

# Five glycosylation-related gene signatures predict the prognostic risks of lung adenocarcinoma

Xiao-Hui Liu<sup>1</sup>, Yi-Ming Yang<sup>2</sup>, Ming-Ming-Zhang<sup>1</sup> and Hao-Yu Fu<sup>3</sup>

<sup>1</sup> Department of Thoracic Surgery, Tangshan People's Hospital, Hebei, P.R. China

<sup>2</sup> Department of Breast Surgery, Tangshan People's Hospital, Hebei, P.R. China

<sup>3</sup> Department of Radiation Oncology, Tangshan People's Hospital, Hebei, P.R. China

**Abstract.** This study aimed to identify glycosylation-related genes associated with lung adenocarcinoma (LUAD) prognosis through comprehensive bioinformatic analysis. Glycosylation-related genes were identified from the Human Gene Nomenclature Committee, and LUAD prognostic genes were screened from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO)-GSE68465 datasets. Glycosylation risk score (GLRS) was calculated to predict LUAD prognostic risk. Samples were grouped into GLRS-high and GLRS-low and compared. The Tumor Immune Dysfunction and Exclusion (TIDE) score was computed to assess the antitumor immune escape possibility after immunotherapy. From 213 glycosylation-related genes, five gene signatures served as prognostic LUAD predictors using univariate and stepwise Cox regression analyses. GLRS-based models were constructed using TCGA and GSE68465 samples; their sensitivity and specificity in predicting LUAD prognosis were confirmed. GLRS was an independent LUAD prognostic factor and contributed to the nomogram to predict patient survival. High GLRS was associated with advanced tumor stage and higher mutation frequencies, estimate scores, and TIDE scores. GLRS-high and GLRS-low patients differed in immune cell infiltration and epithelial-mesenchymal transition (EMT)-related gene expression. Thus, we propose five glycosylation-related gene signatures to predict overall survival and prognostic risks of LUAD. Their regulatory roles may be related to immune invasion, immunotherapy response, mutation, and EMT.

**Key words:** Glycosylation — Lung adenocarcinoma — Prognostic risk — Immune microenvironment — Epithelial-mesenchymal transition

**Abbreviations:** DEGs, differentially expressed genes; EMT, epithelial-mesenchymal transition; FDR, false discovery rate; GEO, Gene Expression Omnibus; GLRS, glycosylation risk score; GSEA, gene set enrichment analysis; HR, hazard ratio; LUAD, lung adenocarcinoma; ROC, receiver operator characteristic; TCGA, The Cancer Genome Atlas; TIDE, tumor immune dysfunction and exclusion; TMB, tumor mutation burden; TME, tumor microenvironment.

## Highlights

- Five prognostic gene signatures of LUAD were identified from glycosylation-related screening.
- GLRS constructed by five gene signatures is an independent prognostic factor of LUAD.
- GLRS-based prognostic model and nomogram can predict LUAD prognostic risks.
- LUAD patients with high and low GLRS present different clinical features, immune landscape, mutation status, and EMT-related gene expression.

**Electronic supplementary material.** The online version of this article (doi: 10.4149/gpb\_2023025) contains Supplementary material.

**Correspondence to:** Xiao-Hui Liu, Department of Thoracic Surgery, Tangshan People's Hospital, Tangshan 063001, Hebei, P.R. China  
E-mail: lxh467373246@163.com

© The Authors 2023. This is an **open access** article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Introduction

With an estimated 1.8 million new cases and 1.5 million deaths *per year*, non-small cell lung cancer (NSCLC) contributes to 85% of lung cancer-related morbidity and mortality (Thai et al. 2021). Lung adenocarcinoma (LUAD) is the most common histological subtype, accounting for 78% of all diagnoses (Thai et al. 2021), followed by squamous cell carcinoma that accounts for 18% of total diagnoses. The 5-year survival rate of early-stage patients with primary and local tumors after surgical resection is greater than 70%; however, 75% of patients with NSCLC are diagnosed at the advanced stage (III/IV) (Knight et al. 2017). In the last decade, advances in therapy have contributed to improved survival, including targeted therapy for common driver mutations; the 2-year survival rate of distant metastatic NSCLC is still less than 20% (Siegel et al. 2021). The likely reason may be the non-universality of several mutations that confers beneficial therapeutic effects only to a minority of patients (Puderecki et al. 2020; Spella and Stathopoulos 2021). Recent studies have shown that genes and long chain noncoding RNA is many cancer prognostic factor including pancreatic ductal adenocarcinoma, ovarian cancer, gastric cancer, lung cancer, glioblastoma, breast cancer, colorectal cancer (Qiu et al. 2022), hepatocellular carcinoma (Qiu et al. 2021; Xie et al. 2022), bladder cancer, head and neck squamous cell carcinoma, renal clear cell carcinoma, endometrial carcinoma, and Cutaneous melanoma (Xie et al. 2021). However, the reproducibility, robustness and clinical validity of the currently developed prognostic markers remain controversial (Tang et al. 2017). Therefore, further development of predictive markers that can accurately identify candidates for targeted therapy is needed.

Glycosylation is the covalent attachment of a monosaccharide or glycan to selected residues of a protein; this is a common but complex post-translational modification (Eichler 2019). Glycosylation of specific proteins is altered with the development of cancer and is involved in multiple cancer-related biological processes, including tumor cell communication, cell-matrix interaction, tumor angiogenesis, and immune regulation (Pinho and Reis 2015). Several tumor-related glycosylated products such as cancer antigen 19-9 (CA19-9), CA125, carcinoembryonic antigen, prostrate-specific antigen, and alpha-fetoprotein are secreted or shed into the bloodstream and can be used as biomarkers for cancer diagnosis, detection, and prognosis (Silsirivanit 2019; Thomas et al. 2021). Studies have analyzed glycosylation patterns in LUAD patients and found that N-glycosylation levels in tissues can help identify and predict LUAD, even at different tumor stages (Ruhaak et al. 2015; Lattová et al. 2020). Several studies have explored the relationship between glycosyltransferase (GT) expres-

sion and LUAD prognosis. For instance, Gu et al. (2004) found that downregulation of *GalNAc-T3* expression was an independent factor in predicting poor prognosis and early recurrence of LUAD. In addition, overexpression of *GALNT2/14* is considered to be related to poor prognosis of LUAD (Yu et al. 2021). Furthermore, *GalNAc-T6* expression is significantly correlated with TNM staging and can independently predict unfavorable overall survival in LUAD patients (Li et al. 2016). These studies confirmed that abnormal expression of GT family genes may affect glycosylation modification and lead to tumor invasion and recurrence after treatment; however, comprehensive studies exploring glycosylation-related genes that may affect the prognosis of LUAD are still lacking.

The present study compiled a list of genes related to glycosylation from public databases and identified gene signatures significantly associated with LUAD prognosis. Based on these genes, we constructed a glycosylation risk score (GLRS)-dependent prognostic model and compared the clinical characteristics, mutation status, immune landscape, participating pathways, and epithelial-mesenchymal transition (EMT)-related gene expression in patients with different prognostic risks. Our study proposes novel glycosylation-related genes that can be used as prognostic markers of LUAD and highlights potential regulatory mechanisms to explain their prognostic predictabilities.

## Materials and Methods

### *Data capturing and pre-processing*

The Cancer Genome Atlas (TCGA) database was used to download the normalized expression matrix and clinical follow-up data of patients with LUAD. After excluding samples with lost survival information, 504 tumor samples and 59 normal controls were retained as the training set. The GSE68465 dataset from the Gene Expression Omnibus (GEO) (Shedden et al. 2008), that contains 442 LUAD samples and is detected on the GPL96[HG-U133A] Affymetrix Human Genome U133A Array platform, was used as the validation set for external validation. Detailed clinical information of patients from TCGA and GSE68465 cohorts is summarized in Table S1 and S2 in Supplementary material, respectively.

### *Screening of glycosylation-related genes*

To explore the glycosylation mechanism, the major family genes encoding GT were analyzed. Therefore, we obtained glycosylation-related genes from the Human Gene Nomenclature Committee (HGNC) (Mohamed Abd-El-Halim et al. 2021), and a list containing 213 genes was downloaded.

### Screening of glycosylation-related differentially expressed genes (DEGs)

For TCGA samples, the expression profiles of LUAD and normal tissues were compared using limma package Version 3.34.7 (Ritchie et al. 2015) to select DEGs. These DEGs were then visualized by a heatmap using R. pheatmap Version 1.0.8 (Wang et al. 2014). By considering the intersection of glycosylation-related genes and DEGs, glycosylation-related DEGs were selected using a Venn diagram for further analysis.

### Selection of prognostic glycosylation-related DEGs

Considering the expression data of glycosylation-related DEGs and survival information of TCGA-LUAD samples, a univariate Cox regression analysis was performed to screen genes associated with prognosis based on their expression levels. Kaplan-Meier curves were plotted to evaluate the association between the expression of prognostic glycosylation-related DEGs and survival status using the R3.6.1 survival package Version 2.41-1 (Rizvi et al. 2018).

