A hypoxia-related gene pairs signature for predicting overall survival in lung adenocarcinoma

Kang Li

First People's Hospital of Wenling

Bo Zhang (✉ zhangbo_wenling@163.com)

First People's Hospital of Wenling <https://orcid.org/0000-0002-8944-1034>

Research

Keywords: lung adenocarcinoma, prognosis, hypoxia-related gene pairs

Posted Date: April 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-361989/v1>

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

[Read Full License](#)

Abstract

Background: Increasing evidence has depicted the clinical importance of the correlation between hypoxia and immune status in lung adenocarcinoma environment (LUAD). However, the reliable prognostic signatures based on the interaction of hypoxia and immune status are still limited. Therefore, we strived to construct a hypoxia-immune-related gene pairs signature for risk assessment and stratification for patients with LUAD.

Methods: Gene expression profiles and clinical data of patients with lung adenocarcinoma were acquired from two public data sets, used as training and validation cohorts respectively. Different bioinformatics and statistical methods were applied to construct a robust hypoxia-related gene pairs (HRGPs) signatures for predicting overall survival in lung adenocarcinoma. Furthermore, we explored the correlation between HRGPs signature and infiltrating immune cells using the CIBERSORT algorithm in LUAD samples.

Results: Among 146 hypoxia-related genes, a 13 HRGPs signature was built that was significantly associated with OS in the training cohort ($P < 0.01$). In the validation cohort, the HRGPs signature stratified patients into high- and low-risk groups with a significant OS difference ($P = 0.04$). After adjusting the other clinical factors, the developed HRGPs signature remained independent in multivariate analysis. Plasma cells and NK cells resting were significantly correlated with the OS of patients with LUAD. Functional analysis showed that the high-risk group was enriched in terms of DNA replication, lymphocyte apoptotic process, and cell cycle-related pathways.

Conclusion: The HRGPs signature represents a promising risk model for patient's stratification in LUAD. It might provide new insights into clinical decision-making regarding individualized treatment.

Background

LUAD is the most common subtype of non-small cell lung cancer (NSCLC)[1]. Currently, the standard treatment for individual LUAD patients is based on a combination of surgery, postoperative chemotherapy and radiotherapy[2]. Despite significant progress in treatment, the prognosis for LUAD remains dismal[3]. Conventional survival classifiers for LUAD patients, such as differentiation grade and tumor-node-metastasis (TNM) staging, only considers anatomical and clinicopathological factors[4]. However, LUAD is a heterogeneous malignancy, and its prognosis can be significantly different even for patients with similar clinical features, suggesting that current classifiers are insufficient[5]. Therefore, development of novel markers with higher predictive value is urgently needed.

Hypoxia, or lack of oxygen, is a characteristic feature of most solid tumor[6]. Increasing evidence has found that the hypoxic environment in tumors directly stimulated the malignant properties of cancers, including advanced aggressiveness and resistance to therapies[7-10]. Meanwhile, the infiltration of immune cells in the tumor is also a crucial component that can affect the tumor invasion and metastasis[11]. Interestingly, there was a direct or indirect interaction of immune status and hypoxia in the lung cancer microenvironment, although their underlying mechanisms remain unclear[9, 12].

Researchers have identified several molecular subtypes of LUAD based on gene expression data, which were used as new markers for survival prediction[13-15]. However, due to the large-scale gene expression data, none of them have been applied to daily clinical practice. Recently, a new method based on relative ranking of gene expression levels to eliminate the shortcomings of data scaling and normalization has been reported, and has produced robust results in various studies[16, 17].

In this study, we explore the correlation between hypoxia and immune in LUAD, and identified an individualized prognostic signature for LUAD by incorporating HRGPs, aimed to improve the prognostication of LUAD patients.

