Prognostic model based on circular RNA circPDK1 for resected lung squamous cell carcinoma

Xiao Sun[1#], Maolong Wang[1#], Rongjian Xu[1#], Dongyang Zhang[1#], Ao Liu[1], Yuanyong Wang[1], Tong Lu[1], Yanlu Xin[2], Yandong Zhao[1], Yunpeng Xuan[1], Tong Qiu[1], Hao Wang[3], Shicheng Li[1], Yang Wo[1], Dahai Liu[4], Jinpeng Zhao[5], Bo Fu[6], Yaliang Lan[1], Yudong Han[1], Wenjie Jiao[1]

[1]Department of Thoracic Surgery, [2]Department of Endocrinology and Metabolism, [3]Administrative Office, [4]Medical Examination Center, Affiliated Hospital of Qingdao University, Qingdao 266003, China; [5]Department of Thoracic Surgery, Laiyang Central Hospital, Yantai 264000, China; [6]Otorhinolaryngology Head and Neck Surgery, Affiliated Qilu Hospital of Shandong University, Jinan 250000, China

Contributions: (I) Conception and design: X Sun, M Wang; (II) Administrative support: W Jiao; (III) Provision of study materials or patients: Y Zhao, Y Xuan, T Qiu, J Zhao, B Fu, Y Han; (IV) Collection and assembly of data: T Lu, Y Xin, H Wang, S Li, Y Wo; (V) Data analysis and interpretation: X Sun, A Liu, Y Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Wenjie Jiao. Department of Thoracic Surgery, Affiliated Hospital of Qingdao University, NO. 16 Jiangsu Road, Shinan District, Qingdao 266003, China. Email: jiaowj@qduhospital.cn.

Background: Circular RNA has been revealed as a potential biomarker in multiple malignancies. However, few studies have focused on its potential to be prognostic markers in lung squamous cell carcinoma (LSCC). In this work, we aimed to build a prognostic model of resected LSCC based on circular RNA pyruvate dehydrogenase kinase 1 (circPDK1) and other clinicopathological factors.

Methods: circPDK1 was identified via next-generation sequencing. Three hundred two cases of LSCC tissue and their adjacent normal lung tissues were obtained from multiple medical centers and divided into study cohort (n=232) and validation cohort (n=70). The expression of circPDK1 was detected for analyzing its potential prognostic value for recurrence-free survival (RFS) and overall survival (OS) in LSCC. Finally, combined with circPDK1, T staging, lymph nodes (LN) metastasis status, age, and serum squamous cell Carcinoma Antigen (SCCAg), we built a prognostic model by nomograms method and confirmed it in the validation cohort.

Results: CircPDK1 was identified to be overexpressed (P<0.01) in LSCC. Through analysis in study cohort, circPDK1^{low} patients (less than the mean expression, n=124) showed more lymph nodes metastasis (P=0.025), more vascular invasion (VI) (P=0.047), more visceral pleural invasion (VPI) (P=0.015) and poorer prognosis (P=0.003) than circPDK1^{high} ones (n=108). Univariate and multivariate analysis showed that circPDK1, T staging, LN status, age, and SCCAg were significant prognostic factors for RFS and OS. The prognostic model based on these factors showed the concordance index (C-index) of 0.8214 and 0.8359 for predicting 5-year RFS and OS, respectively. Finally, the calibration curves were performed in the study cohort and a validation cohort to evaluate the model's efficiency.

Conclusions: circPDK1 was identified as a potential biomarker of resected LSCC. The prognostic model including circPDK1, T staging, LN status, age, and SCCAg could effectively predict prognosis of resected LSCC.

Keywords: Circular RNA; lung squamous cell carcinoma (LSCC); nomograms; prognostic model

Introduction

Lung squamous cell carcinoma (LSCC) is the second most common pulmonary malignancy worldwide, accounting for 20–30% of non-small cell lung cancer (NSCLC) cases (1,2). Although the poor prognosis of LSCC has raised the concern of researchers for decades, it was still unable to be predicted accurately by now (3).

Serum tumor markers have long been investigated as prognostic factors because they are simple parameters to measure. The used lung cancer markers for diagnosis and prognosis include carcinoembryonic antigen (CEA), Cytokeratin 19 fragment (CK19), Carbohydrate antigen 125 (CA125), Serum gastrin-releasing peptide precursor (ProGRP), SCCAg and Neuron-specific enolase (NSE) (4-9). However, the sensitivity and specificity of theses markers were always too low to predict lung malignancies. Besides, the serum level of these markers fluctuates in different duration of disease. Therefore, more stable, and accurate biomarkers were needed for LSCC.

Circular RNAs are a group of circular-stated transcripts associated with the progression of multiple human malignancies and diseases (10). Compared with linear mRNAs, these circular products are highly stable *in vivo* and more suitable as biomarkers (11). So far, circular RNA has been revealed as a potential biomarker in multiple malignancies, including gastric, colorectal, pancreatic and breast cancer (12-16). However, few studies have revealed the relationship between LSCC and circular RNAs.

In this work, we identified a circular RNA, circPDK1, which is overexpressed in LSCC and associated with good prognosis of resected LSCC patients. We aimed to build a prognostic model including circPDK1 and other clinicopathological factors, thereby revealing its potential as a prognostic biomarker for LSCC.

Methods

Next-generation sequencing

Ten cases of LSCC tumor tissues and their adjacent normal tissues were collected intraoperatively for transcriptome-wide associated sequencing. The sequencing was performed by BioMarker Technologies (Beijing, China).

Study population and follow-up

Three hundred two cases of LSCC samples and their adjacent normal tissues were collected from four Chinese

medical centers during October 2013 and December 2014. All the individuals were primary lung squamous cell cancer patients underwent R0 resection and systematic lymph node dissection. The patients who had received preoperative radiotherapy, chemotherapy, or other adjuvant therapies, or those with recurrence and distant metastasis were excluded. Normal lung tissues were collected from a distance more than 5 cm from the tumor margin and used as negative controls. All collected tissue samples were immediately stored at –80 ℃ until use. The clinicopathological information of the patients was retrospectively collected from the electronic medical records. The serum levels of tumor markers (CEA, SCCAg, NSE, CK19, CA125, and ProGRP) were routinely measured preoperatively. The cutoff value of these tumor markers were set according to other literatures: CEA, 5.5 ng/mL; SCCAg, 1.5 ng/mL; NSE, 16.3 U/mL; CK19, 3.3 ng/mL; CA125, 35 U/mL; ProGRP, 50 pg/mL (8,17,18). These cutoff values were also following the values showed in Survival status was collected via outpatient clinic revisit records or telephone call. The mean follow-up period of the patients was 62.2 months (range, 56–70 months). The patient loss to follow-up was excluded. For all patients, the postsurgical pathological staging was determined according to the eighth edition TNM classification. Individuals were divided into study cohort (n=232) and validation cohort (n=70) randomly. The baseline information and clinicopathologic characteristics of the two cohorts were seen in *Table 1*. Informed consents have been obtained from all the patients or their families.