### Generation of a GLRS-based model and performance verification

To identify the optimal gene set from prognostic glycosylation-related DEGs, a stepwise Cox regression analysis was performed using the Survminer package Version 0.4.9, and the GLRS was then calculated based on the following formula:

$$\text{GLRS} = h_0(t) * \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)$$

Here,  $\beta$  indicates the regression coefficient,  $h_0(t)$  indicates the baseline risk function, and  $h(t,X)$  indicates the risk function associated with  $X$  at time  $t$ . Thereafter, the GLRS of each sample was calculated, and TCGA and GSE68465 LUAD samples were assigned to GLRS-high and GLRS-low groups based on their respective median values of GLRS. Kaplan-Meier analysis was performed to assess the difference in prognosis between the GLRS-high and GLRS-low groups.

### Prognostic independence analysis and nomogram model construction

The prognostic independence of GLRS and clinical characteristics (including age, sex, pathologic TNM, and tumor stage) were further analyzed using univariate and multivariate Cox regression analyses. Independent prognostic factors were visualized by a forest plot. These independent prognostic factors were then incorporated to establish a nomogram model using the rms package Version 5.1-2 (Zhang et al. 2020) to predict the survival probability of patients with LUAD.

### Correlation analysis of clinical factors and GLRS

To analyze the correlation between clinical factors and GLRS, patients were grouped according to their clinical characteristics, and the difference in GLRS was then compared.

### Mutation analysis

The number of mutations in each gene *per* sample was counted and sorted. The top 20 mutated genes were collected and their mutation frequencies were calculated using the maftools package Version 2.6.05 (Zhang et al. 2019). Furthermore, the tumor mutation burden (TMB) was computed and compared between GLRS-high and GLRS-low groups.

### Correlation analysis of tumor microenvironment (TME) and GLRS

Based on the expression data of TCGA-LUAD samples, we analyzed the proportions of 22 types of immune cells using CIBERSORT (Chen et al. 2018) and then compared differential immune cells (DICs) between GLRS-high and GLRS-low groups. Then, the immune score, stromal score, estimate score, and tumor purity were calculated using the estimate package (<http://127.0.0.1:29606/library/estimate/html/estimateScore.html>) (Hu et al. 2019) and the values were compared between GLRS-high and GLRS-low groups.

### Immune checkpoint response analysis

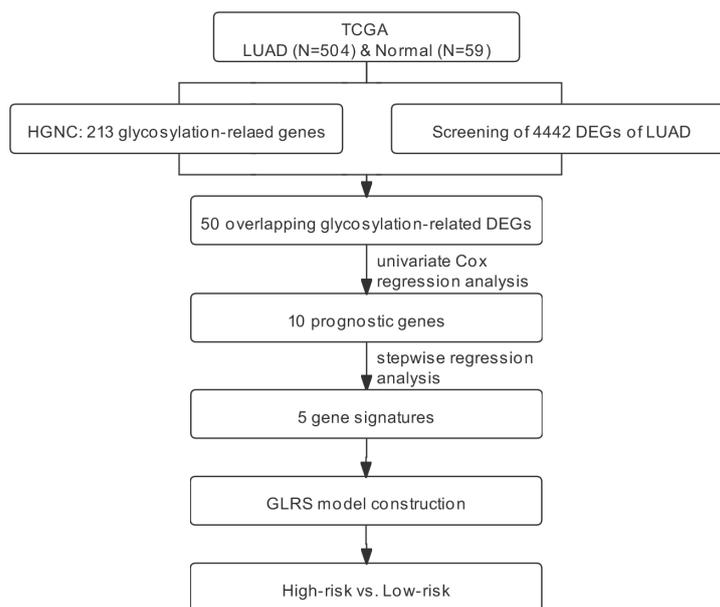
The Tumor Immune Dysfunction and Exclusion (TIDE) database was used to predict the response of LUAD patients to immune checkpoint therapy. The TIDE score of each sample was calculated and the values were compared between GLRS-high and GLRS-low groups.

### Analysis of differences in EMT-related gene expression

Glycosylation can affect EMT in pulmonary interstitial fibrosis and trigger tumor invasion and metastasis. Therefore, we generated an EMT-related gene list through published articles (Tao et al. 2020) and compared the differences in EMT-related gene expression between the two GLRS groups.

### Gene set enrichment analysis (GSEA)

GSEA (Reimand et al. 2019) was used to assess whether a set of genes presented statistical differences between the two biological states. Therefore, GSEA was performed to analyze significantly enriched pathways among GLRS-high and GLRS-low groups.



**Figure 1.** The design and workflow of this study.

### Statistical analysis

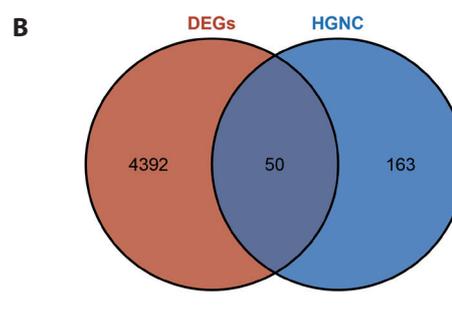
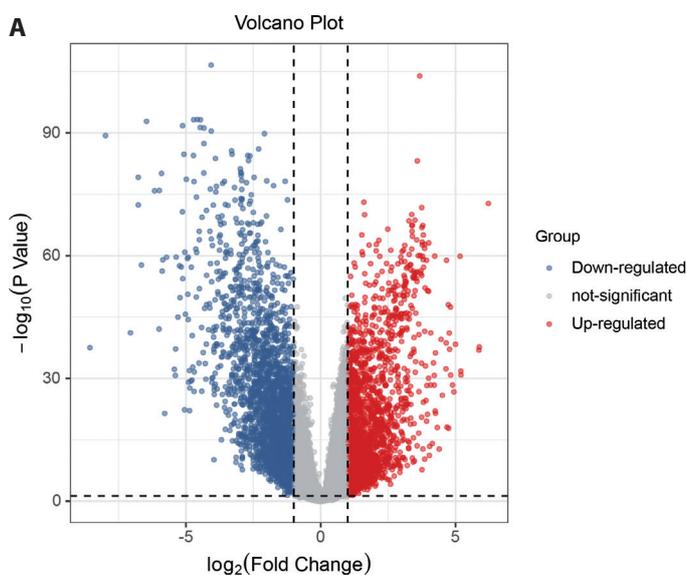
DEGs between LUAD and normal samples were defined at false discovery rate (FDR) < 0.05 and  $|\log_2 \text{fold-change (FC)}| > 1$ . Univariate Cox regression analysis was used to screen glycosylation-related DEGs associated with prognosis, and stepwise Cox regression analysis was applied for screening prognostic signatures to construct the GLRS-based model. Independent prognostic factors were identified using univariate and multivariate Cox regression analyses. Significantly different pathways between GLRS-high and GLRS-low groups were selected with  $|\text{normalized enrichment score (NES)}| > 1$ . Comparisons of

TMB, immune score, stromal score, TIDE score, and EMT-related gene expression between GLRS groups were performed by the Wilcoxon test. Statistical significance was set at  $p < 0.05$ .

