

Prognostic Implication of a Metabolism-Associated Gene Signature in Lung Adenocarcinoma

Lulu He,¹ Jiaxian Chen,¹ Feng Xu,² Jun Li,³ and Jun Li^{1,4}

¹State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310003, China; ²Key Laboratory of Respiratory Disease of Zhejiang Province, Department of Respiratory and Critical Care Medicine, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310003, China; ³Department of Pathology, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310003, China; ⁴Precision Medicine Center of Taizhou Central Hospital, Taizhou University Medical School, Taizhou 318000, China

Lung cancer is the most common cancer worldwide, leading to high mortality each year. Metabolic pathways play a vital role in the initiation and progression of lung cancer. We aimed to establish a prognostic prediction model for lung adenocarcinoma (LUAD) patients based on a metabolism-associated gene (MTG) signature. Differentially expressed (DE)-MTGs were screened from The Cancer Genome Atlas (TCGA) LUAD cohorts. Univariate Cox regression analysis was performed on these DE-MTGs to identify genes significantly correlated with prognosis. Least absolute shrinkage and selection operator (LASSO) regression was performed on the resulting genes to establish an optimal risk model. Survival analysis was used to assess the prognostic ability of the model. The prognostic value of the gene signature was further validated in independent Gene Expression Omnibus (GEO) datasets. A gene signature with 13 metabolic genes was identified as an independent prognostic factor. Kaplan-Meier survival analysis demonstrated the good performance of the risk model in both TCGA training and GEO validation cohorts. Finally, a nomogram incorporating clinical parameters and the metabolic gene signature was constructed to help individualize outcome predictions. The calibration curves showed excellent agreement between the actual and predicted survival.

INTRODUCTION

Lung cancer is the most prevalent cancer in the world. Non-small cell lung cancer (NSCLC), including adenocarcinoma, squamous cell carcinoma, bronchioloalveolar carcinoma, and large cell carcinoma, comprises 85% of primary lung cancer cases. It is estimated that 1.4 million patients die of lung cancer each year, the mortality of which is among the highest of all types of cancers.¹ The major reason for the high mortality is that lung cancer is diagnosed at advanced stages in the majority of patients. Traditional prognostic approaches, such as histopathological diagnosis and tumor staging systems, have limited usefulness, and early detection continues to be an elusive goal. With the development of biomedical technology in recent years, we have obtained a better understanding of tumor biology.

Metabolism refers to the ordered chemical reactions that occur in an organism to sustain life. Metabolic pathways show evolutionarily conserved features in cells. Past studies have revealed that the integration of metabolic pathways with diverse signal transduction pathways plays a central role in many disorders.² Otto Warburg first proposed the concept of metabolic reprogramming in solid tumors in 1924.³ With advances in modern science, we have gone further in the research area of the intersections between metabolism and tumor biology. Transformed cells adapt their metabolism to support the biological process of neoplasia. Specific metabolic pathways directly participate in transformation and tumor progression. The blocking of these pathways or restoration of altered metabolic pathways has been proven to be promising therapeutic target strategies.⁴ The phosphatidylinositol 3-kinase (PI3K)/protein kinase B (AKT)/mammalian target of rapamycin (mTOR) and mitogen-activated protein kinase kinase (MEK)/extracellular-regulated protein kinase (ERK)/5' AMP-activated protein kinase (AMPK) pathways participate in the metabolic reprogramming of lung cancer, affecting endogenous fatty acid metabolism, glycolysis, and the tricarboxylic acid cycle and activating the proliferation, invasion, and metastasis of lung cancer as a result.^{5,6} Metabolic reprogramming is becoming an important feature of tumor biology.

Metabolism is a complicated process involving multiple genes. A model based on multiple metabolism-related genes should perform better than a single gene in prognosis prediction. With the development of large-scale genome sequencing technologies, the integration of prognosis-associated gene signatures and traditional clinical parameters show advantages in improving the accuracy of early diagnosis for cancers. In this study, we screened prognosis-related metabolic genes from The Cancer Genome Atlas (TCGA) lung

Received 10 April 2020; accepted 30 September 2020;
<https://doi.org/10.1016/j.omto.2020.09.011>.

Correspondence: Jun Li, MD, PhD, State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, Zhejiang University School of Medicine, 79 Qingchun Rd., Hangzhou 310003, China.
E-mail: lijun2009@zju.edu.cn



adenocarcinoma (LUAD) cohorts, the most common histological subtype of primary lung cancer. A metabolism-related multiple gene signature was established, and its prognostic value was validated in LUAD patients. We also constructed a nomogram based on the integration of the metabolic gene signature and clinical characteristics to predict individual overall survival (OS). In conclusion, our work might contribute to the early diagnosis of LUAD patients.

RESULTS

Identification of Differentially Expressed (DE)-Metabolism-Associated Genes (MTGs) and Functional Analysis

We conducted our study as illustrated in the flow chart (Figure 1). TCGA cohorts consisted of 499 LUAD patients with survival statistics (Table S1). The Kaplan-Meier (K-M) survival curves and log-rank test for clinicopathological parameters, including tumour (T), node (N), metastasis (M), and overall stage, are shown in Figure S1. A total of 1,857 MTGs with a relevance score >8 was identified from the GeneCards website, 86 of which were DE in TCGA cohorts, as shown in the volcano plot (Figure 2A). The intersection of DE genes (DEGs) and MTGs is visualized in a Venn diagram (Figure 2B).

These 86 DE-MTGs were selected for functional analysis, including Gene Ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis. In the KEGG pathway analysis, the DE-MTGs were mainly enriched in malaria, the peroxisome proliferator-activated receptor (PPAR) signaling pathway, and the advanced glycation endproduct (AGE)-receptor for AGE (RAGE) signaling pathway in diabetic complications (Figure S2A). Gene Ontology revealed that in the cellular component category, the DE-MTGs were mainly enriched in the apical plasma membrane, apical part of the cell, and membrane raft (Figure S2B). In biological processes, the DE-MTGs were mainly enriched in transcytosis, regulation of monocyte extravasation, and positive regulation of dendritic cell differentiation (Figure S2C). Regarding molecular function, the DE-MTGs were mainly enriched in amide binding, peptide binding, and amyloid-beta binding (Figure S2D).

Construction of a Prognostic Gene Signature in TCGA Training Cohorts

Univariate Cox regression analysis identified 37 DE-MTGs significantly associated with OS (Figure 2C; risk genes are in red and protective genes in green). The differential expression of these genes between normal subjects and LUAD patients is visualized in Figures 2D and 2E. These significant genes were subjected to least absolute shrinkage and selection operator (LASSO) Cox regression analysis to construct the prognostic model. The calculation of the regression coefficient is visualized in Figure 3A. The prognostic model performed best when 13 genes were included (Figure 3B). The functions of the genes in the prognostic gene signature involved glycometabolism, lipid metabolism, and vitamin metabolism (Table 1).

We checked the genetic alterations of the 13 genes in TCGA cohorts via the cBioPortal for Cancer Genomics (<http://www.cbioportal.org/>) website. These genes were altered in 119 (23%) of 507 patients in the PanCancer Atlas for the LUAD dataset (Figure 3C). The Firehose

Legacy for the LUAD dataset also showed that 54 (23%) of the 230 queried patients had a mutation in these genes (Figure S3A). In the larger dataset of Nat Genet 2016 for NSCLC, which contains 1,144 patients, these genes were altered in 304 (27%) patients (Figure S3B). The frequent mutations, to some extent, indicated the contribution of these genes to the development of tumors. We also checked the genetic alterations of the prognostic genes in six other different cancer types, including breast-invasive carcinoma (BRCA), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), pancreatic adenocarcinoma (PAAD), skin cutaneous melanoma (SKCM), stomach adenocarcinoma (STAD), and liver hepatocellular carcinoma (LIHC). These genes also showed frequent mutations in the other cancer types: 179 (18%) of the 996 BRCA patients, 43 (15%) of the 278 CESE patients, and 62 (18%) of the 353 LIHC patients had a mutation in these genes; 16 (9%) of the 175 PAAD patients, 187 (52%) of the 363 SKCM patients, and 117 (27%) of the 434 STAD patients showed gene alterations (Figure S4).

