

# Establishment and Validation of a Prognostic Signature for Lung Adenocarcinoma Based on Metabolism-Related Genes

**Zhihao Wang**

Huazhong University of Science and Technology Tongji Medical College

**Kidane Siele Embaye**

Huazhong University of Science and Technology Tongji Medical College

**Qing Yang**

Qingdao University of Technology

**Lingzhi Qin**

Huazhong University of Science and Technology Tongji Medical College

**Chao Zhang**

Huazhong University of Science and Technology Tongji Medical College

**Liwei Liu**

Huazhong University of Science and Technology Tongji Medical College

**Xiaoqian Zhan**

Huazhong University of Science and Technology Tongji Medical College

**Fengdi Zhang**

Huazhong University of Science and Technology Tongji Medical College

**Xi Wang**

Huazhong University of Science and Technology Tongji Medical College

**Shenghui Qin** (✉ [2015tj0147@hust.edu.cn](mailto:2015tj0147@hust.edu.cn))

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

<https://orcid.org/0000-0002-0275-133X>

---

## Primary research

**Keywords:** Lung adenocarcinoma, Metabolism-related genes, Prognostic, The Cancer Genome Atlas

**Posted Date:** August 21st, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-62222/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Cancer Cell International on April 15th, 2021. See the published version at <https://doi.org/10.1186/s12935-021-01915-x>.

**Establishment and validation of a prognostic signature for lung adenocarcinoma  
based on metabolism-related genes**

Zhihao Wang<sup>1</sup>, Kidane Siele Embaye<sup>1</sup>, Qing Yang<sup>2</sup>, Lingzhi Qin<sup>1</sup>, Chao Zhang<sup>1</sup>, Liwei Liu<sup>1</sup>, Xiaoqian Zhan<sup>1</sup>, Fengdi Zhang<sup>3</sup>, Xi Wang<sup>1</sup>, Shenghui Qin<sup>1</sup>

<sup>1</sup> Institute of Pathology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China;

<sup>2</sup> Department of Pharmacy, Hiser Medical Center of Qingdao, Qingdao 266033, China;

<sup>3</sup> Department of Pathology, Wuhan Third Hospital (Tongren Hospital of Wuhan University), Wuhan 430030, China;

**Correspondence to :** Shenghui Qin;

**Email:** 2015tj0147@hust.edu.cn;

## **Abstract**

**Background:** Given that dysregulated metabolism has been recently identified as a hallmark of cancer biology, this study aims to establish and validate a prognostic signature of lung adenocarcinoma (LUAD) based on metabolism-related genes (MRGs).

**Methods:** The gene sequencing data of LUAD samples with clinical information and the metabolism-related gene set were obtained from The Cancer Genome Atlas (TCGA) and Molecular Signatures Database (MSigDB), respectively. The differentially expressed MRGs were identified by Wilcoxon rank sum test. Then, univariate Cox regression analysis were performed to identify MRGs that related to overall survival (OS). A prognostic signature was developed by multivariate Cox regression analysis. Furthermore, the signature was validated in the GSE31210 dataset. In addition, a nomogram that combined the prognostic signature was created for predicting the 1-, 3- and 5-year OS of LUAD. The accuracy of the nomogram prediction was evaluated using a calibration plot. Finally, cox regression analysis was applied to identify the prognostic value and clinical relationship of the signature in LUAD.

**Results:** A total of 116 differentially expressed MRGs were detected in the TCGA dataset. We found that 12 MRGs were most significantly associated with OS by using the univariate regression analysis in LUAD. Then, multivariate Cox regression analyses were applied to construct the prognostic signature, which consisted of six MRGs (ALDOA, CAT, ENTPD2, GNPAT1, LDHA, and TYMS). The prognostic value of this signature was further successfully validated in the GSE31210 dataset. Furthermore, the calibration curve of the prognostic nomogram demonstrated good

agreement between the predicted and observed survival rates for each of OS. Further analysis indicated that this signature could be an independent prognostic indicator after adjusting to other clinical factors. Finally, the signature was found to be associated with various clinicopathological features.

**Conclusions:** We established a prognostic signature based on MRGs for LUAD and validated the performance of the model, which may provide a promising tool for the diagnosis and prognosis in patients with LUAD.

**Keywords:** Lung adenocarcinoma, Metabolism-related genes, Prognostic, The Cancer Genome Atlas

## **Background**

Lung cancer is one of the most commonly diagnosed cancer types with high mortality worldwide in men and women[1]. Lung adenocarcinoma(LUAD), which is considered a highly molecular heterogeneous disease, is a prevalent pathological subtype of lung cancer with an average 5-year survival rate of only 15%[2-4]. Molecular targeted therapy for LUAD has been widely accepted in recent years, and the epidermal growth factor receptor (EGFR) gene, the anaplastic lymphoma kinase (ALK) gene, and the Kirsten rat sarcoma viral oncogene (KRAS) gene is an important target of LUAD[5-7]. Despite great clinical improvements in the molecular basis, diagnosis and treatment modalities of LUAD, the recurrence rate still remains high, and the survival rate remains poor[4, 8]. As LUAD has the tendency of early metastasis, and most of them are found at advanced stages, which may be the most important cause of high mortality in LUAD patients[9, 10]. There is an urgent need, therefore, to develop more reliable

and more effective biomarkers for the early detection, diagnosis, prognosis and monitoring of LUAD.

Dysregulated metabolism has been recently identified as a well-recognized hallmark of cancer biology, meeting cancer cell the requirements of rapid proliferation and preferential survival[11, 12]. In the 1920s, Otto Warburg first discovered that cancer cells vigorously take up glucose and convert pyruvate into lactate despite the presence of sufficient oxygen, a phenomenon now widely termed aerobic glycolysis, or the Warburg effect[13, 14]. This phenomenon not only provide a niche for the survival and proliferation of tumor cells, but also has a profound effect on the tumor microenvironment[15]. In addition, it has recently been reported that high concentrations of lactate in the tumor microenvironment are associated with distant metastasis and poor prognosis in a multitude of cancers, including LUAD[16, 17]. There is general agreement that the metabolic processes plays an important role in the pathogenesis and progression of lung cancer. However, few studies have comprehensively analyzed the relationship between metabolism-related genes(MRGs) and the diagnosis, risk stratification, prognosis, and survival of LUAD by high-throughput biomarker sequencing.

In the present study, we constructed a prognostic signature based on MRGs from The Cancer Genome Atlas (TCGA) database, which was further validated in the GSE31210 dataset to explore an efficient metabolic biomaker for the more accurate stratification management of LUAD. In addition, a nomogram that combined the prognostic signature was created for predicting the 1-, 3- and 5-year OS of LUAD. The accuracy

of the nomogram prediction was evaluated using a calibration plot. Finally, cox regression analysis was applied to identify the prognostic value and clinical relationship of the signature in LUAD.

## **Methods**

### **Data collection**

The transcriptomic and the corresponding clinical data of patients with lung adenocarcinoma (LUAD) were downloaded from The Cancer Genome Atlas (TCGA; <https://portal.gdc.cancer.gov/>) database and the Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) database. The RNA-seq data, including 497 LUAD and 54 adjacent non-tumor cases from TCGA database and 174 LUAD cases from GSE31210 dataset were examined. The metabolism-related genes (MRGs) were identified from the metabolic pathway-related gene sets of “c2.cp.kegg.v7.0.symbols” in gene set enrichment analysis (GSEA). MRGs can be further analyzed only when they are included in the above data sets.

### **Differentially expressed MRGs and enrichment analysis**

The differentially expressed metabolism-related genes (MRGs) in LUAD and normal tissues were detected using the R package limma and the Wilcoxon test method [18].  $|\log_{2}FC| > 1$  and adjusted  $P < 0.05$  were considered as significant. To explore the characteristic biological function and potential pathways of these MRGs, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were carried out with R package clusterProfiler [19]. Functional categories with a false discovery rate (FDR) smaller than 0.05 were

presented.

### **Construction of the metabolism-related signature for LUAD**

To avoid the interference of irrelevant factors, patients with follow-up time less than 2000 days and more than 0 day were included. The 426 LUAD samples with survival information in the TCGA dataset were taken as the training set for constructing the prognosis risk model, and the 174 LUAD samples with survival information in the GSE31210 dataset were explored for external validation. Firstly, univariate Cox analysis was used to screen out MRGs associated with the overall survival (OS) of patients with LUAD, and only MRGs with a P value < 0.001 were selected for subsequent analyses. To avoid the prognostic signature overfitting and narrow the genes for prediction of the OS, Lasso Cox regression was carried out using R “glmnet” package. MRGs detected via Lasso algorithm were evaluated by step wise multivariate Cox regression analysis. By weighting the estimated Cox regression coefficients, the model of tumor-related metabolism genes risk was constructed[20]. The prognostic metabolism-related gene signatures were shown as risk score =  $\sum (\beta_i \times \text{Exp}_i)$ , where  $\beta_i$ , the coefficients, represented the weight of the respective signature and  $\text{Exp}_i$  represented the expression value. Based on the risk score formula, patients were assigned into low-risk group and high-risk group with the median risk score as the cut-off point. The Kaplan-Meier (K-M) survival curve was used the log-rank test to evaluate the differences in survival rate between the two groups. Furthermore, the receiver operating characteristic (ROC) curve was implemented by R “survivalROC” package[21] and the corresponding area under the ROC curve (AUC) was measured to assess the sensitivity

and specificity of the metabolism-related signature.

