Int J Clin Exp Med 2019;12(10):12086-12093
www.ijcem.com /ISSN:1940-5901/IJCEM0097666

Original Article

The identification of genes associated with NSCLC prognosis based on an integrated bioinformatics analysis

Kan Li[1], Yan Xiong[2], Bin Li[3], Jiafu He[4]

[1]Department of Basic Medicine, Yangtze University, Jingzhou 434000, Hubei Province, PR China; [2]Department of Obstetrics and Gynecology, The Second Hospital of Jingzhou, Jingzhou 434000, Hubei Province, PR China; [3]Department of Cardiovascular Medicine, Xiangyang Central Hospital of Hubei Province, Xiangyang 441021, Hubei Province, PR China; [4]Department of Respiratory, Xiangyang Central Hospital of Hubei Province, Xiangyang 441021, Hubei Province, PR China

Received May 28, 2019; Accepted August 7, 2019; Epub October 15, 2019; Published October 30, 2019

Abstract: Objective: To investigate the pathway and the prognostic significance of differently expressed genes in NSCLC using a bioinformatics analysis. Methods: Up-regulated genes between cancer tissue and normal lung tissue from NSCLC patients were screened in the Oncomine database. The overlapping up-regulated genes in all the datasets were included for further analysis. The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEEG) databases were used to identify the genes' biological functions and relevant pathways. The protein-protein interaction (PPI) network was constructed using the STRING database. And hub genes from the included genes were identified using cytoHubba. The correlation between gene expression and prognosis was evaluated using the Kaplan-Meier Plotter database. Results: Twelve up-regulated genes were finally identified (*PYCR1, TOP2A, KIAA0101, COL11A1, TPX2, ZWINT, MELK, CCNB1, TK1, UBE2C, PAICS, KDM5B*). These genes were enriched in the positive regulation of the apoptotic process, nucleus and ATP binding in terms of the biological process (BP), cellular component (CC), and molecular function (MF) respectively. The KEGG pathway analysis indicated the 12 genes were enriched in the signal pathway of the cell cycle. *CCNB1* was identified from the PPI network as the hub gene which can lead to uncontrolled cell growth by binding to CDKs. All except for *PAICS* and *KDM5B* of the 12 included genes were found to be correlated with the patients' overall survival (OS) and progression-free survival (PFS). Conclusion: *PYCR1, TOP2A, KIAA0101, COL11A1, TPX2, ZWINT, MELK, CCNB1, TK1, and UBE2C* were enriched in the positive regulation of the apoptotic process, nucleus, and ATP binding and were correlated with poor survival of NSCLC, and can be used as biomarkers for NSCLC patients' prognoses.

Keywords: Non-small cell lung cancer, bioinformatics analysis, prognosis

Introduction

Lung cancer, including non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) is the most common cause of cancer associated death [1, 2]. NSCLC is the main type of lung cancer, accounting for 70-80% of all lung cancers [2]. The general prognosis of NSCLC is poor because patients tend to be in an advanced stage when fist diagnosed [3, 4]. Chemoradiation is the main treatment method for advanced NSCLC patients [5-7]. One of the main reasons for the poor treatment response and prognosis is due to an unclear understanding of the molecular mechanisms of lung cancer development and progression [8, 9]. Therefore, to elucidate the molecular mechanisms of lung cancer development and its signaling pathway is important for improving the treatment response and prognosis. With the development of molecular biology and bioinformatics, lots of data from microarray and gene expression sequencing can be accessed from open gene expression databases [10, 11]. These databases provided us with free platforms for lung cancer analysis, including carcinogenesis, metastasis, invasion, metastasis, prognosis, etc.