A prognostic risk score for each patient was calculated based on the mRNA expression levels of the 13 genes and the coefficients from LASSO Cox regression analysis. We performed univariate and multivariate Cox regression analyses to evaluate the prognostic value of the risk score. Univariate Cox regression analysis revealed that the risk score ($p < 0.001$, hazard ratio [HR] = 4.035, 95% confidence interval [CI] = 2.828–5.759) and clinicopathological parameters, including T stage ($p < 0.001$, HR = 1.544, 95%CI = 1.281–1.862), N stage ($p < 0.001$, HR = 1.744, 95%CI = 1.457–2.088), M stage ($p = 0.024$, HR = 1.917, 95%CI = 1.088–3.380), and overall stage ($p < 0.001$, HR = 1.611, 95%CI = 1.396–1.859), were significantly associated with OS in TCGA LUAD cohorts. Multivariate Cox regression analysis proved the risk score to be an independent prognostic variable ($p < 0.001$, HR = 3.639, 95%CI = 2.510–5.277).

The distribution of the risk scores and the correlation between the risk scores and survival data are illustrated in scatterplots (Figure 4A). The patients were divided into a low-risk group and high-risk group, according to the median value of the risk scores in TCGA LUAD cohorts. The gene-expression profiles of the prognostic risk genes between the high-risk group and low-risk group are displayed in the heatmap in Figure 4B. K-M survival analysis revealed a significantly higher survival probability in the low-risk group ($p < 0.0001$) (Figure 5A). The area under the receiver operating characteristic (ROC) curve (AUC) of the risk scores for the survival probability at 1, 2, 3, 4, and 5 years is displayed in Figure 5B. The maximum AUC value reached 0.72, which indicated good sensitivity and specificity.

Evaluation of the Prognostic Gene Signature in Independent Gene Expression Omnibus (GEO) Validation Cohorts

To validate the prognostic value of the risk score, the GEO validation cohorts were divided into high- and low-risk groups, according to the same cut-off value of TCGA cohorts. The distribution of the risk scores and the correlation between the risk scores and survival data of Okayama's cohort are illustrated in Figure 4C. The same

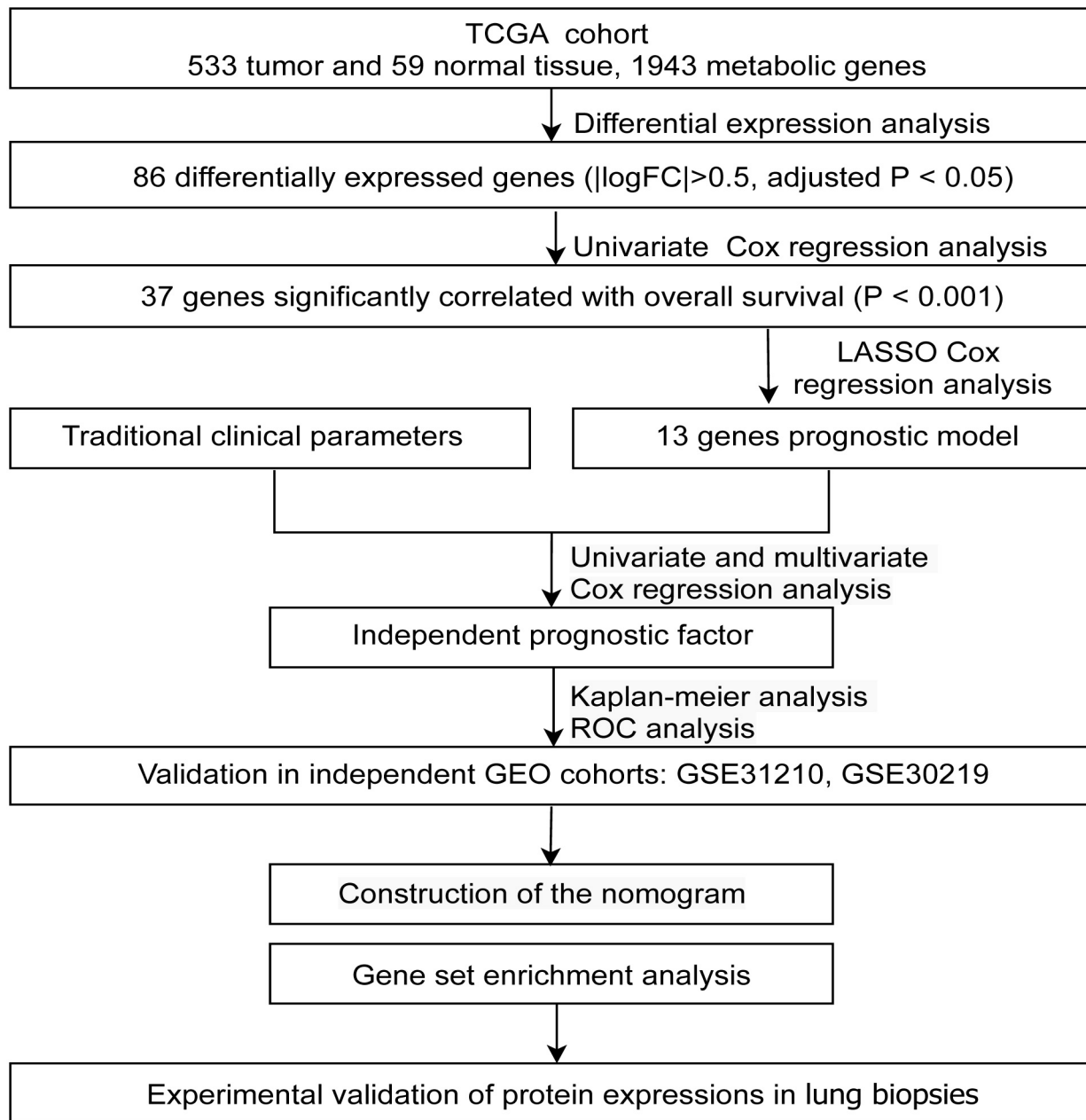
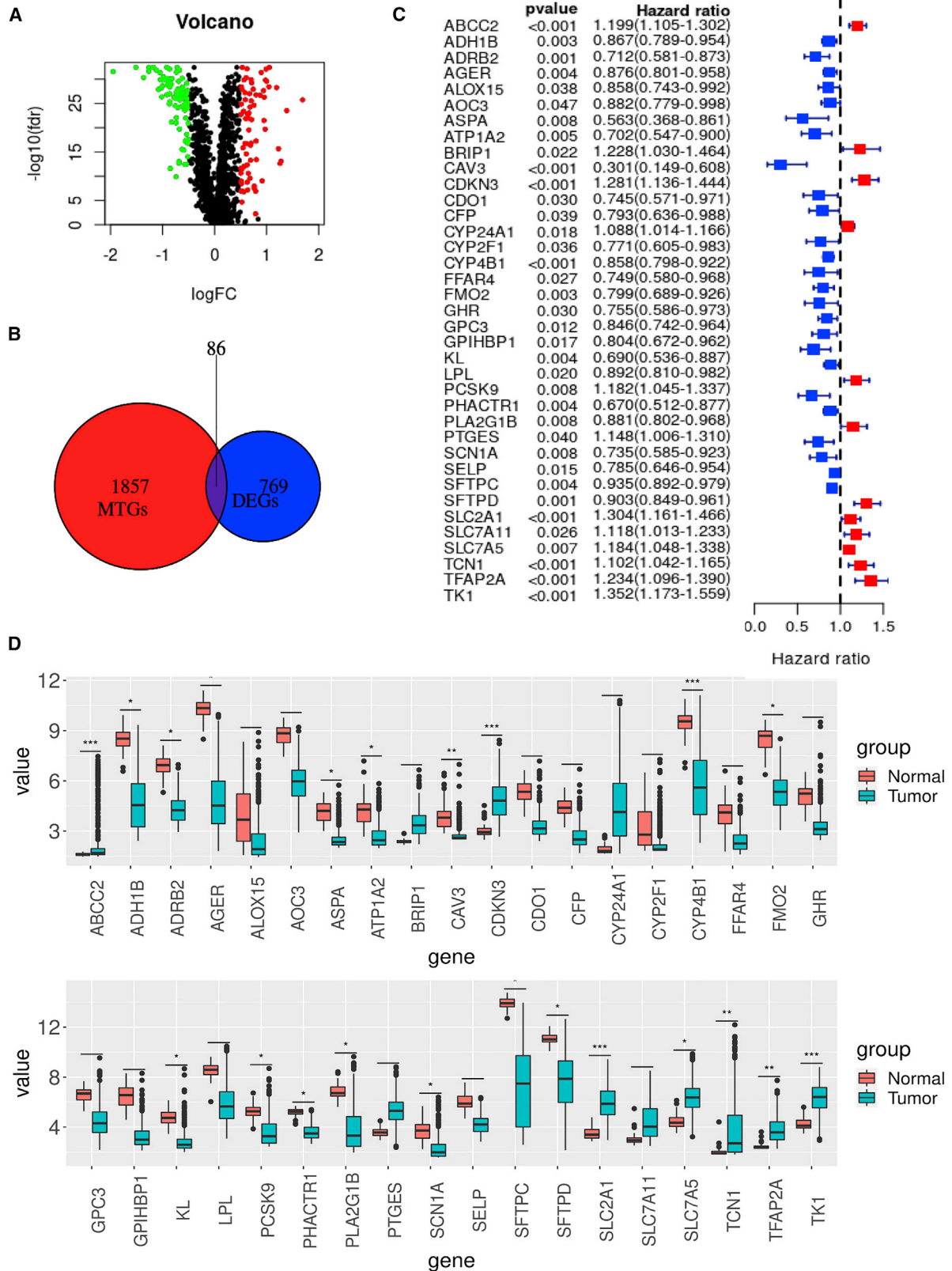