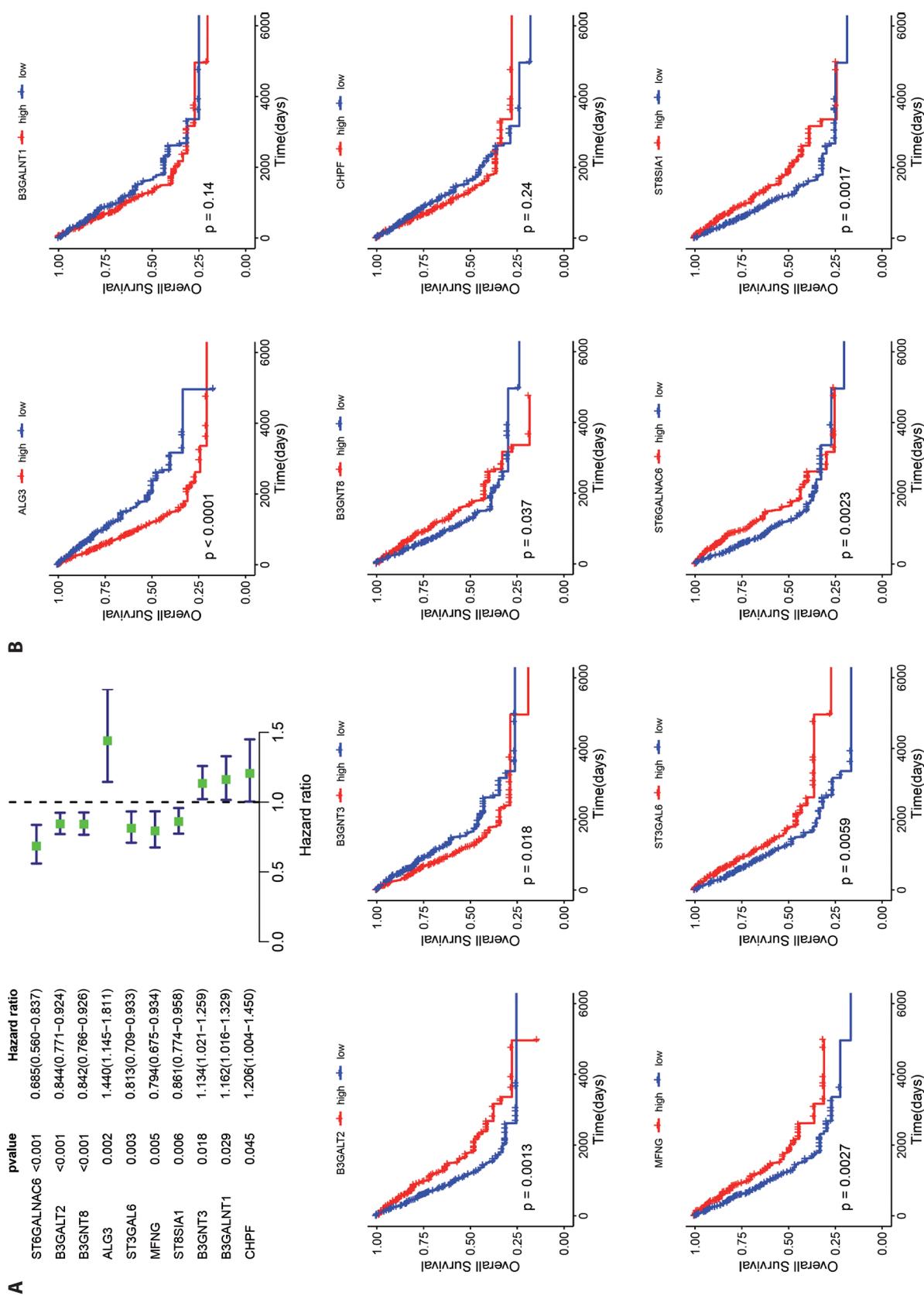
### Results

#### Screening of 50 glycosylation-related DEGs

A flowchart of this study is presented in Figure 1. The comparison of the expression of the genes from LUAD and normal subjects from TCGA revealed 4,442 DEGs. Of these, 1,836 DEGs



**Figure 2.** A total of 50 glycosylation-related DEGs were obtained by identifying the overlap between glycosylation-related genes and DEGs. **A.** The volcano plot shows the upregulated and downregulated DEGs selected at FDR < 0.05, and  $|\log_2 \text{FC}| > 1$ . **B.** The Venn diagram shows the intersection of glycosylation-related genes and DEGs.

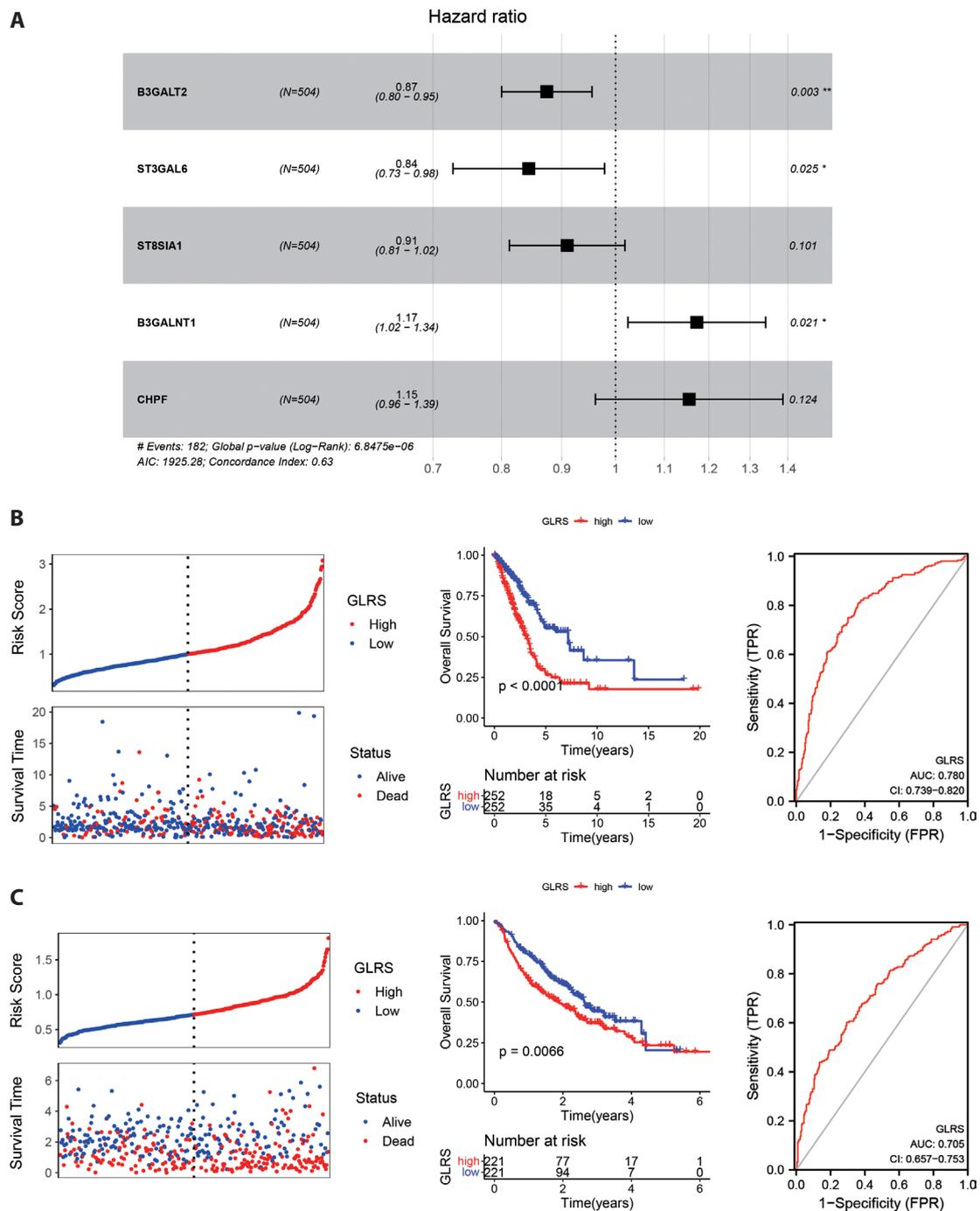


**Figure 3.** Analysis of the prognostic values of the 10 prognostic glycosylation-related DEGs. **A.** Screening of 10 genes associated with prognosis using a univariate Cox regression analysis; six of these genes were protective factors (HR < 1) and four were risk factors (HR > 1). **B.** Kaplan-Meier curves show the relationships between the expression of 10 candidate genes and LUAD prognosis.

were upregulated and 2,606 were downregulated (Fig. 2A; Fig. S1 in Supplementary materials). We identified the genes that overlapped with the 213 glycosylation-related genes and selected 50 glycosylation-related DEGs for further analysis (Fig. 2B).

*Analysis of prognostic genes*

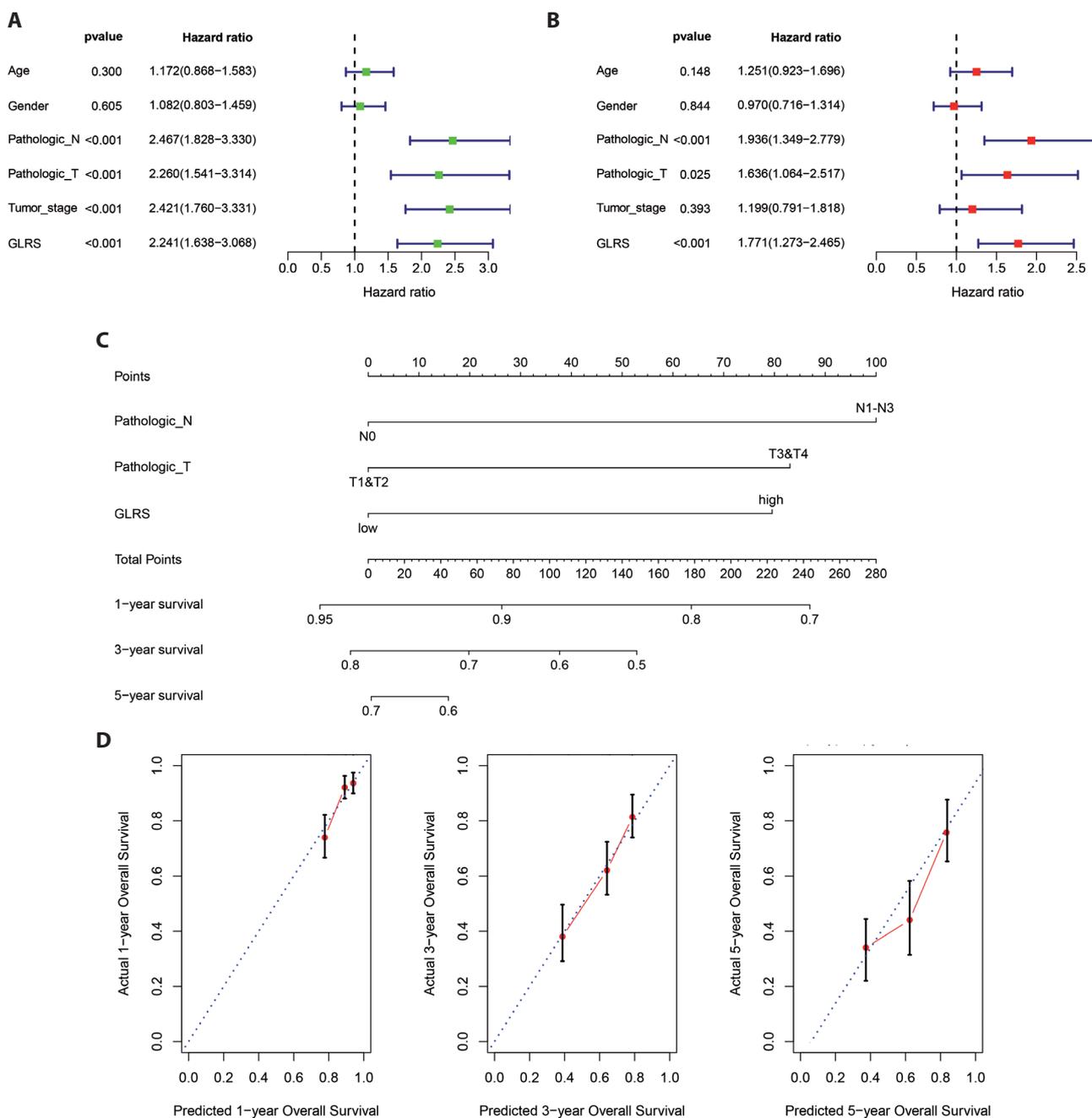
Univariate Cox regression analysis was performed on 50 glycosylation-related DEGs, of which 10 were found to