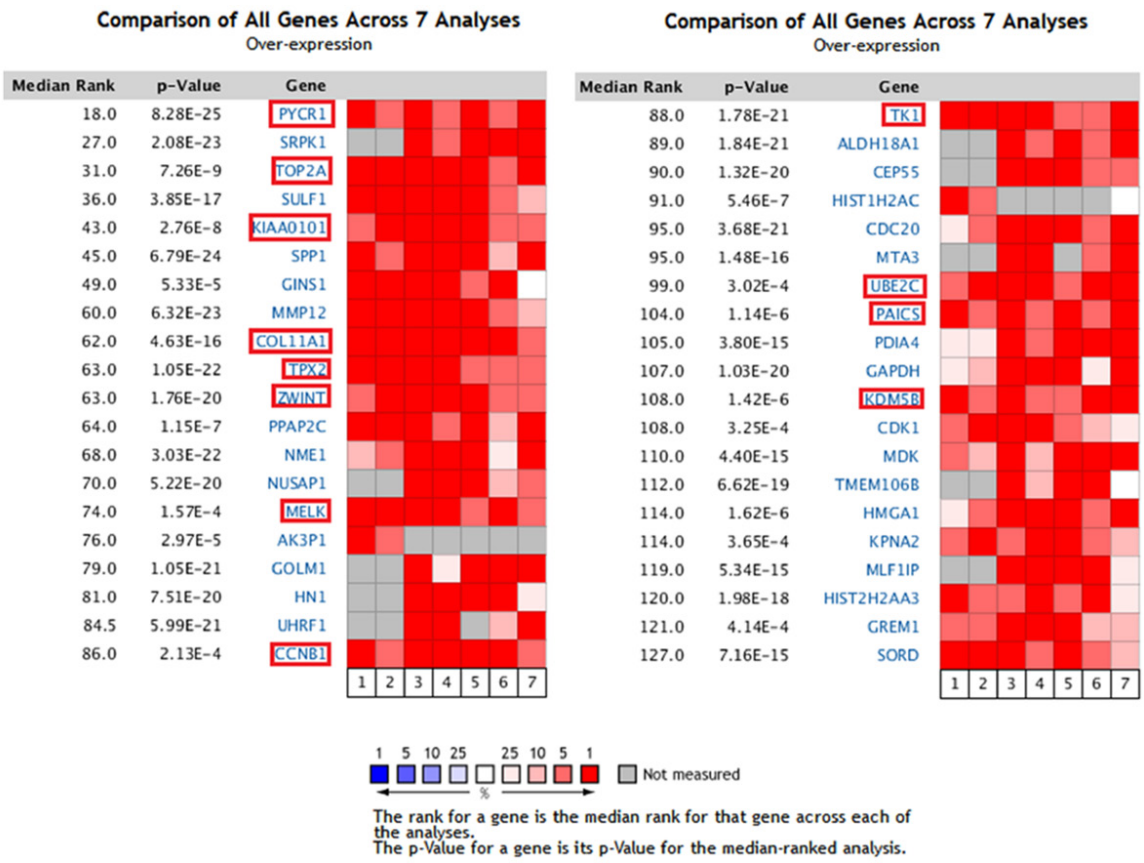

Figure 1. Up-regulated gene identification from the Oncomine database.

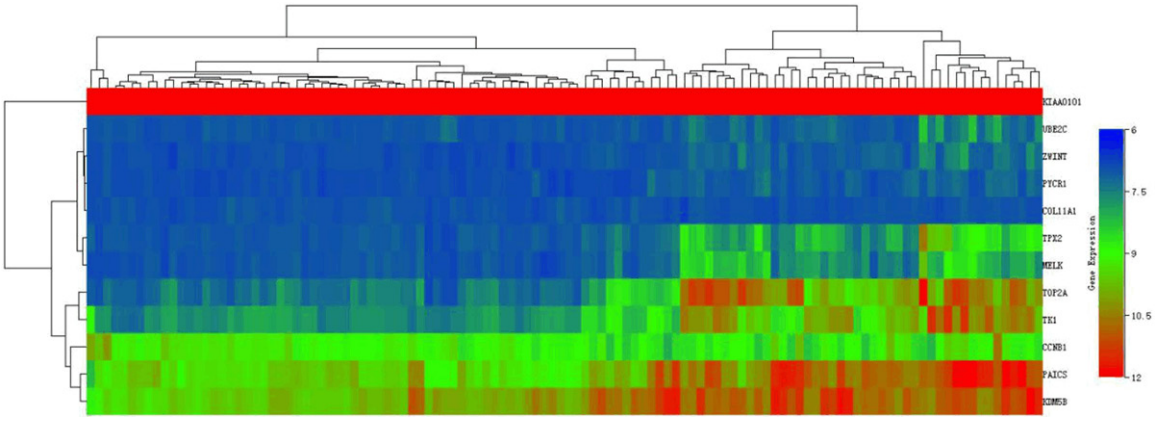

Figure 2. The heat map of the up-regulated 12 genes between cancer tissue and normal lung tissue of NSCLC (original data from GSE32863).