Figure 1. The Flow Chart Summarizes the Scheme Performed to Construct Prognostic Gene Signatures of Lung Adenocarcinoma (LUAD)

scatterplots of Rousseaux’s cohort are displayed in Figure 4E. The gene-expression profiles of the two cohorts are visualized in Figures 4D and 5F. Okayama’s cohort included 226 patients with pathological stage I–II LUAD. The K-M survival curves revealed a higher survival probability of the low-risk group ($p = 0.0011$) in this validation cohort (Figure 5C), the maximum AUC of which reached 0.83 (Figure 5D). Rousseaux’s cohort consisted of 292 patients with pathological stage I–IV lung cancer, including 71 adenocarcinomas and other different kinds of lung cancers. The prognostic gene signature also performed

well in this mixed lung cancer validation cohort ($p < 0.0001$, maximum AUC = 0.64) (Figures 5E and 5F). In general, the 13 metabolic gene-based prognostic signature was proven to be valuable in risk stratification.

Similar approaches were used to evaluate the specificity of this prognostic gene signature in six different types of cancer, including BRCA, CESC, PAAD, SKCM, STAD, and LIHC, in TCGA cohorts. Among these cancers, the low-risk groups in the LIHC ($p = 0.046$) and



(legend on next page)

PAAD ($p = 0.011$) cohorts were consistently associated with higher survival probability using the K-M analysis (Figure S5).

We predicted transcription factor (TF) for our gene signature via the ChEA3 (<https://amp.pharm.mssm.edu/chea3/>) website. The top 10 TFs were listed in Table S2. Forkhead box M1 (*FOXM1*), followed by tumor protein p73 (*TP73*), and cyclic AMP (cAMP)-responsive element-binding protein 3-like 4 (*CREB3L4*) were predicted to be the most related. The protein-protein interaction (PPI) network was constructed and visualized by STRING (Figure S6A).

Construction of the Nomogram

The nomogram is an efficient tool that integrates multiple risk factors for clinical application. We established a nomogram for the prediction of 3-year and 5-year OS in TCGA LUAD cohorts. Seven independent risk factors, including age, sex, stage, T stage, M stage, N stage, and metabolism signature, were included in the model (Figure 6A). The points of the factors indicate their corresponding contribution to the survival probability. The total points of each patient provided the estimated 3-year and 5-year survival times. The C-index of our nomogram was 0.702 (95%CI = 0.679–0.724). The actual OS and nomogram-predicted OS matched well at 3 years and 5 years, as shown by the calibration curves (Figures 6B and 6C).

Exploration of Signaling Pathways

Gene set enrichment analysis (GSEA) has an advantage in exploring the involved signaling pathways from an overall perspective. GSEA revealed that the genes in the high-risk group of TCGA cohorts were significantly enriched in the cell cycle (normalized enrichment score [NES] = 2.43, $p < 0.01$), P53 signaling pathway (NES = 2.26, $p < 0.01$), pyrimidine metabolism (NES = 2.24, $p < 0.01$), and proteasome (NES = 2.24, $p < 0.01$). In contrast, the low-risk group genes were significantly enriched in pathways, such as vascular smooth muscle contraction (NES = -1.84 , $p = 0.008$), aldosterone-regulated sodium reabsorption (NES = -1.84 , $p = 0.008$), asthma (NES = -1.82 , $p = 0.036$), and primary bile acid biosynthesis (NES = -1.67 , $p = 0.020$) (Figure 6D).

Immunohistochemistry Staining of Representative Prognostic Genes

To validate the protein expression of the prognostic gene signature, we performed immunohistochemical analysis on the top 4 genes of Table 1 using lung biopsies. The characters of patients enrolled were illustrated in Table S3. Solute carrier family 2-facilitated glucose transporter member 1 protein (*SLC2A1*) expression was increased in NSCLC and SCLC lung biopsies. PCSK9 and ABCC2 were weakly positive in lung biopsies of lung cancer. Klotho (KL) was not detected in NSCLC or SCLC lung tissues (Figure 7).

Analysis of Gene Expression at the Single-Cell Level

Thienpont and coworkers⁷ identify 52 different stromal cell subclusters within seven major cell types in the lung tumor microenvironment. To further verify the expression of these gene signatures in single cells, we applied single-cell RNA sequencing (RNA-seq) datasets for lung cancer. As Figure S6B illustrated, genes like *SLC2A1*, cytochrome P450 4 (CYP4) B1 (*CYP4B1*), *TFAP2A*, transcobalamin 1 (*TCN1*), *CDKN3*, and *TK1* were expressed in most cell types, whereas *FFAR4* mostly expressed in myeloid cells. *CAV3* only expressed in one cell subcluster within fibroblast cells. Genes like *CYP2F1* and *SCN1A* were expressed at lower levels in specific cell types.

DISCUSSION

Gene signatures based on specific cell activities, such as the cell cycle,⁸ autophagy,⁹ and immune signature,¹⁰ show good advantage in prognosis prediction in cancers. Given the importance of metabolic processes and the superiority of multiple gene-based models than single genes, it is reasonable to expect that metabolism-related prognostic models would perform well in cancers.⁴ To our knowledge, prognostic gene signatures based on MTGs have not yet been reported in LUAD. In this study, we found 86 DE-MTGs from TCGA LUAD cohorts, 37 of which were significantly associated with survival probability in LUAD patients. Our study focused on the altered metabolic genes but was not limited to glycolysis-related genes, as in formal studies.^{11,12} Then, we established a prognostic signature with 13 metabolic genes and constructed a new nomogram that integrated the metabolic signature and clinical parameters. Our gene signature efficiently stratified patient outcomes in the LUAD cohorts and was validated in independent datasets. Recently, Liu et al.¹³ constructed an MTG prognostic model in hepatocellular carcinoma. The combination of Liu et al.'s¹³ MTG risk score and clinical parameters outperforms the traditional tumour, node and metastasis (TNM) staging system. However, there are few common genes between our gene signature and Liu et al.'s,¹³ suggesting that the metabolic contribution-related genes are tumor specific.

