A glycolysis-based three-gene signature predicts survival in patients with lung squamous cell carcinoma

Guichuan Huang  
The third affiliated hospital of Zunyi Medical University

Jing Zhang  
The affiliated hospital of Zunyi Medical University

Ling Gong  
the third affiliated hospital of Zunyi Medical University

Yi Huang  
the third affiliated hospital of Zunyi Medical University

Daishun Liu  
ldsjwtg@126.com  
the third affiliated hospital of Zunyi Medical University

Research article

Keywords: lung cancer, glycolysis, prognosis, gene signature

Posted Date: November 9th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-42550/v2

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 BMC Cancer on May 27th, 2021. See the published version at https://doi.org/10.1186/s12885-021-08360-z.
Abstract

**Background:** Lung cancer is one of the most lethal and most prevalent malignant tumors worldwide, and lung squamous cell carcinoma (LUSC) is one of major histological subtypes. Although, numerous biomarkers were found to be associated with prognosis in LUSC, the prediction effect of a single gene biomarker is not sufficient, especially for glycolysis-related genes. Therefore, we aimed to develop a novel glycolysis-related gene signature to predict survival of patients with LUSC.

**Methods:** The mRNA expression files and clinical information of LUSC were obtained from The Cancer Genome Atlas (TCGA) dataset.

**Results:** Based on Gene set enrichment analysis (GSEA), we found 5 glycolysis-related gene sets were significantly enriched in LUSC tissues. Univariate and multivariate Cox proportional regression models were conducted to choose prognostic-related gene signature. Based on Cox proportional regression model, a risk score of three-gene signature (including HKDC1, ALDH7A1, and MDH1) was established to divide patients into high-risk and low-risk subgroups. We found that a risk score of three-gene signature was an independent of prognostic indicator in LUSC using multivariate Cox regression analysis. Additionally, based on the cBioPortal database, the rate of alterations in HKDC1, ALDH7A1, and MDH1 genes were 1.9%, 1.1%, and 5% in LUSC patients, respectively.

**Conclusion:** In conclusion, a glycolysis-based three-gene signature could serve as a novel biomarker in predicting prognosis of patients with LUSC, which provided more gene targets to cure LUSC patients.

Background

Lung cancer is the leading cause of cancer-related mortality worldwide. There are two clinical subtypes for lung cancer: non-small cell lung cancer (NSCLC) (approximately 85%), and small cell lung cancer (SCLC) (approximately 15%).\(^1\) Based on pathological and molecular features, NSCLC is divided into the following major subtypes: lung squamous cell carcinoma (LUSC), lung adenocarcinoma (LUAD), and large cell lung cancer.\(^2\) Recently, advances in the targeted treatments have obviously improved the overall survival (OS) of patients with LUAD, such as epidermal growth factor receptor (EGFR) kinase inhibitors.\(^3\) However, no specific biomarker and relatively optimal targeted therapies were identified for LUSC patients, and the 5-year survival rate of LUSC is less than 20%.\(^4\) Therefore, it is necessary to explore specific diagnostic and prognostic biomarkers for LUSC.

Energy metabolism reprogramming, a process to promote cell growth and proliferation via adjusting energy metabolism, has been regarded as an emerging hallmark of cancer.\(^5\) Under aerobic conditions, the normal cells obtain energy through mitochondrial oxidative phosphorylation. Under anaerobic conditions, the cells get energy via glycolysis instead of oxygen-consuming mitochondrial metabolism.\(^6\) Glycolysis, also known as Warburg effect, often observed in human cancer cells, in which the cancer cells favor glucose metabolism via glycolysis even in the presence of oxygen.\(^7\) This phenomenon is a unique energy
metabolism exist in cancer cells. In recent years, many biomarkers, including glycolysis-associated genes, for LUSC have been determined, such as kininogen 1 (KNG1) and tripartite motif-containing protein 59 (TRIM59). With the development of high-throughput sequencing, various patient genome databases were constructed, which makes us have a deep understanding of genomic changes. Based on database mining, an increasing number of biomarkers were identified that were related to survival of patients with cancer. However, a single gene cannot have good predictive effects. Multigene prognostic model from original tumor biopsy can guided clinicians to choose more effective treatment strategies. Thus, a signature based on multigene expression associated with glycolysis should be established to predict the prognosis of LUSC patients.

In the present study, we used a genome-wide analysis of mRNA expression profiles in LUSC patients from The Cancer Genome Atlas (TCGA) to construct a glycolysis-related gene signature, which could effectively predict the prognosis in LUSC patients.

