

Proliferation genes in lung development associated with the prognosis of lung adenocarcinoma but not squamous cell carcinoma

Ping Li¹  | Lei Zhang² | Xuexin Yu¹ | Run Tong³ | Xuebing Di¹ | Yousheng Mao⁴ | Yanning Gao¹ | Kaitai Zhang¹ | Lin Feng¹ | Shujun Cheng¹

¹State Key Laboratory of Molecular Oncology, Department of Etiology and Carcinogenesis, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

²Department of Endoscopy, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

³Department of Pulmonary and Critical Care Medicine, Center for Respiratory Diseases, China-Japan Friendship Hospital, Beijing, China

⁴Department of Thoracic Surgery, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Correspondence

Lin Feng and Shujun Cheng, State Key Laboratory of Molecular Oncology, Department of Etiology and Carcinogenesis, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.

Emails: fenglin@cicams.ac.cn; chengshj@cicams.ac.cn

Funding information

CAMS Innovation Fund for Medical Sciences, Grant/Award Number: 2016-I2M-3-005; National Key Laboratory Independent Innovation Project, Grant/Award Number: SKL-2017-04; National Key Research and Development Program of the Ministry of Science and Technology of China, Grant/Award Number: 2016YFC0905301; National Key Basic Research Program, Grant/Award Number: 2015CB553901

There are many similarities between embryonic development and tumorigenesis, and gene expression profiles show that certain correlations exist between the gene signature during development and the clinical phenotypes of different cancers. Our group previously reported the gene expression profiles of human lung development, and the expression of one group of proliferation-related genes (PTN1 genes) steadily decreased during lung development. Here, we examined the prognostic value of PTN1 genes in 5 independent lung adenocarcinoma (ADC) and 5 lung independent squamous cell carcinoma (SCC) microarray datasets and found that the expression levels of PTN1 genes were associated with survival in lung ADC but not lung SCC. All of the lung ADC datasets contained a set of highly correlated genes from PTN1 genes, but the lung SCC datasets had no similar set of genes. We identified 63 unique core genes from the PTN1 genes in the 5 lung ADC datasets: 17 of these core genes appeared in at least 4 of the lung ADC datasets, and the 17 corresponding proteins clearly interacted more strongly with each other in lung ADC than in lung SCC. Moreover, 16 of the 17 core genes play major roles in the G₂/M phase of the cell cycle. These data indicate that proliferation-related genes in lung development have a significant prognostic value for lung ADC; the synergistic effects of the 17 core genes play an important role in lung ADC prognosis. These genes may have significant clinical implications for the treatment and prognosis of lung ADC.

KEYWORDS

development, expression profile, lung cancer, proliferation gene, tumor prognosis

1 | INTRODUCTION

The intimate connection between tumorigenesis and developmental processes has been studied for many years. There are clear analogies between cancer and development. Morphologically, cancerous tissues appear as undifferentiated or poorly differentiated, with some tumor types even exhibiting embryonic tissue.¹ Behaviorally, the ability of tumor cells to grow, to infiltrate, and to suppress the immune system is similar to that of embryonic cells.^{2,3}

Several studies have explored the relationship between cancer and development on the gene expression level and have suggested that there are some associations between gene expression during development and the clinical phenotypes of different cancers. Liu et al⁴ studied the molecular associations of 4 human lung cancer subtypes and the developing mouse lung and found that there was a link between those associations and clinical outcome. A novel gene expression signature derived from mouse embryonic development could predict the metastatic competence of human breast tumor cells and forecast the survival rate of breast cancer patients.⁵ Not only the gene expression during mouse embryonic development but also the gene expression of human embryonic stem cells (ESCs) is related to the clinical phenotypes of cancer. Current evidence indicates that certain ESC genes that were identified in human blastocysts, such as *Polycomb*, *Nanog*, *Oct4*, and *Sox2*, are expressed in different human cancer types.⁶⁻⁸ Histologically poorly differentiated tumors show an increase in the expression of ESC genes, and this ESC-like signature is associated with worse overall survival in breast cancer patients.^{9,10} Moreover, one study explored the correlation between the ESC profile and clinical variables in lung cancer, and the analysis showed that an increased expression of the ESC gene set was associated with poorly differentiated and worse overall survival in lung adenocarcinoma (ADC); however, there was no correlation between the ESC gene signature and the differentiation status or outcome in lung squamous cell carcinoma (SCC).¹¹

Our group previously reported on the gene expression profile signatures of human lung tissues at 4 developmental stages (including whole embryos at postovulatory weeks [PW] 3-5, early embryonic lung at 6-8 PW, middle embryonic lung at 16-24 PW, and adult cancer-free peripheral lung tissue subject to surgery for benign lung diseases [MatureL]) and lung ADC tissue samples.¹² According to the gene expression trend observed during the 4 developmental stages, the genes were divided into 27 distinct groups (PTN1-27). The expression of genes in the PTN1 group (including 213 genes) decreased steadily during lung development; these 213 genes are called PTN1 genes in this study (Table S1). A 12-gene demonstrator from the PTN1 group was shown have prognostic value for lung ADC.

In this study, we explored the association between the expression of all the PTN1 genes and prognosis in lung ADC. Inspired by reports that the ESC gene signature was associated with worse overall survival in lung ADC but not lung SCC, the association between the expression levels of all the PTN1 genes and prognosis in lung SCC was also explored. We identified a group of core genes that

played an important role in the prognosis of lung ADC but not lung SCC. This finding provides a new clue for evaluating the prognosis of lung ADC and might be helpful in developing novel treatment approaches for lung ADC.

