Human sperm-associated antigen 4 as a potential prognostic biomarker of lung squamous cell carcinoma

**Yongheng Wang[1,2,]*, Yao Tang[3,]*, Jianhui Li[1,2],
Danfang Wang[1,4] and Wenhan Li[1,2]** (iD)

Abstract

Objective: Lung cancer (LC) is one of the most prevalent malignant tumors worldwide. As a subtype of LC, lung squamous cell carcinoma (LUSC) has a 5-year survival rate of less than 15%. In this study, we aimed to evaluate the prognostic value of a glycolysis-related gene signature in LUSC patients.

Methods: We obtained RNA-Seq data from The Cancer Genome Atlas (TCGA) database. Prognosis-related genes were screened out by Gene Set Enrichment Analysis (GSEA) and Cox proportional regression models. Quantitative reverse transcription polymerase chain reaction (RT-qPCR) was used to verify the mRNA expression levels in relevant tissues.

Results: We found that sperm-associated antigen 4 (SPAG4) overexpression was an independent risk factor for overall survival (OS) in LUSC. Patients with high-risk scores had higher mortality rates than those with low-risk scores. Moreover, by using RT-qPCR, we validated that SPAG4 mRNA was overexpressed in LUSC tissue samples compared with their paired para-cancerous histological normal tissues.

Conclusions: Analysis of aberrantly overexpressed SPAG4 may provide a further useful approach to complement existing methods and predict prognosis in LUSC patients.

[4]Department of Surgical Oncology, The Affiliated Hospital of Xi'an Medical University, Xi'an Medical University, Xi'an, Shaanxi, China

*These authors contributed equally to this work.

Corresponding author:
Wenhan Li, Department of Surgical Oncology, Shaanxi Provincial People's Hospital, No. 256 Yanta West Road, Xi'an, Shaanxi 710068, China.
Email: Dr_whli@hotmail.com

[1]Department of Surgical Oncology, Shaanxi Provincial People's Hospital, Xi'an, Shaanxi, China
[2]Department of Surgical Oncology, The Third Affiliated Hospital, School of Medicine, Xi'an Jiaotong University, Xi'an, Shaanxi, China
[3]Department of General Surgery, Xi'an No. 3 Hospital, Xi'an, Shaanxi, China

Introduction

Lung cancer (LC) has become the most commonly diagnosed malignancy and the leading cause of cancer-related death worldide.[1,2] Lung squamous cell carcinoma (LUSC) is a pathological subtype of LC with a less than 15% 5-year survival rate.[3] Unlike in lung adenocarcinoma, the application of molecular−targeted therapies in LUSC is very limited.[4,5] Accumulating evidence suggests that reprogrammed energy metabolism may be involved in tumorigenesis and development.[6,7] Aerobic glycolysis, also known as the "Warburg effect," is considered to be a hallmark of almost all cancer cells.[8] In this pathway, glucose in cancer cells is mainly converted to lactic acid regardless of oxygen availability. This process results in a highly acidic microenvironment, as well as glycolytic intermediates, that may contribute to malignant transformation and tumor progression.[9,10] However, few studies have explored the mechanism of glycolysis in LUSC. Moreover, there has been no attempt to analyze integrated database data to assess consistency and evaluate the magnitude of the clinical value of glycolysis in LUSC.[11] Thus, by using a bioinformatics approach and subsequent experimental verification, we sought to identify reliable glycolysis-related biomarkers for LUSC.

Sperm-associated antigen 4 (SPAG4) was first identified in mammalian sperm tails and plays an important role in spermatogenesis and sperm motility. Recent studies have demonstrated that SPAG4 may be involved in aerobic glycolysis and serve as a biomarker in glioblastoma,[12] colorectal cancer,[13] and pancreatic ductal adenocarcinoma.[14] Functional experiments revealed that the expression of SPAG4 is closely related to hypoxia in a hypoxia inducible factor (HIF)-1 and Von Hippel–Lindau (VHL)-dependent manner.[15] Moreover, Ying et al.[16] reported that SPAG4 exerts its tumor-promoting functions by interacting with Nesprin3 in lung carcinoma. However, the role of SPAG4 in LUSC is still not clear. In the present study, we extracted genome expression and corresponding clinical data from The Cancer Genome Atlas (TCGA) database. Using the Gene Set Enrichment Analysis (GSEA) method, we identified that SPAG4 is closely correlated with patients' survival outcome and may serve as an independent prognostic biomarker for LUSC. We further validated the expression levels of SPAG4 in LUSC patients by analyzing clinicopathological data in our center.