Results

Construction of the prognostic HRGPs signature

A total of 146 HRGs were measured on all platforms and fulfilled the criteria of the training cohort ($MAD > 0.5$). From these 146 HRGs, 1645 HRGPs were constructed. After removing HRGPs with relatively small variation ($MAD=0$), 677 HRGPs were kept as candidates. Then we developed the HRGPs index using Lasso Cox regression on the training cohort and selected 13 HRGPs in the final risk model (**Figure1 A-B**). This HRGPs signature consisted of 24 unique HRGs (**Table 2**). Next, we used the HRGPs index to calculate the risk score for each patient in the training cohort. The patients were stratified into high- or low-risk groups based on the optimal cut-off point which was set at 0.101 using time-dependent ROC curve analysis (**Figure 1C**). The survival curves suggested that the high-risk group presented a significantly poor OS than the low-risk group in the training cohort ($P < 0.001$, **Figure 2A**). We further evaluated the prognostic power of HRGPs signature based on other clinical factors in the TCGA-LAUD dataset using the univariate and multivariate Cox proportional hazards regression analysis. The results of the univariate analysis showed that the clinical stage and risk model presented a prognostic effect on the OS of patients with LUAD (**Figure 3A**). After adjusting for other clinical variables, the risk model remained an independent prognostic factor in multivariate analysis ($HR=3.3178$, 95% CI 2.351-4.296, $P < 0.001$, **Figure 3B**).

Validation of the HRGPs signature for prognostic prediction

To validate the constant prognostic value of HRGPs signature, we applied the same formula to other independent populations. GSE31210 ($n=226$) served as an external validation cohort and were divided into high- and low-risk groups based on the optimal cut-off value (HRGPs index= 0.101). when compared the survival difference between these two groups, patients in the high-risk group had a poorer OS than those in the low-risk group ($P=0.04$, **Figure 2B**). The risk model also remained independent from other clinical variables in univariate and multivariate Cox analyses (**Figure 3C-D**). Besides, The HRGPs signature presented a higher C-index when compared with other clinically applicable biomarkers, including RNA-binding proteins signature, immune-related signature, and autophagy-related signature (**Figure 4**) [13, 18, 19].

Correlation of the HRGPs signature and immune cells infiltration

Increasing evidence has been shown that tumor-infiltrating immune cells had an impact on tumors' prognosis[20, 21]. We applied the CIBERSORT algorithm to calculate the abundance of 22 types of immune cells for each sample within different risk groups. **Figure 5A** depicted that several immune cells, such as plasma cells, T-cells CD4 memory activated, T cells follicular helper, T cells regulatory (Tregs), NK cells resting, NK cells activated, Macrophages M0, dendritic cells resting, mast cells resting, mast cells activated, and neutrophils, were significantly enriched between different risk groups ($P < 0.05$). Next, we performed the survival curves based on these immune cells. The results showed the plasma cells and NK cells resting were significantly correlated with the OS of patients with LUAD ($P < 0.05$, **Figure 5B-C**).

Functional analysis of the HRGPs signature

To characterize the biological process determined by the HRGPs signature, we performed the GO analysis and GSEA. The results of the GO analysis showed that the HRGPs signature was mainly enriched in the terms of cornification, glycolmetabolism, and chromosome-related process (**Figure 6**). GSEA demonstrated that DNA replication, lymphocyte apoptotic process, and cell cycle-related pathways were significantly altered in high-risk groups ($FDR < 0.05$, **Figure 7**).

Discussion

As a hallmark of solid tumors, hypoxia is an imbalance between increased oxygen demand and oxygen supply inadequate, which is associated with high proliferative rates[22]. Previously, numerous studies reported that hypoxic phenotypes in solid tumors caused various biological processes, including malignant metastasis, angiogenesis, metabolic alteration, and inducing resistance to treatment[23-25]. These findings support that hypoxia has prognostic value and can be a therapeutic target for tumors. In recent years, some new prognostic biomarkers for LUAD have already appeared, which were used to select patients who are at high-risk and who may benefit from the personalized therapy. However, their accuracy of survival estimation remains uncertain. Therefore, we strived to identify a robust prognostic model for LUAD patients by detecting the HRGs systematically and comprehensively.