2 | MATERIALS AND METHODS

2.1 | Lung SCC and paracancerous sample collection

Lung SCC tissue and paired paracancerous samples (69 lung SCC and 60 paracancerous samples; 9 paracancerous samples were not collected) were collected from patients who underwent surgical resection at the Cancer Hospital, Chinese Academy of Medical Sciences (Beijing, China). Paracancerous samples refer to morphologically normal lung tissue located more than 5 cm from the tumor edge. Fresh tissues were snap frozen in liquid nitrogen and stored at -80°C before RNA preparation. The use of human samples in this study was reviewed and approved by the Ethics Committee of the Cancer Hospital, Chinese Academy of Medical Sciences, and all patients signed informed consent forms. Our group has already completed another study with the 69 lung SCC tissues. The gene expression profile data are available in the Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) (GSE67061),¹³ and the 69 lung SCC samples are named "SCC_Ours" in this study. Clinical information of the SCC_Ours dataset is provided in Table S2, and the median follow-up period was 2.42 years.

2.2 | RNA extraction and gene expression microarray

Total RNA from the paracancerous tissues was isolated with TRIzol reagent (Invitrogen, Carlsbad, CA, USA) and purified using an RNeasy kit (Qiagen, Germantown, MD, USA). RNA was then quantified using an ND-1000UV-VIS Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA), and RNA integrity was determined using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). All RNA samples used in this study showed optical density 260/280 ratios >1.9 and RNA integrity >6 . Paracancerous tissue samples were analyzed using an Agilent Whole Human Genome Microarray $4 \times 44\text{K}$ (G4112F). All sample labeling, hybridization, washing, and scanning steps were carried out following the manufacturer's specifications. The slides were scanned with an Agilent Microarray Scanner System (G2505B), and the fluorescence intensities were extracted and preprocessed using Agilent Feature Extraction Software version 9.1. Data extraction and annotation were carried out using GeneSpring GX version 12.6.1 (Agilent). The raw data were normalized by the median scale method using the R package "limma" (www.r-project.org). Screening probes represented the same gene, and only the probe showing the greatest mean intensity was retained. All raw microarray data and normalized data are publicly available on the GEO website (accession no. GSE101420).

2.3 | Analysis of public microarray datasets

Four independent sets of lung ADC microarray data (ADC_GSE13213,¹⁴ ADC_GSE8894,¹⁵ ADC_PNAS,¹⁶ and ADC_NCI¹⁷), 4 independent sets of lung SCC microarray data (SCC_GSE19188,¹⁸ SCC_GSE8894,¹⁵ SCC_GSE4573,¹⁹ and SCC_GSE14814²⁰) (Table S3) and their corresponding clinical information were collected from existing publications for validation. Our previous study on lung development included 4 developmental stages and 69 lung ADC samples, and their gene expression profile data are available from the GEO database (GSE43767)¹²; these 69 lung ADC samples were named “ADC_Ours” in this study. Clinical information of ADC_Ours is provided in Table S4. The median follow-up period was 3.34 years. The raw data were normalized using the same method used for the paracancerous group, except GSE8894. Because the raw data for GSE8894 were not provided, the GC robust multi-array average processed data were downloaded and analyzed directly.

2.4 | Statistical analysis

The gene set enrichment analysis was carried out using DAVID tools (<http://david.abcc.ncifcrf.gov/>). All statistical analyses in this study were undertaken using R software. In detail, differential gene expression analyses were carried out using an unpaired Student's *t*-test (R package “stats”). Prognosis was evaluated by the log-rank test (R package “survival”). Correlations between genes were evaluated using the correlation test (R package “stats”). Significant differences in the correlation coefficients between lung ADC and SCC were evaluated using a 2-way ANOVA (R package “stats”). Graphical visualization of the protein-protein interaction (PPI) networks were generated using Cytoscape software.²¹

3 | RESULTS

3.1 | Characteristics of PTN1 genes during human lung development and in lung ADC and SCC

A box diagram was drawn to show the average expression levels of PTN1 genes during the 4 lung developmental stages and in lung ADC, lung SCC, and paracancerous samples (Figure 1A). The expression of PTN1 genes decreased gradually during the 4 lung developmental stages, and the levels of expression in the paracancerous samples were similar to those in the MatureL samples but were significantly increased in the lung ADC and SCC samples. The changes in expression indicated that the PTN1 genes were progressively repressed during lung development and were generally reactivated during lung tumorigenesis. Compared with the MatureL samples as the control, PTN1 gene expression showed 80.8% and 88.3% increases in our lung ADC samples (ADC_Ours) and SCC samples (SCC_Ours), respectively (Figure 1B). Gene set enrichment analysis of PTN1 genes was carried out to identify biological themes. Figure 1C shows the top 30 Gene Ontology terms from the smallest to

largest *P*-value ($P < .05$); these terms show that the PTN1 genes were mainly related to cell proliferation, including cell division, mitotic nuclear division, DNA repair, DNA replication, and the G₂/M transition of the mitotic cell cycle.