Table 1. GO and KEGG analysis of the identified 12 genes

Category	Term	Count	*P*-value
GOTERM_BP_DIRECT	Positive regulation of apoptotic process	2	8.6E-2
GOTERM_CC_DIRECT	nucleus	6	9.7E-3
GOTERM_MF_DIRECT	ATP binding	4	2.5E-2
KEGG_PATHWAY	Cell cycle	6	3.0E-3

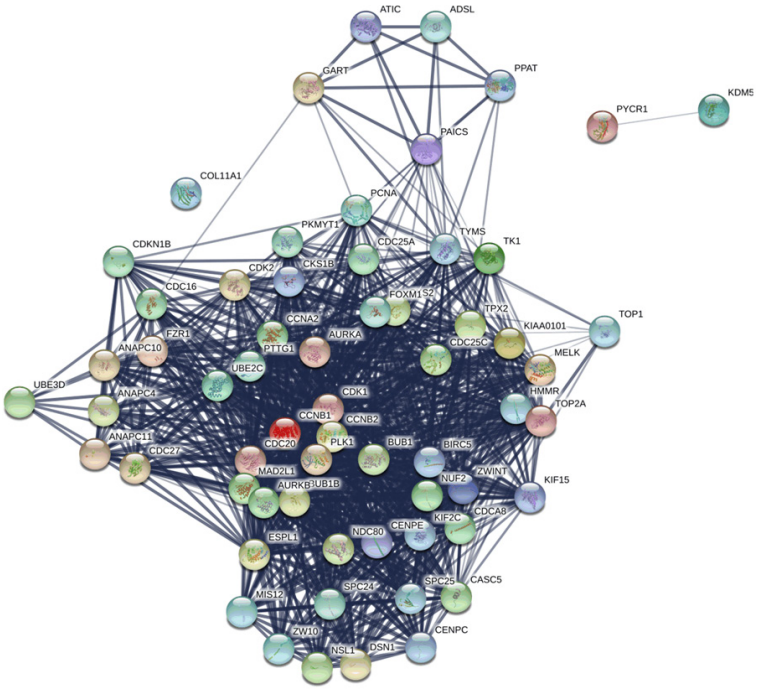

Figure 3. The PPI network of the included 12 genes constructed using STRING.

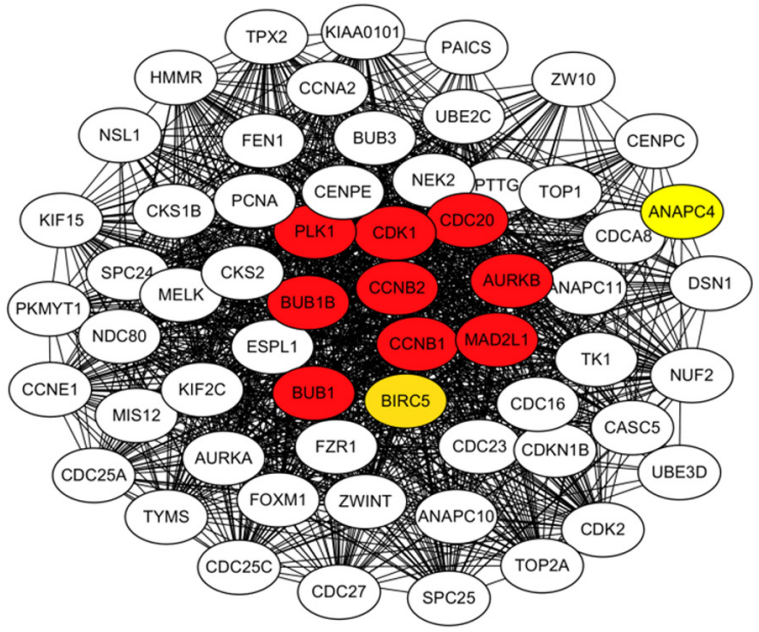

Figure 4. The identified hub gene of the network using a cytoHubba analysis.

