The Prognostic Value of a Glycolysis -Related lncRNA Signature in Non-Small Cell Lung Cancer

Yongquan Dong

Ningbo Yinzhou No 2 Hospital

Wenping Xia (✉ wenpingx2014@163.com)

Ningbo Yinzhou No 2 Hospital

Hefei Zhu

Ningbo Yinzhou No 2 Hospital

Feijie Lu

Ningbo Yinzhou No 2 Hospital

Lijuan Wei

Ningbo Yinzhou No 2 Hospital

Qianjun Li

Ningbo Yinzhou No 2 Hospital

Qiong Zhao

affiliated to Shulan international hospital college of Zhejiang University

Research Article

Keywords: Long non-coding RNA(lncRNA) , Non-small cell lung cancer(NSCLC), glycolysis-related lncRNAs, prognosis

Posted Date: August 30th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-819919/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Long non-coding RNA(lncRNA) is a new modulatory factor for glycolysis, which affect the development and progression of varieties of cancers. However, poor attention has been drawn to the role of glycolysis-related lncRNAs in Non-small cell lung cancer(NSCLC) .

Methods: The RNA sequencing(RNA-seq) data of NSCLC patients was downloaded from The Cancer Genome Atlas(TCGA) database. Glycolysis-related genes were identified using Gene Set Enrichment Analysis(GSEA). Cox regression analysis was used to construct the glycolysis-related signature. GSEA analysis was used for function assessment.

Results: Firstly, a signature consisting of 8 glycolysis-related lncRNAs(AC090559.1, AC099850.3,AL365181.3, AL049555.1, AC024075.3, LINC01843, ATP13A4-AS1 and LINC01133)were estimated. Moreover, on the basis of this signature, we found that overall survival(OS) in high-risk group was markedly lower than that in low-risk group, and the prognostic prediction accuracy was further verified by Multi-parameter ROC curve analysis. Lastly, the 8 glycolysis-related lncRNAs were found to be closely associated with many cancer-related pathways.

Conclusion: The 8 glycolysis-related lncRNAs could be promising prognostic and potential therapeutic targets for NSCLC.

Introduction

Lung cancer is one of the most common tumor-related death in the world[1]. Non-small cell lung cancer(NSCLC) consists of squamous cell carcinoma, adenocarcinoma and larger cell carcinoma. Although the advancement in treatment, including surgery resection, chemotherapy, radiotherapy, the prognosis of lung cancer remains not satisfied, due to its recurrence, metastasis and lack of effective tumor biomarkers[2]. Therefore, identifying efficient biomarkers for its early diagnosis become urgent.

Oxidative phosphorylation and aerobic glycolysis(Warburg effect) are the main ways for human cells to obtain energy[3]. The “Warburg effect” showed that the rapid growth of cancer cells mainly relied on glycolysis to produce energy[4]. Increasing studies have shown that glycolysis genes were significantly dysregulated in various cancers, and closely associated with development and progression of cancers[5, 6].

Long non-coding RNA(lncRNA) is a type of non-coding RNAs with more than 200 nucleotides(nt) in length does not have the capability of protein-coding[7]. Increasing evidences showed that lncRNAs have been reported to be involved in aerobic glycolysis, which result in distant metastasis and chemoresistance of cancer cells[8–10].In addition, lncRNAs play a critical role in multiple biological processes in many cancer types and may be utilized as a potential biomarker to predict the prognosis of cancers patients[11–13].

In our study, we analyzed the lncRNAs of NSCLC patients in The Cancer Genome Atlas(TCGA) database, and identified glycolysis-related lncRNAs associated with the prognosis of NSCLC, and estimated a novel glycolysis-related lncRNA prognostic signature for NSCLC. Our work might open up new ideas for the diagnosis of NSCLC.

Materials And Methods

Sample sources and processing

The RNA sequencing data of NSCLC and corresponding clinical features were downloaded from TCGA database(<https://portal.gdc.cancer.gov/>). The glycolysis-related gene expression profiles were obtained from gene set enrichment analysis(GSEA). We set $FDR < 0.05$ and $|\log_2 FC| > 0.5$ as threshold to identify the differentially expressed glycolysis-related genes and lncRNAs using edge R package.

Pearson correlation analysis was used to glycolysis-related lncRNAs. The selection criteria were $|R| > 0.5$ and $p < 0.05$. Univariate cox regression analysis was employed prognosis-related lncRNAs.

