixty-nine LMRGs were obtained after merging and eliminating duplicates. The overlap of GRGs and LMRGs yielded 58 genes associated with both glycolysis and lactate metabolism, 33 of which were present in all three GEO datasets (Table S5).

We used the R package "sva" (v3.50.0)[53]to correct for batch effects and integrated the GEO datasets into a single dataset comprising 67 HNSCC samples and 45 controls. The integrated dataset was standardized, including probe annotation and normalization, using the R package "limma" (v3.58.1)[54]. PCA[55] was conducted before and after batch correction to assess its effectiveness.

Differential expression analysis of GALMRGs

Differential gene expression analysis between the HNSCC and control groups was performed using the R package "limma" (v3.58.1). Genes were considered differentially expressed when $|\log FC| > 0.5$ and adj.$P < 0.05$, with adjustments applied via the Benjamini–Hochberg method. We intersected the DEGs with GALMRGs to identify the GALMRDEGs in HNSCC. The overlap was visualized using a Venn diagram, and the resulting GALMRDEGs were displayed on a heatmap (R package "pheatmap") and mapped by chromosomal location (R package "RCircos"[56]).

SM and CNV analysis

Somatic mutations in GALMRGs were examined using "Masked Somatic Mutation" data from TCGA, processed via VarScan software (https://varscan.sourceforge.net), and visualized using the R package "maftools" (Version 2.18.0)[57]. CNVs were assessed using "Masked Copy Number Segment" data from the same source.

GO and KEGG pathways

We used the R package "clusterProfiler" (v4.10.0)[58]for the GO[59]and KEGG[60] pathway enrichment analyses of GALMRDEGs. Using the Benjamini–Hochberg correction, GO terms and pathways were considered significant if their adj.*P* and false discovery rate (FDR) were below 0.05.

Phenotype score development, WGCNA, and expression correlation analysis

The GALM.Score for all samples in the TCGA-HNSC dataset were computed using the R package "GSVA" and ssGSEA based on the expression of GALMRDEGs. These scores were visualized using "ggplot2" for group comparisons and assessed for predictive ROC curves from the "pROC" package, with AUC values ranging from 0.5 to 1, indicating diagnostic performance.

We used WGCNA[61] to identify gene modules that were correlated with GALM.Score. The modules are represented as branches in a color-coded hierarchical clustering tree. Module genes within significant modules were identified and compared to GALMRDEGs using a Venn diagram to identify overlapping genes. In TCGA-HNSC dataset, differences in the expression of module genes between the HNSCC and control groups were analyzed using group comparison plots and ROC curves to evaluate diagnostic performance. Spearman's correlation analysis was performed and visualized using "ggplot2," with correlation coefficients categorized as weak ($|r| < 0.3$), moderate ($0.3 \le |r| < 0.5$), or strong ($|r| \ge 0.8$).

Construction of a prognostic risk model for HNSCC

Using univariate and multivariate Cox regression analyses via the R package "survival"[62], we developed a prognostic risk model for TCGA-HNSC dataset to evaluate the prognostic significance of GALMRDEGs. Variables with $P < 0.10$ in the univariate Cox analysis were selected for LASSO regression to reduce overfitting and enhance model generalization. LASSO regression was conducted with the R package "glmnet"[63] (set.seed = 2024,

family = "binomial"), and the outcomes were illustrated using risk and coefficient trajectory plots. GALMRDEGs identified through LASSO regression were used in multivariate Cox regression analysis to determine the key genes for the prognostic model. A forest plot was generated to illustrate the results, highlighting gene expression. The RiskScore was calculated using multivariate Cox regression coefficients as follows:

$$\text{Risk } Score = \sum_i Coefficient\ (gene_i) \times mRNA\ Expression\ (gene_i)$$

Prognostic risk model for HNSCC

The patients were stratified into high- and low-risk groups using the median RiskScore. Kaplan–Meier survival curves, created with the R package "survival," were used to compare the OS between the groups. Univariate and multivariate Cox regression analyses were conducted to assess the associations between RiskScore, clinical variables, and OS, with the results shown in forest plots. A nomogram[64] was developed using the R package "rms" to illustrate survival probabilities based on the RiskScore and clinical factors, and calibration curves were plotted to assess the predictive accuracy.

