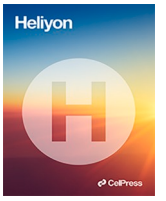
Heliyon 10 (2024) e29383

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Identification and validation of a glycosyltransferase gene signature as a novel prognostic model for lung adenocarcinoma

Jiejun Zhou [1], Kun Zhang [1], Tian Yang, Anqi Li, Meng Li, Xiaojing Peng, Mingwei Chen [*]

Department of Respiratory and Critical Care Medicine, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, 710000, China

A R T I C L E I N F O

Keywords:
Lung adenocarcinoma
Glycosyltransferase
Bioinformatics
Biomarker

A B S T R A C T

Background: The role of glycosyltransferase (GT) genes in lung adenocarcinoma (LUAD) needs further elucidation. Thus, our study aims to identify the prognostic gene signature of LUAD and explore its molecular functions.

Methods: We initially extracted GT gene sets from the database, and obtained mRNA expression levels and clinical data from The Cancer Genome Atlas (TCGA) database. For constructing a prognostic model for GT genes, we utilized univariate, least absolute shrinkage and selection operator (LASSO), and multivariate Cox regression analyses. Using the model, patients were categorized into high- and low-risk groups. Additionally, we evaluated differences in tumor immune infiltration between these groups and identified potential therapeutic drugs. Finally, we experimentally validated the expression levels of these crucial prognostic genes.

Results: We developed a risk score comprising nine GT genes (C1GALT1, FUT1, GALNT2, PLOD2, POMK, PYGB, ST3GAL6, UGT2B11, UGT3A1). Patients were then categorized into low- and high-risk groups based on this score. The low-risk group showed superior overall survival (OS) compared to the high-risk group. There were significantly distinct tumor immune microenvironment statuses observed between the two groups. We identified potential therapeutic drugs, including the MEK inhibitor (PD-184352). Finally, we verified the expression of these nine GT genes through immunohistochemistry (IHC) staining and quantitative real-time PCR (qPCR).

Conclusion: We identified a distinct LUAD GT gene signature, and these differentially expressed mRNAs could serve as valuable prognostic biomarkers and therapeutic targets. Furthermore, we experimentally validated their expression levels and identified potential therapeutic agents.

Abbreviations: AUC, area under the curve; BP, biological processes; CC, cell component; cMAP, the connectivity map database; DEGs, differentially expressed genes; FDR, false discovery rate; GEO, Gene Expression Omnibus; GGDB, the GlycoGene database; GO, gene ontology; GSEA, the Gene Set Enrichment Analysis database; GT, glycosyltransferase; HPA, the human protein map; IHC, immunohistochemistry; KEGG, Kyoto encyclopedia of genes and genomes; LUAD, lung adenocarcinoma; MF, molecular function; OS, overall survival; PCA, principal component analysis; qPCR, quantitative real-time PCR; ROC, receiver operating characteristic; TCGA, cancer genome atlas; TME, tumor microenvironment; T-SNE, T-distributed stochastic neighbor embedding.

* Corresponding author. Department of Respiratory and Critical Care Medicine, the First Affiliated Hospital of Xi'an Jiaotong University, 277#, Yanta West Road, Xi'an, 710061, Shaanxi Province, China.

E-mail address: chenmw36@163.com (M. Chen).

[1] These authors have contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e29383
Received 12 January 2024; Received in revised form 7 April 2024; Accepted 7 April 2024
Available online 9 April 2024

1. Introduction

Lung cancer continues to be the primary cause of cancer-related deaths globally. Adenocarcinoma (LUAD) is the most prevalent type of lung cancer [1]. Despite the development of various treatments for LUAD, the prognosis remains poor, with a 5-year overall survival (OS) rate for LUAD patients below 20% [2]. Thus, investigating predictive variables and exploring novel treatment targets is crucial.

Glycosylation is a critical post-translational modification that occurs in all eukaryotes. Glycobiology regulates protein location, function, and activity in tissues and cells, influencing essential life processes like cell identification, differentiation, signal transmission, and immune response [3,4]. This process primarily involves various glycosylated enzymes' activity, including glycosyltransferase (GT). Glycation, recognized as a significant mechanism contributing to tumor heterogeneity, is widely acknowledged as a cancer marker [5]. Targeted glycosylation has emerged as a potential therapeutic approach. However, the specific functionality of glycosylation in LUAD remains understudied, necessitating the screening of differentially expressed genes (DEGs) associated with GT. Furthermore, analyzing the impact of DEGs on prognosis is crucial for identifying therapeutic targets and enhancing treatment outcomes.

In our prior studies, we identified various glycosylation types in serum and bronchoalveolar lavage samples from lung cancer patients [6,7]. Other studies have reported a significant increase in genes related to mucin O-glycosylation in LUAD, indicating a potential role of abnormal glycosylation in lung cancer initiation and progression [8]. We hypothesize that relevant isoforms can be identified based on abnormal glycosylation patterns. This study aimed to evaluate if LUAD could be categorized based on GT gene expression levels. We developed a prognostic model for GT genes to predict overall survival (OS) in LUAD patients. Furthermore, we explored the immune microenvironment and identified potential drug candidates. Overall, this study enhances our understanding of LUAD prognosis and reveals potential therapeutic targets.

2. Methods

2.1. Data acquisition

The Cancer Genome Atlas database (TCGA, https://portal.gdc.cancer.gov/repository) was utilized to obtain data for LUAD patients (normal = 59 and tumor = 539). Subsequently, the "TCGAbiolinks" R package was used to download RNA-seq gene expression datasets, which were displayed in "Counts" format. Additionally, data related to clinical pathology and survival were retrieved. The obtained RNA-seq data were preprocessed and analyzed for differences using the "DESeq2″ R package and then normalized using Variance Stabilizing Transformation for further analysis. Genes in the TCGA dataset were annotated using the gene annotation file from the GENCODE (https://www.gencodegenes.org/human/) website and Ensemble IDs (Supplementary Table S1). GT genes (Supplementary Table S2) were obtained from the GlycoGene database (GGDB, http://acgg.asia/db/ggdb) and the Gene Set Enrichment Analysis database (GSEA, http://www.gsea-msigdb.org/gsea/msigdb/search.jsp). A total of 223 GT genes were identified from the GGDB. Additionally, 276 genes related to the glycosylation pathway and associated with GT were enriched via GSEA. Finally, by intersecting two datasets, we obtained 348 genes related to GT for further analysis.

2.2. GT gene risk score development and validation

We conducted RNAseq differential expression analysis in TCGA using LUAD tissue and normal lung tissue, $|\log_2(\text{fold change})| > 0.585$ or < -0.585, and false discovery rate <0.05 were the filtering requirements. The differentially expressed genes were analyzed using univariate Cox regression to assess their correlation with survival in LUAD patients. A P-value <0.1 was considered indicative of prognosis in LUAD. Subsequently, we employed the identified prognosis-related genes to perform LASSO and multivariate Cox regression analyses, calculating a risk score using the formula: risk score $= \sum_i^9 Xi * Yi$ (X: gene expression level, Y: coefficient). Patients were categorized as high- or low-risk based on the median risk score. Also, We used the R programming language's "prcomp" function to do a principal component analysis. Survival analysis was conducted using the R packages "survminer" and "survival" to conduct survival analysis to assess differences in OS between the high- and low-risk groups. Subsequently, we explored the relationship between each gene included in the risk score and specific clinical factors, such as age. Furthermore, we evaluated the model's ability to independently predict OS while considering other clinical variables using multivariate and univariate COX regression analysis. Finally, validation was performed using the Gene Expression Omnibus (GEO) database.