Construction of glycolysis-related lncRNAs prognostic signature

To increase the reliability of our work, we randomly separated the entire dataset into a training cohort and a testing cohort using “caret” R package. The least absolute shrinkage and selection operator(LASSO) cox regression analysis was utilized to construct the glycolysis-related lncRNAs prognostic signature. The formula of the risk score of each NSCLC was calculated as follows:

Risk score = $\sum \text{Exp}_i * \beta_i$ (β_i represents the coefficient of each lncRNA and Exp_i represents the expression of each lncRNA)

The median value of risk score served as a cutoff value to divide the NSCLC patients into low risk group and high risk group. Kaplan-Meier curve analysis was employed to detect the overall survival(OS) probability of NSCLC patients between low risk group and high risk group. The “survival ROC” R package was conducted to draw the Receiver Operating Characteristic (ROC) curves to assess the accuracy of the signature.

Gene set enrichment analysis

Genome-wide expression profiles of NSCLC patients were analyzed using gene set enrichment(GSEA) to investigate the signaling pathways and biological process. between low-risk group and high-risk group. Gene sets used in our study were c2.cp.kegg.v7.4.symbols.gmt. The normal(NOM) $p < 0.05$ as well as false discovery rate(FDR) $< 25\%$ were considered noteworthy.

Statistical analysis

The data were processed and analyzed by using R software(Version 4.0.3,<https://www.r-project.org/>) and perl data language. The $p < 0.05$ was considered to be statistically significant.

Results

Screening of glycolysis-related genes

We used GSEA method through five glycolysis-associated gene sets to explore whether there were significant differences between NSCLC samples and normal samples. Four gene sets were found to be enriched in NSCLC samples(Figure 1A-1D). All the genes in these four gene sets were integrated and 295 glycolysis-related genes were obtained for subsequent analysis(Table S1). Then we set false discovery rate(FDR) < 0.05 and $|\log_2 FC| > 0.5$ as the thresholds to identify the differentially expressed glycolysis-related genes, and found that 111 differentially expressed glycolysis-related genes(Table S2).

Coexpression network of glycolysis-related lncRNA for NSCLC

518 lncRNAs were found to be differentially expressed between NSCLC and normal tissues (Table S3), and 109 glycolysis-associated lncRNAs were used to construct the coexpression network by using Spearman's correlation test ($|R| > 0.5$, $p < 0.001$) (Table S4, Figure 2).

Identification of prognostic lncRNA signature in NSCLC

To identify prognostic lncRNAs of NSCLC patients, and 14 lncRNAs were found to be associated with prognosis of NSCLC using univariate cox analysis (Figure 3A). To develop a lncRNA signature for prognostic prediction of NSCLC, LASSO cox regression was performed on these 14 prognostic lncRNAs, we obtained 8 lncRNAs to estimate the risk model, and the coefficients of these lncRNAs were employed to calculate the risk score (Figure 3B). The risk score = $AC090559.1 \times (-0.037) + AC099850.3 \times 0.019 + AL365181.3 \times 0.018 + AL049555.1 \times 0.015 + AC024075.3 \times (-0.016) + LINC01843 \times 0.030 + ATP13A4-AS1 \times (-0.023) + LINC01133 \times 0.002$. Then the NSCLC patients in the training cohort were separated into high-risk group and low-risk group based on cutoff of the median risk score. As shown in Figure 3C, high-risk patients had obviously lower overall survival (OS) than low-risk patients. The risk curve and scatterplot were employed to display the risk score as well as the relevant survival status of NSCLC patients. Our results showed that the mortality accuracy relied on the risk score (Figure 3D, 3E). The heatmap displayed the expression of glycolysis-related lncRNA in the risk model (Figure 3F). Through univariate and multivariate cox regression analysis, the risk score of this signature could be an independent predictor of prognosis in NSCLC patients (Figure 3G). Multi-parameter ROC curve analysis revealed that the AUC value of the signature was 0.710 (Figure 3H).

Validation of prognostic lncRNA signature in NSCLC

Next, the signature was verified by using testing cohort. Consistent with the results from training cohort, the OS of NSCLC patients in the low-risk group was better than the high-risk group on the basis of the signature (Figure 4A). The distribution of risk score and survival status of NSCLC patients was shown in Figure 4B, 4C. The expression of signature-related lncRNAs in the low-risk and high-risk groups were shown using heatmap (Figure 4D). Predicting the prognosis of NSCLC patients on the basis of this signature, the ROC curve analysis displayed a good predictive capability of the signature (Figure 4E). The univariate and multivariate cox regression analysis revealed that the risk score was associated with worse OS of NSCLC patients (Figure 4F). These results verified that the lncRNA signature was an independent prognostic factor for NSCLC patients.