Analysis of key gene expression levels and correlations

We used comparison plots to analyze key gene expression differences between the low- and high-risk groups in TCGA-HNSC dataset. In the combined GEO dataset, multivariate Cox regression coefficients were used to compute the RiskScore, categorizing the samples into low- and high-risk groups according to the median RiskScore. Correlation chord diagrams were used to visualize the Spearman's correlation analysis results for key genes across both datasets. Correlation coefficients were categorized into weak ($|r| < 0.3$), moderate ($0.3 \le |r| < 0.5$), and strong ($|r| \ge 0.8$) groups.

Differential gene expression analysis and GSEA for TCGA-HNSC

TCGA-HNSC samples were classified into high- and low-risk groups based on the median RiskScore. Differential expression analysis using "limma" ($|logFC| > 0.5$, adj. $P < 0.05$) identified DEGs, which were visualized using a volcano plot. The top 20 DEGs were shown in a heatmap using "pheatmap." GSEA[65] was conducted to evaluate the distribution of predefined gene sets between the high- and low-risk groups, thereby emphasizing their phenotypic impact. Genes were ranked based on logFC values, and the analysis utilized the R package "clusterProfiler" (seed = 2024, gene set size range: 10–500) with the c2.all.v2023.2.Hs.symbols gene set from the Molecular Signatures Database. Pathways with adj.$P < 0.05$ and FDR < 0.05 were considered significant.

Immune cell infiltration, TIDE, MSI, and TMB

We analyzed immune-cell infiltration in TCGA-HNSC dataset using CIBERSORT[66]to create immune-cell infiltration matrices. Using the R package "pheatmap" (version 1.0.12), heatmaps were generated to visualize Spearman's correlations between immune cells, and key gene-immune cell relationships were analyzed. We evaluated tumor therapy responses by calculating TIDE scores for HNSCC using the TIDE web tool[67,68] (http://tide.dfci.harvard.edu) based on key gene expressions. The Mann–Whitney U test was used to assess differences in TIDE scores between.

low- and high-risk groups. TMB and MSI data were sourced from the cBioPortal database[69] (https://www.cbioportal.org/) and analyzed across risk groups using a consistent statistical test.

Subtype construction of HNSCC

Consensus clustering[70], a resampling-based algorithm, was employed to classify HNSCC subtypes and validate the clustering stability. Using the R package "ConsensusClusterPlus"[71], we analyzed TCGA-HNSC dataset, focusing on key genes. Clusters 2–9 were tested with 50 resampling iterations on 80% of the samples, employing the "pam" clustering algorithm and "Euclidean" distance metric. Differences in expression among the subtypes were visualized using heatmaps and group comparison plots.

PPI and regulatory networks

A PPI network for key genes with low confidence (0.400) was created using the STRING[72] database (https://cn.string-db.org/)and visualized using Cytoscape[73]. GeneMANIA[74] (https://genemania.org/) predicted functionally similar genes and generated hypotheses regarding gene functions. Key gene–miRNA interactions were analyzed using the relevant miRNAs obtained from StarBase v3.0[75] (https://starbase.sysu.edu.cn/), and the mRNA–miRNA regulatory network was visualized using Cytoscape.

Statistical analysis

Data analysis was performed using R software (version 4.2.0; R Foundation for Statistical Computing, Vienna, Austria). Continuous variables are presented as the mean ± standard deviation. Group comparisons were performed using the Wilcoxon rank-sum test, and molecular associations were evaluated using Spearman's correlation ($P < 0.05$).

Data availability

All data generated or analysed during this study are included in this published article (and its Supplementary Information files).