### **Validation of the metabolism-related signature for LUAD**

To verify the prognostic value of metabolism-related signature, we used the GSE31210 dataset as the validation cohort. The same formula was used to calculate the risk scores for each patient. Survival and ROC curve analyses were implemented as described above. Finally, According to the results of multivariate Cox regression analysis, a nomogram for predicting the likelihood of 1-year, 3-year and 5-year OS was constructed by R “rms” package. The calibration plots were used to evaluate the prognostic accuracy of the nomogram.

### **Analysis of these crucial MRGs expression level**

Differential expression of these hub MRGs at the transcription level were examined by matching cancer and adjacent normal tissues from the TCGA database. For further validation of our analysis, The Human Protein Atlas (HPA) online database (<http://www.proteinatlas.org/>) was applied to identify the expression of these MRGs at the translational level[22].

### **Association of the prognostic signature and clinicopathological features**

In addition, Univariate and multivariate analyses were used to estimate the effect of risk scores on overall survival and the clinicopathologic features(age, gender, clinical stage and pathological grading). We also explored the correlation between the expression of these MRGs and several clinical features.

### **Statistical analysis**

All statistical analyses were performed by version 3.6.1 of R software(<https://www.r->

project.org/) and version 5.28.1 of Perl software (<http://www.perl.org>). The Wilcoxon test was used to compare two paired groups. The Kaplan-Meier survival curves were compared with the log-rank test. If not otherwise stated, data were considered to be statistically significant with P value <0.05.

## **Results**

### **Identification of differentially expressed MRGs in LUAD**

According to the Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathway-related gene sets, a total of nine hundred and forty-four MRGs were obtained from the gene sets of “c2.cp.kegg.v7.0.symbols”. We matched these genes with the sequence data of LUAD related mRNA in the TCGA database and GSE31210 dataset, and only common mRNAs were used. Considering the cutoff criteria (adjusted P value <0.05 and  $|\log FC| > 1.0$ ), 116 differentially expressed MRGs, which consist of 31 downregulated and 85 upregulated MRGs(Fig.1), were selected for subsequent analysis.

### **Functional enrichment of the differentially expressed MRGs**

To investigate the potential functional implication of these MRGs, 116 differentially expressed MRGs were further analyzed by GO functional enrichment analysis and KEGG pathway enrichment analysis. A total of 431 GO terms and 42 pathways were identified (adjusted P<0.05).The top 30 enrichment GO analysis and top 30 enrichment KEGG analysis were displayed in Figure 2. The top enriched GO terms in biological processes were carboxylic acid biosynthetic process and organic acid biosynthetic process, and those in cellular components were mitochondrial matrix, ficolin-1-rich granule lumen, and ficolin-1-rich granule, in terms of molecular function, genes were

mostly enriched in terms of cofactor binding. In the KEGG pathway enrichment analysis, these genes were shown to be significantly associated with signaling pathway related to material synthesis and material metabolism, such as “biosynthesis of amino acids”, “arginine and proline metabolism”, “glycolysis/gluconeogenesis”, “carbon metabolism” and et al.

### **Establishment of metabolism-related prognostic signature for LUAD**

To identify MRGs associated with OS, a univariate Cox proportional hazard regression analysis was initially performed on 116 differentially expressed MRGs in the TCGA database. The result showed that 12 MRGs were significantly associated with the OS (Fig.3a;  $P < 0.001$ ). Of the survival-related MRGs, 10 genes (ALDOA, TPI1, PKM, LDHA, GPI, PFKP, RRM2, TYMS, GNPAT1, and ENTPD2) were considered risk factors (all  $P < 0.001$ ; HRs, 1.0026-1.1103) and that their overexpression may reduce survival; overexpression of the remaining two genes (CAT and FBP1) (all  $P < 0.001$ ; HRs, 0.9747 and 0.9907, respectively) may improve the survival of patients. The Lasso regression analysis was then used to remove MRGs that may be highly related to other MRGs (Fig.3b-c). Furthermore, a prognostic signature model was established based on multivariate Cox regression analysis. Finally, six MRGs were confirmed and applied to establish a metabolism-related signature (Fig.3d). A prognostic model was constructed to evaluate the prognosis of each patient as follows: Risk score =  $(0.001709 \times \text{expression value of ALDOA}) + (-0.01187 \times \text{expression value of CAT}) + (0.082279 \times \text{expression value of ENTPD2}) + (0.030344 \times \text{expression value of GNPAT1}) + (0.003499 \times \text{expression value of LDHA}) + (0.018476 \times \text{expression value of FBP1})$

TYMS).

Then, the risk score of each patient was calculated according to this prognostic model. Based on the median risk score, 426 LUAD patients were classified into a high risk group (n = 213) and low risk group (n = 213). The risk score, survival status and gene expression heatmap of these prognostic MRGs are presented in Fig 4a-c. Kaplan-meier log-rank test indicated that patients in the high risk group showed markedly poorer overall survival than those in the low risk group (Fig. 4d). Areas under the curve value of the signature predicting the 1-, 3- and 5-year OS rates were 0.73, 0.703 and 0.854, indicating that this prognostic model exhibited a good sensitivity and specificity (Fig.4e).

#### **Validation of the metabolism-related prognostic signature for LUAD**

The GSE31210 dataset including 174 LUAD samples were used for the validation of the metabolism-related signature. According to the median risk score, we divided patients into high risk (n = 78) and low risk groups (n = 96). Consistent with the results derived from the TCGA database, the Kaplan-Meier curve demonstrated that patients in the high risk group exhibited markedly poorer overall survival than those in the low risk group (Fig. 5d). The risk score, survival status and gene expression heatmap of these prognostic MRGs are shown in Fig 5a-c. The AUCs for 1-, 3- and 5-year OS rates were 0.654, 0.705 and 0.725 (Fig.5e). A nomogram for predicting 1-, 3- and 5-year OS of patients with LUAD was constructed with the six prognostic genes that had most significant values in multivariate analysis (Fig.6a). In addition, the calibration curve of the prognostic nomogram demonstrated good agreement between the predicted and

observed survival rates for each of OS(Fig.6b-d).

### **Analysis of these crucial MRGs expression level**

To explore the six hub genes at the transcription level, the mRNA expression levels were analyzed using the TCGA database. The results demonstrated that the expression of ALDOA, ENTPD2, GNPAT1, LDHA, and TYMS in LUAD tissues were all higher than those of adjacent tissues, while the expression of CAT was lower than those of adjacent tissues(Fig.7).To assess the six hub genes at the translational level, the protein expression levels were analyzed using the HPA database. The results showed that the protein level of ALDOA, ENTPD2, GNPAT1, LDHA, and TYMS were higher in LUAD tissues than in normal tissues, matched their mRNA expression levels(Fig.8).There is no difference between LUAD tissues and normal tissues for the protein level of CAT(Fig.8b).

### **Clinical value of prognostic signature**

Univariate and multivariate Cox regression analysis was conducted to evaluate the independent prediction ability of metabolism-related prognostic signature between the signature and other common prognostic factors, including age, gender, histological grade, pathological stage and TNM stage. Although univariate Cox analysis indicated that pathologic stage, T stage, N stage and our model were markedly associated with OS (Fig.9a;  $p < 0.001$ ), after the multivariate analysis, only metabolism-related prognostic signature( $p < 0.001$ ) and pathological stage ( $p < 0.007$ )can be used as an independent prognostic factor (Fig.9b). To further evaluate the clinical value of MRGs, the relationship between MRGs prognostic indicators and clinical features were

investigated, and the results indicated that ALDOA, ENTPD2, GNPAT1, LDHA, and CAT were differentially expressed in patients with various clinical features (Fig.10). To validate the clinical value of the metabolism-related prognostic signature, the association between the risk score and clinical characteristics were subsequently assessed, and the results demonstrated that high risk scores were positively associated with survival status, gender, N stage, and pathologic stage in patients with LUAD (Fig.10).