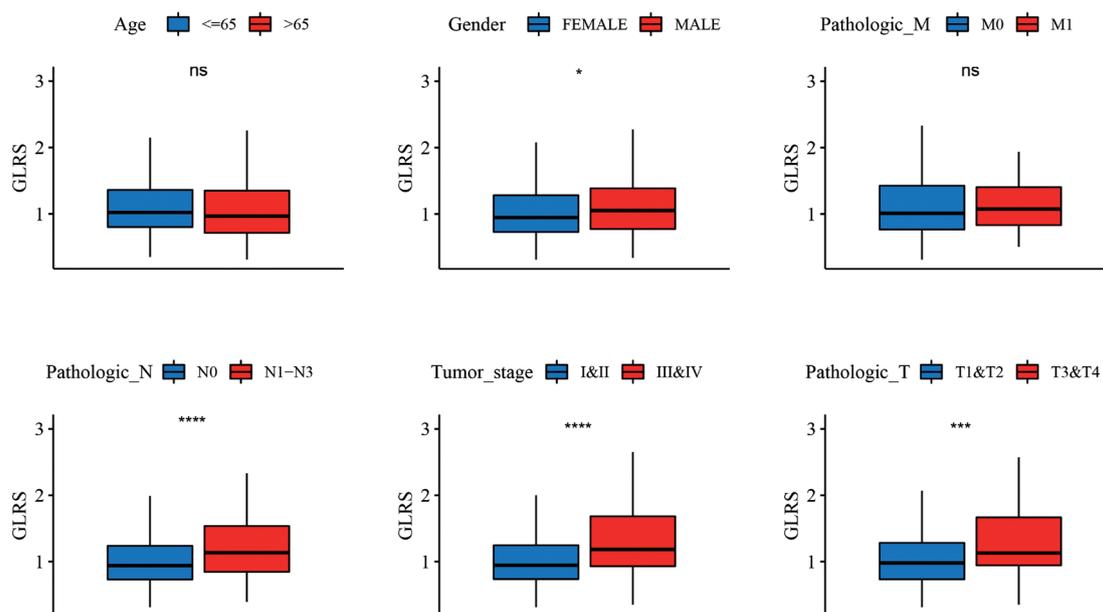
**Figure 4.** Construction of a GLRS-based model in TCGA, the efficiency of which was further validated in the GSE68465 dataset. **A.** The optimal gene set, including five gene signatures, was determined by the stepwise Cox regression analysis. **B., C.** Kaplan-Meier curves present the close relationship between GLRS and LUAD prognosis, and ROC curves were plotted to evaluate the model capability in predicting LUAD prognostic risks, based on TCGA and GSE68465 samples.

correlate with LUAD prognosis with a cutoff of  $p < 0.05$ . Their prognostic values are shown in Figure 3A. Six of these genes are protective factors (hazard ratio (HR) < 1) and four are risk factors (HR > 1). We evaluated the association between the expression levels of 10 prognostic genes and survival

status; Kaplan-Meier curves suggested that patients with high expression levels of *ALG3* and *B3GNT3* had a significantly unfavorable prognosis, whereas overexpression of *B3GALT2*, *B3GNT8*, *MFNG*, *ST3GAL6*, *ST6GALNAC6*, and *ST8SIA1* was associated with a better survival status (Fig. 3B).



**Figure 5.** Screening of independent prognostic factors and construction of a nomogram model to predict LUAD prognosis. **A., B.** Screening of independent prognostic factors of LUAD using univariate (A) and multivariate (B) Cox regression analyses. **C.** An individualized nomogram model was constructed based on independent prognostic risk factors to predict 1-, 3-, and 5-year survival probabilities for patients with LUAD. **D.** Calibration curves were created to verify the performance of the nomogram model.



**Figure 6.** Comparison of GLRS between patients with different clinical characteristics. \*  $p < 0.05$ ; \*\*\*  $p < 0.001$ ; \*\*\*\*  $p < 0.0001$ .

#### Generation and validation of the GLRS-based model

By applying the stepwise Cox regression analysis, we selected the optimal gene set comprising *B3GALT2*, *ST3GAL6*, *ST8SIA1*, *B3GALNT1*, and *CHPF*, which were further identified as gene signatures (Fig. 4A). Based on their regression coefficients and expression levels, the GLRS of samples was calculated and a GLRS-based prognostic model was constructed. According to the median value of GLRS, LUAD patients were divided into GLRS-high and GLRS-low groups. Survival analysis suggested that patients with a high GLRS had a significantly worse survival status than patients with a low GLRS (Fig. 4B). Prognostic model construction and patient grouping were also performed based on the GSE68465 dataset for independent external validation. We found that the patients from GLRS-high and GLRS-low groups presented significantly different survival probabilities (Fig. 4C). Receiver operator characteristic (ROC) curves were plotted based on TCGA and GSE68465 samples, and the area under the curves (AUCs) were found to be 0.780 and 0.705, respectively, indicating high sensitivity and specificity of the GLRS-based model in predicting LUAD prognostic risk (Fig. 4B,C).

#### Construction of an individualized nomogram model to predict prognosis

We further assessed the prognostic independence of the clinical characteristics and GLRS status in TCGA-LUAD samples using univariate and multivariate Cox regression

analyses. The results identified pathological N, pathological T, and GLRS as independent prognostic risk factors for LUAD (Fig. 5A,B). These factors were then incorporated to generate a nomogram model for predicting 1-, 3-, and 5-year survival rates (Fig. 5C). We calculated the consistency index (C-index) for this model; the closer the value is to 1, the higher is the prediction accuracy of the model. The C-index of this nomogram model was 0.702, suggesting the high accuracy of this model in predicting LUAD prognosis. Thereafter, calibration curves were plotted to verify the performance of the nomogram model. As shown in Figure 5D, the predicted 1-, 3-, and 5-year overall survival fitted well with the actual values. Thus, the GLRS may contribute to the superior predictive ability of the nomogram model.

#### Associations between the clinical characteristics and GLRS

We performed stratified analysis of TCGA-LUAD patients based on their clinical characteristics and compared the GLRS status among patients from different subgroups. Male patients had significantly higher GLRS than female patients ( $p < 0.05$ , Fig. 6). In addition, patients with advanced tumor stage and pathologic N and T stages had significantly higher GLRS than early-stage patients ( $p < 0.001$ , Fig. 6).