Considering the technical bias caused by sequencing platforms and biological heterogeneity of LUAD, previous risk models need to standardize gene expression data, which is difficult for clinical practice. Instead, we used a new model to perform the comparing of gene expression without addressing the technical bias of different platforms[26]. This method has been considered to have reliable results in many studies, including cancer molecular classification[27, 28]. The developed prognostic model is achieved through pairing comparison of gene expression levels and relative ranking, rather than data pretreatments, such as normalization and scaling.

In this study, we found that both the immune status and hypoxia of the tumor microenvironment were associated with the survival of LUAD patients. Moreover, the adverse effects of hypoxia status were significantly correlated with prognosis, even after adjusting clinicopathologic risk factors. Finally, an

HRGPs signature consisted of 24 unique HRGs was constructed as a prognosis classifier, with high predictive value in risk stratification both in the training cohort and validation cohort. Most of the genes involved in HRGPs are related to oxygen metabolism. Within these 24 HRGs, elevated expression of NDRG1 contributed to chemoresistance and cancer growth in NSCLC[29]. EFNA1 (ephrinA1) is a member of the ENF family, and it has been studied that over-expression of EFNA1 induced hypoxic environments and pro-angiogenic function in various cancer cells[30]. Decreased expression of FBP1 was associated with cell epithelial-mesenchymal transition, invasion, and metastasis, which present poor prognosis in prostate cancer[31]. Plasma cells had been considered to be related to better survival in diverse cancer[32]. We found a significantly elevated infiltrating of plasma cells in the low-risk group, which was associated to better OS. Besides, the function of NK cells resting in LUAD prognosis remains unknown. In our study, the abundance of NK cells resting was high in the high-risk group and correlated with a poorer prognosis for patients with LUAD. Moreover, we also found that glycometabolism, chromosome-related process, and cell cycle-related pathways were related to HRGPs signature. These pathways were well-known risk factors that affect tumor progression[33-35]. These results suggest the crucial role of HRGPs signature in the development of LUAD patients.

Several limitations exist in this study. First, the prognostic signature is based on the gene expression profiles produced by RNA-seq or microarray platforms, which is difficult to popularize in daily clinical applications due to its high price, long conversion cycle, and requirement of bioinformatics expertise. Second, considering that the microenvironment characteristics might be distinct in different tumor regions, such as tumor core and invasive margin. Samples used for analyses were all collected from the core of the tumor, and it is impossible to evaluate the immune and hypoxia status in different tumor regions. Therefore, findings in this study are waiting for further validation by multicenter, well-designed, prospective studies.

Conclusion

In conclusion, we systematically explored the prognostic value of the HRGPs, which could assess the risk stratification and modify the immune status for LUAD patients. These biological functions of the hypoxia-related genes provide a deep insight into the role of our signature in the progression of LUAD. This signature might serve as a prognostic classifier for clinical decision-making regarding individualized treatment and follow-up schedule.

Materials And Methods

Data collection and processing

In total, two independent public datasets were used to do this comprehensive analysis. The RNA-seq data of 494 LUAD patients were downloaded from The Cancer Genome Atlas (TCGA) data portal (<https://portal.gdc.cancer.gov/>) and used as the training cohort. GSE31210 was the other RNA-seq data of 226 LUAD patients obtained from the Gene Expression Omnibus (GEO) and served as a validation cohort[36].

For TCGA-LUAD datasets, the fragments per kilobase of exon per million fragments mapped (FPKM) value was used to measure all the mRNA expression. For GSE31210 datasets, the original expression levels were normalized using the robust multi-array average (RMA) method from the “affy” R package[37]. The corresponding clinicopathological information of included cohorts was listed in **Table 1**.