## **Discussion**

LUAD, which is highly heterogeneous in morphological characteristics and remarkably variable in prognosis, is the most prevalent subtype of non-small cell lung cancer (NSCLC)[4, 23]. More and more attention has been recently paid to the key role of gene signatures based on specific correlation in predicting the prognosis of LUAD because of the rapid advances in high-throughput technologies and bioinformatics methodology[24-27]. Moreover, the identification of novel gene signatures that predict the prognosis of patients is beneficial for the choice of treatment regimens and the improvement of survival rate[28, 29].

In recent years, interest in dysregulated metabolism of cancer has been growing[30]. Accumulating evidence showed that MRGs play a key role in cancer development and progression[31]. Therefore, the identification of novel MRGs has lately become a hotspot in cancer research, both as a biomarker and potential therapeutic target. In this study, a total of 116 differentially expressed MRGs, which consist of 31 downregulated and 85 upregulated MRGs, were detected in the TCGA dataset. We found that 12 MRGs

were most significantly associated with OS by using the univariate regression analysis in LUAD. After conducting the LASSO regression and multivariable Cox regression analyses, a novel prognostic signature which consisted of six MRGs (ALDOA, CAT, ENTPD2, GNPAT1, LDHA, and TYMS) was established. Based on the gene signature, LUAD patients were classified into a high risk group and low risk group. Patients in the high risk group, which had a survival rate lower than 15%, showed markedly poorer overall survival than the low risk group. The time-dependent ROC analysis demonstrated that the area under the curve (AUC) for 1, 3, and 5 years were 0.73, 0.703, and 0.854, respectively, indicating that this prognostic signature had good sensitivity and specificity. The prognostic value of this signature was further successfully validated in the GSE31210 dataset. Moreover, the calibration curve of the prognostic nomogram demonstrated good agreement between the predicted and observed survival rates for each OS. Further analysis indicated that this signature could be an independent prognostic indicator after adjusting to other clinical factors. Finally, the signature was found to be associated with various clinicopathological features.

Furthermore, six genes in this prognostic signature were selected as crucial MRGs including aldolase A (ALDOA), catalase (CAT), ectonucleoside triphosphate diphosphohydrolase-2 (ENTPD2), glucosamine-phosphate N-acetyltransferase 1 (GNPAT1), lactate dehydrogenase A (LDHA), and thymidylate synthetase (TYMS). ALDOA is an important enzyme involved in the glycolysis pathway that is highly expressed in a wide range of cancers[32]. Some studies also proved that the overexpression of ALDOA may contribute to tumorigenesis and the progression of

cancers through modulation of HIF-1 $\alpha$  signaling[33, 34]. Our results showed that ALDOA may be a tumor-promoting gene in LUAD. Abnormal expression or decreased activity of CAT can lead to an increase in intracellular ROS concentration, which directly or indirectly induces tumorigenesis[35, 36]. Consistently, our study found that compared with normal tissues, the mRNA level of CAT in LUAD tissues was down-regulated. ENTPD2 is ectonucleotidase with extracellular facing catalytic domain, which can convert extracellular ATP to ADP. A study by Chiu et al revealed that blockade of ENTPD2 could significantly inhibit cancer cells growth and effectively improve the efficiency and efficacy of immune checkpoint inhibitors in vivo experiments [37]. GNPAT1, a member of the GNAT protein superfamily, was a key enzyme in the metabolic pathway of N-acetylglucosamine(GlcNAc) synthesis[38]. Zhao et al. reported that the overexpression of GNPAT1 could promote the infiltration and adhesion of lung cancer cells[39]. LDHA catalyzes the conversion of pyruvate to lactate with concomitant oxidation of NADH to NAD<sup>+</sup>, which plays an essential role in metabolic pathways of the cancer cells[40]. Recently, accumulating evidence showed that the overexpression of LDHA can promote the production of lactate, thus contributing to the acidification of the tumor microenvironment, which may limit the effect of anti-PD-L1 therapy[41, 42]. TYMS, a rate-limiting enzyme during the DNA synthesis, plays an important role in catalyzing the methylation of deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP)[43]. High levels of TYMS expression are related to worse responses to 5-FU, shorter survival times and other adverse clinical behaviors in a variety of solid tumors[44, 45]. Since only patients

with low expression of TYMS can respond to 5-FU, individualized chemotherapy regimens can be selected according to the expression of TYMS and tumor classification[46].

However, we should acknowledge that there are some obvious limitations in the present study which should be addressed in future studies. First, the potential selection bias could not be ruled out because of the transcriptomic and the corresponding clinical data of patients with LUAD were obtained from public database. Second, the robustness of the prognostic signature must be validated in large prospective studies. Finally, our findings have to be validated by in vivo and in vitro experiments to explore the functional role of the critical genes in LUAD.

## **Conclusions**

In summary, we identified a novel signature based on MRGs that could be applied to analyze the prognostic of patients with LUAD, and verified by the data from the GEO databases. This signature may provide valuable information either for diagnosis or developing novel therapeutic options for LUAD patients in the future.

## **Abbreviations**

TCGA : The Cancer Genome Atlas; GEO: Gene Expression Omnibus; LUAD: Lung adenocarcinoma; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; FDR: False discovery rate. MRGs:metabolism-related genes; ROC:Receiver operating characteristic; HR: Hazard ratio; CI: Confidence interval. TNM:Tumour size/lymph nodes/distance metastasis, a tumour staging system used in oncology and constructed by the American Joint Committee on Cancer and the Union for

International Cancer Control;

### **Acknowledgements**

None.

### **Authors' contributions**

Shenghui Qin conceived of and directed the project, wrote the manuscript. Zhihao Wang performed data bioinformatics analyses. Kidane Siele Embaye et.al helped with part of English writing and checking. All authors read and approved the manuscript.

### **Funding**

This research was supported by a Grant from the National Natural Science Foundation of China (Grant 81700012; to Shenghui Qin)

### **Data availability statement**

All data generated or analyzed during the present study was downloaded from TCGA database, GEO database and HPA database.

### **Ethics approval and consent to participate**

Not applicable.

### **Consent for publication**

Not applicable.

### **Competing interests**

All authors declare no conflict of interests.

### **Author details**

<sup>1</sup> Institute of Pathology, Tongji Hospital, Tongji Medical College, Huazhong University

of Science and Technology, Wuhan 430030, China. <sup>2</sup> Department of Pharmacy, Hiser Medical Center of Qingdao, Qingdao 266033, China. <sup>3</sup> Department of Pathology, Wuhan Third Hospital (Tongren Hospital of Wuhan University), Wuhan 430030, China.

## References

1. Siegel RL, Miller KD, Jemal A: **Cancer statistics, 2019**. *Ca-a Cancer Journal for Clinicians* 2019, **69**(1):7-34.
2. Travis WD: **Reporting lung cancer pathology specimens. Impact of the anticipated 7th Edition TNM Classification based on recommendations of the IASLC Staging Committee**. *Histopathology* 2009, **54**(1):3-11.
3. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Pineros M, Znaor A, Bray F: **Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods**. *International Journal of Cancer* 2019, **144**(8):1941-1953.
4. Cancer Genome Atlas Research N: **Author Correction: Comprehensive molecular profiling of lung adenocarcinoma**. *Nature* 2018, **559**(7715):E12-E12.
5. Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ *et al*: **EGFR mutations in lung cancer: Correlation with clinical response to gefitinib therapy**. *Science* 2004, **304**(5676):1497-1500.
6. Solomon BJ, Mok T, Kim D-W, Wu Y-L, Nakagawa K, Mekhail T, Felip E, Cappuzzo F, Paolini J, Usari T *et al*: **First-Line Crizotinib versus Chemotherapy in ALK-Positive Lung Cancer**. *New England Journal of Medicine* 2014, **371**(23):2167-2177.
7. Kim YT, Kim T-y, Lee DS, Park SJ, Park J-Y, Seo S-J, Choi H-S, Kang HJ, Hahn S, Kang CH *et al*: **Molecular changes of epidermal growth factor receptor (EGFR) and KRAS and their impact on the clinical outcomes in surgically resected adenocarcinoma of the lung**. *Lung Cancer* 2008, **59**(1):111-118.
8. Field JK, Oudkerk M, Pedersen JH, Duffy SW: **Prospects for population screening and diagnosis of lung cancer**. *Lancet* 2013, **382**(9893):732-741.
9. Collins LG, Haines C, Perkel R, Enck RE: **Lung cancer: diagnosis and management**. *American family physician* 2007, **75**(1):56-63.
10. Riihimäki M, Hemminki A, Fallah M, Thomsen H, Sundquist K, Sundquist J, Hemminki K: **Metastatic sites and survival in lung cancer**. *Lung Cancer* 2014, **86**(1):78-84.
11. Hanahan D, Weinberg RA: **Hallmarks of Cancer: The Next Generation**. *Cell* 2011, **144**(5):646-674.
12. Hirschey MD, DeBerardinis RJ, Diehl AME, Drew JE, Frezza C, Green MF, Jones LW, Ko YH, Le A, Lea MA *et al*: **Dysregulated metabolism contributes to oncogenesis**. *Seminars in Cancer Biology* 2015, **35**:S129-S150.
13. Warburg O: **On the origin of cancer cells**. *Science* 1956, **123**(3191):309-314.
14. Upadhyay M, Samal J, Kandpal M, Singh OV, Vivekanandan P: **The Warburg effect: insights from the past decade**. *Pharmacology & therapeutics* 2013, **137**(3):318-330.
15. Ruocco MR, Avagliano A, Granato G, Vigliar E, Masone S, Montagnani S, Arcucci A: **Metabolic flexibility in melanoma: A potential therapeutic target**. *Semin Cancer Biol*