Materials and methods

Data collection

The mRNA expression profiles and clinical data were extracted from TCGA data portal (https://cancergenome.nih.gov/). A total of 551 samples (including 502 LUSC and 49 healthy controls) with sufficient data were included in our study.

Patients and clinical specimens

We retrospectively collected 13 tumor tissue samples and their paired para-cancerous histological normal tissues (PCHNTs) during curative surgery in our medical center, Shaanxi Provincial People's Hospital. None of the experimental subjects had received prior lung resection or preoperative chemotherapy/radiation therapy. All samples were immediately frozen and stored at $-80°C$ until total RNA was extracted. To reduce bias, samples were randomly coded before processing. This study strictly followed principles and guidelines for reporting preclinical research. All patients voluntarily joined this study with written informed consent to have their biologic specimens analyzed. This study was approved by the Ethical Committee of the Shaanxi Provincial People's Hospital on 12 November 2020 (reference number: 2020-026).

RNA extraction and quantitative reverse transcription polymerase chain reaction (RT-qPCR)

Total RNA was extracted from tissues using TRIzol reagent (Ambion, Thermo Fisher Scientific, Waltham, MA, USA). PrimeScript RT-polymerase (Takara, Kusatsu, Japan) was used to reverse-transcribe RNA into cDNA. RT-qPCR was performed using SYBR Premix Ex Taq^{TM} II (Tli RNaseH Plus) (Takara) with specific PCR primers by Sangon Biotech (Shanghai, China). GAPDH was used as an internal control. The $2^{-\Delta\Delta Ct}$ method was used to calculate relative fold changes. Primer sequences are as follows: SPAG4, forward 5′-ATGAGGATTTTG TGCGGAAG-3′ and reverse 5′-GCGTAA TCGTGGGATGTCTT-3′; GAPDH, forward 5′-TCGACAGTCAGCCGCATCTT CTTT-3′ and reverse 5′-ACCAAATCC GTTGACTCCGACCTT-3′. All reactions were performed in triplicate.

Gene set enrichment analysis

We performed gene set enrichment analysis (http://www.broadinstitute.org/gsea/index.jsp) to identify glycolysis-related genes in LUSC patients. The HALLMARK_GLYCOLYSIS gene set, REACTOME_GLYCOLYSIS gene set, KEGG_GLYCOLYSIS gene set, and BIOCARTA_GLYCOLYSIS_PATHWAY gene set were downloaded from the Molecular Signatures Database (https://www.gseamsigdb.org/gsea/msigdb/genesets.jsp). For each analysis, the number of permutations was set to 1000 times and we defined a false discovery rate (FDR) value <0.25 as statistically significant.

Statistical analyses

We used univariate Cox regression analysis to identify genes associated with patient overall survival (OS), which were then subjected to multivariable Cox regression analysis. Subsequently, we divided patients into high-risk and low-risk groups based on the median value of the gene expression level. Kaplan–Meier (KM) curves and the log-rank method were used to validate the prognostic significance of the risk score in stratified analysis. The sensitivity and specificity of the prognostic gene model were tested by receiver operating characteristic (ROC) curves. The above analyses were conducted using R software (version 3.6.1, http://www.rproject.org). Two-sided p values <0.05 were considered to be statistically significant. Prognosis-related gene alterations were analyzed by cBioPortal web software (http://www.cbioportal.org/). Immunohistochemistry (IHC) staining of SPAG4 protein was analyzed in the Human Protein Atlas (HPA) database (https://www.proteinatlas.org/).