RNA extraction and quantitative real-time polymerase chain reaction (QRT-PCR)

Total RNA was extracted from tissues with TRIZOL reagent (Solarbio, China). The SuperScript® III First-Strand Synthesis System for RT-PCR (Thermo Fisher Scientific, USA) was used to reverse-transcribe RNA into cDNA. QRT-PCR was performed on an Applied Biosystems 7500 Real-Time PCR system (Applied Biosystems, USA) with SYBR® Green PCR Master Mix (Thermo Fisher Scientific, USA). The primers were as follows: circPDK1 forward: 5'-TCGAAAACACATTGGAAGCA and reverse: 3'-TGCCGCAGAAACATAAATGA. Relative RNA levels were calculated with $2^{-\Delta CT}$ [where $\Delta Ct = Ct(gene) – Ct(\beta\text{-actin})$]. mRNA PDK1 forward: 5'-CAGCAGTCCTTGGCATTTGT and reverse: 3'-ATACGTCTCAGGCTCTGCTC. Additionally, the fold change of RNA expression in tumor tissues *vs.* adjacent

Table 1 Clinicopathologic characteristics according to the circPDK1 expression in the study and validation cohorts

Characteristics	Study cohort				Validation cohort			
	Total (n=232), n (%)	CircPDK1low (n=124), n (%)	CircPDK1high (n=108), n (%)	P value	Total (n=70), n (%)	CircPDK1low (n=36), n (%)	CircPDK1high (n=34), n (%)	P value
Gender				0.610				0.411
Male	223 (96.1)	118 (95.2)	105 (97.2)		67 (95.7)	36 (100.0)	31 (91.2)	
Female	9 (3.9)	6 (4.8)	3 (2.8)		3 (4.3)	0 (0.0)	3 (8.8)	
Age				0.561				1.000
≤65	116 (50.0)	60 (48.4)	56 (51.9)		35 (50.0)	18 (50.0)	17 (50.0)	
>65	116 (50.0)	64 (51.6)	52 (48.1)		35 (50.0)	18 (50.0)	17 (50.0)	
Smoking				0.106				0.112
Former/current	204 (87.9)	118 (95.2)	86 (79.6)		60 (85.7)	34 (94.4)	26 (76.5)	
Never	28 (12.1)	6 (4.8)	22 (20.4)		10 (14.3)	2 (5.6)	8 (23.5)	
Surgery				0.456				0.608
Lobectomy	141 (60.8)	72 (58.1)	69 (63.9)		37 (52.9)	20 (55.6)	17 (50.0)	
Bilobectomy	16 (6.9)	7 (5.6)	9 (8.3)		10 (14.3)	4 (11.1)	6 (17.6)	
Pneumonectomy	24 (10.3)	14 (11.3)	10 (9.3)		12 (17.1)	6 (16.7)	6 (17.6)	
Bronchial sleeve resection	49 (21.1)	29 (23.4)	20 (18.5)		11 (15.7)	6 (16.7)	5 (14.7)	
Adjuvant therapies				0.253				0.203
Chemotherapy or others	56 (24.1)	32 (25.8)	24 (22.2)		21 (30.0)	14 (38.9)	7 (20.6)	
None	120 (51.7)	55 (44.4)	65 (60.2)		33 (47.1)	14 (38.9)	19 (55.9)	
Unknown	56 (24.1)	37 (29.8)	29 (26.9)		16 (22.9)	8 (22.2)	8 (23.5)	
Tumor size (cm)				0.477				0.238
0–3	80 (34.5)	37 (29.8)	43 (39.8)		18 (25.7)	8 (22.2)	10 (29.4)	
3–5	93 (40.1)	48 (38.7)	45 (41.7)		26 (37.1)	16 44.4)	10 (29.4)	
5–7	41 (17.7)	20 (16.1)	21 (19.4)		17 (24.3)	7 (19.4)	10 (29.4)	
>7	18 (7.8)	11 (8.9)	7 (6.5)		9 (12.9)	5 (13.9)	4 (11.8)	
Mean ± SD	5.73±2.22	5.83±2.51	5.52±2.31	0.464	5.62±2.31	5.67±2.38	5.59±2.46	0. 643
Lymph nodes metastasis				0.011				0.043
Positive	79 (36.2)	62 (50.0)	17 (15.7)		27 (38.6)	18 (50.0)	9 (26.5)	
Negative	153 (63.8)	62 (50.0)	91 (84.3)		43 (61.4)	18 (50.0)	25 (73.5)	
Pathologic T stage				0.275				0.622
1	58 (25.0)	27 (21.8)	31 (28.7)		14 (20.0)	7 (19.4)	7 (20.6)	
2	100 (43.1)	57 (46)	43 (39.8)		28 (40.0)	16 (44.4)	12 (35.3)	
3	44 (19.0)	21 (16.9)	23 (21.3)		17 (24.3)	8 (22.2)	9 (26.5)	
4	34 (14.7)	19 (15.3)	11 (10.2)		11 (15.7)	5 (13.9)	6 (17.6)	
Pathologic N stage				0.025				0.022
0	153 (63.8)	62 (50.0)	91 (84.3)		43 (61.4)	18 (50.0)	25 (73.5)	
1	33 (14.2)	25 (20.2)	8 (7.4)		9 (12.9)	5 (13.9)	4 (11.8)	
2	46 (19.8)	37 (29.8)	9 (8.3)		18 (25.7)	13 (36.1)	5 (14.7)	

Table 1 (*continued*)

910

Sun et al. Prognostic model for lung cancer

Table 1 (*continued*)

Characteristics	Study cohort				Validation cohort			
	Total (n=232), n (%)	CircPDK1low (n=124), n (%)	CircPDK1high (n=108), n (%)	P value	Total (n=70), n (%)	CircPDK1low (n=36), n (%)	CircPDK1high (n=34), n (%)	P value
TNM staging				0.124				0.168
I	70 (30.2)	36 (29)	34 (31.5)		19 (27.1)	8 (22.2)	11 (32.4)	
II	83 (35.8)	28 (22.6)	55 (50.9)		24 (34.3)	8 (22.2)	16 (47.1)	
III	79 (34.1)	60 (48.4)	19 (17.6)		27 (38.6)	20 (55.6)	7 (20.6)	
Grading				0.087				0.355
G1-2	88 (37.9)	37 (29.8)	51 (47.2)		26 (37.1)	12 (33.3)	14 (41.2)	
G3-4	102 (44.0)	64 (51.6)	38 (35.2)		33 (47.1)	19 (52.8)	14 (41.2)	
Unknown	42 (18.1)	23 (18.5)	19 (17.6)		11 (15.7)	5 (13.9)	6 (17.6)	
Vascular invasion				0.047				0.202
Present	172 (74.1)	100 (80.6)	72 (66.7)		41 (58.6)	23 (63.9)	18 (52.9)	
Absent	60 (25.9)	24 (19.4)	36 (33.3)		29 (41.4)	13 (36.1)	16 (47.1)	
Visceral pleural invasion				0.015				0.011
Present	134 (57.8)	95 (76.6)	39 (36.1)		41 (58.6)	30 (83.3)	11 (32.4)	
Absent	98 (42.2)	29 (23.4)	69 (63.9)		29 (41.4)	6 (16.7)	23 (67.6)	
Serum CEA (ng/mL)				0.124				0.242
≤5.5	147 (63.4)	68 (54.8)	69 (63.9)		46 (65.7)	21 (58.3)	25 (73.5)	
>5.5	85 (36.6)	56 (45.2)	29 (26.9)		24 (34.3)	15 (41.7)	9 (26.5)	
Serum SCCAg (ng/mL)				0.264				0.377
≤1.5	74 (31.9)	46 (37.1)	38 (35.2)		20 (28.6)	12 (33.3)	8 (23.5)	
>1.5	158 (68.1)	78 (62.9)	80 (74.1)		50 (71.4)	24 (66.7)	26 (76.5)	
Serum CK19 (ng/mL)				0.033				0.042
≤3.3	111 (47.8)	52 (41.9)	69 (63.9)		35 (50)	13 (36.1)	22 (64.7)	
>3.3	121 (52.2)	72 (58.1)	39 (36.1)		35 (50)	23 (63.9)	12 (35.3)	
Serum CA125 (U/mL)				0.156				0.231
≤35	207 (89.2)	108 (87.1)	99 (91.7)		62 (88.6)	29 (80.6)	33 (97.1)	
>35	25 (10.8)	16 (12.9)	9 (8.3)		8 (11.4)	7 (19.4)	1 (2.9)	
Serum NSE (U/mL)				0.318				0.419
≤16.3	80 (34.5)	46 (37.1)	34 (31.5)		19 (27.1)	9 (25)	10 (29.4)	
>16.3	152 (61.2)	78 (62.9)	74 (68.5)		51 (72.9)	27 (75)	24 (70.6)	
Serum ProGRP (pg/mL)				0.261				0.748
≤50	211 (90.9)	111 (89.5)	100 (92.6)		65 (92.9)	33 (91.7)	32 (94.1)	
>50	21 (9.1)	13 (10.5)	8 (7.4)		5 (7.1)	3 (8.3)	2 (5.9)	