- 2019, **59**:187-207.
16. Gallo M, Sapio L, Spina A, Naviglio D, Calogero A, Naviglio S: **Lactic dehydrogenase and cancer: an overview**. *Frontiers in bioscience (Landmark edition)* 2015, **20**:1234-1249.
  17. Gottfried E, Kreutz M, Mackensen A: **Tumor metabolism as modulator of immune response and tumor progression**. *Semin Cancer Biol* 2012, **22**(4):335-341.
  18. Noble WS: **How does multiple testing correction work?** *Nature biotechnology* 2009, **27**(12):1135-1137.
  19. Yu G, Wang L-G, Han Y, He Q-Y: **clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters**. *Omics-a Journal of Integrative Biology* 2012, **16**(5):284-287.
  20. Zhang Z, Reinikainen J, Adeleke KA, Pieterse ME, Groothuis-Oudshoorn CGM: **Time-varying covariates and coefficients in Cox regression models**. *Annals of Translational Medicine* 2018, **6**(7).
  21. Heagerty PJ, Lumley T, Pepe MS: **Time-dependent ROC curves for censored survival data and a diagnostic marker**. *Biometrics* 2000, **56**(2):337-344.
  22. Thul PJ, Akesson L, Wiking M, Mahdessian D, Geladaki A, Blal HA, Alm T, Asplund A, Bjork L, Breckels LM *et al*: **A subcellular map of the human proteome**. *Science* 2017, **356**(6340).
  23. Matsuda T, Machii R: **Morphological distribution of lung cancer from Cancer Incidence in Five Continents Vol. X**. *Japanese journal of clinical oncology* 2015, **45**(4):404.
  24. Wang X, Yao S, Xiao Z, Gong J, Liu Z, Han B, Zhang Z: **Development and validation of a survival model for lung adenocarcinoma based on autophagy-associated genes**. *Journal of translational medicine* 2020, **18**(1):149.
  25. Li W, Gao LN, Song PP, You CG: **Development and validation of a RNA binding protein-associated prognostic model for lung adenocarcinoma**. *Aging* 2020, **12**(4):3558-3573.
  26. Xue L, Bi G, Zhan C, Zhang Y, Yuan Y, Fan H: **Development and Validation of a 12-Gene Immune Relevant Prognostic Signature for Lung Adenocarcinoma Through Machine Learning Strategies**. *Frontiers in oncology* 2020, **10**:835.
  27. Li W, Li N, Gao L, You C: **Integrated analysis of the roles and prognostic value of RNA binding proteins in lung adenocarcinoma**. *PeerJ* 2020, **8**:e8509.
  28. Wang Z, Wang Z, Niu X, Liu J, Wang Z, Chen L, Qin B: **Identification of seven-gene signature for prediction of lung squamous cell carcinoma**. *OncoTargets and therapy* 2019, **12**:5979-5988.
  29. Ma X, Ren H, Peng R, Li Y, Ming L: **Identification of key genes associated with progression and prognosis for lung squamous cell carcinoma**. *PeerJ* 2020, **8**:e9086.
  30. Porta C, Sica A, Riboldi E: **Tumor-associated myeloid cells: new understandings on their metabolic regulation and their influence in cancer immunotherapy**. *The FEBS journal* 2018, **285**(4):717-733.
  31. Pavlova NN, Thompson CB: **The Emerging Hallmarks of Cancer Metabolism**. *Cell metabolism* 2016, **23**(1):27-47.
  32. Asaka M, Kimura T, Meguro T, Kato M, Kudo M, Miyazaki T, Alpert E: **Alteration of aldolase isozymes in serum and tissues of patients with cancer and other diseases**. *Journal of clinical laboratory analysis* 1994, **8**(3):144-148.
  33. Ji S, Zhang B, Liu J, Qin Y, Liang C, Shi S, Jin K, Liang D, Xu W, Xu H *et al*: **ALDOA functions as an oncogene in the highly metastatic pancreatic cancer**. *Cancer letters* 2016,

- 374(1):127-135.
34. Saito Y, Takasawa A, Takasawa K, Aoyama T, Akimoto T, Ota M, Magara K, Murata M, Hirohashi Y, Hasegawa T *et al*: **Aldolase A promotes epithelial-mesenchymal transition to increase malignant potentials of cervical adenocarcinoma**. *Cancer science* 2020.
  35. Hasegawa Y, Takano T, Miyauchi A, Matsuzuka F, Yoshida H, Kuma K, Amino N: **Decreased expression of catalase mRNA in thyroid anaplastic carcinoma**. *Japanese journal of clinical oncology* 2003, **33**(1):6-9.
  36. Glorieux C, Dejeans N, Sid B, Beck R, Calderon PB, Verrax J: **Catalase overexpression in mammary cancer cells leads to a less aggressive phenotype and an altered response to chemotherapy**. *Biochem Pharmacol* 2011, **82**(10):1384-1390.
  37. Chiu DK, Tse AP, Xu IM, Di Cui J, Lai RK, Li LL, Koh HY, Tsang FH, Wei LL, Wong CM *et al*: **Hypoxia inducible factor HIF-1 promotes myeloid-derived suppressor cells accumulation through ENTPD2/CD39L1 in hepatocellular carcinoma**. *Nature communications* 2017, **8**(1):517.
  38. Peneff C, Mengin-Lecreux D, Bourne Y: **The crystal structures of Apo and complexed Saccharomyces cerevisiae GNA1 shed light on the catalytic mechanism of an amino-sugar N-acetyltransferase**. *The Journal of biological chemistry* 2001, **276**(19):16328-16334.
  39. Zhao M, Li H, Ma Y, Gong H, Yang S, Fang Q, Hu Z: **Nanoparticle abraxane possesses impaired proliferation in A549 cells due to the underexpression of glucosamine 6-phosphate N-acetyltransferase 1 (GNPNAT1/GNA1)**. *International journal of nanomedicine* 2017, **12**:1685-1697.
  40. Johnson KP, Hillman JD: **Competitive properties of lactate dehydrogenase mutants of the oral bacterium Streptococcus mutans in the rat**. *Archives of oral biology* 1982, **27**(6):513-516.
  41. Hinzman CP, Aljehane L, Brown-Clay JD, Kallakury B, Sonahara F, Goel A, Trevino J, Banerjee PP: **Aberrant expression of PDZ-binding kinase/T-LAK cell-originated protein kinase modulates the invasive ability of human pancreatic cancer cells via the stabilization of oncoprotein c-MYC**. *Carcinogenesis* 2018, **39**(12):1548-1559.
  42. Zhang J, Wolfgang CL, Zheng L: **Precision Immuno-Oncology: Prospects of Individualized Immunotherapy for Pancreatic Cancer**. *Cancers* 2018, **10**(2).
  43. Gangjee A, Yu J, McGuire JJ, Cody V, Galitsky N, Kisliuk RL, Queener SF: **Design, synthesis, and X-ray crystal structure of a potent dual inhibitor of thymidylate synthase and dihydrofolate reductase as an antitumor agent**. *Journal of medicinal chemistry* 2000, **43**(21):3837-3851.
  44. Lee SW, Chen TJ, Lin LC, Li CF, Chen LT, Hsing CH, Hsu HP, Tsai CJ, Huang HY, Shiue YL: **Overexpression of thymidylate synthetase confers an independent prognostic indicator in nasopharyngeal carcinoma**. *Experimental and molecular pathology* 2013, **95**(1):83-90.
  45. Formentini A, Henne-Bruns D, Kornmann M: **Thymidylate synthase expression and prognosis of patients with gastrointestinal cancers receiving adjuvant chemotherapy: a review**. *Langenbeck's archives of surgery* 2004, **389**(5):405-413.
  46. Qiu LX, Tang QY, Bai JL, Qian XP, Li RT, Liu BR, Zheng MH: **Predictive value of thymidylate synthase expression in advanced colorectal cancer patients receiving**

fluoropyrimidine-based chemotherapy: evidence from 24 studies. *Int J Cancer* 2008, 123(10):2384-2389.