Results

Clinical features and RNA sequencing data from 551 samples (including 502 LUSC and 49 healthy individuals) were obtained from TCGA database. Overall, the mean age of the LUSC patients was 67.27 ± 8.62 years and 132 (26.2%) patients were female. Additionally, 407 (81%) patients were in stage I–II. Four glycolysis-related gene sets (HALLMARK_GLYCOLYSIS, REACTOME_GLYCOLYSIS, KEGG_GLYCOLYSIS, and BIOCARTA_GLYCOLYSIS_PATHWAY) were obtained from the Molecular Signatures Database v4.0 (http://www.broadinstitute.org/gsea/msigdb/index.jsp). Then, GSEA was performed using the abovementioned data to explore differentially expressed genes between LUSC and healthy tissues. As shown in Figure 1, we found that all four gene sets were significantly enriched in the cancer samples and 236 glycolysis-related genes were identified for subsequent analysis.

We next conducted univariate and multivariate Cox proportional-hazards regression analyses to identify independent prognostic variables based on clinicopathological data. The results revealed that SPAG4 overexpression was an independent risk factor for OS in LUSC patients ($p = 0.04$, hazard ratio (HR): 1.038). We then divided those patients into two groups based on the median expression level of SPAG4. The distributions of risk score and survival status for each patient are displayed in Figure 2, suggesting that patients with high-risk scores had higher mortality rates than those with low-risk scores. Similarly, KM analysis was used to examine the prognostic effect of SPAG4 expression in the high-risk and low-risk groups (Figure 3a). We found that the high-risk group patients had significantly worse survival than those in the low-risk group ($p < 0.05$). ROC curve analysis was performed and the area under the curve (AUC) was 0.693 (Figure 3b). Additionally, univariate and multivariate analyses were performed to test the predictive ability of risk score with clinicopathological factors. We found that risk score (HR = 1.46; 95% confidence interval (CI) = 1.06–2.15; $p < 0.05$) and TNM stage (HR = 1.49; 95% CI = 1.06–2.09; $p < 0.05$) were two independent prognostic factors (Figure 4). The abovementioned results demonstrate moderate and independent prognostic power of our model in predicting LUSC prognosis.

We next analyzed the genetic alteration of our target gene via the cBioPortal database (http://cbioportal.org). The results suggest that the queried genes were altered in 12 (2.6%) of the sequenced cases (including nine case of gene amplification, two case of missense mutation, and one case of truncation mutation) (Figure 5a). In addition, we used RT-qPCR to validate the mRNA expression levels of SPAG4 in 13 LUSC patient tissue samples and their corresponding PCHNTs, with GAPDH used as the internal standard. Consistent with the bioinformatics results, SPAG4 mRNA expression levels were significantly higher in LUSC tissues than those in corresponding nontumorous tissues (n = 13, $p < 0.01$) (Figure 5b). IHC staining images from the HPA database (https://www.proteinatlas.org/) showed that SPAG4 intensity in LUSC tissues was stronger than that in normal pulmonary specimens (Figure 5c). The results shown here suggest that SPAG4 is overexpressed in LUSC tissue samples and therefore may be considered as a biomarker for LUSC patient prognosis.