Received: 19 December 2024; Accepted: 17 March 2025

Published online: 28 March 2025

References

1. Sharkey Ochoa, I. et al. The role of HPV in determining treatment, survival, and prognosis of head and neck squamous cell carcinoma. *Cancers (Basel)*. **14**, 4321 (2022).
2. Johnson, D. E. et al. Head and neck squamous cell carcinoma. *Nat. Rev. Dis. Primers*. **6**, 92 (2020).
3. Orland, M. D., Ullah, F., Yilmaz, E. & Geiger, J. L. Immunotherapy for head and neck squamous cell carcinoma: present and future approaches and challenges. *JCO Oncol. Pract.* https://doi.org/10.1200/OP.24.00041 (2024).
4. Ngwa, V. M., Edwards, D. N., Philip, M. & Chen, J. Microenvironmental metabolism regulates antitumor immunity. *Cancer Res.* **79**, 4003–4008 (2019).
5. Fukushi, A., Kim, H. D., Chang, Y. C. & Kim, C. H. Revisited metabolic control and reprogramming cancers by means of the Warburg effect in tumor cells. *Int. J. Mol. Sci.* **23**, 10037 (2022).
6. Boschert, V., Teusch, J., Müller-Richter, U. D. A., Brands, R. C. & Hartmann, S. PKM2 modulation in head and neck squamous cell carcinoma. *Int. J. Mol. Sci.* **23**, 775 (2022).
7. Yang, Y. et al. Research progress of Warburg effect in hepatocellular carcinoma. *Front. Bioscience-Landmark*. **29**, 178 (2024).
8. Yang, L. et al. Tumor-Specific peroxynitrite overproduction disrupts metabolic homeostasis for sensitizing melanoma immunotherapy. *Adv. Mater.* **35**, 2301455 (2023).
9. Sung, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
10. Wise-Draper, T. M., Bahig, H., Tonneau, M., Karivedu, V. & Burtness, B. Current therapy for metastatic head and neck cancer: evidence, opportunities, and challenges. *Am. Soc. Clin. Oncol. Educational Book*. 527–540. https://doi.org/10.1200/EDBK_350442 (2022).
11. Wang, Y. et al. The HIF-1α/PKM2 feedback loop in relation to EGFR mutational status in lung adenocarcinoma. *J. Invest. Surg.* **37**, 2301081 (2024).
12. Liu, J. et al. CCND1 amplification profiling identifies a subtype of melanoma associated with poor survival and an immunosuppressive tumor microenvironment. *Front. Immunol.* **13**, 725679 (2022).
13. Qannita, R. A. et al. Targeting Hypoxia-Inducible Factor-1 (HIF-1) in cancer: emerging therapeutic strategies and pathway regulation. *Pharmaceuticals* **17**, 195 (2024).
14. Hsieh, C. Y., Lin, C. C. & Chang, W. C. Taxanes in the treatment of head and neck squamous cell carcinoma. *Biomedicines* **11**, 2887 (2023).
15. Chaves, P. et al. Preclinical models in head and neck squamous cell carcinoma. *Br. J. Cancer*. **128**, 1819–1827 (2023).
16. Simón, L. et al. Inhibition of Glycolysis and Src/Akt signaling reduces Caveolin-1-enhanced metastasis. *Biomed. Pharmacother.* **176**, 116841 (2024).
17. Tang, Y. et al. LDHA: The obstacle to T cell responses against tumor. *Front. Oncol.* **12**, 1036477 (2022).
18. Cai, H. et al. LDHA promotes oral squamous cell carcinoma progression through facilitating Glycolysis and Epithelial–Mesenchymal transition. *Front. Oncol.* **9**, 1446 (2019).
19. Gao, C. et al. Comprehensive analysis reveals the potential roles of CDKN3 in Pancancer and verification in endometrial cancer. *Int. J. Gen. Med. Volume*. **16**, 5817–5839 (2023).
20. Long, W., Gong, X., Yang, Y. & Yang, K. Downregulation of PER2 promotes tumor progression by enhancing Glycolysis via the phosphatidylinositol 3-Kinase/Protein kinase B pathway in oral squamous cell carcinoma. *J. Oral Maxillofac. Surg.* **78**, 1780e1–1780e14 (2020).