Development and validation of HRGPs signature for LUAD

The cases from TCGA-LUAD datasets were used to develop the prognostic model. The hallmark gene sets of hypoxias comprised of 200 genes were collected from the molecular signature (MSigDB version 6.0) of the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (<https://www.kegg.jp/>; ID:04066), including 15 genes involved in “Increase oxygen delivery” and 11 genes related to “Reduce oxygen consumption”. We measured the hypoxia-related genes (HRGs) on all the cohorts included in this study and with relatively high variation (determined by median absolute deviation > 0.5). Next, each HRGP was computed by pairwise comparison of the gene expression level for every sample. The output is 1 if the first gene is larger than the later one and 0 for the reverse order. We removed the HRGPs with a low variation (median absolute deviation, MAD =0) and kept those remaining as initial candidate HRGPs signature for prognosis prediction. Finally, from the initial candidate HRGPs, an index was calculated using the Lasso Cox proportional hazards regression with 10-fold cross-validation (by “glmnet” R package)[38]. As a result, the selected gene pairs were used to construct the final prognostic model. Based on the optimal cut-off value of the HRGPs index determined by the time-dependent receiver operating characteristic (ROC) curve (using “survivalROC” R package) at three years for overall survival (OS), the patients were then stratified into high- and low-risk groups.

The prognostic capability of the HRGPs signature was further validated in GSE31210 datasets using a log-rank test. Besides, we also assessed the independent predictive value of HRGPs signature from other clinical factors by univariate and multivariate Cox analysis in each cohort.

Correlation of HRGPs with the tumor-infiltration immune cells

Immune cells are the main components of the tumor microenvironment and correlated to the prognosis of malignancy. We strived to estimates the relative abundance of various immune cells in different risk groups using the CIBERSORT algorithm, which was a machine learning method based on support vector regression in gene expression data of each TCGA-LUAD sample[39]. The gene expression data of TCGA were uploaded to the online analytical platform CIBERSORT portal website(<http://cibersort.stanford.edu/>), using the default signature matrix at 1,000 permutations. CIBERSORT quantified the proportions of 22 kinds of infiltration immune cells, including B cells, T cells, neutrophils, and macrophages, amongst others.

Gene functional enrichment analysis

To elucidate the molecular function of the HRGPs signature, we conducted the Gene Ontology (GO) annotations using the “clusterProfiler” (version 3.12.0) R package[40]. Gene set enrichment analysis

(GSEA) was performed using the Bioconductor package “fgsea” with 10,000 permutations[41]. The whole RNA-seq of all TCGA-LUAD samples was used for GSEA based on the C5 collection in the molecular signature database. The differential gene sets from high- and low-risk groups were compared, and those with FDR $q < 0.05$ and NOM $p < 0.05$ were defined as significant.

Statistical analysis

All the statistical tests were performed using R (version 3.6.0). The differences among groups were compared using the student’s t-tests. Survival analysis was conducted using the Kaplan-Meier method and analyzed using the log-rank test from the “survival” package. The Cox proportional-hazards regression model was used to perform the univariate and multivariate analyses. The concordance (C)-index was applied to test the predictive value of signature using the “survcomp” package[42]. The primary endpoint was OS and the p-value of two-sided less than 0.05 was considered as a statistically significant difference.

Declarations

Acknowledgements

The authors gratefully acknowledge the Gene Expression Omnibus (GEO) database, and the Cancer Genome Atlas (TCGA) database, which made the data available.

Authors' Contributions: Study design: Kang Li and Bo Zhang; Data extraction and analysis, as well as manuscript writing: Kang Li; Manuscript edition: Kang Li and Bo Zhang. All authors read and approved the final manuscript.

Funding: This work was supported by the Wenling Science and Technology Project (No. 2015C3BA0057).

Availability of data and materials: The raw data of this study are derived from the TCGA database (<https://portal.gdc.cancer.gov/>) and GEO data portal (<https://www.ncbi.nlm.nih.gov/geo/>), which are publicly available databases.