Discussion

Recent studies have shown that metabolism deregulation is an emerging hallmark of cancer. In the 1920s, Warburg et al.[8]
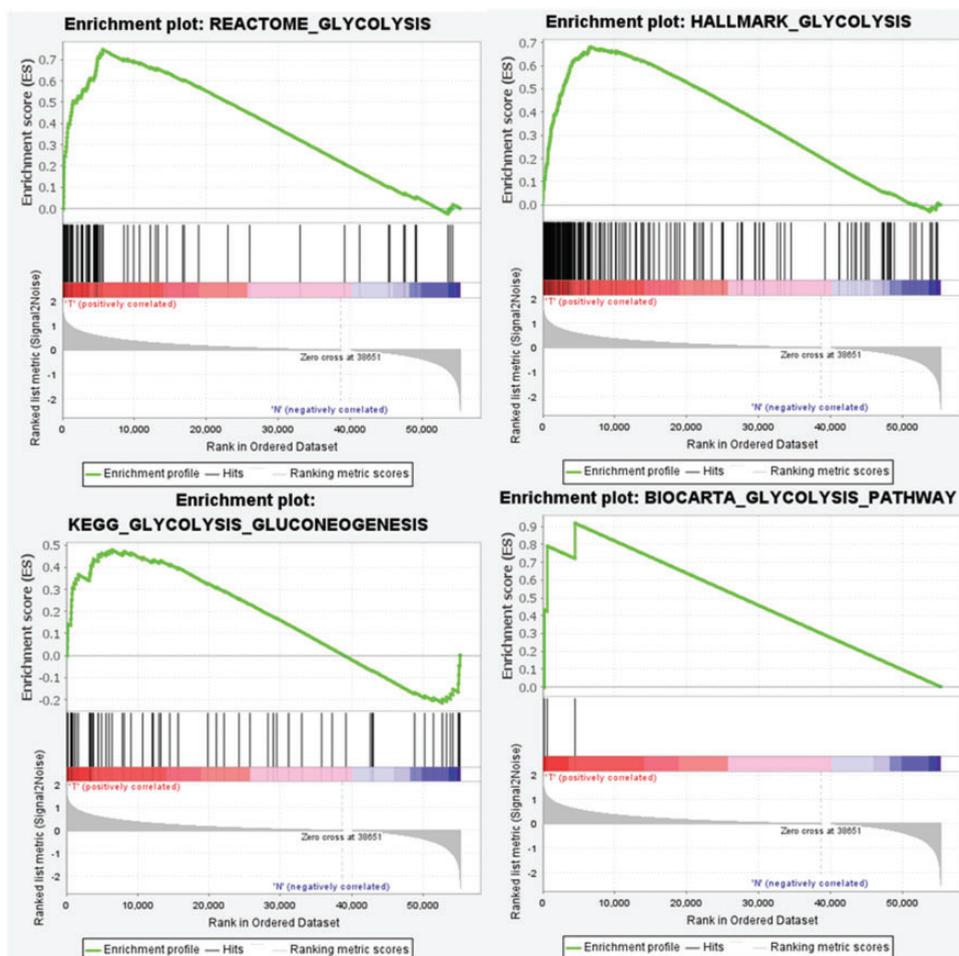
Figure 1. Enrichment plots of four glycolysis-related gene sets between lung squamous cell carcinoma and paired normal tissues identified by gene set enrichment analysis (GSEA).

found that tumor cells mainly rely on glycolysis for energy, even in the presence of sufficient oxygen. Although aerobic glycolysis is less efficient than glucose oxidation, it can provide energy more quickly to tumor cells, therefore providing them an advantage over stromal cells in the competition to use limited nutrients.[17] Moreover, as an intermediate product of this process, excessive accumulation of lactic acid creates an acidic microenvironment, driving tumor