21. Zhang, Z. et al. PER2 binding to HSP90 enhances immune response against oral squamous cell carcinoma by inhibiting IKK/NF-κB pathway and PD-L1 expression. *J. Immunother Cancer*. **11**, e007627 (2023).
22. Bicker, A. et al. The role of myoglobin in epithelial cancers: insights from transcriptomics. *Int. J. Mol. Med.* https://doi.org/10.3892/ijmm.2019.4433 (2019).
23. Burgy, M. et al. Cav1/EREG/YAP axis in the treatment resistance of Cav1-Expressing head and neck squamous cell carcinoma. *Cancers (Basel)*. **13**, 3038 (2021).
24. Zhai, Q. et al. Identification of ferroptosis-associated tumor antigens as the potential targets to prevent head and neck squamous cell carcinoma. *Genes Dis.* **11**, 101212 (2024).
25. Zhang, H. et al. Restoration of caveolin-1 expression suppresses growth and metastasis of head and neck squamous cell carcinoma. *Br. J. Cancer*. **99**, 1684–1694 (2008).
26. Zhu, G. et al. Small extracellular vesicles containing miR-192/215 mediate hypoxia-induced cancer-associated fibroblast development in head and neck squamous cell carcinoma. *Cancer Lett.* **506**, 11–22 (2021).
27. King, B. et al. MicroRNA-510 mediated negative regulation of Caveolin-1 in fibroblasts promotes aggressive tumor growth. *Front. Immunol.* **14**, 1116644 (2023).
28. Jin, Y. et al. Identify and validate circadian regulators as potential prognostic markers and immune infiltrates in head and neck squamous cell carcinoma. *Sci. Rep.* **13**, 19939 (2023).
29. Xiao, L. et al. miR-3187-3p enhances migration and invasion by targeting PER2 in head and neck squamous cell carcinomas. *J. Cancer*. **12**, 5231–5240 (2021).
30. Hsu, T. W. et al. FOXO3a/miR-4259-driven LDHA expression as a key mechanism of gemcitabine sensitivity in pancreatic ductal adenocarcinoma. *Cancer Metab.* **13**, 7 (2025).
31. Kadkhoda, S., Eslami, S., Mahmud Hussen, B. & Ghafouri-Fard S. A review on the importance of miRNA-135 in human diseases. *Front. Genet.* **13**, 973585 (2022).
32. Liu, S. et al. Methylation status of the *Nanog* promoter determines the switch between cancer cells and cancer stem cells. *Adv. Sci.* **7**, 1903035 (2020).
33. Higgins, B. W., McHeyzer-Williams, L. J. & McHeyzer-Williams, M. G. Programming Isotype-Specific plasma cell function. *Trends Immunol.* **40**, 345–357 (2019).
34. Cillo, A. R. et al. Immune landscape of Viral- and Carcinogen-Driven head and neck cancer. *Immunity* **52**, 183–199e9 (2020).
35. Tzorakoleftheraki, S. E. & Koletsa, T. The complex role of mast cells in head and neck squamous cell carcinoma: A systematic review. *Med. (B Aires)*. **60**, 1173 (2024).
36. Cho, J. H. & Lim, Y. C. Prognostic impact of regulatory T cell in head and neck squamous cell carcinoma: A systematic review and meta-analysis. *Oral Oncol.* **112**, 105084 (2021).
37. Li, B., Ren, M., Zhou, X., Han, Q. & Cheng, L. Targeting tumor-associated macrophages in head and neck squamous cell carcinoma. *Oral Oncol.* **106**, 104723 (2020).
38. Cruz-Gregorio, A., Martínez-Ramírez, I., Pedraza-Chaverri, J. & Lizano, M. Reprogramming of energy metabolism in response to radiotherapy in head and neck squamous cell carcinoma. *Cancers (Basel)*. **11**, 182 (2019).
39. Zhang, C., Shen, Q., Gao, M., Li, J. & Pang, B. The role of Cyclin dependent kinase inhibitor 3 (CDKN3) in promoting human tumors: literature review and pan-cancer analysis. *Heliyon* **10**, e26061 (2024).
40. Li, B. Analysis of the relationship between GLUT family in the progression and immune infiltration of head and neck squamous carcinoma. *Diagn. Pathol.* **18**, 88 (2023).

