Predicting the survival of patients with lung adenocarcinoma using a four-gene prognosis risk model

WEI ZHANG, YANG SHEN and GANZHU FENG

Department of Respiratory Medicine, The Second Affiliated Hospital of
Nanjing Medical University, Nanjing, Jiangsu 210011, P.R. China

Abstract. Lung adenocarcinoma (LAD) is difficult to diagnose as it tends to be small in size and metastasize early. The aim of the present study was to investigate prognostic factors for patients with LAD and establish a prognosis risk model. A training set consisting of clinical and RNA sequencing data from 503 patients with LAD, as well as expression data from a further 59 LAD and adjacent tissues, was obtained from The Cancer Genome Atlas. Additionally, a validation dataset was acquired from the Gene Expression Omnibus database (GSE26939), which included clinical and gene expression data from 115 patients. Using the DESeq2 package to compare expression between LAD and adjacent tissues, differentially expressed genes (DEGs) were identified. On the basis of survival and the random forests for survival, regression and classification package, genes for constructing the prognosis risk model were selected. The prognosis risk model was constructed and validated using the survival package. Subsequently, high- and low-risk groups were compared using the Limma package to identify DEGs, and enrichment analysis was performed using the web-based gene set analysis toolkit. A protein-protein interaction network was visualized using Cytoscape software. There were 18,567 DEGs between the LAD samples and the adjacent tissues, and 363 DEGs between the high- and low-risk groups. Of these, four genes were selected for constructing the prognosis risk model, myosin IE (*MYO1E*), endoplasmic reticulum oxidoreductase 1α (*ERO1L*), C1q and tumor necrosis factor-related protein 6 (*C1QTNF6*) and family with sequence similarity 83, member A (*FAM83A*). The survival time of high- and low-risk groups in the validation set were significantly different. Functional enrichment revealed that the genes that interacted with *MYO1E*, *ERO1L*, *C1QTNF6* and *FAM83A* separately were enriched in 'cell cycle regulation', 'synthesis and assembly of nucleic acids', 'histone modification and cell cycle progression' and 'cell secretion process'. The four-gene prognosis risk model could potentially be used for predicting the survival of patients with LAD.

Introduction

Lung adenocarcinoma (LAD), also termed pulmonary adenocarcinoma, is a common type of lung cancer that often occurs in the peripheral lung tissue (1,2). LAD is induced by smoking, but is also the most frequent type of lung cancer in non-smokers (3,4). The occurrence of LAD in the periphery of the lung may be due to the nature of cigarette filters, which keep larger particles out of the lungs (5). LAD is difficult to diagnose owing to its small mass, and it readily enters the regional lymph nodes and tends to metastasize at an early stage (6). LAD is a type of non-small cell lung carcinoma, and its response to radiation therapy is worse compared with that of small cell lung carcinoma (7). Further understanding of the mechanism of LAD development may result in improved prognosis for patients.

Sex-determining region box 2 can serve as a marker predicting poor outcome and may assist in identifying a high risk of recurrence following surgery in patients with stage I LAD (8,9). Zhang *et al* (10) reported an analysis of the clinico-pathological features of LAD that identified the upregulation of programmed death 1 ligand 1 (*PD-L1*) and *PD-L2*, indicating that *PD-L* expression is an independent predictor of LAD (10). The high expression of metastasis-associated in colon cancer 1 may be a promising predictor of postoperative recurrence in patients with LAD following surgery (11,12). Upregulation of the cytochrome P450 family 24 subfamily A member 1 (*CYP24A1*) in patients with LAD is associated with poor survival, which may be due to the abrogation of anti-proliferative effects of the active form of vitamin D, $1\alpha,25$-dihydroxyvitamin D_3 (13). Upregulation of histone deacetylase 1 in tumor cells can be used as a marker for poor prognosis in patients with LAD (14). Nevertheless, the previous studies have not comprehensively revealed prognostic factors for LAD. In the present study, RNA-sequencing data of LAD tissues was obtained from The Cancer Genome Atlas (TCGA), and the genes associated with the survival rate of patients with

Correspondence to: Dr Ganzhu Feng, Department of Respiratory Medicine, The Second Affiliated Hospital of Nanjing Medical University, 121 Jiangjiayuan Road, Nanjing, Jiangsu 210011, P.R. China
E-mail: zhangwei3125@sina.com

LAD were explored. On this basis, a prognosis risk model was constructed, and validated using microarray data from a separate cohort of patients with LAD. Furthermore, the biological significance of the prognostic differences in patients with LAD was investigated.

Materials and methods

Expression profile data. RNA-sequencing data, including raw count and fragments per kilobase of transcript per million fragments mapped (FPKM) and the relevant clinical data were downloaded from the Genomic Data Commons database (GDC; https://portal.gdc.cancer.gov). There were expression data and clinical data for 503 tissues from patients with LAD in the TCGA dataset (TCGA-LAD), which were taken as the training set for constructing the prognosis risk model for LAD. Additionally, the expression data of a further 59 pairs of LAD tissues and adjacent tissues were obtained.

Gene expression data and clinical data were acquired from the Gene Expression Omnibus database (GEO; http://www.ncbi.nlm.nih.gov/geo) (GSE26939). A total of 115 samples were selected as the validation set for the prognosis risk model. The clinical data of the training set and the validation set are presented in Table I.

Screening for differentially expressed genes (DEGs) between LAD and adjacent tissues. Using raw data from 59 pairs of LAD and adjacent tissues, DEGs between the two groups were identified using the R 3.1.0 software package DESeq2 1.4.5 (http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html) (15). The thresholds were $\log_2$ fold change ($|\log_2 FC|$) >1 and false discovery rate (FDR) <0.05.

Identification of genes for constructing the prognosis risk model. In combination with the clinical data and the gene expression data (FPKM values) for each patient, the Cox univariate regression analysis in the survival analysis package for R (https://cran.r-project.org/web/packages/survival/index.html) (16) was used to identify the survival time-associated genes from the DEGs. The log-rank test (17), sc-test (Rao's score test) (18) and Wald test (19) were utilized to screen genes with P<0.01. Subsequently, the random forests for survival, regression and classification package in R (https://cran.r-project.org/web/packages/randomForestSRC/index.html) (20) was applied for ranking the survival time-associated DEGs by survival contribution degrees. Finally, the genes with survival contribution degree >0.002 were selected as the genes for constructing the prognosis risk model.

Construction and evaluation of prognosis risk model. According to the FPKM values of the selected genes and the survival status of the patients, the risk scoring formula for prognosis of LAD was constructed using the Cox multivariate regression analysis in R package survival (16). Subsequently, the risk scores of the 503 patients with LAD were calculated. To assess the predictive independence of the risk scores in the training set, the risk scores and eight clinical factors, including age, sex, smoking status, stage, epidermal growth factor receptor (EGFR), KRAS proto-oncogene,

Table I. Clinical characteristics of patients in the training set and validation set.