Adjuvant therapies include postoperative chemotherapy (neoadjuvant and adjuvant ones), radiotherapy or target therapy.

normal tissues was calculated with the $2^{-\Delta\Delta CT}$ method. GAPDH was used as an internal control.

Prognostic analysis

Overall survival (OS) was defined as the period from after operation to patient death or the last follow-up. Relapse-free survival (RFS) was defined as the period from surgery to tumor recurrence or patient death from a non-cancerous cause. The univariate and multivariate survival analyses were performed by Cox proportional hazards regression models to analyze the potential prognostic factors. The 5-year survival was analyzed with Kaplan-Meier plotter (K-M plot) survival curve. The prognostic model was built based on the results of multivariate analysis. Nomograms (19) was built to calculate the 5-year RFS and OS. C-index and calibration curves were used to achieve internal validation. External validation was performed in the validation cohort.

Statistical analysis

All data were analyzed with the statistical package SPSS 22.0 (SPSS Inc., Chicago, IL, USA) and R language (version 3.5.1, the "survival" and "rms" packages). Data are presented as mean ± SD. Graphs were made with GraphPad Prism 6 (GraphPad Software, CA, USA). Student's t-test, chi-square test, multiple logistic regression, and Pearson correlation analysis were used. P values <0.05 were considered statistically significant.

Results

CircPDK1 was identified as a prognostic biomarker of LSCC

Through next-generation sequencing, we identified a unique circular RNA, circPDK1, which was highly expressed in LSCC tissues but was scarcely detected in normal tissues (*Figure 1A,B*).

Next, we detected the relative expression of circPDK1 in each LSCC specimen. The results of qRT-PCR proved that circPDK1 was significantly higher in LSCC tissues than that in normal tissues (5.05 folds change, P<0.01) (*Figure 1C*). However, the expression of circPDK1 fluctuated among different patients.

To explore the correlation between circPDK1 and 5-year RFS/OS of LSCC patients, we divided the cases into circPDK1^{high} and circPDK1^{low} groups according to its

mean expression ($2^{-\Delta\Delta CT}$ =3.7), then, K-M plot analysis was performed in the study cohort. The OS of circPDK1^{low} and circPDK1^{high} was 29.0% and 48.2%, respectively (P=0.0003, median survival time (MST) was 26 and 55 months) (*Figure 1D*). The RFS of circPDK1^{low} and circPDK1^{high} was 17.7% and 25.9%, respectively [P=0.0034, median survival time (MST) was 23 and 45 months] (*Figure 1E*). To our data, elevated circPDK1 indicated better RFS and OS of resected LSCC.

Correlations between circPDK1 expression and clinicopathological characteristics

The study cohort included 223 males and 9 females with a median age of 65 years old (range, 47–82). Most patients had a smoking history (204 patients; 87.9%). According to the 8th TNM staging, 70 patients (30.2%) were in stage I, 83 ones (35.8%) were in stage II, and 79 ones (34.1%) were in stage III. The number of patients who underwent lobectomy, bilobectomy, pneumonectomy, and bronchial sleeve resection was 141 (60.8%), 16 (6.9%), 24 (10.3%) and 49 (21.1%), respectively. Fifty-six patients (24.1) received adjuvant therapies postoperatively, while 120 ones (51.7%) were not. The detail clinicopathological features among the enrolled patients are listed in *Table 1*.

In the study cohort, circPDK1^{low} showed more lymph nodes metastasis (50.0% *vs.* 15.7%, P=0.025), more VI (80.6% *vs.* 66.7%, P=0.047), more VPI (76.6% *vs.* 36.1%, P=0.015) and more cases with serum CK19 higher than 2.2 ng/mL (58.1% *vs.* 36.1%, P=0.033) than circPDK1^{high}. In the validation cohort, the results were similar, except for VI (P=0.202) (*Table 1*).

Survival analysis in the study cohort

According to univariate analysis in study cohort with the Cox proportional hazards regression model, we found multiple clinicopathologic factors were significant to both RFS and OS, including older than 65 years (P=0.021 and 0.014, respectively), larger tumor size (P=0.008 and 0.012, respectively), more advanced pathologic stage (P<0.001, respectively), more advanced pathologic T stage (P<0.001, respectively), more advanced pathologic N stage (P<0.001, respectively), LN metastasis status (P<0.001, respectively), higher tumor grading (P=0.017 and 0.008, respectively), presence of VPI (P=0.004 and 0.001, respectively), less circPDK1 expression in tumor (P=0.001 and 0.002, respectively), higher level of serum CEA (P=0.011 and

912

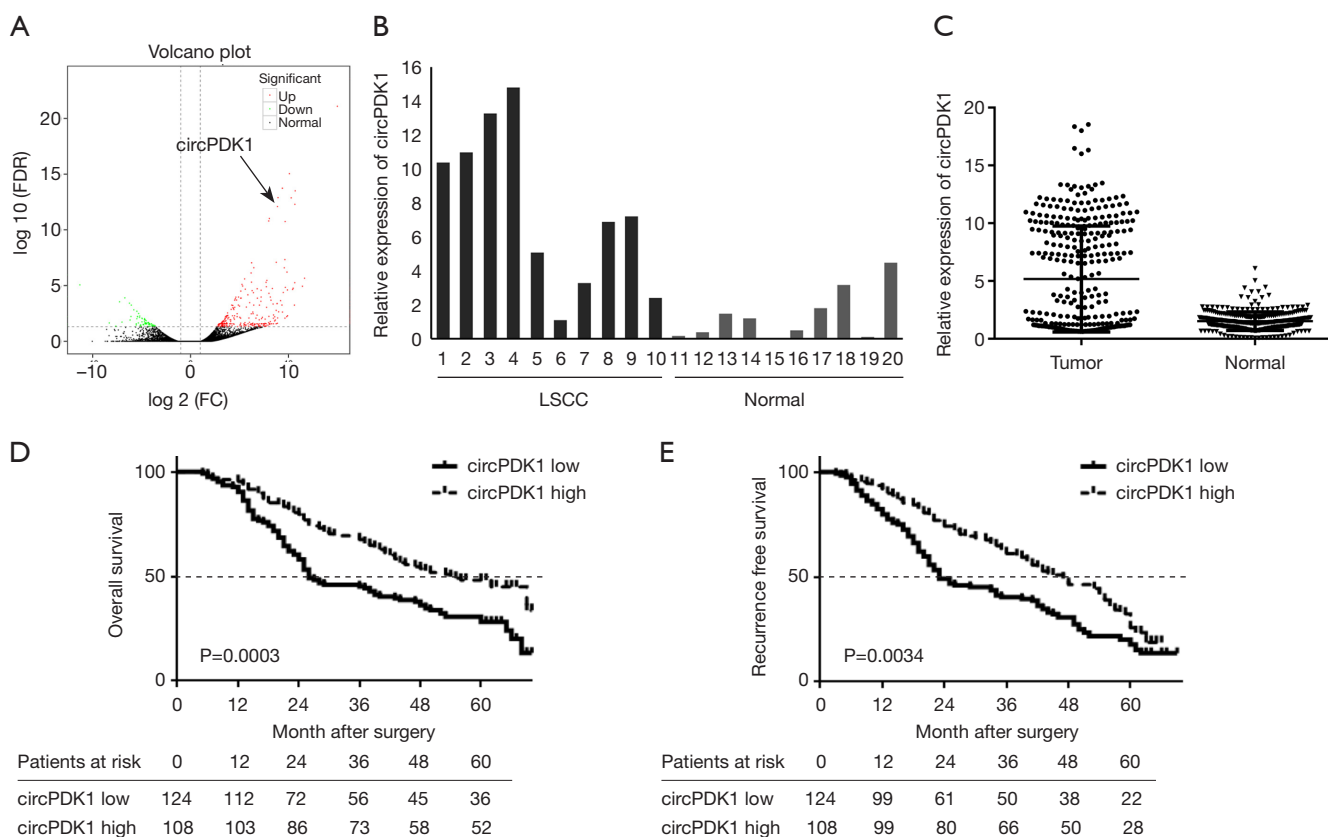Sun et al. Prognostic model for lung cancer

