

PRIMARY RESEARCH

Open Access



Identification of four differentially methylated genes as prognostic signatures for stage I lung adenocarcinoma

Wei-Ming Luo¹, Zheng-Yu Wang^{2*} and Xin Zhang^{3*}

Abstract

Background: Lung adenocarcinoma (LUAD) is the main subtype of non-small cell lung cancer with a low survival prognosis. We aimed to generate a prognostic model for the postoperative recurrence of LUAD.

Methods: The methylated DNA data of LUAD patients were downloaded from the Cancer Genome Atlas (TCGA). The differentially methylated genes were identified and protein–protein interacting network was constructed, with which prognostic signature of this cancer was generated. Survival and functional pathways analysis was used to evaluate the clustering ability of the prognostic signature.

Results: We identified 151 differentially methylated genes related to relapse-free survival of patients with LUAD. Nine hub genes were identified in PPI network, with which 4 gene pair signature was selected as prognostic signature. The potential functions of 6 genes (JDP2, SERPINA5, PLG, SEMG2, RFX5, and POLR3B) in the 4-gene pair signature were enriched in intracellular protein synthesis and transportation.

Conclusion: The four gene pair signature can predict the prognosis of patients with stage I LUAD. Our study provides a reference for patients with postoperative adjuvant therapy.

Keywords: Lung adenocarcinoma, Prognostic signature, DNA methylation, Survival analysis

Background

Lung cancer has two main types known as small-cell lung carcinoma (SCLC) and non-small-cell lung carcinoma (NSCLC) [1, 2]. Currently, no treatment can cure this kind of cancer, making lung cancer the leading cause of the death from cancer worldwide. Hereinto, lung adenocarcinoma (LUAD) is the main subtype of non-small cell lung cancer (NSCLC). With an increasingly high incidence, LUAD is always a threat to human beings regardless of gender and smoking condition over the past few decades in many countries [3]. Previous reports showed that patients with LUAD had a shorter survival time

among patients with other types of non-small cell lung cancer (NSCLC) [4–6].

As a common histological subtype of NSCLC, most of clinical treatments of LUAD are basing on the cancer staging system of American Joint Committee on Cancer (AJCC) [7, 8]. Surgical resection is usually used to treat patients with stage I LUAD, but, achieved a poor prognosis. The recurrence rate of postoperative patients is as high as 35–50% [9]. For the large amount of patients suffering for reoccurrence, adjuvant chemotherapy could be an effective approach to significantly improve their survival time [10]. Thus, it is indeed urgent to discover an accurate and reliable clinical method to evaluate and/or predict the prognosis of LUAD, especially at early stage.

The most popular approach in recent years is to screen oncogenes or/and microRNAs as biomarkers for clinical diagnosis and treatment [11]. For diagnosis and treatment, the application of biomarkers can improve the survival of patients following with more personalized

*Correspondence: zheywang@163.com; 973306782@qq.com

² Department of Pharmacy, The Affiliated Huai'an Hospital of Xuzhou Medical University and The Second People's Hospital of Huai'an, 62 South Huai'hai Rode, Huai'an, China

³ Department of Medical Imaging, The Fourth People's Hospital of Huai'an, Huai'an, Jiangsu, China

Full list of author information is available at the end of the article



treatments [12, 13]. In addition to the genetic causes, of cancer, epigenetic changes also contribute to the development of cancer. Hereinto, DNA methylation is a kind of mark in a variety of tumors including lung cancer [14]. In DNA methylation, hypermethylation is recognized not only able to repress transcription of tumor suppressor genes, but also to trigger oncogenesis. It has been studied that tumor-suppressor genes are inactivated by the hypermethylations occurring on cytosine–guanine (CpG) island which locates in the promoter regions [15]. In addition, the enlarging public database for DNA methylation also provide wider and sufficient resources to study the mechanism and to explore methylated gene biomarkers [16, 17]. Taken these, identifying the target genes that can be silenced by DNA methylation or screen the cancer related methylations have a great impact on early diagnosis, cancer staging, and prognosis of cancer patients. However, less study is focusing on the relationships of DNA methylation and recurrence of LUAD patients [18].

Here selected the Cancer Genome Atlas (TCGA) database and downloaded methylated DNA data of LUAD patients. Then we identified 151 differentially methylated genes related to relapse-free survival (RFS) of patients with LUAD. Nine hub genes were identified in PPI network, with which 4 gene pair signature was selected as prognostic signature, which successfully clustered LUAD patients at stage I into high- and low-risk groups with significant differences. The potential functions of 6 genes (JDP2, SERPINA5, PLG, SEMG2, RFX5, and POLR3B) in the 4-gene pair signature were enriched in intracellular protein synthesis and transportation. The prognostic signature generated in the present study can predict the prognosis of patients with stage I LUAD, providing reference on the recurrence risk. The application of this signature will as well contribute to improve the overall survival of LUAD patients with postoperative adjuvant therapy. It should be noted that the signatures identified in this study are generally applicable. But for the personalized cases, specific methods would be applied in practice.