2.3. Functional enrichment analysis

To investigate the underlying biological processes of the GT gene model, we utilized the R package "clusterProfiler" to undertake Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) investigations. GSEA was conducted using WebGestalt [9]. A false discovery rate (FDR) of 0.05 was utilized to determine statistical significance.

2.4. Immune cell infiltration analysis

The R package "CIBERSORT" was used to quantify the percentage of 22 immune cell types in lung adenocarcinoma tissues from

patients in the high- and low-risk groups. Significant results ($p < 0.05$) were selected for follow-up analysis.

2.5. Identify potential small molecule drugs

We utilize the Connectivity Map database (cMAP) to predict potential drugs for LUAD, accessible at https://clue.io. The top 150 upregulated and 150 downregulated DEGs were submitted to the cMAP database to identify small molecular compounds that could be potentially used for LUAD treatment. Scores ranged from −100 to 100, with a negative score indicating the potential benefit of the drug for LUAD treatment.

2.6. Verification of gene expression

Human lung epithelial cell lines (BEAS-2B) and four LUAD cell lines (A549, PC-9, H1299, H1650) were cultured in Dulbecco's Modified Eagle Medium or Roswell Park Memorial Institute 1640 medium containing 10% fetal bovine serum, following standard cell culture protocols. The total RNA was extracted by using RNA Plus (Takara, Otsu, Shiga, Japan) following the manufacturer's protocol. Then, 1 μg of total RNA was reverse-transcribed using PrimeScript™ RT Master Mix (TaKaRa, Japan), and quantitative real-time PCR (qPCR) was performed using a ViiA 7 Real-Time PCR System (Applied Biosystems, USA). SYBR Green-based three-step RT–qPCR was performed using TB Green® Premix Ex Taq™ II (TaKaRa, Japan). Information on the primers is summarized in Supplementary Table S3. The human Protein map (HPA, https://www.proteinatlas.org/) was utilized to compare the protein expression levels of gene signatures between normal and tumor tissues.

2.7. Statistical analyses

Data analysis was performed using R (Version 4.1.2). Unless otherwise specified, a p-value <0.05 was considered statistically significant. The "Deseq2" package was used to identify DEGs between tumor and normal samples. The "survival" package was used to evaluate the association of each gene with survival. The predictive accuracy of the risk model for survival was evaluated using time-dependent ROC curve analysis, and the survival rate was calculated the using Kaplan-Meier method. The log-rank test was conducted

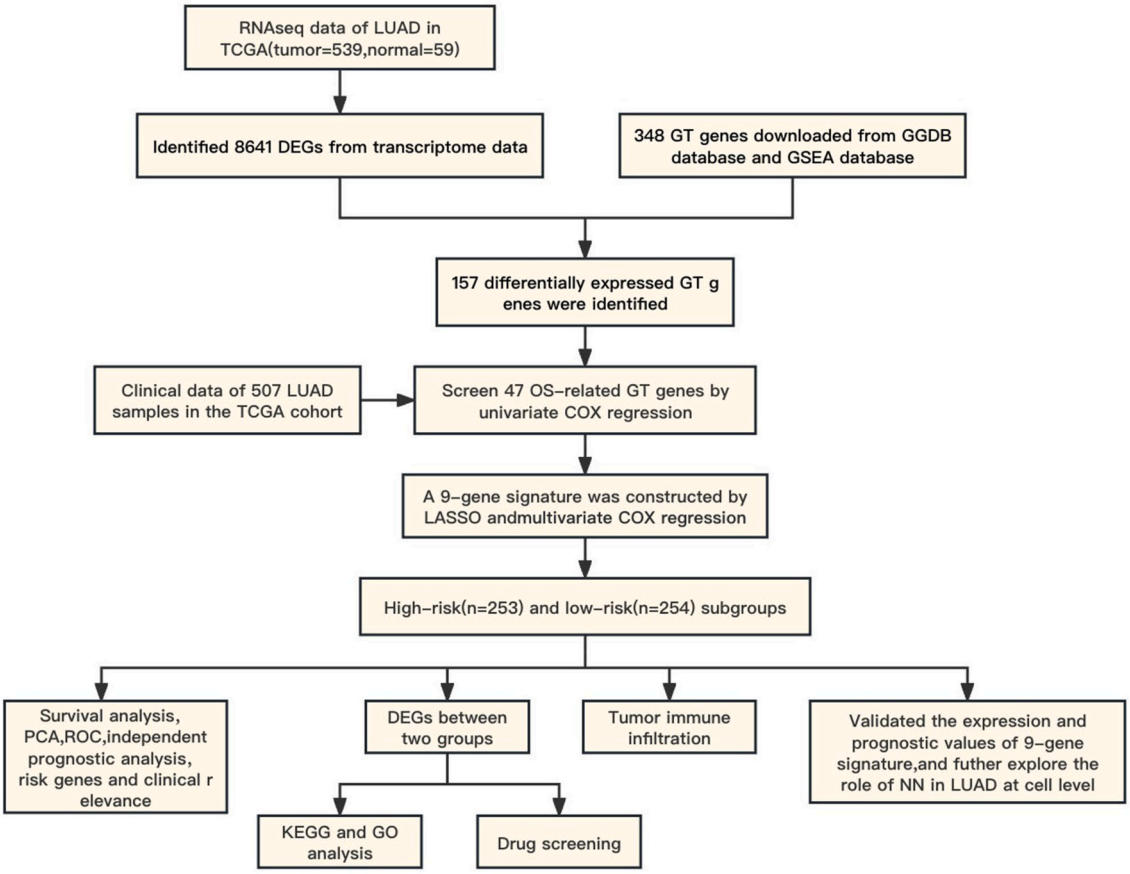


Fig. 1. Flowchart.

to assess the significance of differences in survival curves. The two-tailed Student's t-test was used to determine the statistical significance.

3. Results

3.1. Development of a risk score

Fig. 1 illustrates the basic procedure of this study. We initially conducted a differential expression analysis of RNAseq data from tumor and normal tissues in the TCGA cohort, identifying a total of 8641 DEGs, including 23 up-regulated genes and 50 down-regulated genes (Fig. 2A). We then retrieved 348 GT genes from the GSEA database and GGDB database. The intersection of these 348 GT genes with the previously identified DEGs resulted in a total of 157 GT DEGs (Fig. 2B). The heat map in Fig. 2C depicts the expression levels of the TOP30 GT DEGs in tumor and normal tissues.

Univariate Cox regression analysis of the 157 GT DEGs showed that 48 of them significantly correlated with overall survival (OS) (Fig. 2D). Lasso regression analysis, a common method for multiple regression, was used to prevent overfitting and effectively identify predictors. As a result, 22 GT DEGs were selected through LASSO regression analysis for further analysis (Fig. 2E and F). Following this, multivariate Cox regression analysis identified 9 GT DEGs associated with OS, including C1GALT1, FUT1, GALNT2, PLOD2, POMK, PYGB, ST3GAL6, UGT2B11, and UGT3A1 (Fig. 2G). These 9 GT DEGs were used to construct a prognostic model for evaluating prognosis in LUAD patients. The risk score is calculated as the sum of the coefficient of each risk gene multiplied by its gene expression level.