## **Figure Legends**

### **Fig.1 Differentially expressed metabolism-related genes (MRGs) in lung adenocarcinoma (LUAD).**

Heatmap of MRGs between LUAD and nontumor tissues in TCGA database (a); The color from blue to red represents the progression from low expression to high expression. Volcano plot of MRGs in TCGA database(b); The red dots in the plot represents upregulated genes and blue dots represents downregulated genes with statistical significance. Gray dots represent no differentially expressed genes.

### **Fig2. Gene functional enrichment analysis of differentially expressed MRGs.**

(a)The top 30 significant terms of GO function enrichment. BP biological process, CC cellular component, MF molecular function. (b)The GO circle shows the scatter map of the logFC of the specified gene. (c)The top 30 significant terms of KEGG analysis. (d)The KEGG circle shows the scatter map of the logFC of the specified gene. The higher the Z-score value indicated, the higher expression of the enriched pathway.

### **Fig3. Establishment of metabolism-related prognostic signature.**

(a) Univariate cox regression analysis identified these differentially expressed MRGs related to the LUAD risks. P values <0.001 are considered to be statistically significant. (b)Screening of optimal parameter (lambda) at which the vertical lines were drawn. (c)Lasso coefficient profiles of the seventeen MRGs with non-zero coefficients determined by the optimal lambda. (d) Multivariate cox analysis to developing a prognostic index based on these MRGs.

**Fig.4 Construction of the metabolism-based prognostic risk signature in the TCGA cohort.**

(a)The risk score distribution of LUAD patients; (b)Survival status and duration of patients;(c)Heatmap of the metabolism-related genes expression; (d)Survival curves for the low risk and high risk groups; (e)Time-independent receiver operating characteristic (ROC) analysis of risk scores for prediction the overall survival in the TCGA set.

**Fig.5 Validation of the metabolism-based prognostic risk signature in the GSE31210 cohort.**

(a)The risk score distribution of LUAD patients; (b)Survival status and duration of patients;(c)Heatmap of the metabolism-related genes expression; (d)Survival curves for the low risk and high risk groups; (e)Time-independent receiver operating characteristic (ROC) analysis of risk scores for prediction the overall survival in the GSE31210 set.

**Fig.6 Construction of a nomogram based on the metabolism-related signature in the TCGA cohort.**

(a) The nomogram based on the signature in LUAD patients at 1, 3, and 5 years. (b-d) Calibration curves of nomogram for the signature at 1, 3, and 5 years.

**Fig.7 Comparison of the crucial genes mRNA levels in paired adjacent normal tissues and LUAD tissues from TCGA. (a): ALDOA, (b):CAT, (c): ENTPD2, (d):GNPNAT1, (e): LDHA , (f): TYMS.**

**Fig.8 Verification of hub MRGs expression in LUAD and normal liver tissue using the HPA database. (a): ALDOA, (b):CAT, (c): ENTPD2, (d):GNPNAT1, (e): LDHA , (f): TYMS.**

**Fig.9 Univariate (a) and multivariate (b) independent prognostic analysis of independent risk factors for overall survival (OS) in patients with LUAD.**

**Fig.10 Relationships between MRGs expression and clinicopathological factors in LUAD (P < 0.05).**

# Figures

Fig. 1

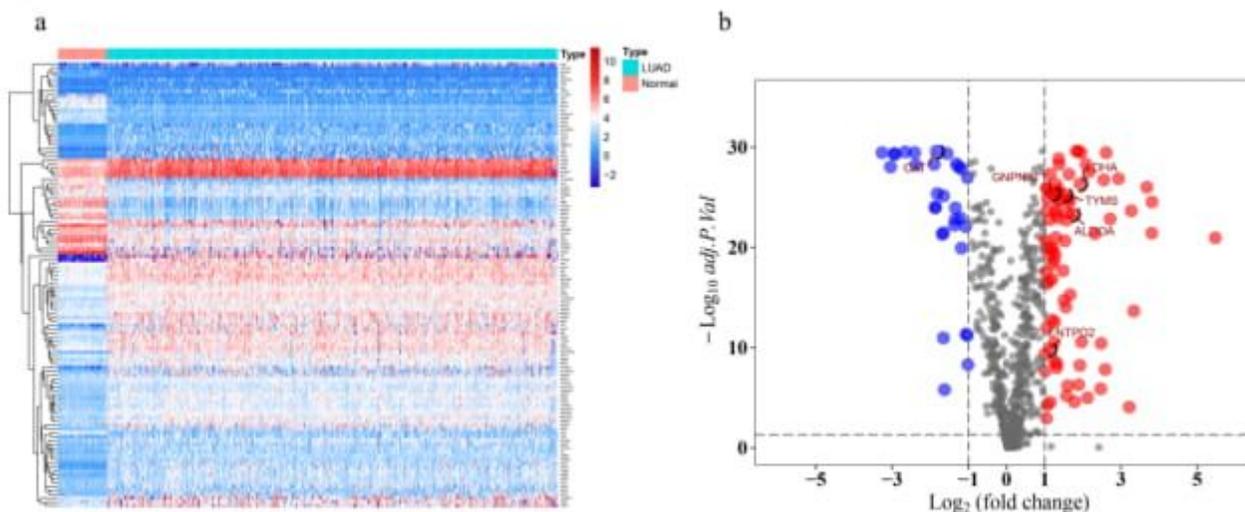


Figure 1

Differentially expressed metabolism-related genes (MRGs) in lung adenocarcinoma (LUAD). Heatmap of MRGs between LUAD and nontumor tissues in TCGA database (a); The color from blue to red represents the progression from low expression to high expression. Volcano plot of MRGs in TCGA database(b); The red dots in the plot represents upregulated genes and blue dots represents downregulated genes with statistical significance. Gray dots represent no differentially expressed genes.

Fig. 2

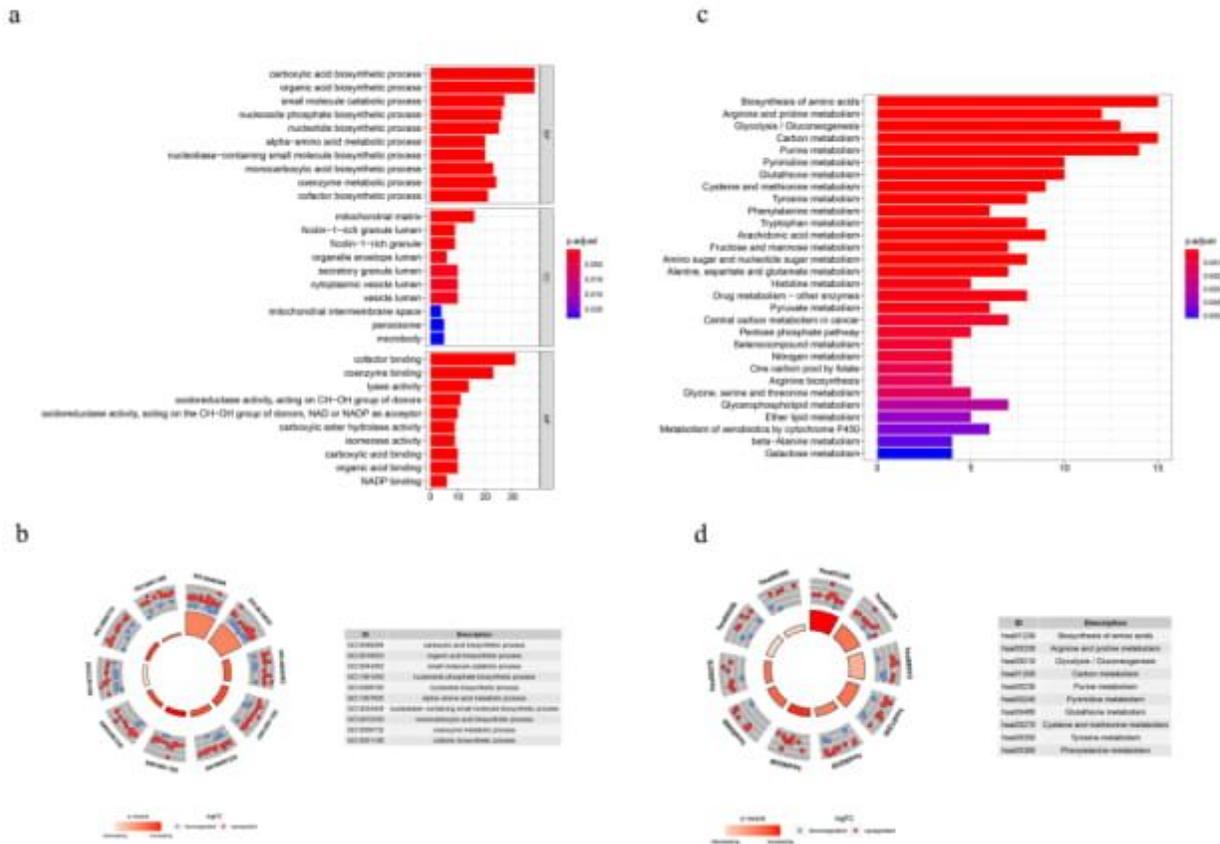


Figure 2

Gene functional enrichment analysis of differentially expressed MRGs. (a) The top 30 significant terms of GO function enrichment. BP biological process, CC cellular component, MF molecular function. (b) The GO circle shows the scatter map of the logFC of the specified gene. (c) The top 30 significant terms of KEGG analysis. (d) The KEGG circle shows the scatter map of the logFC of the specified gene. The higher the Z-score value indicated, the higher expression of the enriched pathway.