3.2 | Expression of PTN1 genes associated with survival in lung ADC patients

To further investigate the relationship between PTN1 genes and clinical phenotypes of lung cancer, we examined the prognostic value of PTN1 genes in the ADC_Ours and SCC_Ours microarray datasets. As shown in Figure 2, the patients of each dataset were classified into 2 groups by *k*-means clustering according to the expression of the PTN1 genes. The low-risk group (good prognosis) is indicated in black, and the high-risk group (poor prognosis) is indicated in red. The log-rank test results indicate a significant difference in prognosis between the 2 groups in the ADC_Ours dataset ($n = 69$, $P = .039$), but no significant difference was noted in the SCC_Ours dataset ($n = 69$, $P = .48$). The survival analysis suggested that PTN1 genes were associated with the overall survival of patients with lung ADC but not patients with lung SCC, suggesting that PTN1 genes might be valuable for lung ADC. To validate the prognostic potential of PTN1 genes for lung ADC but not for lung SCC, we analyzed 4 independent sets of lung ADC microarray data and 4 independent sets of lung SCC microarray data from the GEO website by *k*-means clustering and the log-rank test. The results showed a significant difference in prognosis between the low-risk and high-risk groups in all 4 independent sets of lung ADC (ADC_GSE13213, $n = 117$, $P = .012$; ADC_GSE8894, $n = 62$, $P = .0023$; ADC_NCI, $n = 282$, $P = .00018$; and ADC_PNAS, $n = 190$, $P = .024$). In contrast, no significant difference in prognosis was noted between the low-risk and high-risk groups in 4 independent sets of lung SCC (SCC_GSE19188, $n = 27$, $P = .099$; SCC_GSE8894, $n = 76$, $P = .6$; SCC_GSE4573, $n = 130$, $P = .99$; and SCC_GSE14814, $n = 52$, $P = .93$) (Figure S1).

3.3 | Group of highly correlated genes exists in lung ADC

Because PTN1 genes share similar biological functions, mainly related to cell proliferation (Figure 1C), there might be a correlation among the PTN1 genes. To explore correlations among PTN1 genes, correlation tests were carried out on ADC_Ours, SCC_Ours and 4 independent sets each of lung ADC and SCC microarray data. Because the test statistics (*t*-value) of the correlation tests followed a Student's *t*-distribution, we used test statistics to observe the correlations among PTN1 genes. To better display the correlations, we determined the 90th percentile of the test statistics for each gene in relation to the other genes and then used the 90th percentile as an index to re-rank the genes, finally generating heat maps to illustrate the correlations. As shown in Figure 3, a group of highly related genes exists in the ADC_Ours dataset (the red region), but the