3.2. Evaluation of the risk score

LUAD patients were divided into low- and high-risk using the median risk score. The Principal Component Analysis (PCA) plot showed a clear distinction between the two groups (Fig. 3A). Fig. 3B and C describe the risk grade distribution, survival state, and survival time of two different risk groups, respectively. In survival analysis, OS was considerably better in the low-risk group compared to the high-risk group ($P < 0.001$) (Fig. 3D). Time-dependent receiver operating characteristic (ROC) curves for 1, 3, and 5-year survival were constructed to assess the prediction accuracy of the prognostic signature, and the area under the curve (AUC) values were determined using the survival ROC software. Results show that the 1, 3, and 5-year survival ROC curves showed an area under the ROC curve of 0.740,0.740, and 0.658, respectively. (Fig. 3E). Univariate and multivariate Cox regression analyses were conducted on the entire TCGA dataset to assess the independence of nine GT gene models. Univariate Cox regression analysis revealed that the HR

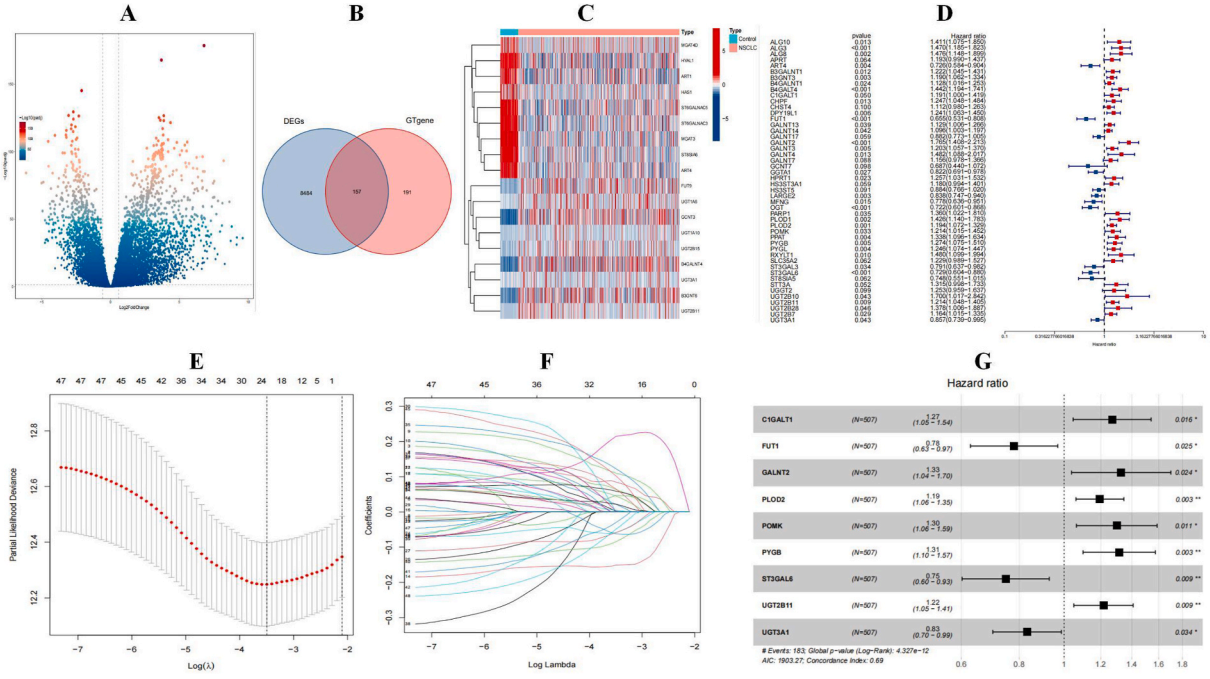


Fig. 2. Construction of the GT gene risk score to evaluate the prognosis of LUAD. The volcano diagram of the differential expression analysis of RNAseq from tumor and normal tissues in the TCGA cohort (Fig. 2A). The Venn diagram of the intersection of 348 GT genes with the above DEGs (Fig. 2B). The heat map of expression levels of TOP30 GT DEGs in tumor and normal tissues(Fig. 2C). The univariate Cox regression analysis of GT DEGs for OS of LUAD(Fig. 2D). GT DEGs screened by LASSO regression analysis(Fig. 2E and F). Multivariate Cox regression analysis identified 9 GT DEGs associated with OS(Fig. 2G).

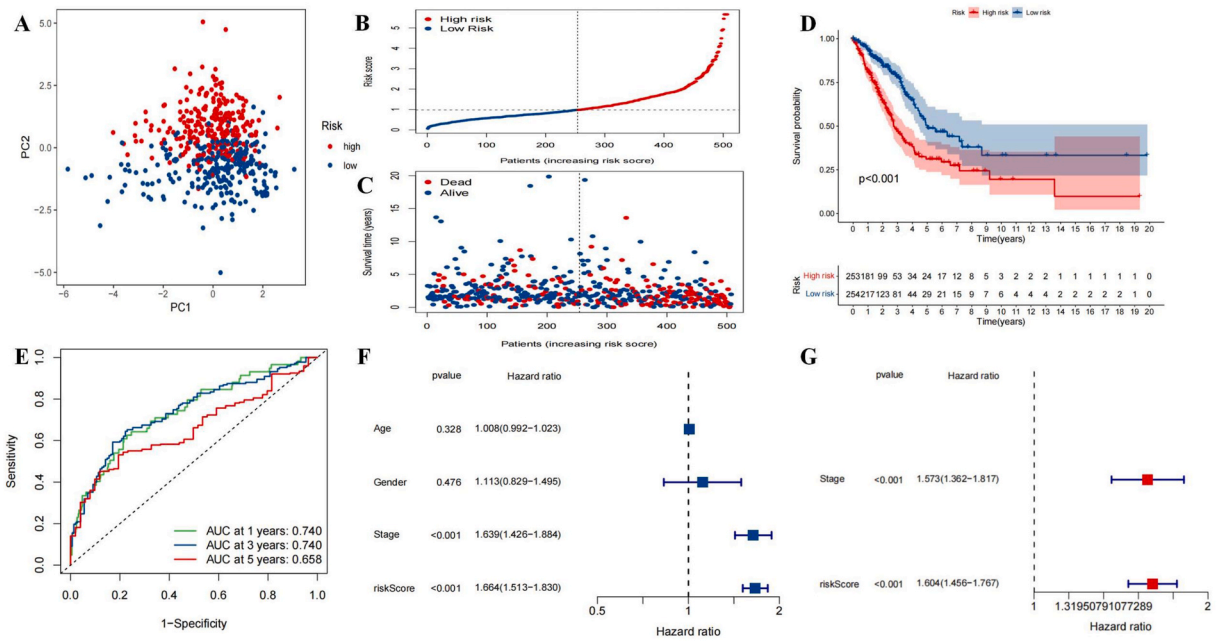


Fig. 3. Evaluation of the efficacy of the risk score. LUAD patients are split into low- and high-risk groups by the median risk score. The PCA plot of the distinction between the two groups (Fig. 3A). The two groups' risk grade distribution, survival state, and survival time (Fig. 3B and. C). Survival analysis of OS in the two groups (Fig. 3D). The ROC curve and AUC (Fig. 3E) examined the risk score's prediction accuracy. Univariate Cox regression analysis of the risk score and other clinical indicators (Fig. 3F). Multivariate Cox regression analysis for found independent prognostic variables (Fig. 3G).