invasion and metastasis and conferring resistance to radiation therapy.[9,10] Glycolysis-related gene signatures were also considered as prognostic prediction tools with favorable performance for multiple types of solid tumors. Increased aerobic glycolysis is also essential for tumor aggressiveness in LC cells.[18–20] However, studies concerning glycolysis-related biomarkers for LUSC are very limited. In this study, by using bioinformatics analysis and
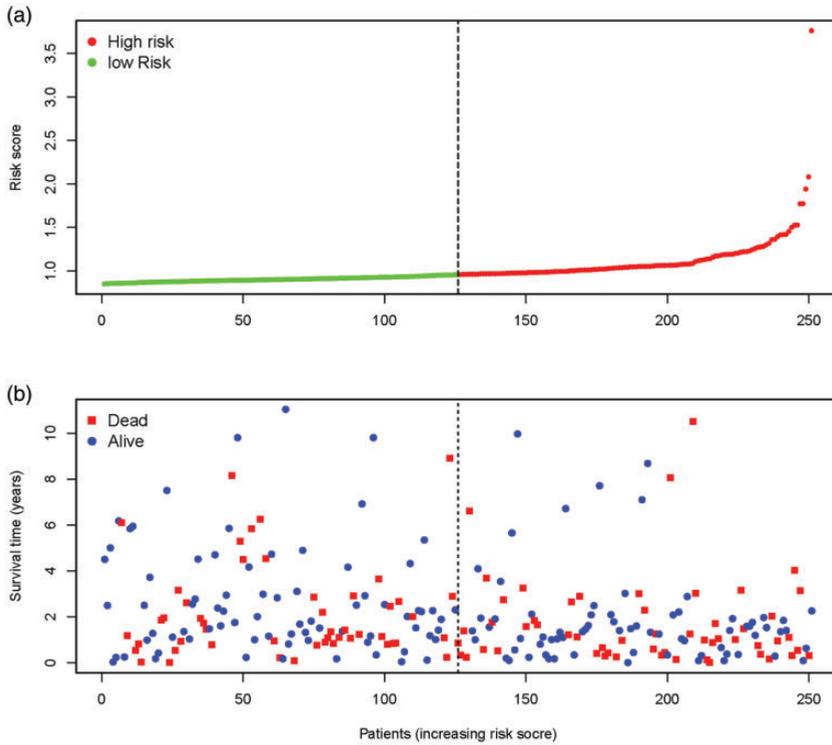
Figure 2. Glycolysis-related gene signature predicts overall survival (OS) in lung squamous cell carcinoma (LUSC) patients. (a) Distribution of risk scores per patient. (b) Relationship between survival days and survival status of each patient.
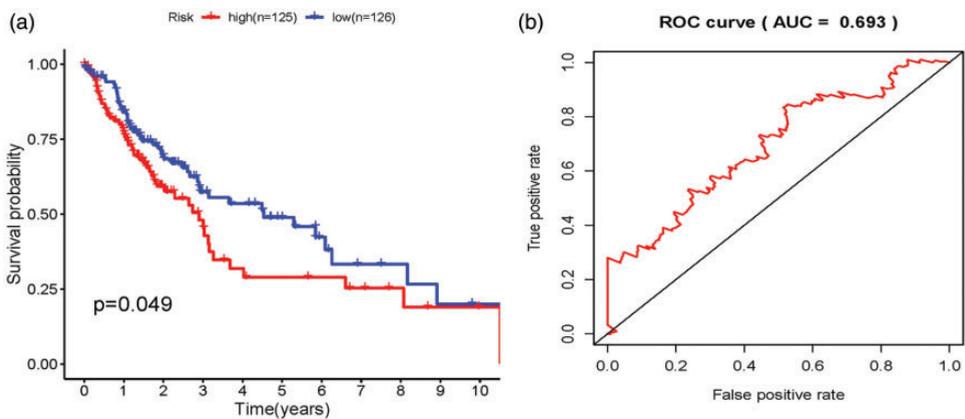

Figure 3. Glycolysis-related gene signature predicts overall survival (OS) in lung squamous cell carcinoma (LUSC) patients. (a) Kaplan–Meier (KM) curve to verify the predictive effect of the gene signature. (b) Receiver operating characteristic (ROC) curve analysis to evaluate the diagnostic efficacy of the gene signature.
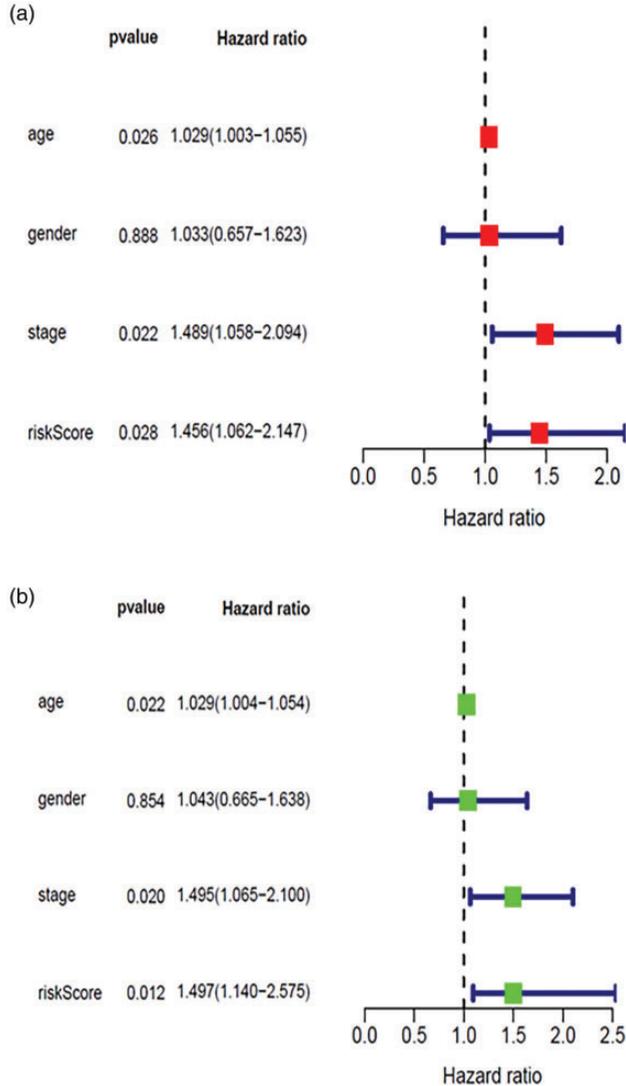
Figure 4. Univariable (a) and multivariable (b) analyses for the risk score and each clinical feature.