Lung cancers, including LUAD, are heterogeneous, and the prognosis varies even in patients with the same pathological stage: some fortunately get over the disease, but some are afflicted by recurrence. There are still aspects of the disease that the staging system cannot explain. The elucidation of the molecular method helps to uncover the underlying mechanisms and predict outcomes. The identification of the MTG signature shows clinical implications, as it is significantly related to the outcomes of LUAD patients. Gene-targeted therapy is a novel treatment that is effective in some lung cancer patients with gene mutations, but it is expensive.¹⁴ Compared to traditional treatments, patients at high risk might benefit from innovative treatments, such as DNA- and RNA-based

Figure 2. Identification of Differentially Expressed (DE)-MTGs and Selection of MTGs Associated with the Survival of TCGA LUAD Patients

(A and B) Volcano plot (A) and Venn diagram (B) of the 86 DE-MTGs in the LUAD cohorts of TCGA database. (C) Forest plot of the univariate Cox regression analysis with MTGs. (D and E) Expression levels of survival-related genes in tumor and normal tissues. * $p < 0.01$, ** $p < 0.001$ and *** $p < 0.0001$. Nineteen genes are shown in D panel and eighteen genes in E panel.

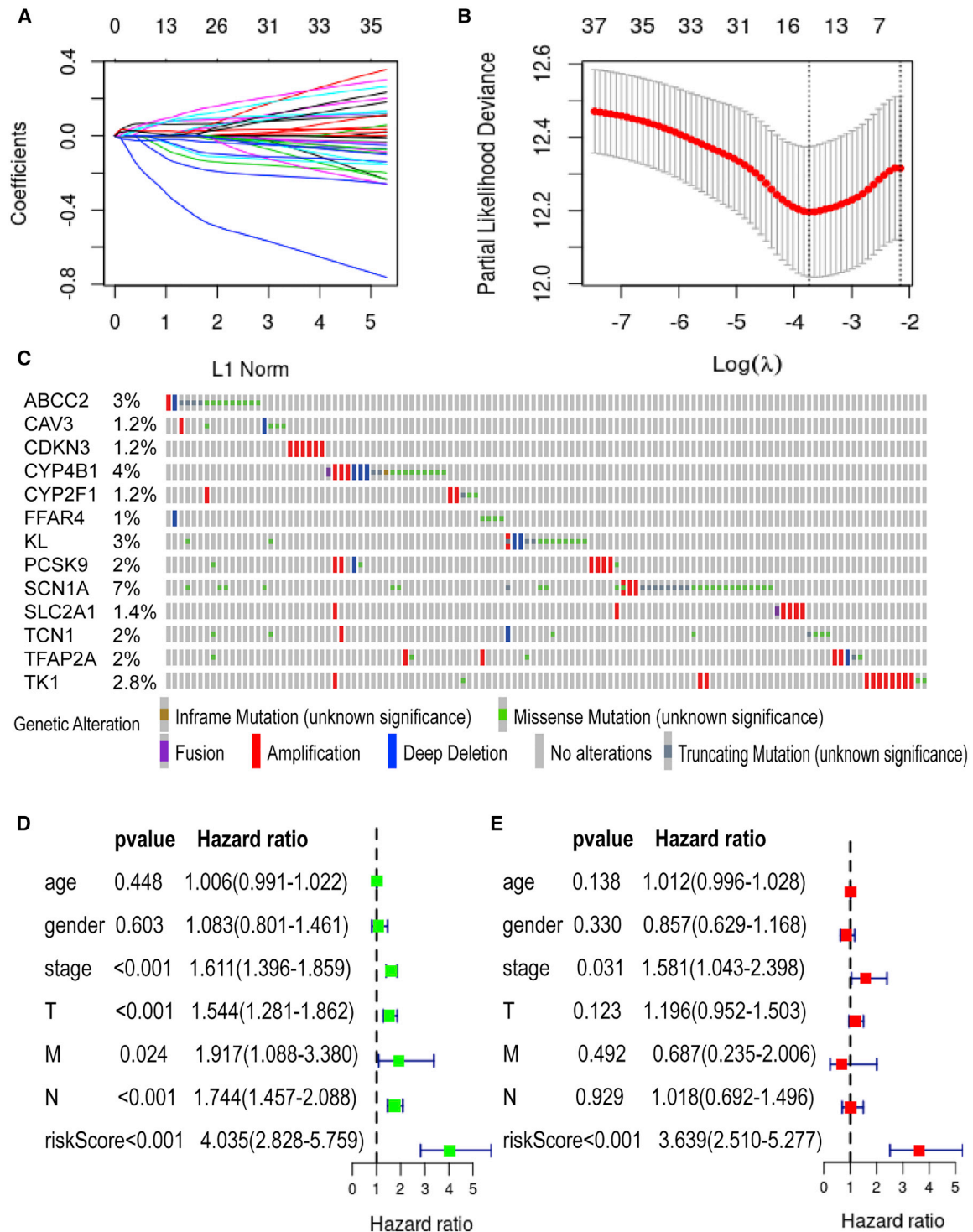


Figure 3. The Construction of Metabolic Gene Signature and the Evaluation of Its Independent Prognostic Value

(A) LASSO coefficient profiles of the 37 survival-related genes. (B) A coefficient profile plot was produced against the log (λ) sequence in the LASSO model. The optimal parameter (λ) was selected as the first black dotted line indicated. (C) Genetic alteration profiles of the prognostic genes in TCGA LUAD RNA-seq dataset (TCGA, PanCancer Atlas). (D and E) Forest plots of the (D) univariate and (E) multivariate Cox regression analyses in TCGA LUAD cohorts.

Table 1. Functions of the Genes in the Prognostic Gene Signature

Number	Gene Symbol	Full Name	Function	Risk Coefficient	Relevance Score
1	SLC2A1	solute carrier family 2-facilitated glucose transporter member 1	glucose metabolism-related gene	0.02516	34.97
2	PCSK9	proprotein convertase subtilisin/kexin type 9	plays a role in cholesterol and fatty acid metabolism	0.06971	33.16
3	KL	klotho	participates in the carbohydrate metabolic process	-0.02761	24.39
4	ABCC2	ATP-binding cassette subfamily C member 2	promotes biliary metabolism	0.06233	21.08
5	CAV3	caveolin 3	glycometabolism-related gene	-0.29684	18.03
6	TCN1	transcobalamin 1	vitamin metabolism-related gene	0.07042	14.52
7	CDKN3	cyclin-dependent kinase inhibitor 3	diabetes-related gene	0.07383	13.63
8	FFAR4	free fatty acid receptor 4	participates in insulin sensitizing	-0.03365	11.56
9	CYP2F1	cytochrome P450 family 2 subfamily F member 1	catalyzes reactions involved in drug metabolism and synthesis of lipids	-0.06792	11.49
10	SCN1A	sodium voltage-gated channel alpha subunit 1	metabolism-related gene	-0.11240	11.4
11	CYP4B1	cytochrome P450 family 4 subfamily B member 1	catalyzes reactions involved in drug metabolism and synthesis of lipids	-0.02138	10.93
12	TK1	thymidine kinase 1	metabolism-related gene	0.00644	9.53
13	TFAP2A	transcription factor AP-2 alpha	diabetes-related gene	0.09059	8.53

therapeutics, whereas those with low-risk gene signatures could temporarily postpone undergoing those methods. The prognostic model could contribute to patient classification, support personalized therapeutic strategies in clinical practice for LUAD patients, and ultimately contribute to reducing mortality.