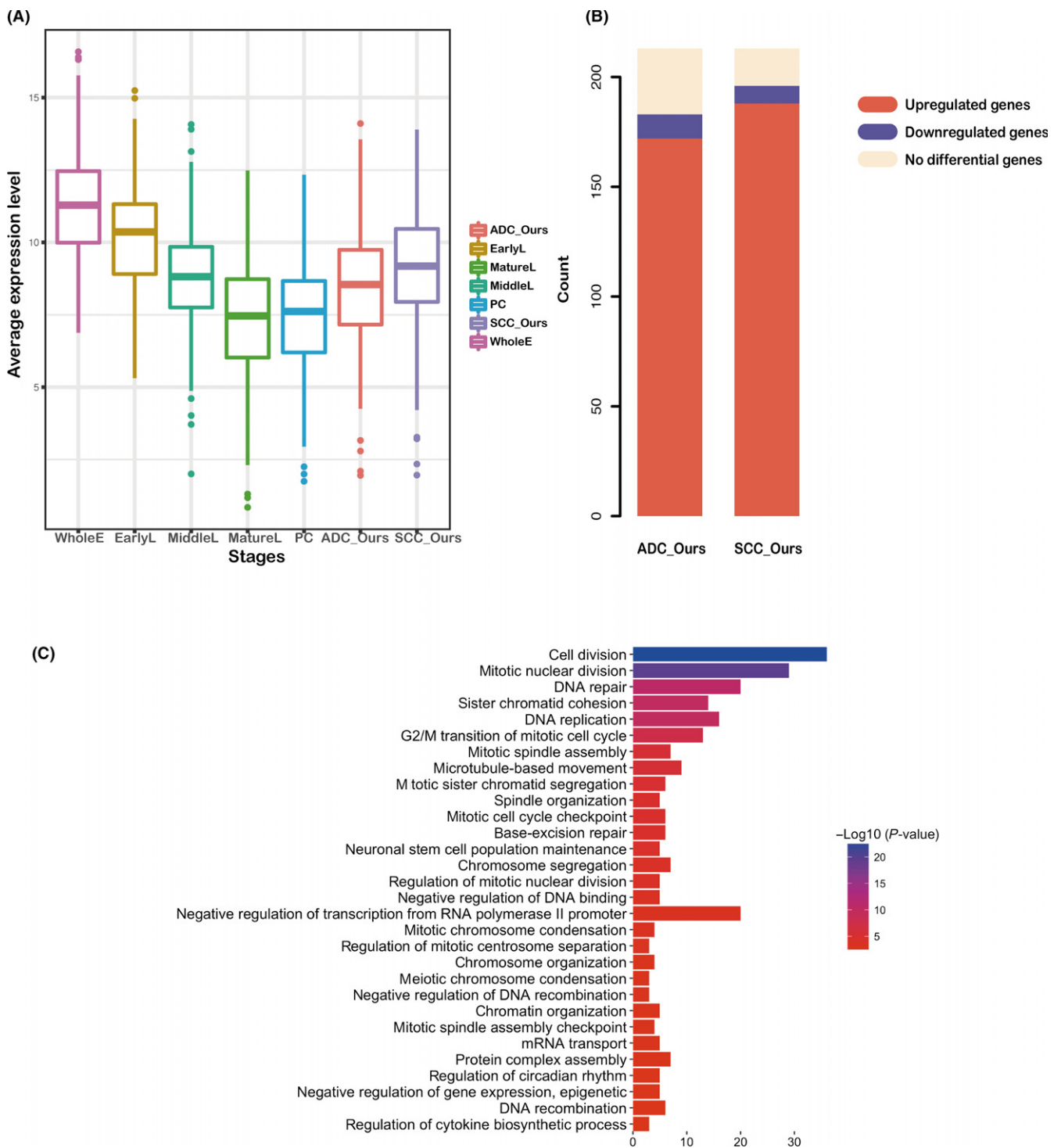


FIGURE 1 Expression features and functional annotations of proliferation-related genes involved in lung development (PTN1 genes). A, Expression levels of PTN1 genes are presented as box diagrams. The sample types are plotted on the x-axis, and the normalized average gene expression level in every sample type is plotted on the y-axis. EarlyL, early embryonic lung at 6-8 postovulatory weeks (PWs); MatureL, adult cancer-free peripheral lung tissue subject to surgery for benign lung diseases; MiddleL, middle embryonic lung at 16-4 PWs; PC, paracancerous; WholeE, whole embryos at 3-5 PWs. B, Histogram displays the distribution of PTN1 genes in 69 lung adenocarcinoma (ADC) samples (ADC_Ours) and 69 lung squamous cell carcinoma (SCC) samples (SCC_Ours), with MatureL as the control. The count of PTN1 genes is plotted on the y-axis, and the proportions of upregulated (red), downregulated (blue), and unregulated (yellow) genes are marked on the columns. C, The most significantly enriched functional categories of PTN1 genes are mainly related to cell proliferation ADC: lung adenocarcinoma; SCC: lung squamous cell carcinoma

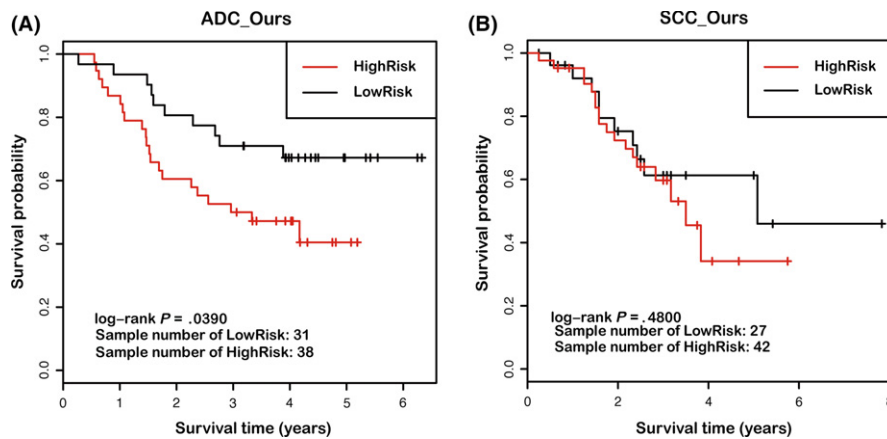


FIGURE 2 Prognostic value of proliferation-related genes (PTN1 genes) for lung adenocarcinoma (ADC) and lung squamous cell carcinoma (SCC) patients. (A, B) Kaplan-Meier survival curves and log-rank tests were used to estimate the survival of patients included in the ADC_Ours and SCC_Ours datasets, containing gene expression data from 69 lung ADC samples and 69 lung SCC samples, respectively. Patients from each dataset were classified into high-risk and low-risk groups by k-means clustering

