Identification of a six-gene metabolic signature predicting overall survival for patients with lung adenocarcinoma

Yubo Cao [Corresp., 1], **Xiaomei Lu** [2], **Yue Li** [1], **Jia Fu** [1], **Hongyuan Li** [1], **Xiulin Li** [1], **Ziyou Chang** [1], **Sa Liu** [Corresp. 1]

[1] Department of Medical Oncology, The Fourth Affiliated Hospital of China Medical University, Shenyang, China

[2] Department of Pathophysiology, China Medical University, Shenyang, China

Corresponding Authors: Yubo Cao, Sa Liu
Email address: ybcao@cmu.edu.cn, liusa118@163.com

Background: Lung cancer is the leading cause of cancer-related deaths worldwide. Lung adenocarcinoma (LUAD) is one of the main subtypes of lung cancer. Hundreds of metabolic genes are altered consistently in LUAD; however, their prognostic role remains to be explored. This study aimed to establish a molecular signature that can predict the prognosis in patients with LUAD based on metabolic gene expression. **Methods**: The transcriptome expression profiles and corresponding clinical information of LUAD were obtained from the Cancer Genome Atlas and Gene Expression Omnibus databases. The differentially expressed genes (DEGs) between LUAD and paired non-tumor samples were identified by the Wilcoxon rank sum test. Univariate Cox regression analysis and the lasso Cox regression model were used to construct the best-prognosis molecular signature. A nomogram was established comprising the prognostic model for predicting overall survival. To validate the prognostic ability of the molecular signature and the nomogram, the Kaplan-Meier survival analysis, Cox proportional hazards model, and receiver operating characteristic analysis were used. **Results**: The six-gene molecular signature (*PFKP, PKM, TPI1, LDHA, PTGES, and TYMS*) from the DEGs was constructed to predict the prognosis. The molecular signature demonstrated a robust independent prognostic ability in the training and validation sets. The nomogram including the prognostic model had a greater predictive accuracy than previous systems. Furthermore, a gene set enrichment analysis revealed several significantly enriched metabolic pathways, which suggests a correlation of the molecular signature with metabolic systems and may help explain the underlying mechanisms. **Conclusions**: Our study identified a novel six-gene metabolic signature for LUAD prognosis prediction. The molecular signature could reflect the dysregulated metabolic microenvironment, provide potential biomarkers for predicting prognosis, and indicate potential novel metabolic molecular-targeted therapies.

Identification of a six-gene metabolic signature predicting overall survival for patients with lung adenocarcinoma

Yubo Cao[1], Xiaomei Lu[2], Yue Li[1], Jia Fu[1], Hongyuan Li[1], Xiulin Li[1], Ziyou Chang[1], Sa Liu[1]

[1] Department of Medical Oncology, The Fourth Affiliated Hospital of China Medical University, Shenyang, Liaoning, China

[2] Department of Pathophysiology, China Medical University, Shenyang, Liaoning, China

Corresponding Authors:
Yubo Cao
No.4 Chongshan East Road, Huanggu District, Shenyang, Liaoning, Zip code: 110032, China
Email address: ybcao@cmu.edu.cn
Sa Liu
No.4 Chongshan East Road, Huanggu District, Shenyang, Liaoning, Zip code: 110032, China
Email address: liusa118@163.com

Abstract

Background: Lung cancer is the leading cause of cancer-related deaths worldwide. Lung adenocarcinoma (LUAD) is one of the main subtypes of lung cancer. Hundreds of metabolic genes are altered consistently in LUAD; however, their prognostic role remains to be explored. This study aimed to establish a molecular signature that can predict the prognosis in patients with LUAD based on metabolic gene expression.

Methods: The transcriptome expression profiles and corresponding clinical information of LUAD were obtained from the Cancer Genome Atlas and Gene Expression Omnibus databases. The differentially expressed genes (DEGs) between LUAD and paired non-tumor samples were identified by the Wilcoxon rank sum test. Univariate Cox regression analysis and the lasso Cox regression model were used to construct the best-prognosis molecular signature. A nomogram was established comprising the prognostic model for predicting overall survival. To validate the prognostic ability of the molecular signature and the nomogram, the Kaplan-Meier survival analysis, Cox proportional hazards model, and receiver operating characteristic analysis were used.

Results: The six-gene molecular signature (*PFKP, PKM, TPI1, LDHA, PTGES, and TYMS*) from the DEGs was constructed to predict the prognosis. The molecular signature demonstrated a robust independent prognostic ability in the training and validation sets. The nomogram including the prognostic model had a greater predictive accuracy than previous systems. Furthermore, a gene set enrichment analysis revealed several significantly enriched metabolic pathways, which suggests a correlation of the molecular signature with metabolic systems and may help explain the underlying mechanisms.

Conclusions: Our study identified a novel six-gene metabolic signature for LUAD prognosis prediction. The molecular signature could reflect the dysregulated metabolic microenvironment, provide potential biomarkers for predicting prognosis, and indicate potential novel metabolic molecular-targeted therapies.

Introduction

Lung cancer is the leading cause of cancer-related deaths worldwide, accounting for nearly 20% of all cancer deaths (*Bray et al., 2018*). Lung adenocarcinoma (LUAD) is one of the main subtypes of lung cancer (*Travis, 2020*), accounting for more than 40% of lung cancer cases (*Hutchinson et al., 2019*), and its relative frequency is increasing (*Twardella et al., 2018*). Despite great improvements in the treatment of LUAD, the prognosis in patients with LUAD remains poor owing to the lack of early detection and effective individual therapies (*Dolly et al., 2017*). Therefore, exploring prognostic biomarkers is a critical need to help predict prognosis in LUAD and to design individual therapies. Until now, most prognostic models were based on clinical characteristics (e.g., age, sex, TNM stage, vascular tumor invasion, and organization classification) or a single molecular biomarker, such as carcinoembryonic antigen and epidermal growth factor receptor. However, these prognostic models have limited power for predicting prognosis because of the complicated molecular mechanisms of LUAD development and

progression. Therefore, it is important to explore the mechanism of LUAD pathology in more depth using bioinformatics to construct prognostic models that predict the patient' prognosis more accurately.

Metabolic reprogramming is one of the hallmarks of cancer (*Faubert et al., 2020*), which takes place from the onset and throughout the development of cancer (*Chang, Fang & Gu, 2020*). It plays an important role in the progression, metastasis, depressed immunity, and therapy resistance of cancer (*Lane et al., 2019*). Metabolic reprogramming has been widely accepted as the basis for the discovery of novel tumor biomarkers. *Satriano et al. (2019)* observed that metabolic rearrangement played an important role in predicting the prognosis in patients with primary liver cancers. *Chen et al. (2019)* revealed that reprogrammed tumor glucose metabolism could promote cancer stemness and result in poor prognosis in breast cancer patients. There are hundreds of metabolic genes that consistently have an altered expression in LUAD (*Asavasupreechar et al., 2019; Vanhove et al., 2019*); however, their roles and mechanisms of action remain unclear. This study investigated the role of abnormal metabolism in predicting the prognosis in patients with LUAD.

With the development of genome sequencing and bioinformatics, new data have emerged. Prognosis-related gene signatures that were constructed using these new tools have made great contributions to tumor prognosis prediction. This study aimed to use bioinformatic methods to establish a prognostic metabolic-gene molecular model that can predict prognosis in patients with LUAD. This model could potentially guide personalized therapy for such patients.

Materials & Methods

Data expression datasets

The transcriptome expression profiles and corresponding clinical information for LUAD were downloaded from the Cancer Genome Atlas (TCGA; http://portal.gdc.cancer.gov/) and Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) databases. From the TCGA, gene expression data were of the HTSeq-FPKM type, obtained from 497 LUAD and 54 non-tumor samples. From the GEO, the GSE68465 dataset included 443 LUAD and 19 non-tumor samples, using the GPL96 platform (Affymetrix Human Genome U133A Array). The metabolic genes in the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway were extracted from Gene Set Enrichment Analysis (GSEA) (http://software.broadinstitute.org/gsea/index.jsp/), and the overlapping metabolism-related genes were identified from TCGA and GSE68465 (*Possemato et al., 2011; Zhu et al., 2020*).

Construction and validation of the prognostic metabolic gene signature

The clinical cases from the TCGA database were used to assess the prognostic associations of the metabolic genes with clinical outcomes. The differentially expressed genes (DEGs) between LUAD and paired non-tumor samples were obtained by the Wilcoxon rank sum test using the R package called "limma", and the adjusted P-value < 0.05 and absolute log2 fold change (FC) >1 were considered as the selection criterion. Univariate Cox regression analysis was used to

identify prognosis-related metabolic genes, and adjusted *P*-values < 0.001 were considered statistically significant. The lasso penalty for Cox proportional hazards model (1,000 iterations) was used to construct the prognostic gene-expression signature utilizing an R package called "glmnet." The prognostic gene-expression signature was designed using a risk scoring method with the following formula:

$$\boldsymbol{Risk\ score} = \sum_{i}^{n}(\boldsymbol{x_i} * \boldsymbol{\beta_i})$$

where x_i indicates the expression of gene i and β_i indicates the coefficient of gene i generated from the Cox multivariate regression.

The R package "survminer" was used to explore the cutoff point of the risk score, which divided patients into high- and low-risk groups. The R package "survival" was used to draw the Kaplan-Meier survival curves to demonstrate the overall survival (OS) in the high- and low-risk groups. The R package "survival ROC" was used to evaluate the prognostic value of the gene-expression signature.

Independence of the prognostic gene signature from other clinical characteristics

To determine whether the predictive power of the prognostic gene-expression signature could be independent from other clinicopathological variables in patients with LUAD (including age, sex, TNM stage, T stage, N stage, and M stage), univariate and multivariate Cox regression analyses were performed. The hazard ratio (HR), 95% confidence intervals (Cis), and *P*-values were calculated.

Construction and validation of a predictive nomogram

The nomogram was constructed using all the independent prognostic factors of the Cox regression analyses using R package "rms." Validation of the nomogram was assessed by discrimination and calibration using the concordance index (C-index) by Harrell et al. (1996) (bootstraps with 1,000 resamples) and the calibration plot, respectively.

External validation of the prognostic metabolic gene signature

To verify the prognostic metabolic-gene molecular signature in the GEO dataset, the risk score of patients was calculated directly with the gene-expression signature constructed from the TCGA dataset for further analysis. The receiver operating characteristic (ROC) and Kaplan-Meier analyses were performed identically with the gene signature in the TCGA dataset. The mRNA expression levels of the signature genes were analyzed further using online databases (the Oncomine database [http://www.oncomine.org/] and TIMER database [http://cistrome.shinyapps.io/timer/]). The protein expression levels associated with the signature genes were validated using the Human Protein Atlas database (http://www.proteinatlas.org/). The known genetic alterations of the signature genes were investigated using cBioPortal for Cancer Genomics (http://www.cbioportal.org/).

Gene Set Enrichment Analysis

Enrichment analysis of the KEGG pathways of the signature genes was performed using GSEA on the TCGA dataset. The nominal (NOM) P-value < 0.05 and the False Discovery Rate (FDR) q-value < 0.25 indicated statistical significance.

Statistical analysis

All analyses were performed using R software v3.6.3 (R Foundation for Statistical Computing, Vienna, Austria). Two-tailed P-values < 0.05 were considered statistically significant.

Results

Clinical characteristics

The TCGA dataset included 486 patients with LUAD (**Table S1**). The GEO dataset included 443 patients with LUAD (**Table S1**). Patients with a survival time of less than 30 days were omitted. For the study, 454 and 439 patients remained in the TCGA and GEO datasets, respectively. The detailed clinical characteristics of all patients are listed in **Table 1**.

Building and validation of the prognostic metabolic gene signature

To clarify our study design, a flow chart of the analysis procedure is presented in **Fig. 1**. A list of 994 genes in the KEGG pathway was identified from GSEA (**Table S2**), and 633 overlapping metabolism-related genes were abstracted from TCGA and GSE68465 (**Table S3**). The 96 DEGs (72 up-regulated genes and 24 download genes) between LUAD and paired non-tumor samples were identified from the further analysis (**Fig. 2; Table S4**). Seven significant genes associated with OS were identified using univariate analysis (**Table S4**). Furthermore, six genes were selected to build the prognostic model using a lasso-penalized Cox analysis (**Table 2**). The six genes were phosphofructokinase platelet (*PFKP*), pyruvate kinase muscle (*PKM*), triosephosphate isomerase 1 (*TPI1*), lactate dehydrogenase A (*LDHA*), prostaglandin E synthase (*PTGES*), and thymidylate synthase (*TYMS*). Risk score = (0.00005× *PFKP* mRNA level) + (0.00173× *PKM* mRNA level) + (0.00038× *TPI1* mRNA level) + (0.00379× *LDHA* mRNA level) + (0.00292 × *PTGES* mRNA level) + (0.02490× *TYMS* mRNA level).

The 445 patients with LUAD were divided into the high-risk or low-risk group based on the median risk score of 0.861 in the TCGA dataset. Patients in the high-risk group had significantly poorer OS than those in the low-risk group (P < 0.001; **Fig. 3A**). The distribution of the risk score and survival status of the patients is presented in **Fig. 3C**, which showed a higher mortality in the high-risk group than in the low-risk group. The expression of the six prognostic genes is shown in the heatmap. All the six genes had a significant positive correlation with the high-risk group (**Fig. 3E**). The area under the curve (AUC) of the time-dependent ROC curve was used to identify the prognostic ability of the six-gene molecular signature. The AUCs of the six-gene signature model were 0.693, 0.655, and 0.565 for the 1-, 3-, and 5-year OS, respectively, suggesting that the prediction model had a good performance in predicting the OS in patients with LUAD (**Fig. 3G**).