Despite the limitation of retrospective research, our MTG signature was validated in two independent cohorts. Our gene signature efficiently stratified patient outcomes even in the mixed cohorts consisting of different kinds of lung cancer. Thus, we believe that the results we obtained using the MTG signature are reliable. Most of the genes in our signature have been previously reported to be involved in cancers previously. For example, *TCN1* regulates the homeostasis of vitamin B12 as the binding protein. Elevated *TCN1* expression has been reported in breast cancer and hepatocellular carcinoma.¹⁵ *SLC2A1*, also called glucose transporter type 1 (*GLUT1*), is overexpressed in cancers, such as gastric, liver, and lung cancers. The expression of *SLC2A1* is significantly related to the histopathological grades. Given the role of *SLC2A1* in glucose utilization, it is supposed that overexpressed *SLC2A1* meets the great needs of energy, contributes to the acidification of the tumor microenvironment, and ultimately promotes the growth and metastasis of tumors.¹⁶ Consistent with our results, *KL* acts as a suppressive gene in cancers. Downregulated *KL* expression has been reported in several types of cancers, including breast, gastric, bladder, and lung cancer.¹⁷ Enzymes in the CYP4 family have been implicated in the metabolism of drugs, fatty acids, and signaling molecules. With the consideration of the role of CYP4 enzymes in the maintenance of fatty acid homeostasis, the enzymes are suggested to be involved in the process of carcinogenesis. As a member of the CYP4 superfamily of enzymes, *CYP4B1* is involved in tumor angiogenesis.¹⁸ Metabolism is a complex process in which thousands of genes are considered to participate. However, there

has been great expansion of our knowledge of the role of metabolic pathways in carcinogenesis in recent years. Overall, our metabolic gene signature might indicate a dysregulated metabolic microenvironment and reveal targets for the development of therapy in lung cancer.

Interestingly, functional analysis revealed that the intersection of metabolic genes and DEGs was mainly associated with malaria. Lung cancer and malaria seem to be diseases of distinct patterns. However, there are hidden connections between cancer and human malaria, which involve a multiplicity of metabolic routes.¹⁹ The infection of malaria is associated with the highly mutated gene in cancers—P53.²⁰ GSEA of the high- and low-risk groups in LUAD indicates the involvement of different pathways. Consistent with the clinical outcome, the high-risk group was mainly enriched in KEGG pathways significantly related to cancer, including the cell-cycle and P53 signaling pathway, and further discerned the participation of pyrimidine metabolism, the proteasome, homologous recombination, and mismatch repair. Consistent with Liu et al's¹³ results, the low-risk group was more associated with metabolic pathways, such as aldosterone-regulated sodium reabsorption and bile acid biosynthesis. Therefore, we speculate that the competition between carcinogenic factors and body resistibility results in disturbed metabolic microenvironments. The low-risk group represents the early period of compensation and might benefit more from metabolism-related treatment than the high-risk group. These hypotheses need further investigation.

Moreover, we established a nomogram for clinical-decision support. A nomogram incorporates assumed risk factors, calculates the proportion of each factor, constructs and visualizes the statistical predictive model, and ultimately generates a numerical possibility for

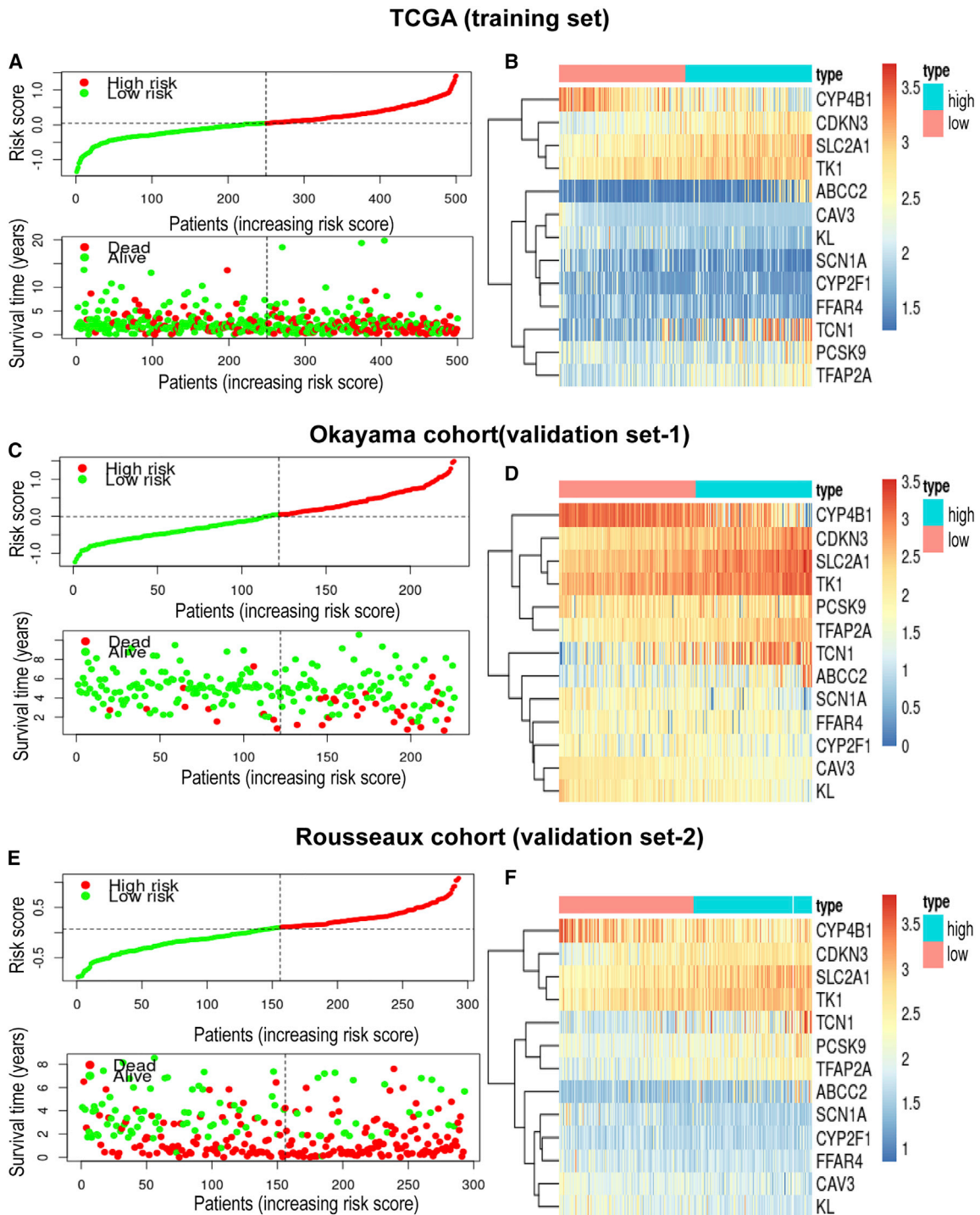


Figure 4. Characteristics of the Risk Score and Heatmap of the Metabolic Gene Signature

(A, C, and E) The distributions of the risk score, survival time, and status of patients in TCGA training cohorts (A), GEO validation set 1, and (C) GEO validation set 2 (E). The dotted lines indicate the optimal cut-off value between the low- and high-risk groups. (B, D, and F) Heatmap of the gene-expression profiles of the metabolic gene signature in TCGA training cohorts (B), GEO validation set 1 (D), and GEO validation set 2 (F).