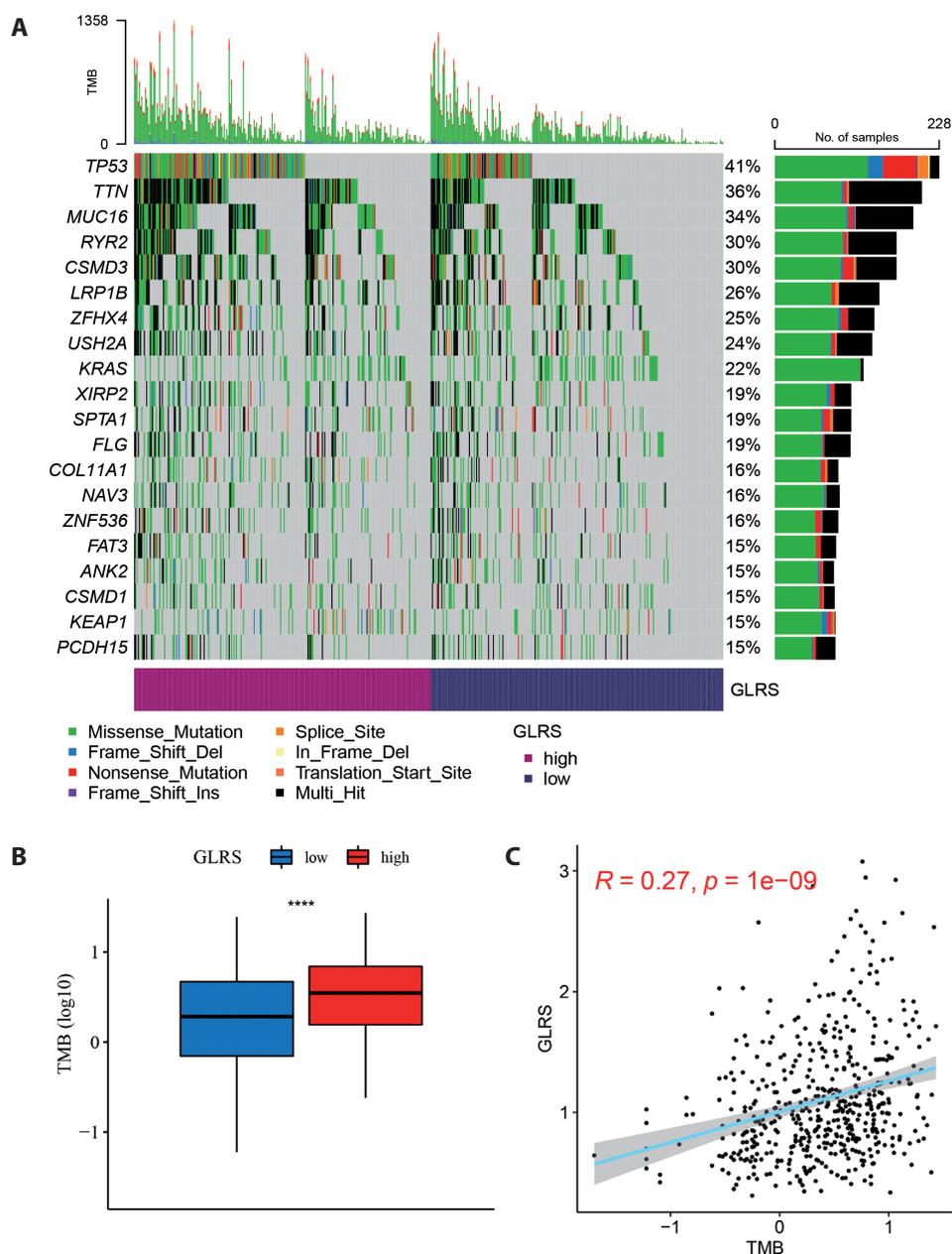
#### Differences in mutations between risk groups

Gene mutation frequencies in TCGA-LUAD samples were calculated, and the top 20 mutated genes are shown

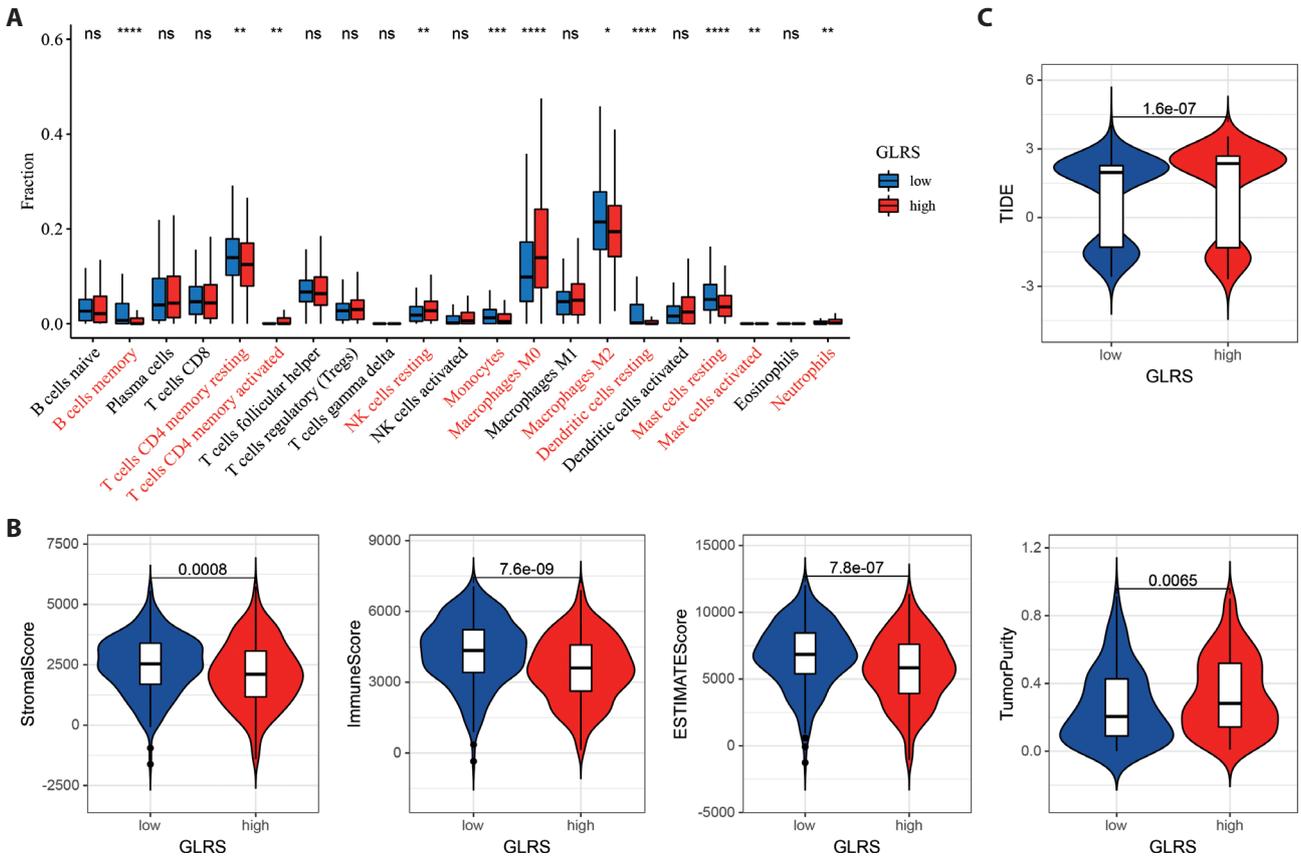
in Figure 7A. The mutation frequencies of *TP53*, *TTN*, *MUC16*, *RYR2*, and *CSMD3* were 30% or higher, and the main mutation type was missense mutation. We also observed that mutation frequencies of these genes were higher in the GLRS-high group than in the GLRS-low group. Patients with high GLRS exhibited significantly higher TMB than those with low GLRS ( $p < 0.0001$ , Fig. 7B). There was a significant positive correlation between TMB and GLRS (Fig. 7C).

#### Analysis of association between TME and GLRS

CIBERSORT was used to assess the proportion of 22 types of immune cells. We compared their infiltration abundance between GLRS-high and GLRS-low groups. With a cutoff of  $p < 0.05$ , 11 DICs were obtained; the two groups showed the most significant differences in the infiltrations of memory B cells, M0 macrophages, resting dendritic cells, and resting mast cells ( $p < 0.0001$ , Fig. 8A). We employed the estimate algo-



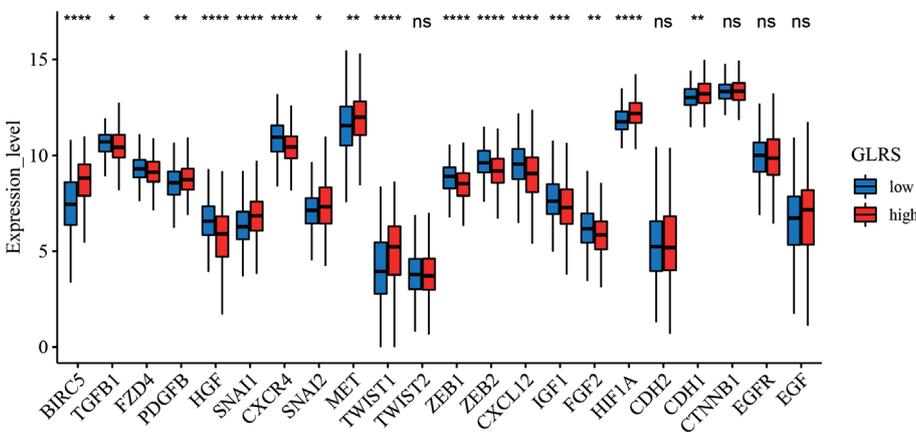
**Figure 7.** Comparison of gene mutation frequencies and TMB between GLRS-high and GLRS-low groups. **A.** Top 20 mutated genes and their mutation frequencies in GLRS-high and GLRS-low groups. **B.** The TMB was compared between GLRS-high and GLRS-low groups. \*\*\*\*  $p < 0.0001$ . **C.** Correlation analysis indicated a positive correlation between TMB and GLRS.