Clinical characteristic	Training set (n=503)	Validation set (n=115)
Age, years ± SD	63.0±16.4	63.9±10.9
Sex, n		
Female	233	50
Male	270	49
Unknown	0	16
Smoking status, n		
Smoker	173	100
Non-smoker	28	12
Unknown	302	3
Stage, n		
I	270	55
II	119	16
III	81	14
IV	25	2
Unknown	8	28
Epidermal growth factor receptor, n		
WT	189	70
Mutated	31	11
Unknown	283	34
KRAS proto-oncogene, n		
WT	154	88
Mutated	66	20
Unknown	283	7
Serine/threonine kinase 11, n		
WT	182	75
Mutated	38	6
Unknown	283	34
Subtype, n		
Bronchioid	85	46
Magnoid	61	40
Squamoid	74	29
Unknown	283	0
Survival time, months ± SD	30.3±29.9	40.5±35.1
Survival status, n		
Alive	320	49
Deceased	183	66

WT, wild-type; SD, standard deviation.

serine/threonine kinase 11 and subtype were subjected to Cox univariate and multivariate regression analyses. In addition, stratified analysis was performed for other survival-associated factors. Thereafter, the patients were divided into high- and low-risk groups according to the median of risk scores. The Kaplan-Meier (KM) survival curve (21) was used to evaluate the survival difference between the two groups of patients. Significant differences in the tumor stages of patients in the

Table II. Univariate and multivariate Cox regression analyses for *FAM83A*, *MYO1E*, *C1QTNF6* and *ERO1L*.

Gene	Univariate Cox regression			Multivariate Cox regression		
	Coefficient	HR	P-value	Coefficient	HR	P-value
FAM83A	9.94×10^{-3}	1.01	1.29×10^{-13}	6.35×10^{-3}	1.01	9.05×10^{-5}
MYO1E	4.18×10^{-2}	1.04	2.86×10^{-7}	2.11×10^{-2}	1.02	5.87×10^{-2}
C1QTNF6	8.59×10^{-2}	1.09	1.28×10^{-8}	4.23×10^{-2}	1.04	2.22×10^{-2}
ERO1L	1.27×10^{-2}	1.01	2.00×10^{-8}	6.34×10^{-3}	1.01	2.29×10^{-2}

FAM83A, family with sequence similarity 83, member A; *MYO1E*, myosin IE; *C1QTNF6*, C1q and tumor necrosis factor-related protein 6; *ERO1L*, endoplasmic reticulum oxidoreductase 1α; HR, hazard ratio.

high- and low-risk groups were identified using a χ^2 test with a threshold value of P<0.05.

Validation of the prognosis risk model. Using the expression and clinical data from the GSE26939 dataset, the risk scores of 115 patients with LAD were calculated using the risk scoring formula. Similarly, the 115 patients were classified into high- and low-risk groups. The KM survival curve (21) was used to assess the survival difference between the two groups of patients.

Biological analysis of DEGs between the high- and low-risk groups. On the basis of the prognosis risk model, the 503 LAD samples were divided into two groups. Using the R package Limma (http://www.bioconductor.org/packages/release/bioc/html/limma.html) (22), DEGs between the two groups were identified, using the thresholds of |log$_2$ FC| >1 and FDR <0.05. Next, the functional and pathway enrichment analyses were conducted for the DEGs using the web-based gene set analysis toolkit (WebGestalt) (http://www.webgestalt.org/option.php) (23).

Protein-protein interaction (PPI) network analysis. Using the PPI information in the InWeb_InBioMap database (https://www.intomics.com/inbio/map/#home) (24), the PPI network for the genes involved in the prognosis risk model was constructed using Cytoscape software (http://www.cytoscape.org) (25). The WebGestalt tool (23) was used to perform functional enrichment analysis for the genes identified to interact with the genes of the prognosis risk model.

Results

Screening for DEGs between LAD and adjacent tissues. There were a total of 18,567 DEGs between the LAD and adjacent tissues, including 15,483 genes that were significantly upregulated in the LAD tissues and 3,084 genes that were significantly downregulated.

Identification of genes for constructing the prognosis risk model. Cox univariate regression analysis showed that a total of 279 DEGs were correlated with survival time of patients. Subsequently, the 279 DEGs were ranked by survival contribution degrees, and the top four genes were selected for constructing the prognosis risk model (Table II).

These four genes were family with sequence similarity 83, member A (*FAM83A*; contribution degree=0.00309), myosin IE (*MYO1E*; contribution degree=0.00231), C1q and tumor necrosis factor-related protein 6 (*C1QTNF6;* contribution degree=0.00226) and endoplasmic reticulum oxidoreductase 1α (*ERO1L*; contribution degree=0.00211).

Construction and evaluation of prognosis risk model. According to the FPKM values of the four selected genes and the survival status of the patients, the following risk scoring formula for prognosis of LAD was constructed: Risk score = (0.00635 x FAM83A) + (0.02114 x MYO1E) + (0.04234 x C1QTNF6) + (0.00634 x ERO1L).

The risk scores of the 503 patients with LAD were calculated, and then the risk scores and the eight clinical factors were subjected to correlation and regression analyses. The results indicated that the risk score and the stage of LAD were significantly correlated with patient survival time (Table III). Based on the median risk score, the patients were divided into high- and low-risk groups. The risk scores and survival status of the 2 groups were statistically analyzed, showing that the high-risk group had more mortality cases (high-risk group, 114/252 patients; low-risk group, 68/251 patients) and a significantly lower overall survival time (high-risk group, 25.9±24.8 months; low-risk group, 34.7±33.6 months; P=4.44×10^{-4}) compared with the low-risk group (Fig. 1). Using KM survival analysis, the survival difference between the 2 groups of patients was evaluated (Fig. 2A).

Tumor stage can also result in a difference in the patient survival time (Fig. 2B); therefore, the number of patients at each stage was evaluated. There were 109, 74, 53 and 15 patients in stage I, II, III and IV in the high-risk group, respectively. In the low-risk group, there were 161, 45, 28 and 10 patients at stage I, II, III and IV, respectively. The tumor stages of patients in the high- and low-risk groups were identified to be significantly different using a χ^2 test (P=1.10×10^{-5}).

In addition, stratified analysis was performed for the tumor stage (stage I, II, III and IV). The risk scores of the patients at each stage were calculated, and the patients were divided into high- and low-risk groups. The survival difference between the two groups at each stage was assessed using KM survival analysis (Fig. 2C-F). The results suggested that the prognosis risk model could significantly differentiate between patients at early stages (P<0.05), but could not distinguish patients at

Table III. Univariate and multivariate Cox regression analyses for risk score and eight clinical factors in the training set.