Material and methods

Up-regulated gene screening

Up-regulated genes between cancer tissue and normal lung tissue of NSCLC patients were screened in the Oncomine database (https://www.oncomine.org/resource/login.html) [11]. The gene identification criteria were as follows: The search terms were "non-small cell lung cancer", and the cancer type was non-small cell lung cancer (adenocarcinoma of the lung and squamous cell lung carcinoma); the data type was mRNA; the sample size of the original studies was more than 100 subjects. The overlapping up-regulated genes in all the datasets were included for further analysis.

Biological function and pathway enrichment

The Database for Annotation, Visualization, and Integrated Discovery (DAVID) (https://david.ncifcrf.gov/) [12] online analysis tool was used for biological function and pathway enrichment. All the included genes were put into the gene list and electronic searching was carried out in the homo specimen.

PPI analysis

The STRING database (http://string-db.org/cgi/input.pl) [13] was applied for a protein-protein interaction network analysis. The network was constructed under the following conditions: a minimum required interaction score of 0.4; active interaction sources of text mining, experiments, databases, co-expression, neighborhood, gene fusion and co-recurrence; the maximum number of interactors was no more than 50. The hub genes were identified using Cytoscape software [14].

Survival analysis

The overall survival (OS) and progression free survival (PFS) of NSCLC patients were analyzed using the Kaplan-Meier Plotter database [15]. The hazard ration (HR) and corresponding 95%

Table 2. Correlation between included 12 expression and patients' prognosis

Gene	Median OS(months)		HR (95% CI)	*p*-value	Median PFS (months)		HR (95% CI)	*p*-value
	Low	High			Low	High		
PYCR1	77.6	61.2	1.22 (1.08-1.39)	0.0016	27.0	12.0	1.69 (1.39-2.06)	7.7e-8
TOP2A	92.97	47.00	1.55 (1.36-1.76)	1.6e-11	104.9	50.0	1.53 (1.26-1.86)	1.4e-5
KIAA0101	95.00	45.00	1.71 (1.15-1.94)	1.1e-16	104	52	1.59 (1.31-1.93)	1.7e-6
COL11A1	46.00	64.10	1.14 (1.01-1.30)	0.04	22.53	15.37	1.46 (1.2-1.77)	0.00011
TPX2	96.20	42.00	1.87 (1.64-2.12)	P<1e-16	31.00	10.50	1.8 (1.48-2.18)	2.1e-9
ZWINT	87.70	47.63	1.50 (1.32-1.71)	3e-10	104.9	50.56	1.58 (1.3-1.91)	3.3e-6
MELK	92.60	47.63	1.60 (1.41-1.82)	3.6e-13	104.9	45.08	1.74 (1.43-2.11)	1.5e-8
CCNB1	96.00	44.00	1.77 (1.56-2.02)	P<e-16	102.00	62.00	1.35 (1.12-1.64)	0.0019
TK1	125.77	45.47	2.07 (1.75-2.46)	P<e-16	40.21	17.07	1.86 (1.41-2.45)	7.6e-6
UBE2C	96.2	44.94	1.77 (1.55-2.01)	P<e-16	27.73	11.24	1.65 (1.36-2.00)	3.5e-7
PAICS	79.0	54.93	1.30 (1.14-1.47)	5.5e-5	84.00	88.77	1.11 (0.92-1.35)	0.27
KDM5B	69.01	70.30	1.04 (0.91-1.18)	0.57	102.00	70.27	1.22 (1.01-1.47))	0.043

confidence interval (95% CI) were calculated between the high expression and low expression groups.