Subgroup analysis

We further separated NSCLC patients into different subgroups in accordance with clinical features, including age, gender, T stage, N stage, M stage and TNM stage. The results revealed that patients with low-risk had a significantly better OS than patients with high-risk in most subgroups except for M1 subgroups (Figure 5). It showed that the lncRNA signature was suitable for multiple categories of NSCLC patients.

Gene set enrichment analysis

To investigate the biological function of the signature in NSCLC patients, we determined a total of 50 gene sets by using GSEA, which were markedly enriched with FDR q value < 0.25 and $p < 0.05$. The results showed that 46 pathways were obviously enriched in the high-risk group (Table S5), some of them are cancer-related pathways, such as cell cycle, p53 signaling pathway, DNA replication, Glycolysis/Gluconeogenesis pathway (Figure 6). Our

results revealed that these glycolysis-related lncRNAs contributed to cancer-related pathways, which might provide the evidence for the targeted therapy of NSCLC.

Discussion

Previous studies have shown that glycolysis plays an critical role in the occurrence and development of many malignancies[14–16]. In our study, 296 glycolysis-related genes were obtained though using GSEA analysis and 111 differentially expressed glycolysis-related genes were identified. Accumulating studies had shown the importance of lncRNAs in NSCLC, such as tumor suppressor, tumor carcinogenic functions and prognostic biomarkers[17–19]. Recently, lncRNAs were emerging as important regulators in aerobic glycolysis[20, 21]. Therefore, it was urgent to investigate the role of glycolysis-related lncRNAs in NSCLC.

Honghao Cao, et al. [22]constructed a four glycolysis-related lncRNA signature and the signature might be used to predict the survival in renal carcinoma patients. Yang Bai, et al.[23] Estimated a four glycolysis-related signature, which might contribute to clinical strategies as well as clinical outcome prediction of hepatocellular carcinoma. Herein, we conducted to this work to estimate an glycolysis-related lncRNA signature in order to predict the prognosis of NSCLC patients by using the TCGA database.

In our study, we TCGA data sets were collected and divided into two groups(training group and testing group). In the training group, glycolysis-related lncRNAs were conducted to explore the prognosis of NSCLC patients. We also identified a signature of glycolysis-associated lncRNAs and separated the NSCLC patients into low-risk and high-risk groups on the basis of median risk score. Kaplan-Meier curve and ROC analysis determined the prognostic value of this signature in the training group. The similar results were also verified in the testing group. Furthermore, univariate and multivariate cox regression analyses were performed and revealed that the signature could be an independent prognostic factor for NSCLC.

To further prove the prognostic value of the signature associated with glycolysis, NSCLC patients were classified by age(≤ 65 or > 65), gender(male or female), T stage(T1-2 or T3-4), N stage(N0 or N1-3), M stage(M0 or M1) and TNM stage(I-II or III-IV), and found that patients with low-risk had a significantly better OS than patients with high-risk in most subgroups except for M1 subgroups.

To explore role of glycolysis-related lncRNAs in tumor behavior, GSEA method was performed. The results showed that the differentially expressed genes in the high-risk group were involved in many cancer-related pathways, such as cell cycle, p53 signaling pathway, DNA replication, Glycolysis/Gluconeogenesis pathway. These results may provide a new target for NSCLC therapy.

Conclusion

In summary, our study suggested that the 8 glycolysis-related lncRNA signature might not only contribute to predict the prognosis of NSCLC patients, but also could be the new targets that could be potentially utilized to cure NSCLC patients.

Abbreviations

lncRNA:Long non-coding RNA; NSCLC:Non-small cell lung cancer; OS:overall survival; GSEA:Gene set enrichment analysis; LASSO: least absolute shrinkage and selection operator.

Declarations

Acknowledgements

Not applicable

Author contribution

Conception and design: YQD, WPX and QZ. Data collection and analysis: WPX, HFZ, FJL and LJW. Drafting: YQD, QJL. Review and proof-reading: QZ. All authors agree to the submitted version on your respected journal.