Clinical factor	Univariate Cox regression				Multivariate Cox regression			
	P-value	HR	Lower 95% CI	Upper 95% CI	P-value	HR	Lower 95% CI	Upper 95% CI
Risk score	<0.001	1.66	1.48	1.86	<0.001	1.73	1.36	2.19
Age	0.05	1.01	1.00	1.02	0.02	1.01	1.00	1.03
Sex (female/male)	0.78	1.04	0.78	1.39	0.49	1.20	0.71	2.04
Smoking status (smoker/non-smoker)	0.30	0.73	0.42	1.29	0.15	0.63	0.33	1.19
Stage (I/II/III/IV)	<0.001	3.33	2.14	5.20	0.02	2.68	1.29	5.72
Epidermal growth factor receptor (WT/mutated)	0.42	0.79	0.44	1.39	0.07	0.55	0.28	1.06
KRAS proto-oncogene (WT/mutated)	0.39	1.23	0.76	2.02	0.73	1.10	0.64	1.89
Serine/threonine kinase 11 (WT/mutated)	0.21	0.72	0.44	1.18	0.14	0.63	0.34	1.16
Subtype (bronchioid/magnoid/squamoid)	0.07	1.69	1.01	2.83	0.55	0.81	0.42	1.57

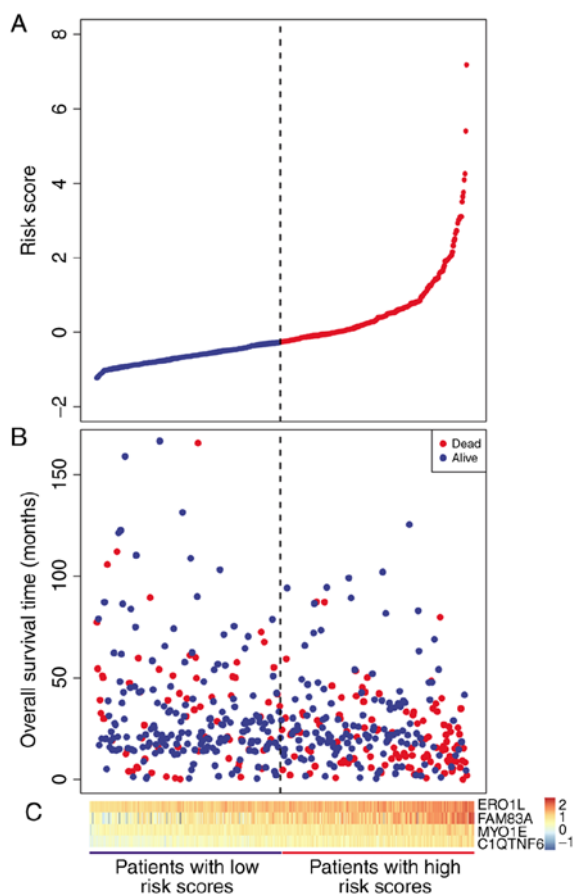HR, hazard ratio; CI, confidence interval; WT, wild-type.

Figure 1. Risk scores and survival status of the patients in the high- and low-risk groups of the training set. (A) Risk scores of the patients in the high (red) and low (blue) -risk groups. (B) Survival status of the patients in the two groups. Blue and red represent surviving and deceased patients, respectively. (C) Expression of *ERO1L*, *FAM83A*, *MYO1E* and *C1QTNF6* in each patient. *ERO1L*, endoplasmic reticulum oxidoreductase 1α; *FAM83A*, family with sequence similarity 83, member A; *MYO1E*, myosin IE; *C1QTNF6*, C1q and tumor necrosis factor-related protein 6.

later stages (P>0.05). Thus, the prognosis risk model was more efficient for patients in the early stages of LAD.

Validation of the prognosis risk model. Using the prognosis risk model, the risk scores of the 115 patients with LAD from GSE26939 were calculated. Similarly, regression analyses were performed for the risk scores and the eight clinical factors, demonstrating that the risk score was significantly correlated with the patient survival time (Table IV).

On the basis of the risk scores, the patients were divided into high- and low-risk groups. Statistical analysis revealed a significantly lower overall survival time in the high-risk group compared with that in the low-risk group (high-risk group, 34.2±31.3 months; low-risk group, 46.9±37.9 months; P=0.026; Fig. 3). There was a significant difference in the survival time between the high- and low-risk groups in the validation set, indicating that the prognosis risk model was efficient and reliable (Fig. 4).

Biological analysis of DEGs between the high- and low-risk groups. On the basis of the prognosis risk model, the 503 LAD samples were divided into high- and low-risk groups. A total of 363 DEGs between high- and low-risk groups were identified, including 188 genes that were upregulated in the high-risk group and 175 genes that were downregulated.