Fig. 3

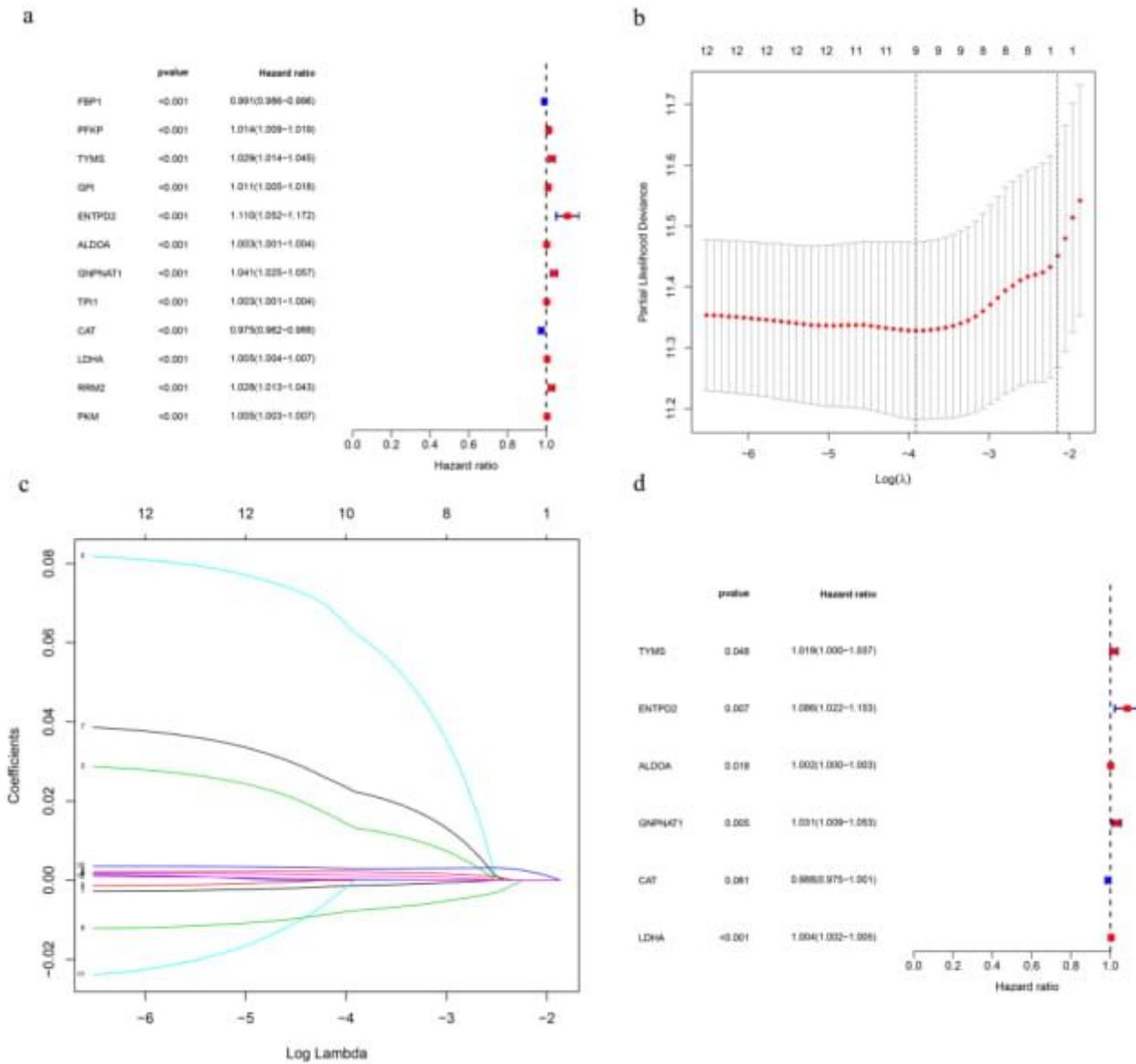


Figure 3

Establishment of metabolism-related prognostic signature. (a) Univariate cox regression analysis identified these differentially expressed MRGs related to the LUAD risks. P values <0.001 are considered to be statistically significant. (b) Screening of optimal parameter (lambda) at which the vertical lines were drawn. (c) Lasso coefficient profiles of the seventeen MRGs with non-zero coefficients determined by the optimal lambda. (d) Multivariate cox analysis to developing a prognostic index based on these MRGs.

Fig. 4

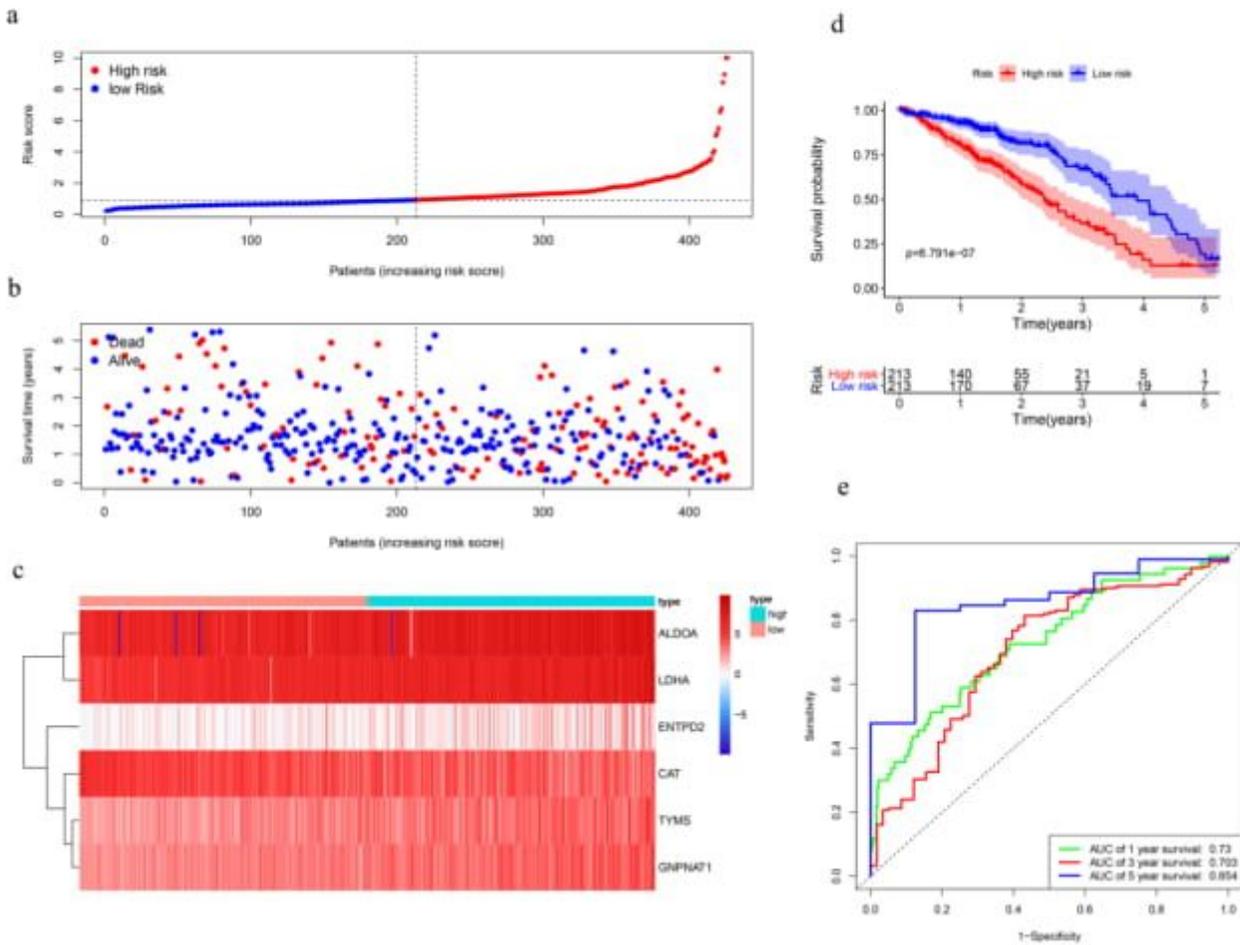


Figure 4

Construction of the metabolism-based prognostic risk signature in the TCGA cohort. (a) The risk score distribution of LUAD patients; (b) Survival status and duration of patients; (c) Heatmap of the metabolism-related genes expression; (d) Survival curves for the low risk and high risk groups; (e) Time-independent receiver operating characteristic (ROC) analysis of risk scores for prediction the overall survival in the TCGA set.