The prognostic model was validated in the GSE68465 dataset. The 439 patients with LUAD were divided into the high-risk or low-risk group based on the median risk score of 0.861. Patients in the high-risk group had a poor OS compared with those in the low-risk group ($P < 0.001$; **Fig. 3B**). The distribution of the risk score and survival status showed a higher mortality in the high-risk group than in the low-risk group (**Fig. 3D**). The expression heatmap of the six prognostic genes showed that all the six genes had a significant positive correlation with the high-risk group (**Fig. 3F**). The AUCs of the six-gene signature model were 0.728, 0.654, and 0.618 for the 1-, 3-, and 5-year OS, respectively (**Fig. 3H**). Taken together, these results suggested that the prognostic model had a high sensitivity and specificity in predicting the OS in patients with LUAD.

The prognostic gene signature was independent from other clinicopathological factors

Univariate and multivariate Cox regression analyses were conducted to assess the independent predictive value of the six-gene prognostic signature. In the TCGA dataset, univariate Cox regression analysis demonstrated that the prognostic model (HR: 2.845, $P < 0.001$), TNM stage (HR: 1.666, $P < 0.001$), T stage (HR: 1.605, $P < 0.001$), and N stage (HR: 1.806, $P < 0.001$) had a prognostic value for OS (**Fig. 4A**). Multivariate Cox regression analysis demonstrated that the only prognostic model (HR: 2.448, $P < 0.001$) and TNM stage (HR: 1.950, $P < 0.01$) were independent prognostic factors for OS (**Fig. 4A**). In the GSE68465 dataset, the prognostic model, T stage, N stage, and age had a prognostic value in the univariate and multivariate Cox regression analyses (**Fig. 4B**). Gender was the only independent prognostic factor for OS in the univariate Cox regression analysis (**Fig. 4B**).

In addition, the time-dependent ROC curve was used to identify the predictive ability of the prognostic model compared with the other clinicopathological characteristics. In the TCGA dataset, the AUCs of the prognostic model were 0.693, 0.655, and 0.565 for the 1-, 3-, and 5-year OS, respectively, which were higher than most of the other clinicopathological characteristics including age (0.498, 0.511, 0.485), gender (0.579, 0.485, 0.451), T stage (0.673, 0.613, 0.608), N stage (0.685, 0.666, 0.628), and M stage (0.508, 0.527, 0.530) (**Fig. 5A**). Furthermore, in the GSE68465 dataset, the AUCs of the prognostic model were 0.728, 0.654, and 0.618 for the 1-, 3-, and 5-year OS, respectively, which were higher than most of the other clinicopathological characteristics including age (0.593, 0.568, 0.581), gender (0.539, 0.549, 0.547), grade (0.580, 0.571, 0.548), T stage (0.647, 0.606, 0.606), and N stage (0.690, 0.680, 0.655) (**Fig. 5B**). The prognostic model had a larger AUC value compared with other clinicopathological characteristics. These results indicated that the model was an excellent prognostic model for LAUD patients, especially for the 1- and 3-year OS.

These results suggested that our prognostic model could be an independent predictor of prognosis in patients with LAUD.

Building and validating a predictive nomogram

A nomogram was built to predict the survival probability in patients with LAUD in the TCGA dataset. The nomogram was constructed using four prognostic factors (the TNM stage, T stage, N stage, and prognostic model; **Fig. 6A**). The C-index was calculated to evaluate the predictive ability of the nomogram for OS. The C-index for the nomogram was 0.754 (95% CI: 0.561–0.947). Calibration plots indicated that the nomogram had a good accuracy in predicting the 1- and 3-year OS (**Fig. 6B**).

To predict the survival probability more accurately, the combined prognostic model was built based on the nomogram. The combined prognostic model consisted of the TNM stage, T stage, N stage, and prognostic model. A time-dependent ROC curve was used to identify the predictive ability of the combined prognostic model. The AUCs of the combined prognostic models were 0.782, 0.717, and 0.688 for the 1-, 3-, and 5-year OS, respectively, which were higher than other clinical models including the TNM stage model (0.732, 0.687, 0.681), T stage model (0.671, 0.612, 0.613), N stage model (0.686, 0.661, 0.648), and the prognostic model (0.692, 0.634, 0.576). The combined model had the largest AUC value compared with other factors, which indicated that the combined model had a good predictive accuracy for survival. These results suggested that the predictive ability of the combined model built with the nomograms is better than other models, especially for predicting 1- and 3-year survival (**Fig. 6C**).

Gene Set Enrichment Analysis

To recognize signaling pathways that are differentially activated in LUAD, a GSEA was used, and a total of 49 significantly enriched KEGG pathways were found in the high-risk group and low-risk group (**Table S5**) of the TCGA dataset (FDR q-val < 0.25, NOM p-val < 0.05). Among them, many enriched pathways were related to metabolism and some highly dysregulated pathways including cell cycle, p53 signaling pathway, and basal transcription factors were also contained in these results (**Table S5**). We chose the top five significantly enriched metabolism-signaling pathways depending on the normalized enrichment score from the high-risk group or low-risk group. We found that the top five most significantly enriched metabolism-related pathways of the high-risk group were the cysteine and methionine, fructose and mannose, glyoxylate and dicarboxylate, purine, and pyrimidine pathways (**Fig. 7A**). The top five most significantly enriched metabolism-related pathways of the low-risk group were the alpha linolenic acid, arachidonic acid, ether lipid, glycerophospholipid, and linoleic acid pathways (**Fig. 7B**). Most of the metabolism-related pathways in the high-risk group mainly focused on amino acid and glycolysis metabolism, while the pathways in the low-risk group mainly focused on lipid metabolism. The results of the ten representative enriched metabolism-related KEGG pathways are given in **Table 3**. Furthermore, all the six metabolic genes of the prognostic model enriched these metabolism pathways significantly. LDHA enriched the cysteine and methionine pathway (**Table S6**); PFKP and TPI1 enriched the fructose and mannose pathway (**Table S6**); PKM enriched the purine pathway (**Table S6**); TYMS enriched the pyrimidine pathway (**Table S6**); and PTGES enriched the arachidonic acid pathway (**Table S6**). The results further

elucidated the role of metabolism in LUAD and the value of the six-gene signature in predicting the prognosis of LUAD.

External validation using online databases

To further identify the role of the six metabolic genes in LUAD, we compared the mRNA expression levels of the six metabolic genes (*PFKP*, *PKM*, *TPI1*, *LDHA*, *PTGES*, and *TYMS*) in the LAUD tissues with those in the normal lung tissues using data from the Oncomine database (**Fig. 8**). Obviously, all the six genes were overexpressed in lung cancer in all the datasets from the Oncomine database with the threshold of fold change=2, *P*-value=0.001(**Fig. 8A**). Furthermore, the mRNA levels of all the six genes in LUAD were significantly upregulated than those in normal tissues in the combined LUAD datasets from the Oncomine database **(Fig. 8B; Table 4)**. To further validate the overexpression of the six genes in LUAD, we analyzed the expression of the six genes using TIMER databases (**Fig. 9**). The results revealed that all the mRNA expression of the six genes in LUAD were significantly higher than in normal tissues. All the results from the Oncomine and TIMER databases were consistent with our results for the TCGA and GEO datasets. In addition, the mRNA expression of the six genes was also higher in esophageal carcinoma, head and neck squamous cell carcinoma, lung squamous cell carcinoma, and stomach adenocarcinoma from the TIMER databases (**Fig. 9**). The protein expressions of these six genes were analyzed using clinical specimens from the Human Protein Profiles (**Fig. 10A and 10B; Table 5**). The representative images of the six gene protein levels from the Human Protein Profiles are shown in **Fig. 10A**. Compared with the expression level in normal lung tissue, *LDHA* (100%, n=7) and *TYMS* (80%, n=5) showed a significantly higher percentage of high/medium expression levels in the LAUD tissue (**Fig. 10B; Table 5**). *PKM* (50%, n=6), *PFKP* (33.33%, n=6), and *PTGES* (16.67%, n=6) showed a significant moderate percentage of high/medium expression levels in the LAUD tissue (**Fig. 10B; Table 5**). However, *TPI1* showed no detected expression both in the LAUD and normal lung tissue (**Fig. 10B; Table 5**). The genetic alterations were explored in the cBioPortal database. Amplifications and mutations were the most common alterations in the six metabolic genes (**Fig. 10C**). The aberrant genetic alterations might elucidate the overexpression of these six genes in LUAD.

Altogether, the correlation of the aberrant expression of these six genes with LAUD cancer was further validated using multiple online databases.

Discussion

LUAD is the most common histological subtype of primary lung cancer. The incidence of LUAD has been increasing rapidly, and mortality has not significantly decreased despite great improvements in research and treatment. Therefore, exploring the molecular mechanisms of LUAD progression and constructing a valid and accurate molecule-based tool for evaluating the prognosis in patients is urgently needed. This could help design more efficient therapeutic strategies for LUAD. Metabolic reprogramming in cancers could lead to their development and progression (*Nwosu et al., 2017; Liu et al., 2020*). Characterization of the changes in metabolic

gene expression in LUAD would allow development of novel prognostic biomarkers. However, a single biomarker is not a robust measure for predicting patient prognosis. Thus, constructing a robust multiple-biomarker signature for predicting the prognosis in cancer patients is necessary.

We identified and designed a novel six-gene prognostic molecular signature based on the TCGA database and validated its efficiency in the GSE68465 dataset. The results indicated that the molecular signature was significantly associated with OS in patients with LUAD in the training and validation sets. These results indicate that the molecular signature has a robust prognostic value, especially for predicting short-term survival in patients with LUAD. These results also demonstrated that the prognostic signature was independent of other clinicopathological characteristics, which further supports the prognostic value of this signature.

To increase the accuracy of the prediction of prognosis, we constructed a nomogram built with the combination of genetic and clinically related variables of patients with LUAD. The nomogram included the prognostic model, TNM, T stage, and N stage. Its predictive accuracy was verified using calibration plots, the C-index, and the AUC, which indicated that the nomogram had a greater predictive value than the previous systems. The Gene Set Enrichment Analysis showed that many significantly enriched pathways were metabolism-related pathways. The different risk groups possessed different metabolic pathway features. The metabolism-related pathways in the high-risk group were mainly associated with amino acid and glycolysis metabolism, while the pathways in the low-risk group were mainly associated with lipid metabolism. These results revealed that the different risk groups possessed the different metabolic features, which might provide the underlying metabolic mechanisms of promoting the prognosis of LUAD. All these results further suggest a strong association between the molecular signature and metabolic systems and might reflect the dysregulated metabolic microenvironment of cancers.

Most of the six genes in our prognostic signature are suggested to be related to cancer development. *PFKP* is a major isoform of cancer-specific phosphofructokinase-1, an enzyme that catalyzes the phosphorylation of fructose-6-phosphate to form fructose-1,6-bisphosphate. Recently, *PFKP* was noted to have an aberrant upregulation in many cancers, such as breast cancer, prostate cancer, and glioblastoma. The dynamic upregulation of *PFKP* promotes metabolic reprogramming and cancer cell survival (*Bjerre et at., 2019; Kim et al., 2017*). As a key regulator enzyme in glycolysis, PFKP enriched the fructose and mannose metabolism pathway. Recent studies showed that PFKP is highly expressed in lung cancer and promotes lung cancer development via fructose and mannose metabolism (*Shen et al., 2020; Wang, et al., 2015*). *PKM* is a rate-limiting enzyme in the final step of glycolysis, that is considered as one of the metabolic hallmarks of cancer (*Prakasam et al., 2017*). The abnormal expression of *PKM* promoted cancer growth, invasion, and metastasis by governing aerobic glycolysis (*Prakasam et al., 2017; Zahra et al., 2020*) and induced cancer treatment resistance (*Calabretta et al., 2016*). Furthermore, *PKM* is overexpressed in non-small cell lung cancer (NSCLC) and involved in the development and prognosis of NSCLC (*Luo et al., 2018*). *TPI1* is a crucial enzyme in carbohydrate metabolism, catalyzing the interconversion of dihydroxyacetone phosphate and d-

glyceraldehyde-3-phosphate during glycolysis and gluconeogenesis. *TPI1* is abnormally expressed in different kinds of cancers, such as breast cancer, gastric cancer, and lymphoma and is associated with a poor prognosis in patients with neuroblastoma and pancreatic cancer through dysregulating glycometabolism (*Ludvigsen et al., 2018; Applebaum et al., 2016; Follia et al., 2019*). *LDHA* is an enzyme that catalyzes the interconversion of pyruvate and lactate. *LDHA* was enriched in cysteine and methionine metabolism, and its aberrant metabolism regulation promoted many pathological processes in tumors, such as cell proliferation, survival, invasion, metastasis, and immunity (*Dorneburg et al., 2018*). Overexpressed *LDHA* is associated with poor prognosis in many tumors, including NSCLC, breast cancer, gallbladder carcinoma, and gastrointestinal cancer (*Mizuno et al., 2020; Guddeti et al., 2019*). *PTGES* is a key enzyme in the arachidonic acid metabolism pathway. An abnormally high expression of *PTGES* is correlated with proliferation, invasion, and metastasis in many cancer cells (*Kim et al., 2016; Delgado-Goñi et al., 2020*). The dysregulated PTGES promoted tumor migration and metastasis of lung cancer cells and played an important role in lung cancer progression (*Wang et al., 2019*). *TYMS is* a rate-limiting enzyme, which plays an important role in regulating the pyrimidine metabolism signaling pathway (*Yeh et al., 2017*). *TYMS* is overexpressed frequently in different kinds of cancers, such as NSCLC, pancreatic, colorectal, and breast cancers, and it has resulted in a poor cancer prognosis and chemotherapy resistance via dysregulating pyrimidine metabolism (*Troncarelli Flores et al., 2019; Wu et al., 2019*). In our study, we constructed a six-gene signature for a prognostic model based on the TCGA database. This novel six-gene signature had a higher survival prediction, and the predictive ability of this signature was further validated by the GSE68465 dataset and multiple online databases. To our knowledge, the six-gene signature for prognosis prediction in LUAD has not been reported yet. Compared with the traditional prognostic models such as clinical characteristics (e.g., TNM stage, vascular tumor invasion, and organization classification) or a single molecular biomarker, a multi-gene signature can predict the prognosis more accurately and provide a clearer molecular mechanism for personalized LUAD therapy.