**Methods**

**Patient dataset**

The mRNA expression profiles of LUSC patients and their corresponding clinical information were obtained from the TCGA database (https://portal.gdc.cancer.gov/).

A total of 501 patients with LUSC and 49 adjacent normal samples were included for the following study. Clinical information, including age, gender, clinical stage, T, M, and N was included in present study (Table 1).

**Gene set enrichment analysis (GSEA)**

GSEA was performed to determine whether the identified gene sets were significant difference between the LUSC and normal groups. The expression levels of 443 glycolysis-related genes were analyzed in LUSC samples and in adjacent non-cancerous tissues. Normalized \( p \) value less than 0.05 was considered statistically significant.

**Prognostic analysis**

We conducted univariate Cox proportional hazard regression analysis to determine the relationship between glycolysis-related genes and overall survival in LUSC patients.

If the \( p<0.01 \), the corresponding glycolysis-related genes were retained and regarded as the candidate prognostic genes of LUSC. Then, the multivariate Cox proportional hazards regression analysis was performed among the pool candidate prognostic glycolysis-related genes to establish the prognostic model. These analyses were performed with the use of R package of survival.

**Statistical analysis**
The selected mRNAs were divided into the risky [hazard ratio (HR)>1] and protective (0<HR<1) types. Based on a linear combination of the expression level of filtered mRNAs weighted by the regression coefficient (\(\beta\)), the formula of risk score was illustrated as follows: Risk score = expression of gene 1 \(\times\) \(\beta_1\) + expression gene 2 \(\times\) \(\beta_2\) + ... + expression of gene \(n\) \(\times\) \(\beta_n\). The \(\beta\) represents the regression coefficient of the corresponding gene obtained from the multivariate cox regression model. According the median value of risk score, patients were divided into high-risk and low-risk groups. Kaplan-Meier curves and log-rank test were utilized to validate the prognostic significance of the risk score. The Student's t test or Mann-Whitney u test was conducted to explore the differential expression of selected genes in LUSC tissues and adjacent normal tissue. If the expression data of selected genes follows normal distribution, student's t test was used to analyze differences between LUSC tissues and adjacent normal tissue; otherwise, Mann-Whitney u test was utilized. Filtered gene alterations in LUSC were explored using cBioPortal database (http://www.cbioportal.org/). All statistical analyses were conducted with the use of SPSS 23.0 and GraphPad Prism 8.0 software.

Results

Initial screening of genes by GSEA

The mRNA expression data set and clinical information of 501 patients with LUSC were obtained from the TCGA database (Figure 1). We found five glycolysis-related gene sets on the Molecular Signatures Database v7.0, including (1) BIOCARTA_GLYCOLYSIS_PATHWAY, (2) GO_GLYCOLYTIC_PROCESS, (3) HALLMARK_GLYCOLYSIS, (4) KEGG_GLYCOLYSIS_GLUONEOGENESIS, (5) REACTOME_GLYCOLYSIS. We performed GSEA to explore whether the identified gene sets were significant difference between LUSC and normal tissues. We found these 5 gene sets were significantly enriched (Figure 2 and Table 2). Then, we collected 443 genes from 5 gene sets for further analysis.

Identification of glycolysis-related genes associated with patient survival

First, univariate Cox proportional hazard regression analysis was conducted to 443 genes that were significantly enriched in LUSC samples from GSEA. A total of 4 genes were obtained which were significantly correlated to the survival of patients (\(p<0.01\)). Next, we performed multivariate Cox regression analysis to further explore the association between the 4 mRNA expression profiles and the overall survival of patients. Finally, 3 genes, including hexokinase domain-containing protein 1 (HKDC1), aldehyde dehydrogenase 7A1 (ALDH7A1), and malate dehydrogenase 1 (MDH1), were included to construct prognostic model. As shown in Table 3, two of the three genes were verified as independent prognostic markers in LUSC. Among three genes, one gene (MDH1) was considered as protective factor owing to 0<HR<1, whereas the remaining two genes (HKDC1 and ALDH7A1) might be prognostic risky factors with their HR>1.

Subsequently, we explored the alterations of three selected genes in 501 LUSC samples using cBioPotral database. The results showed that the rate of alterations in HKDC1, ALDH7A1, and MDH1 genes were 1.9%, 1.1%, and 5%, respectively (Supplementary Figure 1).
The expression level of three genes was conducted between adjacent normal tissues and LUSC tissues. We found that all the three genes were upregulated in LUSC tissues compared with in normal tissues (Figure 3).

