


PRIMARY RESEARCH

Open Access



A novel 4-gene signature for overall survival prediction in lung adenocarcinoma patients with lymph node metastasis

Yanfang Wang^{1†}, Quanli Zhang^{2†}, Zhaojia Gao³, Shan Xin^{1,4}, Yanbo Zhao⁵, Kai Zhang⁵, Run Shi^{1*} and Xuanwen Bao^{6,7*} 

Abstract

Background: Lung adenocarcinoma (LUAD) patients experiencing lymph node metastasis (LNM) always exhibit poor clinical outcomes. A biomarker or gene signature that could predict survival in these patients would have a substantial clinical impact, allowing for earlier detection of mortality risk and for individualized therapy.

Methods: With the aim to identify a novel mRNA signature associated with overall survival, we analysed LUAD patients with LNM extracted from The Cancer Genome Atlas (TCGA). LASSO Cox regression was applied to build the prediction model. An external cohort was applied to validate the prediction model.

Results: We identified a 4-gene signature that could effectively stratify a high-risk subset of these patients, and time-dependent receiver operating characteristic (tROC) analysis indicated that the signature had a powerful predictive ability. Gene Set Enrichment Analysis (GSEA) showed that the high-risk subset was mainly associated with important cancer-related hallmarks. Moreover, a predictive nomogram was established based on the signature integrated with clinicopathological features. Lastly, the signature was validated by an external cohort from Gene Expression Omnibus (GEO).

Conclusion: In summary, we developed a robust mRNA signature as an independent factor to effectively classify LUAD patients with LNM into low- and high-risk groups, which might provide a basis for personalized treatments for these patients.

Keywords: Transcriptome, Lung adenocarcinoma (LUAD), Lymph node metastasis (LNM), mRNA signature, Weighted gene co-expression network analysis (WGCNA), Overall survival (OS)

Background

Lung cancer is the leading cause of cancer-related death worldwide, with adenocarcinoma being the most common histological type [1]. Despite advances in cancer therapy in recent decades, the prognosis of lung adenocarcinoma (LUAD) patients is still unfavourable, with an

overall 5-year survival rate less than 15% [2]. The main reason for this low overall survival rate is that LUAD patients have a high frequency of lymph node metastasis (LNM) or even distant metastases at diagnosis [3–6]. Therefore, the identification of a high-risk subset from these patients who have greater need for additional systematic therapy is urgently needed.

In recent years, the development of gene expression profile technologies, such as microarray and next generation sequencing (NGS), have provided further opportunities to comprehensively characterize the molecular features of cancer [7, 8]. Considering that individual biomarkers usually have little statistical power, the current approach is to identify novel molecular signatures to offer better

*Correspondence: shirun0122@gmail.com; shirun@outlook.com; xuanwen.bao@helmholtz-muenchen.de

[†]Yanfang Wang and Quanli Zhang contributed equally to this work

¹ Ludwig-Maximilians-Universität München (LMU), 80539 Munich, Germany

⁶ Institute of Radiation Biology, Helmholtz Center Munich, German Research Center for Environmental Health, 85764 Neuherberg, Germany
Full list of author information is available at the end of the article



prediction in various cancers [9–11]. A number of studies have proposed gene expression-based signatures for survival stratification in patients with lung cancer [12–16]. However, few studies have focused on the prognostic prediction for LUAD patients with LNM.

In this study, based on The Cancer Genome Atlas (TCGA) LUAD mRNA-seq and clinical datasets, we sought to develop a gene expression signature to predict overall survival for LUAD patients with LNM, and then the proposed gene signature was validated in an external cohort from the Gene Expression Omnibus (GEO) database.

Methods

Data download and processing

TCGA RNA-seq datasets and clinical data for LUAD were downloaded by UCSC Xena browser (<https://xenabrowse.rnet/>). GSE68465 was download from the GEO database. The LUAD patients with LNM were filtered by the criteria that N stage of patients was I–IV.

Co-expression gene network based on RNA-seq data

The Weighted Correlation Network Analysis (WGCNA) was used to construct the gene co-expression network [17]. The co-expression similarity $s_{i,j}$ was defined as the absolute value of the correlation coefficient between the profiles of nodes i and j :

$$s_{i,j} = |cor(x_i, x_j)|$$

where x_i and x_j are expression values of for gene i and j , and $s_{i,j}$ represent the Pearson’s correlation coefficients of gene i and gene j , respectively.

A weighed network adjacency was defined by raising the co-expression similarity to a power β :

$$a_{i,j} = s_{i,j}^\beta$$

with $\beta \geq 1$ [18]. We selected the power of $\beta = 5$ and scale free $R^2 = 0.95$ as the soft-thresholding parameters to ensure a signed scale-free co-expression gene network. Briefly, network construction, module detection, feature selection, calculations of topological properties, data simulation, and visualization were performed. Modules were identified via hierarchical clustering of the weighting coefficient matrix. The module membership of node i in module q can be defined as:

$$K_{cor,i}^{(q)} := cor(x_i, E^{(q)})$$

where x_i is the profile of node i , and $E^{(q)}$ is the module eigengene (the first principal component of a given module) of module q [19]. The module membership measure $K_{cor,i}^{(q)}$ lies in $[-1, 1]$ and specifies how close node i is to module q , $q = 1, \dots, Q$.