Figure 1 CircPDK1 overexpressed in LSCC tissues and led to a better prognosis. (A) In RNA expression volcano map, circPDK1 was pointed out by the arrow. (B) Expression profile of circPDK1 in LSCC tumor and normal tissues [Tumor (n=10) *vs.* Normal (n=10), detected by the next-generation sequencing]. (C) CircPDK1 expression in LSCC tumor tissues and their adjacent normal tissues detected with qRT-PCR (n=232). (D) 5-year OS of the circPDK1-low group [circPDK1 expression ≤3.50 (the mean value), n=124] and the circPDK1-high group (circPDK1 >3.5, n=108). (E) 5-year RFS of the two groups.

0.004, respectively) and SCCAg (P=0.003 and 0.023, respectively). Also, the presence of VI was significantly associated with RFS (P=0.012) but not to OS (P=0.220). At the multivariate analysis, age 65 years or older (P=0.029 and 0.011, respectively), pathologic T stage III and IV (P=0.001 and 0.007, respectively), LN metastasis status (P=0.008 and 0.014, respectively), SCCAg >1.5 ng/mL (P=0.032 and 0.023, respectively) and relative circPDK1 expression >3.7 (P=0.015 and 0.031, respectively) were significant independent predictors for shorter DFS and OS (*Table 2*).

The prognostic value of circPDK1 has been proved via K-M plot analysis before (*Figure 1D,E*). The survival rates according to the presence of other four independent predictors were analyzed by K-M plot, either. To the results, patients older than 65 years showed significantly poorer OS and RFS than the patients 65 years old or

younger (P<0.001, respectively) (*Figure 2A,B*). While to the patients with higher serum SCCAg (>1.5 ng/mL), a significantly lower curve was observed in RFS (P=0.0424), but not in OS (P=0.0640) (*Figure 2C,D*). Furthermore, T3-4 patients had a significantly poorer OS and RFS than T1-2 ones (P<0.0001, respectively) (*Figure 2E,F*). Finally, the LN positive patients also showed significantly poorer OS and RFS than negative ones (P=0.0321 and 0.0385, respectively) (*Figure 2G,H*).

A prognostic model based on circPDK1 and multiple clinicopathological characteristics

According to the outcomes of the multivariate analysis, a prognostic model was built with circPDK1, T staging, LN status, age, and SCCAg. The nomograms were performed

Table 2 Univariate and multivariate analysis for RFS and OS in the study cohort

Characteristics, factor	Univariate analysis						Multivariate analysis					
	RFS			OS			RFS			OS		
	HR	95% CI	P value	HR	95% CI	P value	HR	95% CI	P value	HR	95% CI	P value
Clinical factors												
Gender, male	1.262	0.301–2.101	0.511	1.278	0.592–2.710	0.436	–	–	–	–	–	–
Age (>65 vs. ≤65)	2.450	1.676–3.251	0.021	2.745	1.597–3.650	0.014	1.409	0.836–2.396	0.029	1.592	0.889–2.316	0.011
Smoking history, former/current	2.201	1.532–3.144	0.347	2.306	1.512–3.367	0.303	–	–	–	–	–	–
Chemotherapy or others vs. none	1.011	1.135–3.111	0.424	1.230	0.593–2.882	0.211	–	–	–	–	–	–
Surgery, each method	0.971	0.487–1.522	0.601	0.913	0.479–1.501	0.622	–	–	–	–	–	–
Pathologic factors												
Tumor size (per cm)	4.686	2.953–6.376	0.008	4.040	2.846–6.882	0.012	–	–	–	–	–	–
Pathologic T stage (per stage)	3.233	2.081–4.778	0.002	3.044	1.500–4.659	0.002	–	–	–	–	–	–
Pathologic T stage, T1-2 vs. T3-4	7.212	5.676–9.434	<0.001	7.377	5.926–9.458	<0.001	4.203	2.972–5.480	0.001	4.224	2.879–5.496	0.007
LN positive vs. negative	5.022	2.996–7.002	<0.001	5.095	2.760–7.383	<0.001	3.223	1.997–4.428	0.008	3.450	1.755–4.433	0.014
Pathologic N stage (per stage)	2.516	1.178–4.097	<0.001	2.621	1.296–4.096	<0.001	–	–	–	–	–	–
Pathologic stage (per stage)	4.083	1.931–6.243	<0.001	3.473	2.004–5.430	<0.001	1.231	0.570–3.001	0.156	1.344	0.558–2.988	0.093
Pathologic stage, I vs. II, III	2.256	1.134–3.448	<0.001	2.433	1.107–3.522	<0.001	–	–	–	–	–	–
Pathologic stage, I, II vs. III	4.377	2.211–6.876	<0.001	4.825	2.034–6.733	<0.001	–	–	–	–	–	–
Grading (G1-2 vs. G3-4)	2.067	1.208–2.939	0.017	2.214	0.989–3.152	0.008	1.450	0.679–2.483	0.268	1.331	0.656–2.486	0.337
Vascular invasion, presence	2.485	1.518–3.272	0.012	2.112	1.392–3.300	0.220	–	–	–	–	–	–
Visceral pleural invasion, presence	3.234	1.803–4.956	0.004	3.147	1.657–5.009	0.001	–	–	–	–	–	–
Tumor markers												
CEA (>5.5 vs. ≤5.5)	3.019	1.820–4.270	0.011	2.586	1.586–4.298	0.004	2.022	1.079–2.885	0.138	2.047	0.976–2.983	0.152
SCCAg (>1.5 vs. ≤1.5)	3.692	1.881–5.406	0.003	3.351	1.691–5.348	0.023	2.011	1.104–3.023	0.032	2.029	0.999–3.066	0.023
CK19 (>3.3 vs. ≤3.3)	1.606	0.529–3.055	0.080	1.637	0.576–2.965	0.103	–	–	–	–	–	–
CA125 (>35 vs. ≤35)	1.258	0.616–2.127	0.487	1.374	0.552–2.019	0.172	–	–	–	–	–	–
NSE (>16.3 vs. ≤16.3)	1.818	0.773–3.313	0.323	2.046	1.131–2.950	0.251	–	–	–	–	–	–
ProGRP (>50 vs. ≤50)	2.592	1.738–3.592	0.024	2.685	1.673–3.878	0.058	–	–	–	–	–	–
CircPDK1 (>3.7 vs. ≤3.7)	3.802	1.894–5.374	0.001	3.734	1.504–5.355	0.002	2.277	1.329–3.350	0.015	1.959	1.294–2.675	0.031

HR, hazard ratio; CI, confidence interval.