Results

Up-regulated genes identification

Twelve up-regulated genes were identified from 7 comparisons of 5 studies in the Oncomine database when comparing cancer tissue to the normal lung tissue of the non-small cell lung cancer patients. All the selected 12 genes (*PYCR1, TOP2A, KIAA0101, COL11A1, TPX2, ZWINT, MELK, CCNB1, TK1, UBE2C, PAICS, KDM5B*) were highly expressed in all 5 studies (**Figure 1**). The differentially expressed genes between NSCLC cancer tissue and lung normal tissue are shown in **Figure 2** from GSE 32863 [16].

GO and KEGG analysis

For the gene ontology analysis, the included 12 genes were enriched in the positive regulation of the apoptotic process, nucleus, and ATP binding in terms of the biological process (BP), cellular component (CC), and molecular function (MF) respectively. The KEGG pathway analysis indicated that the 12 genes were enriched in the signal pathway of the cell cycle See **Table 1**.

Protein-protein interaction network

The protein-protein interaction network of the 12 included genes was constructed by STRING. 62 nodes and 1055 edges, and an average node degree of 34 was established in the 12 up-regulated genes. The local clustering coefficient was 0.84 with the PPI enrichment *p* value of 1.0E-16 (**Figure 3**). One hub gene (*CCNB1*) from the 12 up-regulated genes was identified from the PPI network, which can lead to uncontrolled cell growth by binding to CDKs (**Figure 4**).

Correlation with prognosis

The correlation between gene expression and patients' prognosis was analyzed using the Kaplan-Meier Plotter database (**Table 2**). Ten out of the 12 included genes were found to be correlated with the patients' overall survival (OS), (**Figure 5**) and progression-free survival (DFS) (**Figure 6**).

Discussion

In our present work, 12 up-regulated genes were ultimately included for further analysis in the aspects of biological function (biological process, cellular component, molecular function) and the KEGG pathway. These genes were generally enriched in the positive regulation of the apoptotic process, nucleus, and ATP binding in terms of the biological process (BP), cellular component (CC), and molecular function (MF) respectively. The KEGG pathway analysis indicated the 12 genes were enriched in the signal pathway of the cell cycle. This gene function and pathway all correlated with cancer development or metastasis. Of the included 12 genes, we found that, *CCNB1* (Cyclin B1) was a