Gene Ontology enrichment analysis demonstrated that the DEGs were involved in extracellular structure (extracellular matrix organization, extracellular structure organization, collagen fibril organization), the circulatory system (circulatory system development, blood vessel development, vasculature development, angiogenesis), cell adhesion and cell migration (Fig. 5A). Pathway enrichment analysis indicated that the DEGs were mainly enriched in focal adhesion, ECM-receptor interaction and protein digestion and absorption pathways (Fig. 5B).
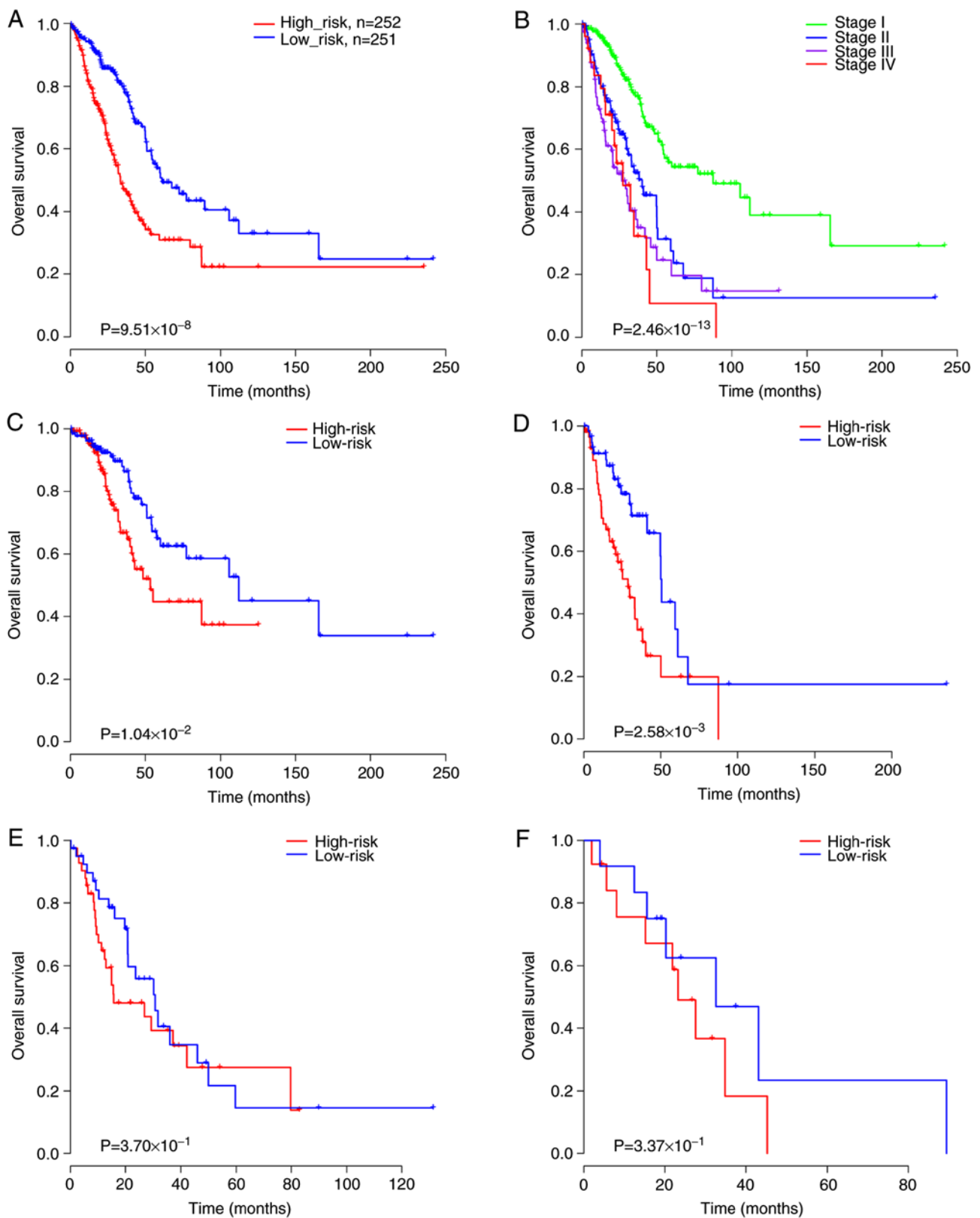
Figure 2. Kaplan-Meier survival curves comparing overall survival rates on the basis of tumor stage and risk score. (A) Patients in the high- and low-risk groups on the basis of median risk score of the prognosis risk model. (B) Patients with tumors at all stages. Patients with tumors at (C) stage I, (D) stage II, (E) stage III and (F) stage IV in the high- and low-risk groups according to risk score.

No significant difference was identified by χ^2 test between the high- and low-risk groups in terms of the occurrence of relapse (high-risk group, 82/252 patients; low-risk group, 70/251 patients; P=0.2989). There was also no significant difference between the high- and low-risk groups in terms of the occurrence of metastasis (high-risk group, 17/188 patients with relevant information; low-risk group, 7/170 patients with relevant information; P=0.09914). This may be due to limitations in the clinical factors.

Table IV. Univariate and multivariate Cox regression analyses for risk score and eight clinical factors in the validation set.

Clinical factor	Univariate Cox regression				Multivariate Cox regression			
	P-value	HR	Lower 95% CI	Upper 95% CI	P-value	HR	Lower 95% CI	Upper 95% CI
Risk score	<0.001	248.55	25.21	2450.60	0.02	73.67	2.22	2445.24
Age	0.02	1.03	1.00	1.05	0.15	1.03	0.99	1.07
Sex (female/male)	0.07	1.64	0.96	2.78	0.18	1.67	0.79	3.51
Smoking status (smoker/nonsmoker)	0.96	0.98	0.44	2.15	0.87	1.21	0.12	11.89
Stage (I/II/III/IV)	0.03	3.16	1.06	10.76	0.16	4.50	0.99	23.49
Epidermal growth factor receptor (WT/mutated)	0.22	0.58	0.23	1.47	1.00	0.00	0.00	Infinity
KRAS proto-oncogene (WT/mutated)	0.41	1.30	0.70	2.41	0.93	0.95	0.28	3.24
Serine/threonine kinase 11 (WT/mutated)	0.40	0.63	0.19	2.03	0.81	0.84	0.21	3.36
Subtype (bronchioid/ magnoid/squamoid)	0.84	1.13	0.63	2.06	0.46	0.67	0.23	1.94

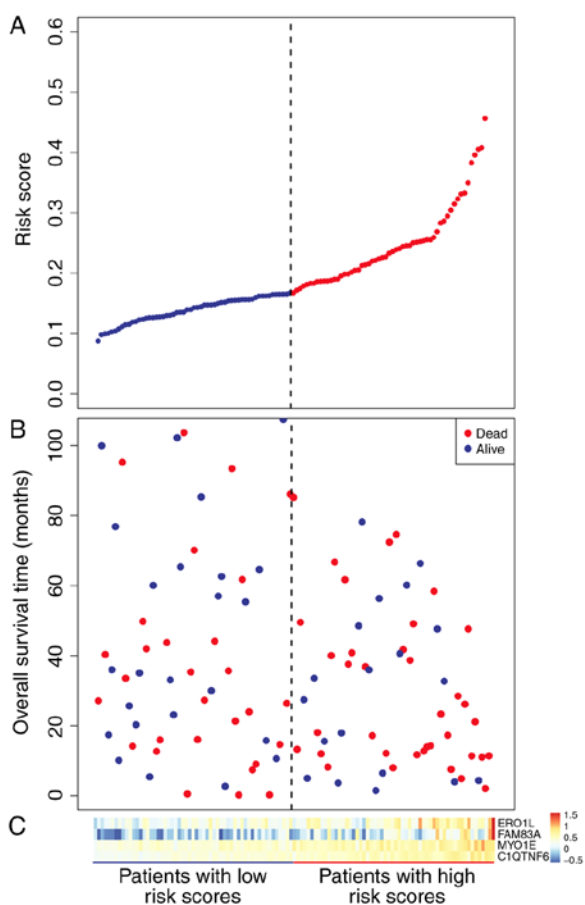HR, hazard ratio; CI, confidence interval; WT, wild-type.