914
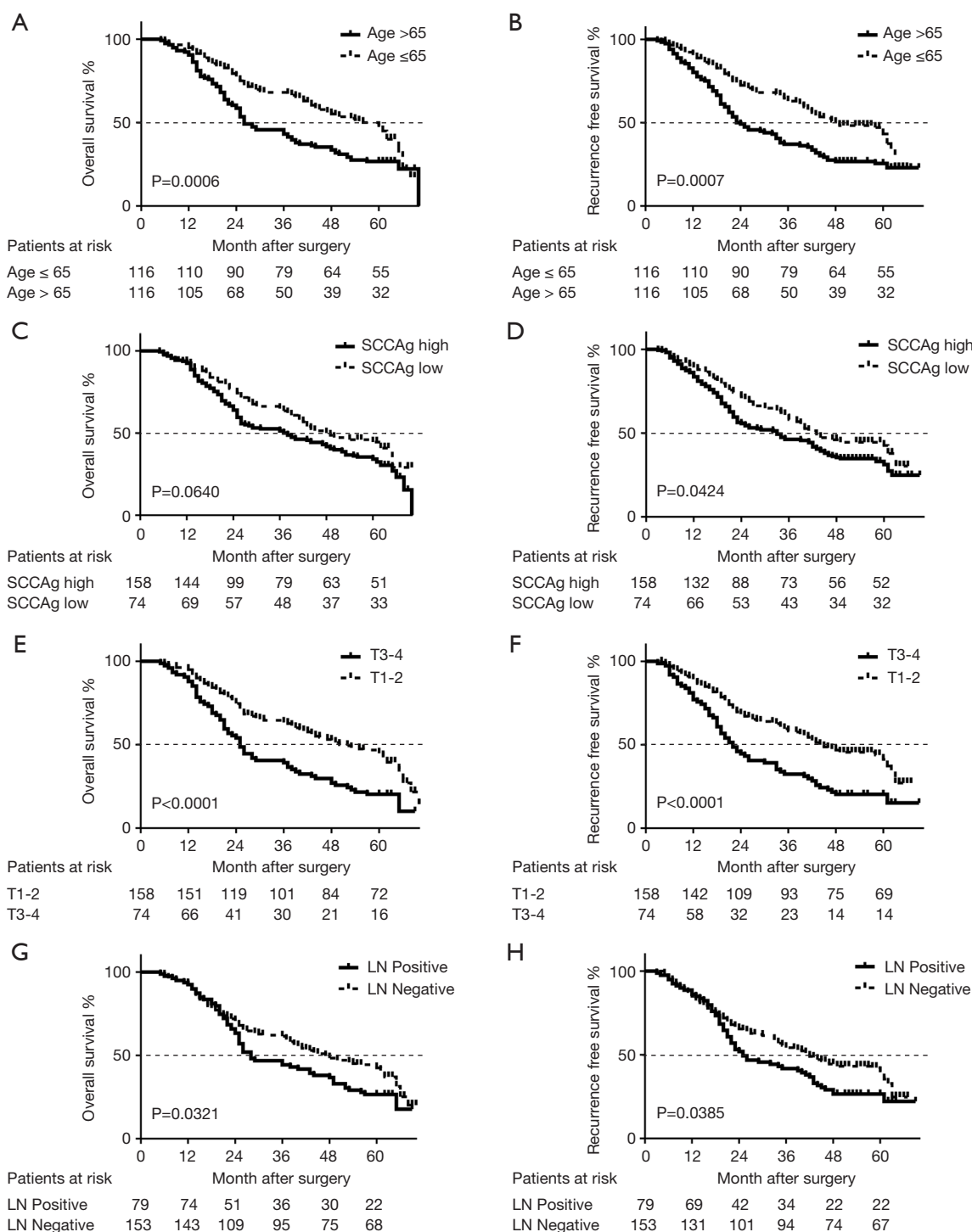
Sun et al. Prognostic model for lung cancer

Figure 2 Kaplan-Meier survival analysis of OS and RFS according to the age, serum SCCAg level, T descriptor according to TNM 8th edition and lymph nodes metastasis in the study cohort. (A) OS in patients ≤65 years old (n=116) *vs.* patients >65 years old (n=116). (B) RFS of the patients younger than 65 *vs.* the patients older than 65. (C) OS in patients with low level of serum SCCAg (≤1.5 ng/mL, n=114) *vs.* patients with high SCCAg (>1.5 ng/mL, n=118). (D) RFS in patients with low SCCAg *vs.* those with high SCCAg. (E) OS in T1-2 patients (n=158) *vs.* T3-4 ones (n=74). (F) RFS in T1-2 patients *vs.* T3-4 ones. (G) OS in patients with lymph nodes positive (n=79) *vs.* lymph nodes negative (n=153). (H) RFS in patients with lymph nodes positive *vs.* lymph nodes negative.
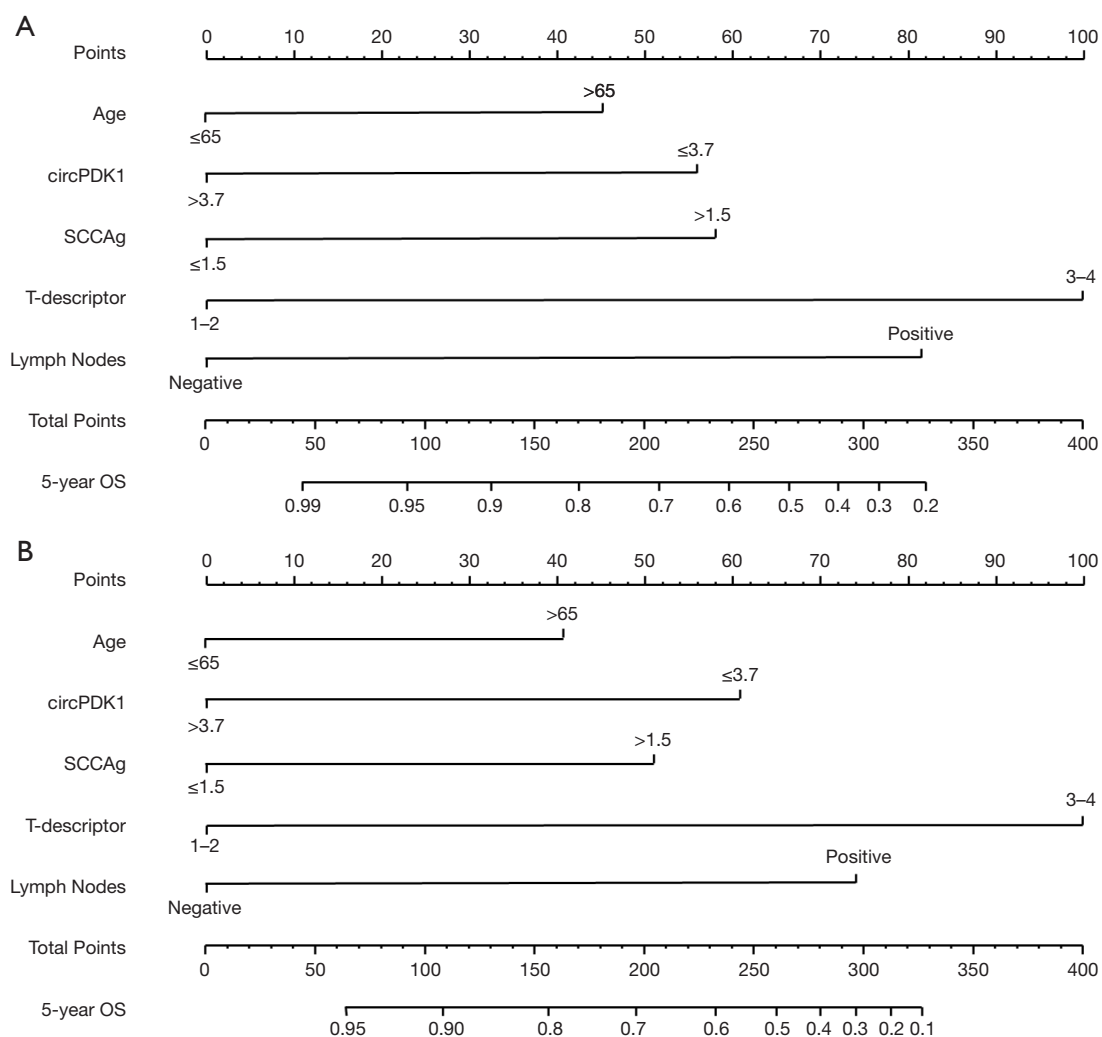
Figure 3 Nomograms for 5-year RFS and OS in the study cohort. (A) Nomogram of 5-year RFS. (B) Nomogram of 5-year OS. In the nomogram, a line is drawn upwards to determine the number of points received for each risk factor. The sum of these numbers is located on the total points axis, and a line is then drawn downwards to the survival axis to predict the 5-year RFS and OS attributing an individual patient.