Materials and methods

Data source

Expressing profiles of gene-specific DNA methylation data and the follow-ups of corresponding patients were downloaded from database The Cancer Genome Atlas (TCGA, <https://tcga-data.nci.nih.gov/tcga/>) before April 17th, 2017. The clinic information of the patients was listed in detail (Table 2). The DNA methylation data was measured with Illumina Human Methylation 27 Beadchip (27K array) and Illumina Human Methylation 450 Beadchip (450K array). We focused on analyzing the 25,978 CpG sites measured by both 27K and 450K arrays.

The probes were annotated to genes according to the annotation table of 27K platform.

Data preprocessing and differentially methylated sites

To clear the DNA methylation data, CPG sites would be deleted if the CPGs were in absence in more than 50% of the total samples. CPG sites deficient in less than 50% samples would be weighted by using k-Nearest-Neighbor values [19].

Methylated CPG sites differentially expressed in samples with LUAD was screened by using SAM [20]. This process was repeated 1000 times for each disturbance. For the multiple binomial tests, the *p*-values are adjusted by the Benjamin and Hochberg method to control the false discovery rate (FDR) [21].

Protein and protein interactions (PPIs)

To analyze the biological functions of the selected genes, protein–protein interacting (PPI) network was constructed to screen the hub genes associating with prognosis of LUAD [22]. All the protein networks used in the present study were obtained in database signor [23].

Survival analysis

The correlation between each differentially methylated gene pair and the RFS was evaluated in univariate survival analysis. To estimate the independent prognostic value of this signature basing on several clinical factors including age, gender and smoking, the multivariate survival analysis was performed subsequently. Cox proportional-hazards regression model was applied in all the above survival analysis [24]. At the same time, the significance of the correlation between the ridge expansion osteotomy (REO) of a gene pair and the poor RFS was validated. Survival curves were drawn by using the Kaplan–Meier method and were compared in the log-rank test [25].

Enrichment analysis

The KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways of selected methylated genes were analyzed by using DAVID (The Database for Annotation, Visualization and Integrated Discovery) [26]. Functional key genes were determined by Spearman's correlation coefficient $FDR < 0.01$, $r > 0.6$.

Results

Data source

All gene-specific DNA methylation data were obtained from 578 patient samples from lung adenocarcinoma (LUAD), including 56 normal samples next to cancer tissues. There were 191 survival information containing RFS from 578 patients with stage I LUAD.

Differentially methylated sites

In identification of the differentially methylated CPG sites (1000 times, $FDR < 0.01$), a total of 5029 CPG sites were identified up-regulated and 3269 ones were down-regulated. Then these differentially methylated CPG sites were mapped onto corresponding genes.

One gene would be defined as differentially methylated if one differentially methylated CPG site occurred on its promoter. On the contrary, if there were two methylated sites occurred on the promoter with different directions, this gene would be deleted. Finally, we obtained 3498 hypermethylation of genes and 2465 ultra-low methylated genes. In univariate Cox analysis basing on survival information, genes related to non-recurrence of postoperative patients with stage I LUAD were screened as well ($p < 0.01$), 61 hypermethylation and 90 ultra-low methylated genes.

Hub genes screening in PPI network

With these 151 genes associating with non-recurrence of postoperative patients, PPI networks were constructed to identify hub genes. Totally, 17 hub genes were obtained, indicating 9 key genes (KLK3, GUCY2F, KLK2, SERPINA5, PLG, SEMG2, RFX5, POLR3B, JDP2) relating to non-recurrence of postoperative patients with LUAD (Fig. 1).

Survival-associated gene pairs were generated with these 9 key genes. Five gene pairs were identified significantly associating with survival of LUAD ($p < 0.01$) in univariate Cox analysis. By using the C-index of gene pairs

and forward-stepwise algorithm, a 4-gene pair prognostic signature, consisting of 6 genes (Table 1), were generated for postoperative patients with stage I LUAD (Fig. 2). Postoperative patients with on gene pair of the 4-gene pair signature were defined as poor prognosis of recurrence. In Fig. 2, patients with stage I LUAD were able to be classified into high- and low-risk group with significant RFS.

Survival analysis

We applied Kaplan–Meier method and log-rank test to evaluate the impacts of prognostic signature casting on the RFS of stage I LUAD patients. The relationships of RFS time with age, gender and smoking condition were further analyzed (Table 2). Figure 3 indicated that the 4 gene-pair signature was able to classify the stage I LUAD patients into two groups with significant RFS (univariate COX $p < 0.0057$, multivariate COX $p < 0.0069$). However, in the relationships of RFS and age, gender and smoking condition, no factors showed influences on our prognostic signature.