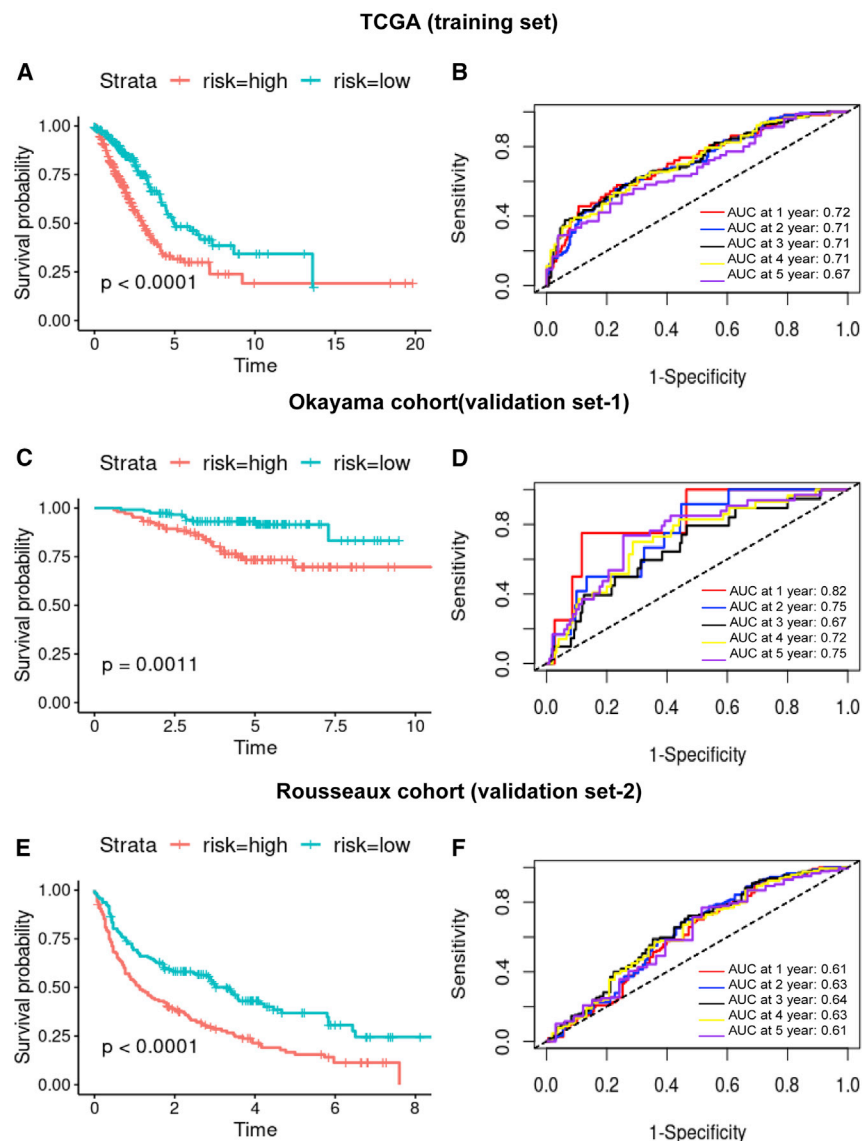


Figure 5. Kaplan-Meier and Time-Dependent ROC Analysis of the Prognostic Gene Signature

(A, C, and E) Kaplan-Meier curves of the gene signature in TCGA training cohorts (A), GEO validation set 1 (C), and GEO validation set 2 (E). (B, D, and F) The time-dependent ROC curves of the prognostic gene signature in TCGA training cohorts (B), GEO validation set 1 (D), and GEO validation set 2 (F).

Despite the underlying clinical significance of our results, several limitations need to be considered. First, the clinical features extracted from TCGA and GEO databases are limited and incomplete. Potential prognostic factors, such as personal history, treatment, and background diseases, are missing in our nomogram. It is not clear how the environmental factors, including smoking, exposure to certain toxins, or treatments such as chemotherapy, radiotherapy, and targeted drug therapy, affect the identified gene signatures. Second, our metabolic gene signature and derived cut-off value were constructed based on RNA-seq data. The procedure of sample treatment, RNA extraction, reverse transcription, and detection needs to be standardized. Third, independent prospective cohorts are needed to verify the prognostic model developed in this study. The value of those genes as potential pharmacological targets also needs further investigations.

In summary, we identified a reliable prognostic MTG signature based on TCGA database. Our gene signature indicates a dysregulated metabolic microenvironment and might suggest therapeutic targets for LUAD patients. A nomogram based on the MTG signature and clinical parameters could accurately predict the 3- and 5-year survival probability of individual LUAD patients. Our findings might favor personalized therapeutic strategies.

individual clinical outcomes.²¹ Owing to their intuitive visual presentation and personalized application, nomograms have become a popular tool for oncology prognosis. Liu et al.⁹ reported a nomogram to predict 3- and 5-year OS in NSCLC with the incorporation of risk scores derived from an autophagy-related gene signature. The combination of clinicopathological features and the autophagy-related gene signature performed better than each alone. Long and colleagues²² developed a nomogram that included a TP53-associated immune prognostic signature for hepatocellular carcinoma. The adoption of an immune prognostic signature and prognostic factors, including hepatitis C infection and vascular tumor invasion, exhibits high prognostic accuracy, as demonstrated by ROC curves and calibration curves. Consistently, our nomogram could well predict the 3- and 5-year survival probabilities for LUAD patients by incorporating the MTG signature and prognostic factors.

5-year survival probability of individual LUAD patients. Our findings might favor personalized therapeutic strategies.

MATERIALS AND METHODS

Acquisition of MTGs

MTGs were collected from the GeneCards (<https://www.genecards.org/>) database,²³ which provides comprehensive information on human genes. The term “metabolism” was used as the key word for the search, and genes with relevance scores >8 were taken as MTGs.

Collection of Datasets

The RNA-seq data and clinical characteristics of TCGA LUAD, PAAD, BRCA, SKCM, LIHC, STAD, and CESC cohorts were obtained from TCGA website (<https://portal.gdc.cancer.gov/>) for training. The large-scale genome sequencing was performed before treatment in those patients, as