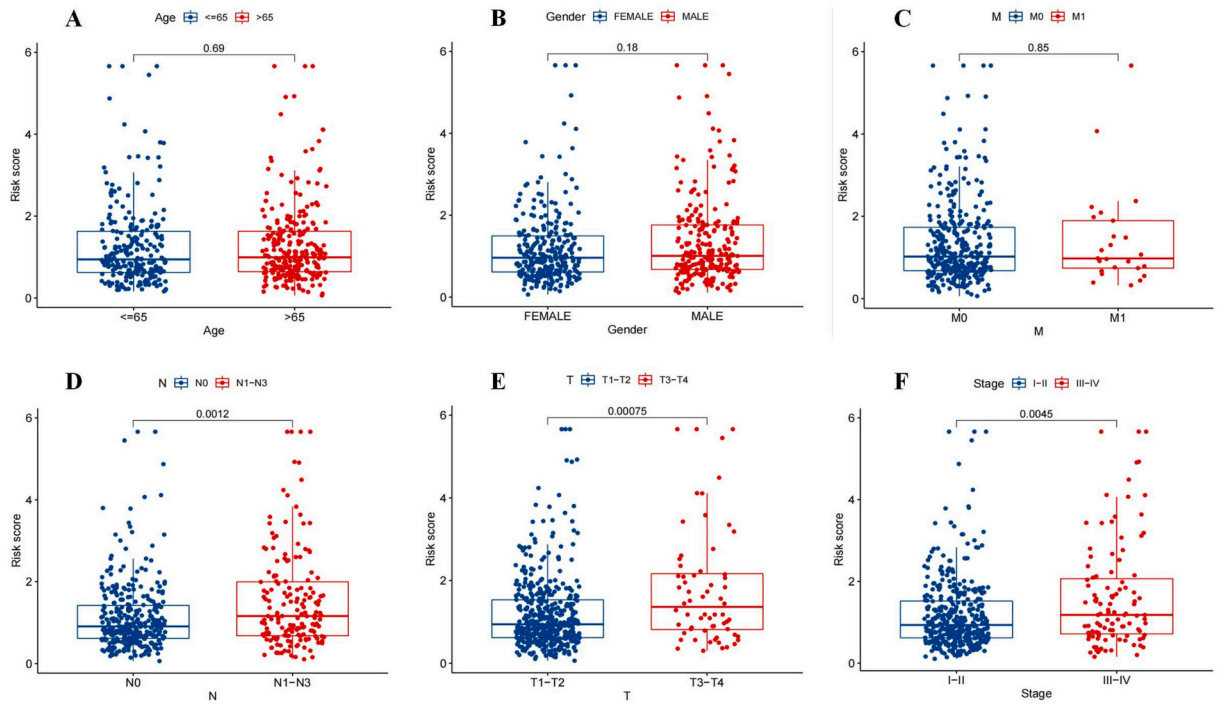


Fig. 4. The link between the risk score and each clinicopathological condition. The link between the risk score and age (Fig. 4A). The link between the risk score and gender (Fig. 4B). The link between the risk score and metastasis (Fig. 4C). The link between the risk score and node (Fig. 4D). The link between the risk score and tumor (Fig. 4E). The link between the risk score and clinical stage (Fig. 4F).

and 95% CI of risk scores were 1.664 and 1.513–1.830, respectively ($P < 0.001$), indicating that the HR of risk score was greater than age, gender, or stage (Fig. 3F). Fig. 3G shows that independent prognostic variables in multivariate Cox regression analysis included stage and risk score, with HR of 1.604 and 95%CI of 1.456–1.767 ($P < 0.001$), the HR of risk score was greater than the stage. As mentioned, our model is both dependable and accurate. Finally, we analyzed the validation cohort in the GEO database. The results, shown in Fig. 1, roughly align with the TCGA dataset analysis. However, the incomplete gene collection in the GEO datasets limits the adequacy of the validation cohort.

3.3. Clinical features and the risk score

Additionally, our study investigated the relationship between the risk score and various clinicopathological conditions. The results showed that the risk score does not correlate with age, gender, and distant metastases (M) (Fig. 4A and. B, and Fig. 4C, $P > 0.05$). In contrast, a higher risk score is significantly associated with primary tumor size (T), lymph node invasion (N), and clinical stage (Fig. 4D–. E, and Fig. 4F, $P < 0.05$).

3.4. Functional enrichment analysis

To investigate the differences between the high- and low-risk groups. We conducted differential expression analysis between the two groups and identified a total of DEGs. We then performed a functional enrichment analysis of these DEGs. GO analysis shows that the top three biological processes (BP) terms are passive trans membrane transporter activity, channel activity, and signaling receptor activator activity. The top three terms of cell component (CC) are collagen-containing extracellular matrix, apical plasma membrane,