Table 1 The composition of the signature

Signature	RMOs (Ma > Mb) ^a		Hazard ratio	p value
Gene pair 1	JDP2	SERPINA5	2.4315	0.0005
Gene pair 2	JDP2	PLG	2.4974	0.0035
Gene pair 3	JDP2	SEMG2	2.5136	0.0020
Gene pair 4	RFX5	POLR3B	2.2728	0.0051

^a Represents the relative methylation ordering (RMOs) of gene pair (Ma > Mb), Ma and Mb represent the methylation value of genes a and b, respectively

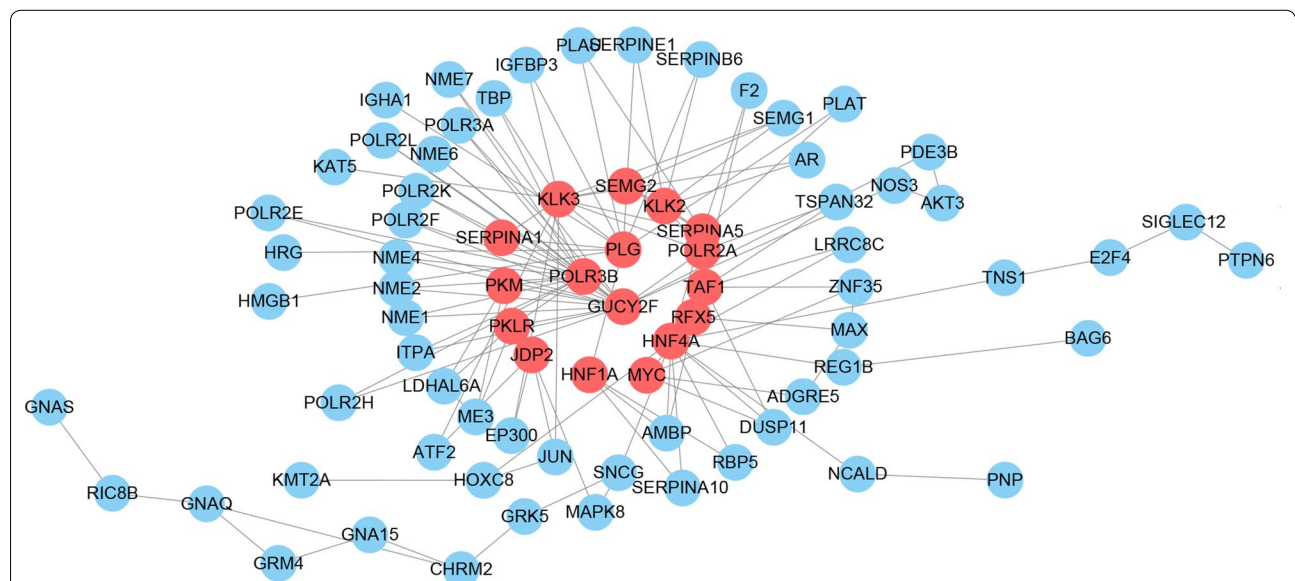


Fig. 1 PPI network based on the survival 151 genes. Node represents the gene in the PPI network. Red represents the degree of the gene is over 3; blue represents the degree of the gene less than 3