to predict the 5-year RFS and OS (*Figure 3*). The internal validation of the accuracy to predict RFS and OS has been shown with calibration curves (*Figure 4A,B*). According to the curves, there was a great correlation between the nomogram-estimated prediction and the actual observed result. The C-index of our model to 5-year RFS and OS was 0.8214 and 0.8359, respectively.

Predictive accuracy of the model in the validation cohort

To perform external validation, the predictive accuracy of this prognostic model to 5-year OS and RFS of resected

LSCC was evaluated in the validation cohort, either. In the validation cohort, the C-index of our model for predicting 5-year RFS and OS was 0.8098 and 0.7924, respectively, which showed the excellent discriminative ability of prognosis. Finally, the nomograms were confirmed in the validation cohort, and the calibration curves proved high accuracy of prognosis prediction (*Figure 4C,D*).

Discussion

In this work, we identified a circular RNA, circPDK1, which is overexpressed in LSCC tumor tissues and exhibited
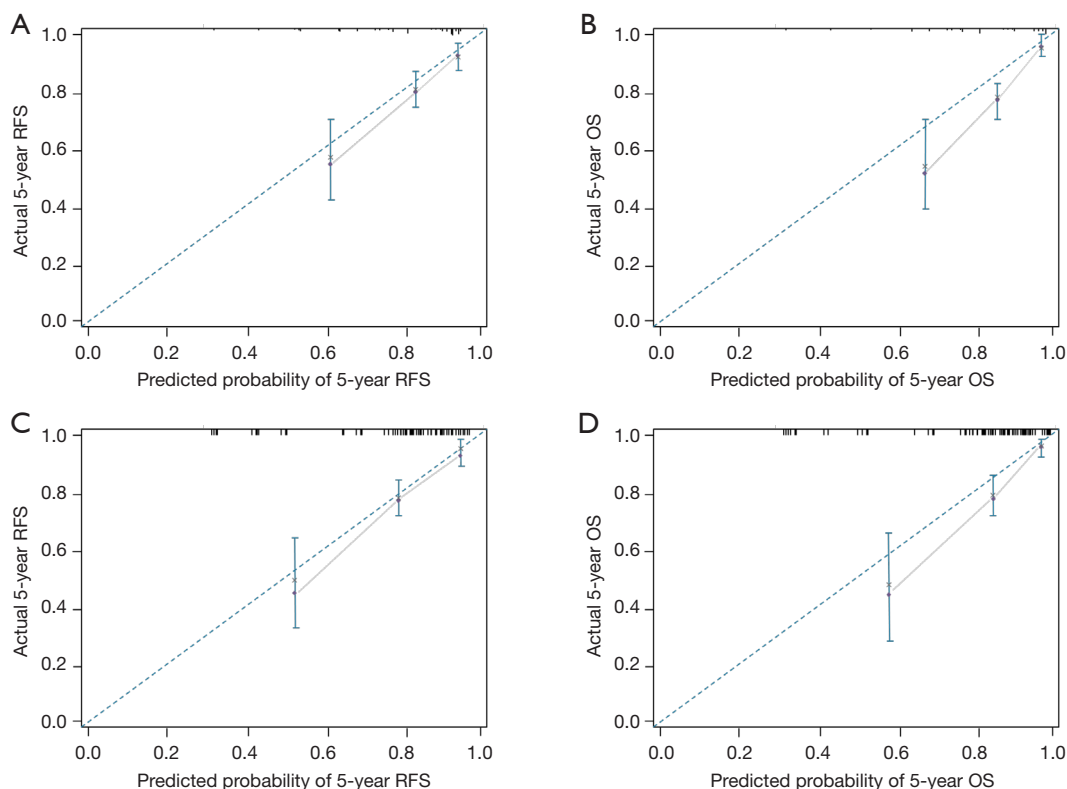
Figure 4 Calibration curves of the nomograms for 5-year RFS and OS in the study and validation cohorts. (A) The calibration curve for the prediction of 5-year RFS in the study cohort. (B) The calibration curve for the prediction of 5-year OS in the study cohort. (C) The calibration curve for the prediction of 5-year RFS in the validation cohort. (D) The calibration curve for the prediction of 5-year OS in the validation cohort.

well susceptibility and specificity for LSCC prognosis. Eventually, a new prognostic model was built with circPDK1 and several clinicopathologic factors (T staging, LN status, age, and SCCAg) and verified to be efficient for 5-year RFS/OS prediction of resected LSCC patients.

Nomograms model were used to calculate the probability of 5-year OS and RFS in this work, which has been validated as a reliable method for prognostic prediction (19). Our model consisted of five factors: circPDK1, T staging, LN status, age, and SCCAg that were independent prognostic factors according to the multivariate analysis. According to the hazard ratio (HR) of multivariate analysis, we integrated these clinicopathologic factors into the nomograms, thereby providing an individualized prognostic assessment for each individual. For example, a 68-year-old LSCC patient who underwent R0 resection and systematic LN dissection showed the clinicopathologic features of T3 without mediastinal LN metastasis; his preoperative serum SCCAg concentration is 1.3 ng/mL; the relative expression

of circPDK1 in his tumor tissues is 2.7. The nomograms demonstrate that his 5-year RFS and OS is 67% and 72%, respectively. It is indicated that after 5 years, the recurrence risk of this patient is 33%, and the cancer-related mortality risk is 28%. C-index shows the probability that the predicted results are consistent with the observed ones, which is always used to assess the predictive capability of the prognostic model (19). In the study cohort, the C-index of the nomograms for RFS and OS was 0.8214 and 0.8098, respectively. While in the validation cohort, the C-index for RFS and OS was 0.8359 and 0.7924, respectively. Hence, our model that based on circPDK1, T staging, LN status, age, and SCCAg could accurately predict the 5-year RFS and 5-year OS of resected LSCC.