**Construction of the three-gene signature to predict patient prognosis**

To predict patient’s prognosis using glycolysis-related genes expression, a prognostic risk model was developed based on the regression coefficients of multivariate Cox regression model to weight the expression level of each gene in the three-gene signature:

\[
\text{risk score} = 0.1598 \times \text{expression value of HKDC1} + 0.1571 \times \text{expression value of ALDH7A1} + (-0.2636) \times \text{expression value of MDH1}.
\]

Owing to 6 of 501 patients lacked the data of survival time, a total of 495 patients were included in the survival analysis. According to the risk score formula, patients were classified into the high-risk (n=247) and the low-risk group (n=248) with a medial value of risk score as a cut-off (Figure 4A). The survival time of each patient was shown in Figure 4B. As shown in Figure 4D, patients in high-risk group had a shorter survival than in low-risk group \((p<0.001)\). The 3-year and 5-year survival rates of patient in high-risk group were 45.4% and 35.0%, respectively. However, the 3-year and 5-year survival rates of low-risk group were 71.9% and 58.1%, respectively. Additionally, a heatmap presented the expression profiles of three mRNAs (Figure 4C). As the risk score increased in the patients with LUSC, the mRNA expression of HKDC1 and ALDH7A1 was obviously upregulated; in contrast, the mRNA expression of MDH1 was downregulated. The area under the receiver operating characteristic (ROC) curve (AUC) for the risk score at 1, 3, and 5-year overall survival were 0.629, 0.665, and 0.636, respectively. (Figure 5).

**Risk score from three-gene signature is an independent prognostic indicator**

Univariate and multivariate Cox regression analysis were performed to evaluate the independent risk factors in patients with LUSC. Several clinicopathological parameters, including age, gender, clinical stage, T, N, and M, as well as risk score were included. The results showed that only risk score was associated with prognosis in the univariate Cox analysis \((HR=2.553, 95\%CI: 1.710-3.811, p<0.0001)\) (Table 4). In the following multivariate Cox analysis, we found age and risk score as independent prognostic indicators (Table 4). These results indicated that the risk score was reliable in predicting the prognosis of patients with LUSC.

**Validation of three-gene signature for survival prediction by Kaplan-Meier curve analysis**

To further verify the prognostic value of the risk score of the three-gene signature associated with glycolysis, patients with LUSC were stratified by age \((\leq 65 \text{ or }>65)\), gender (Female or Male), clinical stage (I+II or III+IV), T (T1+T2 or T3+T4), N (N0 or N1+N2+N3), and M (M0 or M1) (Figure 6). We found no significant difference between high-risk and low-risk in patients with tumor remote metastasis (Figure 6). However, in the subgroup of patients without tumor remote metastasis, the risk score of the three-gene signature was still an independent prognostic indicator. Additionally, regardless of the age, gender, clinical
stage, T, N, patients in the high-risk group based on the risk score had a poor prognosis than patients in
the low-risk group. These findings demonstrated that the three-gene signature predicts effectively the
survival of LUSC patients.

**Discussion**

Recently, numerous genes were considered as biomarkers for cancer prognosis and the clinical
significance of the biomarkers has been explored. For example, a study made by Tang and his colleagues
found that the overexpression of dipeptidyl peptidase 9 (DPP9) was a significant independent factor for
poor prognosis in patients with NSCLC. Similarly, Feng et al. reported that high expression of
forkhead box Q1 (FoxQ1) was associated with the poor prognosis in patients with NSCLC. However, the
expression level of single gene can be influenced by multiple factors, making these biomarkers hard to be
reliable and independent prognosis indications in clinical. Therefore, a statistical model based on the
combination of multiple genes was used to improve the prediction of prognosis in cancer patients.
Studies have shown that the pool of multiple genes was more accurate than single gene in predicting the
prognosis of patients with cancer.

In the present study, we obtained mRNA expression profiles in 501 LUSC patients from TCGA database.
We found that 5 glycolysis-related gene sets were significantly enriched in LUSC samples using GSEA.
Univariate and multivariate Cox regression analysis were carried out to identify the risk score of the three-
gene signature with prognostic value for patients with LUSC. Kaplan-Meier curve analysis indicated that
patients with high risk score had a poor prognosis than patients with low risk score. In additional, in
stratified analysis, the risk score of three-gene signature could effectively predict the prognosis of LUSC
patients in all subgroups except for the subgroup of patients with tumor remote metastasis. The reason
for this discrepancy might be that the number of patients with tumor remote metastasis was too small
(n=7). These results demonstrated that the risk score of three-gene signature could be as an independent
prognostic indicator for LUSC patients. Moreover, measuring the risk score of patients might help
clinicians choose optimal therapy methods.