There are limitations in our study. First, our nomogram was not validated further in the GEO database because the GSE68465 lacked detailed TNM stage data. Thus, the nomogram should be externally validated using larger datasets from multicenter clinical trials and perspective studies. Second, functional experiments should be further performed to explore the molecular mechanisms predicted by the metabolic gene expression.

Conclusions

We concluded from our research results that the six-gene metabolic prognostic signature could accurately predict the prognosis in patients with LUAD. The molecular signature may provide potential biomarkers for metabolic therapy and prognosis prediction of LUAD.

References

Applebaum, M. A., Jha, A. R., Kao, C., Hernandez, K. M., DeWane, G., Salwen, H. R., Chlenski, A., Dobratic, M., Mariani, C. J., Godley, L. A., Prabhakar, N., White, K., Stranger, B. E., & Cohn, S. L. 2016. Integrative genomics reveals hypoxia inducible genes that are associated with a poor prognosis in neuroblastoma patients. *Oncotarget* **7(47)**:76816–76826 DOI 10.18632/oncotarget.12713.

Asavasupreechar, T., Chan, M., Saito, R., Miki, Y., Boonyaratanakornkit, V., & Sasano, H. 2019. Sex steroid metabolism and actions in non-small cell lung carcinoma. *The Journal of steroid biochemistry and molecular biology* **193**:105440 DOI 10.1016/j.jsbmb.2019.105440.

Beer, D. G., Kardia, S. L., Huang, C. C., Giordano, T. J., Levin, A. M., Misek, D. E., Lin, L., Chen, G., Gharib, T. G., Thomas, D. G., Lizyness, M. L., Kuick, R., Hayasaka, S., Taylor, J. M., Iannettoni, M. D., Orringer, M. B., & Hanash, S. 2002. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nature medicine* **8(8)**:816–824 DOI 10.1038/nm733.

Bjerre, M. T., Strand, S. H., Nørgaard, M., Kristensen, H., Rasmussen, A. K., Mortensen, M. M., Fredsøe, J., Mouritzen, P., Ulhøi, B., Ørntoft, T., Borre, M., & Sørensen, K. D. 2019. Aberrant DOCK2, GRASP, HIF3A and PKFP Hypermethylation has Potential as a Prognostic Biomarker for Prostate Cancer. *International journal of molecular sciences* **20(5)**:1173 DOI 10.3390/ijms20051173.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* **68(6)**:394–424 DOI 10.3322/caac.21492.

Calabretta, S., Bielli, P., Passacantilli, I., Pilozzi, E., Fendrich, V., Capurso, G., Fave, G. D., & Sette, C. 2016. Modulation of PKM alternative splicing by PTBP1 promotes gemcitabine resistance in pancreatic cancer cells. *Oncogene* **35(16)**:2031–2039 DOI 10.1038/onc.2015.270.

Chang, L., Fang, S., & Gu, W. 2020. The Molecular Mechanism of Metabolic Remodeling in Lung Cancer. *Journal of Cancer* **11(6)**:1403–1411 DOI 10.7150/jca.31406.

Chen, C., Bai, L., Cao, F., Wang, S., He, H., Song, M., Chen, H., Liu, Y., Guo, J., Si, Q., Pan, Y., Zhu, R., Chuang, T. H., Xiang, R., & Luo, Y. 2019. Targeting LIN28B reprograms tumor glucose metabolism and acidic microenvironment to suppress cancer stemness and metastasis. *Oncogene* **38(23)**:4527–4539 DOI 10.1038/s41388-019-0735-4.

Delgado-Goñi, T., Galobart, T. C., Wantuch, S., Normantaite, D., Leach, M. O., Whittaker, S. R., & Beloueche-Babari, M. 2020. Increased inflammatory lipid metabolism and anaplerotic mitochondrial activation follow acquired resistance to vemurafenib in BRAF-mutant melanoma cells. *British journal of cancer* **122(1)**:72–81 DOI 10.1038/s41416-019-0628-x.

Dolly, S. O., Collins, D. C., Sundar, R., Popat, S., & Yap, T. A. 2017. Advances in the Development of Molecularly Targeted Agents in Non-Small-Cell Lung Cancer. *Drugs* **77(8)**: 813–827 DOI 10.1007/s40265-017-0732-2.

Dorneburg, C., Fischer, M., Barth, T., Mueller-Klieser, W., Hero, B., Gecht, J., Carter, D. R., de Preter, K., Mayer, B., Christner, L., Speleman, F., Marshall, G. M., Debatin, K. M., & Beltinger, C. 2018. LDHA in Neuroblastoma Is Associated with Poor Outcome and Its Depletion Decreases Neuroblastoma Growth Independent of Aerobic Glycolysis. *Clinical cancer research* **24(22)**:5772–5783 DOI 10.1158/1078-0432.CCR-17-2578.

Exposito, F., Villalba, M., Redrado, M., de Aberasturi, A. L., Cirauqui, C., Redin, E., Guruceaga, E., de Andrea, C., Vicent, S., Ajona, D., Montuenga, L. M., Pio, R., & Calvo, A. 2019. Targeting of TMPRSS4 sensitizes lung cancer cells to chemotherapy by impairing the proliferation machinery. *Cancer letters* **453**: 21–33 DOI 10.1016/j.canlet.2019.03.013.

Faubert, B., Solmonson, A., & DeBerardinis, R. J. 2020. Metabolic reprogramming and cancer progression. *Science* **368(6487)**:eaaw5473 DOI 10.1126/science.aaw5473.

Follia, L., Ferrero, G., Mandili, G., Beccuti, M., Giordano, D., Spadi, R., Satolli, M. A., Evangelista, A., Katayama, H., Hong, W., Momin, A. A., Capello, M., Hanash, S. M., Novelli, F., & Cordero, F. 2019. Integrative Analysis of Novel Metabolic Subtypes in Pancreatic Cancer Fosters New Prognostic Biomarkers. *Frontiers in oncology* **9**:115 DOI 10.3389/fonc.2019.00115.

Garber, M. E., Troyanskaya, O. G., Schluens, K., Petersen, S., Thaesler, Z., Pacyna-Gengelbach, M., van de Rijn, M., Rosen, G. D., Perou, C. M., Whyte, R. I., Altman, R. B., Brown, P. O., Botstein, D., & Petersen, I. 2001. Diversity of gene expression in adenocarcinoma of the lung. *Proceedings of the National Academy of Sciences of the United States of America* **98(24)**:13784–13789 DOI 10.1073/pnas.241500798.

Guddeti, R. K., Bali, P., Karyala, P., & Pakala, S. B. 2019. MTA1 coregulator regulates LDHA expression and function in breast cancer. *Biochemical and biophysical research communications* **520(1)**:54–59 DOI 10.1016/j.bbrc.2019.09.078.

Hou, J., Aerts, J., den Hamer, B., van Ijcken, W., den Bakker, M., Riegman, P., van der Leest, C., van der Spek, P., Foekens, J. A., Hoogsteden, H. C., Grosveld, F., & Philipsen, S. 2010. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. *PloS one* **5(4)**:e10312 DOI 10.1371/journal.pone.0010312.

Hutchinson, B. D., Shroff, G. S., Truong, M. T., & Ko, J. P. 2019. Spectrum of Lung Adenocarcinoma. *Seminars in ultrasound, CT, and MR* **40(3)**:255–264 DOI 10.1053/j.sult.2018.11.009.

Kim, N. H., Cha, Y. H., Lee, J., Lee, S. H., Yang, J. H., Yun, J. S., Cho, E. S., Zhang, X., Nam, M., Kim, N., Yuk, Y. S., Cha, S. Y., Lee, Y., Ryu, J. K., Park, S., Cheong, J. H., Kang, S. W., Kim, S. Y., Hwang, G. S., Yook, J. I., Kim, H. S. 2017. Snail reprograms glucose metabolism by repressing phosphofructokinase PFKP allowing cancer cell survival under metabolic stress. *Nature communications* **8**:14374 DOI 10.1038/ncomms14374.

Kim, S. B., Bozeman, R. G., Kaisani, A., Kim, W., Zhang, L., Richardson, J. A., Wright, W. E., & Shay, J. W. 2016. Radiation promotes colorectal cancer initiation and progression by inducing senescence-associated inflammatory responses. *Oncogene* **35(26)**:3365–3375 DOI 10.1038/onc.2015.395.

Landi, M. T., Dracheva, T., Rotunno, M., Figueroa, J. D., Liu, H., Dasgupta, A., Mann, F. E., Fukuoka, J., Hames, M., Bergen, A. W., Murphy, S. E., Yang, P., Pesatori, A. C., Consonni, D., Bertazzi, P. A., Wacholder, S., Shih, J. H., Caporaso, N. E., & Jen, J. 2008. Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PloS one* **3(2)**: e1651 DOI 10.1371/journal.pone.0001651.

Lane, A. N., Higashi, R. M., & Fan, T. W. 2019. Metabolic reprogramming in tumors: Contributions of the tumor microenvironment. *Genes & diseases* **7(2)**:185–198 DOI 10.1016/j.gendis.2019.10.007.

Liu, G. M., Xie, W. X., Zhang, C. Y., & Xu, J. W. 2020. Identification of a four-gene metabolic signature predicting overall survival for hepatocellular carcinoma. *Journal of cellular physiology* **235(2)**:1624–1636 DOI 10.1002/jcp.29081.

Ludvigsen, M., Bjerregård Pedersen, M., Lystlund Lauridsen, K., Svenstrup Poulsen, T., Hamilton-Dutoit, S. J., Besenbacher, S., Bendix, K., Møller, M. B., Nørgaard, P., d'Amore, F., & Honoré, B. 2018. Proteomic profiling identifies outcome-predictive markers in patients with peripheral T-cell lymphoma, not otherwise specified. *Blood advances* **2(19)**:2533–2542 DOI 10.1182/bloodadvances.2018019893.

Luo, M., Luo, Y., Mao, N., Huang, G., Teng, C., Wang, H., Wu, J., Liao, X., & Yang, J. 2018. Cancer-Associated Fibroblasts Accelerate Malignant Progression of Non-Small Cell Lung Cancer via Connexin 43-Formed Unidirectional Gap Junctional Intercellular Communication. *Cellular physiology and biochemistry* **51(1)**:315–336 DOI 10.1159/000495232.

Mizuno, Y., Hattori, K., Taniguchi, K., Tanaka, K., Uchiyama, K., & Hirose, Y. 2020. Intratumoral heterogeneity of glutaminase and lactate dehydrogenase A protein expression in colorectal cancer. *Oncology letters* **19(4)**:2934–2942 DOI 10.3892/ol.2020.11390.

Nwosu, Z. C., Megger, D. A., Hammad, S., Sitek, B., Roessler, S., Ebert, M. P., Meyer, C., & Dooley, S. 2017. Identification of the Consistently Altered Metabolic Targets in Human Hepatocellular Carcinoma. *Cellular and molecular gastroenterology and hepatology* **4(2)**:303–323.e1 DOI 10.1016/j.jcmgh.2017.05.004.

Possemato, R., Marks, K. M., Shaul, Y. D., Pacold, M. E., Kim, D., Birsoy, K., Sethumadhavan, S., Woo, H. K., Jang, H. G., Jha, A. K., Chen, W. W., Barrett, F. G., Stransky, N., Tsun, Z. Y., Cowley, G. S., Barretina, J., Kalaany, N. Y., Hsu, P. P., Ottina, K., Chan, A. M., … Sabatini, D. M. 2011. Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature* **476(7360)**:346–350 DOI 10.1038/nature10350.

Prakasam, G., Singh, R. K., Iqbal, M. A., Saini, S. K., Tiku, A. B., & Bamezai, R. 2017. Pyruvate kinase M knockdown-induced signaling via AMP-activated protein kinase promotes

mitochondrial biogenesis, autophagy, and cancer cell survival. *The Journal of biological chemistry* **292(37)**:15561–15576 DOI 10.1074/jbc.M117.791343.

Satriano, L., Lewinska, M., Rodrigues, P. M., Banales, J. M., & Andersen, J. B. 2019. Metabolic rearrangements in primary liver cancers: cause and consequences. *Nature reviews. Gastroenterology & hepatology* **16(12)**:748–766 DOI 10.1038/s41575-019-0217-8.

Selamat, S. A., Chung, B. S., Girard, L., Zhang, W., Zhang, Y., Campan, M., Siegmund, K. D., Koss, M. N., Hagen, J. A., Lam, W. L., Lam, S., Gazdar, A. F., & Laird-Offringa, I. A. 2012. Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. *Genome research* **22(7)**:1197–1211 DOI 10.1101/gr.132662.111.