**Figure 8.** Comparison of the immune landscape, including immune cell infiltration, immune scores, and TIDE scores, between GLRS-high and GLRS-low groups. **A.** The infiltration abundances of 22 types of immune cells were evaluated by CIBERSORT and then compared between the two groups. \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ ; \*\*\*\*  $p < 0.0001$ ; ns, not significant. **B.** By applying the estimate algorithm, the stromal score, immune score, estimate score, and tumor purity were evaluated and compared between the two groups. **C.** The TIDE score was compared to assess differences in immune checkpoint responses.

rithm and found that stromal, immune, and estimated scores were significantly higher and tumor purity was significantly lower in the GLRS-high group than in the GLRS-low group (Fig. 8B). We calculated the TIDE score to predict immune

checkpoint response, and found it to be significantly higher for the GLRS-high group than for the GLRS-low group (Fig. 8C). These results indicate that patients with different GLRS presented different immune landscapes.



**Figure 9.** Comparison of the expression levels of 22 EMT-related genes between GLRS-high and GLRS-low groups.

**Table 1.** Enriched KEGG pathways of GLRS-high and GLRS-low groups

Terms	Size	ES	NES	NOM <i>p</i> value	GLRS
KEGG_SPLICEOSOME	114	0.53633744	2.1494124	0	high
KEGG_CELL_CYCLE	118	0.661843	2.0438285	0	high
KEGG_DNA_REPLICATION	36	0.7265612	1.8697815	0.00209205	high
KEGG_MISMATCH_REPAIR	23	0.68897027	1.9233372	0.004065041	high
KEGG_HOMOLOGOUS_RECOMBINATION	26	0.69105154	1.7048123	0.012219959	high
KEGG_P53_SIGNALING_PATHWAY	67	0.4916378	1.7475405	0.01713062	high
KEGG_NUCLEOTIDE_EXCISION_REPAIR	44	0.48294634	1.747494	0.041322313	high
KEGG_BASE_EXCISION_REPAIR	33	0.52697057	1.7044103	0.04831933	high
KEGG_GNRH_SIGNALING_PATHWAY	100	-0.5272094	-1.8053851	0	low
KEGG_FC_EPSILON_RI_SIGNALING_PATHWAY	78	-0.5531674	-1.750391	0.006097561	low
KEGG_ALPHA_LINOLENIC_ACID_METABOLISM	18	-0.6991381	-1.5868899	0.012958963	low
KEGG_ETHER_LIPID_METABOLISM	32	-0.56197333	-1.5557991	0.03177966	low
KEGG_ARACHIDONIC_ACID_METABOLISM	57	-0.56098104	-1.4829572	0.037037037	low
KEGG_GLYCEROPHOSPHOLIPID_METABOLISM	76	-0.40054628	-1.4754204	0.03956044	low
KEGG_FATTY_ACID_METABOLISM	40	-0.51944405	-1.5524104	0.040899795	low
KEGG_MAPK_SIGNALING_PATHWAY	265	-0.37193426	-1.4349014	0.04208417	low

KEGG, Kyoto Encyclopedia of Genes and Genomes; GLRS, glycosylation risk score; ES, enrichment score; NES, normalized enrichment score; NOM, nominal.

#### *Difference in EMT gene expression between risk groups*

To further understand the biological significance of glycation disorders, we generated a list of genes related to EMT from published articles. Their expression levels in TCGA-LUAD samples were analyzed and compared between the GLRS groups. The results suggest that among 22 EMT-related genes, 17 showed significant expression differences between the two groups (Fig. 9). Thus, there seems to exist a potential relationship between EMT and glycosylation in LUAD prognosis.

#### *Difference in enriched pathways between risk groups*

By applying GSEA, we screened 16 significantly different Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways between GLRS-high and GLRS-low groups at selection thresholds of nominal  $p < 0.05$  and  $|NES| > 1$  (Table 1). Eight KEGG pathways related to the GLRS-high group were mainly involved in DNA repair, while eight pathways related to the GLRS-low group were mainly related to metabolism.

## Discussion

Asn-linked N-glycosylation and Ser/Thr-linked O-glycosylation are two modalities of galactosylation. Abnormal glycosylation detected in tumor cells is often associated with adverse outcomes in cancer patients, and GT plays an important role in the synthesis of glycans and the promo-

tion of tumor metastasis (Fu et al. 2016). Considering the potential effects of GT in LUAD metastasis and recurrence, we identified 213 GT family genes. We exploited the clinical information of TCGA samples to identify a gene set containing five gene signatures that significantly correlated with the prognosis of LUAD using univariate and stepwise Cox regression analyses. ROC curves also indicated that the prognostic model constructed based on these five genes had high sensitivity and specificity for predicting LUAD prognosis in both the TCGA training set and the GSE68465 validation set. As an independent predictor, GLRS contributed in predicting the 1-, 3-, and 5-year survival probabilities for patients with LUAD in the nomogram model. Furthermore, patients with higher GLRS were mainly in the advanced tumor stage and had higher gene mutation frequencies and lower survival. These results are consistent with the clinical reality because patients with advanced tumors face more severe survival challenges (Kuhn et al. 2018). Our observations further indicate that GLRS is reliable in predicting the prognosis of LUAD and that the prognostic value of the five gene signatures is more definite.

The five gene signatures we identified include *B3GALT2*, *ST3GAL6*, *ST8SIA1*, *B3GALNT1*, and *CHPF*. Of these, *B3GALT2* belongs to the family of GT that may affect the prognosis of LUAD through methylation-related molecular mechanisms (Meng et al. 2021). *ST3GAL6* encodes sialic acid transferase, and the activity in its promoter region is negatively correlated with tumor size, lymph node metastasis, distant metastasis, and tumor stage (Hu et al. 2019). *ST3GAL6* was found to be an excellent prognostic predictor of LUAD

by affecting metabolic pathways (Ma et al. 2021). These findings are consistent with our results, and *B3GALT2* and *ST3GAL6* both showed significant prognostic associations for LUAD prognosis. Higher expression levels of *B3GALT2* and *ST3GAL6* indicated lower prognostic risks in patients, and these genes may be regulated at their promoter regions to inhibit tumor cell growth and metastasis. In addition, we found that another galactosyl transferase family gene, *B3GALNT1*, is a prognostic risk factor for LUAD. Based on its protein structure, Umeyama et al. (2014) speculated that *B3GALNT1* is a potential key gene mediating NSCLC metastasis. This finding provides a basis for *B3GALNT1* as a potential drug treatment target, as downregulation of *B3GALNT1* expression can inhibit tumor metastasis and consequently reduce the prognostic risk of recurrence in patients with LUAD.

The infiltration of key immune cells was compared among patients with different prognoses, and the results indicated that the proportion of memory B cells, resting dendritic cells, and resting mast cells significantly decreased in the GLRS-high group, whereas the infiltration of M0 macrophages significantly increased in GLRS-high patients as compared to that in GLRS-low group. Liu et al. (2017) evaluated tumor-infiltrating immune cells in 492 LUAD patients and found that memory B cell loss or M0 macrophage proliferation was associated with poor prognosis in early LUAD, which is highly consistent with our results. Another study indicated that the affinity maturation process of B cells depends on the selection of N-glycosylation sites (Koers et al. 2019). In addition, there is extensive N-glycosylation in the B cell receptor variable domain during the process of B cell response that plays an important role in autoimmunity (Vergoesen et al. 2019). Furthermore, M0 macrophages are considered to have the potential to polarize into the M2 phenotype and act as an antitumor immunosuppressive factor in lung cancer (Pritchard et al. 2020). Hu et al. found an increase in the proportion of M0 in lung cancer samples as compared to that in paracarcinoma samples (Hu et al. 2020). Both O-GlcNAcylation and extracellular glycosylation are thought to be involved in the tumor-related macrophage function and polarization (Mantuano et al. 2019). Therefore, we can speculate that glycosylation may increase the prognostic risk of LUAD by regulating tumor immune invasion and activating tumor cell regeneration and metastasis. TIDE score is used to assess tumor immune evasion and has been found to be more accurate than programmed death ligand 1 (PD-L1) expression and TMB in predicting survival outcomes in cancer patients treated with immune checkpoint blockade (Kaderbhai et al. 2019; Keenan et al. 2019). The TIDE score also reflects T lymphocyte dysfunction and immunosuppressive factor rejection (Wang et al. 2020). Our results suggest that the TIDE score increased in patients with high GLRS, probably owing to the activation of an antitumor immune

escape mechanism that resulted in a reduced response to immune checkpoint blockade therapy and adverse clinical outcomes.

EMT is a developmental process where tumor cells are reactivated and lose polarity and adhesion to epithelial cells, leading to cell migration and mesenchymal phenotypic transformation. This phenomenon is supported by abnormal glycosylation and hyper O-GlcNAcylation (Da Fonesca et al. 2016; Carvalho-Cruz et al. 2017). In this study, EMT-related genes were differentially expressed between GLRS-high and GLRS-low groups, suggesting that the prognostic value of glycosylation-related genes may be associated with EMT in tumor progression. The overexpression of *BIRC5* was significantly associated with poor overall survival in patients with LUAD (Cao et al. 2019). *BIRC5* also rapidly responds to N-glycation inhibition under endoplasmic reticulum stress (Maldonado-Agurto et al. 2019). *TWIST1* is another EMT-related gene that is significantly overexpressed in the high prognostic risk group. *TWIST1* is a transcription factor of EMT that is essential for oncogene-driven NSCLC tumorigenesis (Yochum et al. 2017). Co-overexpression of *HIF-1 $\alpha$*  and *TWIST1* is considered as a predictive factor for tumor recurrence in NSCLC patients, and *HIF-1 $\alpha$*  can regulate *TWIST1* to promote tumor metastasis (Hung et al. 2009). In the present study, the level of *HIF1A* significantly increased in GLRS-high samples. Interestingly, *TWIST1* expression can be activated by galectin-4, and parallel changes in galectin-4 and O-glycosylation facilitate progeny renewal in distal tumors, leading to more aggressive cancers (Tsai et al. 2016). Regulation of *HIF1A* by O-GlcNAcylation also affects glycolytic metabolism and survival stress signal transduction in cancer cells (Ferrer et al. 2014). Therefore, abnormal expression of these genes may be involved in the regulation of glycation and glycolysis metabolism under EMT activation of tumor cells, leading to tumor metastasis and adverse clinical outcomes in patients with LUAD.