Similar methods have been used in other studies. In the year 2015, Pilotto *et al*. built a risk stratification model with age, T staging, LN metastasis and tumor grading for prognostic prediction of resected LSCC and performed the external multicenter validation of their model in the year

2017. Moreover, Wang *et al.* exhibited their nomograms model based on CEA, CK19, TNM staging, and Karnofsky performance status (KPS) for 1-, 3-, and 5-year OS of resected LSCC in 2018. Recently, Liu *et al.* verified the prognostic value of the tumor spread through air spaces (STAS) in lung adenocarcinoma (LUAD) and then set up a prognostic model based on STAS, VPI, VI, and histological subtype (8,20-22). Compared with those studies, there were more clinicopathological characteristics in our data, such as postoperative adjuvant therapies and multiple serum tumor markers. Besides, our model firstly included a circular RNA, which exhibited excellent potential to be a prognostic marker of LSCC. In considering the stability and abundance in tissues and blood serum, circular RNAs would become the clinical frequently-used biomarkers in the future.

In earlier studies, many serum lung cancer markers have been showed the prognostic value in LSCC (7-9,23). However, the views about the biomarkers' prediction efficiency were inconsistent. In the year 2013, Nagashima *et al.* (24) proposed that older age, high serum CEA level, and nodal metastasis are significant prognostic factors to LSCC. While in 2014, Kinoshita *et al.* (17) reported that high serum SCCAg (>5.5 ng/mL) indicated the poorer prognosis of early-stage LSCC when combined with VPI status and VI status. However, the prognostic role of CEA was not observed in the latter study. One hand, the cut off values of tumor markers in different studies were inconsistent, that had a significant impact on the diagnostic usability of the test results (4,9,25-27). In our institution, the standard cut off values of CEA, SCCAg, NSE, CK19, CA125, ProGRP were 3.4 ng/mL, 1.5 ng/mL, 17 ng/mL, 3.3 ng/mL, 35 U/mL and 63 pg/mL, respectively. However, to enhance the rigor and persuasiveness of this work, we set the cut off values according to several pieces of authoritative literature similar to our study (CEA, 5.5 ng/mL; NSE, 16.3 U/mL and ProGRP, 50 pg/mL) (8,17,18). In the other hand, there were inevitable confounding factors in retrospective data that lead to statistical bias. CEA and pathologic stage have been reported as independent risk factors of LSCC by multiple studies. To our data, no statistical significances were observed in CEA (RFS and OS: P=0.138 and 0.152, respectively) and pathologic stage (RFS and OS: P=0.156 and 0.093, respectively) to the multivariate analysis. In considering that the P values of these items were close to 0.05; thus, the outcomes might be interfered by some confounding factors.

According to the next-generation sequencing data, circPDK1 was transcribed by *pyruvate dehydrogenase kinase 1* (*PDK1*) gene (28). The function of this circular RNA has never been revealed in NSCLC before. To our data, elevated circPDK1 was associated with fewer lymph nodes metastasis, less VI and VPI presence and lower serum concentration of CK19. These results indicated an antitumor function of it in LSCC. However, although transcribed from the same gene, linear PDK1 and circPDK1 exert distinct functions in lung cancer. According to previous studies, PDK1 promotes cancer malignant behaviors, whereas inhibition of PDK1 increases the anti-cancer effect of EGFR tyrosine kinase inhibitors in non-small cell lung cancer (29). We also detected the expression of mRNA PDK1 in LSCC, while there was no significant difference between the tumor and normal tissues (*Figure S1A*). In addition, we observed a negative correlation between circPDK1 and linear PDK1 in p-LSCC (r=−0.71, P=0.009; *Figure S1B*). We assumed that, when more circPDK1 was transcribed, the production of linear PDK1 decreased. We have also analyzed the role of linear RNA PDK1 in LSCC in public database (Data source: K-M plot database, http://kmplot.com), and found that its association with LSCC prognosis was not as strong as that of circPDK1 (*Figure S1C,D,E*). All these findings were worthy of further studies.

This study has some limitations. First, although it is a multicenter study, the sample size was still small for a prognostic model. However, to detect the expression of circPDK1, we must get well-preserved LSCC tissues of all the study objects, while most of the LSCC specimens were poor preserved or not collected 5 years ago. Second, the function of circPDK1 in LSCC has never been reported by previous studies. Hence, it is unconvincing to build a prognostic model with this RNA. However, in our analysis, we have confirmed its potential as a prognostic marker in LSCC. We should reveal the function of circPDK1 in LSCC and further explore its underling molecular biological mechanisms in next work.

Footnote

Conflicts of Interest: The authors have no conflicts of interest

918

Sun et al. Prognostic model for lung cancer

to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The medical ethics for this research were approved by the Ethic Committee of the Affiliated Hospital of Qingdao University (No. QYFYKY 2018-10-11-2).