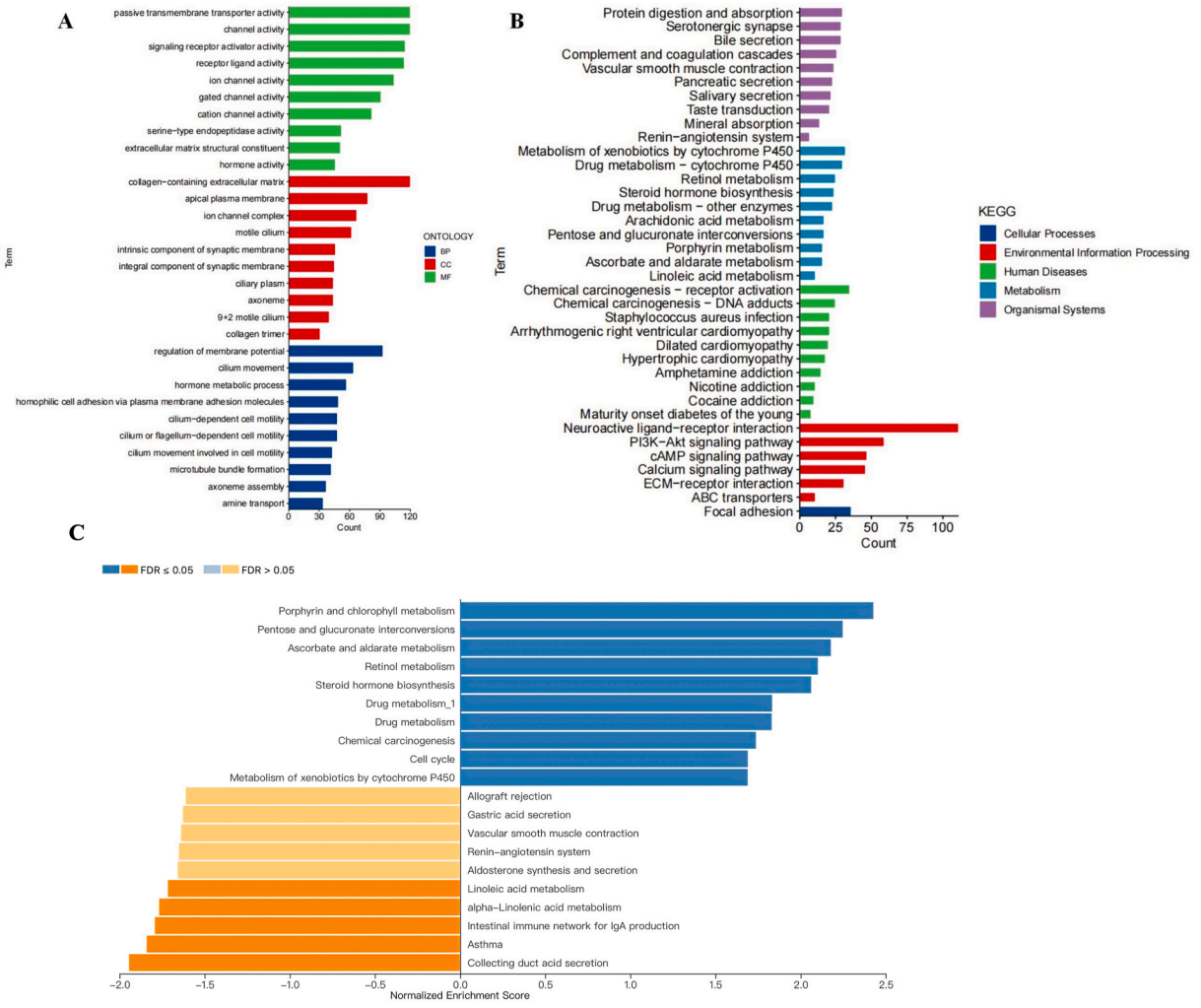


Fig. 5. Functional enrichment analysis of DEGs in low- and high-risk groups. GO analysis (Fig. 5A). KEGG analysis (Fig. 5B). GSEA analysis (Fig. 5C).

and ion channel complex. The top three terms for molecular function (MF) are regulation of membrane potential, cilium movement, and hormone metabolic process. Fig. 5A shows the top 10 GO-BP, GO-CC, and GO-MF terms. KEGG analysis showed that PI3K-Akt, cAMP, and calcium signaling pathways were three significant enrichment pathways. The KEGG pathway is shown in Fig. 5B. GSEA analysis revealed that the genes involved in pentose and glucuronate interconversions, the cell cycle, and other pathways were significantly up-regulated. Conversely, the genes associated with the intestinal immune network for IgA production, asthma, and other signaling pathways were significantly down-regulated (Fig. 5C).

3.5. Immune cell infiltration analysis

The immune status of the tumor microenvironment (TME) is closely associated with the tumor. As shown in Fig. 6A, the abundance ratio of 22 immune cells in LUAD samples from high- and low-risk groups was analyzed by CIBESORT. The results the of immune cell correlation analysis showed that T cell CD4 memory activated was significantly related to T cells CD8, while NK cells resting was adversely associated with both NK cells activated and Mast cells resting (Fig. 6B). Additionally, the percentage of 22 different immune cell types varied between the two risk groups (Fig. 6C). The low-risk group had higher levels of infiltration of B cell native, monocytes, and dendritic cells resting ($p < 0.05$). B cells memory, T cells CD4 memory activated, and Macrophages M0 were lower in the low-risk group compared to the high-risk group ($p < 0.05$).

4. Identify potential small molecule drugs

The top 150 up- and down-regulated DEGs between the two groups were uploaded to the cMAP database for analysis. Table 1 shows the top ten identified small molecule drugs with anticancer properties against LUAD progression.

4.1. The validation by intro experiments

The qPCR results showed differential expression of the 9 GT genes in our model between the normal bronchial epithelial cell line BEAS-2B and LUAD cell lines (A549, PC-9, H1650, H1299) (Fig. 7). Subsequently, we utilized the HPA database to assess the protein levels of these 9 genes in clinical LUAD and normal tissue, and the findings were largely consistent with the mRNA study (Fig. 8).

5. Discussion

Glycation plays a role in fundamental biological processes associated with cancer and is widely accepted as a cancer marker due to its significant contribution to tumor heterogeneity [5]. Our previous research identified diverse changes in serum glycosylation types in lung cancer patients [6]. Additionally, we assessed protein glycosylation types in 281 bronchoalveolar lavage fluid samples,

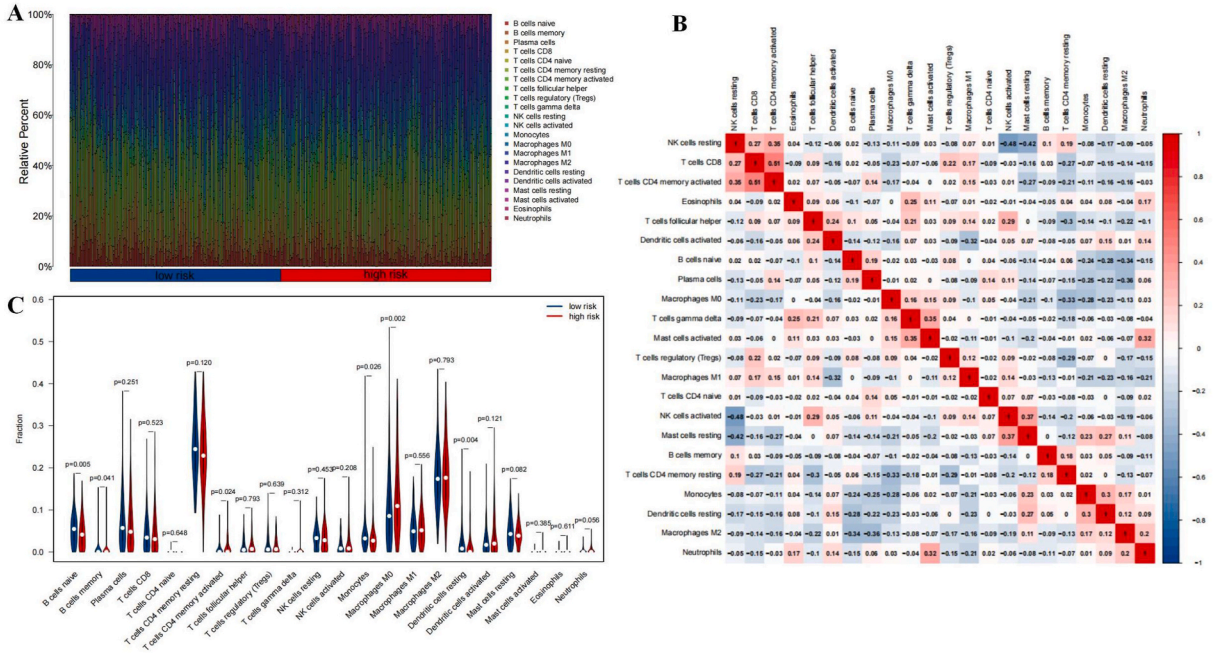


Fig. 6. The difference of tumor immune infiltration in high- and low-risk groups. The abundance ratio of 22 immune cells in LUAD (Fig. 6A). The immune cell correlation analysis (Fig. 6B). The percentage of 22 different immune cell types varied between the two risk groups (Fig. 6C).