developing a statistical model, we sought to identify glycolysis-related biomarkers for the prognosis of LUSC patients.

In the current research, we performed GESA analysis to identify glycolysis-related genes based on TCGA mRNA expression data. We then used univariate and multivariate Cox regression analyses to evaluate the prognostic value of these identified genes. The results suggested that only SPAG4 was significantly correlated with LUSC patient prognosis. Subsequently, by using KM analysis and ROC curves, we found that the prognostic value of SPAG4 was acceptable. Additionally, the SPAG4 mRNA expression levels were examined in LUSC tissue samples from our medical center,
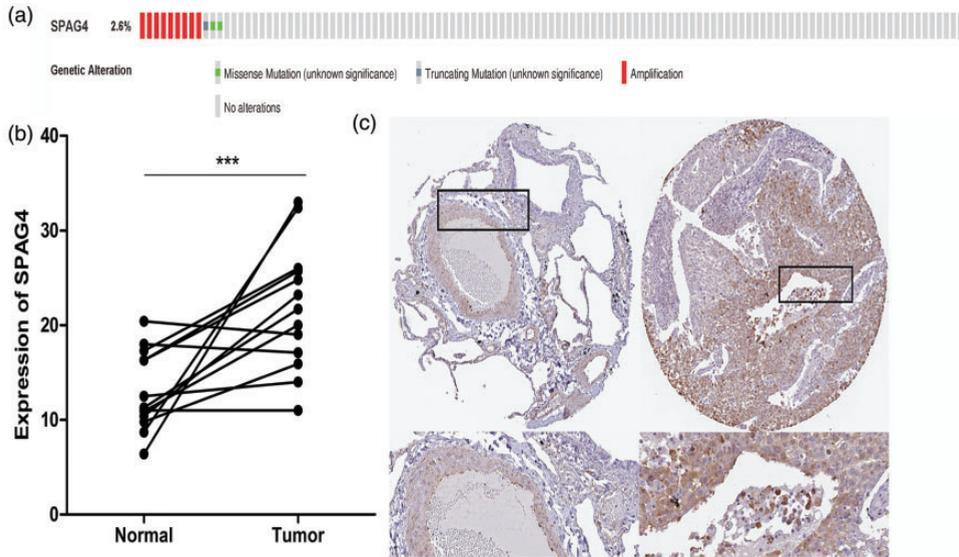
Figure 5. Identification of the sperm-associated antigen 4 (SPAG4) gene in lung squamous cell carcinoma (LUSC) samples. (A) The proportion of alteration for the SPAG4 gene in LUSC clinical samples in the cBioPortal database. (B) Expression of the SPAG4 in the LUSC samples (n = 13) and paired adjacent normal samples (n = 13) detected by quantitative real-time-PCR (qRT-PCR). (C) Immunohistochemistry (IHC) staining of SPAG4 protein in normal lung tissues (left) and LUSC tissues (right) in the Human Protein Atlas (HPA) database. ***$P < 0.001$.

suggesting its favorable performance for prediction.