References

1. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2009. CA Cancer J Clin 2009;59:225-49.
2. Lopci E, Rossi S. Tumor metabolism and prognostic role of EZH2 in non-small cell lung cancer. Transl Cancer Res 2017;6:S982-8.
3. Martínez-Terroba E, Behrens C, Agorreta J, et al. 5 protein-based signature for resectable lung squamous cell carcinoma improves the prognostic performance of the TNM staging. Thorax 2019;74:371-9.
4. Pollan M, Varela G, Torres A, et al. Clinical value of p53, c-erbB-2, CEA and CA125 regarding relapse, metastasis and death in resectable non-small cell lung cancer. Int J Cancer 2003;107:781-90.
5. Saintigny P, Coulon S, Kambouchner M, et al. Real-time RT-PCR detection of CK19, CK7 and MUC1 mRNA for diagnosis of lymph node micrometastases in non small cell lung carcinoma. Int J Cancer 2005;115:777-82.
6. Watanabe S, Oda M, Go T, et al. Should mediastinal nodal dissection be routinely undertaken in patients with peripheral small-sized (2 cm or less) lung cancer? Retrospective analysis of 225 patients. Eur J Cardiothorac Surg 2001;20:1007-11.
7. Johnson DH, Marangos PJ, Forbes JT, et al. Potential utility of serum neuron-specific enolase levels in small cell carcinoma of the lung. Cancer Res 1984;44:5409-14.
8. Wang J, Jiang W, Zhang T, et al. Increased CYFRA 21-1, CEA and NSE are Prognostic of Poor Outcome for Locally Advanced Squamous Cell Carcinoma in Lung: A Nomogram and Recursive Partitioning Risk Stratification Analysis. Transl Oncol 2018;11:999-1006.
9. Hatzakis KD, Froudarakis ME, Bouros D, et al. Prognostic value of serum tumor markers in patients with lung cancer. Respiration 2002;69:25-9.
10. Salzman J. Circular RNA Expression: Its Potential Regulation and Function. Trends Genet 2016;32:309-16.
11. Barrett SP, Salzman J. Circular RNAs: analysis, expression and potential functions. Development 2016;143:1838-47.
12. Vo JN, Cieslik M, Zhang Y, et al. The Landscape of Circular RNA in Cancer. Cell 2019;176:869-81.e13.
13. Hansen TB, Kjems J, Damgaard CK. Circular RNA and miR-7 in cancer. Cancer Res 2013;73:5609-12.
14. Chen J, Li Y, Zheng Q, et al. Circular RNA profile identifies circPVT1 as a proliferative factor and prognostic marker in gastric cancer. Cancer Lett 2017;388:208-19.
15. Chen B, Huang S. Circular RNA: An emerging non-coding RNA as a regulator and biomarker in cancer. Cancer Lett 2018;418:41-50.
16. Li Z, Yanfang W, Li J, et al. Tumor-released exosomal circular RNA PDE8A promotes invasive growth via the miR-338/MACC1/MET pathway in pancreatic cancer. Cancer Lett 2018;432:237-50.
17. Kinoshita T, Ohtsuka T, Hato T, et al. Prognostic factors based on clinicopathological data among the patients with resected peripheral squamous cell carcinomas of the lung. J Thorac Oncol 2014;9:1779-87.
18. Lang D, Horner A, Brehm E, et al. Early serum tumor marker dynamics predict progression-free and overall survival in single PD-1/PD-L1 inhibitor treated advanced NSCLC-A retrospective cohort study. Lung Cancer 2019;134:59-65.
19. Iasonos A, Schrag D, Raj GV, et al. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol 2008;26:1364-70.
20. Pilotto S, Sperduti I, Novello S, et al. Risk Stratification Model for Resected Squamous-Cell Lung Cancer Patients According to Clinical and Pathological Factors. J Thorac Oncol 2015;10:1341-8.
21. Pilotto S, Sperduti I, Leuzzi G, et al. Prognostic Model for Resected Squamous Cell Lung Cancer: External Multicenter Validation and Propensity Score Analysis exploring the Impact of Adjuvant and Neoadjuvant Treatment. J Thorac Oncol 2018;13:568-75.
22. Liu A, Hou F, Qin Y, et al. Predictive value of a prognostic model based on pathologic features in lung invasive adenocarcinoma. Lung Cancer 2019;131:14-22.
23. Kulpa J, Wojcik E, Reinfuss M, et al. Carcinoembryonic antigen, squamous cell carcinoma antigen, CYFRA 21-1, and neuron-specific enolase in squamous cell lung cancer patients. Clin Chem 2002;48:1931-7.
24. Nagashima T, Sakao Y, Mun M, et al. A clinicopathological study of resected small-sized squamous cell carcinomas of the peripheral lung: prognostic significance of serum

carcinoembryonic antigen levels. Ann Thorac Cardiovasc Surg 2013;19:351-7.

25. Yang DW, Zhang Y, Hong QY, et al. Role of a serum-based biomarker panel in the early diagnosis of lung cancer for a cohort of high-risk patients. Cancer 2015;121 Suppl 17:3113-21.

26. Sigari N, Mohsenpour B, Nikkhoo B, et al. Determination of the best prognostic value of serum tumor markers in patients with suspected lung cancer in an Iranian population. Clin Lab 2014;60:23-7.

27. Ishibashi N, Maebayashi T, Aizawa T, et al. Serum tumor marker levels at the development of intracranial metastasis in patients with lung or breast cancer. J Thorac Dis 2019;11:1765-71.

28. Sradhanjali S, Reddy MM. Inhibition of Pyruvate Dehydrogenase Kinase as a Therapeutic Strategy against Cancer. Curr Top Med Chem 2018;18:444-53.

29. Yang Z, Zhang SL, Hu X, et al. Inhibition of pyruvate dehydrogenase kinase 1 enhances the anti-cancer effect of EGFR tyrosine kinase inhibitors in non-small cell lung cancer. Eur J Pharmacol 2018;838:41-52.
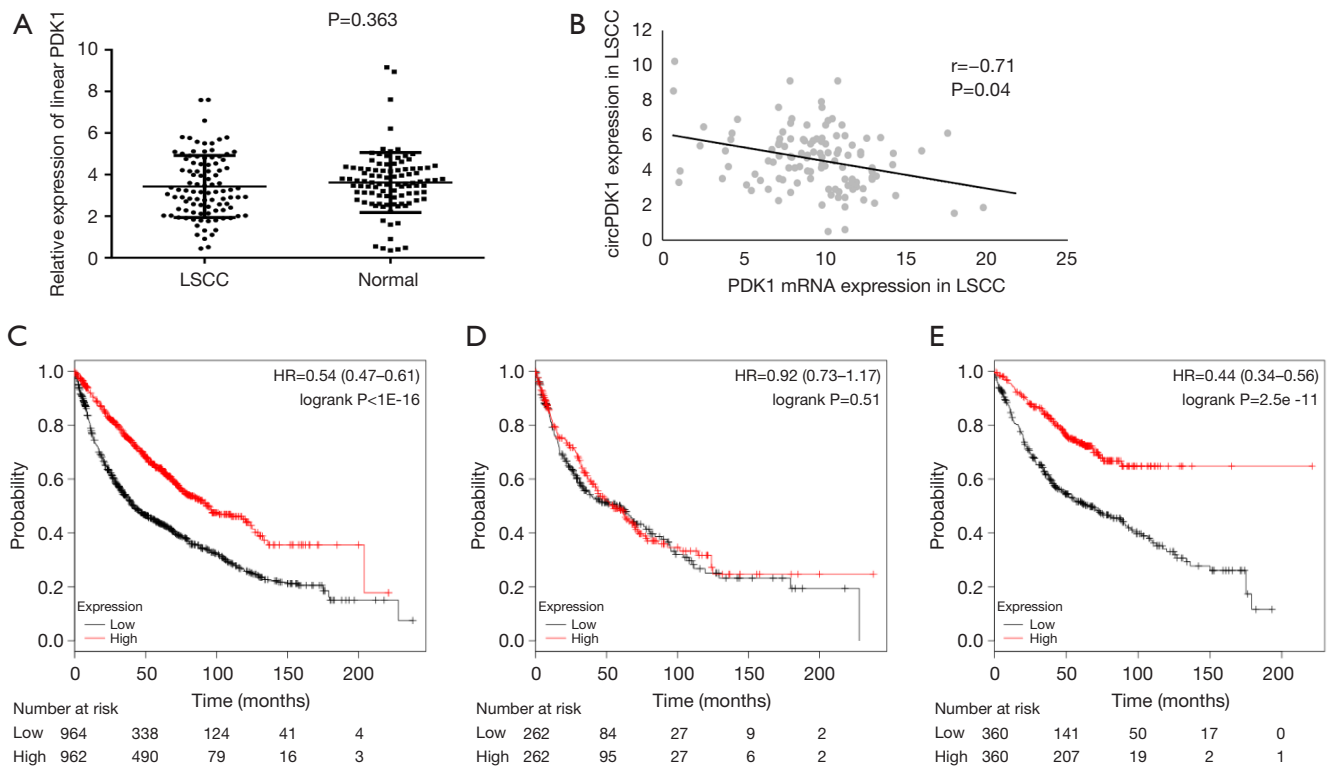
Figure S1 Correlation between linear PDK1 and lung cancer. (A) Relative expression of mRNA PDK1 in LSCC tumor and normal tissues. (B) Correlation between the expression of circPDK1 and PDK1. (C) The association between PDK1 and lung cancer (n=964 *vs.* 962). (D) The association between PDK1 and LSCC (n=262 *vs.* 262). (E) The association between PDK1 and LUAD (n=360 *vs.* 360).