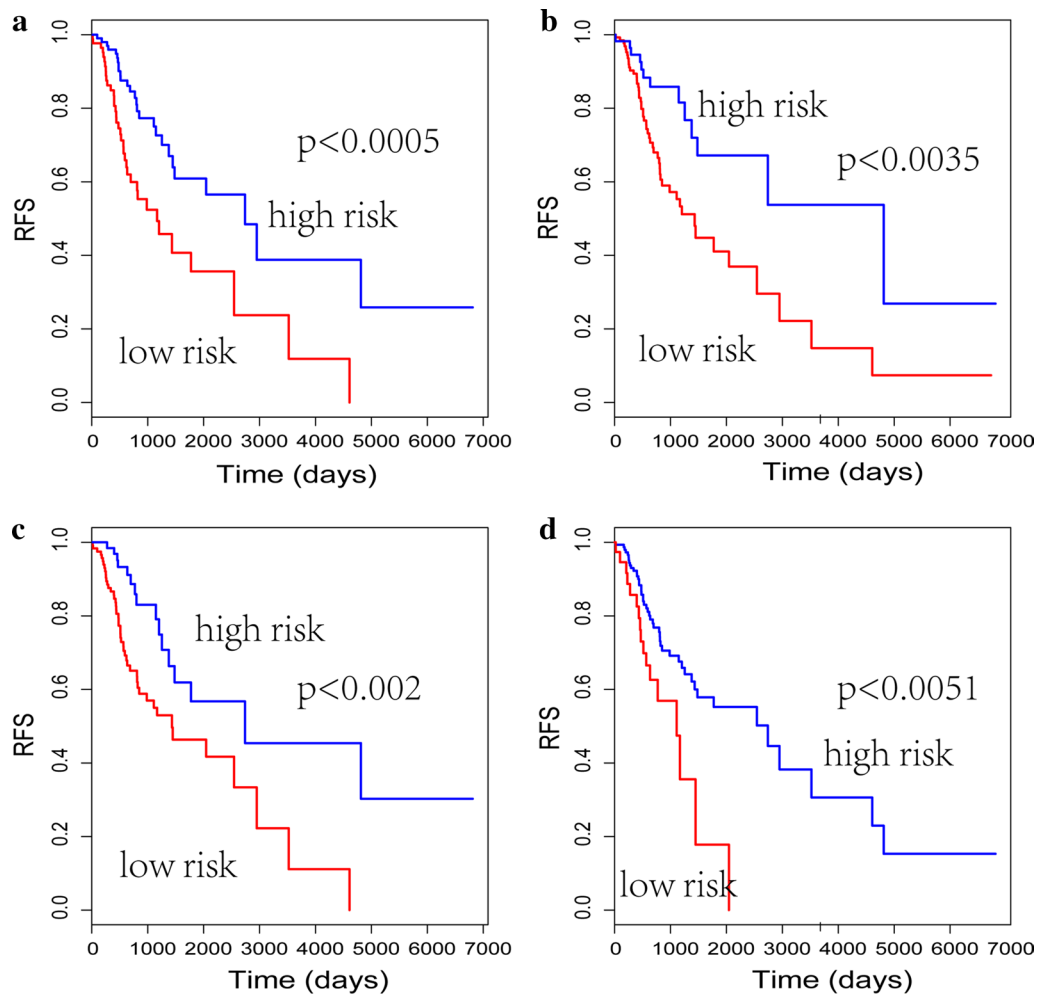


Fig. 2 The association between four gene-pair signature and the RFS of 191 LUAD patients. The log rank p value is displayed separately. **a** High- and low-risk group clustered by gene pair 1; **b** high- and low-risk group clustered by gene pair 2; **c** high- and low-risk group clustered by gene pair 3; **d** high- and low-risk group clustered by gene pair 4

Table 2 The clinical data for the 191 early stage LUAD

Covariates	Category	Total
Age years ^a	< 60	54
	≥ 60	131
Gender	Male	83
	Female	108
RFS status	Relapse	61
	Non-relapse	130
Smoker ^a	Non-smoker	141
	Smoker	46

^a Information is partially lacked

Enrichment analysis

The potential functions of 6 genes (JDP2, SERPINA5, PLG, SEMG2, RFX5, and POLR3B) in the 4-gene pair

signature were further analyzed in enrichment analysis together with all 5963 differentially methylated genes. A total of 560 hypermethylation genes were screened according to the threshold of $FDR < 0.01$ and $r > 0.6$. In GO term, all these genes were associated with intracellular protein synthesis and transportation. All the enriched pathways were listed in Table 3.

Discussion

In non-small cell lung cancer (NSCLC), lung adenocarcinoma (LUAD) is a common histological subtype with high mortality but poor outcome in both male and female, smokers and non-smokers [27]. To date, the current therapy of stage I LUAD patients is still surgery. However, almost half of the patients after surgical treatment suffer from recurrence or even death, resulting in a low 5-year survival rate [28, 29]. Therefore, an effective