The topological overlap measure (TOM) plots visualized inter-connectivity patterns and suggests the presence of large modules. This property of dense connections among the genes of module q can be measured using the concept of module density, which is defined as the average adjacency of the module genes:

$$Density(A^{(q)}) = \frac{\sum_i \sum_{j \neq i} a_{i,j}^{(q)}}{n^{(q)}(n^{(q)} - 1)}$$

where $A^{(q)}$ denotes the $n^{(q)} \times n^{(q)}$ adjacency matrix corresponding to the sub-network formed by the genes of module q . By evaluating the correlations between the LNM status of LUAD and their module memberships, highly correlated modules can be identified. The modules that had correlation coefficient with N status larger than 0.15 or less than -0.15 were selected.

Differentially expressed gene (DEG) analysis

DEG analysis was performed by the Limma package [20]. The tissue samples were separated into para-tumour group and tumour group. The DEGs were defined as genes with Q value (adjusted p value between two groups) less than 0.05.

Cox regression

Cox regression, also called Proportional Hazards Regression, is a survival analysis model [21]. It can be used to analyse relationships between different features and the survival time. The Cox model is based on the proportional hazards condition, which assumes that features have a proportional relationship to the exponential change of a hazard. Thus, the model is formulated as a multiplication of a baseline hazard function with a sole time variable, t , and an exponential function of the linear combination of all of the features as an input. Given a set of n samples $\{(X_i, Y_i, s_i) | 0 \leq i \leq n, i \in \mathbf{R}\}$, where $X_i = (x_{i0}, x_{i1}, \dots, x_{ik})$ and stands for the i -th sample of all the k features, Y_i is the observation time and s_i is the survival status, the hazard function is

$$H_i(t) = H_0(t)e^{X_i^T \beta}$$

$\beta = (\beta_0, \beta_1, \dots, \beta_k)$ is the coefficient vector weighing the contribution of the features. The partial likelihood of all the samples is

$$\begin{aligned} L(\beta) &= \prod_{i=1}^n L_i(\beta) \\ &= \prod_{i=1}^n \frac{H_i(Y_i|X_i)}{\sum_{j: Y_j \geq Y_i} H_j(Y_i|X_j)} \\ &= \prod_{i=1}^n \frac{e^{X_i^T \beta}}{\sum_{j: Y_j \geq Y_i} e^{X_j^T \beta}} \end{aligned}$$

By penalizing $-\log(L(\beta))$, the optimal β can be uncovered.

LASSO regularization

LASSO (Least Absolute Shrinkage and Selection Operator) is an important regularization in many regression analysis methods (e.g., COX regression, logistic regression) [22]. The idea behind LASSO is that an L1-norm is used to penalize the weight of the model parameters. Assuming a model has a set of parameters $\{w_0, w_1, \dots, w_n\}$, the LASSO regularization can be defined as

$$\lambda \cdot \sum_{i=0}^n \|w_i\|_1$$

It can also be expressed as a constraint to the targeted objective function

$$\sum \|Y - Y^*\|_2, \quad s.t. \|w_i\|_1 < t$$

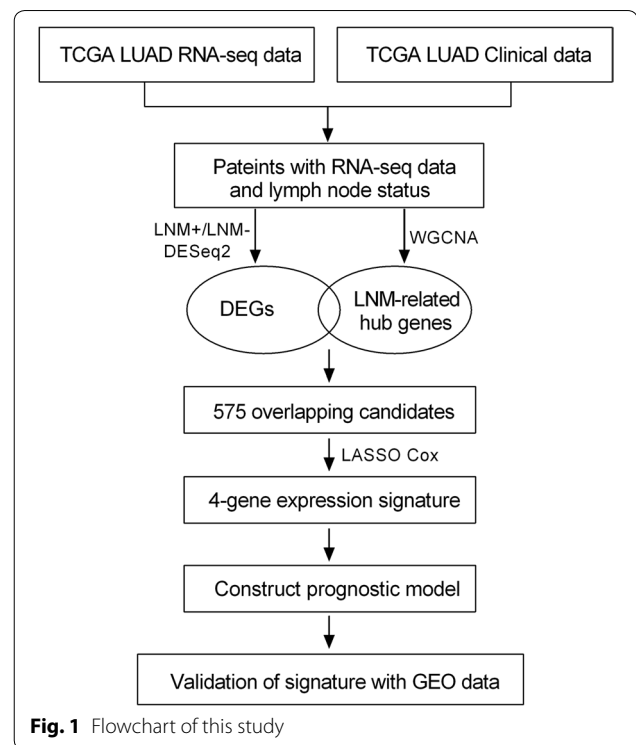
An important property of the LASSO regularization term is that it can force the parameter values to be 0, thus generating a sparse parameter space, which is a desirable characteristic for feature selection. In our analysis, the overlapping genes from DEGs and selected modules were used as the input of Lasso Cox regression. The nomogram was done by rms package [23]. The GSEA was done by GSEA software from Broad institute [24].

Results

The flowchart of our study is shown in Fig. 1. By integrating the TCGA LUAD mRNA-seq dataset with the clinical dataset, 575 promising candidates were filtered out and submitted for LASSO Cox regression analysis to identify robust markers to construct a prognostic signature. Then, a GEO dataset (GSE68465) was used to validate the model.