Shen, J., Jin, Z., Lv, H., Jin, K., Jonas, K., Zhu, C., & Chen, B. 2020. PFKP is highly expressed in lung cancer and regulates glucose metabolism. *Cellular oncology (Dordrecht)* **43(4)**:617–629 DOI 10.1007/s13402-020-00508-6.

Stearman, R. S., Dwyer-Nield, L., Zerbe, L., Blaine, S. A., Chan, Z., Bunn, P. A., Jr, Johnson, G. L., Hirsch, F. R., Merrick, D. T., Franklin, W. A., Baron, A. E., Keith, R. L., Nemenoff, R. A., Malkinson, A. M., & Geraci, M. W. 2005. Analysis of orthologous gene expression between human pulmonary adenocarcinoma and a carcinogen-induced murine model. *The American journal of pathology* **167(6)**:1763–1775 DOI 10.1016/S0002-9440(10)61257-6.

Su, L. J., Chang, C. W., Wu, Y. C., Chen, K. C., Lin, C. J., Liang, S. C., Lin, C. H., Whang-Peng, J., Hsu, S. L., Chen, C. H., & Huang, C. Y. 2007. Selection of DDX5 as a novel internal control for Q-RT-PCR from microarray data using a block bootstrap re-sampling scheme. *BMC genomics* **8**:140 DOI 10.1186/1471-2164-8-140.

Travis W. D. 2020. Lung Cancer Pathology: Current Concepts. *Clinics in chest medicine* **41(1)**:67–85 DOI 10.1016/j.ccm.2019.11.001.

Troncarelli Flores, B. C., Souza E Silva, V., Ali Abdallah, E., Mello, C., Gobo Silva, M. L., Gomes Mendes, G., Camila Braun, A., Aguiar Junior, S., & Thomé Domingos Chinen, L. 2019. Molecular and Kinetic Analyses of Circulating Tumor Cells as Predictive Markers of Treatment Response in Locally Advanced Rectal Cancer Patients. *Cells* **8(7)**:641 DOI 10.3390/cells8070641.

Twardella, D., Geiss, K., Radespiel-Tröger, M., Benner, A., Ficker, J. H., & Meyer, M. 2018. Trends der Lungenkrebsinzidenz nach histologischem Subtyp bei Männern und Frauen in Deutschland : Analyse von Krebsregisterdaten unter Einsatz von multipler Imputation [Trends in incidence of lung cancer according to histological subtype among men and women in Germany : Analysis of cancer registry data with the application of multiple imputation techniques]. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz* **61(1)**:20–31 DOI 10.1007/s00103-017-2659-x.

Vanhove, K., Graulus, G. J., Mesotten, L., Thomeer, M., Derveaux, E., Noben, J. P., Guedens, W., & Adriaensens, P. 2019. The Metabolic Landscape of Lung Cancer: New Insights in a Disturbed Glucose Metabolism. *Frontiers in oncology* **9**:1215 DOI 10.3389/fonc.2019.01215.

Wang, T., Jing, B., Sun, B., Liao, Y., Song, H., Xu, D., Guo, W., Li, K., Hu, M., Liu, S., Ling, J., Kuang, Y., Feng, Y., Zhou, B. P., & Deng, J. 2019. Stabilization of PTGES by deubiquitinase USP9X promotes metastatic features of lung cancer via PGE2 signaling. *American journal of cancer research* **9(6)**:1145–1160.

Wang, Y., Mei, Q., Ai, Y. Q., Li, R. Q., Chang, L., Li, Y. F., Xia, Y. X., Li, W. H., & Chen, Y. 2015. Identification of lung cancer oncogenes based on the mRNA expression and single nucleotide polymorphism profile data. *Neoplasma* **62(6)**:966–973 DOI 10.4149/neo_2015_117.

Wu, Q., Zhang, B., Sun, Y., Xu, R., Hu, X., Ren, S., Ma, Q., Chen, C., Shu, J., Qi, F., He, T., Wang, W., & Wang, Z. 2019. Identification of novel biomarkers and candidate small molecule drugs in non-small-cell lung cancer by integrated microarray analysis. *OncoTargets and therapy* **12**:3545–3563 DOI 10.2147/OTT.S198621.

Yamagata, N., Shyr, Y., Yanagisawa, K., Edgerton, M., Dang, T. P., Gonzalez, A., Nadaf, S., Larsen, P., Roberts, J. R., Nesbitt, J. C., Jensen, R., Levy, S., Moore, J. H., Minna, J. D., & Carbone, D. P. 2003. A training-testing approach to the molecular classification of resected non-small cell lung cancer. *Clinical cancer research: an official journal of the American Association for Cancer Research* **9(13)**:4695–4704.

Yeh, H. W., Lee, S. S., Chang, C. Y., Hu, C. M., & Jou, Y. S. 2017. Pyrimidine metabolic rate limiting enzymes in poorly-differentiated hepatocellular carcinoma are signature genes of cancer stemness and associated with poor prognosis. *Oncotarget* **8(44)**:77734–77751 DOI 10.18632/oncotarget.20774.

Zahra, K., Dey, T., Ashish, Mishra, S. P., & Pandey, U. 2020. Pyruvate Kinase M2 and Cancer: The Role of PKM2 in Promoting Tumorigenesis. *Frontiers in oncology* **10**:159 DOI 10.3389/fonc.2020.00159.

Zhu, Z., Li, L., Xu, J., Ye, W., Chen, B., Zeng, J., & Huang, Z. 2020. Comprehensive analysis reveals a metabolic ten-gene signature in hepatocellular carcinoma. *PeerJ* **8**: e9201 DOI 10.7717/peerj.9201.

Figure 1

Overall flowchart of steps used in the construction of the prognostic metabolic gene signature.

The TCGA dataset was utilized to construct the prognostic metabolic gene signature. The TCGA clinical information, GSE68465 dataset and online databases from international platforms were further utilized to validate the prognostic model. TCGA, The Cancer Genome Atlas; OS, overall survival; ROC, the receiver operating characteristic.

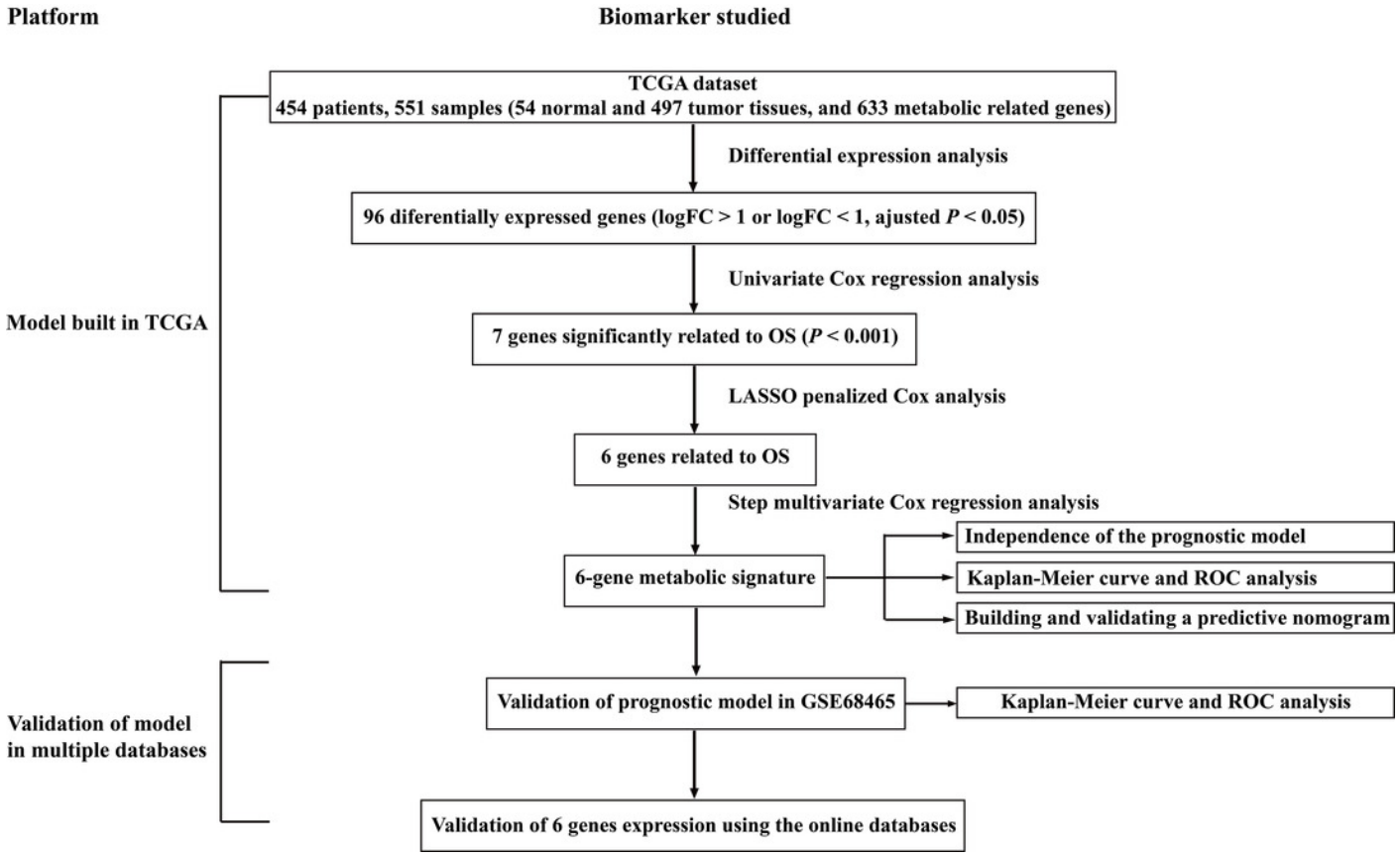


Figure 2

Heatmap and Volcano plot of metabolism-related DEGs.

(A) The heatmap of metabolism-related DEGs. The red color represented high expression genes, the green color represented low expression genes, and the black color represented the expression genes with no significant difference (FDR < 0.05, absolute log FC > 1). (B) Volcano plot of metabolism-related DEGs. The red, green and black dots represented the high expression genes, low expression genes, and the expression genes with no significant difference (FDR < 0.05, absolute log FC > 1). DEGs, differentially expressed genes; FDR, false discovery rate.

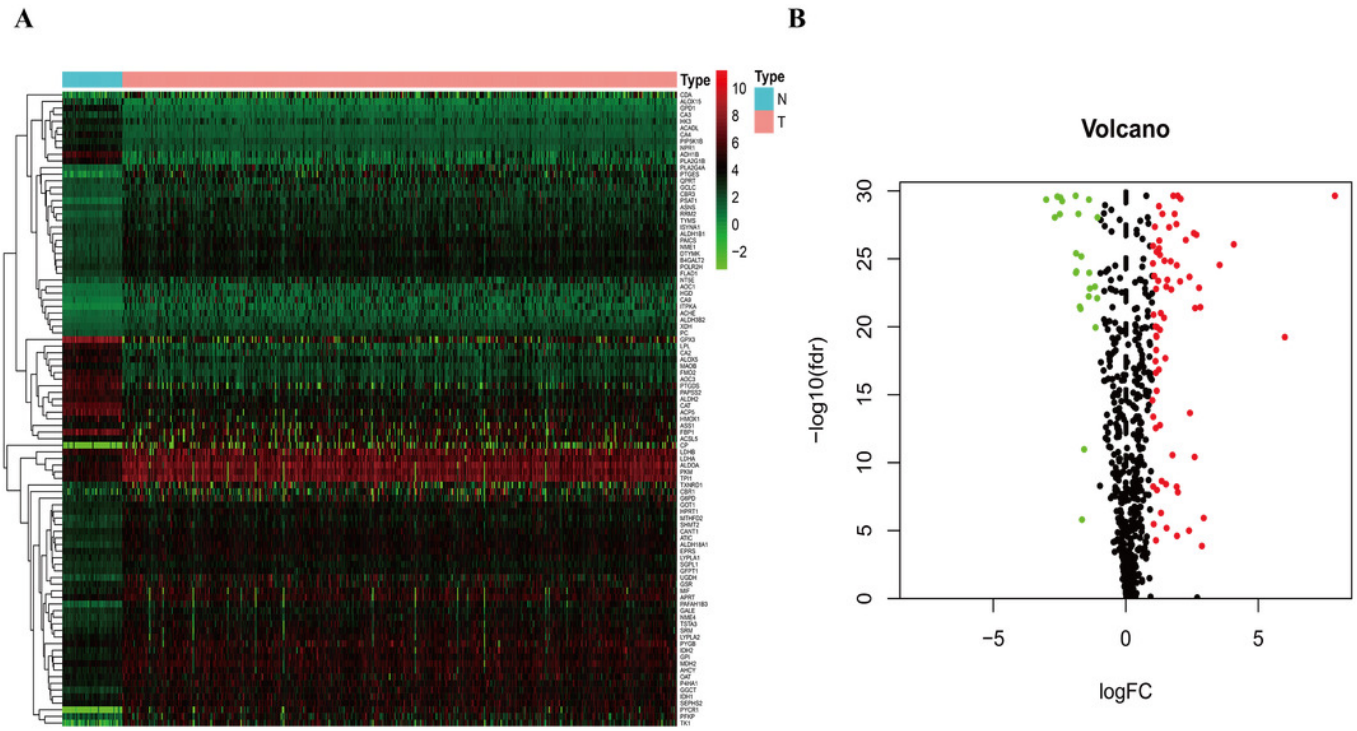


Figure 3

Identification of the prognostic model in lung adenocarcinoma.