The lack of experimental validation is a clear limitation of this study. How candidate genes affect the prognosis of LUAD through glycosylation modification and the involvement of immune cell infiltration in this process remain unknown. Therefore, collection of clinical solid tumor samples and evaluation of the underlying mechanism are warranted to compare the differences in glycosylation patterns and immune regulatory pathways in tumor tissues under different prognostic risks.

## Conclusion

We identified five gene signatures, *B3GALT2*, *ST3GAL6*, *ST8SIA1*, *B3GALNT1*, and *CHPF*, that form a glycosylation-related gene set and constructed GLRS-based and nomogram models to predict the prognosis of LUAD. High

GLRS was associated with advanced tumors, high TMB, and high TIDE score. There were also significant differences in immune cell infiltration and EMT-related gene expression between GLRS-high and GLRS-low groups. Our findings provide a broader perspective for exploring the recurrence and metastasis of tumors associated with glycosylation modification in LUAD.

**Acknowledgments.** We acknowledge TCGA database for providing their platforms and contributors for uploading their meaningful datasets.

**Conflict of interest.** The authors declare that no competing interests exist.

**Author contributions.** Yi-Ming Yang: contributed to conception and design; Xiao-Hui Liu: contributed to all experimental work, data and statistical analysis, and interpretation of data; Ming-Ming Zhang: was responsible for overall supervision; Hao-Yu Fu: drafted the manuscript, which was revised by Xiao-Hui Liu. All authors read and approved the final manuscript.

**Data availability.** Previously reported gene expression and clinical data were used to support this study and are available at The Cancer Genome Atlas (TCGA, <https://portal.gdc.cancer.gov/>). These prior studies (and datasets) are cited at relevant places within the text.

## References

- Cao Y, Zhu W, Chen W, Wu J, Hou G, Li Y (2019): Prognostic value of BIRC5 in lung adenocarcinoma lacking EGFR, KRAS, and ALK mutations by integrated bioinformatics analysis. *Dis. Markers* **2019**, 5451290  
<https://doi.org/10.1155/2019/5451290>
- Carvalho-Cruz P, Alisson-Silva F, Todeschini AR, Dias WB (2017): Cellular glycosylation senses metabolic changes and modulates cell plasticity during epithelial to mesenchymal transition. *Dev. Dyn.* **247**, 481-491  
<https://doi.org/10.1002/dvdy.24553>
- Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA (2018): Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol. Biol.* **1711**, 243  
[https://doi.org/10.1007/978-1-4939-7493-1\\_12](https://doi.org/10.1007/978-1-4939-7493-1_12)
- Da Fonesca LM, Da Silva VA, Freire-de-Lima L, Previato JP, Mendonça-Previato L, Capella MAM (2016): Glycosylation in cancer: Interplay between multidrug resistance and epithelial-to-mesenchymal transition? *Front. Oncol.* **6**, 158  
<https://doi.org/10.3389/fonc.2016.00158>
- Eichler J (2019): Protein glycosylation. *Curr. Biol.* **29**, R229-R231  
<https://doi.org/10.1016/j.cub.2019.01.003>
- Ferrer CM, Lynch TP, Sodi VL, Falcone JN, Schwab LP, Paacock DL, Vocadlo DJ, Seagroves TN, Reginato MJ (2014): O-GlcNAcylation regulates cancer metabolism and survival stress signaling via regulation of the HIF-1 pathway. *Mol. Cell* **54**, 820-831  
<https://doi.org/10.1016/j.molcel.2014.04.026>
- Fu C, Zhao H, Wang Y, Cai H, Xiao Y, Zeng Y, Chen H (2016): Tumor-associated antigens: Tn antigen, sTn antigen, and T antigen. *HLA* **88**, 275-286  
<https://doi.org/10.1111/tan.12900>
- Gu C, Oyama T, Osaki T, Li J, Takenoyama M, Izumi H, Sugio K, Kohno K, Yasumoto K (2004): Low expression of polypeptide GalNAc N-acetylgalactosaminyl transferase-3 in lung adenocarcinoma: impact on poor prognosis and early recurrence. *Br. J. Cancer* **90**, 436-442  
<https://doi.org/10.1038/sj.bjc.6601531>
- Hu B, Shi X, Du X, Xu M, Wang Q, Zhao H (2020): Pattern of immune infiltration in lung cancer and its clinical implication. *Clin. Chim. Acta* **508**, 47-53  
<https://doi.org/10.1016/j.cca.2020.04.036>
- Hu D, Zhou M, Zhu X (2019): Deciphering immune-associated genes to predict survival in clear cell renal cell cancer. *Biomed. Res. Int.* **2019**, 1-10  
<https://doi.org/10.1155/2019/2506843>
- Hu J, Shan Y, Ma J, Pan Y, Zhou H, Jiang L, Jia L (2019): LncRNA ST3Gal6-AS1/ST3Gal6 axis mediates colorectal cancer progression by regulating  $\alpha$ -2,3 sialylation via PI3K/Akt signaling. *Int. J. Cancer* **145**, 450-460  
<https://doi.org/10.1002/ijc.32103>
- Hung JJ, Yang MH, Hsu HS, Hsu WH, Liu JS, Wu KJ (2009): Prognostic significance of hypoxia-inducible factor-1 $\alpha$ , TWIST1 and Snail expression in resectable non-small cell lung cancer. *Thorax* **64**, 1082-1089  
<https://doi.org/10.1136/thx.2009.115691>
- Kaderbhai C, Tharin Z, Ghiringhelli F (2019): The role of molecular profiling to predict the response to immune checkpoint inhibitors in lung cancer. *Cancers* **11**, 201  
<https://doi.org/10.3390/cancers11020201>
- Keenan TE, Burke KP, Van Allen EM (2019): Genomic correlates of response to immune checkpoint blockade. *Nat. Med.* **25**, 389-402  
<https://doi.org/10.1038/s41591-019-0382-x>
- Knight SB, Crosbie PA, Balata H, Chudziak J, Hussell T, Dive C (2017): Progress and prospects of early detection in lung cancer. *Open Biol.* **7**, 170070  
<https://doi.org/10.1098/rsob.170070>
- Koers J, Nil D, Ooijevaar-De HP, Nota B, van de Bovenkamp FS, Vidarsson G, Rispens T (2019): Biased N-glycosylation site distribution and acquisition across the antibody V region during B cell maturation. *J. Immunol.* **202**, 2220-2228  
<https://doi.org/10.4049/jimmunol.1801622>
- Kuhn E, Morbini P, Cancellieri A, Damiani S, Cavazza A, Comin CE (2018): Adenocarcinoma classification: patterns and prognosis. *Pathologica* **110**, 5-11
- Lattová E, Skříčková J, Hausnerová J, Frola L, Křen L, Ichnatová I, Zdráhal Z, Bryant J, Popovič M (2020): N-Glycan profiling of lung adenocarcinoma in patients at different stages of disease. *Mod. Pathology* **33**, 1146-1156  
<https://doi.org/10.1038/s41379-019-0441-3>
- Li Z, Yamada S, Wu Y, Wang KY, Liu YP, Uramoto H, Kohno K, Sasaguri Y (2016): Polypeptide N-acetylgalactosaminyltransferase-6 expression independently predicts poor overall survival in patients with lung adenocarcinoma after curative resection. *Oncotarget* **7**, 54463-54473