Identification and selection of prognostic biomarkers in LNM-positive patients

The whole transcriptome of LNM-positive samples was included to perform DEG analysis, and the volcano plot showed 974 differentially expressed genes (DEGs) in 173 LNM-positive (LNM+) tumour samples compared to 335 LNM-negative (LNM-) tumour samples (Fig. 2a). To construct gene co-expression modules, RNA-seq data from the whole genome was subjected to WGCNA. Genes were assigned to different modules by cluster dendrogram trees, and unassigned genes were categorized into the grey module (Fig. 2b). The relationships between clinical traits and gene modules are presented in Fig. 2c. Absolute values of correlation coefficients between LNM status and modules greater than 0.15 were considered

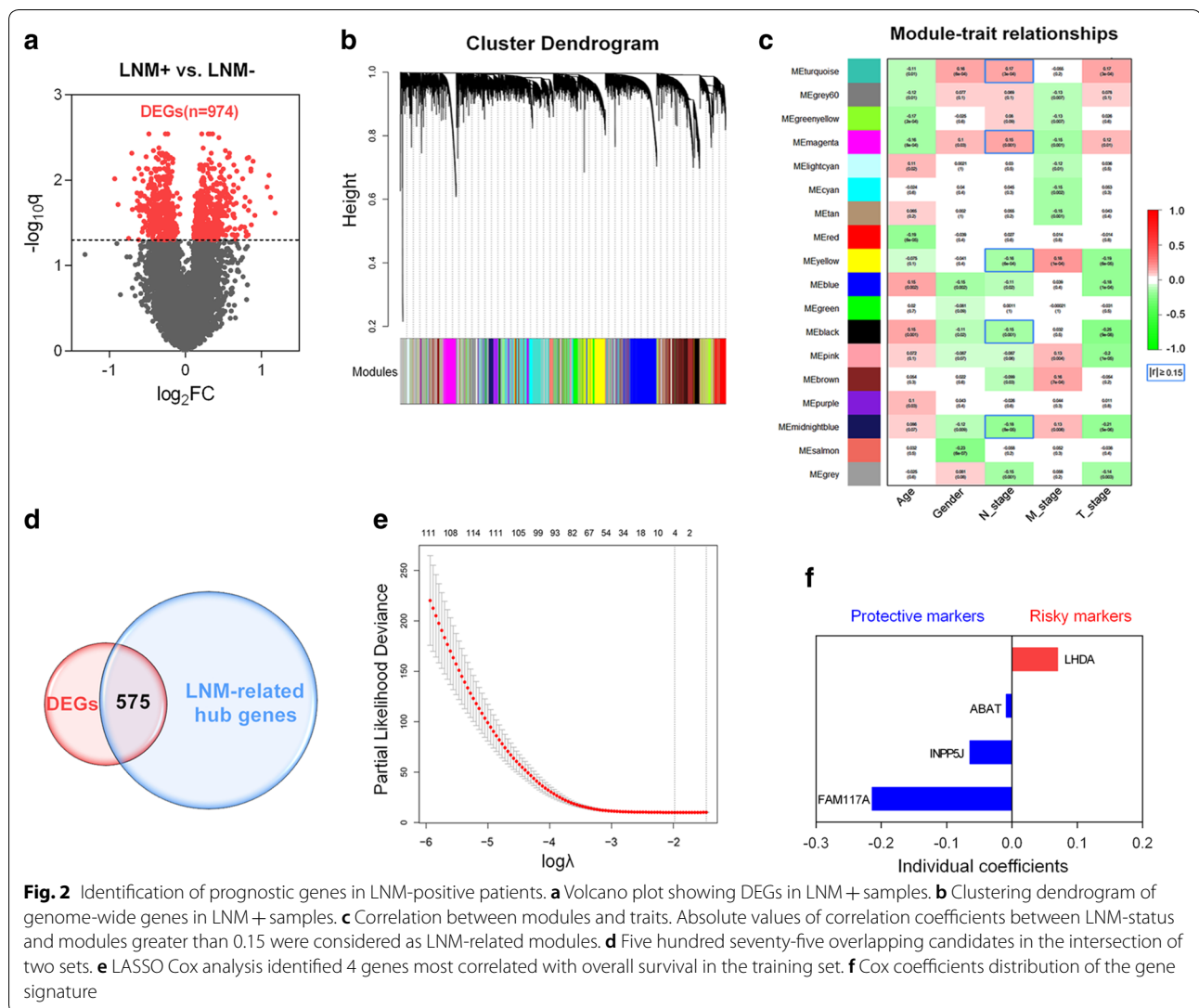


as LNM-related modules, and genes in these modules were extracted for further study. As shown in Fig. 2d, we obtained 575 overlapping genes in the intersection of DEGs and LNM-related hub genes.

Then, LASSO Cox regression analysis was performed to identify robust markers among the 575 candidates. By forcing the sum of the absolute value of the regression coefficients to be less than a fixed value, certain coefficients were shrunk to exactly zero, and the most powerful prognostic markers were identified with relative regression coefficients. Cross-validation was applied to prevent the over-fitting of the LASSO Cox model (Fig. 2e). Figure 2f shows individual coefficient distributions of the 4 filtered markers: LDHA was associated with high risk (HR > 1), while ABAT, INPP5J and FAM117A were shown to be protective (HR < 1).

Risk score and survival prediction based on the 4-gene signature

To comprehensively investigate the association between the 4 identified genes and prognosis in these patients, we developed a 4-gene signature-based prognostic model according to their individual coefficients. Then, we calculated the risk score for each LNM+ patient in the training set and ranked them. Thus, LNM+ patients with follow-up information were divided into two groups: a low-risk group (n=78) and a high-risk group (n=78)



based on median cut-off value (Fig. 3a). Figure 3b shows the survival overview in the training cohort. A heatmap showed that patients in the high-risk group have a tendency to have higher expression of LDHA and lower expression of ABAT, FAM117A and INPP5J (Fig. 3c). The Kaplan–Meier curve and log-rank test suggested that patients in the high-risk group have significantly worse overall survival compared to those in the low-risk group (HR=1.884, p=0.0035) (Fig. 3d). As shown in Fig. 3e, GSEA showed the top 5 hallmarks correlated with the high-risk group: E2F targets, EMT, Hypoxia, MTORC1 signalling and MYC targets (FDR q<0.001).