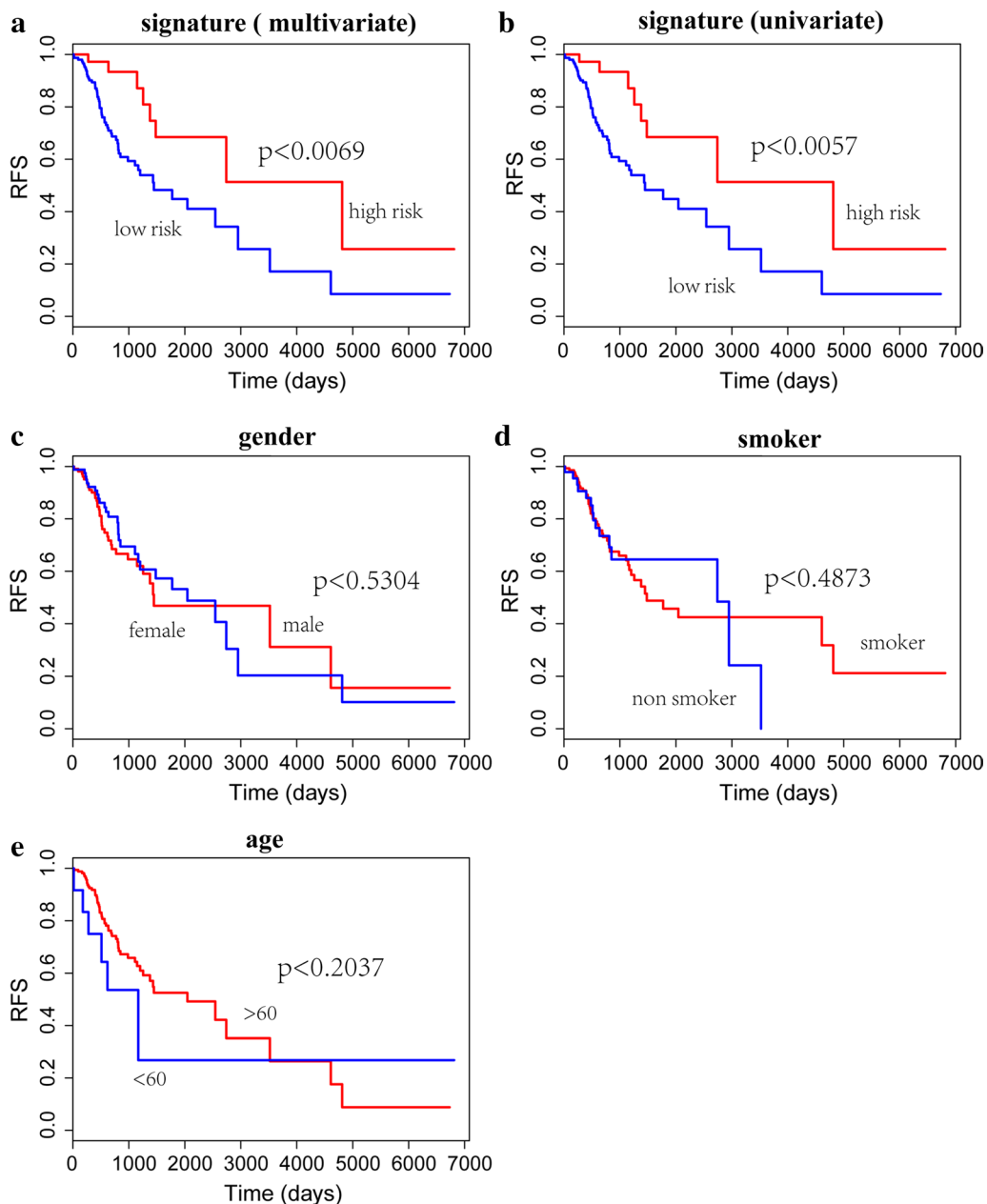


Fig. 3 Survival analysis of 4 gene-pair signature. This signature was able to classify the stage I LUAD patients into two groups with significant RFS (univariate COX $p < 0.0057$, multivariate COX $p < 0.0069$). **a** Multivariate survival analysis of the signature; **b** univariate survival analysis of the signature; **c** survival analysis on the impacts of gender; **d** survival analysis on the impacts of smoking; **e** survival analysis on the impacts of age. The log rank p value is displayed separately

prognostic signature to evaluate the survival outcomes or to predict the risk of postoperative recurrence is in need.

In the present research, we screened out 560 hypermethylation genes ($FDR < 0.01$ and $r > 0.6$). But for further construction of prognostic signature, 151 genes associating with non-recurrence of postoperative patients were used to analyze for hub genes. Finally, 17 hub genes were

obtained, indicating 9 key genes (KLK3, GUCY2F, KLK2, SERPINA5, PLG, SEMG2, RFX5, POLR3B, JDP2) relating to non-recurrence of postoperative patients with LUAD. In the selecting process of prognostic methylated genes, threshold of differential significance $p < 0.01$ was referred in univariate Cox analysis. In the present research, we generated a 4-gene-pair signature consisting of 6 genes

Table 3 Enriched pathways of methylated genes in signature

Pathway name	p-value
Cellular protein localization	5.47E-07
Cellular macromolecule localization	6.42E-07
Intracellular protein transport	8.81E-07
Intracellular transport	4.37E-05
Macromolecule localization	5.02E-05
Cellular protein metabolic process	7.15E-05
Protein localization	0.000123001
Cellular localization	0.000128524
Protein transport	0.000204877
Negative regulation of cellular macromolecule biosynthetic process	0.000244491
Primary metabolic process	0.000251601
Protein metabolic process	0.000260279
Negative regulation of nucleobase nucleoside nucleotide and nucleic acid metabolic process	0.000341604
Cellular macromolecule metabolic process	0.000350947
Establishment of protein localization	0.000404266
Negative regulation of nitrogen compound metabolic process	0.000458234
Negative regulation of biosynthetic process	0.000512494
Protein modification process	0.000516455
Negative regulation of macromolecule biosynthetic process	0.000562752
Negative regulation of cellular biosynthetic process	0.000615668
Protein targeting	0.000632614
Establishment of localization in cell	0.000691823

(JDP2, SERPINA5, PLG, SEMG2, RFX5 and POLR3B) with potential functions in LUAD. All these genes were enriched in intracellular protein synthesis and transportation pathways in GO term. However, further functional analysis specifically on every gene was in need to explore the mechanism of this prognostic signature.