(A, B) Kaplan-Meier curves of overall survival of the high-risk and low-risk groups stratified by the six-gene signature- based risk score in the TCGA or GEO dataset. (C, D) Risk score distribution, survival status distribution in the TCGA or GEO dataset. (E, F) The expression heatmap of the six prognostic genes in the TCGA or GEO dataset. (G, H) Time-dependent ROC curves of the six-gene signature in the TCGA or GEO dataset. TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus; ROC, receiver operating characteristic.

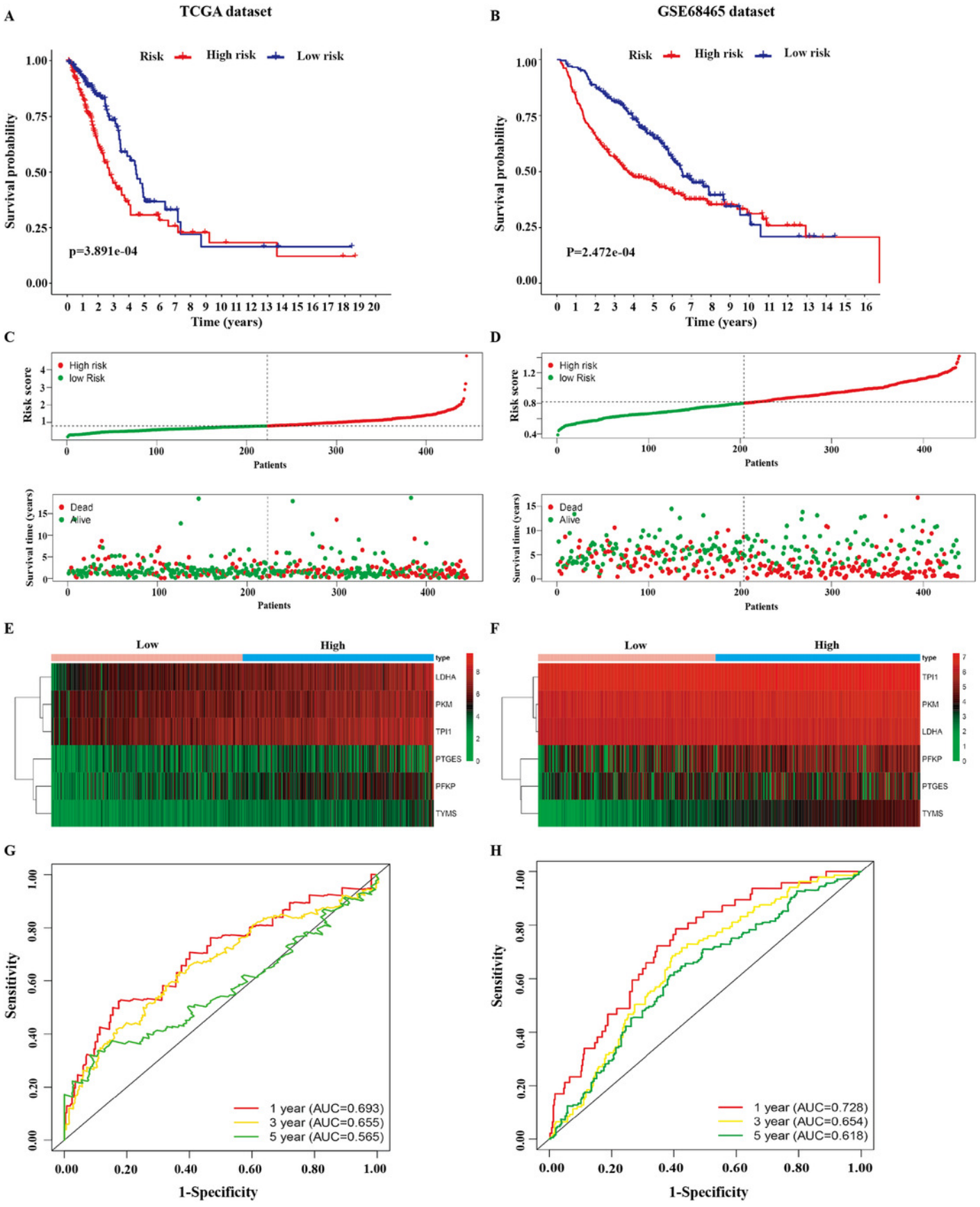

A
TCGA dataset
Risk
High risk
Low risk
Survival probability
p=3.891e-04
Time (years)
B
GSE68465 dataset
Risk
High risk
Low risk
Survival probability
P=2.472e-04
Time (years)
C
High risk
low Risk
Risk score
Patients
Dead
Alive
Survival time (years)
Patients
D
High risk
low Risk
Risk score
Patients
Dead
Alive
Survival time (years)
Patients
E
Low
High
type
LDHA
PKM
TPI1
PTGES
PFKP
TYMS
F
Low
High
type
TPI1
PKM
LDHA
PFKP
PTGES
TYMS
G
Sensitivity
1-Specificity
1 year (AUC=0.693)
3 year (AUC=0.655)
5 year (AUC=0.565)
H
Sensitivity
1-Specificity
1 year (AUC=0.728)
3 year (AUC=0.654)
5 year (AUC=0.618)

Figure 4

Cox regression analysis of the associations between the prognostic model and clinicopathological characteristics with overall survival in LAUD.

Univariate and multivariate Cox regression analyses in the TCGA dataset (A) and GEO dataset (B). LUAD, lung adenocarcinoma; TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus.

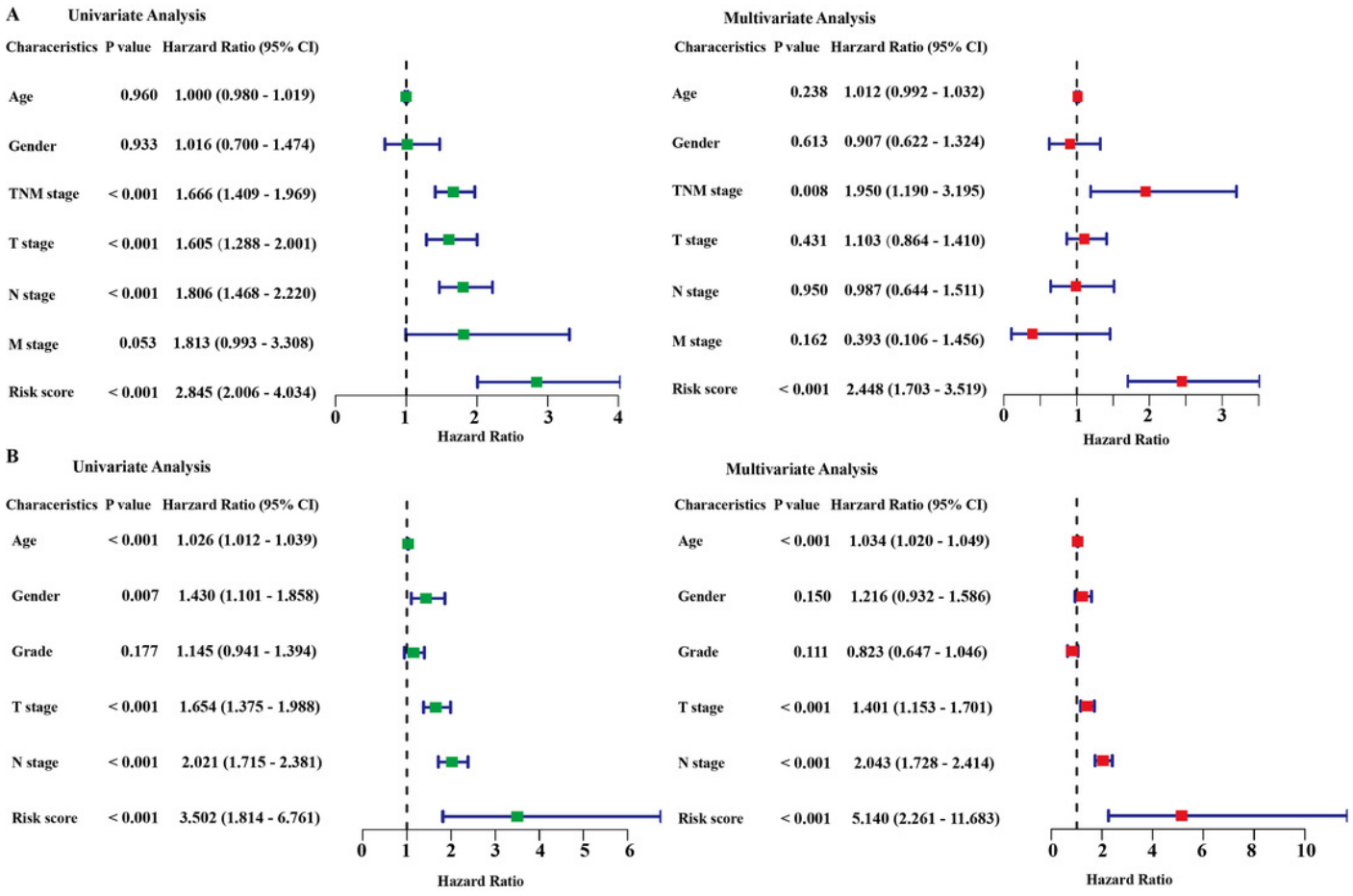


Figure 5

The time-dependent receiver operating characteristic (ROC) analysis for the prognostic model and clinicopathological characteristics in LAUD.

(A) The time-dependent ROC curves of risk score, age, gender, TNM stage, T stage, N stage, and M stage in the TCGA dataset. (B) The time-dependent ROC curves of risk score, age, gender, grade, T stage, and N stage in the GEO dataset. LUAD, lung adenocarcinoma.

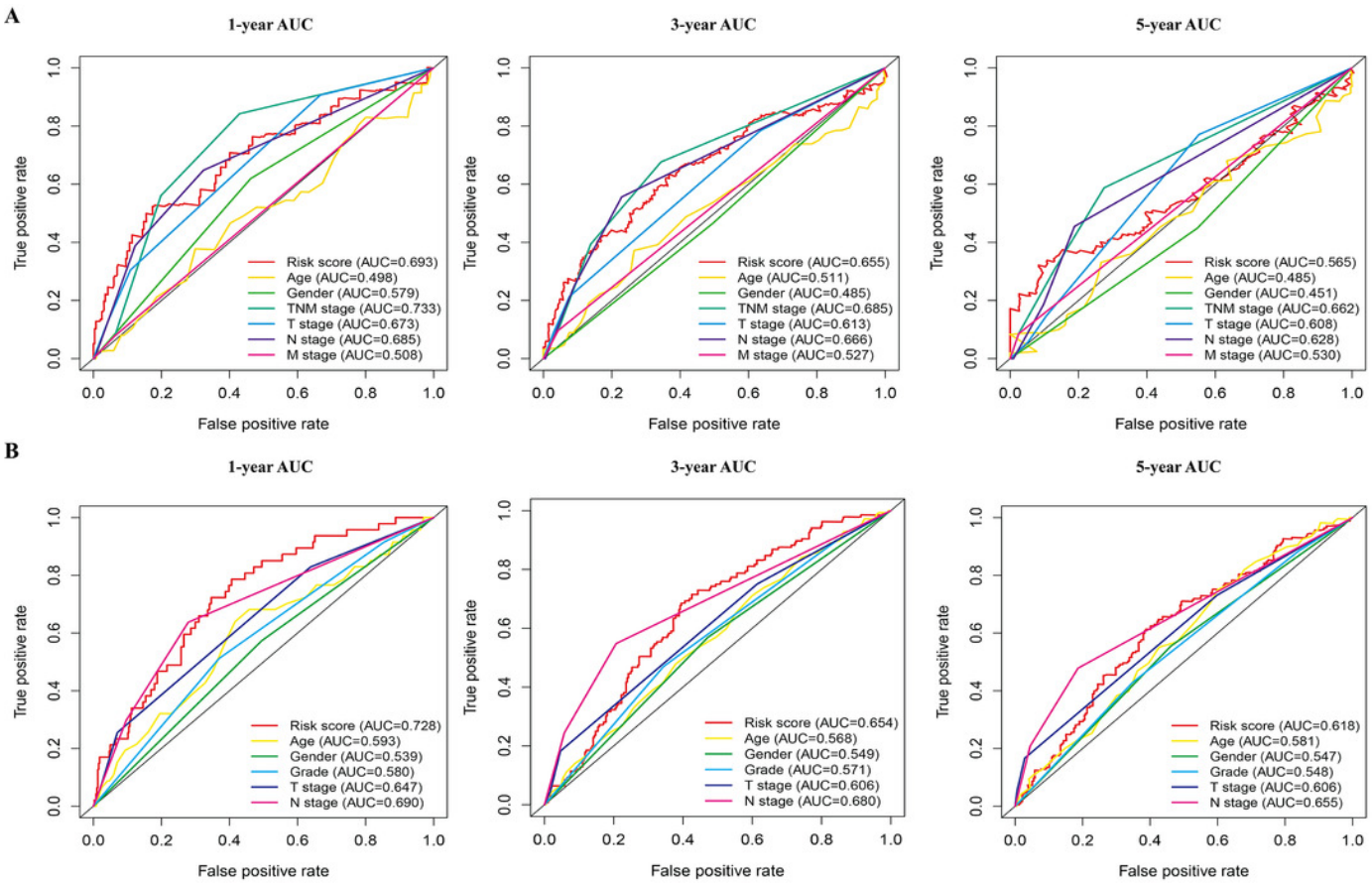


Figure 6

Construction and validation of a nomogram for survival prediction in LUAD from the TCGA dataset.