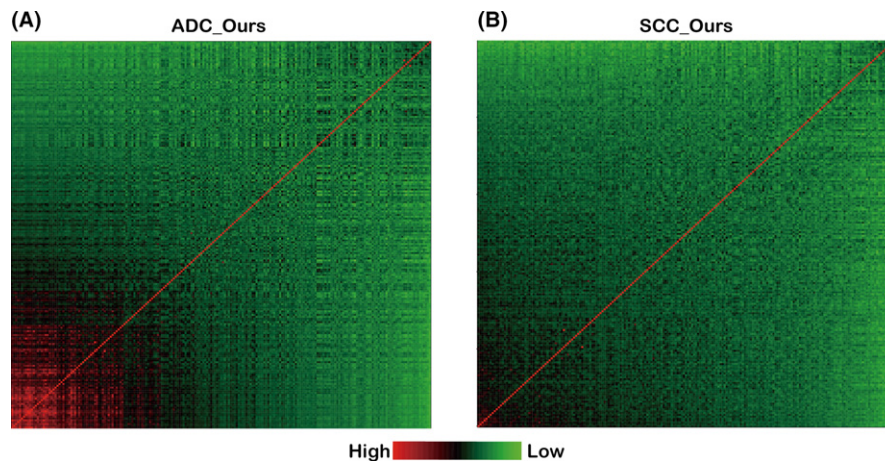


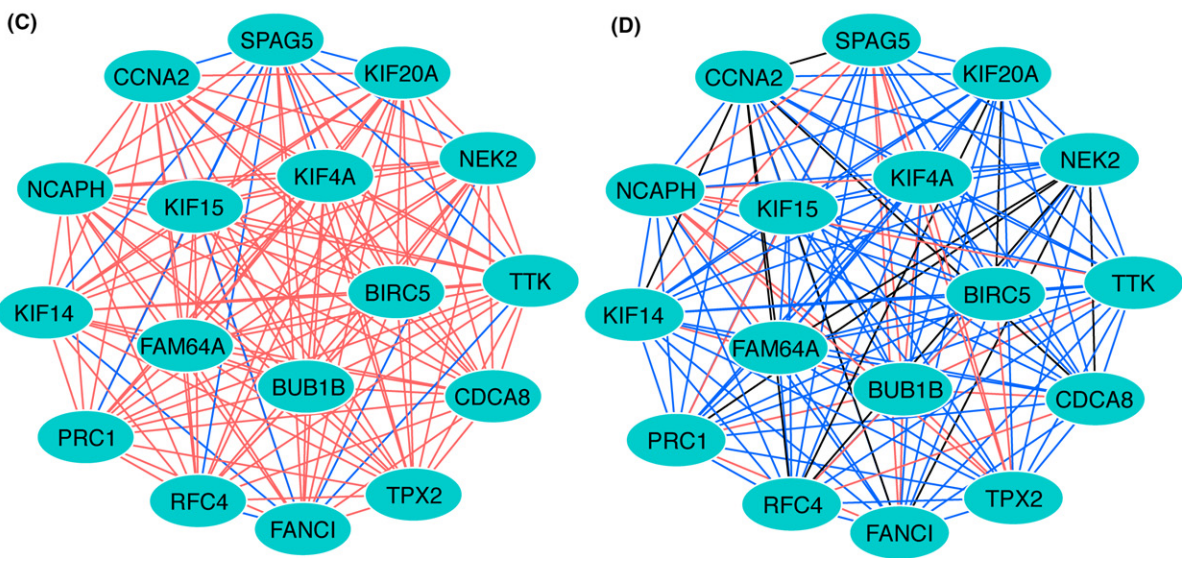
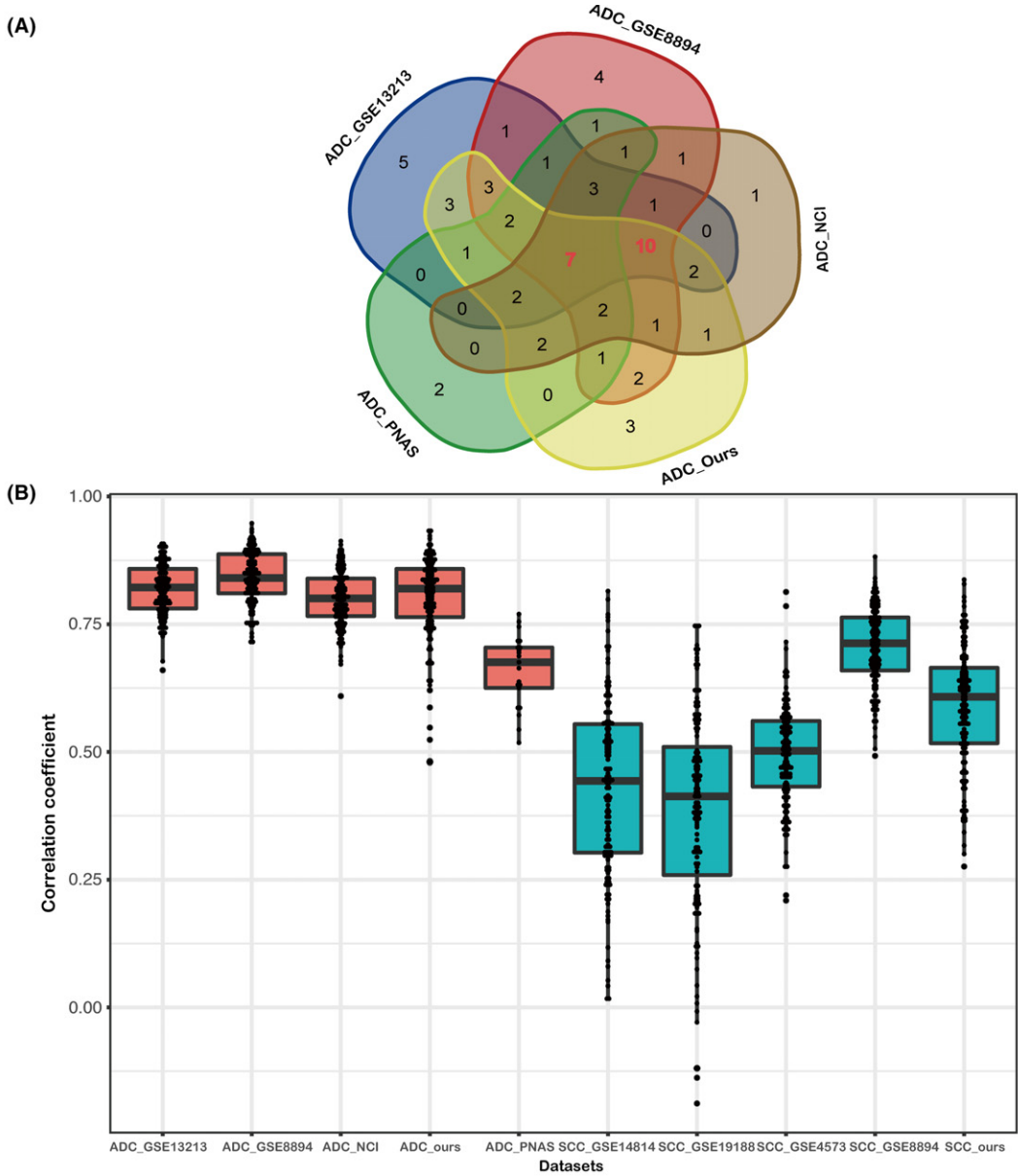
FIGURE 3 Correlated genes in lung adenocarcinoma (ADC_Ours) (A) and squamous cell carcinoma (SCC_Ours) (B) datasets, containing gene expression data from 69 lung ADC samples and 69 lung SCC samples, respectively. The red region indicates highly correlated genes, and the green region indicates poorly correlated genes. The red diagonal represents the relationship of a gene with itself