Subsequently, the signature was applied to evaluate the impacts of prognostic signature casting on the RFS of stage I LUAD patients and to detect the relationships between RFS and age, gender and smoking condition. It is satisfied that the 4 gene-pair signature was able to classify the stage I LUAD patients into two groups with significant RFS (univariate COX $p < 0.0057$, multivariate COX $p < 0.0069$). Meanwhile, the clustering ability of this signature was not influence by the age, gender and smoking condition. All these indicated that this signature was reliable and stable for evaluation of the prognosis of LUAD.

To get further knowledge of the 6 hub genes involved in the prognostic signature, we carried out literature review to explore the functions or association in cancer. Among these 6 genes, several were identified related to cancer.

That is, most of the genes in the present study were firstly identified related to lung cancer.

Gene *JDP2* encodes Jun dimerization protein-2 that functions as AP-1 transcription factor. This protein can repress the transactivation mediated by the Jun family of proteins [30]. It has been identified that decreased-expression of *JDP2* was related to lymph node metastasis and distant metastasis. Besides, its depression was as well found casting strongly relationships with the post-surgery survival time. Yuanhong et al. stated that there might be a possible relationship between the expression of *JDP2* and metastasis in pancreatic carcinoma, suggesting *JDP2* as a prognostic biomarker for patients with pancreatic carcinoma [31]. Besides, research also discovered an association of epithelial-mesenchymal transition (EMT) with Jun dimerization protein 2 (*JDP2*) in pancreatic cancer [32]. Hereinto, EMT is considered to contribute to the invasion and metastasis of a variety of malignant tumors. Taken all these previous studies which conform to our discovery, we suggest the *JDP2* might play an important role in invasion and metastasis of LUAD as well. The signature model containing *JDP2* is reasonable when clustering the LUAD patients. Further research focusing on the specific functions of this gene in lung is still in need to complete the mechanism of the recurrence of LUAD at gene level. Besides, the probable cooperating genes of *JDP2* also deserve more attention.

Another cancer related gene discovered in our prognostic signature is *SERPINA5*, also called Protein C Inhibitor (PCI), belongs to the serine protease inhibitor super family as well. This gene is known to prevent metastasis and anti-angiogenesis in tumor cells, including renal, breast, prostate and ovarian cancers [33–37]. However, little research indentified this depression of *SERPINA5* in LUAD till now. The association between *SERPINA5* and this subtype of lung cancer is an inspiring result, but the expression status (at transcriptional or post-transcriptional level) and biological function of this gene in LUAD are largely unknown.

Semenogelin-2 (*SEMG2*) was ever found being catalyzed to degrade by the proteolytic activity of the active PSA-enzyme in prostate cancer [38] and that the functions in HER2+ cellular models of breast cancer [39]. As to RFX5 (Regulatory Factor X), being able to bind DNA and lend promoter specificity [40], was identified as a transcriptional activator of the *TPP1* gene in hepatocellular carcinoma [41]. Besides, RFX5 was tested to express differentially in breast cancer [42]. *POLR3B* (Polymerase RNA III beta subunit) is reported to code for the other subunit that forms Pol III's catalytic site, which transcribes small untranslated RNAs [43]. Researchers were searching the mutations in *POLR3*, so that its functions in several tumors could be revealed at mRNA levels by

using RNA-Seq analysis [44]. Thus, the identification of this gene in LUAD might provide a reference for further studies.

According to the discussions above, we supposed that the 4-gene-pair signature could be used to evaluate the prognosis of LUAD patients generally. However, considering the gene diversity of cancer patients and even within a pedigree, personalized insight is also the key point for cancer managements including the prognosis and therapy.

Conclusion

All in all, in the present study, we generated a 4-gene pair signature with 6 methylated genes. The signature was able to classify the stage I LUAD patients into two groups with significant RFS. However, age, gender and smoking condition did not influence the prediction of our prognostic signature. The potential functions of 6 genes were associated with intracellular protein synthesis and transportation. This signature might be used to provide clinical reference for postoperative chemotherapy of patients with stage I LUAD. In addition, our signature may independently predict the prognosis of LUAD without depending on data standardization.

Authors' contributions

YW, ZFM, and ZYW collected the data; YW, ZFM, and ZYW performed the statistical analysis; YW, ZFM, and ZYW wrote the paper; YW, ZFM, and ZYW conceived the study. All authors read and approved the final manuscript.

Author details

¹ Department of Radiation Oncology, Shanghai Minhang District Cancer Hospital, 106 Ruili Road, Shanghai 200240, China. ² Department of Pharmacy, The Affiliated Huai'an Hospital of Xuzhou Medical University and The Second People's Hospital of Huai'an, 62 South Huai'hai Rode, Huai'an, China. ³ Department of Medical Imaging, The Fourth People's Hospital of Huai'an, Huai'an, Jiangsu, China.

Acknowledgements

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The authors declare that the data supporting the findings of this study are available within the article.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

Not applicable.