Figure 3. Risk scores and survival status of patients in the high- and low-risk groups of the validation set. (A) Risk scores of the patients in the high (red) and low (blue) -risk groups. (B) Survival status of the patients in the two groups. Blue and red represent surviving and deceased patients, respectively. (C) Expression of *ERO1L*, *FAM83A*, *MYO1E* and *C1QTNF6* in each patient. *ERO1L*, endoplasmic reticulum oxidoreductase 1α; *FAM83A*, family with sequence similarity 83, member A; *MYO1E*, myosin IE; *C1QTNF6*, C1q and tumor necrosis factor-related protein 6.
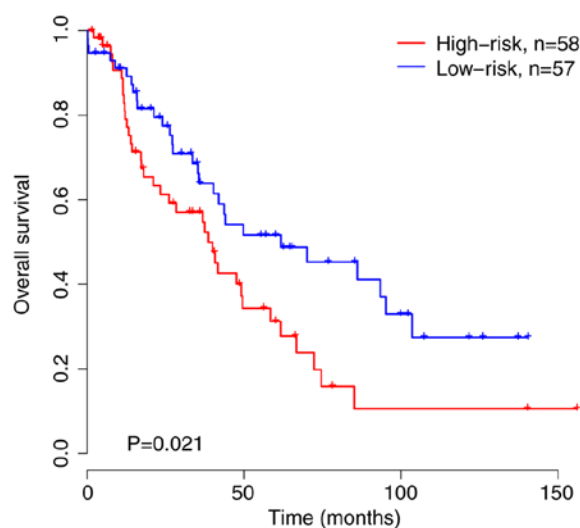

Figure 4. Kaplan-Meier survival curve for the high- and low-risk groups of the validation set.

PPI network analysis. In the PPI network for the four genes involved in the prognosis risk model, there were 266, 154, 10 and 8 nodes separately interacting with *MYO1E*, *ERO1L*, *C1QTNF6* and *FAM83A* (Fig. 6). Functional enrichment analysis indicated that the interacting genes of *MYO1E*, *ERO1L*, *C1QTNF6* and *FAM83A* were mainly involved in M/G1 transition of mitotic cell cycle, ribosome biogenesis, regulation of centrosome duplication and mitotic cell cycle, respectively (Fig. 7).

Discussion

In the present study, a total of 18,567 DEGs (including 15,483 upregulated and 3,084 downregulated genes) between LAD
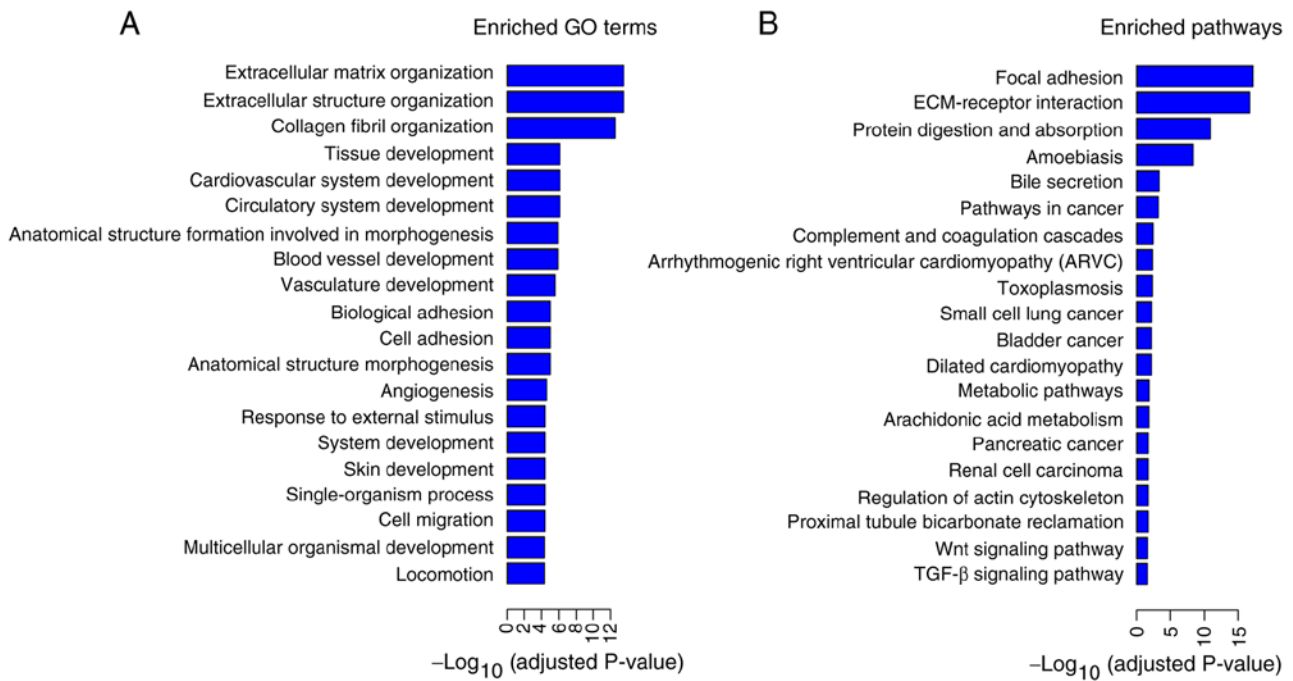
Figure 5. (A) GO terms and (B) pathways enriched for differentially expressed genes between the high- and low-risk groups. GO, Gene Ontology; ECM, extracellular matrix.

and adjacent tissues were screened. According to the survival contribution degrees, the top four genes (MYO1E, ERO1L, C1QTNF6 and FAM83A) were selected to construct the prognosis risk model. The survival time of the high- and low-risk groups in the validation set were significantly different, indicating that the prognosis risk model was efficient and reliable. The 503 LAD samples were divided into high- and low-risk groups on the basis of the prognosis risk model, and 363 DEGs (including 188 upregulated and 175 downregulated genes in the high-risk group) between the two groups were identified. There was no significant difference between the high- and low-risk groups in terms of the occurrence of relapse or metastasis, suggesting that the prognosis risk model is limited.

The oxidizing enzyme encoded by the ERO1L gene is expressed in the endoplasmic reticulum, and its expression in a variety of cancer cells is higher compared with that in normal tissue (26). ERO1L contributes to the formation of disulfide bonds on the cell surface and secreted proteins, in combination with the protein disulfide-isomerase, and may be a potential target for cancer immunotherapy (27). ERO1L overexpression is detected in various tumor types, and it suppresses the T cell response in the tumor by promoting the production of myeloid-derived suppressor cells through the regulation of cytokines and chemokines (28,29). ERO1L serves an important role in mediating tumor growth factors, including vascular endothelial growth factor, and its expression has been correlated with poor prognosis in patients with cancer (30). Therefore, ERO1L may be involved in the prognosis of patients with LAD.