41. Miller, Z. A. et al. GLUT1 inhibitor BAY-876 induces apoptosis and enhances anti-cancer effects of bitter receptor agonists in head and neck squamous carcinoma cells. *Cell. Death Discov*. **10**, 339 (2024).
42. Marabelle, A. et al. Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 KEYNOTE-158 study. *Lancet Oncol.* **21**, 1353–1365 (2020).
43. Deja, S. et al. Hepatic malonyl-CoA synthesis restrains gluconeogenesis by suppressing fat oxidation, pyruvate carboxylation, and amino acid availability. *Cell. Metab.* **36**, 1088–1104e12 (2024).
44. Kiesel, V. A. et al. Pyruvate carboxylase and cancer progression. *Cancer Metab.* **9**, 20 (2021).
45. Colaprico, A. et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71–e71 (2016).
46. Goldman, M. J. et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat. Biotechnol.* **38**, 675–678 (2020).
47. Davis, S. & Meltzer, P. S. GEOquery: a Bridge between the gene expression omnibus (GEO) and bioconductor. *Bioinformatics* **23**, 1846–1847 (2007).
48. Sacconi, A. et al. TMPRSS2, a SARS-CoV-2 internalization protease is downregulated in head and neck cancer patients. *J. Experimental Clin. Cancer Res.* **39**, 200 (2020).
49. Kuriakose, M. A. et al. Selection and validation of differentially expressed genes in head and neck cancer. *Cell. Mol. Life Sci.* **61**, 1372–1383 (2004).
50. Shimizu, S. et al. Identification of a novel therapeutic target for head and neck squamous cell carcinomas: A role for the neurotensin-neurotensin receptor 1 oncogenic signaling pathway. *Int. J. Cancer*. **123**, 1816–1823 (2008).
51. Stelzer, G. et al. The genecards suite: from gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinf.* **54**, 1–30 (2016).
52. Zhao, F., Wang, Z., Li, Z., Liu, S. & Li, S. Identifying a lactic acid metabolism-related gene signature contributes to predicting prognosis, immunotherapy efficacy, and tumor microenvironment of lung adenocarcinoma. *Front. Immunol.* **13**, 980508 (2022).
53. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The <tt>sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
54. Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).
55. Ben Salem, K. & Ben Abdelaziz, A. Principal component analysis (PCA). *Tunis Med.* **99**, 383–389 (2021).
56. Zhang, H., Meltzer, P. & Davis, S. RCircos: an R package for circos 2D track plots. *BMC Bioinform.* **14**, 244 (2013).
57. Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **28**, 1747–1756 (2018).
58. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
59. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**, D419–D426 (2019).
60. Kanehisa, M. K. E. G. G. Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
61. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
62. Therneau, T. M. survival: Survival Analysis. *CRAN: Contributed Packages* Preprint at https://doi.org/10.32614/CRAN.package.survival (2001).
63. Engebretsen, S. & Bohlin, J. Statistical predictions with Glmnet. *Clin. Epigenetics*. **11**, 123 (2019).
64. Wu, J. et al. A nomogram for predicting overall survival in patients with low-grade endometrial stromal sarcoma: A population-based analysis. *Cancer Commun.* **40**, 301–312 (2020).
65. Subramanian, A. et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).
66. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods.* **12**, 453–457 (2015).
67. Jiang, P. et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* **24**, 1550–1558 (2018).
68. Fu, J. et al. Large-scale public data reuse to model immunotherapy response and resistance. *Genome Med.* **12**, 21 (2020).
69. Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1 (2013).
70. Lock, E. F. & Dunson, D. B. Bayesian consensus clustering. *Bioinformatics* **29**, 2610–2616 (2013).
71. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573 (2010).
72. Szklarczyk, D. et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
73. Shannon, P. et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
74. Franz, M. et al. GeneMANIA update 2018. *Nucleic Acids Res.* **46**, W60–W64 (2018).
75. Li, J. H., Liu, S., Zhou, H., Qu, L. H. & Yang, J. H. StarBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* **42**, D92–D97 (2014).

Author contributions

JHY: Conceptualization, Formal analysis, Methodology, Software, Validation, Writing – original draft, Writing – review and editing.ZLJ and ZHD: Data curation, Validation.YZK: Supervision, Writing – review and editing.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-94843-z.

Correspondence and requests for materials should be addressed to Z.Y.