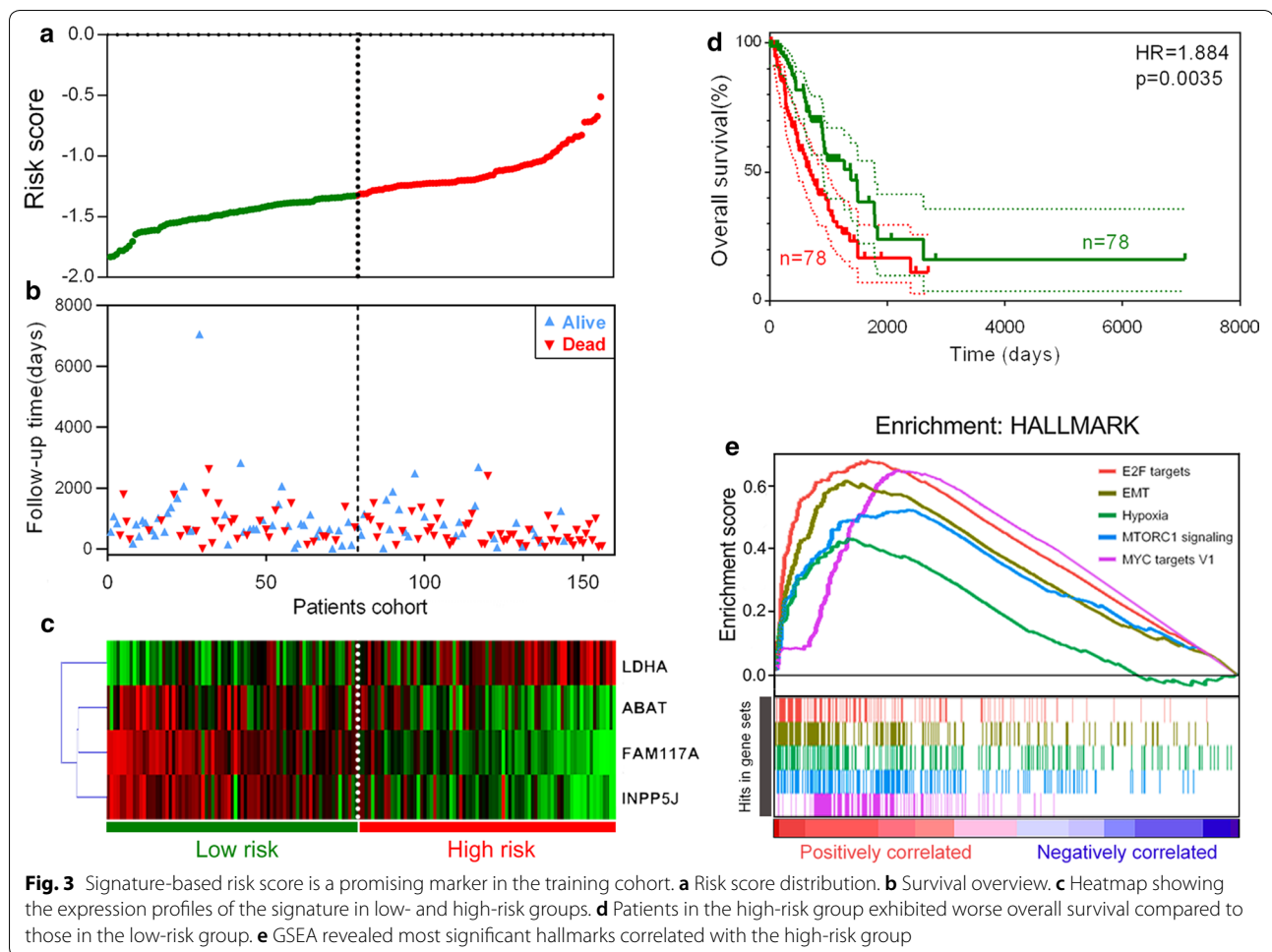
Expression profiles of the 4-gene signature and subgroup analysis

We investigated the 4 genes’ expression profiles in different AJCC-TNM stages in the TCGA cohort. As shown

in Fig. 4a, one-way ANOVA test showed that LDHA mRNA expression was significantly upregulated, while ABAT, FAM117A and INPP5J were significantly downregulated, in more advanced stages. As shown in Fig. 4b, the signature-based risk score also serves as a promising marker to predict overall survival in different subgroups, including stage II (HR=3.015, p=0.0006), stage III-IV (HR=3.321, p<0.0001), EGFR-wild-type (EGFR-WT) (HR=2.240, p=0.0013), EGFR-mutated (EGFR-Mut) (HR=4.094, p=0.0060), KRAS-wild type (KRAS-WT) (HR=2.044, p=0.0089) and KRAS-mutated (KRAS-Mut) (HR=3.433, p=0.0003) patients, respectively.

Construction of a nomogram to predict overall survival in LNM-positive patients

Then, we constructed a nomogram that integrated clinicopathological features with the 4-gene signature to



predict survival probability of LNM+ patients (Fig. 5a). Calibration plot showed that the nomogram-predicted 3-year and 5-year survival probabilities and corresponded closely to the actual observed proportions (Fig. 5b). The AUC(t) functions of the multivariable models were developed to indicate how well these features serve as prognostic markers. Compared to other features, such as signature-based risk score, AJCC-TNM stage and age, the nomogram showed the highest predictive power for overall survival in the training cohort, with an average AUC above 0.7 in the follow-up period (Fig. 5c).