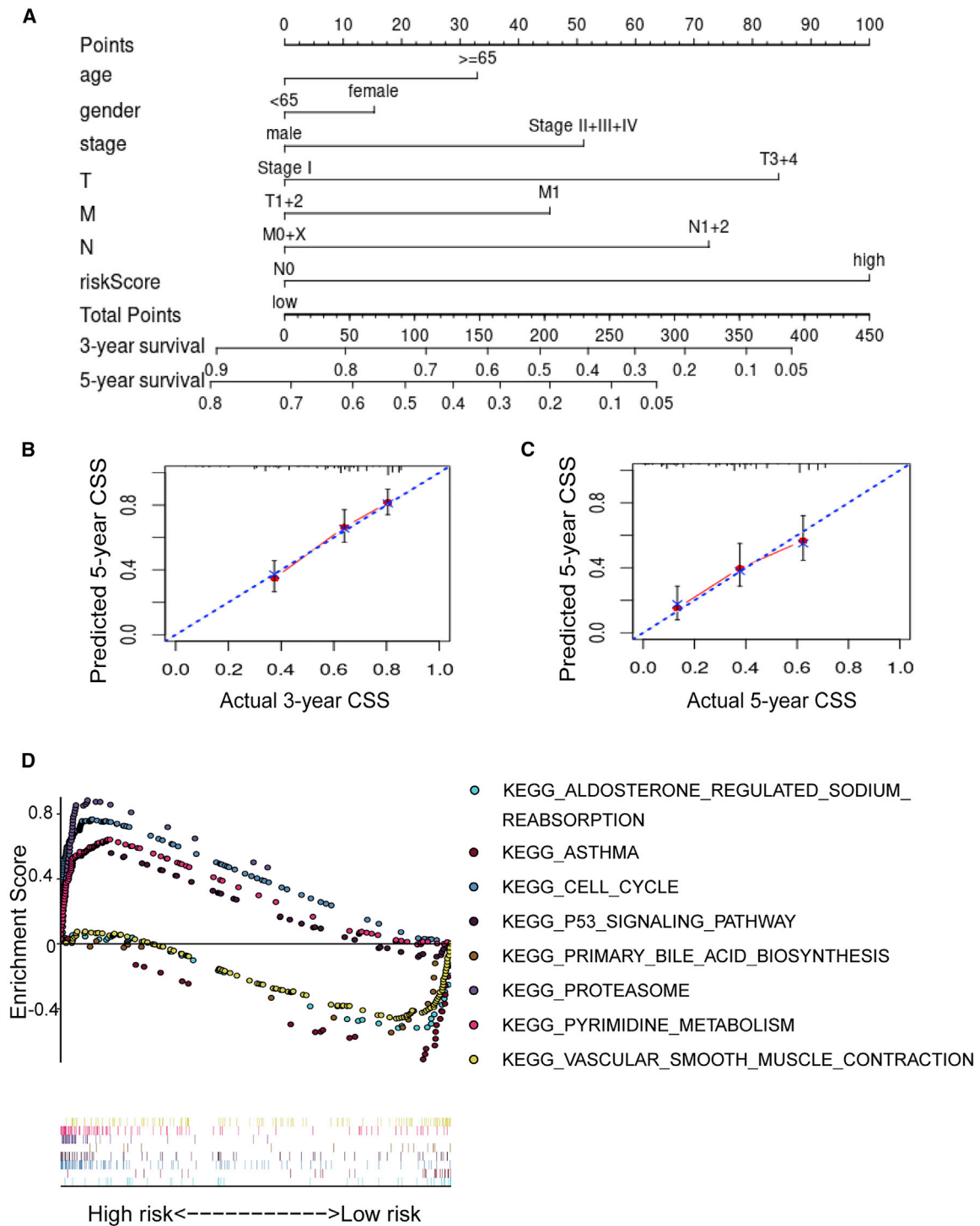


Figure 6. Nomogram Predicting OS for LUAD Patients in TCGA Cohorts

(A) The nomogram was constructed based on seven independent prognostic factors. (B and C) The calibration plots for the internal validation of the nomogram predicting 3-year (B) and 5-year (C) OS. The x axis represents the nomogram actual survival, and the y axis represents the predicted survival. (D) Enrichment plot of the DEGs between the high- and low-risk groups using GSEA.

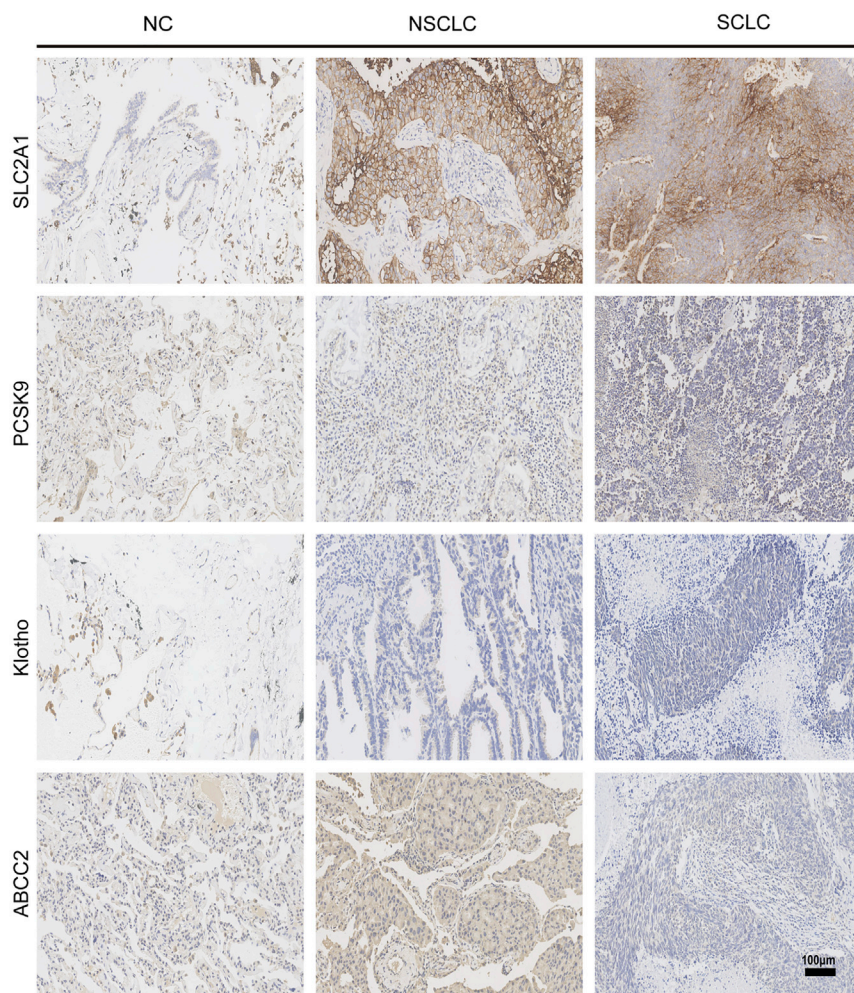


Figure 7. Immunohistochemistry Staining of the Prognostic Genes in Lung Cancer Biopsies

First panel on the left shows the expression of gene signatures in normal controls, whereas the last two panels represent the expression in NSCLC and SCLC lung biopsies, respectively.

cance.^{26,27} The clusterProfiler package in R software was used to perform statistical analysis and to visualize the functional profiles of the DE-MTGs, including Gene Ontology and KEGG analysis.²⁸ adj. p value < 0.05 was considered the cut-off value for significance.

Construction of the Prognostic Gene Signature

Univariate Cox proportional hazards regression analysis was performed on each DE-MTG to screen genes significantly associated with OS in TCGA training set.^{29,30} Then, the LASSO Cox regression method was applied to those identified genes.³¹ A multivariable model with the metabolism-related genes was constructed. Those genes with nonzero coefficients were screened out to calculate the risk score. A prognostic risk score was generated for each patient with the following formula: risk score = expression level of gene1 \times j1 + expression level of gene2 \times j2 + ... + expression level of genex \times jx, where j represents the coefficient. The median risk score was considered the cut-off value to divide TCGA LUAD patients

TCGA focuses on untreated primary cancers.²⁴ The gene-expression profile matrixes of the LUAD cohort GEO: GSE31210 and the mixed lung cancer cohort GEO: GSE30219 were downloaded from the GEO website (<https://www.ncbi.nlm.nih.gov/geo/>) for validation. Log₂ transformation and normalization were employed for the expression profiles. The average expression level was retained for duplicated genes. The sva package (<http://bioconductor.org/packages/release/bioc/html/sva.html>) in R software 3.6.0 (<https://www.r-project.org/>) was used to eliminate batch effects.

DEG Analysis and Functional Analysis

Those MTGs mentioned above were investigated with differential expression analysis in TCGA cohort by the edgeR package (<https://bioconductor.org/packages/release/bioc/>) and visualized as volcano plots. Adjusted p value (adj. p) < 0.05 and |fold change (FC)| > 0.5 were considered statistically significant for identifying DEGs.²⁵ The intersection of the DEGs and MTGs (DE-MTGs), as visualized in a Venn diagram, were selected for further analysis.

Gene functional analysis is a critical step in translating molecular findings from high-throughput methods into biological signifi-

into a high-risk group and a low-risk group. The same formula and same cut-off value were applied to two GEO datasets for validation.

Univariate and multivariate Cox proportional hazards regression analyses were performed to test whether the MTG-based prognostic model was an independent prognostic factor. A K-M survival curve was constructed, and the log-rank test was used to assess the survival differences between groups. The sensitivity and specificity of the prognostic performance were examined by ROC curve analysis. The AUC values indicated discrimination.