Ethics approval and consent to participate

Not necessary.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests

References

1. Siegel RL, Miller KD, Jemal A: **Cancer statistics, 2019**. *CA Cancer J Clin* 2019, **69**(1):7-34.
2. Tseng JS, Hsu KH, Zheng ZR, Yang TY, Chen KC, Huang YH, Su KY, Yu SL, Chang GC: **Primary Tumor Resection Is Associated with a Better Outcome among Advanced EGFR-Mutant Lung Adenocarcinoma Patients Receiving EGFR-TKI Treatment**. *Oncology-Basel* 2020:1-9.
3. Hua X, Zhao W, Pesatori AC, Consonni D, Caporaso NE, Zhang T, Zhu B, Wang M, Jones K, Hicks B *et al*: **Genetic and epigenetic intratumor heterogeneity impacts prognosis of lung adenocarcinoma**. *Nat Commun* 2020, **11**(1):2459.
4. Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman J, Chirieac LR, D'Amico TA, DeCamp MM, Dilling TJ, Dobelbower M *et al*: **Non-Small Cell Lung Cancer, Version 5.2017, NCCN Clinical Practice Guidelines in Oncology**. *J Natl Compr Canc Netw* 2017, **15**(4):504-535.
5. Wang B, Chen L, Huang C, Lin J, Pan X, Shao Z, Hu S, Zhang X, Wang X: **The homogeneous and heterogeneous risk factors for occurrence and prognosis in lung cancer patients with bone metastasis**. *J Bone Oncol* 2019, **17**:100251.
6. Ruan K, Song G, Ouyang G: **Role of hypoxia in the hallmarks of human cancer**. *Journal of cellular biochemistry* 2009, **107**(6):1053-1062.
7. Steinberger KJ, Forget MA, Bobko AA, Mihalik NE, Gencheva M, Roda JM, Cole SL, Mo X, Hoblitzell EH, Evans R *et al*: **Hypoxia-Inducible Factor alpha Subunits Regulate Tie2-Expressing Macrophages That Influence Tumor Oxygen and Perfusion in Murine Breast Cancer**. *J Immunol* 2020.
8. Sureshbabu SK, Chaukar D, Chiplunkar SV: **Hypoxia regulates the differentiation and anti-tumor effector functions of gammadeltaT cells in oral cancer**. *Clin Exp Immunol* 2020, **201**(1):40-57.
9. Vito A, El-Sayes N, Mossman K: **Hypoxia-Driven Immune Escape in the Tumor Microenvironment**. *Cells* 2020, **9**(4).
10. Shao C, Yang F, Miao S, Liu W, Wang C, Shu Y, Shen H: **Role of hypoxia-induced exosomes in tumor biology**. *Mol Cancer* 2018, **17**(1):120.
11. Edlund K, Madjar K, Mattsson JSM, Djureinovic D, Lindskog C, Brunnstrom H, Koyi H, Branden E, Jirstrom K, Ponten F *et al*: **Prognostic Impact of Tumor Cell Programmed Death Ligand 1 Expression and Immune Cell Infiltration in NSCLC**. *J Thorac Oncol* 2019, **14**(4):628-640.
12. Terry S, Engelsen AST, Buart S, Elsayed WS, Venkatesh GH, Chouaib S: **Hypoxia-driven intratumor heterogeneity and immune evasion**. *Cancer Lett* 2020, **492**:1-10.
13. Guo D, Wang M, Shen Z, Zhu J: **A new immune signature for survival prediction and immune checkpoint molecules in lung adenocarcinoma**. *J Transl Med* 2020, **18**(1):123.
14. 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**. *J Transl Med* 2020, **18**(1):149.
15. Li W, Gao LN, Song PP, You CG: **Development and validation of a RNA binding protein-associated prognostic model for lung adenocarcinoma**. *Aging (Albany NY)* 2020, **12**(4):3558-3573.