(A) The nomogram was built in the TCGA dataset. (B) Calibration plots revealed the nomogram-predicted survival probabilities. (C) The time-dependent ROC analysis evaluated the accuracy of the nomogram. TCGA, The Cancer Genome Atlas; ROC, receiver operating characteristic; LUAD, lung adenocarcinoma.

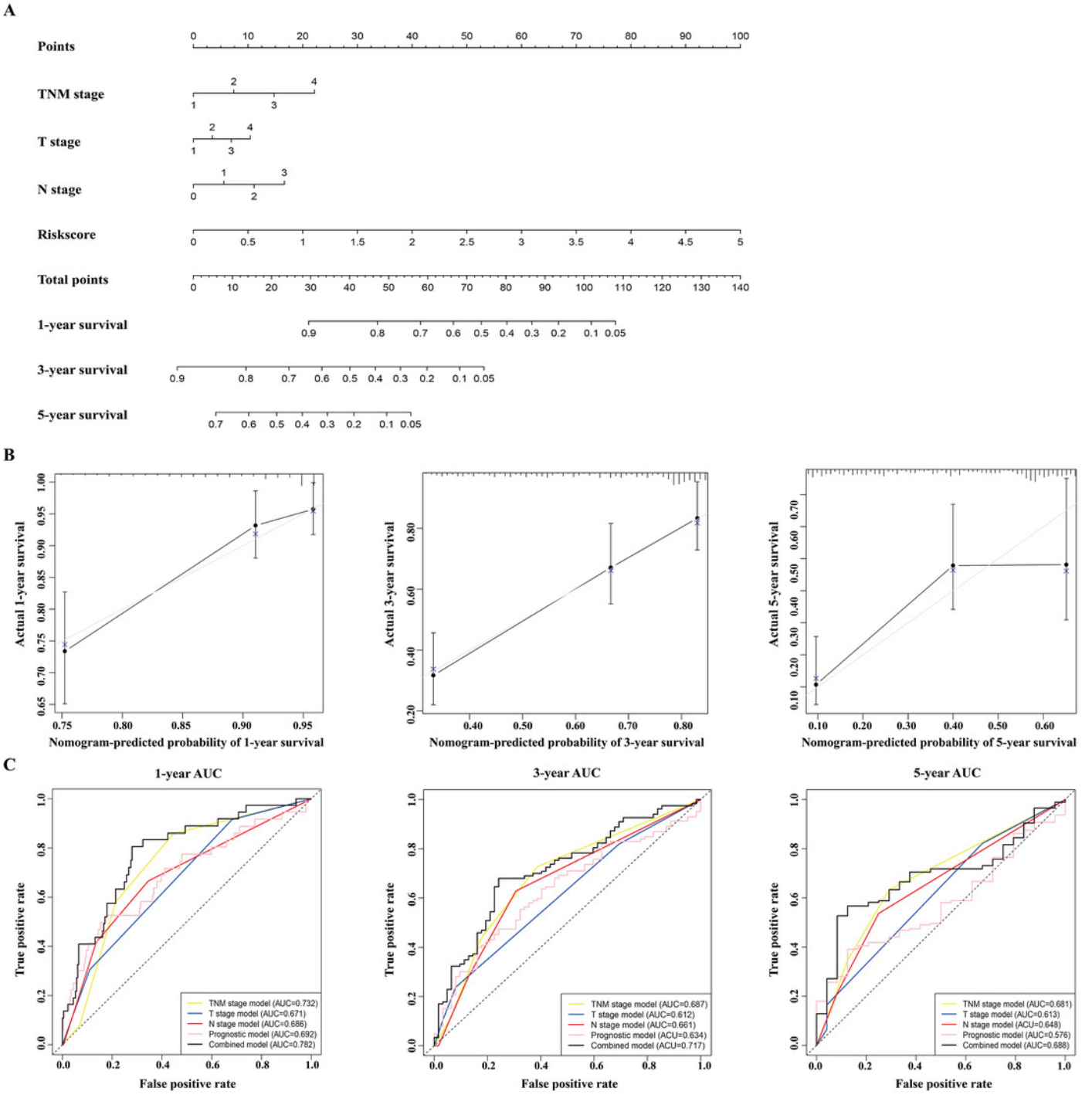

A
Points
TNM stage
T stage
N stage
Riskscore
Total points
1-year survival
3-year survival
5-year survival
B
Actual 1-year survival
Nomogram-predicted probability of 1-year survival
Actual 3-year survival
Nomogram-predicted probability of 3-year survival
Actual 5-year survival
Nomogram-predicted probability of 5-year survival
C
1-year AUC
True positive rate
False positive rate
TNM stage model (AUC=0.732)
T stage model (AUC=0.671)
N stage model (AUC=0.686)
Prognostic model (AUC=0.692)
Combined model (AUC=0.782)
3-year AUC
TNM stage model (AUC=0.687)
T stage model (AUC=0.612)
N stage model (AUC=0.661)
Prognostic model (ACU=0.634)
Combined model (ACU=0.717)
5-year AUC
TNM stage model (AUC=0.681)
T stage model (AUC=0.613)
N stage model (ACU=0.648)
Prognostic model (AUC=0.576)
Combined model (AUC=0.688)

Figure 7

The representative enriched metabolism-related KEGG pathways in the TCGA dataset by GSEA.

(A) The top five significantly representative enriched metabolism-related KEGG pathways in the high-risk group. (B) The top five significantly representative enriched metabolism-related KEGG pathways in the low-risk group. Related parameters for the ten representative enriched metabolism-related KEGG pathways are given in Table 3. GSEA, Gene Set Enrichment Analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; TCGA, The Cancer Genome Atlas.

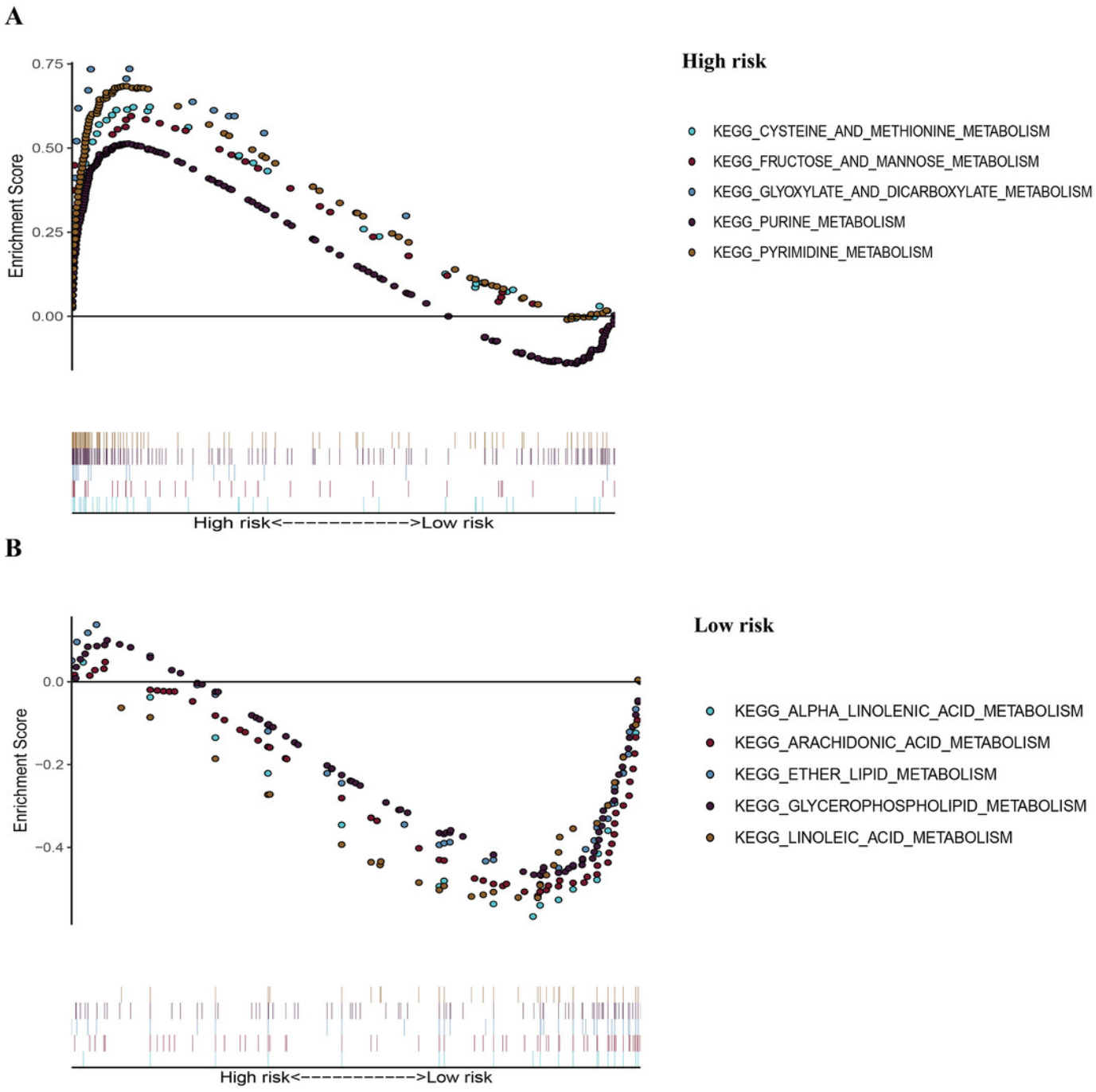

A
0.75
0.50
0.25
0.00
Enrichment Score
High risk
KEGG_CYSTEINE_AND_METHIONINE_METABOLISM
KEGG_FRUCTOSE_AND_MANNOSE_METABOLISM
KEGG_GLYOXYLATE_AND_DICARBOXYLATE_METABOLISM
KEGG_PURINE_METABOLISM
KEGG_PYRIMIDINE_METABOLISM
High risk<----------->Low risk
B
0.0
−0.2
−0.4
Enrichment Score
Low risk
KEGG_ALPHA_LINOLENIC_ACID_METABOLISM
KEGG_ARACHIDONIC_ACID_METABOLISM
KEGG_ETHER_LIPID_METABOLISM
KEGG_GLYCEROPHOSPHOLIPID_METABOLISM
KEGG_LINOLEIC_ACID_METABOLISM
High risk<----------->Low risk

Figure 8

mRNA expression levels of the six prognostic genes from online databases.

(A) mRNA expression levels of the six genes in the Oncomine database (http://www.oncomine.org/). The threshold is shown at the bottom (*P* value < 0.001 and fold change > 2 were utilized for screening). The figure in the colored cell represents the number of datasets complying with the threshold. The red cells indicate that the genes were overexpressed in the cancer, while the blue cells indicate that the genes were overexpressed in the normal tissues. (B) Comparisons of the mRNA expression levels of the six genes between LUAD and normal tissues in the combined LUAD datasets from the Oncomine database. *PFKP*, phosphofructokinase platelet; *PKM*, pyruvate kinase muscle; *TPI1*, triosephosphate isomerase 1; *LDHA*, lactate dehydrogenase A; *PTGES*, prostaglandin E synthase; *TYMS*, thymidylate synthase; LUAD, lung adenocarcinoma.

A

Analysis Type by Cancer	PFKP Cancer vs. Normal		PKM Cancer vs. Normal		TPI1 Cancer vs. Normal		LDHA Cancer vs. Normal		PTGES Cancer vs. Normal		TYMS Cancer vs. Normal	
Bladder Cancer	1		2		3		1				4	
Brain and CNS Cancer	1	4					2	1			2	1
Breast Cancer	1		3	1	2			2		1	14	
Cervical Cancer									1		3	
Colorectal Cancer			5						3		2	
Esophageal Cancer				1		1			1		3	
Gastric Cancer			1		1		2				3	
Head and Neck Cancer	1						1		1		4	
Kidney Cancer	10		9		4	1	9				5	
Leukemia	11			3	1			4		3	1	4
Liver Cancer	1		3								4	
Lung Cancer	9		4		5		4		5		12	
Lymphoma	5		1		9		7			2	5	
Melanoma			1								3	
Myeloma		2	1				2					
Other Cancer		5	2		5		1			1	7	2
Ovarian Cancer	2		6		2		3				1	
Pancreatic Cancer	3		3		1		3		1		2	1
Prostate Cancer							1	1				
Sarcoma	1	1	5								12	
Significant Unique Analyses	45	12	46	5	33	2	36	8	12	7	86	8
Total Unique Analyses	453		410		393		461		428		458	

1 5 10 % 10 5 1

THRESHOLD (P-VALUE: 0.001; FOLD CHANGE: 2; GENE RANK: Top 10%)

Cell color is determined by the best gene rank percentile for the analysis within the cell.
NOTE: An analysis may be counted in more than one cancer type.