Validation of the 4-gene signature for survival prediction

To confirm our findings in the training set, we validated the prognostic function of the 4-gene signature in an independent GEO cohort (GSE68465). After extracting the microarray data of 140 LNM+ patients with follow-up information, we calculated the risk score for each patient by using the same formula in the training set. Figure 6a shows the risk score distribution, and Fig. 6b shows the survival overview in the validation cohort. According

to the median cut-off value, the cohort of patients were divided into high- (n=70) and low-risk (n=70) groups. The Kaplan–Meier curve suggested a significant better overall survival in the low-risk group compared to the high-risk group (HR=1.632, p=0.0106) (Fig. 6c). The result was consistent with our previous findings in the training cohort based on TCGA dataset, indicating the gene signature was validated as a reliable predictor for overall survival in an independent LNM+LUAD patient cohort.

Discussion

Accumulating evidence shows that lung adenocarcinoma patients with local invasion or lymph node metastasis always exhibit poor responses to standard treatments and thus tend to have poor clinical outcomes [3–5, 25]. In clinical practice, these patients need more intensive monitoring and aggressive therapy, and robust biomarkers are urgently needed to stratify high-risk groups among these patients. However, individual biomarkers usually have very little predictive power. The established clinical

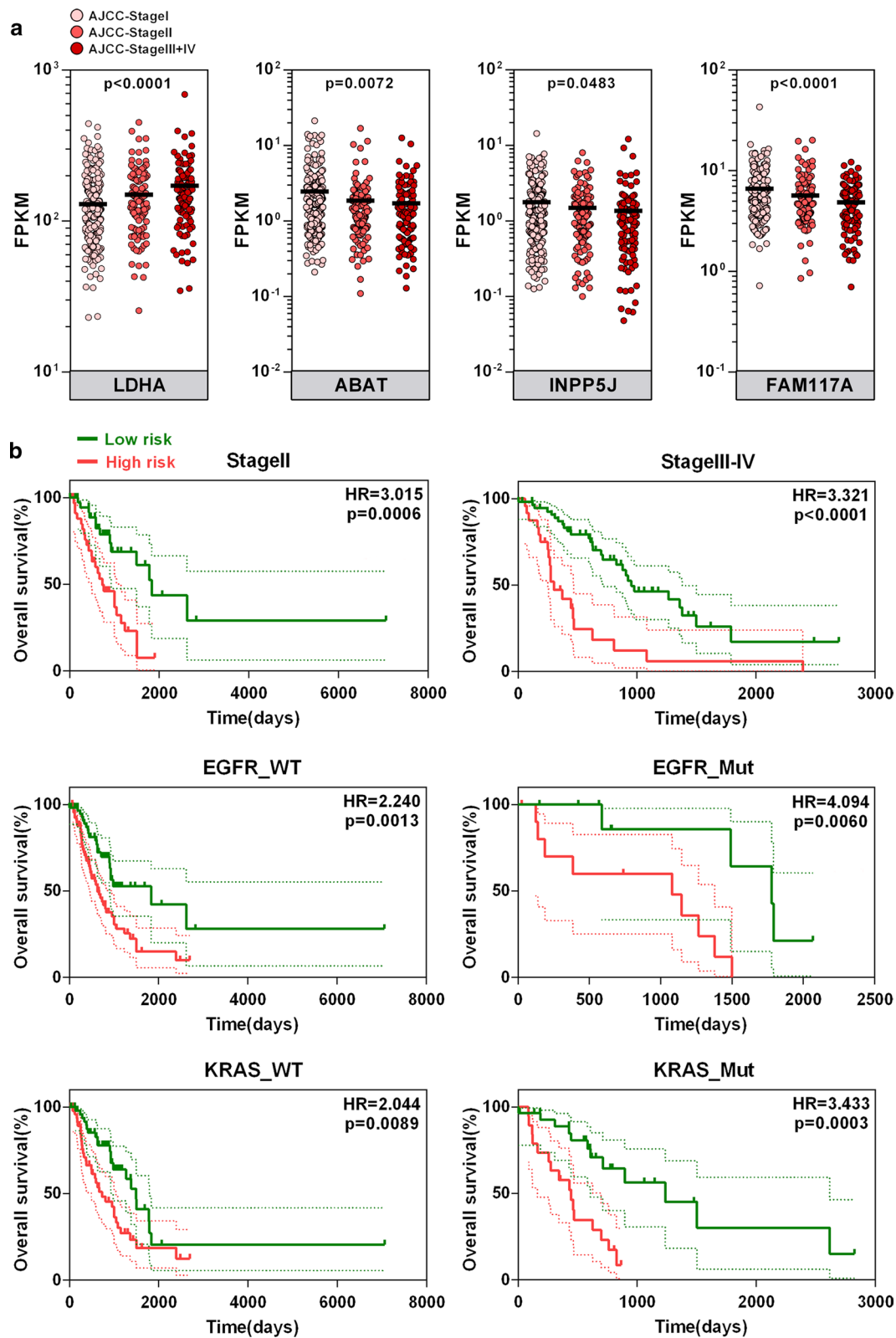


Fig. 4 Expression and survival analysis in subgroups. **a** Expression pattern of the gene signature in different AJCC-TNM stages. **b** Signature-based risk score is a promising marker for overall survival in subgroups with different tumour stages and EGFR and KRAS statuses