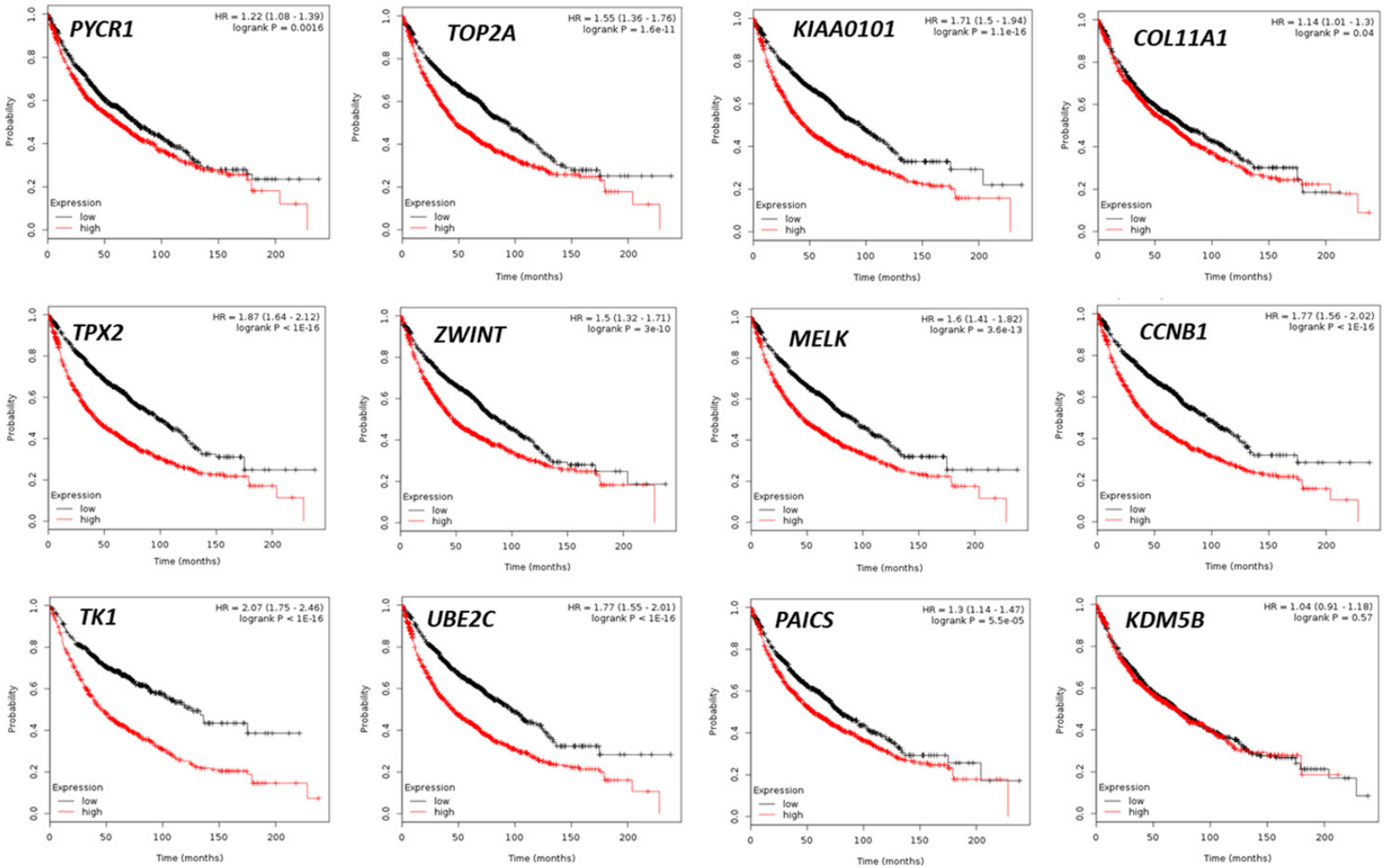

Figure 5. The correlation between gene expression and patients' overall survival (OS).