Table 1
The top ten identified small-molecule drug.

Rank	Score	ID	Name	Description
1	−97	BRD-K05104363	PD-184352	MEK inhibitor
2	−96.79	BRD-K41859756	NVP-AUY922	HSP inhibitor
3	−96.26	BRD-K67578145	GDC-0879	RAF inhibitor
4	−95.91	BRD-K92093830	doxorubicin	Topoisomerase inhibitor
5	−95.12	BRD-K46056750	AZD-7762	CHK inhibitor
6	−93.17	BRD-K49865102	PD-0325901	MEK inhibitor
7	−92.98	BRD-K99545815	PF-562271	Focal adhesion kinase inhibitor
8	−92.49	BRD-K12502280	TG-101348	FLT3 inhibitor
9	−91.37	BRD-K94176593	TWS-119	Glycogen synthase kinase inhibitor
10	−90.87	BRD-K51575138	TPCA-1	IKK inhibitor

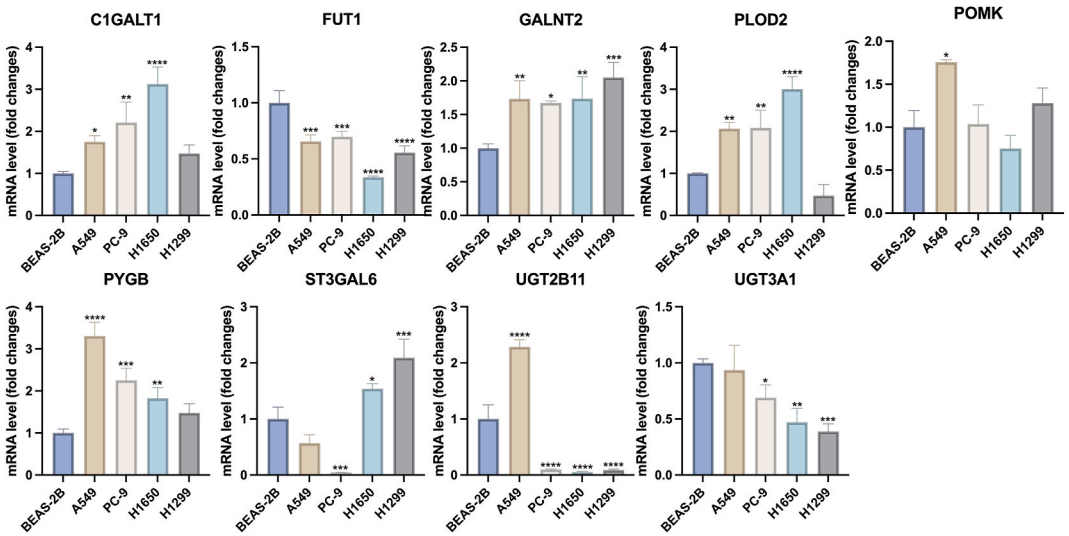


Fig. 7. The qPCR verified the expression levels of 9 GT genes in our model by normal bronchial epithelial cell line and LUAD cell lines mRNA.

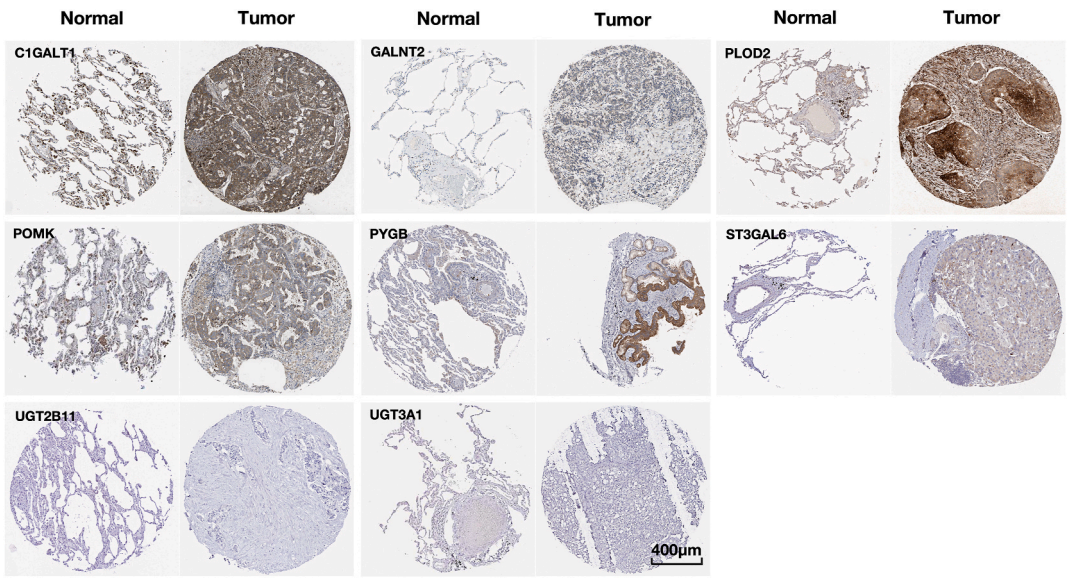


Fig. 8. The representative images of protein levels of GT genes in our model by clinical LUAD and normal tissues detected by IHC staining in the HPA database.

revealing their utility in identifying lung cancer biomarkers [7]. Erika Lattova observed substantial differences in protein glycosylation among different histological stages of LUAD, suggesting a link between glycosylation and LUAD onset and progression [10]. Other studies have indicated a notable increase in mucin O-glycosylation genes in LUAD, implying a potential contribution of aberrant glycosylation to lung cancer biogenesis and development [8]. Hence, we propose the feasibility of identifying lung cancer subtypes based on differences in glycosylated gene expression patterns. However, the relevance of GT genes in LUAD remains unknown. Thus, we conducted a screening of prognosis-related GT genes, crucial for identifying therapeutic targets and enhancing LUAD prognosis. In this study, we established a risk score comprising 9 GT genes (C1GALT1, FUT1, GALNT2, PLOD2, POMK, PYGB, ST3GAL6, UGT2B11, UGT3A1). Among them, C1GALT1, GALNT2, PLOD2, POMK, PYGB, and UGT2B11 were identified as risk factors for LUAD, while FUT1, ST3GAL6, and UGT3A1 were identified as protective factors.

Some of the GT genes identified in this study have previously been shown to have important functions in malignancies, including LUAD. Prior research has linked elevated FUT1 expression to a favorable EGFR-TKI response [11]. In our study, LUAD patients with high FUT1 expression exhibited a better prognosis. Studies have validated that C1GALT1 [12], GALNT2 [13], PYGB [14,15], and PLOD2 [16,17] are associated with a poorer prognosis in LUAD, aligning with our conclusions. Presently, there are limited studies on POMK, with only one speculating on its potential cancer-promoting effect [18], and no relevant study in LUAD. However, POMK expression levels were significantly higher in the LUAD cell line A549 compared to the normal lung epithelial cell line BEAS-2B. Immunohistochemical staining also indicated high PMOK expression in lung cancer tissue, suggesting a potential role in the onset and progression of LUAD. Regarding ST3GAL6, only one study has explored its correlation with LUAD, reporting decreased ST3GAL6 levels in LUAD and its inability to promote LUAD cell aggressiveness through activated EGFR/MAPK signaling [19]. In contrast, our study identifies ST3GAL6 as a risk factor for LUAD development, necessitating further exploration of the relationship between ST3GAL6 and LUAD. Although no studies have investigated UGT2B11 in lung cancer, it has been linked to the prognosis of breast [20] and prostate cancer [21]. Our study reveals increased UGT2B11 expression in LUAD, associating it with a higher risk of lung cancer. A study suggested that rs10045685 in UGT3A1 was associated with erlotinib adverse drug reactions ($P = 0.015$) [22], and our results indicate UGT3A1 as a protective factor in LUAD. These findings collectively support the interpretability of our results.