Funding

This work was supported by the Public welfare projects of Zhejiang Province (LGF19H010009)

Availability of data and materials

The datasets used or/and analyzed during this study are available from the corresponding author.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not application

Competing interesting

None

References

1. RL Siegel, KD Miller, A Jemal. Cancer statistics, 2019. CA Cancer J Clin 2019, 69:7-34.
2. S Chheang, K Brown. Lung cancer staging: clinical and radiologic perspectives. Semin Intervent Radiol. 2013, 30:99-113.
3. J Lu, M Tan, Q Cai. The Warburg effect in tumor progression: mitochondrial oxidative metabolism as an anti-metastasis mechanism. Cancer Lett. 2015, 356:156-64.
4. DL Rothman, RG Shulman. Two transition states of the glycogen shunt and two steady states of gene expression support metabolic flexibility and the Warburg effect in cancer. Neoplasia. 2021, 23:879-886.
5. B Ou, Y Liu, X Yang, X Xu, Y Yan, J Zhang. C5aR1-positive neutrophils promote breast cancer glycolysis through WTAP-dependent m6A methylation of ENO1. Cell Death Dis. 2021, 12:737.

6. S Lin, Y Li, D Wang, C Huang, D Marino, O Boltt, C Wu, MD Taylor, W Li, GM DeNicola, et al. Fascin promotes lung cancer growth and metastasis by enhancing glycolysis and PFKFB3 expression. *Cancer Lett.* 2021, 518:230-242.
7. F Kopp, JT Mendell. Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell.* 2018, 172:393-407.
8. F Chen, J Chen, L Yang, J Liu, X Zhang, Y Zhang, Q Tu, D Yin, D Lin, PP Wong, et al. Extracellular vesicle-packaged HIF-1alpha-stabilizing lncRNA from tumour-associated macrophages regulates aerobic glycolysis of breast cancer cells. *Nat Cell Biol.* 2019, 21:498-510.
9. P Malakar, I Stein, A Saragovi, R Winkler, N Stern-Ginossar, M Berger, E Pikarsky, R Karni. Long Noncoding RNA MALAT1 Regulates Cancer Glucose Metabolism by Enhancing mTOR-Mediated Translation of TCF7L2. *Cancer Res.* 2019, 79:2480-2493.
10. J Chen, Y Yu, H Li, Q Hu, X Chen, Y He, C Xue, F Ren, Z Ren, J Li, et al. Long non-coding RNA PVT1 promotes tumor progression by regulating the miR-143/HK2 axis in gallbladder cancer. *Mol Cancer.* 2019, 18:33.
11. Z Xue, C Cui, Z Liao, S Xia, P Zhang, J Qin, Q Guo, S Chen, Q Fu, Z Yin, et al. Identification of LncRNA Linc00513 Containing Lupus-Associated Genetic Variants as a Novel Regulator of Interferon Signaling Pathway. *Front Immunol.* 2018, 9:2967.
12. Q Han, L Xu, W Lin, X Yao, M Jiang, R Zhou, X Sun, L Zhao. Long noncoding RNA CRCMSL suppresses tumor invasive and metastasis in colorectal carcinoma through nucleocytoplasmic shuttling of HMGB2. *Oncogene.* 2019, 38:3019-3032.
13. P Zhang, G Liu, L Lu. N6-Methyladenosine-Related lncRNA Signature Is a Novel Biomarkers of Prognosis and Immune Response in Colon Adenocarcinoma Patients. *Front Cell Dev Biol.* 2021, 9:703629.
14. XS Liu, LM Zhou, LL Yuan, Y Gao, XY Kui, XY Liu, ZJ Pei. NPM1 Is a Prognostic Biomarker Involved in Immune Infiltration of Lung Adenocarcinoma and Associated With m6A Modification and Glycolysis. *Front Immunol.* 2021, 12:724741.
15. B Ou, Y Liu, X Yang, X Xu, Y Yan, J Zhang: C5aR1-positive neutrophils promote breast cancer glycolysis through WTAP-dependent m6A methylation of ENO1. *Cell Death Dis.* 2021, 12:737.
16. S Wang, N Wang, Y Zheng, B Yang, P Liu, F Zhang, M Li, J Song, X Chang, Z Wang. Caveolin-1 inhibits breast cancer stem cells via c-Myc-mediated metabolic reprogramming. *Cell Death Dis.* 2020, 11:450.
17. C Wang, X Meng, Y Zhou, J Yu, Q Li, Z Liao, Y Gu, J Han, S Linghu, Z Jiao, et al. Long Noncoding RNA CTD-2245E15.3 Promotes Anabolic Enzymes ACC1 and PC to Support Non-Small Cell Lung Cancer Growth. *Cancer Res* 2021, 81:3509-3524.
18. RQ Wang, XR Long, NN Zhou, DN Chen, MY Zhang, ZS Wen, LJ Zhang, FZ He, ZL Zhou, SJ Mai, et al. Lnc-GAN1 expression is associated with good survival and suppresses tumor progression by sponging mir-26a-5p to activate PTEN signaling in non-small cell lung cancer. *J Exp Clin Cancer Res.* 2021, 40:9.