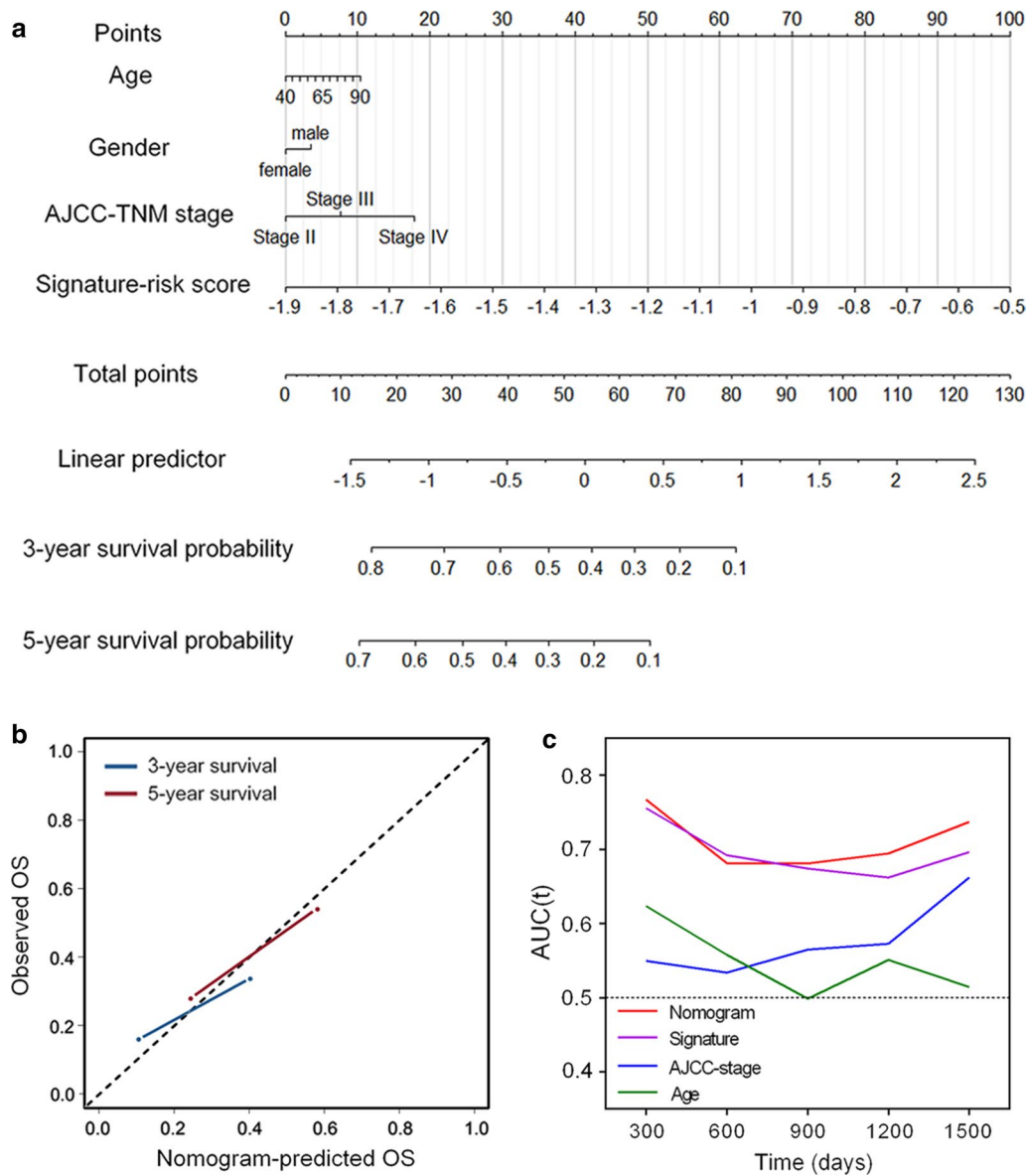
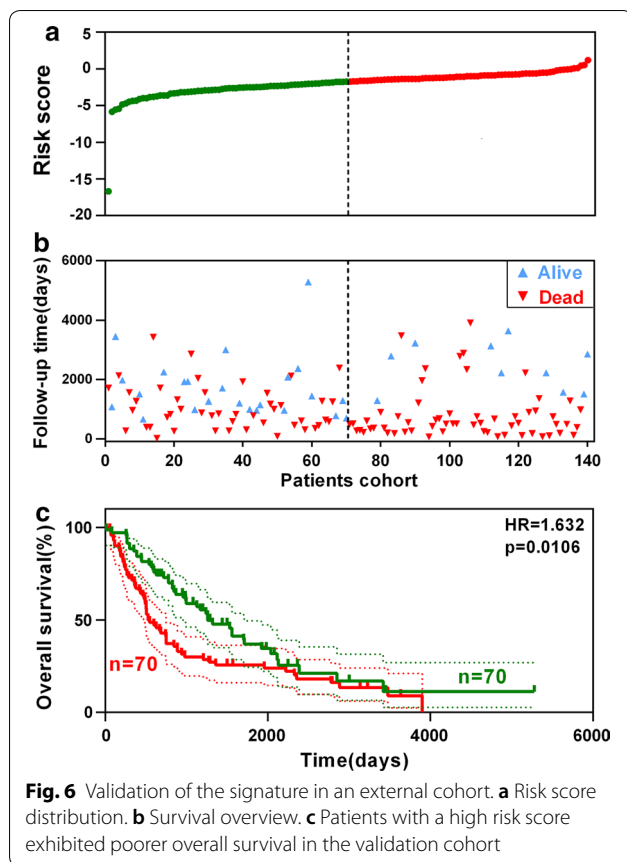


Fig. 5 Construction of a nomogram for survival prediction. **a** Nomogram combining signature with clinicopathological features. **b** Calibration plot showing that nomogram-predicted survival probabilities corresponded closely to the actual observed proportions. **c** The AUC(t) of multivariable models indicated the nomogram had the highest predictive power for overall survival

markers for survival are primarily based on patient- and tumour-related factors, such as AJCC-TNM stage, while the accuracy and specificity are also limited. Therefore, our study aimed to identify novel molecular signatures integrated with established clinicopathological features to predict overall survival in LUAD patients with lymph node metastasis.