B

Comparison of PFKP Across 5 Analyses
Over-expression

Median Rank	p-Value	Gene
280.0	1.64E-17	PFKP

1 2 3 4 5

Legend

1. Lung Adenocarcinoma vs. Normal *Garber Lung, Proc Natl Acad Sci U S A, 2001*
2. Lung Adenocarcinoma vs. Normal *Hou Lung, PLoS One, 2010*
3. Lung Adenocarcinoma vs. Normal *Landi Lung, PLoS ONE, 2008*
4. Lung Adenocarcinoma vs. Normal *Selamat Lung, Genome Res, 2012*
5. Lung Adenocarcinoma vs. Normal *Stearman Lung, Am J Pathol, 2005*

Comparison of LDHA Across 2 Analyses
Over-expression

Median Rank	p-Value	Gene
85.5	3.15E-4	LDHA

1 2

Legend

1. Lung Adenocarcinoma vs. Normal *Selamat Lung, Genome Res, 2012*
2. Lung Adenocarcinoma vs. Normal *Yamagata Lung, Clin Cancer Res, 2003*

Comparison of PTGES Across 4 Analyses
Over-expression

Median Rank	p-Value	Gene
253.0	5.62E-7	PTGES

1 2 3 4

Legend

1. Lung Adenocarcinoma vs. Normal *Garber Lung, Proc Natl Acad Sci U S A, 2001*
2. Lung Adenocarcinoma vs. Normal *Hou Lung, PLoS One, 2010*
3. Lung Adenocarcinoma vs. Normal *Selamat Lung, Genome Res, 2012*
4. Lung Adenocarcinoma vs. Normal *Stearman Lung, Am J Pathol, 2005*

1 5 10 25 % 25 10 5 1 Not measured

Comparison of TYMS Across 6 Analyses
Over-expression

Median Rank	p-Value	Gene
150.0	1.09E-7	TYMS

1 2 3 4 5 6

Legend

1. Lung Adenocarcinoma vs. Normal *Beer Lung, Nat Med, 2002*
2. Lung Adenocarcinoma vs. Normal *Hou Lung, PLoS One, 2010*
3. Lung Adenocarcinoma vs. Normal *Landi Lung, PLoS ONE, 2008*
4. Lung Adenocarcinoma vs. Normal *Selamat Lung, Genome Res, 2012*
5. Lung Adenocarcinoma vs. Normal *Stearman Lung, Am J Pathol, 2005*
6. Lung Adenocarcinoma vs. Normal *Su Lung, BMC Genomics, 2007*

The rank for a gene is the median rank for that gene across each of the analyses.
The p-Value for a gene is its p-Value for the median-ranked analysis.

Figure 9

mRNA expression levels of the six prognostic genes extracted from online database.

The mRNA expression levels of the six genes in different tumour types from the TIMER database (http://cistrome.shinyapps.io/timer/) (*P < 0.05, **P < 0.01, ***P < 0.001). *PFKP*, phosphofructokinase platelet; *PKM*, pyruvate kinase muscle; *TPI1*, triosephosphate isomerase 1; *LDHA*, lactate dehydrogenase A; *PTGES*, prostaglandin E synthase; *TYMS*, thymidylate synthase.

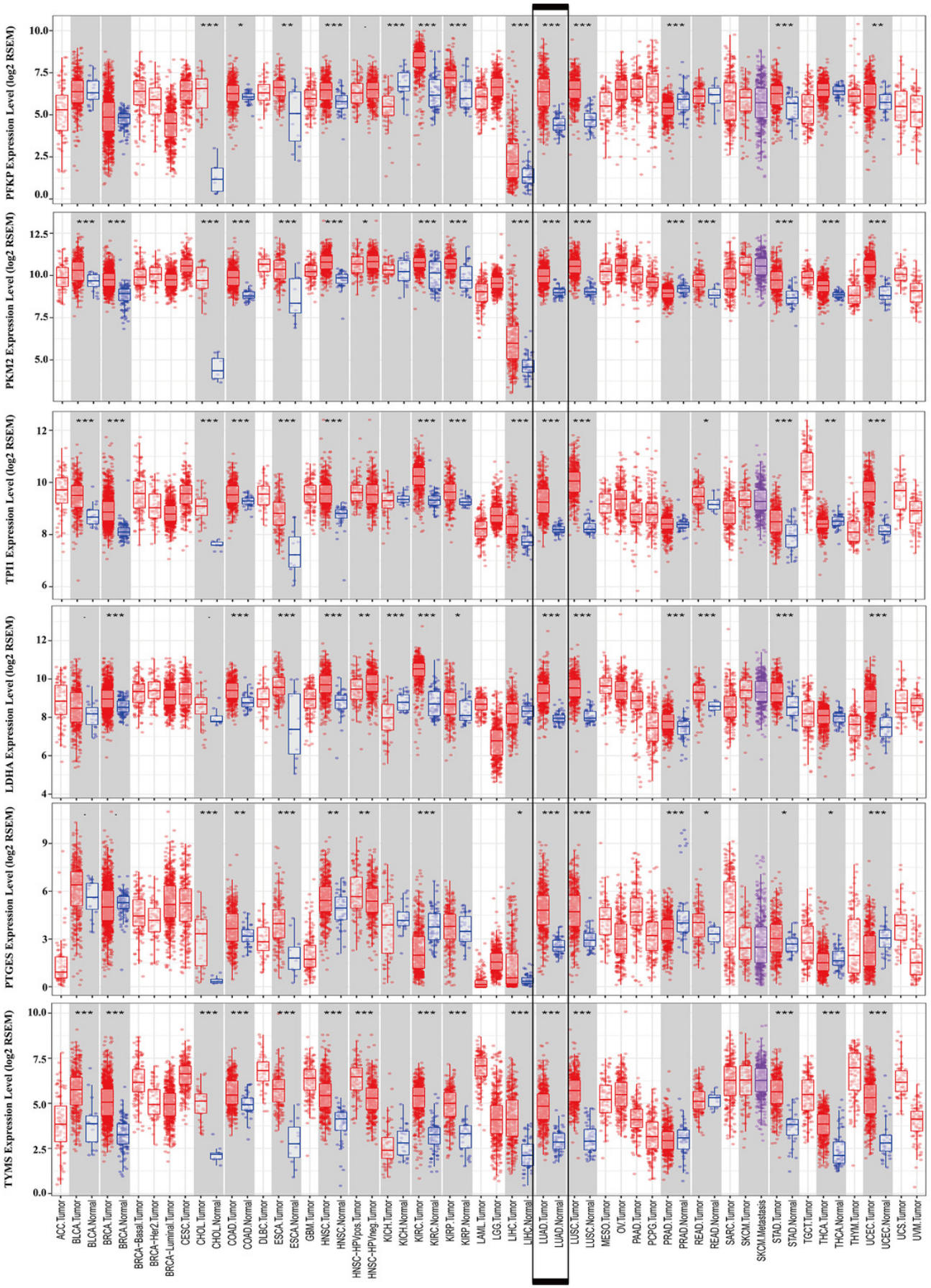
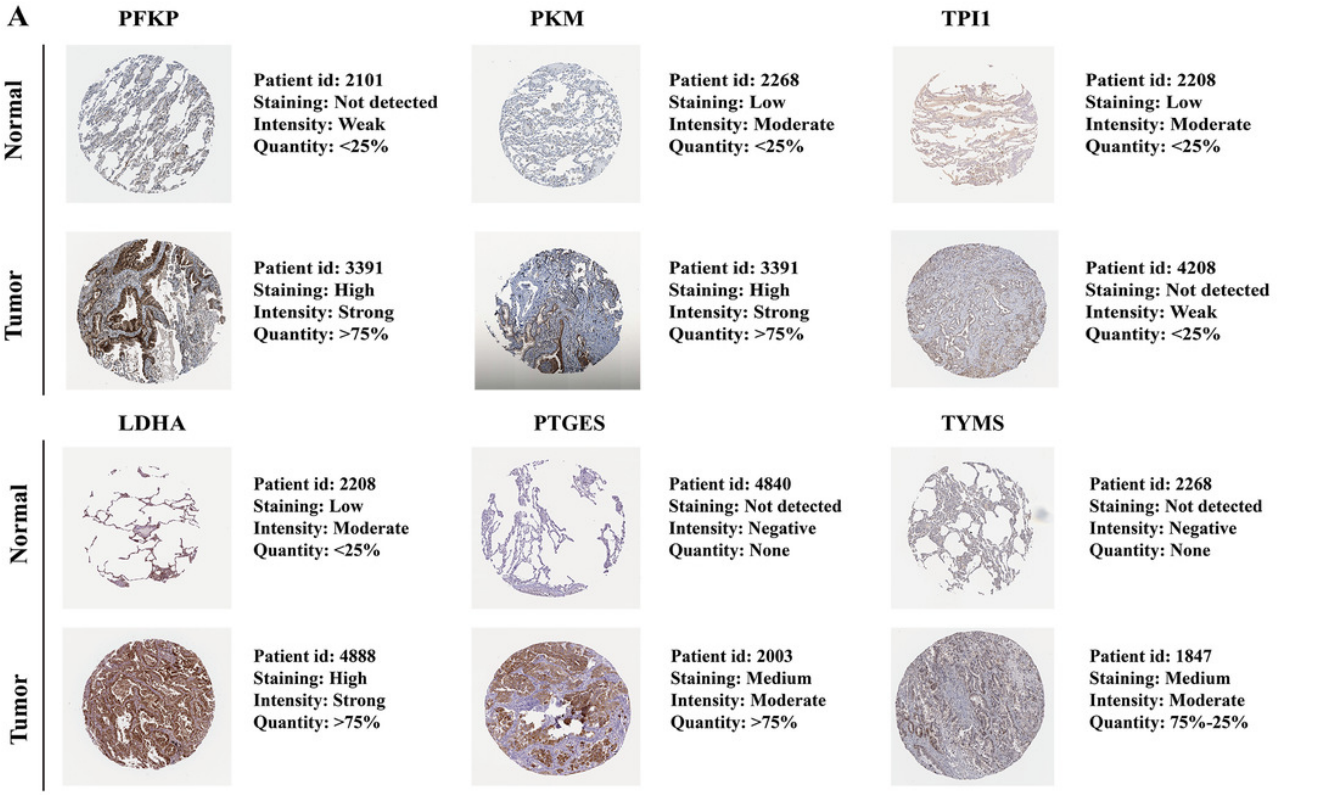
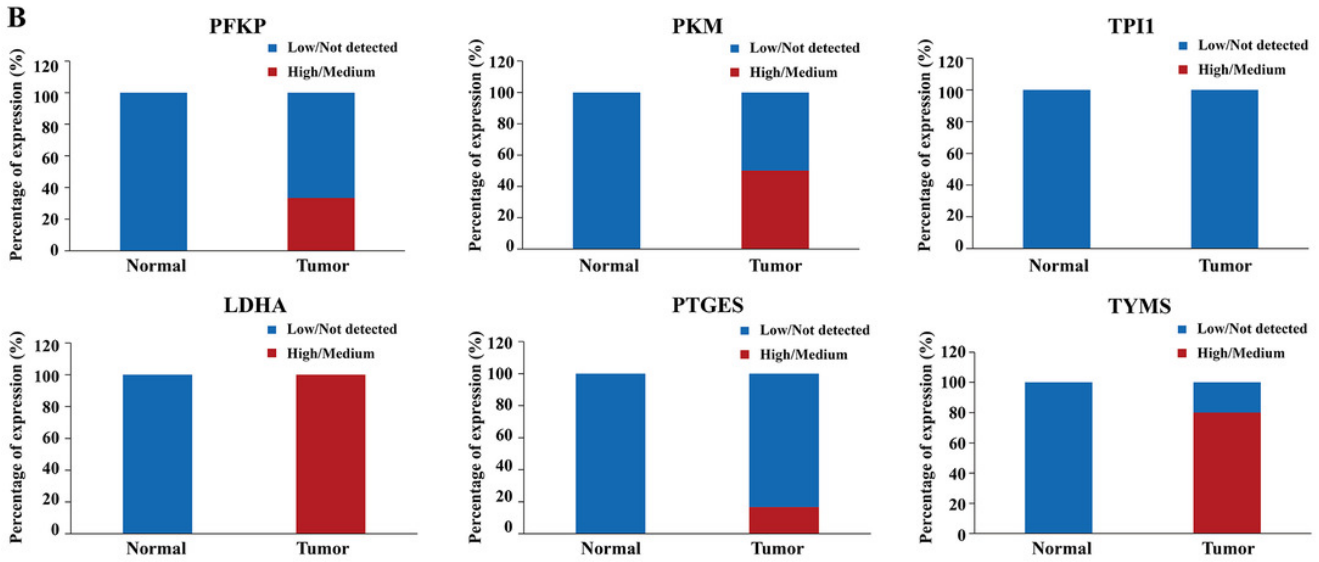

PFKP Expression Level (log2 RSEM)
PKM2 Expression Level (log2 RSEM)
TPI1 Expression Level (log2 RSEM)
LDHA Expression Level (log2 RSEM)
PTGES Expression Level (log2 RSEM)
TYMS Expression Level (log2 RSEM)
ACC.Tumor
BLCA.Tumor
BLCA.Normal
BRCA.Tumor
BRCA.Normal
BRCA-Basal.Tumor
BRCA-Her2.Tumor
BRCA-Luminal.Tumor
CESC.Tumor
CHOL.Tumor
CHOL.Normal
COAD.Tumor
COAD.Normal
DLBC.Tumor
ESCA.Tumor
ESCA.Normal
GBM.Tumor
HNSC.Tumor
HNSC.Normal
HNSC-HPVpos.Tumor
HNSC-HPVneg.Tumor
KICH.Tumor
KICH.Normal
KIRC.Tumor
KIRC.Normal
KIRP.Tumor
KIRP.Normal
LAML.Tumor
LGG.Tumor
LIHC.Tumor
LIHC.Normal
LUAD.Tumor
LUAD.Normal
LUSC.Tumor
LUSC.Normal
MESO.Tumor
OV.Tumor
PAAD.Tumor
PCPG.Tumor
PRAD.Tumor
PRAD.Normal
READ.Tumor
READ.Normal
SARC.Tumor
SKCM.Tumor
SKCM.Metastasis
STAD.Tumor
STAD.Normal
TGCT.Tumor
THCA.Tumor
THCA.Normal
THYM.Tumor
UCEC.Tumor
UCEC.Normal
UCS.Tumor
UVM.Tumor

Figure 10

Protein expression levels and genetic alterations of the corresponding six prognostic genes obtained from online databases.