The metabolism of tumor cells is more active than the normal cells, thus tumor cells need more energy to
keep their higher proliferation. Glycolysis and oxidative phosphorylation are the two important
metabolic pathways related to energy supply. Glycolysis is a relatively low-energy-providing pathway
compared with oxidative phosphorylation. In 1920s, Warburg have found that cancer cells are very active
in glycolysis and require a large amount of glucose to obtain ATP for metabolic activities. This aberrant
phenomenon of glucose metabolism was called aerobic glycolysis, also known as the Warburg
effect. After that, the main genes and enzymes related to glycolysis were begun to explore and further
researched for understanding their functions in metabolism of tumor cells. In recent years, studies have
shown that aerobic glycolysis plays a significant role in tumorigenesis, tumor progression and
metastasis. For example, enolase1 (ENO1) was proved to promote cell glycolysis, growth, migration, and
invasion in NSCLC. Glucose transporter 1 (GLUT1) facilitated increased transport of glucose into cancer
cells to maintain an elevated rate of glycolysis under aerobic conditions.\textsuperscript{20} A high expression of GLUT1 significantly associated with a poor prognosis in lung cancer patients.\textsuperscript{21} However, no set of glycolysis-related genes for predicting LUSC prognosis has been established.

HKDC1, a recently identified fifth hexokinase, plays an important role in cellular glucose metabolism.\textsuperscript{22} Aberration expression HKDC1 was associated with various cancers, including colorectal cancer,\textsuperscript{23} liver cancer,\textsuperscript{24} and breast cancer.\textsuperscript{25} Additionally, Wang and his colleagues reported that HKDC1 was overexpressed in LUAD tissues, and high expression of HKDC1 promoted proliferation, migration, invasion in LUAD.\textsuperscript{26} Furthermore, HKDC1 could serve as a prognostic biomarker for LUAD patients.\textsuperscript{26} Aldehyde dehydrogenase (ALDH) superfamily comprises 19 enzymes which play a vital role in maintaining the epithelial homeostasis. ALDH activity has been implicated in detoxification, cell proliferation, differentiation, drug resistance, and response to oxidative stress.\textsuperscript{27,28} Thus, deregulation of these enzymes could result in various cancers, including esophageal squamous cell carcinoma,\textsuperscript{29} and breast cancer.\textsuperscript{30} Giacalone \textit{et al.}\textsuperscript{31} reported that ALDH7A1, one of ALDH superfamily members, was correlated with overall survival and recurrence in patients with surgically resected stage I NSCLC. MDH1, an NAD(H)-dependent enzyme, is an important part of the malate/aspartate shuttle (MAS).\textsuperscript{32} This metabolic cycle contributes to maintain intracellular NAD(H) redox homeostasis as it transfers reducing equivalent NAD(H) across the mitochondrial membrane.\textsuperscript{32} It has been reported that abnormal expression of MDH1 was related to the tumor occurrence and progression.\textsuperscript{33} For example, MDH1 promoted pancreatic ductal adenocarcinoma cell proliferation and metabolism through NAD production to support glycolysis.\textsuperscript{33,34} Zhang \textit{et al.}\textsuperscript{35} reported that MDH1 expression was elevated in NSCLC tissue compared with normal lung tissue. However, there were no combination of these three glycolysis-related gene (HKDC1, ALDH7A1, and MDH1) to predict the prognosis of LUSC.

This study was the first to report that a glycolysis-based three-gene signature could serve as a prognostic indicator for patients with LUSC. A higher risk score indicates a worse prognosis. Of course, there still exist some limitations: First, the risk score model was constructed using TCGA database and should be verified in other cohorts in future researches. Second, studies on the three predicted genes should be made to explore the concrete mechanism in the occurrence and development of LUSC.

**Conclusions**

This study suggested that the three-gene signature associated with glycolysis might not only help to predict prognosis of LUSC patients, but also could provide more gene targets to cure LUSC patients.

**Abbreviations**

LUSC: lung squamous cell carcinoma; TCGA: The Cancer Genome Atlas; GSEA: Gene set enrichment analysis; HKDC1, hexokinase domain-containing protein 1; ALDH7A1: aldehyde dehydrogenase 7A1; MDH1: malate dehydrogenase 1; NSCLC: non-small cell lung cancer; SCLC: small cell lung cancer; LUAD:
Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Availability of data and material
All TCGA related data can be obtained from the TCGA database (https://tcga-data.nci.nih.gov/).