FAM83 family members (FAM83A, FAM83B and FAM83D) exhibit oncogenic potential and are involved in tumor development, and may serve as promising targets for cancer therapy (31). FAM83 proteins are transforming oncoproteins that regulate EGFR/RAS signaling, and they may prove to be targets for novel tumor-targeting therapeutic strategies (32). Overexpression of FAM83 proteins leads to higher tumor grade and decreased overall survival, and thus FAM83 proteins are oncoproteins that can be targeted for tumor treatment aimed at inhibiting mitogen-activated protein kinase signaling (33,34). The mRNA expression of FAM83B in lung squamous cell carcinoma (LSCC) is significantly higher compared with that in LAD or normal lung, indicating that FAM83B may serve as a diagnostic and prognostic biomarker for LSCC (35). This indicates that FAM83A may serve an important role in the prognosis of patients with LAD.

MYO1E regulates the migration, adhesion, endocytosis and invadosome dynamics in podocyte cells, which may be used as a prognostic biomarker in clinical applications (36). Cyclin-dependent kinases mediate the onset of mitosis, cell cycle commitment and DNA synthesis, which are involved in tumor development and are potential targets for cancer treatment (37,38). Tumor-secreted factors are precursors for tumor diffusion and migration, which serve critical roles in tumor metastasis and deterioration (39). Functional enrichment analysis revealed that the genes interacting with MYO1E, ERO1L, C1QTNF6 and FAM83A were involved in cell cycle regulation, synthesis and assembly of nucleic acids, histone modification and cell cycle progression, and cell secretion process, suggesting that MYO1E, ERO1L, C1QTNF6 and FAM83A may be prognostic biomarkers for LAD.

There are a number of limitations to the present study. The results obtained in the present study were not validated through in vitro or in vivo laboratory experiments or clinical study. Therefore, further studies should be designed and performed to confirm these results. Future studies are required to analyze
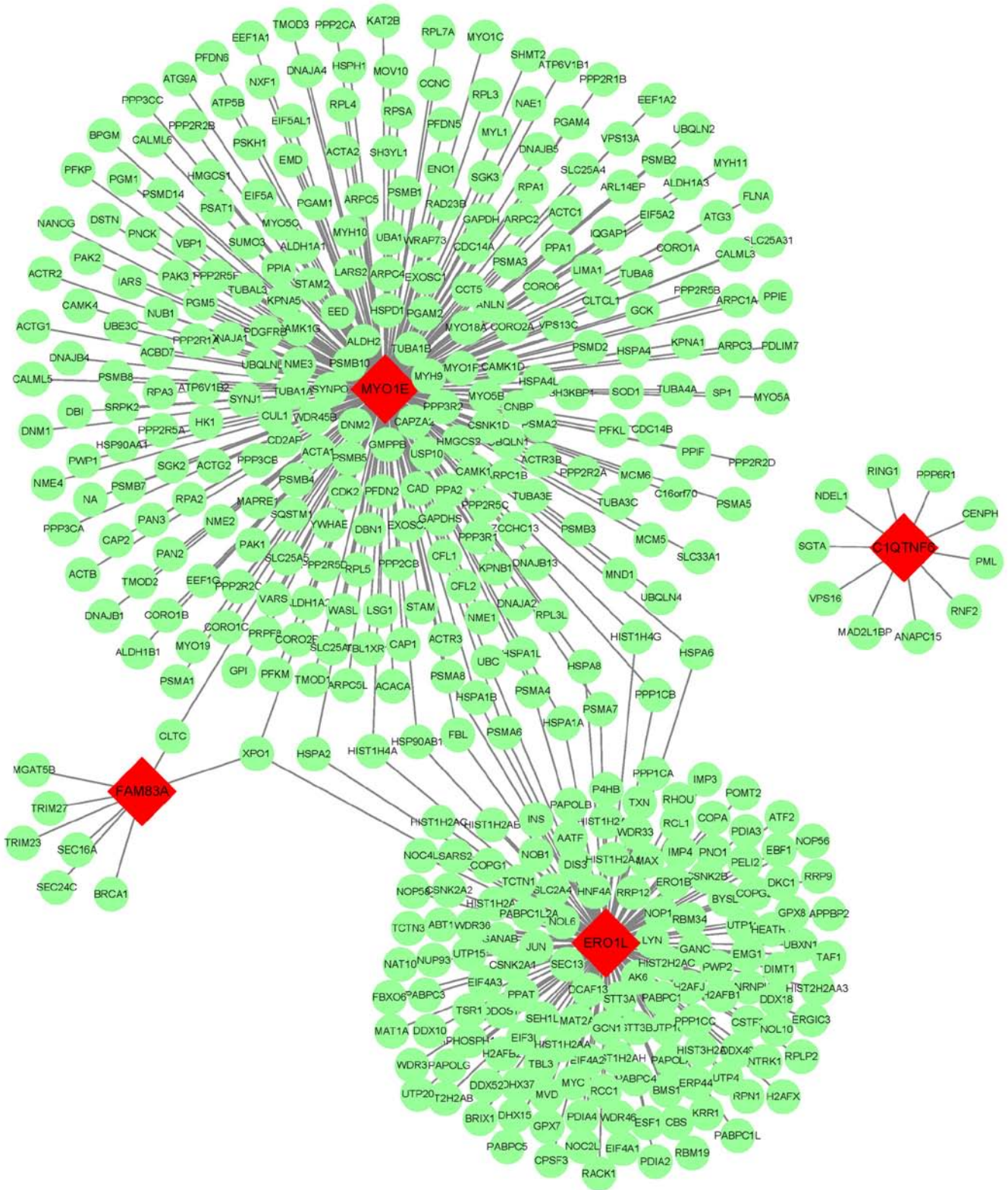
Figure 6. Protein-protein interaction network constructed for the four genes, *ERO1L*, *FAM83A*, *MYO1E* and *C1QTNF6*. Red diamonds and green circles represent the four genes and the nodes interacting with them, respectively. The lines represent interaction between the nodes. *ERO1L*, endoplasmic reticulum oxidoreductase 1α; *FAM83A*, family with sequence similarity 83, member A; *MYO1E*, myosin IE; *C1QTNF6*, C1q and tumor necrosis factor-related protein 6.

the correlation between *MYO1E*, *ERO1L*, *C1QTNF6* and *FAM83A* expression and prognosis of lung adenocarcinoma in clinical samples.