16. Zhao E, Zhou C, Chen S: **A signature of 14 immune-related gene pairs predicts overall survival in gastric cancer.** *Clinical & translational oncology : official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico* 2020.
17. Wu J, Zhao Y, Zhang J, Wu Q, Wang W: **Development and validation of an immune-related gene pairs signature in colorectal cancer.** *Oncoimmunology* 2019, **8**(7):1596715.
18. 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.
19. Zhou M, Shao W, Dai H, Zhu X: **A Robust Signature Based on Autophagy-Associated LncRNAs for Predicting Prognosis in Lung Adenocarcinoma.** *Biomed Res Int* 2020, **2020**:3858373.
20. Tian X, Xu W, Wang Y, Anwaier A, Wang H, Wan F, Zhu Y, Cao D, Shi G, Zhu Y *et al*: **Identification of tumor-infiltrating immune cells and prognostic validation of tumor-infiltrating mast cells in adrenocortical carcinoma: results from bioinformatics and real-world data.** *Oncoimmunology* 2020, **9**(1):1784529.
21. Zhang Y, Zhang Z: **The history and advances in cancer immunotherapy: understanding the characteristics of tumor-infiltrating immune cells and their therapeutic implications.** *Cell Mol Immunol* 2020, **17**(8):807-821.
22. Brahimi-Horn MC, Chiche J, Pouyssegur J: **Hypoxia and cancer.** *J Mol Med (Berl)* 2007, **85**(12):1301-1307.
23. Zhang Q, Huang R, Hu H, Yu L, Tang Q, Tao Y, Liu Z, Li J, Wang G: **Integrative Analysis of Hypoxia-Associated Signature in Pan-Cancer.** *iScience* 2020, **23**(9):101460.
24. Yaghoubi S, Najminejad H, Dabaghian M, Karimi MH, Abdollahpour-Alitappeh M, Rad F, Mahi-Birjand M, Mohammadi S, Mohseni F, Sobhani Lari M *et al*: **How hypoxia regulate exosomes in ischemic diseases and cancer microenvironment?** *IUBMB Life* 2020, **72**(7):1286-1305.
25. Yu T, Tang B, Sun X: **Development of Inhibitors Targeting Hypoxia-Inducible Factor 1 and 2 for Cancer Therapy.** *Yonsei Med J* 2017, **58**(3):489-496.
26. Eddy JA, Sung J, Geman D, Price ND: **Relative expression analysis for molecular cancer diagnosis and prognosis.** *Technology in cancer research & treatment* 2010, **9**(2):149-159.
27. Huo J, Wu L, Zang Y: **Development and Validation of a Novel Immune-Gene Pairs Prognostic Model Associated with CTNNB1 Alteration in Hepatocellular Carcinoma.** *Medical science monitor : international medical journal of experimental and clinical research* 2020, **26**:e925494.
28. Xu F, Zhan X, Zheng X, Xu H, Li Y, Huang X, Lin L, Chen Y: **A signature of immune-related gene pairs predicts oncologic outcomes and response to immunotherapy in lung adenocarcinoma.** *Genomics* 2020.
29. Du A, Jiang Y, Fan C: **NDRG1 Downregulates ATF3 and Inhibits Cisplatin-Induced Cytotoxicity in Lung Cancer A549 Cells.** *Int J Med Sci* 2018, **15**(13):1502-1507.
30. Hao Y, Li G: **Role of EFNA1 in tumorigenesis and prospects for cancer therapy.** *Biomed Pharmacother* 2020, **130**:110567.

31. Zhang YP, Liu KL, Yang Z, Lu BS, Qi JC, Han ZW, Yin YW, Zhang M, Chen DM, Wang XW *et al*: **The involvement of FBP1 in prostate cancer cell epithelial mesenchymal transition, invasion and metastasis by regulating the MAPK signaling pathway.** *Cell Cycle* 2019, **18**(19):2432-2446.
32. Yeong J, Lim JCT, Lee B, Li H, Chia N, Ong CCH, Lye WK, Putti TC, Dent R, Lim E *et al*: **High Densities of Tumor-Associated Plasma Cells Predict Improved Prognosis in Triple Negative Breast Cancer.** *Front Immunol* 2018, **9**:1209.
33. Huang Y, Yang X, Sun F, Lu T, Bi G, Liang J, Sui Q, Zhan C, Shi Y, Wang Q: **Prognostic effects of glycometabolism changes in lung adenocarcinoma: a prospective observational study.** *Transl Lung Cancer Res* 2019, **8**(6):808-819.
34. Eifler K, Vertegaal ACO: **SUMOylation-Mediated Regulation of Cell Cycle Progression and Cancer.** *Trends Biochem Sci* 2015, **40**(12):779-793.
35. Kar S: **Unraveling Cell-Cycle Dynamics in Cancer.** *Cell Syst* 2016, **2**(1):8-10.
36. Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraishi K, Iwakawa R, Furuta K, Tsuta K, Shibata T, Yamamoto S *et al*: **Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas.** *Cancer Res* 2012, **72**(1):100-111.
37. Gautier L, Cope L, Bolstad BM, Irizarry RA: **affy-analysis of Affymetrix GeneChip data at the probe level.** *Bioinformatics* 2004, **20**(3):307-315.
38. Engebretsen S, Bohlin J: **Statistical predictions with glmnet.** *Clin Epigenetics* 2019, **11**(1):123.
39. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA: **Robust enumeration of cell subsets from tissue expression profiles.** *Nature methods* 2015, **12**(5):453-457.
40. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, Vilo J: **g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update).** *Nucleic Acids Res* 2019, **47**(W1):W191-W198.
41. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *et al*: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci U S A* 2005, **102**(43):15545-15550.
42. Schroder MS, Culhane AC, Quackenbush J, Haibe-Kains B: **survcomp: an R/Bioconductor package for performance assessment and comparison of survival models.** *Bioinformatics* 2011, **27**(22):3206-3208.