Competing interests
The authors declare that they have no conflicts of interest.

Funding
This research was supported by the program of “Workstation of Academician in the First People’s Hospital of Zunyi” funded by Zunyi Municipal Science and Technology Bureau (Zunshi Kehe (2015) No.17). The funders were not involved in the study design, data collection, analysis, manuscript writing, decision to publish, or reparation of the work.

Authors’ contributions
GH and DL contributed to the conception of the study. GH, JZ, LG, YH performed the data analyses. GH and JZ wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements
We thank the TCGA database for the valuable public data.

References
1. Osmani L, Askin F, Gabrielson E, Li QK. Current WHO guidelines and the critical role of immunohistochemical markers in the subclassification of non-small cell lung carcinoma (NSCLC): Moving from targeted therapy to immunotherapy. *Semin Cancer Biol.* 2018;52(Pt 1):103-109.
2. Molinier O, Goupil F, Debieuvre D, et al. Five-year survival and prognostic factors according to histology in 6101 non-small-cell lung cancer patients. *Respir Med Res.* 2019;77:46-54.

3. Pirker R. What is the best strategy for targeting EGF receptors in non-small-cell lung cancer? *Future Oncology (London, England).* 2015;11(1):153-167.

4. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* 2019;69(1):7-34.

5. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011;144(5):646-674.

6. Li Z, Zhang H. Reprogramming of glucose, fatty acid and amino acid metabolism for cancer progression. *Cell Mol Life Sci.* 2016;73(2):377-392.

7. SY L, cell VHMJAro, biology d. Aerobic glycolysis: meeting the metabolic requirements of cell proliferation. 2011;27:441-464.

8. Wang W, Wang S, Zhang M. Evaluation of kininogen 1, osteopontin and α-1-antitrypsin in plasma, bronchoalveolar lavage fluid and urine for lung squamous cell carcinoma diagnosis. *Oncology letters.* 2020;19(4):2785-2792.

9. Lou M, Gao Z, Zhu T, et al. TRIM59 as a novel molecular biomarker to predict the prognosis of patients with NSCLC. *Oncology letters.* 2020;19(2):1400-1408.

10. X W, G L, Q L, J X, international GCJCc. Integrated TCGA analysis implicates IncRNA CTB-193M12.5 as a prognostic factor in lung adenocarcinoma. 2018;18:27.

11. H G, Y Y, D W, Y H, OncoTargets TFJ, therapy. Potential role of LINC00996 in colorectal cancer: a study based on data mining and bioinformatics. 2018;11:4845-4855.

12. Tang Z, Li J, Shen Q, et al. Contribution of upregulated dipeptidyl peptidase 9 (DPP9) in promoting tumorigenicity, metastasis and the prediction of poor prognosis in non-small cell lung cancer (NSCLC). *Int J Cancer.* 2017;140(7):1620-1632.

13. Feng J, Zhang X, Zhu H, Wang X, Ni S, Huang J. FoxQ1 overexpression influences poor prognosis in non-small cell lung cancer, associates with the phenomenon of EMT. *PLoS One.* 2012;7(6):e39937.

14. Zhang L, Zhang Z, Yu Z. Identification of a novel glycolysis-related gene signature for predicting metastasis and survival in patients with lung adenocarcinoma. *J Transl Med.* 2019;17(1):423.

15. Liu C, Li Y, Wei M, Zhao L, Yu Y, Li G. Identification of a novel glycolysis-related gene signature that can predict the survival of patients with lung adenocarcinoma. *Cell Cycle.* 2019;18(5):568-579.

16. Abbaszadeh Z, Cesmeli S, Biray Avci C. Crucial players in glycolysis: Cancer progress. *Gene.* 2020;726:144158.

17. Warburg O. On respiratory impairment in cancer cells. *Science (New York, NY).* 1956;124(3215):269-270.

18. WH K, PL B, Cancer DCJNr. Otto Warburg’s contributions to current concepts of cancer metabolism. 2011;11(5):325-337.

19. Fu Q-F, Liu Y, Fan Y, et al. Alpha-enolase promotes cell glycolysis, growth, migration, and invasion in non-small cell lung cancer through FAK-mediated PI3K/AKT pathway. *Journal of Hematology & Oncology.* 2015;8:22.
20. Osugi J, Yamaura T, Muto S, et al. Prognostic impact of the combination of glucose transporter 1 and ATP citrate lyase in node-negative patients with non-small lung cancer. *Lung Cancer*. 2015;88(3):310-318.