(A) The representative immunohistochemistry images of the protein expression of the six genes in the normal lung tissues and LUAD tissues from the Human Protein Atlas database. (http://www.proteinatlas.org/) (B) The percentage of protein expression levels in the normal lung tissues and LUAD tissues analysed based on the Human Protein Atlas database. Anti-PFKP antibody is HPA018257; anti-PKM antibody is CAB019421; anti-TPI1 antibody is HPA053568; anti-LDHA antibody is CAB069404; anti-PTGES is HPA045064; anti-TYMS antibody is CAB002784. (C) Genetic alterations of the six genes in 230 LUAD patients / samples (TCGA, Firehose Legacy). Data were obtained from the cBioportal for Cancer Genomics (http://www.cbioportal.org/). *PFKP*, phosphofructokinase platelet; *PKM*, pyruvate kinase muscle; *TPI1*, triosephosphate isomerase 1; *LDHA*, lactate dehydrogenase A; *PTGES*, prostaglandin E synthase; *TYMS*, thymidylate synthase; TCGA, The Cancer Genome Atlas; LUAD, lung adenocarcinoma.


A
PFKP
PKM
TPI1
Normal
Tumor
Patient id: 2101
Staining: Not detected
Intensity: Weak
Quantity: <25%
Patient id: 2268
Staining: Low
Intensity: Moderate
Quantity: <25%
Patient id: 2208
Staining: Low
Intensity: Moderate
Quantity: <25%
Patient id: 3391
Staining: High
Intensity: Strong
Quantity: >75%
Patient id: 3391
Staining: High
Intensity: Strong
Quantity: >75%
Patient id: 4208
Staining: Not detected
Intensity: Weak
Quantity: <25%
LDHA
PTGES
TYMS
Patient id: 2208
Staining: Low
Intensity: Moderate
Quantity: <25%
Patient id: 4840
Staining: Not detected
Intensity: Negative
Quantity: None
Patient id: 2268
Staining: Not detected
Intensity: Negative
Quantity: None
Patient id: 4888
Staining: High
Intensity: Strong
Quantity: >75%
Patient id: 2003
Staining: Medium
Intensity: Moderate
Quantity: >75%
Patient id: 1847
Staining: Medium
Intensity: Moderate
Quantity: 75%-25%


B
PFKP
PKM
TPI1
LDHA
PTGES
TYMS
Low/Not detected
High/Medium
Percentage of expression (%)
0
20
40
60
80
100
120
Normal
Tumor

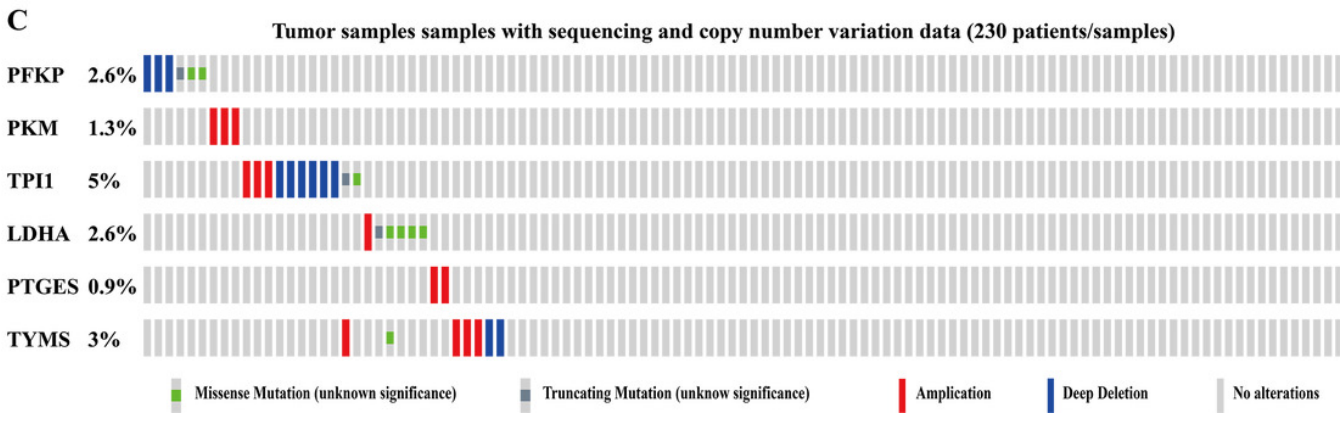

C
Tumor samples samples with sequencing and copy number variation data (230 patients/samples)
PFKP 2.6%
PKM 1.3%
TPI1 5%
LDHA 2.6%
PTGES 0.9%
TYMS 3%
Missense Mutation (unknown significance)
Truncating Mutation (unknow significance)
Amplication
Deep Deletion
No alterations

Table 1(on next page)

Clinical characteristics of the included datasets

1

Characteristics	TCGA (n, %) (n=454)	GSE68465 (n, %) (n=439)
Age		
＜60	133 (29.3%)	128 (29.2%)
≥60	321 (70.7%)	311 (70.8%)
NA	0 (0.0%)	0 (0.0%)
Gender		
Female	248 (54.6%)	218 (49.7%)
Male	206 (45.4%)	221 (50.3%)
NA	0 (0.0%)	0 (0.0%)
Grade		
G1	0 (0.0%)	60 (13.7%)
G2	0 (0.0%)	206 (46.9%)
G3	0 (0.0%)	166 (37.8%)
NA	454 (100%)	7 (1.6%)
TNM stage		
Ⅰ	243 (53.5%)	
Ⅱ	105 (23.1%)	
Ⅲ	74 (16.3%)	
Ⅳ	24 (5.3%)	
NA	8 (1.8%)	439（100%）
T stage		
T1	156 (34.4%)	150 (34.2%)
T2	240 (52.9%)	248 (56.5%)
T3	37 (8.1%)	28 (6.4%)
T4	18 (4.0%)	11 (2.5%)
Tx	3 (0.7%)	2 (0.5%)
N stage		
N0	291 (64.1%)	297 (67.7%)
N1	86 (18.9%)	87 (19.8%)
N2	64 (14.1%)	52 (11.8%)
N3	2 (0.4%)	0 (0.0%)
Nx	11 (2.4%)	3 (0.7%)
M stage		
M0	305 (67.2%)	439 (100%)
M1	23 (5.1%)	0 (0.0%)
Mx	126 (27.8%)	0 (0.0%)
Survival status		
Alive	300 (66.1%)	206 (46.9%)
Dead	154 (33.9%)	233 (53.1%)

Note:

TCGA, The Cancer Genome Atlas.

Table 2(on next page)

Prognostic values for the six-gene metabolic signature in 454 LUAD patients.

Gene	Coef	HR	HR.95L	HR.95H	*P* value
PFKP	0.000050	1.009949	1.004184	1.015747	0.0007
PKM	0.001734	1.005149	1.002545	1.007759	0.000104
TPI1	0.000384	1.003352	1.001479	1.005229	0.000448
LDHA	0.003792	1.005663	1.003774	1.007556	0.00000000396
PTGES	0.002922	1.008392	1.003408	1.0134	0.000946
TYMS	0.024904	1.033141	1.01572	1.050861	0.000172

Note:

LUAD, lung adenocarcinoma; HR, hazard ratio; CI, confidence interval.

Table 3(on next page)

The results of the ten representative enriched metabolism-related KEGG pathways analysed by GSEA.

Pathway	Size	ES	NES	NOM *p*-val	FDR *q*-val
High risk					
KEGG_CYSTEINE_AND_METHIONINE_ METABOLISM	34	0.62	1.98	0.00	0.006
KEGG_FRUCTOSE_AND_MANNOSE_ METABOLISM	33	0.60	1.95	0.002	0.007
KEGG_GLYOXYLATE_AND_DICARBOXYLATE_ METABOLISM	16	0.74	1.89	0.002	0.013
KEGG_PURINE_METABOLISM	157	0.51	2.02	0.000	0.005
KEGG_PYRIMIDINE_METABOLOSM	98	0.68	2.36	0.000	0.000
Low risk					
KEGG_ALPHA_LINOLENIC_ACID_METABOLISM	19	-0.60	-1.82	0.002	0.113
KEGG_ARACHIDONIC_ACID_METABOLISM	58	-0.53	-1.86	0.000	0.101
KEGG_ETHER_LIPID_ METABOLISM	33	-0.51	-1.73	0.011	0.129
KEGG_GLYCEROPHOSPHOLIPID_METABOLISM	77	-0.48	-1.91	0.002	0.127
KEGG_LINOLEIC_ACID_ METABOLISM	29	-0.55	-1.77	0.008	0.138

Note:

KEGG, Kyoto Encyclopedia of Genes and Genomes; GSEA, Gene Set Enrichment Analysis; ES, enrichment score; NOM *p*-val, nominal *p*-value; FDR *q*-val, false discovery rate *q*-value; NES, normalized enrichment score.

Table 4(on next page)

Comparison of mRNA expression levels of the six genes between LUAD and normal tissues from the Oncomine database.

Gene	Analysis Type of lung cancer vs. normal	t-Test	Fold change	*P* value	References
PFKP	LUAD (n=40) vs. Normal (n=6)	7.146	3.469	7.56E-8	Garber et al., 2001
	LUAD (n=58) vs. Normal (n=49)	11.177	2.685	4.35E-19	Landi et al., 2008
	LUAD (n=45) vs. Normal (n=65)	7.946	2.536	3.39E-11	Hou et al., 2010
	LUAD (n=58) vs. Normal (n=58)	10.910	2.883	1.64E-17	Selamat et al., 2012
	LUAD (n=20) vs. Normal (n=19)	5.277	3.149	9.37E-6	Stearman et al., 2005
	Comparison of PFKP expression across 5 Analysis between LUAD and Normal	-	-	1.64E-17	-
PKM	LUAD (n=58) vs. Normal (n=58)	12.037	2.551	3.56E-20	Selamat et al., 2012
TPI1	LUAD (n=40) vs. Normal (n=6)	4.929	2.283	4.03E-4	Garber et al., 2001
LDHA	LUAD (n=9) vs. Normal (n=3)	4.502	4.037	6.29E-4	Yamagata et al. 2003
	LUAD (n=58) vs. Normal (n=58)	11.533	2.179	1.59E-19	Selamat et al., 2012
	Comparison of LDHA expression across 2 Analysis between LUAD and Normal	-	-	3.15E-4	-
PTGES	LUAD (n=20) vs. Normal (n=19)	9.332	5.883	1.54E-11	Stearman et al., 2005
	LUAD (n=40) vs. Normal (n=6)	6.690	4.969	1.12E-6	Garber et al., 2001
	LUAD (n=58) vs. Normal (n=58)	10.267	2.179	5.58E-16	Selamat et al., 2012
	LUAD (n=45) vs. Normal (n=65)	6.513	2.170	6.22E-9	Hou et al., 2010
	Comparison of PTGES expression across 4 Analysis between LUAD and Normal	-	-	5.62E-7	-
TYMS	LUAD (n=45) vs. Normal (n=65)	9.322	3.929	6.92E-15	Hou et al., 2010
	LUAD (n=27) vs. Normal (n=30)	7.395	3.016	2.40E-9	Su et al., 2007
	LUAD (n=58) vs. Normal (n=49)	11.169	2.797	9.86E-20	Landi et al., 2008
	LUAD (n=20) vs. Normal (n=19)	6.509	2.118	2.18E-7	Stearman et al., 2005
	LUAD (n=86) vs. Normal (n=10)	4.191	2.158	3.05E-4	Beer et al., 2002
	LUAD (n=58) vs. Normal (n=58)	8.565	2.040	3.35E-13	Selamat et al., 2012
	Comparison of TYMS expression across 6 Analysis between LUAD and Normal	-	-	1.09E-7	-

Note:

Owing to only one dataset meeting the screening criteria, the comparison of PKM or TPI1 expression in LUAD and normal has not been built based on the combined LUAD datasets. *P* value < 0.001 and fold change > 2 were utilized for screening. LUAD, lung adenocarcinoma.

Table 5(on next page)

Protein expression levels of the six prognostic genes in the normal lung tissues and LUAD tissues obtained from the Human Protein Atlas database.

Gene name	Tissue type	Patients in high/medium staining n (%)	Patients in low/not detected staining n (%)
PFKP	Normal	0 (0%)	3 (100%)
	Tumor	2 (33.33%)	4 (66.67%)
PKM	Normal	0 (0%)	3 (100%)
	Tumor	3 (50%)	3 (50%)
TPI1	Normal	0 (0%)	3 (100%)
	Tumor	0 (0%)	3 (100%)
LDHA	Normal	0 (0%)	3 (100%)
	Tumor	7 (100%)	0 (0%)
PTGES	Normal	0 (0%)	3 (100%)
	Tumor	1 (16.67%)	5 (83.33%)
TYMS	Normal	0 (0%)	3 (100%)
	Tumor	4 (80%)	1 (20%)

Note:

LUAD, lung adenocarcinoma.