In conclusion, a total of 18,567 DEGs between LAD and adjacent tissues were screened. On the basis of the prognosis risk model, 363 DEGs between the high- and low-risk groups were identified. Additionally, the four genes (including *MYO1E*, *ERO1L*, *C1QTNF6* and *FAM83A*) utilized in the construction of the prognosis risk model may be used for predicting the prognosis of LAD. Therefore, in
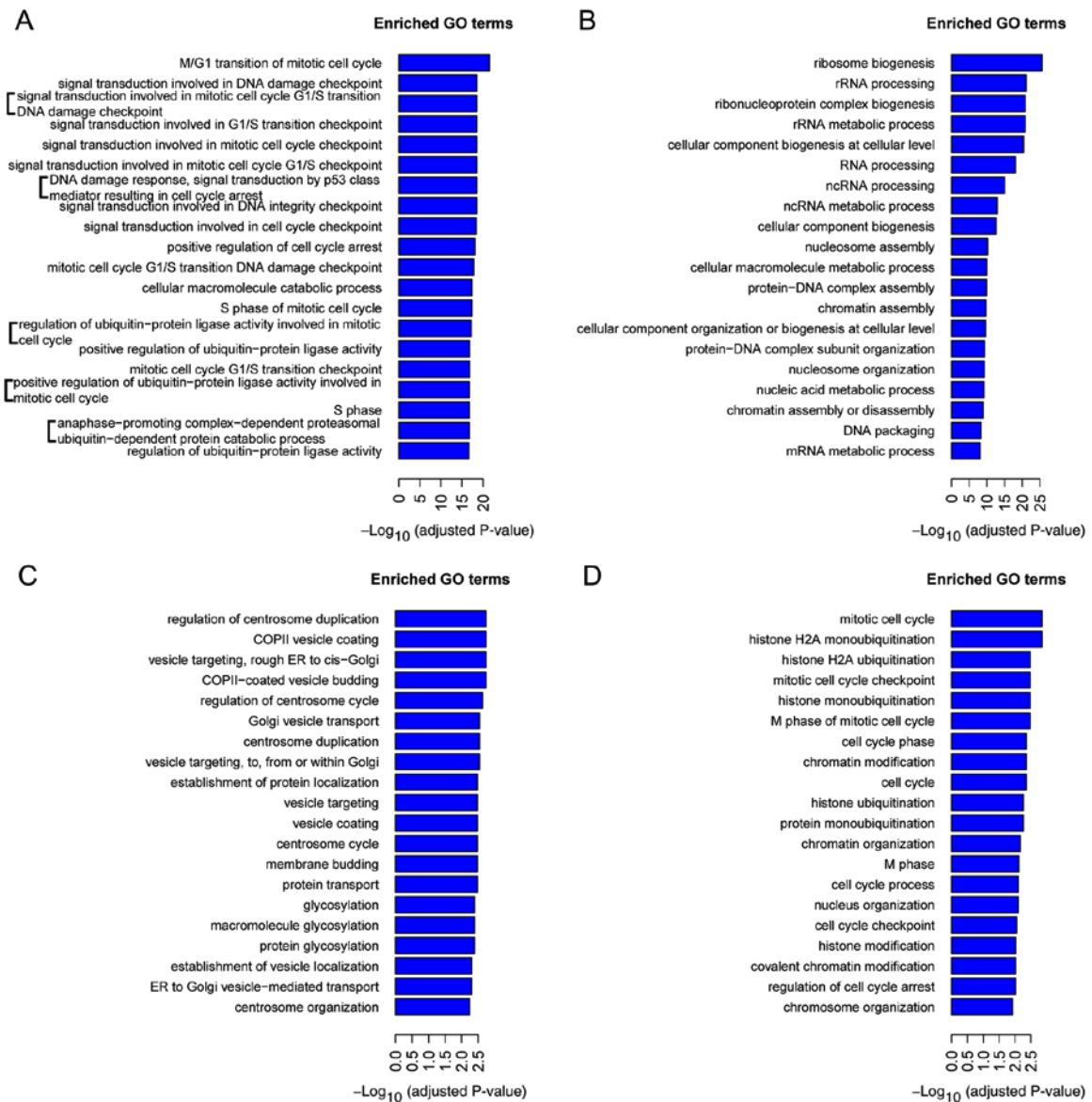
Figure 7. Functional GO terms enriched for the interacting genes of (A) *MYO1E*, (B) *ERO1L*, (C) *FAM83A* and (D) *C1QTNF6*. GO, Gene Ontology; *MYO1E*, myosin IE; *ERO1L*, endoplasmic reticulum oxidoreductase 1α; *FAM83A*, family with sequence similarity 83, member A; *C1QTNF6*, C1q and tumor necrosis factor-related protein 6; ncRNA, non-coding RNA; ER, endoplasmic reticulum; COPII, coatomer protein II.

future clinical practice, it may be possible to predict the prognosis of patients by calculating the risk score using data from surgical specimens. In the future, this risk prognosis model should be further validated using peripheral blood from patients with LAD.