SCC_Ours dataset does not have these genes. These results indicate that a set of genes, which are highly correlated with each other, might play an important role in the development of lung ADC. This finding offers a possible explanation as to why the PTN1 genes were associated with the overall survival of patients with lung ADC, but not with the overall survival of patients with lung SCC. To validate whether lung ADC, but not lung SCC, is associated with a set of genes that originate during lung development and are highly correlated with each other, we analyzed the correlation among the PTN1 genes in the 4 independent sets of both lung ADC and SCC microarray data. All 4 independent lung ADC datasets appear as a group of highly related genes, but no lung SCC datasets show such genes (Figure S2).

3.4 | Set of core genes represents the common molecular characteristics in lung ADC

Genes in the red region (Figures 3A, S2) were present in all 5 lung ADC datasets and were highly correlated with each other, but all 5 lung SCC datasets lacked such genes, indicating that lung ADC patients may have a common molecular characteristic. To identify this common molecular characteristic, we extracted the top 10% (>90th percentile) of the genes that were re-ranked, as described earlier, and named them “core genes”; these genes represent the majority of genes in the red regions. The gene numbers were 42, 41, 34, and 25 in the ADC_Ours, ADC_GSE13213, ADC_GSE8894, ADC_NCI, and ADC_PNAS datasets, respectively. After merging the

FIGURE 4 Characteristics of core genes in lung adenocarcinoma (ADC_Ours) and squamous cell carcinoma (SCC_Ours) datasets, containing gene expression data from 69 lung ADC samples and 69 lung SCC samples, respectively. A, The Venn diagram displays the overlap of the core genes of 5 independent lung ADC datasets. Each dataset is colored, and the numbers indicate the number of overlapping genes. B, The difference in the correlation coefficients of 17 core genes between the lung ADC and SCC datasets. The correlation coefficients of the 17 core genes in the lung ADC datasets are higher than those in the lung SCC datasets; 2-way ANOVA was used to determine the significant difference ($P < .0001$). (C, D) Protein-protein interaction networks of associated proteins encoded by the 17 core genes in the lung ADC_Ours and SCC_Ours datasets. The nodes represent core genes, and the edges represent correlation coefficients, the colors of the edges represent the r value from the correlation test. Red edges represent an r value $\geq .7$ and < 1 ; blue edges represent an r value $\geq .4$ and $< .7$; black edges represent an r value > 0 and $< .4$. The 17 proteins clearly interact more strongly with each other in the lung ADC_Ours dataset (C) than in the lung SCC_Ours dataset (D)



core genes of the 5 lung ADC datasets, we obtained 63 unique core genes (Table S5). Of these core genes, 7 (*TPX2*, *SPAG5*, *CCNA2*, *KIF14*, *TTK*, *RFC4*, and *BUB1B*) belonged to all 5 datasets and 10 (*CDC48*, *FANCI*, *NCAPH*, *PRC1*, *NEK2*, *KIF20A*, *KIF15*, *KIF4A*, *FAM64A*, and *BIRC5*) belonged to 4 datasets (ADC_Ours, ADC_GSE13213, ADC_GSE8894, and ADC_NCI) (Figure 4A).

Because the ADC_PNAS dataset was obtained from an earlier type of microarray (Table S2), which contained fewer genes compared with the microarrays used for the other 4 lung ADC datasets, these 10 genes and the 7 genes common to all 5 datasets were further studied by us. Figure 4B shows that the correlation coefficients of these 17 core genes in the lung ADC datasets were higher than those in the lung SCC datasets, and a 2-way ANOVA revealed a significant difference ($P < .0001$). Using the lung ADC_Ours and lung SCC_Ours datasets as representatives, we constructed PPI networks using Cytoscape to show how these 17 genes were related to each other in terms of interactions between their associated proteins. In Figure 4C,D, the nodes represent core genes, and edges represent correlation coefficients, while the colors of the edges indicate the intensity of the correlation. The 2 PPI networks show that the 17 proteins clearly interact more strongly with each other in the lung ADC_Ours dataset than in the lung SCC_Ours dataset. Significantly, 16 genes (except *RFC4*, which is involved in DNA replication) are involved in microtubule movement and spindle assembly, functions that play a major role in the G₂/M phase of the cell cycle. Thus, these 17 proliferation-related core genes were maintained from the embryonic stage to lung ADC and continued to be highly correlated with each other. The set of proliferation-related core genes represents the common molecular characteristics in lung ADC.

4 | DISCUSSION

Genome-wide transcriptional profiling has identified gene expression signatures that are similar in organogenesis and tumors.^{4,22,23} Signaling, transcriptional, and metabolic pathways that are shared between fetal development and malignant tumors are reflected in the reactivation of embryonic genes in tumors.²⁴⁻²⁷ As the embryonic lung gradually matures, the expression of PTN1 genes, which are mainly related to cell proliferation, steadily decreases, and these proliferation-related genes are increased in lung cancer to adapt to the continuous division and proliferation of tumor cells. These trends in gene expression changes are consistent with previous theories that embryonic genes are reactivated in malignant tumors.