Fig.5

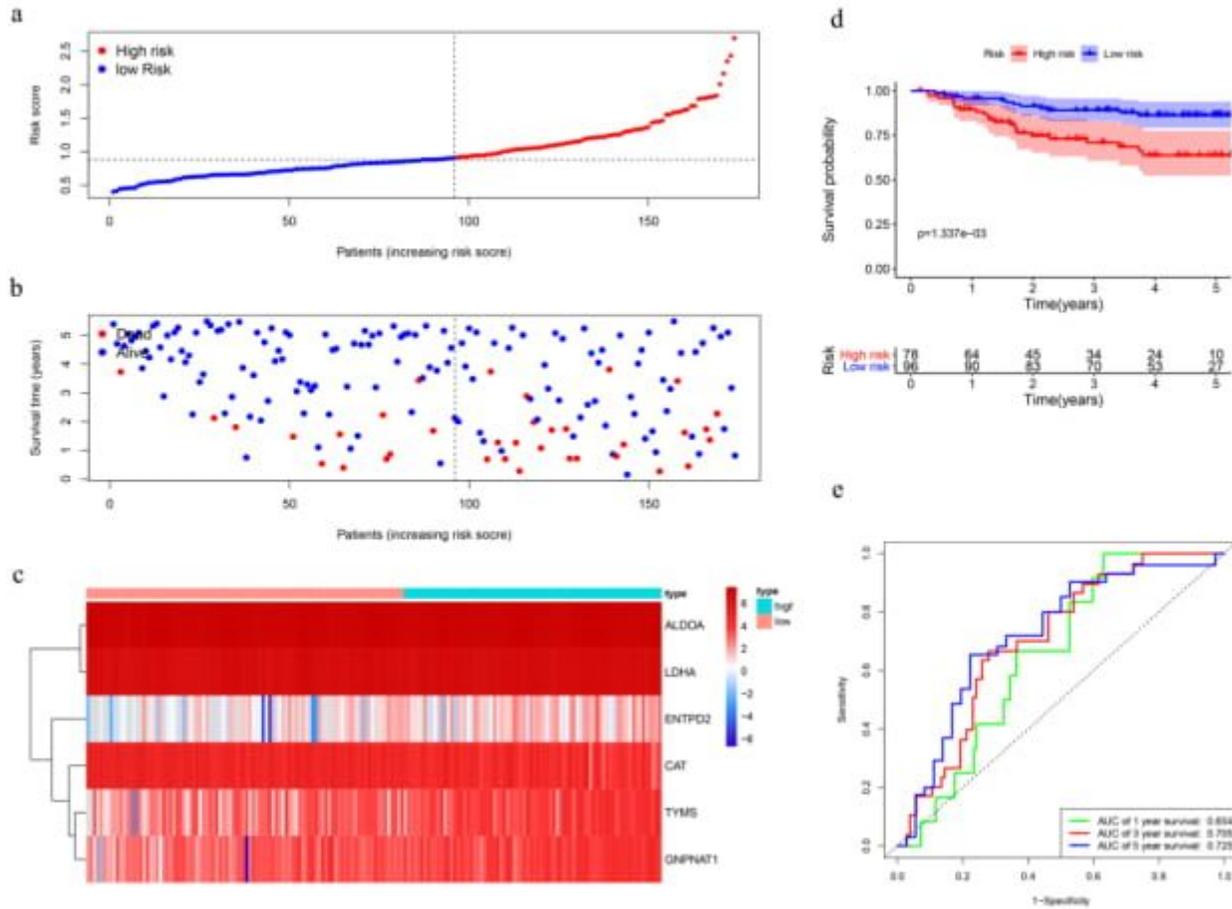
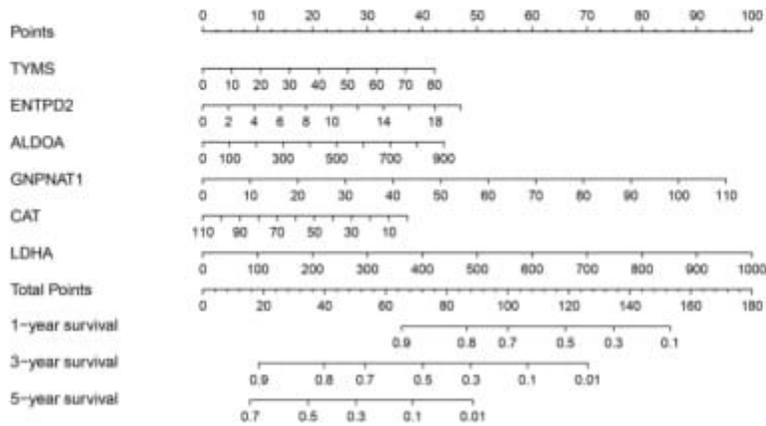


Figure 5

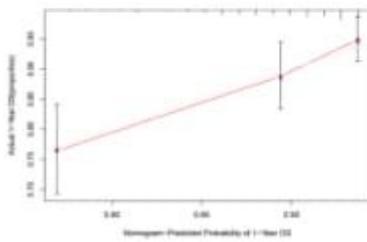
Validation of the metabolism-based prognostic risk signature in the GSE31210 cohort. (a)The risk score distribution of LUAD patients; (b)Survival status and duration of patients;(c)Heatmap of the metabolism-related genes expression; (d)Survival curves for the low risk and high risk groups; (e)Time-independent receiver operating characteristic (ROC) analysis of risk scores for prediction the overall survival in the GSE31210 set.

Fig. 6

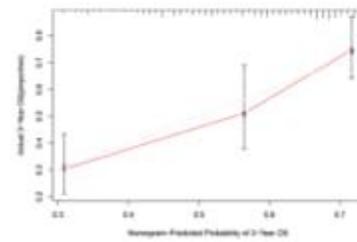
a



b



c



d

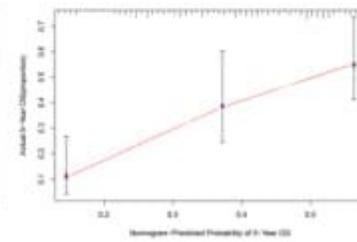
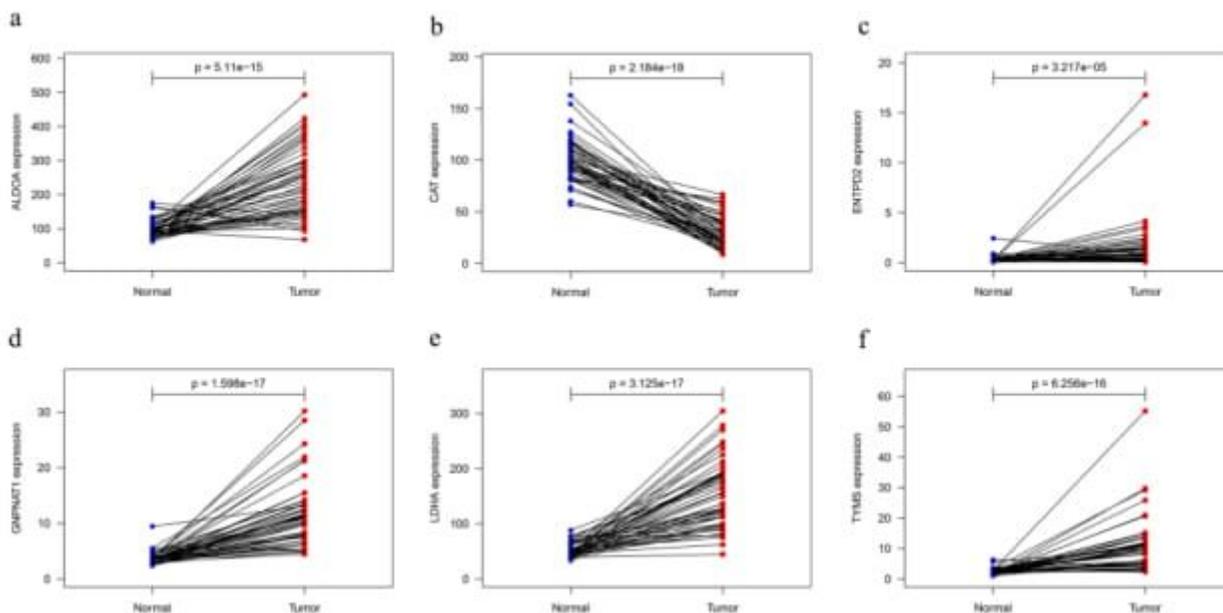


Figure 6

Construction of a nomogram based on the metabolism-related signature in the TCGA cohort. (a) The nomogram based on the signature in LUAD patients at 1, 3, and 5 years. (b-d) Calibration curves of nomogram for the signature at 1, 3, and 5 years.