19. MM Wei, GB Zhou. Long Non-coding RNAs and Their Roles in Non-small-cell Lung Cancer. *Genomics Proteomics Bioinformatics*. 2016, 14:280-288.
20. Y Liu, D He, M Xiao, Y Zhu, J Zhou, K Cao. Long noncoding RNA LINC00518 induces radioresistance by regulating glycolysis through an miR-33a-3p/HIF-1alpha negative feedback loop in melanoma. *Cell Death Dis*. 2021, 12:245.
21. M Liao, W Liao, N Xu, B Li, F Liu, S Zhang, Y Wang, S Wang, Y Zhu, D Chen, et al. LncRNA EPB41L4A-AS1 regulates glycolysis and glutaminolysis by mediating nucleolar translocation of HDAC2. *Ebiomedicine*. 2019, 41:200-213.
22. H Cao, H Tong, J Zhu, C Xie, Z Qin, T Li, X Liu, W He. A Glycolysis-Based Long Non-coding RNA Signature Accurately Predicts Prognosis in Renal Carcinoma Patients. *Front Genet*. 2021, 12:638980.
23. Y Bai, H Lin, J Chen, Y Wu, S Yu. Identification of Prognostic Glycolysis-Related lncRNA Signature in Tumor Immune Microenvironment of Hepatocellular Carcinoma. *Front Mol Biosci*. 2021, 8:645084.

Figures

Figure 1

KEGG pathway enrichment analysis of glycolysis-related genes obtained from GSEA. A. Biocarta glycolysis pathway. B. Glycolytic process pathway. C. Hallmark glycolysis. D. Reactome glycolysis.

Figure 2

Network of lncRNAs and co-expressed glycolysis-related genes in NSCLC. The red node represents glycolysis-related genes, and the green represents lncRNAs. The co-expression network was conducted using igraph package.

Figure 3

Prognostic value of glycolysis-related lncRNAs in the training cohort.(A) Relationship between 14 glycolysis-related lncRNAs and the prognosis of NSCLC patients. (B). Lasso cox regression analysis of 8 glycolysis-related lncRNAs. (C). OS analysis for NSCLC in low/high risk group. (D). The distribution of risk score between low-risk group and high-risk group. (E). The scatter plot showed the associations with risk score, survival status and survival time.(F). Heatmap of glycolysis-related lncRNA expression.(G). univariate cox regression analysis and multivariate cox regression analysis of clinical features related to OS in the training cohort. (H). ROC curve analysis exhibits the prognostic accuracy of clinicopathological factors and glycolysis-related lncRNA prognostic risk score.

Figure 4

Prognostic value of glycolysis-related lncRNAs in the testing cohort.(A). OS analysis for NSCLC in low/high risk group. (B). The distribution of risk score between low-risk group and high-risk group. (C). The scatter plot showed the associations with risk score, survival status and survival time.(D). Heatmap of glycolysis-related lncRNA expression.(E). univariate cox regression analysis and multivariate cox regression analysis of clinical features related to OS in the testing cohort. (F). ROC curve analysis exhibits the the prognostic accuracy of clinicopathological factors and glycolysis-related lncRNA prognostic risk score.

Figure 5

Kaplan-Meier curve analysis for the low-risk group and high-risk group classified by clinical features including Age, Gender, T stage, N stage, M stage and TNM stage.

Figure 6

GSEA analysis of the glycosis-related lncRNA signature in the high-risk group.

Supplementary Files

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

- [TableS1.xlsx](#)
- [TableS2.xls](#)

- [TableS3.xlsx](#)
- [TableS4.xlsx](#)
- [TableS5.xlsx](#)