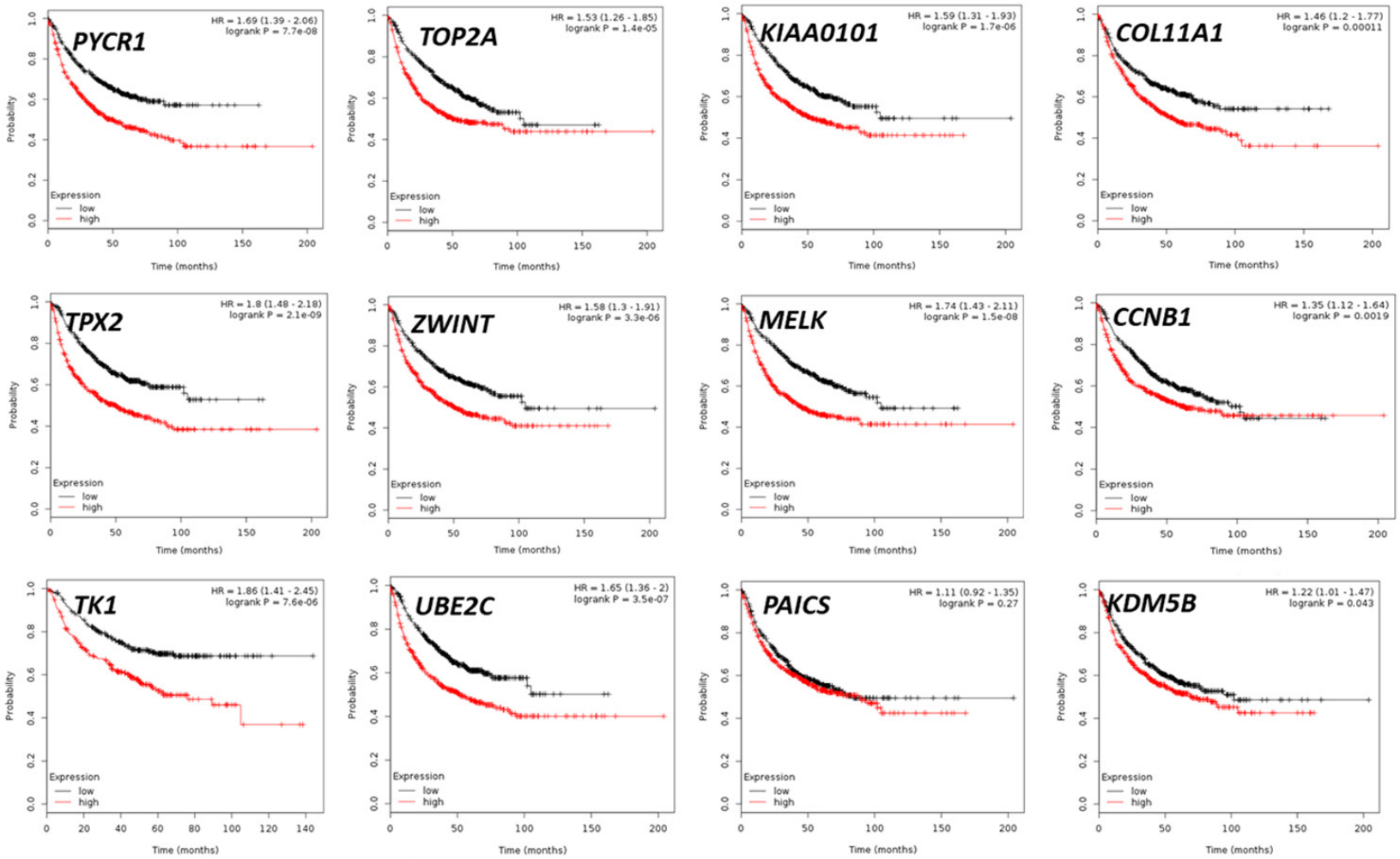

Figure 6. The correlation between gene expression and patients' progression free survival (PFS).

hub gene of the whole protein-protein interaction network constructed by the 12 included genes. The *CCNB1* gene belongs to a highly conserved protein family [17, 18]. It is located on human chromosome 5 q13.2 and contains nine exons [19]. The protein encoded by CCNB1 is a regulatory protein involved in mitosis. Zhou et al. found that CCNB1 was up-regulated in human pancreatic cancer and correlated with patients' poor prognosis [20]. Kosacka and his colleagues evaluated the prognostic value of cyclin B1 expression in patients with resected non-small-cell lung cancer stage I-IIIA, a preliminary report by immunohistochemical assay [21]. In their study, the authors revealed the prognosis was not correlated with cyclin B1 expression in NSCLC. However, Soria [22] found cyclin B1 was up-regulated in NSCLC and its high expression correlated with poor prognosis, which could be used as biomarker for prognosis. In our work, we found *CCNB1* mRNA was up-regulated in cancer tissue compared to normal lung tissue and its high expression correlated with the short OS (HR=1.77, P<0.0001) and PFS (HR=1.35, P<0.001).

Our results were in accordance with Soria's results [22].

Conclusion

PYCR1, TOP2A, KIAA0101, COL11A1, TPX2, ZWINT, MELK, CCNB1, TK1, and *UBE2C* are enriched in the positive regulation of the apoptotic process, nucleus, and ATP binding and correlate with the poor survival of NSCLC patients, and can be used as biomarker for NSCLC patients' prognosis.

Acknowledgements

This work was supported by project of science and technology development plan of Jingzhou City, Hubei Province (NO.2018061).