Publisher's Note

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

Received: 16 August 2017 Accepted: 22 March 2018

Published online: 19 April 2018

References

1. Ettinger DS, Akerley W, Borghaei H, Chang AC, Cheney RT, Chirieac LR, Damico TA, Demmy TL, Govindan R, Grannis FW. Non-small cell lung cancer, version 2.2013. *J Natl Compr Canc Netw*. 2013;11(6):645–53.
2. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. *CA Cancer J Clin*. 2010;60(5):277–300.
3. Cancer IAfRo. GLOBOCAN 2012: estimated cancer incidence, mortality and prevalence worldwide in 2012. Lyon: Cancer IAfRo; 2012. p. 2012.
4. Williams D, Pairolero P, Davis C, Bernatz P, Payne W, Taylor W, Uhlenhopp M, Fontana R. Survival of patients surgically treated for stage I lung cancer. *J Thor Cardiovasc Surg*. 1981;82(1):70–6.
5. Grover FL, Piantadosi S. Recurrence and survival following resection of bronchioloalveolar carcinoma of the lung—the Lung Cancer Study Group experience. *Ann Surg*. 1989;209(6):779.
6. Youlden DR, Cramb SM, Baade PD. The international epidemiology of lung cancer: geographical distribution and secular trends. *J Thorac Oncol*. 2008;3(8):819–31.
7. Balch CM, Buzaid AC, Soong S-J, Atkins MB, Cascinelli N, Coit DG, Fleming ID, Gershenwald JE, Houghton A Jr, Kirkwood JM. Final version of the American Joint Committee on Cancer staging system for cutaneous melanoma. *J Clin Oncol*. 2001;19(16):3635–48.
8. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol*. 2010;17(6):1471–4.
9. Chansky K, Sculier JP, Crowley JJ, Giroux D, Van Meerbeeck J, Goldstraw P, International Staging C, Participating I. The International Association for the Study of Lung Cancer Staging Project. Prognostic factors and pathologic TNM stage in surgically managed non-small cell lung cancer. *Chin J Lung Cancer*. 2010;13(1):9–18.
10. Subramanian J, Simon R. Gene expression-based prognostic signatures in lung cancer: ready for clinical use? *J Natl Cancer Inst*. 2010;102(7):464–74.
11. Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. *Nat Rev Cancer*. 2005;5(11):845–56.
12. Hu Z, Chen X, Zhao Y, Tian T, Jin G, Shu Y, Chen Y, Xu L, Zen K, Zhang C, Shen H. Serum MicroRNA signatures identified in a genome-wide serum MicroRNA expression profiling predict survival of non-small-cell lung cancer. *J Clin Oncol*. 2010;28(10):1721–6.
13. Yu H, Jiang L, Sun C, Guo L, Lin M, Huang J, Zhu L. Decreased circulating miR-375: a potential biomarker for patients with non-small-cell lung cancer. *Gene*. 2014;534(1):60–5.
14. Das PM, Singal R. DNA methylation and cancer. *J Clin Oncol*. 2004;22(22):4632–42.
15. Dehan P, Kustermans G, Guenin S, Horion J, Boniver J, Delvenne P. DNA methylation and cancer diagnosis: new methods and applications. *Expert Rev Mol Diagn*. 2009;9(7):651–7.
16. Grunau C, Renault E, Rosenthal A, Roizes G. MethDB—a public database for DNA methylation data. *Nucleic Acids Res*. 2001;29(1):270.
17. Weisenberger DJ. Characterizing DNA methylation alterations from The Cancer Genome Atlas. *J Clin Invest*. 2014;124(1):17–23.
18. Zhou S, Wang P, Su X, Chen J, Chen H, Yang H, Fang A, Xie L, Yao Y, Yang J. High ECT2 expression is an independent prognostic factor for poor overall survival and recurrence-free survival in non-small cell lung adenocarcinoma. *PLoS ONE*. 2017;12(10):e0187356.
19. Hechenbichler K, Schliep K. Weighted k-nearest-neighbor techniques and ordinal classification. Ludwig: University Munich; 2004.
20. Kobayashi Y, Absher DM, Gulzar ZG, Young SR, McKenney JK, Peehl DM, Brooks JD, Myers RM, Sherlock G. DNA methylation profiling reveals novel biomarkers and important roles for DNA methyltransferases in prostate cancer. *Genome Res*. 2011;21(7):1017–27.
21. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B (Methodological)*. 1995;57:289–300.
22. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504.
23. Peretto L, Briganti L, Calderone A, Perpetuini AC, Iannuccelli M, Langone F, Licata L, Marinkovic M, Mattioni A, Pavlidou T, et al. SIGNOR: a database of causal relationships between biological entities. *Nucleic Acids Res*. 2016;44(D1):D548–54.