21. B Z, Z X, B L. The clinicopathologic impacts and prognostic significance of GLUT1 expression in patients with lung cancer: A meta-analysis. 2019;689:76-83.

22. Irwin DM, Tan H. Molecular evolution of the vertebrate hexokinase gene family: Identification of a conserved fifth vertebrate hexokinase gene. *Comp Biochem Physiol Part D Genomics Proteomics*. 2008;3(1).

23. Fuhr L, El-Athman R, Scrima R, et al. The Circadian Clock Regulates Metabolic Phenotype Rewiring Via HKDC1 and Modulates Tumor Progression and Drug Response in Colorectal Cancer. *EBioMedicine*. 2018;33:105-121.

24. Zhang Z, Huang S, Wang H, et al. High expression of hexokinase domain containing 1 is associated with poor prognosis and aggressive phenotype in hepatocarcinoma. *Biochem Biophys Res Commun*. 2016;474(4):673-679.

25. Chen X, Lv Y, Sun Y, et al. PGC1β Regulates Breast Tumor Growth and Metastasis by SREBP1-Mediated HKDC1 Expression. *Front Oncol*. 2019;9:290.

26. Wang X, Shi B, Zhao Y, et al. HKDC1 promotes the tumorigenesis and glycolysis in lung adenocarcinoma via regulating AMPK/mTOR signaling pathway. *Cancer Cell Int*. 2020;20:450.

27. Muzio G, Maggiora M, Paiuzzi E, Oraldi M, Canuto RA. Aldehyde dehydrogenases and cell proliferation. *Free Radic Biol Med*. 2012;52(4):735-746.

28. Ahmed Laskar A, Younus H. Aldehyde toxicity and metabolism: the role of aldehyde dehydrogenases in detoxification, drug resistance and carcinogenesis. *Drug Metab Rev*. 2019;51(1):42-64.

29. Wang H, Tong L, Wei J, et al. The ALDH7A1 genetic polymorphisms contribute to development of esophageal squamous cell carcinoma. *Tumour Biol*. 2014;35(12):12665-12670.

30. Marcato P, Dean CA, Pan D, et al. Aldehyde dehydrogenase activity of breast cancer stem cells is primarily due to isoform ALDH1A3 and its expression is predictive of metastasis. *Stem Cells*. 2011;29(1):32-45.

31. Giacalone NJ, Den RB, Eisenberg R, et al. ALDH7A1 expression is associated with recurrence in patients with surgically resected non-small-cell lung carcinoma. *Future Oncol*. 2013;9(5):737-745.

32. Broeks MH, Shamseldin HE, Alhashem A, et al. MDH1 deficiency is a metabolic disorder of the malate-aspartate shuttle associated with early onset severe encephalopathy. *Hum Genet*. 2019;138(11-12):1247-1257.

33. Hanse EA, Ruan C, Kachman M, Wang D, Lowman XH, Kelekar A. Cytosolic malate dehydrogenase activity helps support glycolysis in actively proliferating cells and cancer. *Oncogene*. 2017;36(27):3915-3924.

34. Wang Y-P, Zhou W, Wang J, et al. Arginine Methylation of MDH1 by CARM1 Inhibits Glutamine Metabolism and Suppresses Pancreatic Cancer. *Mol Cell*. 2016;64(4):673-687.
35. Zhang B, Tornmalm J, Widengren J, Vakifahmetoglu-Norberg H, Norberg E. Characterization of the Role of the Malate Dehydrogenases to Lung Tumor Cell Survival. *J Cancer.* 2017;8(11):2088-2096.

**Tables**

**Table 1** Clinical characteristic of LUSC (n=501) from TCGA database

| Clinical characteristic | N  | %  |
|-------------------------|----|----|
| **Age**                 |    |    |
| ≤65                     | 190| 37.9|
| >65                     | 302| 60.3|
| NA                      | 9  | 1.8 |
| **Gender**              |    |    |
| Female                  | 130| 25.9|
| Male                    | 371| 74.1|
| **Clinical stage**      |    |    |
| I-II                    | 406| 81.0|
| III-IV                  | 91 | 18.2|
| NA                      | 4  | 0.8 |
| **T**                   |    |    |
| T1-T2                   | 407| 81.2|
| T3-T4                   | 94 | 18.8|
| **N**                   |    |    |
| N0                      | 319| 63.7|
| N1–3                    | 176| 35.1|
| NA                      | 6  | 1.2 |
| **M**                   |    |    |
| M0                      | 411| 82.0|
| M1                      | 7  | 1.4 |
| NA                      | 83 | 16.6|