As a member of SUN family, SPAG4 was originally identified as a testis-specific gene.[21,22] Previous studies showed that it may participate in nuclear remodeling, nuclear membrane integrity maintenance, and sperm tail development[23,24] However, aberrantly overexpressed SPAG4 was also involved in tumorigenesis. Zhao et al.[25] found that SPAG4 regulates glioblastoma progression by activating the MEK/ERK signaling pathway. Shoji et al.[25] demonstrated that SPAG4 plays a crucial role in cytokinesis to defend against hypoxia-induced tetraploid formation. Furthermore, Ji et al.[16] reported that SPAG4 acts as a positive regulator of Nesprin3. Overexpression of SPAG4 could affect the location and expression of Nesprin3, while reduced migration of tumor cells can be observed when Nesprin3 or SPAG4 are knocked out. To date, no study has reported the expression of SPAG4 in LUSC, nor its effect on the prognosis of this disease. In this study, we first investigated the expression levels and prognostic value of SPAG4 in LUSC using publicly available data, and then validated the specific role of SPAG4 in LUSC by experimental techniques. We found that SPAG4 may be a clinically relevant cancer marker for predicting poor survival of LUSC patients. Further experiments are required to elucidate its role in the formation and progression of LUSC.

Certain limitations should be taken into consideration when interpreting these study findings. First, selection bias is inevitable because of the retrospective nature of this study. Additionally, the sample size was relatively small and more clinical data are

required for further validation. Third, the combination of more gene profiles or clinicopathological features may offer more accurate predictive power than using a single biomarker. Mechanistic studies of aerobic glycolysis may allow for real-time surveillance of dynamic molecular and other changes accurately through the course of the disease, enable a better understanding of tumor progression, and guide individualized treatment.

Conclusion

Our study determined that a glycolysis-related gene, SPAG4, is aberrantly overexpressed in LUSC tissue samples compared with normal controls. Furthermore, elevated SPAG4 expression levels are associated with worse survival rates among LUSC patients. These findings may provide insight into the mechanisms of cellular glycolysis and help identify LUSC patients with poor prognosis.

ORCID iD

Wenhan Li https://orcid.org/0000-0002-9412-1166

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; 68: 394–424. doi: 10.3322/caac.21492.

2. Siegel RL, Miller KD and Jemal A. Cancer Statistics, 2017. *CA Cancer J Clin* 2017; 67: 7–30. doi: 10.3322/caac.21387.

3. Heist RS, Mino-Kenudson M, Sequist LV, et al. FGFR1 amplification in squamous cell carcinoma of the lung. *J Thorac Oncol* 2012; 7: 1775–1780. doi: 10.1097/JTO.0b013e31826aed28.

4. Chan BA and Hughes BG. Targeted therapy for non-small cell lung cancer: current standards and the promise of the future. *Transl Lung Cancer Res* 2015; 4: 36–54. doi: 10.3978/j.issn.2218-6751.2014.05.01.

5. Kris MG, Johnson BE, Berry LD, et al. Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. *JAMA* 2014; 311: 1998–2006. doi: 10.1001/jama.2014.3741.

6. Wu T and Dai Y. Tumor microenvironment and therapeutic response. *Cancer Lett* 2017; 387: 61–68. doi: 10.1016/j.canlet.2016.01.043.

7. Hanahan D and Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011; 144: 646–674. doi: 10.1016/j.cell.2011.02.013.

8. Warburg O. On the origin of cancer cells. *Science* 1956; 123: 309–314. doi: 10.1126/science.123.3191.309.

9. Brooks GA. Cell-cell and intracellular lactate shuttles. *J Physiol* 2009; 587: 5591–5600. doi: 10.1113/jphysiol.2009.178350.

10. Peppicelli S, Bianchini F and Calorini L. Extracellular acidity, a 'reappreciated' trait of tumor environment driving malignancy: perspectives in diagnosis and therapy. *Cancer Metastasis Rev* 2014; 33: 823–832. doi: 10.1007/s10555-014-9506-4.