Tables

Table 1. Clinicopathological characteristics of LUAD patients from training and validation cohort.

Characteristics	Training cohort (n=494)		Validation cohort(n=226)	
Age(years)	Number	%	Number	%
median	65 (33-88)		61 (30-76)	
< 65	220	44.5	164	72.6
≥ 65	274	55.5	62	27.4
Gender				
Female	266	53.8	121	53.5
Male	228	46.2	105	46.5
Stage				
Stage1-2	380	78.2	226	100
Stage3-4	106	21.8	0	0
Smoking				
No			115	50.9
Yes			111	49.1

Table 2. Information about HRGPs signature.

Gene pairs	Coefficient
NDRG1 EFNA1	0.170172901835233
NDRG1 FBP1	0.0634434919635711
HK2 GPC4	0.0982613517143679
ANGPTL4 SDC3	0.0940138014698999
JUN DDIT4	-0.0490110591526602
SERPINE1 PPP1R15A	0.104245702792977
DDIT4 CXCR4	0.532067084001365
EGFR GPC4	0.257720738258149
CDKN1A SELENBP1	0.0621460488363104
MIF FBP1	0.00804772592535753
KLF6 FBP1	0.330691556790575
GPC4 TPD52	-0.064536325936734
NEDD4L AKAP12	-0.49933149687839

Figures

Figure 1

Identification of HRGPs signature in the training cohort. (A) The LASSO coefficient profiles of prognostic HRGPs signature with the optimal λ value of minimum criteria; (B) The LASSO Cox regression proportional model was used to construct the robust signature; The partial likelihood deviance is plotted with the $\log(\lambda)$ and error bars representing standard error (SE); The dotted vertical lines are depicted at the

Figure 3

The HRGPs signature was identified as independent risk factor for LUAD patients. Univariate and multivariate Cox regression analyses of different prognostic factors for OS in the training cohort (A-B) and the validation cohort (C-D).

Figure 4

C-index of different established signature for LUAD. The hypoxia-related signature showed the higher C-index (0.718) compared to RNA-binding proteins signature (0.660, PMID: 32071816), immune-related signature (0.655, PMID: 32143735), and autophagy-related signature (0.653, PMID: 32190662), representing a better predictive capability for LUAD.

A

B

Plasma cells(p=0.015)

C

NK cells resting(p=0.038)

Figure 5

The association of infiltrating immune cells and the HRGPs signature in LUAD samples. (A) The abundance of 22 types of immune cells within different risk groups. A red violin represents the high-risk group and a green violin represents the low-risk group. The white dots inside the violin depicts median values. (B-C) Plasma cells and NK cells resting are significantly correlated to OS in patients with LUAD.

Figure 6

GO functional analysis of the 24 hypoxia-related genes in HRGPs signature.

Figure 7

Gene set enrichment analysis showed the upregulated pathways in the high-risk group.