- <https://doi.org/10.18632/oncotarget.9810>
- Ma W, Liang J, Liu J, Tian D, Chen Z (2021): Establishment and validation of an eight-gene metabolic-related prognostic signature model for lung adenocarcinoma. *Aging* **13**, 8688-8705  
<https://doi.org/10.18632/aging.202681>
- Maldonado-Agurto R, Dickson AJ (2019): Multiplexed digital mRNA expression analysis profiles system-wide changes in mRNA abundance and responsiveness of UPR-specific gene expression changes during batch culture of recombinant chinese hamster ovary cells. *Biotechnol. J.* **13**, e1700429  
<https://doi.org/10.1002/biot.201700429>
- Mantuano NR, Oliveira-Nunes MC, Alisson-Silva F, Dias WB, Todeschini AR (2019): Emerging role of glycosylation in the polarization of tumor-associated macrophages. *Pharmacol. Res.* **146**, 104285  
<https://doi.org/10.1016/j.phrs.2019.104285>
- Meng J, Cao L, Song H, Chen L, Qu Z (2021): Integrated analysis of gene expression and DNA methylation datasets identified key genes and a 6-gene prognostic signature for primary lung adenocarcinoma. *Genet. Mol. Biol.* **44**, e20200465  
<https://doi.org/10.1590/1678-4685-gmb-2020-0465>
- Mohamed Abd-El-Halim Y, El Kaoutari A, Silvy F, Rubis M, Bigonnet M, Roques J, Cros J, Nicolle R, Iovanna J, Dussetti N, Mas E (2021): A glycosyltransferase gene signature to detect pancreatic ductal adenocarcinoma patients with poor prognosis. *EBioMedicine* **71**, 103541  
<https://doi.org/10.1016/j.ebiom.2021.103541>
- Pinho SS, Reis CA (2015): Glycosylation in cancer: mechanisms and clinical implications. *Nat. Rev. Cancer* **15**, 540-555  
<https://doi.org/10.1038/nrc3982>
- Pritchard A, Tousif S, Wang Y, Hough K, Khan S, Strenkowski J, Chacko BK, Darley-Usmar VM, Deshane JS (2020): Lung tumor cell-derived exosomes promote M2 macrophage polarization. *Cells* **9**, 1303  
<https://doi.org/10.3390/cells9051303>
- Puderecki M, Szumio J, Marzec-Kotarska B (2020): Novel prognostic molecular markers in lung cancer. *Oncol. Lett.* **20**, 9-18  
<https://doi.org/10.3892/ol.2020.11541>
- Qiu Y, Li H, Xie J, Qiao X, Wu J (2021): Identification of ABCC5 among ATP-binding cassette transporter family as a new biomarker for hepatocellular carcinoma based on bioinformatics analysis. *Int. J. Gen. Med.* **14**, 7235-7246  
<https://doi.org/10.2147/IJGM.S333904>
- Qiu Y, Li H, Zhang Q, Wu J (2022): Ferroptosis-related long non-coding RNAs as prognostic marker for colon adenocarcinoma. *Appl. Bionics Biomech.* **2022**, 5220368  
<https://doi.org/10.1155/2022/5220368>
- Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Ros-tamianfar A, Wadi L, Meyer M, Wong J, Xu C, et al. (2019): Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat. Protoc.* **14**, 482-517  
<https://doi.org/10.1038/s41596-018-0103-9>
- Ritchie ME, Belinda P, Wu D, Hu Y, Law CW, Shi W, Smith GK (2015): Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47  
<https://doi.org/10.1093/nar/gkv007>
- Rizvi AA, Karaesmen E, Morgan M, Preus L, Wang J, Sovic M, Hahn T, Sucheston-Campbell LE (2018): Gwasurvivr: An R package for genome-wide survival analysis. *Bioinformatics* **35**, 1968-1970  
<https://doi.org/10.1093/bioinformatics/bty920>
- Ruhaak LR, Taylor SL, Stroble C, Nguyen UT, Parker EA, Song T, Lebrilla CB, Rom WN, Pass H, Kim K, et al. (2015): Differential N-glycosylation patterns in lung adenocarcinoma tissue. *J. Proteome Res.* **14**, 4538-4549  
<https://doi.org/10.1021/acs.jproteome.5b00255>
- Shedden K, Taylor JMG, Enkemann SA, Tsao MS, Yeatman TJ, Gerald WL, Eschrich S, Jurisica I, Giordano TJ, Misek DE, et al. (2008): Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat. Med.* **14**, 822-827  
<https://doi.org/10.1038/nm.1790>
- Siegel RL, Miller KD, Fuchs HE, Jemal A (2021): Cancer statistics, 2021. *CA Cancer J. Clin.* **71**, 7-33  
<https://doi.org/10.3322/caac.21654>
- Silsirivanit A (2019): Glycosylation markers in cancer. *Adv. Clin. Chem.* **89**, 189-213  
<https://doi.org/10.1016/bs.acc.2018.12.005>
- Spella M, Stathopoulos GT (2021): Immune resistance in lung adenocarcinoma. *Cancers* **13**, 384  
<https://doi.org/10.3390/cancers13030384>
- Tang H, Wang S, Xiao G, Schiller J, Papadimitrakopoulou V, Minna J, Wistuba I, Xie Y (2017): Comprehensive evaluation of published gene expression prognostic signatures for biomarker-based lung cancer clinical studies. *Ann. Oncol.* **28**, 733-740  
<https://doi.org/10.1093/annonc/mdw683>
- Tao C, Huang K, Shi J, Hu Q, Li K, Zhu X (2020): Genomics and prognosis analysis of epithelial-mesenchymal transition in glioma. *Front. Oncol.* **10**, 183  
<https://doi.org/10.3389/fonc.2020.00183>
- Thai AA, Solomon BJ, Sequist LV, Gainor JF, Heist RS (2021): Lung cancer. *Lancet* **398**, 535-554  
[https://doi.org/10.1016/S0140-6736\(21\)00312-3](https://doi.org/10.1016/S0140-6736(21)00312-3)
- Thomas D, Rathinavel AK, Radhakrishnan P (2021): Altered glycosylation in cancer: A promising target for biomarkers and therapeutics. *Biochim. Biophys. Acta* **1875**, 188464  
<https://doi.org/10.1016/j.bbcan.2020.188464>
- Tsai CH, Tzeng SF, Chao TK, Tsai CY, Yang YC, Lee MT, Hwang JJ, Chou YC, Tsai MH, Cha TL, Hsiao PW (2016): Metastatic progression of prostate cancer is mediated by autonomous binding of galectin-4-O-glycan to cancer cells. *Cancer Res.* **76**, 5756-5767  
<https://doi.org/10.1158/0008-5472.CAN-16-0641>
- Umeyama H, Iwadate M, Taguchi YH (2014): TINAGL1 and B3GALNT1 are potential therapy target genes to suppress metastasis in non-small cell lung cancer. *BMC Genomics* **15** (Suppl. 9), S2  
<https://doi.org/10.1186/1471-2164-15-S9-S2>
- Vergroesen RD, Slot LM, van Schaik BDC, Koning MT, Rispens T, van Kampen AHC, Toes REM, Scherer HU (2019): N-glycosylation site analysis of citrullinated antigen-specific B-cell receptors indicates alternative selection pathways during autoreactive B-cell development. *Front. Immunol.* **10**, 2092

- <https://doi.org/10.3389/fimmu.2019.02092>
- Wang L, Cao C, Ma Q, Zeng Q, Wang H, Cheng Z, Qi J, Ma H, Nian H, Wang Y (2014): RNA-seq analyses of multiple meristems of soybean: novel and alternative transcripts, evolutionary and functional implications. *BMC Plant Biol.* **14**, 169  
<https://doi.org/10.1186/1471-2229-14-169>
- Wang Q, Li M, Yang M, Yang Y, Song F, Zhang W, Li X, Chen K (2020): Analysis of immune-related signatures of lung adenocarcinoma identifies two distinct immunophenotypic subtypes: Implications for immune checkpoint blockade therapy. *Aging* **12**, 3312-3339  
<https://doi.org/10.18632/aging.102814>
- Xie J, Chen L, Sun Q, Li H, Wei W, Wu D, Hu Y, Zhu Z, Shi J, Wang M (2022): An immune subtype-related prognostic signature of hepatocellular carcinoma based on single-cell sequencing analysis. *Aging (Albany NY)* **14**, 3276-3292  
<https://doi.org/10.18632/aging.204012>
- Xie J, Li H, Chen L, Cao Y, Hu Y, Zhu Z, Wang M, Shi J (2021): A novel pyroptosis-related lncRNA signature for predicting the prognosis of skin cutaneous melanoma. *Int. J. Gen. Med.* **14**, 6517-6527  
<https://doi.org/10.2147/IJGM.S335396>
- Yochum ZA, Cades J, Mazzacurati L, Neumann NM, Khetarpal SK, Chatterjee S, Wang H, Attar MA, Huang EHB, Chatley SN, et al. (2017): A first-in-class TWIST1 inhibitor with activity in oncogene-driven lung cancer. *Mol. Cancer Res.* **15**, 1764-1776  
<https://doi.org/10.1158/1541-7786.MCR-17-0298>
- Yu Y, Wang Z, Zheng Q, Li J (2021): GALNT2/14 overexpression correlate with prognosis and methylation: a potential therapeutic target for lung adenocarcinoma. *Gene* **790**, 145689  
<https://doi.org/10.1016/j.gene.2021.145689>
- Zhang C, Li Z, Qi F, Hu X, Luo J (2019): Exploration of the relationships between tumor mutation burden with immune infiltrates in clear cell renal cell carcinoma. *Ann. Transl. Med.* **7**, 648-648  
<https://doi.org/10.21037/atm.2019.10.84>
- Zhang S, Tong YX, Zhang XH, Zhang YJ, Xu YS, Xiao AT, Chao TF, Gong JP (2020): A novel and validated nomogram to predict overall survival for gastric neuroendocrine neoplasms. *J. Cancer* **10**, 5944-5954  
<https://doi.org/10.7150/jca.35785>

Received: January 27, 2022

Final version accepted: May 30, 2022