**Abbreviations:** LUSC, lung squamous cell carcinoma; TCGA, The Cancer Genome Atlas; NA, not available.
Table 2 Gene set enriched in LUSC

| Gene sets follow link to MSigDB                      | Size | NES   | NOM p-val | FDR q-val |
|-----------------------------------------------------|------|-------|-----------|-----------|
| BIOCARTA_GLYCOLYSIS_PATHWAY                          | 3    | 1.43  | 0.029     | 0.029     |
| GO_GLYCOLYTIC_PROCESS                                | 106  | 2.01  | <0.0001   | <0.0001   |
| HALLMARK_GLYCOLYSIS                                  | 200  | 2.29  | <0.0001   | <0.0001   |
| KEGG_GLYCOLYSIS_GLUCONEOGENESIS                      | 62   | 1.56  | 0.028     | 0.028     |
| REACTOME_GLYCOLYSIS                                  | 72   | 2.28  | <0.0001   | <0.0001   |

**Abbreviations**: LUSC, lung squamous cell carcinoma; MSigDB, molecular signatures database; NES, normalized enrichment score; NOM p-val, nominal p-value; FDR q-val, false discovery rate q-value.

Table 3 Details of three genes for constructing the prognostic model

| Gene     | Ensemble ID      | Location                        | HR (95%CI) | Coefficient | p value |
|----------|------------------|---------------------------------|------------|-------------|---------|
| HKDC1    | ENSG00000156510  | chr10: 69,220,303-69,267,559    | 1.1733     | 0.1598      | 0.0446  |
| ALDH7A1  | ENSG00000164904  | chr5: 126,531,200-126,595,390   | 1.1701     | 0.1571      | 0.0097  |
| MDH1     | ENSG0000014641   | chr2: 63,588,609-63,607,197     | 0.7682     | -0.2636     | 0.0559  |

**Abbreviation**: HR, hazard ratio.

Table 4 Univariate and multivariate Cox regression analysis of clinicopathologic factors and glycolysis-related genes signature for OS
| Clinical features                  | Univariate analysis | Multivariate analysis |
|-----------------------------------|---------------------|-----------------------|
|                                   | HR      | 95%CI of HR | P value | HR      | 95%CI of HR | P value |
| Age (>65 vs. <=65)                | 1.376   | 1.000-1.892 | 0.050   | 1.489   | 1.078-2.055 | 0.016   |
| Gender (Male vs. Female)          | 1.360   | 0.944-1.958 | 0.099   | 1.300   | 0.901-1.877 | 0.161   |
| Clinical stage (III-IV vs. I-II)  | 1.392   | 0.974-1.989 | 0.070   | 1.185   | 0.716-1.962 | 0.510   |
| T (T3+T4 vs. T1+T2)              | 1.412   | 0.973-2.048 | 0.069   | 1.335   | 0.845-2.110 | 0.216   |
| M (M1 vs. M0)                    | 2.432   | 0.898-6.591 | 0.080   | 2.216   | 0.768-6.399 | 0.141   |
| N (N1+N2+N3 vs. N0)              | 1.062   | 0.780-1.444 | 0.704   | 1.087   | 0.765-1.545 | 0.642   |
| Risk score (high risk vs. low-risk) | 2.553   | 1.710-3.811 | <0.0001 | 2.663   | 1.790-3.962 | <0.0001 |

**Abbreviations:** OS, overall survival; HR, hazard ratio.

**Figures**
Figure 1

The flow chart of study.
Figure 1

The flow chart of study.
Figure 1

The flow chart of study.
Figure 2

GSEA results for enrichment plots of five gene sets which were significantly differentiated between in LUSC and normal tissues. (A), BIOCARTA_GLYCOLYSIS_PATHWAY; (B), GO_GLYCOLYTIC_PROCESS; (C), HALLMARK_GLYCOLYSIS; (D), KEGG_GLYCOLYSIS_GLUCONEOGENESIS; (E), REACTOME_GLYCOLYSIS

Abbreviations: GSEA, gene set enrichment analysis; LUSC, lung squamous cell carcinoma.
Figure 2