In this study, we identified 4 coding genes associated with overall survival in LUAD patients with lymph node metastasis, namely, LDHA, ABAT, FAM117A and

INPP5J, in the training set. Among the 4 coding genes, LDHA was widely reported to promote malignant progress and predict poor survival in various cancer types [26–30]. In addition, Ooms et al. reported that INPP5 J functions as a tumour suppressor to inhibit breast cancer cells’ invasive ability via PI3 K/AKT signalling [31]. However, ABAT and FAM117A remain inadequately investigated in cancer-related research. To some degree, our study might provide some clues for further investigation



on the biological roles and clinical significance of these genes.

Based on multivariate Cox coefficients derived from LASSO analysis, we developed a 4-gene signature-based risk score model. Moreover, we investigated the prognostic value of the signature in different subgroups. In detail, the signature still exhibited powerful prediction for overall survival in LNM+ patients with same TNM stage and genomic alteration (including EGFR and KRAS mutation status), confirming that the signature is a promising marker independent of different clinicopathological features. In addition to survival prediction, GSEA showed that the signature-identified high-risk group was significantly correlated with certain hallmarks of cancer, such as EMT and hypoxia, indicating the potential molecular mechanisms underlying the lethal tendency of these LNM+ patients. By integrating established clinicopathological features with the signature, we developed a nomogram to predict the survival probability of LNM+ patients. The predictive power was measured by the time-dependent area under the ROC curve (AUC(t)), and the result showed that the integrated nomogram model had higher predictive power than individual markers. Lastly, an external GEO

cohort was used to validate the prognostic value of the 4-gene signature, and the survival analysis showed the same tendency in the validation cohort.

The limitations of our study should be acknowledged. This is a retrospectively designed study, and the sample size of the training and validation sets is relatively small.

In summary, the novel 4-gene signature proved to be a robust model with high predictive power in LUAD patients with LNM+. The predictive power was stable over time and showed promising survival prediction in combination with established markers. The use of the signature integrated with clinicopathological features can help to further stratify LNM+ patients into risk groups, thus serving as a predictive tool for clinical outcome, guiding personalized treatment, and resulting in more aggressive therapy for high-risk patients or less aggressive therapy for low-risk patients. Further research is needed to reveal the interplay between these genes, and thereby, we will be able to develop better treatment alternatives for high-risk LUAD patients with lymph node metastasis.

Conclusion

In conclusion, based on publicly available data, we constructed a robust mRNA signature that could serve as a reliable marker to stratify a high-risk group among LUAD LNM+ patients. Subgroup analysis indicated that the signature works effectively independent of other clinical features. Validation in an external cohort from GEO further confirmed the prognostic value of the signature. We hope the identified signature could help to improve the strategies for personalized treatment of LUAD patients with LNM.

Abbreviations

LUAD: lung adenocarcinoma; LNM: lymph node metastasis; TCGA: The Cancer Genome Atlas; tROC: time-dependent receiver operating characteristic; GSEA: Gene Set Enrichment Analysis; GEO: Gene Expression Omnibus; WGCNA: weighted gene co-expression network analysis; OS: overall survival; NGS: next generation sequencing; DEG: differentially expressed gene; LASSO: Least Absolute Shrinkage and Selection Operator.

Authors' contributions

XW B and R S conceived and designed the experiments. XWB, SX, QLZ, ZJG and KZ analysed the data. XWB, QLZ, YBZ and YFW wrote the paper. All authors read and approved the final manuscript.

Author details

¹ Ludwig-Maximilians-Universität München (LMU), 80539 Munich, Germany. ² Jiangsu Key Laboratory of Molecular and Translational Cancer Research, Nanjing 210009, China. ³ Department of Thoracic Surgery, The Affiliated Changzhou No. 2 People's Hospital of Nanjing Medical University, Changzhou 213000, China. ⁴ Institute of Molecular Toxicology and Pharmacology, Helmholtz Center Munich, German Research Center for Environmental Health, 85764 Neuherberg, Germany. ⁵ Department of Cardiology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou 310016, China. ⁶ Institute of Radiation Biology, Helmholtz Center Munich, German Research

Center for Environmental Health, 85764 Neuherberg, Germany. ⁷ Technical University Munich (TUM), 80333 Munich, Germany.

Acknowledgements

We would like to thank Dr. Omid Azimzadeh and Dr. Michael Rosemann for helpful discussions and suggestions.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The datasets supporting the conclusions of this article are available in the Xena browser (<https://xenabrowser.net/>) repository.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

This work was supported by the Zhejiang Provincial Natural Science Foundation (No. LY16H020005).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 4 March 2019 Accepted: 8 April 2019