Several studies have determined that cell proliferation-related genes contribute to prognostic predictive ability,²⁸⁻³⁰ and developmental gene signatures can be used to predict outcomes.^{4,5,31} It has been reported that an ESC gene expression signature was associated with the prognosis of lung ADC patients but was not associated with SCC; the study authors postulated that this occurred because the lung ADC samples expressed a higher percentage of survival-related genes in the ESC gene set than did the SCC samples, suggesting that lung ADC has a molecular signature that is similar to the ESC

profile.¹¹ However, this explanation is inadequate. The PTN1 genes are proliferation-related genes expressed during lung development. In this study, we discovered that PTN1 genes could predict the prognosis in all 5 lung ADC datasets but had no prognostic value for any of the 5 lung SCC datasets. In addition, we provided evidence for this difference from the standpoint of correlated genes. In this study, we used a correlation test to reveal a set of correlated genes in all 5 lung ADC datasets, but none of the lung SCC datasets expressed such genes. This finding might explain why the PTN1 genes were associated with the overall survival of patients with lung ADC but not with patients with lung SCC. The synergistic effects of these correlated genes might play an important role in lung ADC, whereas the synergistic effects of these genes are destroyed in lung SCC, resulting in a loss of ability to predict prognosis. Furthermore, we identified 63 unique core genes from lung development in 5 lung ADC datasets, and 17 of these genes appeared in at least 4 lung ADC datasets. The correlation coefficients of these 17 core genes in lung ADC were significantly higher than those in lung SCC, and the interactions of the 17 corresponding proteins with each other were clearly stronger in lung ADC than those in lung SCC, which might further suggest a reason for the prognostic difference between lung ADC and SCC. The G₂/M phase is the preparative stage for mitosis and the cell division stage. Of these core genes, 16 play a major role in the G₂/M phase of the cell cycle, which might indicate that the synergistic effects of the 17 proliferation-related core genes play an important role during mitosis in lung ADC development.

Other mechanisms might play an important role in causing the different prognostic value of the PTN1 genes for lung ADC and SCC. Based on the histopathological appearance and gene expression signatures, it is speculated that lung ADC and SCC arise from distinct cells. It is generally accepted that lung ADC originates mainly from alveolar epithelial cells, such as the type II pneumocyte lineage, by progressing through a series of stages from atypical adenomatous hyperplasia to adenoma to ADC.^{32,33} However, it was reported that the histopathology and gene expression patterns of lung SCC resembled tracheal basal cell progenitors, and lung SCC expressed the basal cell markers p63 and CK5/14. Therefore, it is speculated that lung SCC is possibly derived from basal cells³³⁻³⁶ and has a stepwise progression from normal bronchial epithelium to squamous metaplasia, dysplasia, and SCC.³⁷ Lung ADC and SCC undergo distinct developmental processes. In particular, lung SCC progresses through a particular step in the form of squamous metaplasia. It may be that lung ADC has molecular characteristics similar to the PTN1 genes, and this similarity is maintained during the progression of lung ADC; however, lung SCC might develop without such molecular characteristics due to its different cellular origins and developmental stages. It is our hope that this study will provide a new clue for investigating the developmental and clinical differences between human lung ADC and SCC.

ACKNOWLEDGMENTS

This research was supported by the CAMS Innovation Fund for Medical Sciences (CIFMS 2016-I2M-3-005), National Key Laboratory

Independent Innovation Project (SKL-2017-04), National Key Research and Development Program of the Ministry of Science and Technology of China (Project No. 2016YFC0905301), and National Key Basic Research Program (Project No. 2015CB553901).

CONFLICT OF INTEREST

The authors have no conflict of interest.