We examined genes exhibiting differential expression between the high- and low-risk groups and identified the top three terms for molecular function as regulation of membrane potential, cilium movement, and hormone metabolic process. KEGG analysis revealed three significantly enriched pathways: PI3K-Akt, cAMP, and calcium signaling pathways. Notably, PI3K-Akt signaling pathways are closely associated with lung cancer progression, while cAMP and calcium signaling pathways are fundamental pathways in tumorigenesis [23,24]. The appeal results further proved that GTgenes were associated with lung adenocarcinoma development.

Utilizing the model formed by the 9 GT genes, we stratified patients into low- and high-risk groups, observing substantial differences in their prognoses. Multivariate Cox regression analysis showed that the prognostic characteristics of these 9 GT gene components were independent factors for OS ($HR = 1.604$, $P < 0.001$). The predictive capability of this signature was validated through ROC curve analysis. Our study introduces novel perspectives to the discourse on patient prognosis and stratification.

Our study explored the relationship between risk score models and immune cells. We found that among the 22 types of immune cells, resting CD4 memory T cells, and M2 macrophages were the most abundant, consistent with immune cell infiltration in tumor tissues. The expression levels of immune cells differed between the two groups, indicating that the imbalanced proportion of immune cell components was linked to poor prognosis in cancer patients. Memory B cells activated CD4 memory T cells, and Macrophages M0 in the high-risk group were higher than in the low-risk group. We propose that the expression of GT genes may stimulate the active differentiation of M0 macrophages into M1 or M2 macrophages, opposing tumor cell activity and enhancing survival outcomes.

We then explored potential therapeutic agents and the top 5 compounds for effectiveness were MEK inhibitor (PD-184352), RAF inhibitor (GDC-0879), HSP inhibitor (NVP-AUY922), Topoisomerase inhibitor (doxorubicin), and CHK inhibitor (AZD-7762). MEK inhibitor has been shown to reduce lung cancer resistance [25]. RAF inhibitor was found to improve the prognosis of lung cancer with RAS or BRAF mutations [26,27]. Previous studies have confirmed that Topoisomerase inhibitor combined with the EGFR inhibitor has improved the anti-tumor effect [28]. NVP-AUY922, as a member of HSP inhibitor, has been previously confirmed to enhance the anti-cancer effect of BCL-2 inhibitor [29]. And AZD-7762, which is a CHK inhibitor, has previously been shown to promote cancer cell death [30]. These studies suggest that the drugs we have identified may have therapeutic effects in patients classified according to GT genes. At the same time, it also reflects that it is feasible to classify LUAD patients according to GT genes.

Through this study, we identified a distinct LUAD GT gene signature, indicating the potential importance of these differentially expressed mRNAs as prognostic biomarkers and offering valuable insights for future investigations. Additionally, we validated the molecular expression levels of GT genes using IHC and qPCR, which largely corroborated our previous data analysis findings. The tumor cell lines utilized in our study comprised four distinct lung adenocarcinoma cell lines. Despite variations in genotype and drug sensitivity among the four cell lines, the molecular expression trends of GT genes were largely consistent. This consistency substantiates the universality of the gene signature we have identified, which exhibits stable correlations with the occurrence and progression of lung cancer. This underscores the value of our gene signature. Therefore, we will further study the GT gene signature genes in our study, to provide new targets for the treatment of lung cancer. However, this research has some limitations: Firstly, the primary information in this survey was sourced from publicly accessible databases, limiting coverage of variations among patients from diverse geographic locations. Secondly, our analysis revealed a preliminary association between GT genes and LUAD prognosis. While we validated expression levels using qPCR and IHC, further experimental studies are necessary to provide a more comprehensive understanding of the biological mechanism.

6. Conclusion

We discovered a new gene signature for LUAD involving glycosyltransferases. These differentially expressed genes could serve as valuable prognostic biomarkers and therapeutic targets. Additionally, we experimentally validated their expression levels and identified potential therapeutic agents.

Funding

This work was co-supported by the Key R&D Project of Shaanxi Province (grant number 2021ZDLSF02-05), the Science and Technology Project of Shaanxi Province (grant number 2022SF-580), and Wu Jieping Medical Foundation (grant number HX202293).

Ethics approval and consent to participate

Not applicable.

Data availability statement

Data included in article.

Additional information

No additional information is available for this paper.

CRediT authorship contribution statement

Jiejun Zhou: Writing – review & editing, Writing – original draft, Formal analysis, Data curation, Conceptualization. **Kun Zhang:** Writing – review & editing, Writing – original draft, Formal analysis, Data curation, Conceptualization. **Tian Yang:** Writing – review & editing, Funding acquisition. **Anqi Li:** Writing – review & editing, Formal analysis. **Meng Li:** Writing – review & editing, Data curation. **Xiaojing Peng:** Writing – review & editing, Formal analysis. **Mingwei Chen:** Writing – review & editing, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors express their gratitude for the utilization of Microsoft Word, R (Version 4.1.2), and Endnote software tools in this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e29383.

References

[1] H. Sung, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA A Cancer J. Clin. 71 (3) (May 2021) 209–249, https://doi.org/10.3322/caac.21660 (in eng).
[2] J. Rotow, T.G. Bivona, Understanding and targeting resistance mechanisms in NSCLC, Nat. Rev. Cancer 17 (11) (Oct 25 2017) 637–658, https://doi.org/10.1038/nrc.2017.84 (in eng).
[3] K. Ohtsubo, J.D. Marth, Glycosylation in cellular mechanisms of health and disease, Cell 126 (5) (Sep 8 2006) 855–867, https://doi.org/10.1016/j.cell.2006.08.019 (in eng).
[4] K.T. Schjoldager, Y. Narimatsu, H.J. Joshi, H. Clausen, Global view of human protein glycosylation pathways and functions, Nat. Rev. Mol. Cell Biol. (Oct 21 2020), https://doi.org/10.1038/s41580-020-00294-x (in eng).
[5] J. Munkley, D.J. Elliott, Hallmarks of glycosylation in cancer, Oncotarget 7 (23) (Jun 7 2016) 35478–35489, https://doi.org/10.18632/oncotarget.8155 (in eng).
[6] Y. Liang, et al., Stage-associated differences in the serum N- and O-glycan profiles of patients with non-small cell lung cancer, Clin. Proteonomics 16 (2019) 20, https://doi.org/10.1186/s12014-019-9240-6 (in eng).
[7] L. Liu, et al., Protein glycopatterns in bronchoalveolar lavage fluid as novel potential biomarkers for diagnosis of lung cancer, Front. Oncol. 10 (2020) 568433, https://doi.org/10.3389/fonc.2020.568433 (in eng).
[8] M. Lucchetta, I. da Piedade, M. Mounir, M. Vabistsevits, T. Terkelsen, E. Papaleo, Distinct signatures of lung cancer types: aberrant mucin O-glycosylation and compromised immune response, BMC Cancer 19 (1) (Aug 20 2019) 824, https://doi.org/10.1186/s12885-019-5965-x (in eng).