Published online: 16 April 2019

References

- Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. *CA Cancer J Clin*. 2016;66(2):115–32.
- Ettinger DS, Wood DE, Akerley W, Bazhenova LA, Borghaei H, Camidge DR, et al. Non-small cell lung cancer, Version 6. 2015. *J Natl Compr Cancer Netw*. 2015;13(5):515–24.
- Nwogu CE, Groman A, Fahey D, Yendamuri S, Dexter E, Demmy TL, et al. Number of lymph nodes and metastatic lymph node ratio are associated with survival in lung cancer. *Ann Thorac Surg*. 2012;93(5):1614–9 (**discussion 9–20**).
- Ou SH, Zell JA, Ziogas A, Anton-Culver H. Prognostic factors for survival of stage I nonsmall cell lung cancer patients: a population-based analysis of 19,702 stage I patients in the California Cancer Registry from 1989 to 2003. *Cancer*. 2007;110(7):1532–41.
- Sakao Y, Okumura S, Mingyon M, Uehara H, Ishikawa Y, Nakagawa K. The impact of superior mediastinal lymph node metastases on prognosis in non-small cell lung cancer located in the right middle lobe. *J Thorac Oncol*. 2011;6(3):494–9.
- Kafka A, Tomas D, Beroš V, Pečina H, Zeljko M, Pečina-Šlaus N. Brain metastases from lung cancer show increased expression of DVL1, DVL3 and beta-catenin and down-regulation of E-cadherin. *Int J Mol Sci*. 2014;15(6):10635–51.
- Hurd PJ, Nelson CJ. Advantages of next-generation sequencing versus the microarray in epigenetic research. *Brief Funct Genomic Proteomic*. 2009;8(3):174–83.
- Meldrum C, Doyle MA, Tothill RW. Next-generation sequencing for cancer diagnostics: a practical perspective. *Clin Biochem Rev*. 2011;32(4):177–95.
- Kim S-K, Kim S-Y, Kim J-H, Roh S, Cho D-H, Kim YS, et al. A nineteen gene-based risk score classifier predicts prognosis of colorectal cancer patients. *Mol Oncol*. 2014;8(8):1653–66.
- Dai W, Li Y, Mo S, Feng Y, Zhang L, Xu Y, et al. A robust gene signature for the prediction of early relapse in stage I–III colon cancer. *Mol Oncol*. 2018;12(4):463–75.
- Bai F, Zhou H, Ma M, Guan C, Lyu J, Meng QH. A novel RNA sequencing-based mi RNA signature predicts with recurrence and outcome of hepatocellular carcinoma. *Mol Oncol*. 2018;12:1125–37.
- Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med*. 2002;8(8):816–24.
- Director's Challenge Consortium for the Molecular Classification of Lung A, Shedden K, Taylor JM, Enkemann SA, Tsao MS, Yeatman TJ, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med*. 2008;14(8):822–7.
- Kratz JR, He J, Van Den Eeden SK, Zhu ZH, Gao W, Pham PT, et al. A practical molecular assay to predict survival in resected non-squamous, non-small-cell lung cancer: development and international validation studies. *Lancet*. 2012;379(9818):823–32.
- Wistuba II, Behrens C, Lombardi F, Wagner S, Fujimoto J, Raso MG, et al. Validation of a proliferation-based expression signature as prognostic marker in early stage lung adenocarcinoma. *Clin Cancer Res*. 2013;19(22):6261–71.
- Chen HY, Yu SL, Chen CH, Chang GC, Chen CY, Yuan A, et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med*. 2007;356(1):11–20.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform*. 2008;9(1):559.
- Horvath S, Zhang B, Carlson M, Lu K, Zhu S, Felciano R, et al. Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proc Natl Acad Sci*. 2006;103(46):17402–7.
- Horvath S, Dong J. Geometric interpretation of gene coexpression network analysis. *PLoS Comput Biol*. 2008;4(8):e1000117.
- Smyth GK. Limma: linear models for microarray data. *Bioinformatics and computational biology solutions using R and bioconductor*. Berlin: Springer; 2005. p. 397–420.
- Cox DR. Regression models and life-tables. *J Roy Stat Soc Ser B*. 1972;34(2):187–202.
- Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med*. 1997;16(4):385–95.
- Harrell FE. Ordinal logistic regression. *Regression modeling strategies*. Berlin: Springer; 2015. p. 311–25.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene Set Enrichment Analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci*. 2005;102(43):15545–50.
- Kuroda H, Sakao Y, Mun M, Uehara H, Nakao M, Matsuura Y, et al. Lymph node metastases and prognosis in left upper division non-small cell lung cancers: the impact of interlobar lymph node metastasis. *PLoS ONE*. 2015;10(8):e0134674.
- Huang X, Li X, Xie X, Ye F, Chen B, Song C, et al. High expressions of LDHA and AMPK as prognostic biomarkers for breast cancer. *Breast*. 2016;30:39–46.
- Jin L, Chun J, Pan C, Alesi GN, Li D, Magliocca KR, et al. Phosphorylation-mediated activation of LDHA promotes cancer cell invasion and tumour metastasis. *Oncogene*. 2017;36(27):3797–806.
- Shi M, Cui J, Du J, Wei D, Jia Z, Zhang J, et al. A novel KLF4/LDHA signaling pathway regulates aerobic glycolysis in and progression of pancreatic cancer. *Clin Cancer Res*. 2014;20(16):4370–80.
- He Y, Chen X, Yu Y, Li J, Hu Q, Xue C, et al. LDHA is a direct target of miR-30d-5p and contributes to aggressive progression of gallbladder carcinoma. *Mol Carcinog*. 2018;57(6):772–83.
- Zhang H, Li L, Chen Q, Li M, Feng J, Sun Y, et al. PGC 1 β regulates multiple myeloma tumor growth through LDHA-mediated glycolytic metabolism. *Mol Oncol*. 2018;12(9):1579–95.
- Ooms LM, Binge LC, Davies EM, Rahman P, Conway JR, Gurung R, et al. The inositol polyphosphate 5-phosphatase PIPP regulates AKT1-dependent breast cancer growth and metastasis. *Cancer Cell*. 2015;28(2):155–69.