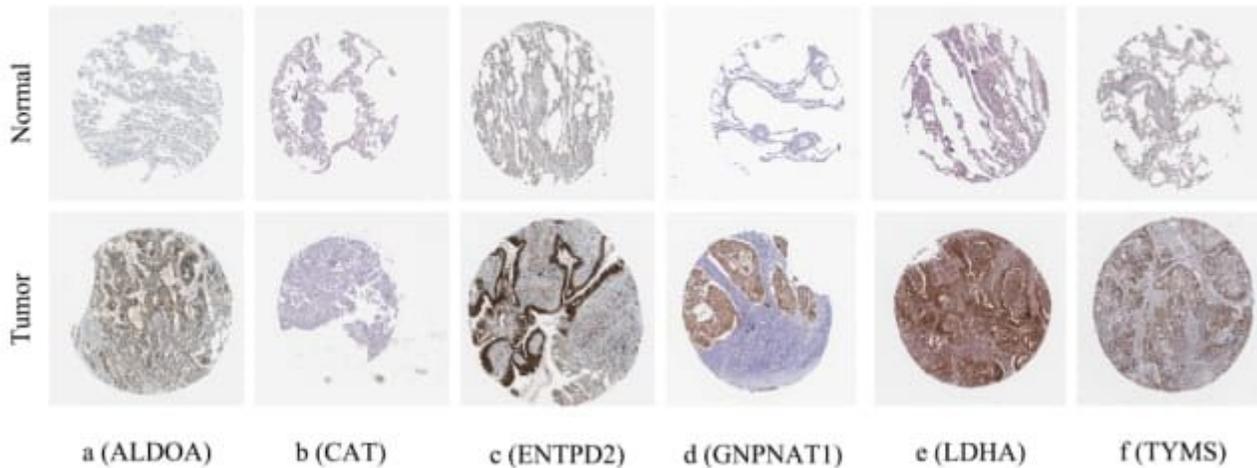
Fig. 7



**Figure 7**

Comparison of the crucial genes mRNA levels in paired adjacent normal tissues and LUAD tissues from TCGA. (a): ALDOA, (b):CAT, (c): ENTPD2, (d):GNPNAT1, (e): LDHA , (f): TYMS.

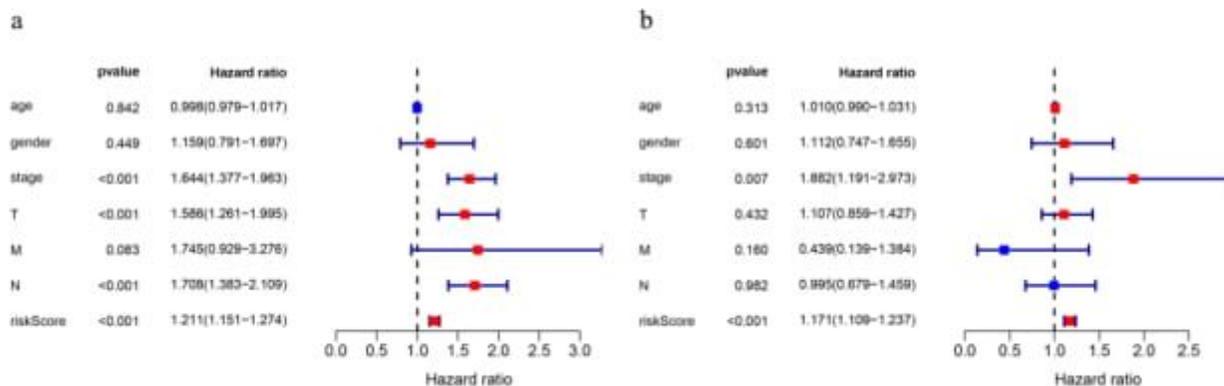
**Fig. 8**



**Figure 8**

Verification of hub MRGs expression in LUAD and normal liver tissue using the HPA database. (a): ALDOA, (b):CAT, (c): ENTPD2, (d):GNPNAT1, (e): LDHA , (f): TYMS.

**Fig. 9**



**Figure 9**

Univariate (a) and multivariate (b) independent prognostic analysis of independent risk factors for overall survival (OS) in patients with LUAD. Fig.10 Relationships between MRGs expression and clinicopathological factors in LUAD (P < 0.05).

Fig. 10

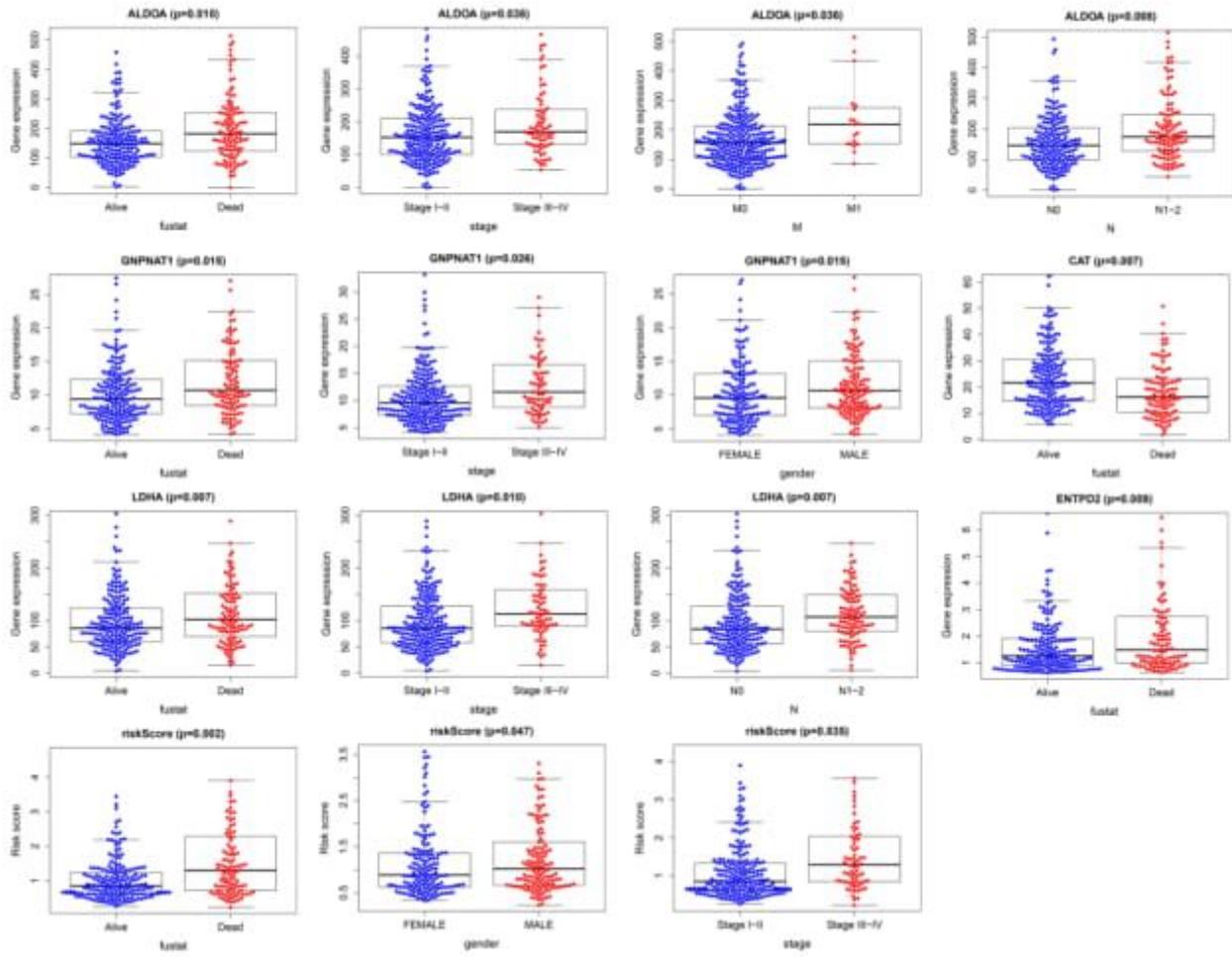


Figure 10

Fig 10