Address correspondence to: Dr. Yan Xiong, Department of Obstetrics and Gynecology, The Second Hospital of Jingzhou, No. 2 Jiangjin Road, Jingzhou 434000, Hubei Province, PR China. E-mail: ihrk8d@163.com

References

[1] Siegel RL, Miller KD and Jemal A. Cancer statistics, 2019. CA Cancer J Clin 2019; 69: 7-34.
[2] Cao M and Chen W. Epidemiology of lung cancer in China. Thorac Cancer 2019; 10: 3-7.
[3] Tang Y, Qiao G, Xu E, Xuan Y, Liao M and Yin G. Biomarkers for early diagnosis, prognosis, prediction, and recurrence monitoring of non-small cell lung cancer. Onco Targets Ther 2017; 10: 4527-4534.
[4] Woodard GA, Jones KD and Jablons DM. Lung cancer staging and prognosis. Cancer Treat Res 2016; 170: 47-75.
[5] Farago AF and Keane FK. Current standards for clinical management of small cell lung cancer. Transl Lung Cancer Res 2018; 7: 69-79.
[6] Mott TF. Lung cancer: management. FP Essent 2018; 464: 27-30.
[7] Tabchi S, Kassouf E, Rassy EE, Kourie HR, Martin J, Campeau MP, Tehfe M and Blais N. Management of stage III non-small cell lung cancer. Semin Oncol 2017; 44: 163-177.
[8] Weller A, O'Brien MER, Ahmed M, Popat S, Bhosle J, McDonald F, Yap TA, Du Y, Vlahos I and deSouza NM. Mechanism and non-mechanism based imaging biomarkers for assessing biological response to treatment in non-small cell lung cancer. Eur J Cancer 2016; 59: 65-78.
[9] Yuan L, Liu LX and Che GW. The advancement of predictive diagnosis and molecular mechanism in multiple primary lung cancer. Chin J Cancer 2010; 29: 575-578.
[10] Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M and Edgar R. NCBI GEO: mining tens of millions of expression profiles--database and tools update. Nucleic Acids Res 2007; 35: D760-765.
[11] Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A and Chinnaiyan AM. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia 2004; 6: 1-6.
[12] Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC and Lempicki RA. DAVID: database for annotation, visualization, and integrated discovery. Genome Biol 2003; 4: P3.
[13] Hoffmann R and Valencia A. A gene network for navigating the literature. Nat Genet 2004; 36: 664.
[14] Treister A and Pico AR. Identifier mapping in cytoscape. F1000Res 2018; 7: 725.
[15] Hou GX, Liu P, Yang J and Wen S. Mining expression and prognosis of topoisomerase isoforms in non-small-cell lung cancer by using Oncomine and Kaplan-Meier plotter. PLoS One 2017; 12: e0174515.
[16] Selamat SA, Chung BS, Girard L, Zhang W, Zhang Y, Campan M, Siegmund KD, Koss MN, Hagen JA, Lam WL, Lam S, Gazdar AF and Laird-Offringa IA. Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. Genome Res 2012; 22: 1197-1211.

[17] Porter LA and Donoghue DJ. Cyclin B1 and CDK1: nuclear localization and upstream regulators. Prog Cell Cycle Res 2003; 5: 335-347.
[18] Takizawa CG and Morgan DO. Control of mitosis by changes in the subcellular location of cyclin-B1-Cdk1 and Cdc25C. Curr Opin Cell Biol 2000; 12: 658-665.
[19] Sartor H, Ehlert F, Grzeschik KH, Müller R and Adolph S. Assignment of two human cell cycle genes, CDC25C and CCNB1, to 5q31 and 5q12, respectively. Genomics 1992; 13: 911-912.
[20] Zhou L, Li J, Zhao YP, Cui QC, Zhou WX, Guo JC, You L, Wu WM and Zhang TP. The prognostic value of Cyclin B1 in pancreatic cancer. Med Oncol 2014; 31: 107.
[21] Kosacka M, Korzeniewska A and Jankowska R. The evaluation of prognostic value of cyclin B1 expression in patients with resected non-small-cell lung cancer stage I-IIIA--preliminary report. Pol Merkur Lekarski 2010; 28: 117-121.
[22] Soria JC, Jang SJ, Khuri FR, Hassan K, Liu D, Hong WK and Mao L. Overexpression of cyclin B1 in early-stage non-small cell lung cancer and its clinical implication. Cancer Res 2000; 60: 4000-4004.