ORCID

Ping Li  <http://orcid.org/0000-0003-3020-3792>

REFERENCES

- Naxerova K, Bult CJ, Peaston A, et al. Analysis of gene expression in a developmental context emphasizes distinct biological leitmotifs in human cancers. *Genome Biol.* 2008;9:R108.
- Ridolfi L, Petrini M, Fiammenghi L, Riccobon A, Ridolfi R. Human embryo immune escape mechanisms rediscovered by the tumor. *Immunobiology.* 2009;214:61-76.
- Ma Y, Zhang P, Wang F, Yang J, Yang Z, Qin H. The relationship between early embryo development and tumorigenesis. *J Cell Mol Med.* 2010;14:2697-2701.
- Liu H, Kho AT, Kohane IS, Sun Y. Predicting survival within the lung cancer histopathological hierarchy using a multi-scale genomic model of development. *PLoS Med.* 2006;3:e232.
- Soundararajan R, Paranjape AN, Barsan V, Chang JT, Mani SA. A novel embryonic plasticity gene signature that predicts metastatic competence and clinical outcome. *Sci Rep.* 2015;5:11766.
- O'Carroll D, Erhardt S, Pagani M, Barton SC, Surani MA, Jenuwein T. The polycomb-group gene *Ezh2* is required for early mouse development. *Mol Cell Biol.* 2001;21:4330-4336.
- Gidekel S, Pizov G, Bergman Y, Pikarsky E. Oct-3/4 is a dose-dependent oncogenic fate determinant. *Cancer Cell.* 2003;4:361-370.
- Santagata S, Ligon KL, Hornick JL. Embryonic stem cell transcription factor signatures in the diagnosis of primary and metastatic germ cell tumors. *Am J Surg Pathol.* 2007;31:836-845.
- Rodriguez-Pinilla SM, Sarrío D, Moreno-Bueno G, et al. Sox2: a possible driver of the basal-like phenotype in sporadic breast cancer. *Mod Pathol.* 2007;20:474-481.
- Ben-Porath I, Thomson MW, Carey VJ, et al. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat Genet.* 2008;40:499-507.
- Hassan KA, Chen G, Kalemkerian GP, Wicha MS, Beer DG. An embryonic stem cell-like signature identifies poorly differentiated lung adenocarcinoma but not squamous cell carcinoma. *Clin Cancer Res.* 2009;15:6386-6390.
- Feng L, Wang J, Cao B, et al. Gene expression profiling in human lung development: an abundant resource for lung adenocarcinoma prognosis. *PLoS ONE.* 2014;9:e105639.
- Tong R, Feng L, Zhang L, et al. Decreased interferon alpha/beta signature associated with human lung tumorigenesis. *J Interferon Cytokine Res.* 2015;35:963-968.
- Tomida S, Takeuchi T, Shimada Y, et al. Relapse-related molecular signature in lung adenocarcinomas identifies patients with dismal prognosis. *J Clin Oncol.* 2009;27:2793-2799.
- Lee ES, Son DS, Kim SH, et al. Prediction of recurrence-free survival in postoperative non-small cell lung cancer patients by using an integrated model of clinical information and gene expression. *Clin Cancer Res.* 2008;14:7397-7404.
- Bhattacharjee A, Richards WG, Staunton J, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A.* 2001;98:13790-13795.
- Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma, Shedden K, Taylor JM, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med.* 2008;14:822-827.
- Hou J, Aerts J, den Hamer B, et al. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. *PLoS ONE.* 2010;5:e10312.
- Raponi M, Zhang Y, Yu J, et al. Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. *Cancer Res.* 2006;66:7466-7472.
- Zhu CQ, Ding K, Strumpf D, et al. Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer. *J Clin Oncol.* 2010;28:4417-4424.
- Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13:2498-2504.
- Couluarn C, Derambure C, Lefebvre G, et al. Global gene repression in hepatocellular carcinoma and fetal liver, and suppression of *dudulin-2* mRNA as a possible marker for the cirrhosis-to-tumor transition. *J Hepatol.* 2005;42:860-869.
- Dekel B, Metsuyanim S, Schmidt-Ott KM, et al. Multiple imprinted and stemness genes provide a link between normal and tumor progenitor cells of the developing human kidney. *Cancer Res.* 2006;66:6040-6049.
- Hu M, Shivdasani RA. Overlapping gene expression in fetal mouse intestine development and human colorectal cancer. *Cancer Res.* 2005;65:8715-8722.
- Lum L, Beachy PA. The Hedgehog response network: sensors, switches, and routers. *Science.* 2004;304:1755-1759.
- Sancho E, Batlle E, Clevers H. Signaling pathways in intestinal development and cancer. *Annu Rev Cell Dev Biol.* 2004;20:695-723.
- Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nat Med.* 2004;10:789-799.
- Dai H, van't Veer L, Lamb J, et al. A cell proliferation signature is a marker of extremely poor outcome in a subpopulation of breast cancer patients. *Cancer Res.* 2005;65:4059-4066.
- Starmans MH, Krishnapuram B, Steck H, et al. Robust prognostic value of a knowledge-based proliferation signature across large patient microarray studies spanning different cancer types. *Br J Cancer.* 2008;99:1884-1890.
- Mosley JD, Keri RA. Cell cycle correlated genes dictate the prognostic power of breast cancer gene lists. *BMC Med Genomics.* 2008;1:11.
- Liu R, Wang X, Chen GY, et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med.* 2007;356:217-226.
- Jackson EL, Willis N, Mercer K, et al. Analysis of lung tumor initiation and progression using conditional expression of oncogenic *K-ras*. *Genes Dev.* 2001;15:3243-3248.
- Han X, Li F, Fang Z, et al. Transdifferentiation of lung adenocarcinoma in mice with *Lkb1* deficiency to squamous cell carcinoma. *Nat Commun.* 2014;5:3261.
- Sutherland KD, Berns A. Cell of origin of lung cancer. *Mol Oncol.* 2010;4:397-403.
- Borthwick DW, Shahbazian M, Krantz QT, Dorin JR, Randell SH. Evidence for stem-cell niches in the tracheal epithelium. *Am J Respir Cell Mol Biol.* 2001;24:662-670.
- Terry J, Leung S, Laskin J, Leslie KO, Gown AM, Ionescu DN. Optimal immunohistochemical markers for distinguishing lung adenocarcinomas from squamous cell carcinomas in small tumor samples. *Am J Surg Pathol.* 2010;34:1805-1811.

37. Nishisaka T, Takeshima Y, Inai K. Evaluation of p53 gene mutation and loss of heterozygosity of 3p, 9p and 17p in precancerous lesions of 29 lung cancer patients. *Hiroshima J Med Sci.* 2000;49:109-116.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Li P, Zhang L, Yu X, et al. Proliferation genes in lung development associated with the prognosis of lung adenocarcinoma but not squamous cell carcinoma. *Cancer Sci.* 2018;109:308–316. <https://doi.org/10.1111/cas.13456>