11. Guo W, Kuang Y, Wu J, et al. Hexokinase 2 Depletion Confers Sensitization to Metformin and Inhibits Glycolysis in Lung Squamous Cell Carcinoma. *Front Oncol* 2020; 10: 52. doi: 10.3389/fonc.2020.00052.

12. Zhang C, Wang M, Ji F, et al. A Novel Glucose Metabolism-Related Gene Signature for Overall Survival Prediction in Patients with Glioblastoma. *Biomed Res Int* 2021; 2021: 8872977. doi: 10.1155/2021/8872977.

13. Zhu J, Wang S, Bai H, et al. Identification of Five Glycolysis-Related Gene Signature and Risk Score Model for Colorectal Cancer. *Front Oncol* 2021; 11: 588811. doi: 10.3389/fonc.2021.588811.

14. Tian G, Li G, Liu P, et al. Glycolysis-Based Genes Associated with the Clinical Outcome of Pancreatic Ductal Adenocarcinoma Identified by The Cancer Genome Atlas Data Analysis. *DNA Cell Biol* 2020; 39: 417–427. doi: 10.1089/dna.2019.5089.

15. Knaup KX, Monti J, Hackenbeck T, et al. Hypoxia regulates the sperm associated antigen 4 (SPAG4) via HIF, which is expressed in renal clear cell carcinoma and promotes migration and invasion in vitro. *Mol Carcinog* 2014; 53: 970–978. doi: 10.1002/mc.22065.

16. Ji Y, Jiang J, Huang L, et al. Sperm–associated antigen 4 (SPAG4) as a new cancer marker interacts with Nesprin3 to regulate cell migration in lung carcinoma. *Oncol Rep* 2018; 40: 783–792. doi: 10.3892/or.2018.6473.

17. Vander Heiden MG, Cantley LC and Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* 2009; 324: 1029–1033. doi: 10.1126/science.1160809.

18. Hua Q, Jin M, Mi B, et al. LINC01123, a c-Myc-activated long non-coding RNA, promotes proliferation and aerobic glycolysis of non-small cell lung cancer through miR-199a-5p/c-Myc axis. *J Hematol Oncol* 2019; 12: 91. doi: 10.1186/s13045-019-0773-y.

19. Zeng C, Wu Q, Wang J, et al. NOX4 supports glycolysis and promotes glutamine metabolism in non-small cell lung cancer cells. *Free Radic Biol Med* 2016; 101: 236–248. doi: 10.1016/j.freeradbiomed.2016.10.500.

20. Faubert B, Li KY, Cai L, et al. Lactate Metabolism in Human Lung Tumors. *Cell* 2017; 171: 358–371.e9. doi: 10.1016/j.cell.2017.09.019.

21. Tarnasky H, Gill D, Murthy S, et al. A novel testis-specific gene, SPAG4, whose product interacts specifically with outer dense fiber protein ODF27, maps to human chromosome 20q11.2. *Cytogenet Cell Genet* 1998; 81: 65–67. doi: 10.1159/000014990.

22. Calvi A, Wong AS, Wright G, et al. SUN4 is essential for nuclear remodeling during mammalian spermiogenesis. *Dev Biol* 2015; 407: 321–330. doi: 10.1016/j.ydbio.2015.09.010.

23. Pasch E, Link J, Beck C, et al. The LINC complex component Sun4 plays a crucial role in sperm head formation and fertility. *Biol Open* 2015; 4: 1792–1802. doi: 10.1242/bio.015768.

24. Zhao J, Liu B, Yang JA, et al. Human sperm-associated antigen 4 as a potential biomarker of glioblastoma progression and prognosis. *Neuroreport* 2019; 30: 446–451. doi: 10.1097/wnr.0000000000001226.

25. Shoji K, Murayama T, Mimura I, et al. Sperm-associated antigen 4, a novel hypoxia-inducible factor 1 target, regulates cytokinesis, and its expression correlates with the prognosis of renal cell carcinoma. *Am J Pathol* 2013; 182: 2191–2203. doi: 10.1016/j.ajpath.2013.02.024.