24. Cd R. Regression models and life-tables. *J R Stat Soc.* 1972;34(2):187–220.
25. Jones MP, Crowley J. A general class of nonparametric tests for survival analysis. *Biometrics.* 1989;45(1):157–70.
26. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 2003;4(5):P3.
27. International Agency for Research on Cancer. Globocan I: estimated cancer incidence, mortality and prevalence worldwide in 2012. Lyon: International Agency for Research on Cancer; 2012.
28. Esposito L, Conti D, Ailavajhala R, Khalil N, Giordano A. Lung cancer: are we up to the challenge? *Curr Genomics.* 2010;11(7):513–8.
29. Fan X, Zhang X, Wang H, Jin B. Reevaluation of survival and prognostic factors in pathologic stage I lung adenocarcinoma by the new 2009 TNM classification. *Tumour Biol J Int Soc Oncodev Biol Med.* 2014;35(6):5905–10.
30. Jin C, Kato K, Chimura T, Yamasaki T, Nakade K, Murata T, Li H, Pan J, Zhao M, Sun K. Regulation of histone acetylation and nucleosome assembly by transcription factor JDP2. *Nat Struct Mol Biol.* 2006;13(4):331–8.
31. Yuanhong X, Feng X, Qingchang L, Jianpeng F, Zhe L, Kejian G. Downregulation of AP-1 repressor JDP2 is associated with tumor metastasis and poor prognosis in patients with pancreatic carcinoma. *Int J Biol Markers.* 2009;25(3):136–40.
32. Xu Y, Liu Z, Guo K. The effect of JDP2 and ATF2 on the epithelial–mesenchymal transition of human pancreatic cancer cell lines. *Pathol Oncol Res.* 2012;18(3):571–7.
33. Asanuma K, Yoshikawa T, Hayashi T, Akita N, Nakagawa N, Hamada Y, Nishioka J, Kamada H, Gabazza EC, Ido M, et al. Protein C inhibitor inhibits breast cancer cell growth, metastasis and angiogenesis independently of its protease inhibitory activity. *Int J Cancer.* 2007;121(5):955–65.
34. Cao Y, Becker C, Lundwall A, Christensson A, Gadaleanu V, Lilja H, Bjartell A. Expression of protein C inhibitor (PCI) in benign and malignant prostatic tissues. *Prostate.* 2003;57(3):196–204.
35. Sil H, Sen T, Chatterjee A. Fibronectin-integrin (alpha5beta1) modulates migration and invasion of murine melanoma cell line B16F10 by involving MMP-9. *Oncol Res.* 2011;19(7):335–48.
36. Jing Y, Jia D, Wong CM, Oi-Lin Ng I, Zhang Z, Liu L, Wang Q, Zhao F, Li J, Yao M, et al. SERPINA5 inhibits tumor cell migration by modulating the fibronectin-integrin beta1 signaling pathway in hepatocellular carcinoma. *Mol Oncol.* 2014;8(2):366–77.
37. Sigurdson AJ, Brenner AV, Roach JA, Goudeva L, Muller JA, Nerlich K, Reiners C, Schwab R, Pfeiffer L, Waldenberger M, et al. Selected single-nucleotide polymorphisms in FOXE1, SERPINA5, FTO, EVPL, TICAM1 and SCARB1 are associated with papillary and follicular thyroid cancer risk: replication study in a German population. *Carcinogenesis.* 2016;37(7):677–84.
38. Bangma CH, van Schaik RH, Blijenberg BG, Roobol MJ, Lilja H, Stenman U-H. On the use of prostate-specific antigen for screening of prostate cancer in European Randomised Study for Screening of Prostate Cancer. *Eur J Cancer.* 2010;46(17):3109–19.
39. Kulkarni YM, Klinke DJ. Protein-based identification of quantitative trait loci associated with malignant transformation in two HER2+ cellular models of breast cancer. *Proteome science.* 2012;10(1):11.
40. Scholl T, Mahanta SK, Strominger JL. Specific complex formation between the type II bare lymphocyte syndrome-associated transactivators CIITA and RFX5. *Proc Natl Acad Sci.* 1997;94(12):6330–4.
41. Zhao Y, Xie X, Liao W, Zhang H, Cao H, Fei R, Wang X, Wei L, Shao Q, Chen H. The transcription factor RFX5 is a transcriptional activator of the TPP1 gene in hepatocellular carcinoma. *Oncol Rep.* 2017;37(1):289–96.
42. Iglesias-Martinez LF, Kolch W, Santra T. Supplementary material: BGRM1: a method for inferring gene regulatory networks from time course gene expression data and its application on breast cancer research. *Sci Rep.* 2016;6:37140.
43. Tétreault M, Choquet K, Orcesi S, Tonduti D, Balottin U, Teichmann M, Fribourg S, Schiffmann R, Brais B, Vanderver A. Recessive mutations in POLR3B, encoding the second largest subunit of Pol III, cause a rare hypomyelinating leukodystrophy. *Am J Hum Genet.* 2011;89(5):652–5.
44. Seidel NE, Farrar J, Vlachos A, Anderson SM, An X, Atsidaftos E, Malech H, Arceci RJ, Ellis SR, Lipton JM. Autosomal recessive diamond-blackfan anemia: identification of mutations in MCM2 and FlnB. *Blood.* 2013;122(21):589.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