GSEA results for enrichment plots of five gene sets which were significantly differentiated between in LUSC and normal tissues. (A), BIOCARTA_GLYCOLYSIS_PATHWAY; (B), GO_GLYCOLYTIC_PROCESS; (C), HALLMARK_GLYCOLYSIS; (D), KEGG_GLYCOLYSIS_GLUCONEOGENESIS; (E), REACTOME_GLYCOLYSIS

Abbreviations: GSEA, gene set enrichment analysis; LUSC, lung squamous cell carcinoma.
Figure 2

GSEA results for enrichment plots of five gene sets which were significantly differentiated between in LUSC and normal tissues. (A), BIOCARTA_GLYCOLYSIS_PATHWAY; (B), GO_GLYCOLYTIC_PROCESS; (C), HALLMARK_GLYCOLYSIS; (D), KEGG_GLYCOLYSIS_GLUCONEOGENESIS; (E), REACTOME_GLYCOLYSIS

Abbreviations: GSEA, gene set enrichment analysis; LUSC, lung squamous cell carcinoma.

Figure 3
Differential expression of three genes in the normal tissues (n=49) and tumor tissues (n=501). (* p<0.05, **p<0.01, ***p<0.001) (A), ALDH7A1; (B), HKDC1; (C), MDH1.

Figure 3

Differential expression of three genes in the normal tissues (n=49) and tumor tissues (n=501). (* p<0.05, **p<0.01, ***p<0.001) (A), ALDH7A1; (B), HKDC1; (C), MDH1.
Figure 4

A risk of three-gene signature predicted the overall survival in patients with LUSC. (A), Distribution of risk score per patient; (B), Survival status of each patient; (C), A heatmap of three genes expression profile; (D), Kaplan-Meier survival curve analysis for LUSC patients divided into the high-risk and low-risk groups. Abbreviation: LUSC, lung squamous cell carcinoma.
Figure 4

A risk of three-gene signature predicted the overall survival in patients with LUSC. (A), Distribution of risk score per patient; (B), Survival status of each patients; (C), A heatmap of three genes expression profile; (D), Kaplan-Meier survival curve analysis for LUSC patients divided into the high-risk and low-risk groups. Abbreviation: LUSC, lung squamous cell carcinoma.
Figure 4

A risk of three-gene signature predicted the overall survival in patients with LUSC. (A), Distribution of risk score per patient; (B), Survival status of each patients; (C), A heatmap of three genes expression profile; (D), Kaplan-Meier survival curve analysis for LUSC patients divided into the high-risk and low-risk groups. Abbreviation: LUSC, lung squamous cell carcinoma.
Figure 5

The time-independent ROC curve of the risk score for prediction the 1, 3, 5-year overall survival.

Figure 5

The time-independent ROC curve of the risk score for prediction the 1, 3, 5-year overall survival.
Figure 5

The time-independent ROC curve of the risk score for prediction the 1, 3, 5-year overall survival.
Figure 6

Validation for prognostic value of the risk score in different subgroups. (A), Subgroups for age ≤ 65; (B) Subgroups for age > 65; (C), Subgroups for Female; (D) Subgroups for Male; (E), Subgroups for clinical stage I-II; (F), Subgroups for clinical stage III-IV; (G), Subgroups for T1+T2; (H), Subgroups for T3+T4; (I), Subgroups for M0; (J), Subgroups for M1; (K), Subgroups for N0; (L), Subgroups for N1+N2+N3.
Figure 6

Validation for prognostic value of the risk score in different subgroups. (A), Subgroups for age $\leq$ 65; (B) Subgroups for age $>$ 65; (C), Subgroups for Female; (D) Subgroups for Male; (E), Subgroups for clinical stage I-II; (F), Subgroups for clinical stage III-IV; (G), Subgroups for T1+T2; (H), Subgroups for T3+T4; (I), Subgroups for M0; (J), Subgroups for M1 ; (K), Subgroups for N0; (L), Subgroups for N1+N2+N3.
Figure 6

Validation for prognostic value of the risk score in different subgroups. (A), Subgroups for age≤65; (B) Subgroups for age>65; (C), Subgroups for Female; (D) Subgroups for Male; (E), Subgroups for clinical stage I-II; (F), Subgroups for clinical stage III-IV; (G), Subgroups for T1+T2; (H), Subgroups for T3+T4; (I), Subgroups for M0; (J), Subgroups for M1 ; (K), Subgroups for N0; (L), Subgroups for N1+N2+N3.

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure1.tif
- SupplementaryFigure1.tif
- SupplementaryFigure1.tif