[9] J. Wang, S. Vasaikar, Z. Shi, M. Greer, B. Zhang, WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit, Nucleic Acids Res. 45 (W1) (Jul 3 2017) W130–w137, https://doi.org/10.1093/nar/gkx356 (in eng).
[10] E. Lattová, et al., N-Glycan profiling of lung adenocarcinoma in patients at different stages of disease, Mod. Pathol. : an official journal of the United States and Canadian Academy of Pathology, Inc 33 (6) (Jun 2020) 1146–1156, https://doi.org/10.1038/s41379-019-0441-3 (in eng).
[11] J. Zhuge, et al., Construction of the model for predicting prognosis by key genes regulating EGFR-TKI resistance, Front. Genet. 13 (2022) 968376, https://doi.org/10.3389/fgene.2022.968376 (in eng).
[12] X. Dong, et al., C1GALT1, negatively regulated by miR-181d-5p, promotes tumor progression via upregulating RAC1 in lung adenocarcinoma, Front. Cell Dev. Biol. 9 (2021) 707970, https://doi.org/10.3389/fcell.2021.707970 (in eng).
[13] Y. Yu, Z. Wang, Q. Zheng, J. Li, GALNT2/14 overexpression correlate with prognosis and methylation: potential therapeutic targets for lung adenocarcinoma, Gene 790 (Jul 20 2021) 145689, https://doi.org/10.1016/j.gene.2021.145689 (in eng).
[14] L. Xiao, W. Wang, Q. Huangfu, H. Tao, J. Zhang, PYGB facilitates cell proliferation and invasiveness in non-small cell lung cancer by activating the Wnt-β-catenin signaling pathway, Biochem. Cell. Biol. 98 (5) (Oct 2020) 565–574, https://doi.org/10.1139/bcb-2019-0445 (in eng).
[15] K. Lei, et al., Development and clinical validation of a necroptosis-related gene signature for prediction of prognosis and tumor immunity in lung adenocarcinoma, Am. J. Cancer Res. 12 (11) (2022) 5160–5182 (in eng).
[16] Y. Meng, J. Sun, G. Zhang, T. Yu, H. Piao, Clinical prognostic value of the PLOD gene family in lung adenocarcinoma, Front. Mol. Biosci. 8 (2021) 770729, https://doi.org/10.3389/fmolb.2021.770729 (in eng).
[17] F. Kocher, et al., Deregulated glutamate to pro-collagen conversion is associated with adverse outcome in lung cancer and may be targeted by renin-angiotensin-aldosterone system (RAS) inhibition, Lung Cancer 159 (Sep 2021) 84–95, https://doi.org/10.1016/j.lungcan.2021.06.020 (in eng).
[18] C. Quereda, À. Pastor, J. Martín-Nieto, Involvement of abnormal dystroglycan expression and matriglycan levels in cancer pathogenesis, Cancer Cell Int. 22 (1) (Dec 9 2022) 395, https://doi.org/10.1186/s12935-022-02812-7 (in eng).
[19] J. Li, et al., Comprehensive landscape of the ST3GAL family reveals the significance of ST3GAL6-AS1/ST3GAL6 axis on EGFR signaling in lung adenocarcinoma cell invasion, Front. Cell Dev. Biol. 10 (2022) 931132, https://doi.org/10.3389/fcell.2022.931132 (in eng).
[20] J. Wang, et al., Overexpression of lipid metabolism genes and PBX1 in the contralateral breasts of women with estrogen receptor-negative breast cancer, Int. J. Cancer 140 (11) (Jun 1 2017) 2484–2497, https://doi.org/10.1002/ijc.30680 (in eng).
[21] Y. Liu, et al., Comprehensive signature analysis of drug metabolism differences in the White, Black and Asian prostate cancer patients, Aging 13 (12) (Jun 18 2021) 16316–16340, https://doi.org/10.18632/aging.203158 (in eng).
[22] C. Wang, et al., Genetic association of drug response to erlotinib in Chinese advanced non-small cell lung cancer patients, Front. Pharmacol. 9 (2018) 360, https://doi.org/10.3389/fphar.2018.00360 (in eng).
[23] C. Wei, et al., LPCAT1 promotes brain metastasis of lung adenocarcinoma by up-regulating PI3K/AKT/MYC pathway, J. Exp. Clin. Cancer Res. : CR 38 (1) (Feb 21 2019) 95, https://doi.org/10.1186/s13046-019-1092-4 (in eng).
[24] J. Liang, et al., Mex3a interacts with LAMA2 to promote lung adenocarcinoma metastasis via PI3K/AKT pathway, Cell Death Dis. 11 (8) (Aug 13 2020) 614, https://doi.org/10.1038/s41419-020-02858-3 (in eng).
[25] Q. Wang, S. Yang, K. Wang, S.Y. Sun, MET inhibitors for targeted therapy of EGFR TKI-resistant lung cancer, J. Hematol. Oncol. 12 (1) (Jun 21 2019) 63, https://doi.org/10.1186/s13045-019-0759-9 (in eng).
[26] R.A. Okimoto, et al., Preclinical efficacy of a RAF inhibitor that evades paradoxical MAPK pathway activation in protein kinase BRAF-mutant lung cancer, Proc Natl Acad Sci U S A 113 (47) (Nov 22 2016) 13456–13461, https://doi.org/10.1073/pnas.1610456113 (in eng).
[27] S.H. Chen, et al., RAF inhibitor LY3009120 sensitizes RAS or BRAF mutant cancer to CDK4/6 inhibition by abemaciclib via superior inhibition of phospho-RB and suppression of cyclin D1, Oncogene 37 (6) (Feb 8 2018) 821–832, https://doi.org/10.1038/onc.2017.384 (in eng).
[28] T. Shimoyama, et al., Effects of different combinations of gefitinib and irinotecan in lung cancer cell lines expressing wild or deletional EGFR, Lung Cancer 53 (1) (Jul 2006) 13–21, https://doi.org/10.1016/j.lungcan.2006.03.014 (in eng).
[29] H. Yang, et al., HSP90 inhibitor (NVP-AUY922) enhances the anti-cancer effect of BCL-2 inhibitor (ABT-737) in small cell lung cancer expressing BCL-2, Cancer letters 411 (Dec 28 2017) 19–26, https://doi.org/10.1016/j.canlet.2017.09.040 (in eng).
[30] T.H. Park, S.H. Bae, S.M. Bong, S.E. Ryu, H. Jang, B.I. Lee, Crystal structure of the kinase domain of MerTK in complex with AZD7762 provides clues for structure-based drug development, Int. J. Mol. Sci. 21 (21) (Oct 23 2020), https://doi.org/10.3390/ijms21217878 (in eng).