Acknowledgements

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

WZ performed data analysis and wrote the manuscript. YS contributed significantly to the analysis of the data and manuscript revision. GF conceived and designed the study. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Travis WD: Pathology of lung cancer. Clin Chest Med 32: 669-692, 2011.
2. McGuire S: World cancer report 2014. Geneva, Switzerland: World health organization, international agency for research on cancer, WHO press, 2015. Adv Nut 7: 418-419, 2016.
3. Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL and Loscalzo J: Harrison's principles of internal medicine. 18th edition. McGraw Hill Professional, New York, NYC, USA pp 287, 2011.
4. Subramanian J and Govindan R: Lung cancer in never smokers: A review. J Clin Oncol 25: 561-570, 2007.
5. Marugame T, Sobue T, Nakayama T, Suzuki T, Kuniyoshi H, Sunagawa K, Genka K, Nishizawa N, Natsukawa S, Kuwahara O and Tsubura E: Filter cigarette smoking and lung cancer risk; a hospital-based case-Control study in Japan. Br J Cancer 90: 646-651, 2004.
6. Spiro SG, Tanner NT, Silvestri GA, Janes SM, Lim E, Vansteenkiste JF and Pirker R: Lung cancer: Progress in diagnosis, staging and therapy. Respirology 15: 44-50, 2010.
7. Fukui T and Mitsudomi T: Small peripheral lung adenocarcinoma: Clinicopathological features and surgical treatment. Surgery Today 40: 191-198, 2010.
8. Sholl LM, Barletta JA, Yeap BY, Chirieac LR and Hornick JL: Sox2 protein expression is an independent poor prognostic indicator in stage I lung adenocarcinoma. Am J Surg Pathol 34: 1193-1198, 2010.
9. Nakatsugawa M, Takahashi A, Hirohashi Y, Torigoe T, Inoda S, Murase M, Asanuma H, Tamura Y, Morita R, Michifuri Y, *et al*: SOX2 is overexpressed in stem-like cells of human lung adenocarcinoma and augments the tumorigenicity. Lab Invest 91: 1796-1804, 2011.
10. Zhang Y, Wang L, Li Y, Pan Y, Wang R, Hu H, Li H, Luo X, Ye T, Sun Y and Chen H: Protein expression of programmed death 1 ligand 1 and ligand 2 independently predict poor prognosis in surgically resected lung adenocarcinoma. Onco Targets Ther 7: 567-573, 2014.
11. Shimokawa H, Uramoto H, Onitsuka T, Chundong G, Hanagiri T, Oyama T and Yasumoto K: Overexpression of MACC1 mRNA in lung adenocarcinoma is associated with postoperative recurrence. J Thorac Cardiovasc Surg 141: 895-898, 2011.
12. Wang Z, Li Z, Wu C, Wang Y, Xia Y, Chen L, Zhu Q and Chen Y: MACC1 overexpression predicts a poor prognosis for non-small cell lung cancer. Med Oncol 31: 790, 2014.
13. Chen G, Kim SH, King AN, Zhao L, Simpson RU, Christensen PJ, Wang Z, Thomas DG, Giordano TJ, Lin L, *et al*: CYP24A1 is an independent prognostic marker of survival in patients with lung adenocarcinoma. Clin Cancer Res 17: 817-826, 2011.
14. Minamiya Y, Ono T, Saito H, Takahashi N, Ito M, Mitsui M, Motoyama S and Ogawa J: Expression of histone deacetylase 1 correlates with a poor prognosis in patients with adenocarcinoma of the lung. Lung Cancer 74: 300-304, 2011.
15. Love MI, Huber W and Anders S: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15: 550, 2014.
16. Therneau TM and Grambsch PM: Modeling survival data: Extending the cox model. Springer, New York, 2000.
17. Bland JM and Altman DG: The logrank test. BMJ 328: 1073, 2004.
18. Bera AK and Bilias Y: Rao's score, Neyman's C(α) and Silvey's LM tests: An essay on historical developments and some new results. J Stat Plan Inference 97: 9-44, 2001.
19. Harrell FE: Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis. Springer, New York, 2001.
20. Ishwaran H, Kogalur UB, Blackstone EH and Lauer MS: Random survival forests. Ann Appl Stat 2: 841-860, 2008.
21. Stel VS, Dekker FW, Tripepi G, Zoccali C and Jager KJ: Survival analysis I: The Kaplan-Meier method. Nephron Clin Pract 119: c83-c88, 2011.
22. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK: Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43: e47, 2015.
23. Wang J, Duncan D, Shi Z and Zhang B: WEB-based GEne SeT anaLysis Toolkit (WebGestalt): Update 2013. Nucleic Acids Res 41: W77-W83, 2013.
24. Li T, Wernersson R, Hansen RB, Horn H, Mercer J, Slodkowicz G, orkman CT, Rigina O, Rapacki K, Staerfeldt HH, *et al*: A scored human protein-protein interaction network to catalyze genomic interpretation. Nat Methods 14: 61-64, 2017.
25. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker T: Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498-2504, 2003.
26. Kutomi G, Tamura Y, Tanaka T, Kajiwara T, Kukita K, Ohmura T, Shima H, Takamaru T, Satomi F, Suzuki Y, *et al*: Human endoplasmic reticulum oxidoreductin 1-α is a novel predictor for poor prognosis of breast cancer. Cancer Sci 104: 1091-1096, 2013.
27. Tanaka T, Kutomi G, Kajiwara T, Kukita K, Kochin V, Kanaseki T, Tsukahara T, Hirohashi Y, Torigoe T, Okamoto Y, *et al*: Cancer-associated oxidoreductase ERO1-α promotes immune escape through up-regulation of PD-L1 in human breast cancer. Oncotarget 8: 24706-24718, 2017.
28. Tanaka T, Kajiwara T, Torigoe T, Okamoto Y, Sato N and Tamura Y: Cancer-associated oxidoreductase ERO1-α drives the production of tumor-promoting myeloid-derived suppressor cells via oxidative protein folding. J Immunol 194: 2004-2010, 2015.
29. Kajiwara T, Tanaka T, Kukita K, Kutomi G, Saito K, Okuya K, Takaya A, Kochin V, Kanaseki T, Tsukahara T, *et al*: Hypoxia augments MHC class I antigen presentation via facilitation of ERO1-α-mediated oxidative folding in murine tumor cells. Eur J Immunol 46: 2842-2851, 2016.
30. Tanaka T, Kutomi G, Kajiwara T, Kukita K, Kochin V, Kanaseki T, Tsukahara T, Hirohashi Y, Torigoe T, Okamoto Y, *et al*: Cancer-associated oxidoreductase ERO1-α drives the production of VEGF via oxidative protein folding and regulating the mRNA level. Br J Cancer 114: 1227-1234, 2016.
31. Snijders AM, Lee SY, Hang B, Hao W, Bissell MJ and Mao JH: FAM83 family oncogenes are broadly involved in human cancers: An integrative multi-omics approach. Mol Oncol 11: 167-179, 2017.
32. Bartel CA, Parameswaran N, Cipriano R and Jackson MW: FAM83 proteins: Fostering new interactions to drive oncogenic signaling and therapeutic resistance. Oncotarget 7: 52597-52612, 2016.
33. Cipriano R, Miskimen KL, Bryson BL, Foy CR, Bartel CA and Jackson MW: Conserved oncogenic behavior of the FAM83 family regulates MAPK signaling in human cancer. Mol Cancer Res 12: 1156-1165, 2014.
34. Lin B, Chen T, Zhang Q, Lu X, Zheng Z, Ding J, Liu J, Yang Z, Geng L, Wu L, *et al*: FAM83D associates with high tumor recurrence after liver transplantation involving expansion of CD44+ carcinoma stem cells. Oncotarget 7: 77495-77507, 2016.
35. Okabe N, Ezaki J, Yamaura T, Muto S, Osugi J, Tamura H, Imai J, Ito E, Yanagisawa Y, Honma R, *et al*: FAM83B is a novel biomarker for diagnosis and prognosis of lung squamous cell carcinoma. Int J Oncol 46: 999-1006, 2015.
36. Scholz T, Rump A, Uta P, Hinrichs M, Thiel C and Tsiavaliaris G: Myosin-1 function extends to microtubule-dependent processes. Biophys J 98 (Suppl): 561a, 2010.
37. Malumbres M and Barbacid M: Cell cycle kinases in cancer. Curr Opin Genet Dev 17: 60-65, 2007.
38. Asghar U, Witkiewicz AK, Turner NC and Knudsen ES: The history and future of targeting cyclin-dependent kinases in cancer therapy. Nat Rev Drug Discov 14: 130-146, 2015.
39. Peinado H, Lavotshkin S and Lyden D: The secreted factors responsible for pre-metastatic niche formation: Old sayings and new thoughts. Semin Cancer Biol 21: 139-146, 2011.