Construction and Validation of the Nomogram

A prognostic nomogram was established to assess the survival probability for LUAD patients in 3 or 5 years via the rms R package. Age, sex, pathological stage, pathological T stage, pathological N stage, pathological M stage, and risk score were included as independent parameters. The C-index and calibration curves were used to calculate the discrimination and calibration between the nomogram predicted value and the true survival.³²

GSEA

GSEA was conducted to determine the related pathways and molecular mechanisms of the high-risk group and low-risk group in TCGA cohorts (<https://www.gsea-msigdb.org/gsea/index.jsp>). Gene sets with a p value of <0.05 and false discovery rate (FDR) of <0.25 after 1,000 permutations were considered significantly enriched.³³

Immunohistochemistry Staining

Lung samples were obtained from patients who had been diagnosed with NSCLC or SCLC and went through cancer resection surgery (The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China). The normal controls were correspondingly taken from para-carcinoma tissues. The study conformed to the ethical guidelines of the Declaration of Helsinki and was approved by the Ethics Committee at The First Affiliated Hospital of Zhejiang University. Informed consent was obtained from every participant involved. Samples had been fixed, embedded in paraffin, and processed into thin slices. The prognostic genes SLC2A1 (Santa Cruz;1:100), PCSK9 (Proteintech; 1:500), KL (Santa Cruz; 1:300), and ABCC2 (Proteintech; 1:50) were detected by immunohistochemistry in tissue sections, according to the manufacturer's instructions.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.omto.2020.09.011>.

AUTHOR CONTRIBUTIONS

The study was designed by J.L.(last author). The manuscript was written by J.L.(last author) and L.H. The experiment and data analysis were performed by L.H., J.C., F.X., and J.L.(4th author). All authors were involved in critical revision of the manuscript.

CONFLICTS OF INTEREST

The authors declare no competing interests.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (81830073) and National and Zhejiang Provincial special support program for high-level personnel recruitment (Ten-thousand Talents Program).

REFERENCES

- Torre, L.A., Bray, F., Siegel, R.L., Ferlay, J., Lortet-Tieulent, J., and Jemal, A. (2015). Global cancer statistics, 2012. *CA Cancer J. Clin.* 65, 87–108.
- Levine, A.J., and Puzio-Kuter, A.M. (2010). The control of the metabolic switch in cancers by oncogenes and tumor suppressor genes. *Science* 330, 1340–1344.
- Vander Heiden, M.G., Cantley, L.C., and Thompson, C.B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* 324, 1029–1033.
- Vander Heiden, M.G., and DeBerardinis, R.J. (2017). Understanding the Intersections between Metabolism and Cancer Biology. *Cell* 168, 657–669.
- Chang, L., Fang, S., and Gu, W. (2020). The Molecular Mechanism of Metabolic Remodeling in Lung Cancer. *J. Cancer* 11, 1403–1411.
- Vanhove, K., Graulus, G.J., Mesotten, L., Thomeer, M., Derveaux, E., Noben, J.P., Guedens, W., and Adriaensens, P. (2019). The Metabolic Landscape of Lung Cancer: New Insights in a Disturbed Glucose Metabolism. *Front. Oncol.* 9, 1215.
- Lambrechts, D., Wauters, E., Boeckx, B., Aibar, S., Nittner, D., Burton, O., Bassez, A., Decaluwé, H., Pircher, A., Van den Eynde, K., et al. (2018). Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat. Med.* 24, 1277–1289.
- Zhao, L., Jiang, L., He, L., Wei, Q., Bi, J., Wang, Y., Yu, L., He, M., Zhao, L., and Wei, M. (2019). Identification of a novel cell cycle-related gene signature predicting survival in patients with gastric cancer. *J. Cell. Physiol.* 234, 6350–6360.
- Liu, Y., Wu, L., Ao, H., Zhao, M., Leng, X., Liu, M., Ma, J., and Zhu, J. (2019). Prognostic implications of autophagy-associated gene signatures in non-small cell lung cancer. *Aging (Albany NY)* 11, 11440–11462.
- Song, Q., Shang, J., Yang, Z., Zhang, L., Zhang, C., Chen, J., and Wu, X. (2019). Identification of an immune signature predicting prognosis risk of patients in lung adenocarcinoma. *J. Transl. Med.* 17, 70.
- Liu, C., Li, Y., Wei, M., Zhao, L., Yu, Y., and Li, G. (2019). Identification of a novel glycolysis-related gene signature that can predict the survival of patients with lung adenocarcinoma. *Cell Cycle* 18, 568–579.
- Zhang, L., Zhang, Z., and Yu, Z. (2019). Identification of a novel glycolysis-related gene signature for predicting metastasis and survival in patients with lung adenocarcinoma. *J. Transl. Med.* 17, 423.
- Liu, G.M., Xie, W.X., Zhang, C.Y., and Xu, J.W. (2020). Identification of a four-gene metabolic signature predicting overall survival for hepatocellular carcinoma. *J. Cell. Physiol.* 235, 1624–1636.
- Ramalingam, S.S., Owonikoko, T.K., and Khuri, F.R. (2011). Lung cancer: New biological insights and recent therapeutic advances. *CA Cancer J. Clin.* 61, 91–112.
- Lee, Y.Y., Wei, Y.C., Tian, Y.F., Sun, D.P., Sheu, M.J., Yang, C.C., Lin, L.C., Lin, C.Y., Hsing, C.H., Li, W.S., et al. (2017). Overexpression of Transcobalamin 1 is an Independent Negative Prognosticator in Rectal Cancers Receiving Concurrent Chemoradiotherapy. *J. Cancer* 8, 1330–1337.
- Yan, S., Wang, Y., Chen, M., Li, G., and Fan, J. (2015). Deregulated SLC2A1 Promotes Tumor Cell Proliferation and Metastasis in Gastric Cancer. *Int. J. Mol. Sci.* 16, 16144–16157.
- Pshes-Yeloz, N., Ungar, L., Wohl, A., Jacoby, E., Fisher, T., Leitner, M., Nass, D., Rubinek, T., Wolf, I., and Cohen, Z.R. (2019). Role of Klotho Protein in Tumor Genesis, Cancer Progression, and Prognosis in Patients with High-Grade Glioma. *World Neurosurg.* 130, e324–e332.
- Lim, S., Alshagga, M., Ong, C.E., Chieng, J.Y., and Pan, Y. (2020). Cytochrome P450 4B1 (CYP4B1) as a target in cancer treatment. *Hum. Exp. Toxicol.* 39, 785–796.
- Wein, S., Ghezal, S., Buré, C., Maynadier, M., Périgaud, C., Vial, H.J., Lefebvre-Tournier, I., Wengelink, K., and Cerdan, R. (2018). Contribution of the precursors and interplay of the pathways in the phospholipid metabolism of the malaria parasite. *J. Lipid Res.* 59, 1461–1471.
- Nordor, A.V., Bellet, D., and Siwo, G.H. (2018). Cancer-malaria: hidden connections. *Open Biol.* 8, 180127.
- Iasonos, A., Schrag, D., Raj, G.V., and Panageas, K.S. (2008). How to build and interpret a nomogram for cancer prognosis. *J. Clin. Oncol.* 26, 1364–1370.
- Long, J., Wang, A., Bai, Y., Lin, J., Yang, X., Wang, D., Yang, X., Jiang, Y., and Zhao, H. (2019). Development and validation of a TP53-associated immune prognostic model for hepatocellular carcinoma. *EBioMedicine* 42, 363–374.
- Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T.I., Nudel, R., Lieder, I., Mazor, Y., et al. (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr. Protoc. Bioinformatics* 54, 1.30.1–1.30.33.
- Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D., Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V., et al.; Cancer Genome Atlas Research Network (2018). An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* 173, 400–416.e11.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.

26. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al.; The Gene Ontology Consortium (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29.
27. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45 (D1), D353–D361.
28. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287.
29. Linden, A., and Yarnold, P.R. (2017). Modeling time-to-event (survival) data using classification tree analysis. *J. Eval. Clin. Pract.* 23, 1299–1308.
30. Nagashima, K., and Sato, Y. (2017). Information criteria for Firth's penalized partial likelihood approach in Cox regression models. *Stat. Med.* 36, 3422–3436.
31. Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Stat. Med.* 16, 385–395.
32. Alba, A.C., Agoritsas, T., Walsh, M., Hanna, S., Iorio, A., Devereaux, P.J., McGinn, T., and Guyatt, G. (2017). Discrimination and Calibration of Clinical Prediction Models: Users' Guides to the Medical Literature. *JAMA* 318, 1377–1384.
33. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102, 15545–15550.